

Indirekte Vergleiche von Therapieverfahren

Ben Schöttker, Dagmar Lühmann, Dalila Boulkhemair, Heiner Raspe



**Schriftenreihe
Health Technology Assessment (HTA)
in der Bundesrepublik Deutschland**

Indirekte Vergleiche von Therapieverfahren

**Ben Schöttker
Dagmar Lühmann
Dalila Boukhemair
Heiner Raspe**

Institut für Sozialmedizin, Universität zu Lübeck

Wir bitten um Beachtung

Dieser HTA-Bericht ist publiziert in der DAHTA-Datenbank des DIMDI (www.dimdi.de – HTA) und in der elektronischen Zeitschrift GMS Health Technology Assessment (www.egms.de).

Die HTA-Berichte des DIMDI durchlaufen ein unabhängiges, grundsätzlich anonymisiertes Gutachterverfahren. Potentielle Interessenkonflikte bezüglich der HTA-Berichte werden dem DIMDI von den Autoren und den Gutachtern offengelegt. Die durchgeführte Literaturrecherche erhebt keinen Anspruch auf Vollständigkeit. Die Literaturlauswahl erfolgt nach den Kriterien der evidenzbasierten Medizin. Die Verantwortung für den Inhalt des Berichts obliegt den jeweiligen Autoren.

Die Erstellung des vorliegenden HTA-Berichts des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI) erfolgte gemäß gesetzlichem Auftrag nach Artikel 19 des GKV-Gesundheitsreformgesetzes 2000. Das Thema stammt aus dem öffentlichen Vorschlagsverfahren beim DIMDI, durch das Kuratorium HTA priorisiert und vom DIMDI beauftragt. Der Bericht wurde mit Mitteln des Bundes finanziert.

Herausgeben vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI), Köln

Das DIMDI ist ein Institut im Geschäftsbereich des Bundesministeriums für Gesundheit (BMG)

Kontakt

DAHTA@DIMDI
Deutsche Agentur für Health Technology Assessment des
Deutschen Instituts für Medizinische Dokumentation und Information
Waisenhausgasse 36-38a
50676 Köln
Tel: +49 221 4724-525
Fax: +49 2214724-340
E-Mail: dahta@dimdi.de
www.dimdi.de

Schriftenreihe Health Technology Assessment, Bd. 88
ISSN: 1864-9645
1. Auflage 2009
DOI: 10.3205/hta000071L
URN: urn:nbn:de:0183-hta000071L8

Inhaltsverzeichnis

Tabellenverzeichnis	VI
Abbildungsverzeichnis	VII
Abkürzungsverzeichnis	VIII
Glossar	X
Zusammenfassung	1
Abstract	3
Kurzfassung	5
Summary	11
Hauptdokument	16
1 Gesundheitspolitischer Hintergrund	16
2 Wissenschaftlicher Hintergrund	17
2.1 Evidenzbasis für indirekte Vergleiche	17
2.1.1 Einsatzszenarien	17
2.1.2 Geometrie von Evidenznetzwerken	19
2.2 Die Validität von indirekten Vergleichen	20
2.3 Abgrenzung indirekter Vergleich – Cross design-Synthese	21
3 Forschungsfragen	22
4 Methodik	23
4.1 Literaturrecherche	23
4.1.1 Recherche in elektronischen Datenbanken	23
4.1.2 Handsuchen	23
4.2 Literaturselektion	24
4.2.1 Abstractselektion	24
4.2.2 Volltextselektion	24
4.3 Beschreibung der indirekt vergleichenden Methoden	26
4.4 Häufigkeit des Einsatzes der verschiedenen Methoden für indirekte Vergleiche	26
4.5 Methodik der Validitätsprüfung	26
4.5.1 Indirekte Vergleiche ohne Metaanalysen	26
4.5.2 Indirekte Vergleiche mit metaanalytischen Methoden	26
4.5.2.1 Datensätze für die Validitätsprüfung	26
4.5.2.2 Berechnung der Diskrepanzen zwischen direktem und indirektem Vergleich (z-Statistik)	27
4.5.2.3 Test auf systematische Über- oder Unterschätzung	27
4.5.2.4 Quantifizierung des Ausmaßes der Diskrepanz	28
4.5.2.5 Häufigkeit von statistisch signifikant diskrepanten Datensätzen	28
4.5.2.6 Subgruppenanalyse	28
4.6 Beurteilung der Präzision indirekt vergleichender Methoden	29
4.7 Überprüfung der Kongruenz in den Schlussfolgerungen von direktem und indirektem Vergleich	29
5 Ergebnisse	30
5.1 Ergebnisse der Literaturrecherchen und Literaturselektion	30
5.2 Allgemeine Annahmen und Voraussetzungen für indirekte Vergleiche	32
5.2.1 Heterogenität	32
5.2.2 Umgang mit Heterogenität in systematischen Reviews mit indirekten Vergleichen	34
5.3 Beschreibung der indirekt vergleichenden Methoden	35

5.3.1	Nicht-adjustierter indirekter Vergleich	35
5.3.2	Adjustierter indirekter Vergleich	38
5.3.3	Weiterentwicklung der adjustierten indirekten Vergleiche: Einschluss von Head-to-head- Studien	40
5.3.4	Metaregression	41
5.3.5	Weiterentwicklung der Metaregression: Aufnahme von Kovariaten in die Metaregression.....	43
5.3.6	Weiterentwicklung der Metaregression: Einschluss von kontrollierten Head-to-head-Studien (Mixed model)	44
5.3.7	Weiterentwicklung der Metaregression: Metaregression mit dem Bayes'schen Theorem	45
5.3.8	Mixed treatment comparison (MTC).....	45
5.3.9	Sonstige Methoden	50
5.3.10	Methoden ohne Metaanalysen.....	52
5.3.10.1	Indirekter Vergleich von Einzelstudien	52
5.3.10.2	Deskriptive, nicht-quantitative indirekte Vergleiche (narrativ).....	53
5.4	Relative Häufigkeit der indirekt vergleichenden Methoden	53
5.5	Validitätsprüfung indirekt vergleichender Methoden.....	55
5.5.1	Datensätze für die Validitätsprüfung	55
5.5.2	Methodenspezifische Validitätsprüfung	56
5.5.2.1	Nicht-adjustierter indirekter Vergleich	56
5.5.2.2	Adjustierter indirekter Vergleich	58
5.5.2.3	Metaregression.....	62
5.5.2.4	Mixed treatment comparison (MTC).....	62
5.5.2.5	Sonstige Methoden für indirekte Vergleiche	64
5.5.3	Subgruppenanalyse	65
5.5.3.1	Datensätze für die Subgruppenanalyse	65
5.5.3.2	Test auf systematische Über- oder Unterschätzung	66
5.5.3.3	Ausmaß der durchschnittlichen Diskrepanz.....	66
5.5.3.4	Anteil der statistisch signifikant diskrepanten Datensätze bei den verschiedenen Methoden.....	67
5.5.3.5	Ergebnisgegenüberstellung von Hauptanalyse und Subgruppenanalyse und Zusammenfassung der Validitätsprüfung	68
5.6	Präzision indirekt vergleichender Methoden.....	68
5.6.1	Für die Präzisionsbetrachtung herangezogene Datensätze	68
5.6.2	Theoretischer Ansatz	68
5.6.3	Empirischer Ansatz	69
5.6.4	Einfluss der Präzision indirekter Vergleiche auf die Validitätsprüfung.....	70
5.7	Kongruenz in den Schlussfolgerungen von direktem und indirektem Vergleich	70
6	Diskussion und Beantwortung der Forschungsfragen	72
6.1	Literatur für die Methodenbeschreibung und -bewertung.....	72
6.2	Validität von indirekten Vergleichen.....	72
6.2.1	Grundvoraussetzung: Systematischer Review	72
6.2.2	Formale Voraussetzungen für indirekte Vergleiche	73
6.2.3	Diskussion methodenspezifischer Ursachen für systematische Verzerrungen	75
6.3	Zu Forschungsfrage 1 - Methoden für indirekte Vergleiche	78
6.3.1	Welche Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen existieren bisher?	78
6.3.2	Wie können sie eingesetzt werden?	79
6.3.3	Unter welchen Voraussetzungen können sie eingesetzt werden?	79

6.4	Zu Forschungsfrage 2 - Häufigkeit der verschiedenen indirekt vergleichenden Methoden	80
6.5	Zu Forschungsfrage 3 - Validität indirekter Vergleiche.....	80
6.5.1	Kommen indirekter und direkter Vergleich zu identischen Schlussfolgerungen?	80
6.5.2	Wie ist die Validität indirekter Vergleiche gegenüber der von direkten Vergleichen zu beurteilen?	81
6.5.2.1	Validität nicht-adjustierter indirekter Vergleiche	81
6.5.2.2	Validität adjustierter indirekter Vergleiche	82
6.5.2.3	Validität von Metaregressionen für indirekte Vergleiche	83
6.5.2.4	Validität von Mixed treatment comparison (MTC)	83
6.5.2.5	Validität von sonstigen Methoden für indirekte Vergleiche	84
6.5.2.6	Validität von indirekten Vergleichen ohne Metaanalysen	84
6.5.3	Limitationen der Validitätsprüfung	85
6.6	Zu Forschungsfrage 4 - Welchen Einfluss hat es auf die Validität, wenn Head-to-head-Studien mit in den indirekten Vergleich aufgenommen werden?	86
6.7	Zu Forschungsfrage 5 - Lässt sich ein Goldstandard für indirekte Vergleiche identifizieren? ..	87
6.8	Status quo der Implementation von indirekten Vergleichen in Methodenpapiere von HTA-Institutionen	87
6.9	Ergänzungen zum aktuellen Forschungsstand zu indirekten Therapievergleichen durch diesen HTA-Bericht.....	88
6.10	Forschungsbedarf.....	89
7	Empfehlungen und Schlussfolgerung	91
7.1	Einleitung - Methoden, die für die Durchführung indirekter Vergleiche zur Verfügung stehen	91
7.2	Voraussetzungen für einen indirekten Vergleich	91
7.3	Empfehlungen zu den verschiedenen indirekt vergleichenden Methoden	91
7.4	Schlussfolgerung	92
8	Literaturverzeichnis	93
9	Anhang	108
9.1	Formeln für die Methodik der Validitätsprüfung.....	108
9.2	Kurze Einführung in die Bayes'sche Statistik	108
9.3	Durchsuchte Datenbanken	109
9.4	Suchstrategien	112
9.4.1	Suchstrategien in digitalen Datenbanken	112
9.4.2	Suchstrategie der Handsuchen	113
9.5	Ergebnisse der Handsuchen	114
9.6	Nach Durchsicht im Volltext ausgeschlossene Literatur mit Ausschlussgrund	118
9.7	Nach Durchsicht im Volltext eingeschlossene Literatur.....	127
9.8	Checklisten	132
9.8.1	Checkliste zur Beschreibung und Charakterisierung von Methoden zur Durchführung von indirekten Vergleichen	132
9.8.2	Auswertungsbogen für systematische Reviews mit Metaanalyse(n), die nur einen indirekten Vergleich durchführten	133
9.8.3	Auswertungsbogen für systematische Reviews mit Metaanalyse, die einen direkten und indirekten Vergleich durchführten	134
9.9	Ergebnistabellen und -abbildungen	135
9.9.1	Methodische Daten der systematischen Reviews mit indirekten Vergleichen.....	135
9.9.2	Heterogenitätsbetrachtung in systematischen Reviews mit indirekten Vergleichen....	138
9.9.3	Publikationen mit Mixed treatment comparison (MTC).....	148
9.9.4	Gegenüberstellung der Metaanalyseergebnisse des direkten und indirekten Vergleichs aus den Typ-5-Publikationen	151
9.9.5	Übersichtstabellen zu den Ergebnissen der Validitätsprüfung (Hauptanalyse, Typ-5-Publikationen)	211

Tabellenverzeichnis

Tabelle 1:	Möglichkeiten bei Inkonsistenz und/oder Heterogenität mit einem MTC zu verfahren	50
Tabelle 2:	Anzahl der Einsätze der verschiedenen Methoden zur Durchführung indirekter Vergleiche mit Metaanalyse(n) in systematischen Reviews, publiziert im Zeitraum von Januar 1999 bis Februar 2008	55
Tabelle 3:	Übersicht über die in die Haupt- und Subgruppenanalyse eingeschlossenen Datensätze	66
Tabelle 4:	Test auf signifikante Über- oder Unterschätzung des Ergebnisses des direkten Vergleichs durch den indirekten Vergleich (Subgruppenanalyse)	66
Tabelle 5:	Ausmaß der durchschnittlichen Diskrepanz zwischen direktem und indirektem Vergleich (Subgruppenanalyse)	67
Tabelle 6:	Anteil der statistisch signifikant diskrepanten Datensätze (Subgruppenanalyse)	67
Tabelle 7:	Anzahl der kongruenten Schlussfolgerungen aus direktem und indirektem Vergleich	70
Tabelle 8:	Über das DIMDI recherchierte elektronische Datenbanken	109
Tabelle 9:	Im ISI Web of Knowledge® enthaltene Datenbanken	110
Tabelle 10:	HTA-Institutionen (Internetseiten für die Handsuche)	110
Tabelle 11:	Cochrane Collaboration (Internetseiten für die Handsuche)	112
Tabelle 12:	Suchstrategie, DIMDI-Datenbanken	112
Tabelle 13:	Suchstrategie, ISI Web of Knowledge®	113
Tabelle 14:	Ergebnisse der Handsuchen auf den Internetseiten der HTA-Institutionen	114
Tabelle 15:	Ergebnisse der Handsuchen nach Methodenpapieren in der Cochrane Collaboration ..	115
Tabelle 16:	Details über die im Volltextscreening ein- bzw. ausgeschlossenen Publikationen bei der Suche in der Cochrane Collaboration	115
Tabelle 17:	Ergebnisse der Referenzsuche im Kapitel „Statistical methods for indirect comparisons“ des systematischen Reviews von Glenny et al. ¹⁰⁹ (S. 18-25)	116
Tabelle 18:	Übersicht über die weiteren Methodenpapiere, die aus den Referenzen anderer Reviews oder bereits gefundener Methodenpapiere exzerpiert wurden.	117
Tabelle 19:	Systematische Übersichtsarbeiten, denen Glenny et al. ¹⁰⁹ Metaanalysen entnahmen, um mit ihnen sowohl direkte als auch indirekte Vergleiche durchzuführen	117
Tabelle 20:	Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen	118
Tabelle 21:	Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs	127
Tabelle 22:	Methodische Daten systematischer Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen); geordnet nach der verwendeten Methode	135
Tabelle 23:	Verwendete Verfahren zur Ermittlung von Heterogenität in systematischen Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen)	138
Tabelle 24:	Umgang mit vorliegender Heterogenität in den Typ-5-Publikationen	141
Tabelle 25:	Ausmaß der statistischen Heterogenität in Reviews (Typ-5-Publikationen) mit signifikanter Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs	144
Tabelle 26:	Zusätzlich abgefragte Charakteristika in Typ-4- und Typ-5-Publikationen mit einem MTC	148
Tabelle 27:	Erhebung und Umgang mit Heterogenität und Inkonsistenz in Typ-4- und Typ-5-Publikationen mit MTC	149
Tabelle 28:	Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen	151
Tabelle 29:	Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen	165
Tabelle 30:	Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen	180
Tabelle 31:	Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen	191
Tabelle 32:	Test auf signifikante Über- oder Unterschätzung des Ergebnisses des direkten Vergleichs durch den indirekten Vergleich (Hauptanalyse)	211
Tabelle 33:	Ausmaß der durchschnittlichen Diskrepanz zwischen direktem und indirektem Vergleich (Hauptanalyse)	211
Tabelle 34:	Anteil der statistisch signifikant diskrepanten Datensätze (Hauptanalyse)	211

Abbildungsverzeichnis

Abbildung 1: Einsatzszenarien für direkte und indirekte Vergleiche	17
Abbildung 2: Extremata von geometrischen Strukturen in Evidenznetzwerken.....	19
Abbildung 3: Ergebnisse der Literaturrecherchen und Handsuchen	31
Abbildung 4: Verwendungszwecke der gesamten eingeschlossenen Literatur.....	32
Abbildung 5: Wahlmöglichkeiten bei heterogenen Therapieeffekten.....	35
Abbildung 6: Nicht-adjustierter indirekter Vergleich	36
Abbildung 7: Adjustierter indirekter Vergleich	38
Abbildung 8: Adjustierter indirekter Vergleich mit Einschluss von kontrollierten Head-to-head-Studien	41
Abbildung 9: Indirekter Vergleich durch Metaregression	43
Abbildung 10: Beispiel für zwölf Studien mit Indizien zur Wirksamkeit der Interventionen A und B.....	46
Abbildung 11: Darstellung des Beispiels als Evidenznetzwerk	46
Abbildung 12: Evidenznetzwerk aus drei Therapieoptionen	49
Abbildung 13: Indirekter Vergleich nach Moore et al. und Hind et al.	52
Abbildung 14: Anzahl der Einsätze der verschiedenen Methoden zur Durchführung indirekter Vergleiche mit Metaanalyse(n) in systematischen Reviews, publiziert im Zeitraum von Januar 1999 bis Februar 2008	53
Abbildung 15: Häufigkeit von Publikationen mit indirekten Vergleichen von 2000 bis 2007.....	54
Abbildung 16: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von nicht-adjustierten indirekten Verfahren.....	57
Abbildung 17: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und dichotomen Daten_Teil 1	59
Abbildung 18: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und dichotomen Daten_Teil 2	60
Abbildung 19: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und kontinuierlichen Daten	61
Abbildung 20: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von Metaregressionen für den indirekten Vergleich.....	62
Abbildung 21: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von MTC und dichotomen Daten.....	63
Abbildung 22: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von MTC und kontinuierlichen Daten	64
Abbildung 23: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von sonstigen Methoden (in gefundenen Beispielen wurden nur dichotome Daten benutzt).....	65
Abbildung 24: Häufigkeitsverteilung der prozentualen Zu- bzw. Abnahme der Konfidenzintervallweite im indirekten Vergleich im Verhältnis zum direkten Vergleich in einer Stichprobe von 198 Datensätzen mit direktem und indirektem Vergleich.....	69

Abkürzungsverzeichnis

ACE	Angiotensin converting enzyme
ACR	Krankheitsaktivitätsscore des American College of Rheumatology
AHRQ	Agency for Healthcare Research and Quality
ANOVA	Analysis of variance
ARR	Absolute Risikoreduktion
BMS	Arzneistofffreier Metallstent; engl.: Bare metal stent
Chi ² ; X ²	Ergebnis der Chi ² -Verteilung im Rahmen eines Chi ² -Tests
CI; KI	Confidence interval; dt.: Konfidenzintervall
Cochrane Collaboration	Cochrane Collaboration of Systematic Reviews
CONSORT	Consolidated standards for reporting of trials
COX	Cyclooxygenase
DAHTA	Deutsche Agentur für Health Technology Assessment
df	Anzahl Freiheitsgrade
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
FU	Fluorouracil + Leucovorin
GPS	Allgemeine pathopsychologische Skala, engl.: General psychopathology scale
GRADE	Grades of Recommendation Assessment, Development and Evaluation
HbA1c	Hämoglobin A1c
H ₂ RA	Histamin-2-Rezeptorantagonist
HAART	Hochaktive antiretrovirale Therapie
HIV	Humanes Immundefizienzvirus
HR	Hazard ratio
HTA	Health Technology Assessment
ICD	Implantierbarer Cardioverter/Defibrillator
INAHTA	International network of agencies for Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI; CI	Konfidenzintervall; engl.: Confidence interval
MeSH	Medical subject heading
MTC	Mixed treatment comparison
n	Stichprobenumfang
n. b.	Nicht berichtet
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NLB	Nationalbibliothek für Medizin

Fortsetzung: Abkürzungsverzeichnis

NNT	Number-needed-to-treat
NRTI	Nukleosidischer Reverse-Transkriptase-Inhibitor
NNRTI	Nicht-nukleosidischer Reverse-Transkriptase-Inhibitor
NSAR	Nicht-steroidale Antirheumatika
OR	Odds ratio
PANSS	Positiv- und Negativsyndromskala
PCI	Perkutane koronare Intervention; engl.: Percutaneous coronary intervention
PES	Paclitaxel freisetzender Stent
PPI	Protonenpumpeninhibitor
QUORUM	Quality of reporting of Metaanalyses
RCT	Randomisierte kontrollierte Studie; engl.: Randomized controlled trial
RD	Risikodifferenz
ResRR	Ratio der Ansprechraten, engl.: Response rate ratio
RR	Relatives Risiko, engl.: Risk ratio
SAS [®]	Statistical Analysis System (Statistiksoftware)
SE	Standardfehler; engl.: Standard error
SES	Sirolimus freisetzender Stent; engl.: Sirolimus eluting stent
SGB	Sozialgesetzbuch
SMD	Standardisierte mittlere Differenz; engl.: Standardised mean difference
SPID %	Summe der Schmerzintensitätsdifferenzen; engl.: Sum of pain intensity difference
SPSS [®]	Statistical package for the social sciences (Statistiksoftware)
TOTPAR %	Gesamtschmerzreduktion; engl.: Total pain relief
UK	Vereinigtes Königreich; engl.: United Kingdom
USA	Vereinigte Staaten von Amerika, engl.: United States of America
WHO	Weltgesundheitsorganisation; engl.: World health organisation
WinBUGS [®]	Windows Bayesian Inference Using Gibbs Sampling (Statistiksoftware)
WMD	Gewichtete mittlere Differenz; engl.: Weighted mean difference
95 %-KI	95 %-Konfidenzintervall

Glossar

Absolute Risikoreduktion (ARR). Synonym: Risikodifferenz	Effektmaß für dichotome Endpunkte; beschreibt die absolute Differenz der Rate an Zielereignissen in der experimentellen Gruppe (E) im Vergleich zur Kontrollgruppe (K) ($ARR = K - E$) ¹⁰ .
Abstract	Kurzzusammenfassung einer Studie.
Adjustierter indirekter Vergleich	Ein adjustierter indirekter Vergleich berücksichtigt bei der metaanalytischen Zusammenfassung der Ergebnisse von randomisierten kontrollierten Studien die in den Kontrollgruppen erzielten Effekte, indem er diese von den Therapieeffekten der Verumgruppe(n) subtrahiert ¹⁰⁴ .
Adjustierung	Statistisches Verfahren zur Kontrolle des Einflusses wichtiger Störgrößen auf das Studienergebnis (z. B. Alter ⁴).
Bayes'sche Verfahren Bayes'scher Ansatz	Ein statistischer Ansatz, der auf der Anwendung von Bayes' Theorem basiert. Eine Bayes'sche Analyse nutzt das Bayes-Theorem, um eine Priorverteilung einer unbekanntem Schätzung (z. B. einer Odds ratio) durch Verarbeitung von Ergebnissen einer oder mehrerer Studien in eine Posteriorverteilung zu überführen. Die Priorverteilung kann dabei auf externer Evidenz, allgemein gültigen Annahmen oder subjektiver Meinung beruhen. Statistische Schlussfolgerungen werden auf Basis von Informationen der Posteriorverteilung getroffen und können als Punktschätzer mit Credibility intervals (Bayes'sches Äquivalent zu Konfidenzintervallen) berichtet werden ⁴ .
Beobachtungsstudie (Englisch: Observational study)	Eine Studie, in der Forscher nicht intervenieren, sondern nur Expositionen und Zielgrößen erfassen. Veränderungen oder Unterschiede in einer Exposition werden in Relation zu Veränderungen oder Unterschieden in einer oder mehreren Zielgrößen gesehen ⁴ .
Betrag (des z-Wertes)	Abstand einer reellen Zahl von Null. Der Betrag eines Wertes ist immer positiv.
Bias (systematischer Fehler)	Tendenz der Studienergebnisse, systematisch von den wahren Ergebnissen abzuweichen. Bias führt entweder zu einer Über- oder Unterschätzung der wahren Wirkung einer Maßnahme oder Exposition. Die Ursachen dafür liegen vor allem im Design und der Durchführung der Studie und führen zu systematischen Unterschieden zwischen den Vergleichsgruppen, z. B. bei der Auswahl der Teilnehmer (Selektionsbias), der Erhebung der Endpunkte (Messungsbias) oder dem Verlust von Teilnehmern in der Studie (attrition bias oder Verschleiß-Bias). Ergebnisse aus Studien mit geringem Risiko für Bias werden als valide angesehen ¹⁰ .
Binäre Daten	Synonym zu dichotome Daten ⁴ .
Cochrane Methodology Group	Arbeitsgruppe der Organisation Cochrane Collaboration, die sich mit Methoden zur Erstellung systematischer Reviews befasst.

Fortsetzung: Glossar

Confidence profile-Methode	Von David M. Eddy entwickelte Bayes'sche Methode für indirekte Therapievergleiche. Sie kann neben kontrollierten Studien auch Fall-Kontroll- und andere Beobachtungsstudien einschließen und bezüglich ihrer Biasquellen adjustieren ⁹³ .
Confounder/ Confounding	Faktor, der nicht direkt Gegenstand einer Untersuchung ist, aber sowohl mit der Intervention/Exposition als auch mit der Zielgröße assoziiert ist und dadurch bei Aussagen über die Beziehung zwischen Intervention/Exposition und Zielgröße störende Einflüsse ausüben kann. Häufige Confounder sind z. B. Alter, Geschlecht oder Nikotingenuss. Confounding lässt sich durch ein entsprechendes Studiendesign (z. B. Randomisation oder Matching) oder durch die Anwendung bestimmter statistischer Verfahren bei der Analyse (Stratifizierung, multivariate Analyse) kontrollieren ¹⁰ .
Kovariate	Kovariate sind in einem statistischen Modell (z. B. einem Regressionsmodell) die Variable(n), die neben der zu analysierenden Exposition die Zielgrößen beeinflussen und daher bei der Auswertung zu berücksichtigen sind ¹⁴ .
Credibility interval (Vertrauensbereich)	Analogon in der Bayes'schen Statistik zu einem Konfidenzintervall der konventionellen Statistik ²⁰³ . Interpretiert als Wahrscheinlichkeit, mit der der wahre Wert in diesem Wertebereich liegt.
Cross design-Synthese	Methode zur Zusammenfassung von Studienergebnissen aus Studien unterschiedlicher Designs. Im Zusammenhang mit Aussagen zur Wirksamkeit von Behandlungsmethoden ist insbesondere für potenzielle Biasquellen von nicht-experimentellen Studientypen zu adjustieren ⁹⁰ .
Cut-off-Wert	Schwellenwert, Grenzwert.
Dichotome Daten	Merkmal, bei dem nur zwei Ausprägungen möglich sind ¹⁴ .
Direkter Vergleich	Wirksamkeitsvergleich von zwei oder mehr Interventionen auf der Basis von einer oder mehrerer Head-to-head-Studien ¹⁰⁹ .
Diskrepanz	Abweichung. In diesem Health Technology Assessment als Begriff benutzt, der die Differenz der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich beschreibt.
Diskrete Variablen	Diskrete Variablen weisen im Gegensatz zu kontinuierlichen Variablen nur eine begrenzte Zahl eindeutig voneinander abgrenzbarer Zustände auf (z. B. Augenfarbe: blau, grau, braun, grün). Eine Sonderform sind dichotome (Synonym: binäre) Variablen, die lediglich zwei Alternativen aufweisen, z. B. Raucher/Nichtraucher, lebend oder tot, Test-positiv oder Test-negativ ¹⁰ .
(Gesamt-) Effektschätzer, Effekt	Die beobachtete Beziehung zwischen einer Intervention und einer Zielgröße, darstellbar in unterschiedlichen Effektmaßen (Odds ratio, relatives Risiko, gewichtete mittlere Differenz, ...).
Effektmaß	Maßzahl, um die Stärke eines Effekts zu quantifizieren. Gebräuchliche Effektmaße für dichotome Endpunkte sind das relative Risiko oder die Odds ratio, gebräuchliche Effektmaße für kontinuierliche Endpunkte sind in Einzelstudien die standardisierte mittlere Differenz (SMD) und in Metaanalysen die gewichtete mittlere Differenz ¹⁰ .

Fortsetzung: Glossar

Empirie/empirisch	Empirie von griechisch <i>empireia</i> : Erfahrung, Erfahrungswissen. Empirische Daten werden erhoben, um theoretische Annahmen über die Welt zu widerlegen oder zu bestätigen ²²⁵ .
Evidenz	Informationen, auf denen eine Entscheidung oder eine Empfehlung basiert. Evidenz kann aus einer Reihe von Quellen gewonnen werden, einschließlich randomisierten kontrollierten klinischen Studien, Beobachtungsstudien und Expertenmeinungen ¹⁵ .
Evidenzbasierte Medizin	Unter evidenzbasierter Medizin im engeren Sinn versteht man eine Vorgehensweise des medizinischen Handelns, individuelle Patienten auf der Basis der besten zur Verfügung stehenden Daten zu versorgen ¹⁰ .
Evidenzhierarchie	Bildung einer Hierarchie unter Analysemethoden, mit derjenigen, die die höchste Validität aufweist (am wenigsten anfällig für die Generierung systematisch verzerrter Ergebnisse ist), an der Spitze.
Evidenznetzwerk	Grafische Darstellung der Vergleichsmöglichkeiten von zwei oder mehr therapeutischen Interventionen (z. B. für einen Mixed treatment comparison (MTC)). Ein vollständiges Evidenznetzwerk beinhaltet alle randomisierten kontrollierten Studien, die die Wirksamkeit der zu vergleichenden Interventionen geprüft haben, unabhängig davon, welche Vergleichsinterventionen ihnen in diesen Studien gegenübergestellt wurden. Die grafische Darstellung ähnelt einem Spinnennetz, in dem die Knotenpunkte die einzelnen Interventionen oder Komparatoren darstellen und die Fäden Studien, in denen direkte Vergleiche vorgenommen wurden ¹⁶⁰ .
Fixed effects-Modell	Statistisches Modell zur Zusammenfassung von Ergebnissen einzelner Studien, wobei angenommen wird, dass alle Studien den gleichen Effekt schätzen und Unterschiede nur durch zufällige Abweichung bedingt sind. Somit ist die Ungenauigkeit des Gesamteffekts nur durch die Variation innerhalb der einzelnen Studien beeinflusst. Beispiele sind die Inverse-Varianz-Methode, das Peto-Modell und die Mantel-Haenszel Odds ratio ¹⁰ .
Forest-Plot	Eine gebräuchliche Art, Ergebnisse eines systematischen Reviews grafisch zu präsentieren. Die Schätzungen der Therapieeffekte, gemeinsam mit ihren Konfidenzintervallen, werden relativ zu einer vertikalen Linie eingezeichnet, die indiziert, dass kein Therapieeffektunterschied zwischen der Intervention und der Kontrolle in der eingeschlossenen Studie besteht. Aus dieser Ansicht kann ein Bild der Verteilung der Schätzer des Effekts in allen eingeschlossenen Studien gewonnen werden ¹⁵ .
Generalisierbarkeit	Das Ausmaß indem die Resultate einer in einer definierten Patientenpopulation und/oder einem spezifischen Kontext durchgeführten Studie auf eine andere Population und/oder einen anderen Kontext übertragbar sind ¹⁵ .
Gewichtete mittlere Differenz (WMD)	Zusammenfassendes Effektmaß für kontinuierliche Endpunkte. Dabei erhalten die Einzelstudien ein unterschiedliches Gewicht, um wichtige Studieninformationen, wie z. B. Größe der Studie oder Präzision des Effekts, zu berücksichtigen ¹⁰ .

Fortsetzung: Glossar

Goldstandard	Als Goldstandard wird das Verfahren, das die derzeit beste zur Verfügung stehende Methode beschreibt und an dem neue bzw. alternative Methoden gemessen werden, bezeichnet ¹⁰ .
Handsuche	Geplante Suche nach Publikationen zu einer spezifischen Thematik, in ausgewählten Zeitschriften oder Zitationslisten von Publikationen Seite für Seite (z. B. per Hand). Kann ergänzend zu einer elektronischen Datenbankrecherche stattfinden, um alle relevanten Publikationen zu finden.
Head-to-head-Studie	Studien, die z. B. die Wirksamkeit oder den Nutzen zweier oder mehrerer Interventionen in strukturgleichen Studienarmen mit der gleichen Methodik untersuchen ¹⁶⁷ .
Health Technology Assessment-Bericht (HTA-Bericht)	Sie treffen Aussagen zu Nutzen, Risiko, Kosten sowie ethischen, sozialen, juristischen und organisatorischen Auswirkungen medizinischer Verfahren und Technologien mit Bezug zur gesundheitlichen Versorgung der Bevölkerung. Unter Technologie fallen dabei neben Medikamenten auch Instrumente, Geräte, Prozeduren, Verfahren und Organisationsstrukturen ⁹ .
Heterogenität bzw. Homogenität	In systematischen Reviews oder Metaanalysen bezeichnet Homogenität (Heterogenität), inwieweit die in den eingeschlossenen Studien gefundenen Effekte ähnlich (homogen) oder verschieden (heterogen) sind. Mit statistischen Heterogenitätstests kann festgestellt werden, ob die Unterschiede zwischen den Studien größer sind, als zufallsbedingt zu erwarten wäre. Als Ursachen für Heterogenität kommen Unterschiede in den Patientencharakteristika, Interventionen oder Endpunkte zwischen den Studien in Frage, was aus klinischer Sicht beurteilt werden muss. Bei der Durchführung von Metaanalysen wird Heterogenität unterschieden, die in den Einzelstudien (in-study heterogeneity) und zwischen den Einzelstudien vorliegt (between-study heterogeneity). Die Durchführung einer Metaanalyse mit heterogenen Studien ist problematisch ¹⁰ .
Heterogenitätstest	Siehe Heterogenität.
Hypothese	Aussage, Behauptung, oder Theorie, deren Zutreffen oder Nicht-Zutreffen mithilfe einer Studie oder eines Experimentes geprüft werden soll ¹⁰ .
I ² -Test	Ein statistischer Test, der benutzt wird, um Heterogenität zu quantifizieren. Er beschreibt den prozentualen Anteil der Variabilität in den Effektschätzern, die auf Heterogenität und nicht mehr auf Zufallsfehler zurückzuführen ist. Ein Wert größer 50 % kann als substantielle Heterogenität angesehen werden ⁴ .
Indirekter Vergleich	Vergleich der Effekte verschiedener Interventionen mithilfe eines gemeinsamen Komparators. Dabei kann es sich um den Vergleich mit einem aktiven Komparator (meistens Standardtherapie), oder um Vergleiche mit Placebo handeln ⁵⁵ .
Individuelle Patientendaten	In einer Metaanalyse die Verfügbarkeit von Rohdaten von jedem Studienteilnehmer in jeder eingeschlossenen Studie. Das Gegenteil zu aggregierten Daten (summierte Daten für den Vergleich von Gruppen in jeder Studie) ⁴ .

Fortsetzung: Glossar

Inkonsistenz	Im Kontext des Mixed treatment comparison (MTC) ist Inkonsistenz definiert als die Diskrepanz zwischen dem Ergebnis aus einem direkten und einem oder mehreren indirekten Vergleichen, die nicht mehr durch Zufallsfehler oder Heterogenität erklärbar ist ¹⁵⁹ .
Kolinearität Synonym: Multilinearität	Problematik bei Regressionsanalysen, die zu Verzerrungen führen kann. Kovariate, die einzeln nur gering zur Heterogenität beitragen, führen erst durch ihr Zusammenwirken zu statistisch signifikanter Heterogenität ^{14, 40, 41, 178} .
(Gemeinsamer) Komparator	Als Komparator wird die Vergleichsintervention in einer Studie bezeichnet, gegen deren Wirksamkeit die Wirksamkeit der zu prüfenden Intervention getestet wird. Über einen gemeinsamen Komparator in verschiedenen Studien kann indirekt ein Wirksamkeitsvergleich von gegen den gemeinsamen Komparator getesteten Interventionen vorgenommen werden ¹⁰⁴ .
Konfidenzintervall (KI)	Bereich, in dem der wahre Wert einer Messung (Effektgröße) mit einer bestimmten Sicherheit erwartet werden kann (üblicherweise 95 %-Konfidenzintervall). Das Konfidenzintervall beschreibt die Unsicherheit über die Zuverlässigkeit der Aussage zur Effektgröße. Die Breite des Konfidenzintervalls hängt u. a. von der Zahl der in die Studie eingeschlossenen Patienten ab und wird mit zunehmender Patientenzahl enger, d. h. die Effektgröße kann präziser geschätzt werden ¹⁰ .
Kongruenz	Übereinstimmung.
Konsistenz	Siehe Inkonsistenz.
Konsistenzan- nahme Englisch: Transitivity assumption	Grundannahme des Mixed treatment comparison (MTC) ist die Konsistenzannahme. Sie besagt, dass keine (Fixed effects-Modell) oder eine um den Wert Null normalverteilte (Random effects-Modell) Inkonsistenz in dem zu analysierenden Evidenznetzwerk vorliegt ¹³⁴ . In anderen Worten bedeutet dies, dass sie davon ausgeht, dass die direkten und indirekten Vergleiche im Evidenznetzwerk die gleichen Effekte messen ¹⁵⁹ .
Kontinuierliche Variablen	Im Gegensatz zu diskreten Variablen können kontinuierliche Variablen theoretisch eine unendlich große Zahl von Werten entlang eines Kontinuums annehmen. Körpergröße, Gewicht und viele Laborwerte sind kontinuierliche Variablen ¹⁰ .
Kontrollgruppe	Gruppe von Patienten in einer klinischen Studie, die einen explizit definierten Komparator erhält, mit dem die Effekte einer Intervention verglichen werden ¹⁵ .
Korrelation	Die Korrelation misst den Zusammenhang zwischen zwei quantitativen Merkmalen. Eine Maßzahl für die Stärke der Korrelation ist der Korrelationskoeffizient ¹¹ .

Fortsetzung: Glossar

Likelihood-Funktion	Funktion, die angibt, welche(r) geschätzte(n) Parameter bei gegebenen Daten die größte Wahrscheinlichkeit aufweist, dem wahren Parameter in der Grundgesamtheit zu entsprechen. Die Likelihood-Funktion kann aber aus formalen Gründen nicht als Wahrscheinlichkeitsfunktion aufgefasst werden, weshalb es auch sinnvoll ist, in der deutschen Sprache bei dem englischen Namen zu bleiben, um Verwechslungen vorzubeugen ¹⁴ .
Lineares Netzwerk	Begriff aus dem Mixed treatment comparison (MTC): Wenn neue Therapien nur gegen die jeweils vorher entwickelte getestet werden, und ältere Therapien in Studien nicht mehr berücksichtigt werden, entsteht im Evidenznetzwerk eine lineare Struktur ²⁰⁴ .
Logistische (Meta-)Regression	Eine Form der Regressionsanalyse, die ein individuelles Odds einer Erkrankung oder einem Therapieeffekt als eine Funktion von einem Risikofaktor oder einer Intervention modelliert. Sie wird weit verbreitet für dichotome Therapieeffekte eingesetzt ⁴ .
Logit-Funktion	Einsatz im Rahmen der logistischen Regression, da Odds zwar Werte > 1 annehmen können, doch ihr Wertebereich nach unten beschränkt ist (er nähert sich asymptotisch Null an). Eine unbeschränkter Wertebereich wird durch die Transformation der Odds in die sogenannten Logits erzielt; diese können Werte zwischen minus und plus unendlich annehmen ¹⁴ .
Markov-Ketten Monte-Carlo-Simulation	Methode, um in der Bayes'schen Statistik eine Posteriorverteilung zu berechnen, ohne bekannte Werte in der Priorverteilung anzunehmen ²⁰³ .
Median	Der Median ist der mittlere Wert eines (der Größe nach geordneten) Datenbündels, d. h. der Wert, unter- und oberhalb dessen jeweils die Hälfte der Datenwerte liegt. Der Median ist identisch mit dem 50 %-Quantil. Er ist weniger gegen einzelne stark abweichende Werte anfällig als das Arithmetische Mittel und daher oft besser als dieses geeignet, die zentrale Tendenz eines Datenbündels zu charakterisieren ¹⁴ .
MeSH-Begriff	MeSH (= Medical subject heading) ist das von der amerikanischen Nationalbibliothek für Medizin (NLB) herausgegebene kontrollierte Vokabular, das eingesetzt wird, um Artikel in den elektronischen Literaturdatenbanken der NLB (z. B. MEDLINE) zu indexieren ¹⁶ .
Metaanalyse	Statistisches Verfahren, um die Ergebnisse mehrerer Studien, die die gleiche Frage bearbeiten, quantitativ zu einem Gesamtergebnis zusammenzufassen (= Poolen) und dadurch die Aussagekraft (Genauigkeit der Effektschätzer) gegenüber Einzelstudien zu erhöhen. Metaanalysen werden mit zunehmender Häufigkeit in systematischen Reviews eingesetzt. Allerdings beruht nicht jede Metaanalyse auf einem systematischen Review ¹⁰ .
Metaregression	Eine Technik, die benutzt wird, um die Beziehung zwischen Studiencharakteristika und -ergebnissen in einem systematischen Review aufzuklären ⁴ .
Mixed model	Metaregressionstechnik für den indirekten Vergleich, die Head-to-head-Studien einzuschließen vermag, die einen dritten Arm mit dem gemeinsamen Komparator aufweisen ¹⁰⁹ .

Fortsetzung: Glossar

Mixed treatment comparison (MTC)	Eine Analyse, die zwei oder mehr Interventionen miteinander vergleicht und dabei Evidenz aus direkten (aus Head-to-head-Studien der Interventionen von Interesse) und indirekten Vergleichen (aus Studien mit unterschiedlichen Komparatoren) kombiniert ¹⁵ .
Multiparameter-synthese	Ergänzungen der Standardmetaanalysen durch komplexe Syntheseformen, wie z. B. einem Mixed treatment comparison (MTC), Kombinationen von Evidenz aus multiplen oder surrogaten Endpunkten, Kombination von Informationen aus verschiedenen Studiendesigns und Methoden, um Informationen über individuelle Parameter mit Informationen von komplexen Funktionen zu Parametern zu verbinden ²² .
Narrativ	Beschreibend, deskriptiv.
Nicht-adjustierter indirekter Vergleich	Ein nicht-adjustierter indirekter Vergleich berücksichtigt im Rahmen einer Metaanalyse die in den Kontrollgruppen der einzelnen randomisierten kontrollierten Studien (RCT) beobachteten Ergebnisse nicht und fasst die in den Verumgruppen der Einzelstudien beobachteten Effekte so zusammen, als seien es Arme einer einzigen Studie ^{104, 109} .
Normalverteilung	Verteilung einer stetigen Variablen mit Erwartungswert μ und Varianz σ^2 . Die Normalverteilung ist eine symmetrische und glockenförmige Verteilung mit einem Maximum an der Stelle μ . Die Form wird durch den Wert von σ^2 bestimmt ¹⁴ .
Nullhypothese, Alternativhypothese	Beim statistischen Test besagt die Nullhypothese, dass ein postulierter Zusammenhang oder Unterschied nicht besteht. Es wird eine Teststatistik berechnet, die angibt, ob ein in den Daten beobachteter Zusammenhang oder Unterschied mit der Nullhypothese kompatibel ist. Überschreitet die Teststatistik einen bestimmten, vorab festzulegenden Wert, so wird die Nullhypothese verworfen und die Forschungshypothese, die Alternativhypothese, gilt als vorläufig nicht widerlegt ¹⁴ .
Number-needed-to-treat (NNT)	Klinisch intuitives Effektmaß für dichotome Endpunkte, um die Auswirkung einer Behandlung zu beschreiben. Gibt die Anzahl an Patienten an, die behandelt werden müssten, um ein zusätzliches ungünstiges Ereignis zu verhindern. Die NNT wird als $1/\text{Absolute Risikoreduktion}$ berechnet ¹⁰ .
Nutzen	Unter dem therapeutischen Nutzen einer Intervention wird das Ausmaß des Profitierens des Patienten von einer therapeutischen Intervention verstanden.
Odds	Beschreibt in einer Gruppe das Verhältnis zwischen der Anzahl von Teilnehmern mit einem Endpunkt und der Anzahl von Teilnehmern ohne diesen Endpunkt. Wenn also 30 von 100 Teilnehmern den Endpunkt entwickeln (und 70 nicht), beträgt die Odds 30:70 oder 0,42 ¹⁰ .
Odds ratio (OR)	Effektmaß für dichotome Daten. Bezeichnet das Verhältnis (Ratio) der Odds, das ein Ereignis oder Endpunkt in der experimentellen Gruppe eintritt, zu den Odds, dass das Ereignis in der Kontrollgruppe eintritt. Ein OR von 1 bedeutet, dass zwischen den Vergleichsgruppen kein Unterschied besteht. Bei ungünstigen Endpunkten zeigt ein $OR < 1$, dass die experimentelle Intervention wirksam ist, um die Odds für das Auftreten dieser ungünstigen Endpunkte zu senken ¹⁰ .

Fortsetzung: Glossar

Outcome (Ergebnis, Endzustand)	Übergeordneter Begriff für die Auswirkungen von therapeutischen oder präventiven Maßnahmen und anderen Expositionen ¹⁰ .
Placebo	In einer pragmatischen Definition ein meist zu Studienzwecken eingesetztes Schein-Medikament ohne pharmakologisch aktive Substanz. Das Placebo darf hinsichtlich seiner äußeren Eigenschaften nicht von der aktiven Behandlung (dem Verumpräparat) unterscheidbar sein, wenn es seinen Zweck erfüllen soll ¹⁰ .
Polygonales Netzwerk	Begriff aus Mixed treatment comparison: Wenn alle zu vergleichenden Therapieoptionen in Head-to-head-Studien gegeneinander getestet wurden, entsteht ein Evidenznetzwerk mit polygonaler Struktur. Solche Strukturen werden vor allem in Indikationsfeldern mit vielen Therapieoptionen und hoher Forschungsintensität beobachtet. ²⁰⁴
Poolen	Siehe Metaanalyse.
Posteriorverteilung	Das Ergebnis einer Bayes'schen statistischen Analyse. Die Posteriorverteilung ist eine Wahrscheinlichkeitsverteilung, die beschreibt, wie wahrscheinlich verschiedene Werte eines Outcomeparameters (z. B. Therapieeffekt) sind. Sie berücksichtigt die Annahmen vor der Studie (Priorverteilung) und die durch die Studie erhaltenen Daten ⁴ .
Präzision	Genauigkeit, mit der ein Effekt geschätzt wird. Bedingt durch einen möglichst geringen Zufallsfehler und widergespiegelt durch ein enges Konfidenzintervall ¹⁰ .
Prior (-Verteilung)	Gibt die A-priori-Annahmen des Forschers über den wahren Wert der Zielgröße wieder ²⁰³ .
Publikationsbias	Systematischer Fehler aufgrund einer selektiven Publikationspraxis, bei der Studien mit positiven und signifikanten Ergebnissen eine größere Chance haben, publiziert zu werden als Studien mit negativen und nicht-signifikanten Resultaten. Ein systematischer Review oder eine Metaanalyse, die sich ausschließlich auf publizierte Studien stützt, läuft Gefahr, den Effekt der untersuchten Intervention zu überschätzen ¹⁰ .
p-Wert	p-Werte (p von probability) beschreiben die Wahrscheinlichkeit, dass der beobachtete (oder ein noch extremerer) Effekt in einer Studie aufgetreten sein könnte, wenn die Nullhypothese richtig und der Effekt auf den Zufall zurückzuführen ist. Je kleiner der Wert, desto deutlicher spricht das beobachtete Ergebnis gegen die Nullhypothese. Es ist eine Konvention, dass ein p-Wert gleich oder kleiner 0.05 als statistisch signifikant angesehen wird. Wenn die Signifikanz von Effekten interpretiert wird, sollten p-Werte immer im Zusammenhang mit Konfidenzintervallen verwendet werden ¹⁰ .
Quantil	Quantile teilen ein der Größe nach geordnetes Datenbündel in Teile. Das 25 %-Quantil gibt den Wert an, der das untere Viertel der Datenwerte von den oberen drei Vierteln trennt usw. Das 50 %-Quantil ist der Median.

Fortsetzung: Glossar

Random effects-Modell	Statistisches Modell zur Berechnung zusammengefasster (gepoolter) Ergebnisse, bei denen im Gegensatz zum Fixed effects-Modell Effektunterschiede zwischen verschiedenen Studien berücksichtigt werden. In die Genauigkeit der Schätzung des gemeinsamen Effekts geht daher nicht nur die Variation ein, die innerhalb der Studien beobachtet wird, sondern auch die Variation zwischen den Studien. Ein Beispiel ist die Methode nach DerSimonian & Laird ¹⁰ .
Randomisation	Zufällige Verteilung der Patienten in einer Studie zu zwei oder mehr alternativen Gruppen durch ein Zufallsverfahren, wie z. B. (computer-generierte) Zufallszahlen. Damit soll sicher gestellt werden, dass alle Teilnehmer die gleiche Chance haben, der einen oder anderen Gruppe zugeordnet zu werden und es wahrscheinlich ist, dass sich (bei ausreichender Studiengröße) bekannte wie unbekannte Risiko- und Prognosefaktoren ausgeglichen auf die beiden Gruppen verteilen und dadurch systematische Fehler (Bias) und Confounding reduziert wird ^{10, 15} .
Randomisierte kontrollierte Studie (RCT)	Eine vergleichende Studie, in die Probanden zufällig auf Interventions- sowie Kontrollgruppe(n) verteilt und die Unterschiede in den Therapieeffekten zwischen den Gruppen untersucht werden ¹⁵ . Siehe auch die Begriffe Randomisation und Kontrollgruppe.
Regressionsanalyse	Eine statistische Modellierungstechnik, die genutzt wird, um den Einfluss von einer oder mehreren erklärenden Variablen auf eine Outcomevariable zu schätzen oder vorauszusagen. Die logistische Regression und die Metaregression sind Arten von Regressionsanalysen ⁴ .
Relatives Risiko (RR)	Effektmaß für dichotome Variablen. Das relative Risiko in einer Therapiestudie bezeichnet das Verhältnis zwischen dem Risiko in der experimentellen Gruppe und dem Risiko in der Kontrollgruppe. Ein relatives Risiko von 1 bedeutet, dass zwischen den Vergleichsgruppen kein Unterschied besteht. Bei ungünstigen Ereignissen zeigt ein relatives Risiko < 1, dass die experimentelle Intervention wirksam ist, um das Auftreten von ungünstigen Ereignissen zu senken ¹⁰ .
Relative Risikoreduktion	Die relative Senkung der Rate an ungünstigen Ereignissen in der experimentellen Gruppe einer Studie im Vergleich zur Kontrollgruppe ¹⁰ .
Retrieval Bias	Systematischer Fehler in systematischen Reviews, der durch fehlerhafte Suchstrategien und Datenbankauswahl entsteht.
SAS [®] , SPSS [®] , STATA [®] und WinBugs [®]	Verschiedene Statistiksoftwarepakete.
Schiefe Verteilung	Die Schiefe einer Verteilung gibt an, wie stark die Verteilung der Datenwerte von einer symmetrischen Verteilung abweicht ¹⁴ .
Selektionsbias	Systematischer Fehler in Studien, wenn systematische Unterschiede zwischen Vergleichsgruppen in ihrer Prognose oder ihrem Ansprechen auf die Therapie vorliegen. Eine geeignete Zufallsverteilung (Randomisation) von Probanden auf die unterschiedlichen Gruppen einer Studie verhindert einen Selektionsbias ⁴ .

Fortsetzung: Glossar

Sensitivitätsanalyse	Analyseverfahren, das feststellt, wie stabil sich eine Modellrechnung oder eine Metaanalyse gegenüber Schwankungen der Eingangsparameter oder Veränderungen in der Berechnungsmethodik verhält ¹⁰ .
Signifikanzniveau	Das Signifikanzniveau gibt einen Grenzwert für die Sicherheit an, mit der ein beobachtetes (Studien)ergebnis nicht allein auf Zufall beruht. Ein Signifikanzniveau von 0,05 festzulegen bedeutet, dass man ein Ergebnis als signifikant akzeptiert, das rein zufällig nur in 5 % aller Stichprobenziehungen auftreten würde, wenn der vermutete Zusammenhang in Wahrheit, d. h. in der Grundgesamtheit, gar nicht besteht ¹⁴ .
Sprachbias	Systematischer Fehler in systematischen Reviews, der durch den Verzicht auf gezieltes Recherchieren nach nicht-englischsprachiger Literatur entsteht.
Standardabweichung	Maß für die Streuung von Messwerten um den Durchschnittswert ¹⁰ .
Standardfehler (SE)	Streuung von Stichprobenkennwerten um den wahren Wert des gesuchten Parameters in der Grundgesamtheit. Je größer der Standardfehler, desto breiter ist das Konfidenzintervall, das den wahren Parameter mit vorgegebener Wahrscheinlichkeit enthält ¹⁴ .
Sternförmiges Netzwerk	Begriff aus Mixed treatment comparison (MTC): Werden die zu vergleichenden Interventionen immer gegen den gleichen Komparator (z. B. Placebo) getestet, entsteht ein Evidenznetzwerk mit sternförmiger Struktur. Dies ist z. B. bei dem Vergleich von Zulassungsstudien der Fall, in denen gegen die Standardtherapie oder bei ethischer Vertretbarkeit gegen Placebo getestet wird ²⁰⁵ .
Streuung	Siehe Standardabweichung.
Subgruppenanalyse	Analyse, in der z. B. Interventionseffekte in einer definierten Untereinheit der Teilnehmer einer Studie erhoben werden ⁴ . Definiert wird die Untereinheit durch das Vorliegen von Kovariaten mit potenziellem Einfluss auf den Interventionseffekt wie Geschlecht oder Alter.
Surrogatparameter	Synonym: Surrogatendpunkte oder intermediäre Endpunkte. Endpunkte, die selbst nicht von unmittelbarer Bedeutung für die Patienten sind, aber stellvertretend für wichtige Endpunkte stehen können (z. B. Blutdrucksenkung als Surrogatparameter für Vermeidung eines Schlaganfalls). Surrogatendpunkte sind oft physiologische oder biochemische Marker, die sich relativ schnell und einfach messen lassen und denen eine Vorhersagefunktion für spätere klinische Ereignisse zugestanden wird. Voraussetzung für zuverlässige Aussagen über die Wirksamkeit einer Behandlung ist ein enger kausaler Zusammenhang zwischen Surrogatparameter und dem eigentlichen Endpunkt. Für die allermeisten Surrogatparameter ist ein kausaler Zusammenhang nicht belegt, was den Nutzen von Surrogatparametern für die Vorhersagekraft von klinisch relevanten Endpunkten in Frage stellt ¹⁰ .

Fortsetzung: Glossar

Systematischer Review; synonym mit systematischer Literaturübersicht	Sekundärforschung, bei der zu einer klar formulierten Frage alle verfügbaren Primärstudien systematisch und nach expliziten Methoden identifiziert, ausgewählt und kritisch bewertet und die Ergebnisse extrahiert und deskriptiv oder mit statistischen Methoden quantitativ (Metaanalyse) zusammengefasst werden ¹⁰ .
Überlebenszeitanalyse	Eine statistische Analyse, bei der die Zeit bis zum einem bestimmten Ereignis (time to event) zwischen zwei oder mehr Gruppen verglichen wird, um die Wirkung von prognostischen Faktoren, medizinischer Behandlung oder schädlichen Einflüssen zu schätzen. Das Ereignis kann dabei Tod sein, jedoch auch beliebige andere Endpunkte, wie Heilung, Erkrankung oder Eintreten einer Komplikation. Beispiele für eine solche Analyse sind die Kaplan-Meier-Methode oder die Cox-Regression ⁴ .
Validität/valide/ Validitätsprüfung	Validität heißt allgemein Gültigkeit, das heißt die Eigenschaft eines Instrumentes (z. B. Fragebogen, Messinstrument, Studie), das zu messen, was gemessen werden soll ¹⁴ .
Variabilität	Synonym für Streuung.
Varianz	Quadrat der Standardabweichung (siehe Standardabweichung).
Verblindung	Wenn Studienteilnehmer, betreuendes Studienpersonal und Forscher in Unkenntnis darüber sind, welche Interventionen die Studienteilnehmer erhalten haben ¹⁵ .
Verumgruppe	Gruppe von Patienten in einer klinischen Studie, die eine aktive Therapie erhalten.
Wirksamkeit	In der evidenzbasierten Medizin unterscheidet man bei der Wirksamkeit zwischen efficacy und effectiveness. Efficacy (Wirksamkeit unter Idealbedingungen) beschreibt die Wirksamkeit einer Maßnahme unter Idealbedingungen. Efficacy-Studien zeichnen sich durch hohe innere Validität aus, die Ergebnisse sind jedoch möglicherweise nur bedingt auf die Routineversorgung übertragbar. Effectiveness (Wirksamkeit unter Alltagsbedingungen) beschreibt die Wirksamkeit einer Maßnahme unter Routinebedingungen. Im Gegensatz zur efficacy (Wirksamkeit unter Idealbedingungen) untersuchen Effectiveness-Studien die Frage: Wirkt die Maßnahme unter den Bedingungen der Routineversorgung? ¹⁰
Z-Statistik	Teststatistik auf den Gesamteffektunterschied zwischen zwei Verteilungen. Unter der Nullhypothese, dass es keinen Therapieeffektunterschied gibt, folgt z einer Standard-Normalverteilung ⁸ .
Zwei-Schritt-Analyse	Im Kontext von Metaanalysen und indirekten Vergleichen: Die Entscheidung, ob ein Random- oder Fixed effects-Modell für die Metaanalyse verwendet werden soll, wird durch das Ergebnis eines Heterogenitätstest determiniert. Zeigt er statistisch signifikante Heterogenität an, wird ein Random effects-Modell gewählt.
z-Wert	Ergebnis der z-Statistik. Auf 95-prozentigem Konfidenzniveau zeigt ein z-Wert größer 1,96 oder kleiner -1,96 statistisch signifikant an, dass die zwei zu testenden Verteilungen nicht der gleichen Grundgesamtheit entstammen.

Zusammenfassung

Gesundheitspolitischer Hintergrund

Die vergleichende Nutzenbewertung von gesundheitsrelevanten Technologien ist nicht nur gesetzlich festgeschrieben (§ 139 und § 35b, SGB V: SGB = Sozialgesetzbuch), sondern ist auch ein zentrales Element von klinischen Leitlinien bzw. Entscheidungssituationen. Entscheidungsunterstützungsinstrumente wie Health Technology Assessments (HTA) sollten daher über ein valides methodisches Instrumentarium verfügen.

Wissenschaftlicher Hintergrund

Randomisierte kontrollierte Head-to-head-Studien, die Therapien direkt miteinander vergleichen, gelten als Goldstandard für den Wirksamkeitsvergleich. Da Studien dieses Typs nur begrenzt zur Verfügung stehen, sind Wirksamkeitsvergleiche auf indirekt vergleichende Methoden angewiesen, deren Validität allerdings noch kontrovers diskutiert wird.

Fragestellung

Fragestellungen für das vorliegende Assessment sind: Welche (statistischen) Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen gibt es, wie häufig werden sie eingesetzt und wie ist ihre Validität im Vergleich zu den Ergebnissen direkter Vergleiche zu beurteilen?

Methodik

In einer systematischen Literaturrecherche werden die medizinnahen Datenbanken des Deutschen Instituts für medizinische Dokumentation und Information (DIMDI) nach Methodenpapieren und Anwendungen von indirekten Vergleichen in systematischen Reviews durchsucht. Die Literaturauswertung erfolgt qualitativ beschreibend (Methoden) und quantitativ für die Häufigkeit ihres Einsatzes.

Eine Validitätsprüfung der Methoden für indirekte Therapievergleiche ist über den Vergleich ihrer Ergebnisse mit dem Goldstandard – den Ergebnissen von Head-to-head-Studien - möglich. In systematischen Reviews, in denen Therapieverfahren sowohl direkt als auch indirekt verglichen werden, werden diese Ergebnisse mithilfe der z-Statistik auf Übereinstimmung geprüft.

Ergebnisse

29 Methodenpapiere und 106 Methodenanwendungen werden ausgewertet. Aus diesen lassen sich vier Methoden für indirekte Vergleiche identifizieren: 1. Nicht-adjustierte indirekte Vergleiche schließen, unabhängig vom Komparator, alle randomisierten kontrollierten Studien (RCT), die einen Studienarm mit einer der interessierenden Therapieoptionen enthalten, ein. 2. Adjustierte indirekte Vergleiche und 3. Metaregressionen greifen nur auf Studien zurück, die einen Arm mit einer Therapieoption von Interesse und einen Arm mit einem gemeinsamen Komparator aufweisen. 4. Während die genannten Verfahren konventionelle Metaanalysetechniken einsetzen, können die mit Bayes'schen Methoden arbeitenden Mixed treatment comparisons (MTC) ein komplexes Netzwerk aus RCT mit multiplen Komparatoren simultan analysieren. Im Zeitraum von 1999 bis 2008 werden die adjustierten indirekten Vergleiche am häufigsten angewendet. Seit 2006 ist auch ein deutlicher Anstieg der Verwendung des methodisch anspruchsvolleren MTC zu verzeichnen.

Für die Validitätsprüfung stehen 248 Datensätze mit Ergebnisgegenüberstellungen aus direktem und indirektem Vergleich zur Verfügung. Der Anteil diskrepanter Ergebnisse mit statistischer Signifikanz war am größten bei den nicht-adjustierten indirekten Vergleichen (25,5 % [95 %-KI: 13,1; 38]), gefolgt von dem der Metaregressionen (16,7 % [95 %-KI: -13,2; 46,5]), der adjustierten indirekten Vergleiche (12,1 % [95 %-KI: 6,1; 18]) und des MTC (1,8 % [95 %-KI: -1,7; 5,2]). Diskrepante Ergebnisse werden vor allem dann beobachtet, wenn die Voraussetzung für die Durchführung eines indirekten Vergleichs – eine homogene Studienlage – nicht gegeben ist. Eine systematische Über- oder Unterschätzung der Ergebnisse direkter Vergleiche durch den indirekten Vergleich zeigt sich in dieser Stichprobe bei keinem der genannten Verfahren.

Diskussion

Die Auswahl einer geeigneten Methode für einen indirekten Vergleich hat sich an deren Validität, der Anzahl der zu vergleichenden Therapieoptionen sowie an der Qualität und Quantität der verfügbaren Studien zu orientieren. Nicht-adjustierte indirekte Vergleiche weisen in Gegenüberstellung zu direkten Vergleichen eine geringe Validität auf. Adjustierte indirekte Vergleiche und MTC können dagegen unter bestimmten Voraussetzungen Ergebnisse liefern, die in den meisten Fällen denen direkter Vergleiche entsprechen. Die Validität von indirekten Vergleichen mittels Metaregression kann auf Basis der wenigen, bisher verfügbaren Anwendungsbeispiele noch nicht beurteilt werden.

Schlussfolgerung/Empfehlung

Bei Einhaltung der zentralen Voraussetzung – Anwendung an einem Pool homogener, qualitativ hochwertiger RCT – lassen sich die Ergebnisse von qualitativ hochwertigen Head-to-head-Studien durch den Einsatz adjustierter indirekter Vergleiche und MTC abschätzen. Im Kontext von HTA und Leitlinienerstellung stellen sie somit wertvolle Hilfsinstrumente dar, wenn direkte Evidenz für einen Wirksamkeitsvergleich von Therapieverfahren nicht zur Verfügung steht.

Abstract

Health political background

The comparison of the effectiveness of health technologies is not only laid down in German law (Social Code Book V, § 139 and § 35b) but also constitutes a central element of clinical guidelines and decision making in health care. Tools supporting decision making (e. g. Health Technology Assessments (HTA)) are therefore in need of a valid methodological repertoire for these comparisons.

Scientific background

Randomised controlled head-to-head trials which directly compare the effects of different therapies are considered the gold standard methodological approach for the comparison of the efficacy of interventions. Because this type of trial is rarely found, comparisons of efficacy often need to rely on indirect comparisons whose validity is being controversially debated.

Research questions

Research questions for the current assessment are: Which (statistical) methods for indirect comparisons of therapeutic interventions do exist, how often are they applied and how valid are their results in comparison to the results of head-to-head trials?

Methods

In a systematic literature research all medical databases of the German Institute of Medical Documentation and Information (DIMDI) are searched for methodological papers as well as applications of indirect comparisons in systematic reviews. Results of the literature analysis are summarized qualitatively for the characterisation of methods and quantitatively for the frequency of their application.

The validity of the results from indirect comparisons is checked by comparing them to the results from the gold standard – a direct comparison. Data sets from systematic reviews which use both direct and indirect comparisons are tested for consistency by the z-statistic.

Results

29 methodological papers and 106 applications of indirect methods in systematic reviews are being analysed. Four methods for indirect comparisons can be identified: 1. Unadjusted indirect comparisons include, independent of any comparator, all randomised controlled trials (RCT) that provide a study arm with the intervention of interest. 2. Adjusted indirect comparisons and 3. metaregression analyses include only those studies that provide one study arm with the intervention of interest and another study arm with a common comparator. While the aforementioned methods use conventional metaanalytical techniques, 4. Mixed treatment comparisons (MTC) use Bayesian statistics. They are able to analyse a complex network of RCT with multiple comparators simultaneously.

During the period from 1999 to 2008 adjusted indirect comparisons are the most commonly used method for indirect comparisons. Since 2006 an increase in the application of the more methodologically challenging MTC is being observed.

For the validity check 248 data sets, which include results of a direct and an indirect comparison, are available. The share of statistically significant discrepant results is greatest in the unadjusted indirect comparisons (25,5 % [95 % CI: 13,1 %; 38 %]), followed by metaregression analyses (16,7 % [95 % CI: -13,2 %; 46,5 %]), adjusted indirect comparisons (12,1 % [95 % CI: 6,1 %; 18 %]) and MTC (1,8 % [95 % CI: -1,7 %; 5,2 %]). Discrepant results are mainly detected if the basic assumption for an indirect comparison – between-study homogeneity – does not hold. However a systematic over- or underestimation of the results of direct comparisons by any of the indirectly comparing methods was not observed in this sample.

Discussion

The selection of an appropriate method for an indirect comparison has to account for its validity, the number of interventions to be compared and the quality as well as the quantity of available studies. Unadjusted indirect comparisons provide, contrasted with the results of direct comparisons, a low validity. Adjusted indirect comparisons and MTC may, under certain circumstances, give results which are consistent with the results of direct comparisons. The limited number of available reviews utilizing meta-regression analyses for indirect comparisons currently prohibits empirical evaluation of this methodology.

Conclusions/Recommendations

Given the main prerequisite – a pool of homogenous and high-quality RCT – the results of head-to-head trials may be pre-estimated by an adjusted indirect comparison or a MTC. In the context of HTA and guideline development they are valuable tools if there is a lack of a direct comparison of the interventions of interest.

Kurzfassung

Gesundheitspolitischer Hintergrund

Entscheidungen über die Kostenübernahme für medizinische Maßnahmen durch die Solidargemeinschaft werden verstärkt auf Basis von Nutzen- oder Kosten-Nutzen-Bewertungen im Rahmen von Health Technology Assessments (HTA) gestützt.

Randomisierte kontrollierte Head-to-head-Studien, die Therapien direkt miteinander vergleichen, gelten als Goldstandard für den Wirksamkeitsvergleich. Mit dem medizinisch-wissenschaftlichen Fortschritt steigt jedoch die Zahl der Therapieoptionen für eine bestimmte Indikation, ohne dass die regulativen Vorgaben eine sorgfältige Prüfung ihrer Wirksamkeit gegen bereits etablierte Therapieformen vorsehen. In der Pharmakotherapie reicht häufig ein Vergleich der Wirksamkeit neuer Arzneistoffe gegen Placebo für eine Zulassung aus. Daher sehen die Arzneimittelhersteller bislang nur selten Veranlassung, ihre neuen Arzneistoffe gegen bereits im Markt befindliche Arzneistoffe zu testen. Bei multiplen Behandlungsmöglichkeiten für eine Indikation wird es deshalb in der Regel keine randomisierten kontrollierten Studien (RCT) geben, die sämtliche Therapiemöglichkeiten parallel, d. h. direkt miteinander vergleichen. Soll dennoch eine Aussage zur vergleichenden Wirksamkeit gemacht werden, muss auf indirekt vergleichende Verfahren zurückgegriffen werden.

Wissenschaftlicher Hintergrund

Unter indirekten Vergleichen versteht man die Gegenüberstellung der Wirksamkeit verschiedener Interventionen mithilfe eines gemeinsamen Komparators. Hierbei kann es sich um einen aktiven Komparator (meistens Standardtherapie) oder um Placebo handeln.

Viele Fragen zur Validität indirekter Vergleiche sind aber bisher noch ungeklärt. Eine umfassende systematische Übersicht über verfügbare Methoden für indirekte Vergleiche und ihre Validität wird in dem 2005 publizierten britischen HTA-Bericht von Glenny et al. gegeben. Dieser Bericht, der drei Methoden für indirekte Vergleiche (nicht-adjustierte und adjustierte indirekte Vergleiche sowie Metaregressionen) vorstellt, schließt Publikationen bis 1999 ein. Die Autoren kommen aufgrund des damaligen Kenntnis- und Erfahrungsstandes zu dem Schluss, dass zwischen den Ergebnissen direkter und indirekter Vergleiche deutliche Diskrepanzen sichtbar werden, deren Richtung aber nicht voraussehbar ist. Sie weisen auf eine hohe Anfälligkeit für systematisch verzerrte Ergebnisse in nicht-adjustierten indirekten Vergleichen hin. Adjustierte indirekte Vergleiche und Metaregressionen zeigen dagegen eine deutlich höhere Validität.

Auf der Grundlage dieser Vorarbeiten gibt der vorliegende Bericht anhand von fünf Forschungsfragen einen Überblick über den aktuellen Kenntnisstand. Dabei wird unter anderem aus pragmatischen Gründen auf indirekte Vergleiche fokussiert, die Aussagen zur Wirksamkeit therapeutischer Verfahren machen und hierzu die Ergebnisse von RCT heranziehen.

Forschungsfragen

1. Welche Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen existieren bisher (März 2008) und unter welchen Voraussetzungen können sie eingesetzt werden?
2. Welche Methoden des indirekten Wirksamkeitsvergleichs therapeutischer Interventionen wurden wie häufig in bisher publizierten systematischen Reviews eingesetzt?
3. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen und führen sie zu den gleichen Schlussfolgerungen über Therapieeffektunterschiede?
4. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen, wenn in den indirekten Vergleichen zusätzlich Ergebnisse von Head-to-head-Studien eingebracht werden?
5. Lässt sich ein Goldstandard identifizieren, nach dem indirekte Vergleiche der Wirksamkeit von therapeutischen Interventionen vorgenommen werden sollten?

Methodik

Es werden systematische Literaturrecherchen mit zwei (bzw. drei) Zielen durchgeführt:

1. Auffinden von Methodenpapieren zu indirekten Vergleichen.
2. Auffinden von systematischen Literaturübersichten, in denen indirekte Vergleiche eingesetzt werden (ausschließlich, oder in Ergänzung zu den Informationen aus direkten Vergleichen).

Der Grundstock an relevanten Referenzen wird dem systematischen Review von Glenny et al. entnommen, der die Literatur bis 1999 abdeckt. Für die nach diesem Datum erschienene Literatur werden die medizinnahen Datenbanken des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI) und des ISI Web of Knowledge[®] durchsucht.

Zur Ergänzung werden Handsuchen in den Referenzlisten wichtiger Methodenpapiere, sowie eine Recherche auf den Internetseiten der Mitgliedsorganisationen des International network of agencies for Health Technology Assessment (INAHTA), durchgeführt.

Die Beschreibung der einzelnen indirekt vergleichenden Methoden erfolgt, soweit vorhanden, auf Basis der Methodenpapiere und wird ergänzt um Angaben aus den Methodenbeschreibungen von Anwendungsbeispielen. Die Bestimmung der Häufigkeit des Einsatzes der einzelnen indirekt vergleichenden Methoden erfolgt anhand der gefundenen systematischen Reviews mit indirekten Vergleichen.

Indirekt vergleichende Methoden, die mit metaanalytischen Techniken arbeiten, werden einer empirischen Validitätsprüfung unterzogen. Hierzu werden systematische Literaturübersichten analysiert, in denen Therapieoptionen sowohl anhand der Ergebnisse von Head-to-head-Studien als auch indirekt, unter Hinzuziehung eines gemeinsamen Komparators, gegenübergestellt werden.

Für jede indirekt vergleichende Methode wird die Hypothese geprüft, ob sich die Ergebnisse des indirekten von denen des direkten Vergleichs signifikant unterscheiden. Um diese Hypothese zu testen, wird zunächst für jedes Ergebnis eines direkten und indirekten Vergleichs zu dem gleichen Therapievergleich (Datensatz) die Differenz der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich gebildet; in der Folge Diskrepanz genannt. Anschließend werden, um die Diskrepanzen miteinander vergleichbar zu machen, z-Werte berechnet und in vier Schritten die Validität der Ergebnisse der verschiedenen indirekt vergleichenden Methoden denen der direkten Vergleiche gegenübergestellt.

1. Test auf systematische Über- oder Unterschätzung: Sind die z-Werte (Diskrepanzen) normalverteilt um einen Mittelwert $z = 0$ (Kolmogorov-Smirnov-Test, $p \leq 0,05$)?
2. Quantifizierung des Ausmaßes der Diskrepanz: Berechnung der mittleren Beträge von z ($|\bar{z}|$).
3. Bestimmung des Anteils der statistisch signifikant diskrepanten Datensätze ($|\bar{z}| \geq 1,96$) an allen für die Validitätsprüfung einer Methode herangezogenen Datensätze.
4. In statistisch signifikant diskrepanten Datensätzen: Überprüfung der Homogenität des Studienpools für sowohl den direkten als auch für den indirekten Vergleich.

Abschließend wird berichtet, in wie vielen Fällen der direkte und der indirekte Vergleich zu den gleichen Schlussfolgerungen kommen.

Da vermutet wird, dass der Einschluss von Head-to-head-Studien in indirekte Vergleiche deren Ergebnisse nivelliert (d. h. an die Ergebnisse der Head-to-head-Studien annähert), wird die Validitätsprüfung (Hauptanalyse) an einer Subgruppe von Datensätzen (Subgruppenanalyse) wiederholt, in denen keine Head-to-head-Studien in den indirekten Vergleich eingeschlossen wurden.

Ergebnisse

Methodenbeschreibungen

Die Durchsicht der Literatur ergibt, dass allen Methoden für indirekte Vergleiche die gleiche Annahme zugrunde liegt: Die Variabilität zwischen den Ergebnissen der einzuschließenden Studien ist ausschließlich zufällig bedingt, oder – in anderen Worten – es besteht keine bedeutsame Heterogenität zwischen den Studien.

Es werden vier häufig eingesetzte Methoden für die Durchführung indirekter Vergleiche identifiziert, die alle mit metaanalytischen Ansätzen arbeiten:

1. Nicht-adjustierter indirekter Vergleich: Für den Vergleich einer Therapie A mit einer Therapie B werden die Ergebnisse aller mit A behandelten Studienarme in einer Metaanalyse zu einem Gesamteffektschätzer θ_A zusammengefasst. Das gleiche wird in einer zweiten Metaanalyse für alle mit B behandelten Studienarme durchgeführt um θ_B zu erhalten.
Dieses Vorgehen wird als „Nicht-adjustierter indirekter Vergleich“ bezeichnet, was eine Abkürzung der Bezeichnung „Nicht bezüglich der Ergebnisse aus den Kontrollgruppen adjustierter indirekter Vergleich“ darstellt. Für den Vergleich der zusammengefassten Gesamteffektschätzer θ_A und θ_B gibt es im Anschluss vier Möglichkeiten: Die Bildung eines Gesamteffektschätzers $\theta_{A \text{ versus } B}$; die Prüfung der Differenz zwischen θ_A und θ_B auf statistische Signifikanz mithilfe eines statistischen Tests; die Prüfung der Konfidenzintervalle um θ_A und θ_B auf Überlappung oder ein narratives Gegenüberstellen der Therapieeffekte von A und B.
2. Für den (hinsichtlich der Ereignisse in den Komparatorarmen) adjustierten indirekten Vergleich werden zunächst mithilfe konventioneller Metaanalysemethoden die Gesamteffektschätzer $\theta_{A \text{ versus Komparator}}$ und $\theta_{B \text{ versus Komparator}}$ gebildet. Zum Vergleich der beiden Gesamteffektschätzer sind die unter 1. skizzierten vier Methoden einsetzbar.
3. Bei der Metaregression werden zur Bildung der Gesamteffektschätzer θ_A und θ_B die Einzelstudienresultate „A versus Komparator“ und „B versus Komparator“ in zwei Regressionsgleichungen eingesetzt. Die Regressionsmodelle können darüber hinaus die Effekte von Kovariaten (die als Auslöser von Heterogenität gelten – wie z. B. Alter der Studienpopulationen oder Krankheitsschweregrade) bei der Bildung der Gesamteffektschätzer berücksichtigen. Der Vergleich von θ_A und θ_B erfolgt wieder mithilfe der vier genannten Methoden.
4. Mixed treatment comparison (MTC) steht als Sammelbegriff für indirekt vergleichende Methoden, die mehr als zwei Therapieoptionen simultan miteinander vergleichen und eventuell vorhandene Head-to-head-Studien mit aufnehmen können. Sie kann unbegrenzt viele Therapieoptionen in eine Wirksamkeitsrangfolge bringen. Dabei folgt sie mithilfe der Bayes'schen Statistik einem iterativem Konzept, das sukzessive die gesamte verfügbare RCT-Evidenz zu dem/n Therapievergleich/en zusammenfasst.

Indirekte Vergleiche ohne Metaanalyse werden durchgeführt, wenn nur je eine Studie mit Informationen zu Therapie A und eine zu Therapie B zur Verfügung steht, bzw. wenn die vorhandenen Studien hochgradig heterogen sind. Auch die dann durchgeführten deskriptiven Vergleiche verfolgen entweder das Prinzip des nicht-adjustierten oder des adjustierten Vergleichs.

Relative Häufigkeit des Einsatzes der verschiedenen Methoden für indirekte Vergleiche

In dem ungefähr neunjährigen Zeitraum von Januar 1999 bis Februar 2008 werden in den gefundenen systematischen Übersichtsarbeiten 108-mal metaanalytische Methoden für indirekte Vergleiche angewendet. Ein zeitlicher Trend mit einer kontinuierlich ansteigenden Zahl der Publikationen mit indirekten Vergleichen ist erkennbar.

Die mit Abstand am häufigsten verwendete Methode ist der adjustierte indirekte Vergleich mit 60 Einsätzen. Es folgen die Metaregressionen (17 Einsätze), die nicht-adjustierten indirekten Vergleiche (14 Einsätze), der MTC (zwölf Einsätze) und die sonstigen Methoden, die sich nicht in die vier wichtigsten Methodengruppen einordnen lassen (fünf Einsätze). 2006 ist ein sprunghafter Anstieg bei der Häufigkeit der durchgeführten MTC zu verzeichnen (zehn Beispiele von 2006 bis 2007).

Validitätsprüfung

Insgesamt werden 248 Gegenüberstellungen von Metaanalysen mit direktem und indirektem Vergleich (Datensätze) aus 57 systematischen Reviews für die Validitätsprüfungen herangezogen.

Bei der Testung auf systematische Über- oder Unterschätzung stellt sich heraus, dass keine der untersuchten indirekt vergleichenden Methoden systematisch die Ergebnisse des direkten Vergleichs über- oder unterschätzt.

Dennoch sind Diskrepanzen zwischen den Ergebnissen direkter und indirekter Vergleiche nachweisbar, die mit Abstand am stärksten bei den nicht-adjustierten indirekten Vergleichen ausgeprägt sind ($|\bar{z}| = 1,63$ [95 %-KI: 1,20; 2,07]). Die durchschnittlichen Diskrepanzen der adjustierten indirekten Vergleiche ($|\bar{z}| = 0,95$ [95 %-KI: 0,80; 1,09]), der Metaregressionen ($|\bar{z}| = 0,99$ [95 %-KI: 0,20; 1,79]) und der MTC ($n = 57$; $|\bar{z}| = 0,59$ [95 %-KI: 0,45; 0,73]) liegen auf einem vergleichbaren Niveau. Der niedrige mittlere Betrag der MTC nivelliert sich in der Subgruppenanalyse, in der keine Head-to-head-Studien mit in die MTC eingeschlossen sein durften ($n = 12$; $|\bar{z}| = 0,83$ [95 %-KI: 0,40; 1,26]). Bei den anderen Methoden sind die Ergebnisse aus Haupt- und Subgruppenanalyse konkordant.

Es muss allerdings berücksichtigt werden, dass die Streuung um die Mittelwerte der durchschnittlichen Diskrepanz bei den verschiedenen Methoden stark variiert. Es werden unterschiedlich viele extreme Werte mit statistisch signifikanter Diskrepanz ($|\bar{z}| > 1,96$; entspricht $p = 0,05$) gefunden: Beim nicht-adjustierten indirekten Vergleich beträgt der Anteil statistisch signifikant diskrepanter Datensätze an allen ausgewerteten Datensätzen 25,5 % ($n = 47$; 95 %-KI: 13,1 %; 38,0 %), beim adjustierten indirekten Vergleich 12,1 % ($n = 116$; 95 %-KI: 6,1 %; 18,0 %), bei der Metaregression 16,7 % ($n = 6$; 95 %-KI: -13,2 %; 46,5 %) und beim MTC 1,8 % ($n = 57$; 95 %-KI: 2,1 %; 34,3 %). Die Ergebnisse der Haupt- sind mit denen der Subgruppenanalyse konkordant.

Zusammengefasst weisen 32 der 248 Gegenüberstellungen von Ergebnissen aus direktem und indirektem Vergleich statistisch signifikante Diskrepanz auf (12,9 % [95 %-KI: 8,7 %; 17,1 %]).

Für 15 der 32 diskrepanten Datensätze lagen keine Angaben zur Heterogenität des zugrunde liegenden Studienpools vor. Bei elf der diskrepanten Datensätze war unter den eingeschlossenen Studien signifikante Heterogenität nachweisbar und bei sechs der diskrepanten Datensätze nicht.

Kongruenz in den Schlussfolgerungen

In knapp der Hälfte der 248 Gegenüberstellungen liegt sowohl im direkten als auch im indirekten Vergleich kein signifikanter Therapieeffektunterschied vor (49,2 %; 95 %-KI: 43,0 %; 55,4 %). Am zweithäufigsten wird sowohl vom indirekten als auch vom direkten Vergleich die gleiche Therapieoption als signifikant überlegen erkannt (21,8 %; 95 %-KI: 16,6 %; 26,9 %). In den restlichen 29 % (95 %-KI: 23,4 %; 34,7 %) der Fälle erzielen der direkte und der indirekte Vergleich keine Übereinstimmung in ihren Schlussfolgerungen. Der gefürchtete Fall, dass direkter und indirekter Vergleich jeweils das andere Therapieverfahren als signifikant überlegen erachten, kann allerdings nur in fünf Fällen (Anteil 2 %; 95 %-KI: 0,3 %; 3,8 %) beobachtet werden.

Präzision indirekt vergleichender Methoden

In der untersuchten Stichprobe ($n = 248$) besitzt der Effektschätzer des indirekten Vergleichs eine im Median um 9 % geringere Konfidenzintervallweite als der direkte Vergleich (25. Perzentil: -34 %; 50. Perzentil (Median): -9 %; 75. Perzentil: 30 %). Gleichzeitig beinhalten diese indirekten Vergleiche im Median exakt sechsmal so viele Studien wie die direkten Vergleiche (25. Perzentil: 4; 75. Perzentil: 13). In der untersuchten Stichprobe ergibt sich somit, bei einem Verhältnis an eingeschlossenen Studien von sechs zu eins zwischen indirekten Vergleich und direktem Vergleich, ein geringfügig engeres Konfidenzintervall beim indirekten Vergleich. Dies unterstützt die Aussage von Glenny et al., dass ein indirekter Vergleich viermal mehr ungefähr gleich große Studien einschließen müsse als ein direkter, um die gleiche Präzision zu erhalten.

Diskussion

Bei der Entscheidung ob, und wenn ja welche, Methode des indirekten Vergleichs eingesetzt werden kann, sollten vier Kriterien berücksichtigt werden:

1. Validität der Methode

Die geringste Validität im Vergleich zu direkt vergleichenden Methoden weist der nicht-adjustierte indirekte Vergleich auf. Da die Ergebnisse der Kontrollgruppen nicht mit in der Analyse berücksichtigt werden, wird von einem Bruch der Randomisation gesprochen. In der Analyse besteht ein großes Risiko, dass Biasformen, die für unkontrollierte Studien typisch sind (u. a. Selektionsbias und Confounding), die Resultate verzerren.

Der adjustierte indirekte Vergleich, die Metaregression und der MTC berücksichtigen dagegen die Ergebnisse in den Kontrollgruppen der Einzelstudien und bewahren somit die erzielte Strukturgleichheit in den Einzelstudien. Ein Selektionsbias auf der Metaebene kann jedoch auch bei diesen Methoden auftreten, wenn die Studien unterschiedliche Einschlusskriterien aufweisen. Die resultierenden ungleich verteilten Patientencharakteristika können, wenn sie mit dem Therapieeffekt assoziiert sind, zu Confoundern werden. Confounder täuschen einen Therapieeffekt vor, der nicht ursächlich durch die Therapie, sondern durch den Confounder verursacht wird.

Aufgrund dieser Verzerrungsrisiken sollten die genannten indirekt vergleichenden Methoden nicht bei heterogener Studienlage eingesetzt werden. Diese, alle indirekt vergleichenden Methoden betreffende Voraussetzung, gilt auch für konventionelle Metaanalysen und stellt im Prinzip keine Besonderheit der indirekten Vergleiche dar.

Die theoretischen Betrachtungen werden durch die empirische Validitätsprüfung der indirekt vergleichenden Methoden gestützt. Ausreichend Datensätze zur Untermauerung der Hypothese, dass – eine homogene Studienlage vorausgesetzt - adjustierte indirekte Vergleiche die gleiche Ergebnisvalidität wie direkte Vergleiche aufweisen können, liegen vor. Ebenso können MTC, wenn diese Head-to-head-Studien mit den Therapieoptionen von Interesse einschließen, eine hohe Ergebnisvalidität bescheinigt werden.

Die Validität von Metaregressionen, MTC ohne Einschluss von Head-to-head-Studien und der unter dem Punkt „Sonstige Methoden“ gelisteten Techniken, kann aufgrund mangelnder Daten nicht bewertet werden.

2. Anzahl der zu vergleichenden Therapien

Sollen nur zwei Therapieverfahren indirekt miteinander verglichen werden, bietet sich unter Berücksichtigung der Ergebnisvalidität und des methodischen Aufwandes vor allem der Einsatz eines adjustierten indirekten Vergleichs an. Sollen jedoch mehr als zwei Therapien einander gegenüber gestellt werden, ist nur der MTC einsetzbar, um diese in eine Rangfolge hinsichtlich ihrer Wirksamkeit zu bringen.

3. Einschluss von Head-to-head-Studien

Neben den MTC liefern auch die anderen drei indirekt vergleichenden Methoden Ansätze für den Einschluss von Head-to-head-Studien in den indirekten Vergleich. Es liegen allerdings keine ausreichenden Daten für eine Validitätsprüfung vor. Für die bisher publizierten Beispiele gilt, dass der MTC, der Head-to-head-Studien mit einschließt, zu den gleichen Ergebnissen kommen kann, wie die Head-to-head-Studien allein. Sein zusätzlicher Vorteil liegt darin, dass durch Hinzunahme des indirekten Vergleichs die Präzision des Gesamteffektschätzers erhöht werden kann.

4. Heterogene Studienlage

Der indirekte Vergleich mittels Metaregression stellt noch keine ausreichend validierte Methode dar, um mit ihr verlässlich für Heterogenitätsfaktoren adjustieren zu können. Auch die Adjustierung für Kovariaten beziehungsweise die Berücksichtigung von Heterogenität mit einem Inkonsistenzfaktor in MTC wurden noch nicht ausreichend validiert.

Liegt klinisch bedeutsame Heterogenität vor, kann derzeit kein indirekt vergleichendes Verfahren empfohlen werden, da das Risiko für systematische Fehler hoch ist. Bei geringer Heterogenität, kann über ein Random effects-Modell eine konservative Schätzung vorgenommen werden. Fixed effects-Modelle sollten nur bei weitgehend homogenen Studiensätzen eingesetzt werden. Für alle in diesem HTA beschriebenen Methoden für indirekte Vergleiche stehen sowohl Fixed als auch Random effects-Modelle zur Verfügung.

Schlussfolgerungen

Es bleibt festzustellen, dass für indirekte Vergleiche eine Reihe von Methoden zur Verfügung stehen, die sich hinsichtlich ihrer Möglichkeiten, die Evidenz verschiedener Studienlagen zusammenzufassen, unterscheiden.

In der empirischen Überprüfung zeigt sich, dass vor allem die Ergebnisse nicht-adjustierter indirekter Vergleiche häufig von denen direkter Therapievergleiche abweichen. Die übrigen indirekt vergleichenden Methoden können insbesondere dann konkordante Ergebnisse liefern, wenn die zusammengefassten Studien wenig Heterogenität aufweisen.

Aus diesem Grund sollten adjustierte indirekte Vergleiche, Metaregressionen und MTC nur bei einer homogenen Studienlage eingesetzt werden. Zudem muss ein indirekter Vergleich, damit er einen ähnlich präzisen Effektschätzer wie ein direkter Vergleich liefern kann, ca. sechsmal so viele ungefähr gleich große Studien wie der direkte Vergleich enthalten.

Für den Kontext von HTA und Leitlinienerstellung stellen indirekte Vergleiche dann ein wertvolles Hilfsinstrument dar, wenn direkte Evidenz für einen Wirksamkeitsvergleich von Therapieverfahren nicht zur Verfügung steht.

Da eine homogene Studienlage in der Praxis allerdings nur selten vorliegt, bleibt vor einer breiteren Verwendung von indirekten Therapievergleichen zu definieren, bei welchem Ausmaß an Heterogenität (und Inkonsistenz) es aus Validitätsgründen noch vertretbar ist, sie zur Beantwortung von Forschungsfragen heranzuziehen.

Summary

Health political background

In the system of statutory health insurance coverage decisions are increasingly based on the results of effectiveness or cost-effectiveness analyses conducted in the context of Health Technology Assessments (HTA).

Randomised controlled head-to-head trials which directly compare the effects of different therapies are considered the gold standard methodological approach for the comparison of the efficacy of medical interventions. While research progresses, more and more treatment options are being developed for certain indications. As concerns pharmacological interventions, proven positive effects compared to placebo may be sufficient to attain market approval. Therefore manufacturers rarely see the need to test the effects of new interventions against the effects of interventions that are in the market already. Given multiple therapeutic options for an indication, there will hardly be a head-to-head trial testing all options in parallel. Statements on comparative efficacy have to rely on indirect comparisons.

Scientific background

Comparisons are defined as indirect if the effects of interventions are compared to each other by their performance against a common comparator. This may be an active intervention (usually standard care) or placebo. Up to date many questions concerning the validity of indirect comparisons remain unanswered. In 2005 a British HTA report was published, containing a comprehensive systematic overview of available methods for indirect comparisons and their validity. The report, which refers to publications up to 1999, introduces three methodological approaches for indirect comparisons: unadjusted and adjusted indirect comparisons and metaregression-analyses. The authors conclude that discrepancies between the results of direct and indirect comparisons are considerable but their direction cannot be foreseen. It is pointed out that unadjusted indirect comparisons are highly prone to bias. Contrasting, adjusted indirect comparisons and metaregression-analyses provide a higher degree of validity.

On the basis of these results the current report gives an updated review of indirect comparisons by means of five research questions. It focuses on the comparative efficacy of medical interventions on the basis of high-quality randomised controlled trials (RCT).

Research questions

1. What methodological approaches for indirect comparisons of therapeutic interventions are available today (March 2008) and under what circumstances may they be applied?
2. What methodological approaches for indirect comparisons have been applied in systematic reviews and how often?
3. What is the validity of results from indirect comparisons compared to the results of direct comparisons and do both arrive at the same conclusions?
4. What is the validity of results from indirect comparisons compared to the results of direct comparisons if results from head-to-head trials are included in the indirect comparison?
5. Is it possible to identify a "gold standard methodology" for indirect comparisons of competing interventions?

Methods

Systematic literature searches are conducted with two purposes:

1. Identification of papers describing methodological approaches for indirect comparisons.
2. Identification of systematic reviews which apply indirect comparisons (exclusively, or in addition to information from direct comparisons).

The basis of relevant references is extracted from the systematic review of Glenny et al. which covers the relevant literature up to 1999. To identify papers published after 1999 all medical databases of the German Institute of Medical Documentation and Information (DIMDI) and the ISI Web of Knowledge® are searched using the search strategy of Glenny et al. with minor modifications.

In addition, reference lists of the main methodological papers and systematic reviews as well as the homepages of the member institutions of the International network of agencies for Health Technology Assessment (INAHTA) are screened for relevant papers.

The description of the different methodological approaches for indirect comparisons is based as far as possible on information from methodological papers and completed by information from methods chapters of published applications. Their application frequency is calculated by counting the number of applications in all systematic reviews with indirect comparisons published 1999 to 2008.

Indirect comparisons which use metaanalysis techniques are validated empirically on the basis of systematic reviews that report results of direct as well as indirect comparisons. For every methodological approach the following hypothesis is tested: the results of the indirect comparison do not differ significantly from the results of the direct comparison. In order to test this hypothesis the difference in the results of a direct and an indirect comparison for the same intervention is calculated. This difference is named discrepancy. In order to make discrepancies from different reviews comparable, they are transformed into z-scores. The final validity check for the different methodological approaches for indirect comparisons was performed in four steps.

1. Test for systematic over- or underestimation: Are the z-scores normally distributed with an average value of $z = 0$ (Kolmogorov-Smirnov-Test, $p \leq 0.05$)?
2. Quantification of the amount of discrepancy: Calculation of the mean absolute value of z ($|\bar{z}|$).
3. Determination of the share of statistically discrepant z-scores ($|\bar{z}| \geq 1.96$) among all z-scores.
4. For data sets with statistically significant discrepant z-scores: Homogeneity testing of the underlying study pool for the direct and indirect comparisons.

Finally it is reported in how many cases the direct and indirect comparisons arrive at the same conclusions.

While it is assumed that inclusion of head-to-head trials into indirect comparisons may level out discrepancies between direct and indirect comparison, the validity check (main analysis) is repeated in a subgroup of data sets (subgroup analysis), which do not include results from head-to-head trials into indirect comparisons.

Results

Method descriptions

Literature reveals that all methodological approaches for indirect comparisons are based on the same assumption: The observed variability among the results of studies that are going to be included into an indirect comparison is solely due to random error or - in other words - no meaningful between-study heterogeneity is present.

Four frequently applied methodological approaches for indirect comparisons, which use metaanalytical methods, are identified:

1. In an unadjusted indirect comparison the comparison of an intervention A with an intervention B is prepared by metaanalytically pooling the results of all study arms treated with A to get a summary estimate θ_A and by doing the same in a second metaanalysis with all study arms treated with B to get θ_B . This procedure is called "unadjusted indirect comparison" because the indirect comparison is not adjusting for events in the control group. There are four ways of comparing the summary effect estimates θ_A and θ_B : calculation of a summary effect estimate $\theta_{A \text{ versus } B}$; testing the difference between θ_A and θ_B for statistical significance; check the confidence intervals around θ_A and θ_B for overlap or a narrative comparison of the efficacy of A and B.
2. To perform an adjusted (for events in the comparator arms) indirect comparison the summary effect estimates $\theta_{A \text{ versus } \text{comparator}}$ and $\theta_{B \text{ versus } \text{comparator}}$ are calculated by conventional metaanalytic methods. For the comparison of the two summary effect estimates the same four methods as introduced in point 1 are applicable.

3. In metaregression-analyses the summary effect estimates $\theta_{A \text{ versus comparator}}$ and $\theta_{B \text{ versus comparator}}$ are estimated separately in two regression equations. In addition to adjusting for effects in the comparator arms the regression models can adjust for the effects of further covariates (which are regarded as the origin of heterogeneity – like i. e. age of study population or severity of illnesses). Again, the comparison of $\theta_{A \text{ versus comparator}}$ and $\theta_{B \text{ versus comparator}}$ is performed by the above mentioned four methods (see 1.).
4. Mixed treatment comparison (MTC) is a collective term for methodological approaches for indirect comparisons comparing more than two interventions simultaneously and possibly including head-to-head studies. MTC are able to rank an unlimited number of therapeutic options according to their efficacy. For that purpose Bayesian statistics are applied to successively pool all available evidence from RCT in order to gain summary effect estimates for all possible comparisons of the interventions of interest.

Indirect comparisons without metaanalysis are performed if there is only one trial available for the options of interest or if available studies are highly heterogeneous. Indirect comparisons without metaanalysis also follow the principles of adjusted or unadjusted comparisons and may be performed by the four methods introduced in 1.

Application frequency of different methodological approaches for indirect comparisons

In 106 systematic reviews published between January 1999 and February 2008, found by the literature searches, one metaanalytic method of an indirect comparison is applied (exception: Vandermeer et al. 2007 applied three different methods). The considerably most frequently applied method is the adjusted indirect comparison (60 times), followed by metaregression-analyses (17 times), unadjusted indirect comparisons (14 times), MTC (twelve times) and other approaches which cannot be allocated to the four main methodological groups (five times). In 2006 a steep rise in the utilisation of MTC is observed (ten examples from 2006 until 2007).

Validity check

For the validity check of the indirect approaches a total of 248 paired results from direct and indirect comparisons (z-scores) are available from 57 systematic reviews.

The test for systematic over- or underestimation reveals that none of the approaches for indirect comparisons systematically over- or underestimates the results of a corresponding direct comparison. Nevertheless, differences in the mean absolute z-scores are observed among the indirect methods: The largest are found with the unadjusted indirect comparisons ($|\bar{z}| = 1.63$ [95 %-CI: 1.20; 2.07]) while adjusted indirect comparisons ($|\bar{z}| = 0.95$ [95 %-CI: 0.80; 1.09]), metaregression-analyses ($|\bar{z}| = 0.99$ [95 %-CI: 0.20; 1.79]) and MTC ($n = 57$; $|\bar{z}| = 0.59$ [95 %-CI: 0.45; 0.73]) provide lower values. For the MTC a higher average z-score is observed in the subgroup analysis without inclusion of head-to-head trials ($n = 12$; $|\bar{z}| = 0.83$ [95 %-CI: 0.40; 1.26]) while the results of the main and subgroup analyses are concordant for the other methods. It is to be noted though that the variance of the mean absolute z-scores differs considerably across the methods. The number of outstandingly high z-scores ($|\bar{z}| > 1.96$) varies among the indirect methodological approaches: the unadjusted indirect comparison provides a share of 25.5 % ($n = 47$; 95 %-CI: 13.1 %; 38.0 %) of statistically significant discrepant z-scores, the adjusted indirect comparison of 12.1 % ($n = 116$; 95 %-CI: 6.1 %; 18.0 %), the metaregression-analysis of 16.7 % ($n = 6$; 95 %-CI: -13.2 %; 46.5 %) and the MTC of 1.8 % ($n = 57$; 95 %-CI: 2.1 %; 34.3 %). The results from the main and subgroup analysis are concordant. Summarising all indirect methods, 32 of 248 comparisons provide statistically significant discrepancies (12.9 % [95 %-CI: 8.7%; 17.1%]).

For 15 of the 32 statistically significant discrepancies (z-scores) no information concerning heterogeneity of the pooled studies is given by the original review authors. Proof of significant heterogeneity is found by the original review authors in eleven of the statistically significant discrepant comparisons but not in the remaining six.

Congruence of conclusions

In about half of the 248 comparisons of interventions no statistically significant difference in therapeutic efficacy is found - neither by direct nor by indirect comparison (49.2 %; 95 %-CI: 43.0 %; 55.4 %). In 21.8 % (95 %-CI: 16.6 %; 26.9 %) of cases one intervention is found to perform significantly better than the other by both the direct and the indirect comparison. In another 29 % (95 %-CI: 23.4 %; 34.7 %) of the analysed comparisons the conclusions of the direct and indirect comparison are not concordant. However the feared case that the direct comparison prefers the one and the indirect comparison the other intervention with statistical significance is observed rarely (five cases; corresponding to a share of 2 % (95 %-CI: 0.3 %; 3.8 %)) among all cases.

Precision of indirect comparisons

In the analysed sample (n = 248) the confidence intervals around the effect estimates of the indirect comparisons are found to be slightly smaller than those around the direct estimates (median difference: 9 % (25th percentile: -34 %; 75th percentile: 30 %) while the indirect comparisons include six times more studies than the direct comparisons (median: 6 (25th percentile: 4; 75th percentile: 13)). It may therefore be stated that for the analysed sample a six to one ratio of included studies (with an approximately equal number of participants) for the indirect and direct comparison yields almost comparable precision of effect estimates. This supports the claim of Glenny et al. that an indirect comparison must include four times as many studies (of equal size) as a direct comparison to yield the same precision.

Discussion

In decision making whether, and if so, which approach of indirect comparisons should be applied, four criteria should be taken into consideration:

1. Validity of the methodological approach

Compared to the results of head-to-head trials unadjusted indirect comparison provide the lowest validity. Some authors blame the method for breaking the randomisation of the included RCT because effects are not adjusted for events in the control groups. Therefore results are easily distorted by all types of biases that are normally typical for observational studies (i. e. selection bias and confounding).

In contrast the adjusted indirect comparison, the metaregression and the MTC adjust for events in the control groups and hereby preserve the randomisation of the included RCT. However, a selection bias on the meta-level may still appear if the included studies for one intervention use different inclusion criteria than the studies for the other intervention. The resulting unevenly distributed patient characteristics may, if they are associated with the outcome, act as confounders. Therefore the introduced methods for indirect comparisons should be applied only if the results that are going to be pooled are extracted from homogeneous studies. This prerequisite holds not only for the methodological approaches to indirect comparisons but for conventional metaanalyses as well.

These theoretical aspects are supported by the results of the empirical validity check. Adequate numbers of data were available to support the hypothesis that – provided a homogeneous pool of studies – adjusted indirect comparisons may arrive at the same results as direct comparisons.

Likewise a high validity can be ascribed to MTC, if they include head-to-head studies with the interventions of interest. The validity of metaregression-analyses, MTC without included head-to-head trials and the rarely used other methods cannot be appraised yet due to a limited number of available applications.

2. Number of therapies to compare

If only two interventions are to be compared indirectly the adjusted indirect comparison seems to be the most appropriate methodological approach considering the validity data and the limited methodological effort. If more than two interventions are to be compared, only a MTC is applicable to rank them in order of their efficacy.

3. If results from head-to-head trials are to be included

Beside MTC the three other methods for indirect comparisons also provide methodological extensions for the inclusion of head-to-head trials into an indirect comparison. However there haven't been sufficient data for a check of their validity. It can only be stated yet that MTC which include head-to-head trials yield similar results as the head-to-head trial(s) alone. Their additional advantage is the possible increase in precision of the effect estimate by combining the results of direct and indirect comparisons.

4. Heterogeneous trials

The indirect comparison by metaregression-analysis cannot yet be regarded a sufficiently validated method that trustworthily adjusts for factors that cause heterogeneity. Likewise adjusting for covariates in MTC by introduction of inconsistency factors has not been validated due to the limited number of applications. In conclusion: If considerable heterogeneity is present among the trials, the risk of bias in indirect comparisons is high – regardless of what methodological approach is used. In cases of low heterogeneity a conservative estimate may be calculated by a random effects model. Fixed effects models should only be applied in homogenous pools of studies. Both models are applicable in all methodological approaches for indirect comparisons described.

Conclusions

There are a number of methodological approaches available for indirect comparisons which differ in their ability to summarize the evidence from different pools of studies.

The empirical investigation reveals that mainly the results of unadjusted indirect comparisons differ from the results of direct comparisons. The other indirect methods may provide concordant results with direct comparisons, especially if the summarized studies are characterized by low heterogeneity. For that reason adjusted indirect comparisons, metaregression-analyses and MTC should only be used when study results are homogeneous. In the context of HTA and the development of clinical guidelines they are valuable tools, if direct evidence for a comparison of efficacy of interventions is not available.

Before indirect comparisons can be applied more broadly, it remains to be defined at which amount of heterogeneity (and inconsistency) they provide effect estimates of acceptable validity - because a perfectly homogeneous pool of studies is rarely found in real life.

Hauptdokument

1 Gesundheitspolitischer Hintergrund

Entscheidungen über die Kostenübernahme für medizinische Maßnahmen durch die Solidargemeinschaft werden zunehmend auf der Basis von Nutzen- oder Kosten-Nutzen-Bewertungen im Rahmen von Health Technology Assessments (HTA) getroffen (vergleiche Sozialgesetzbuch (SGB) V § 35b und § 139). Dabei ist der Nutznachweis mithilfe von Studien zu führen, deren Ergebnisse wenig anfällig für systematische Verzerrungen (Bias) sind. Derzeit besteht national und international weitgehender Konsens, dass diese Bedingung am ehesten für randomisierte kontrollierte Studien (RCT) erfüllt ist^{5, 7, 12}. RCT sind es dann auch, die herangezogen werden, um vergleichende Wirksamkeitsbewertungen zwischen Therapieoptionen durchzuführen. Goldstandard sind in diesem Fall direkte Vergleiche dieser Therapieoptionen in sogenannten Head-to-head-Studien: Studien, in denen die zu vergleichenden Therapieoptionen direkt gegeneinander getestet werden¹⁶⁷.

Mit dem medizinisch-wissenschaftlichen Fortschritt steigt jedoch die Zahl der Therapieoptionen für eine bestimmte Indikation, ohne dass die regulativen Vorgaben eine sorgfältige Prüfung ihrer Wirksamkeit gegen bereits etablierte Therapieformen vorsehen. In der Pharmakotherapie reicht häufig ein Vergleich der Wirksamkeit neuer Arzneistoffe gegen Placebo für eine Zulassung aus. Daher sehen Arzneimittelhersteller bislang nur selten Veranlassung, ihre neuen Arzneistoffe gegen sich bereits auf dem Markt befindende Arzneistoffe zu testen²¹⁷. Bei multiplen Behandlungsmöglichkeiten für dieselbe Indikation wird es in der Regel keine RCT geben, die alle verfügbaren Therapiemöglichkeiten parallel, d. h. direkt, einander gegenüberstellen. Wirksamkeitsvergleiche sind daher zunehmend auf indirekte Vergleiche angewiesen.

Viele Fragen zur Validität indirekter Vergleiche sind aber bisher noch ungeklärt¹⁰⁴. Sowohl die meisten HTA-Institutionen¹² als auch die Cochrane Collaboration⁸ verzichten daher derzeit auf die Durchführung von formalen indirekten Vergleichen und beschränken ihre Aussagen auf direkt vergleichende Studien bzw. geben narrative Informationszusammenfassungen, wenn keine direkt vergleichenden Studien vorhanden sind^{145, 192, 193, 247}. Dies hat zur Folge, dass Wirksamkeitsvergleiche, wenn keine Head-to-head-Studien vorliegen, nicht erstellt werden können, weil die vorhandene Evidenz nicht nutzbar gemacht wird.

Indirekte Vergleiche gehören zu den jüngsten Weiter- und Neuentwicklungen von Metaanalysetechniken. Die Aufmerksamkeit der wissenschaftlichen Öffentlichkeit für diese Methoden wurde erst durch die Publikation von Bucher et al.⁵⁵ 1997 geweckt. Die erste und bisher einzige umfassende systematische Übersicht und auch Validierung der bis 2001 vorhandenen Methoden zur Durchführung indirekter Vergleiche wurde in einem HTA-Bericht von Glenny et al.¹⁰⁹ 2005 vorgenommen. Die Weiterentwicklung und Verwendung von Methoden zur Durchführung von indirekten Vergleichen hat seit 2001 kontinuierlich zugenommen. Daher soll der vorliegende HTA-Bericht die Arbeit von Glenny et al.¹⁰⁹ fortsetzen und eine Bestandsaufnahme liefern, welche Methoden 2008 zur Verfügung stehen und wie häufig sie bereits eingesetzt wurden. Da seit 2001 Methodenerweiterungen stattgefunden haben und mit dem Mixed treatment comparison (MTC) auch eine neue Methode hinzugekommen ist, ist es von Interesse zu betrachten, ob sie die alten Techniken verdrängt haben, oder ob sie in Ergänzung zu den älteren eingesetzt werden. Alle Methoden für indirekte Vergleiche werden anschließend einer Validitätsprüfung unterzogen, um die Aussagekraft ihrer Ergebnisse zu überprüfen. Als geeigneter Komparator bieten sich Resultate aus direkten Vergleichen an.

Für Methoden, die Evidenz aus direkten Vergleichen mit Evidenz aus indirekten Vergleichen zusammenführen, soll das Ergebnis dem Ergebnis von metaanalytisch zusammengefassten Head-to-head-Studien gegenübergestellt werden. Dadurch wird die Vermutung überprüft, dass durch die Hinzunahme von Evidenz aus indirekten Vergleichen der Punktschätzer erhalten bleibt, während die Präzision des Ergebnisses des direkten Vergleichs erhöht wird¹⁰⁹.

In den Untersuchungen dieses HTA-Berichts wird nur die Wirksamkeit unter RCT-Bedingungen (efficacy) betrachtet. Durch den gewählten Fokus auf den Vergleich der Wirksamkeit von Therapieoptionen werden zudem die Erhebung und Bewertung von unerwünschten Wirkungen (zur Bestimmung des Nettonutzens) sowie Wirtschaftlichkeitsbetrachtungen nicht berücksichtigt^{202, 257}.

2 Wissenschaftlicher Hintergrund

2.1 Evidenzbasis für indirekte Vergleiche

Unter indirekten Vergleichen versteht man die Gegenüberstellung verschiedener Interventionen mithilfe eines gemeinsamen Komparators. Dabei kann es sich um den Vergleich mit einem aktiven Komparator (meistens Standardtherapie), oder um Vergleiche mit Placebo handeln. Indirekte Vergleiche werden herangezogen, um die Wirksamkeit von solchen Interventionen miteinander vergleichbar zu machen, für die es keine direkte Gegenüberstellung (Head-to-head-Studien) gibt⁸. Im Kontext des vorliegenden HTA werden ausschließlich Vergleiche auf der Basis von RCT betrachtet.

2.1.1 Einsatzszenarien

Abbildung 1 stellt Szenarien dar, welche Typen von vergleichenden Studien potenziell zur Verfügung stehen, wenn die Wirksamkeit verschiedener Therapien verglichen werden soll. Sie entstammt unwesentlich modifiziert einer Veröffentlichung der britischen HTA-Organisation NICE (National Institute for Health and Clinical Excellence)²²⁶ und führt in die Thematik der indirekten Vergleiche in geeigneter Weise ein. An diesen Beispielen wird deutlich, welche Möglichkeiten des direkten und indirekten Vergleichs bei spezifischen Konstellationen Anwendung finden können. Es werden bereits verschiedene Methoden für die Durchführung des indirekten Vergleichs in diesen Einsatzszenarien genannt. Sie werden im Ergebnisteil ausführlich vorgestellt (vergleiche Kapitel 5.3 „Beschreibung der indirekt vergleichenden Methoden“).

Die Buchstaben A, B und C stellen die Therapieoptionen dar, deren Wirksamkeit verglichen werden soll. Die grau unterlegten Buchstaben X und Y symbolisieren eine aktive oder eine Placebokontrolle, über die ein indirekter Vergleich ermöglicht wird. Eine verbindende Linie zwischen den Buchstaben symbolisiert das Vorhandensein von RCT, die diese beiden Therapieoptionen direkt miteinander vergleichen.

Diese abstrakte Darstellung ist eine Hilfestellung zur Veranschaulichung von Evidenznetzwerken. Ein vollständiges Evidenznetzwerk beinhaltet alle randomisierten kontrollierten klinischen Studien, die die Wirksamkeit dieser Interventionen jemals getestet haben, unabhängig davon, welche Vergleichsinterventionen ihnen gegenübergestellt wurden¹⁶⁰.

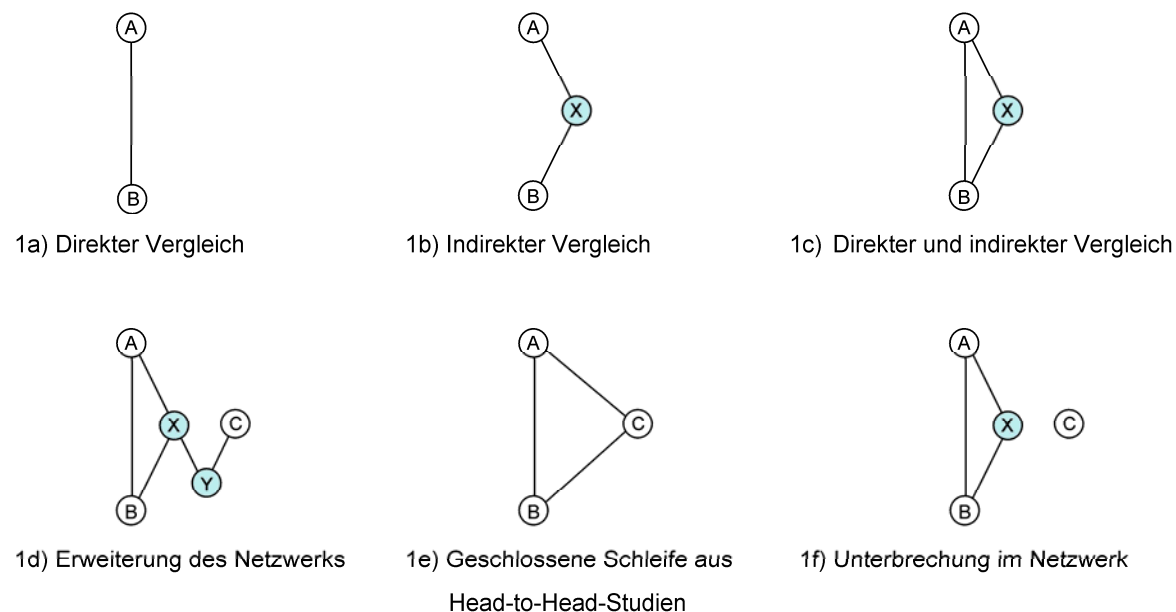


Abbildung 1: Einsatzszenarien für direkte und indirekte Vergleiche

Modifizierte Abbildung aus: Briefing paper for methods review workshop on evidence synthesis (indirect and mixed treatment comparisons) des NICE (National Institute for Health and Clinical Excellence)²²⁶; auch veröffentlicht in Sutton et al.²²⁷.

Direkter Vergleich (Abbildung 1a)

Der bereits beschriebene Fall der Head-to-head-Studie stellt den Goldstandard beim Wirksamkeitsvergleich von therapeutischen Interventionen in der evidenzbasierten Medizin dar. Der Therapievergleich kann auf Basis eines direkten Vergleichs vorgenommen werden. Sind mehrere Head-to-head-Studien verfügbar, können ihre Ergebnisse in einer Metaanalyse zusammengefasst werden.

Indirekter Vergleich (Abbildung 1b)

Dieser Fall tritt bei placebokontrollierten Arzneimittelstudien häufig auf: Es liegen Vergleiche der Medikamente A und B jeweils mit Placebo vor, aber keine direkten Vergleichsstudien von A mit B. Sollen, wie in diesem Beispiel, nur zwei Therapieoptionen miteinander verglichen werden, so können einfache und gut etablierte Methoden des indirekten Vergleichs, wie z. B. der adjustierte indirekte Vergleich nach Bucher⁵⁵, angewendet werden²¹⁷. Sollen mehr als zwei Therapieoptionen in einer Metaanalyse miteinander über einen gemeinsamen Komparator gegenübergestellt werden, so ist dafür eine komplexere und weniger etablierte Methode des indirekten Vergleichs verfügbar, der sogenannte MTC^{104, 157, 160}.

Direkter und indirekter Vergleich (Abbildung 1c)

Dieser Fall ähnelt Abbildung 1a. Zusätzlich zu der Head-to-head-Studie in Abbildung 1a sind jetzt noch zusätzliche Studien für A und B verfügbar, die diese mit einem gemeinsamen Komparator, wie z. B. Placebo, vergleichen. Diese placebokontrollierten Studien würden auch den indirekten Vergleich von A und B über X möglich machen. Kontrovers ist, ob auf diese Informationen über die Wirksamkeit von A und B aus den placebokontrollierten RCT verzichtet werden sollte^{226, 227}. Es sind vier verschiedene Varianten möglich, wie man aus diesem Evidenznetzwerk zu einer vergleichenden Wirksamkeitsbewertung von A versus B kommen kann:

- a) Entscheidung nur auf Basis der Head-to-head-Studie/n fällen⁸.
- b) Entscheidung nur auf Basis des indirekten Vergleichs fällen. Diese Option widerspricht der gängigen Lehrmeinung, hat aber ihre Berechtigung, falls es sich bei der/den Head-to-head-Studie/n um (eine) qualitativ minderwertige Studie/n handelt und die placebokontrollierten Studien von A und B qualitativ hochwertig sind²¹⁶.
- c) Sowohl den direkten Vergleich über die Head-to-head-Studie/n als auch den indirekten Vergleich über X getrennt durchführen und mithilfe eigener Überlegungen die Ergebnisse aus beiden Ansätzen narrativ zu einer Empfehlung über die vergleichende Wirksamkeit zusammenführen²¹⁶.
- d) In einer Metaanalyse die Ergebnisse der RCT, die A und B direkt gegeneinander testen und die Ergebnisse der RCT, die A bzw. B gegen den gemeinsamen Komparator testen, poolen. Für dieses Vorgehen gibt es verschiedene Methoden, in denen dem Charakter der Head-to-head-Studien ein unterschiedlich großes Gewicht zukommt. Im Ergebnis kann die Präzision der Aussage der Head-to-head-Studie durch Hinzunahme der Evidenz auf Basis indirekter Vergleiche erhöht werden^{122, 160}.

Erweiterung des Netzwerks (Abbildung 1d)

In diesem Beispiel wird eine dritte Therapieoption C von Interesse eingeführt. Sie ist über den Komparator Y an das Evidenznetzwerk angeschlossen. Komparator Y ist wie X eine Therapieoption, die für die Forschungsfrage nicht von Interesse ist, und ist nicht identisch mit X. Für den Anschluss von C an das Evidenznetzwerk ist eine direkt vergleichende Studie der beiden Komparatoren X und Y erforderlich.

Es ist mit einer Methode des indirekten Vergleichs (MTC) nun möglich, eine Aussage über die komparative Wirksamkeit von A, B und C zu machen^{226, 227}.

Geschlossene Schleife aus Head-to-head-Studien (Abbildung 1e)

Es ist schwer in einer geschlossenen Schleife einer Evidenzkette zu sagen, welche Teile beim Vergleich von A versus B versus C zu einem direkten oder einem indirekten Vergleich beitragen. Dies ist abhängig von der Forschungsfrage, die festlegt, welche Therapieoptionen miteinander verglichen werden sollen. Soll ein paarweiser Vergleich zweier dieser drei Optionen durchgeführt werden, entspräche dies Szenario 1c. Je nach Fragestellung verkörpert eine der drei Therapieoptionen den gemeinsamen Komparator für den indirekten Vergleich.

Ein direkter Vergleich wäre aber auch möglich. Sollen jedoch alle drei Therapieoptionen in eine Rangfolge hinsichtlich ihrer Wirksamkeit gebracht werden, ist keine direkte Gegenüberstellung mehr möglich, da keine dreiarmlige Studie mit den Armen A, B und C existiert. Für diese Aufgabenstellung gibt es mit dem MTC aber eine Methode des indirekten Vergleichs, um diese Rangfolge zu erstellen.

Es wird deutlich, dass je mehr Therapieoptionen von Interesse sind, es desto unwahrscheinlicher ist, dass die Fragestellung mit direkter Evidenz beantwortet werden kann, da hierfür Head-to-head-Studien mit entsprechend vielen Armen, wie Therapieoptionen, notwendig sind.

Unterbrechung im Netzwerk (Abbildung 1f)

In diesem Beispiel ist die dritte Therapieoption C noch nicht an das Evidenznetzwerk angeschlossen. Diese mag neu und deshalb noch nicht mit A, B oder X in Studien verglichen sein. C steht somit außerhalb des Evidenznetzwerkes. Unabhängig von der Form/Geometrie des Evidenznetzwerkes, gehören Therapien, die nicht in Studien gegen Komparatoren, die im Evidenznetzwerk vertreten sind, getestet wurden, dem Evidenznetzwerk nicht an.

Es existiert keine Methode, die einen Wirksamkeitsvergleich von C mit A und/oder B ermöglichen könnte.

2.1.2 Geometrie von Evidenznetzwerken

Da es für die Durchführung indirekter Vergleiche mit einem MTC hilfreich ist, alle RCT in einem Evidenznetzwerk darzustellen und in einer Analyse zu erfassen, richtet sich aktuelle Forschung auf die Struktur dieser Netzwerke. Georgia Salanti hat mit Co-Autoren zwei aktuelle Publikationen^{204, 683} zu diesem Thema verfasst und führt den Begriff „Geometrie“ von Therapienetzwerken ein: „Die Geometrie eines Netzwerks kann den erweiterten klinischen Kontext der Evidenz wiedergeben und kann geformt sein durch rationale Entscheidungen für Therapievergleiche (...)“²⁰⁵. Sie stellt die drei wichtigsten geometrischen Formen vor, die in Abbildung 2 für ein Netzwerk aus sechs Therapieoptionen A bis F dargestellt sind.

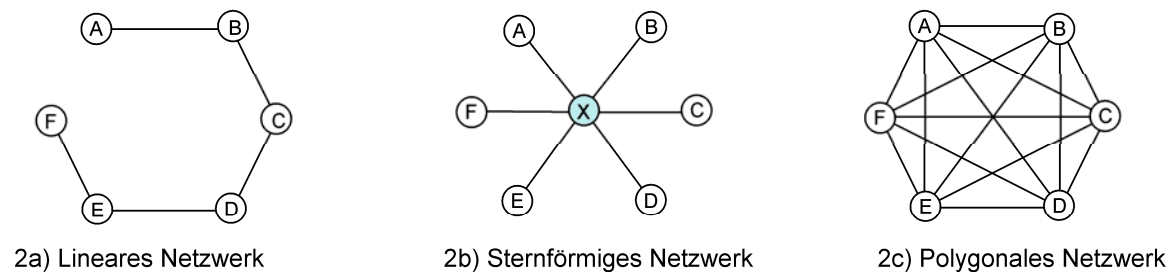


Abbildung 2: Extremata von geometrischen Strukturen in Evidenznetzwerken

Lineares Netzwerk (Abbildung 2a)

Wenn neue Therapien nur gegen das jeweils vorher entwickelte getestet werden und ältere Therapien in Studien nicht mehr berücksichtigt werden, entsteht eine lineare Struktur²⁰⁴. Eine solche Struktur kann sich herausbilden, wenn ein großer therapeutischer Fortschritt durch die jede neue Therapie erzielt wird, sodass es als unakzeptabel angesehen wird, gegen ältere, weniger wirksame Therapien zu testen²⁰⁴. Mit direkten Vergleichen kann die neueste Therapie F nicht mit den Therapien A bis D verglichen werden, da nur Head-to-head-Studien von F versus E vorliegen. Ein konventioneller indirekter Vergleich kann die komparative Wirksamkeit von F versus D ermitteln, indem er E als gemeinsamen Komparator ansieht. Nur der MTC kann das gesamte Netzwerk berücksichtigen und auch den Wirksamkeitsvergleich von F versus C, B oder A ermitteln.

Sternförmiges Netzwerk (Abbildung 2b)

Ein sternförmiges Netzwerk existiert in Indikationsfeldern, in denen entweder immer gegen Placebo (als X in Abbildung 2b dargestellt) getestet wird, oder eine Standardtherapie existiert²⁰⁵, die stets als Referenz für neue Therapien gewählt wird (zweite Möglichkeit für das X).

Dies ist z. B. in Zulassungsstudien der Fall, in denen gegen die Standardtherapie oder bei ethischer Vertretbarkeit gegen Placebo getestet wird. Dafür, dass Head-to-head-Studien, wie z. B. für A versus B nicht durchgeführt werden, können verschiedene Gründe vorliegen. Es kann z. B. die Standardtherapie X sehr etabliert und noch nicht durch die neuen Therapien A bis F verdrängt worden sein²⁰⁴. Ein anderer Grund könnte darin liegen, dass keine Sponsorengelder für die Durchführung komparativer Studien freigegeben werden, da die Hersteller der aktiven Therapien fürchten, dass sich ihre Therapie einer anderen unterlegen zeigen könnte²⁰⁵.

Das Sternnetzwerk stellt eine Erweiterung des unter 1b dargestellten prototypischen Beispiels für einen indirekten Vergleich dar. Neben A und B sind nun auch die Therapieoptionen C, D, E und F bisher nur gegen den gemeinsamen Komparator X getestet worden. Dementsprechend sind für das Sternnetzwerk die Methoden des indirekten Vergleichs geeignet. Methoden des indirekten Vergleichs können paarweise Vergleiche, wie etwa A versus B, B versus C usw. erbringen. Der MTC kann darüber hinausgehend alle Therapieoptionen von A bis F in eine Rangfolge entsprechend ihrer Wirksamkeit bringen.

Poligonales Netzwerk (Abbildung 2c)

Zu einem poligonalem Netzwerk kommt es, wenn alle Therapieoptionen in Head-to-head-Studien gegeneinander getestet werden²⁰⁴. Solche Strukturen entstehen in Indikationsfeldern mit multiplen Therapieoptionen vergleichbarer Wirksamkeit und hoher Forschungsaktivität. Ein Beispiel, das der polygonalen Struktur nahe kommt, ist die Evidenzlage zur Wirksamkeit von Arzneimitteln der ersten Wahl bei Bluthochdruck^{189, 205}. In der polygonalen Struktur fehlen aber meistens einige Verbindungen bzw. Head-to-head-Studien. Das Ergebnis einer zukünftig durchgeführten Head-to-head-Studie kann durch einen indirekten Vergleich abgeschätzt werden²³⁴. Der MTC kann auch in solch komplexen Netzwerken alle Therapieoptionen in eine Rangfolge nach ihrer Wirksamkeit bringen.

Evidenznetzwerke in der Realität

Abbildung 2 stellt die Extremformen von geometrischen Strukturen in Evidenznetzwerken dar. In der Realität werden meistens Mischungen aus den drei verschiedenen Strukturen gefunden, wobei fehlende Bindeglieder ein häufiges Problem sind²⁰⁵. So kann sich ein sternförmiges Netzwerk, nachdem zunächst nur Vergleiche der Therapien mit Placebo durchgeführt wurden, mit der Zeit durch hohes Interesse an komparativen Vergleichen in ein polygonales Netzwerk verwandeln. Der Komplexitätsgrad von Evidenznetzwerken mit multiplen Therapieoptionen kann so schnell erhebliche Ausmaße annehmen. Bis auf den MTC, der auch komplexe Netzwerke als Ganzes erfassen kann, können mit allen übrigen in Kapitel 5.3 „Beschreibung der indirekt vergleichenden Methoden“ vorgestellten Methoden nur indirekte Vergleiche zweier Therapieoptionen vorgenommen werden.

2.2 Die Validität von indirekten Vergleichen

Grundsätzlich ist zu unterscheiden, ob der indirekte Vergleich rein narrativ aus den Studienergebnissen abgeleitet wird, oder ob er unter Zuhilfenahme von Metaanalysetechniken durchgeführt wurde.

Narrative indirekte Vergleiche

Narrative indirekte Vergleiche werden meistens dann durchgeführt, wenn aufgrund einer heterogenen oder begrenzten Studienlage keine Metaanalysen möglich sind. Inwieweit die Ergebnisse von narrativen indirekten Vergleichen als valide genug gelten können, um aus ihnen Entscheidungen zu Therapieempfehlungen abzuleiten, wird in Kapitel 5.5 „Validitätsprüfung indirekt vergleichender Methoden“ diskutiert.

Indirekte Vergleiche mit Metaanalysetechniken – Nicht-adjustierter indirekter Vergleich

Die Validität von indirekten Vergleichen mit Metaanalysetechniken hängt von der Methodik ab, mit der sie durchgeführt werden. Häufig wurde von methodischer Seite argumentiert, dass indirekte Vergleiche von RCT nicht mehr den Status eines randomisierten Vergleichs erfüllen, sondern Beobachtungsstudien darstellen, die auch die Gefahr von für Beobachtungsstudien typischen Biasformen beinhalten¹⁸⁵. Hierdurch könne es zu einer Über- oder Unterschätzung von Therapieeffekten kommen. Für Methoden, die diese Form der Biasanfälligkeit aufweisen, wurde durch Glenny et al.¹⁰⁹ die Nomenklatur nicht- (bzgl. Kontrollgruppe) adjustierter indirekter Vergleiche eingeführt.

Das Hauptkennzeichen dieser Vergleichsart ist, dass nur die Ergebnisse der Verumgruppen der RCT gepoolt werden. Die Ergebnisse der Kontrollgruppen werden nicht mit in der Analyse berücksichtigt¹⁰⁴. Deshalb wird von einem Bruch der Randomisation gesprochen¹⁰⁹. Die Strukturgleichheit der Stichprobe, die bei der zufälligen Verteilung der Probanden auf die Verum- und Kontrollgruppe entsteht, geht bei diesem Verfahren verloren. Effekte von Confoundern werden nicht mehr dadurch abgeschwächt, dass sie sich in der Verum- bzw. in der Vergleichsgruppe gleich stark auswirken und bei Differenzbildung der Therapieoutcomes von beiden Gruppen eliminiert werden.

Indirekte Vergleiche mit Metaanalysetechniken – Adjustierter indirekter Vergleich

Das Gegenstück zum nicht-adjustierten indirekten ist der adjustierte Vergleich. Er adjustiert gegenüber den Kontrollgruppen, indem er für jede Einzelstudie die Differenz des Therapieeffekts aus der Verum- sowie der Kontrollgruppe bildet und erst diese Differenzen im Rahmen einer Metaanalyse poolt⁵⁵. Durch dieses Verfahren wird die methodische Qualität, die durch die Randomisation im RCT gegeben ist, erhalten. Aus diesem Grund empfohlen bereits Glenny et al.¹⁰⁹ die nicht-adjustierte Methode zur Durchführung von indirekten Vergleichen nicht mehr zu verwenden und nur noch adjustierte Verfahren für indirekte Vergleiche einzusetzen¹⁰⁹.

Inwieweit die nach 2001 neu entwickelten Methoden den Randomisationseffekt der Einzelstudien bewahren und ihre Ergebnisse denen eines direkten Vergleichs entsprechen, ist Gegenstand dieses Reviews.

Annahmen in indirekten Vergleichen

Eine Annahme ist allen Metaanalysen und damit auch den indirekten Vergleichen im Rahmen von Metaanalysen gemein: Es wird davon ausgegangen, dass die Variabilität zwischen den Studienergebnissen durch den Zufall bedingt ist. Es darf in anderen Worten keine signifikante Heterogenität zwischen den Studien bestehen. Für indirekte Vergleiche bedeutet dies am Beispiel 1b (siehe Abbildung 1) erklärt, dass der durch direkten Vergleich ermittelte Therapieeffektunterschied zwischen A und B in den Studien von A versus X und B versus X auch ermittelt worden wäre, wenn sie dreiarstig mit den Armen A, B und X gewesen wären. Kann dies angenommen werden, wird von konsistenter Evidenz im Netzwerk gesprochen¹⁶⁰. Kann dies nicht angenommen werden, weist dieses Evidenznetzwerk entsprechend Inkonsistenz in der Evidenz auf. Erste Methoden wurden entwickelt, Inkonsistenz in einem Evidenznetzwerk zu identifizieren, zu quantifizieren und in Analysen zu berücksichtigen¹⁶⁰.

Es soll für die verschiedenen Methoden für indirekte Vergleiche geklärt werden, ob über die Annahme von Konsistenz des Therapieeffekts über die Studien hinaus noch weitere Voraussetzungen vorliegen müssen, damit sie angewendet werden können.

Validität von indirekten Vergleichen gemessen an den Ergebnissen von direkten Vergleichen

Von besonderem Interesse ist grundsätzlich die Validität von indirekten Vergleichen im Verhältnis zu direkten Vergleichen. Direkte Vergleiche werden bisher als der Goldstandard in der vergleichenden Bewertung von Therapien angesehen. Die Validität der Methoden zur Durchführung indirekter Vergleiche soll deshalb der Validität der Methoden zur Durchführung direkter Vergleiche gegenüber gestellt werden. Für die Methoden, die Evidenz aus direkten Vergleichen mit Evidenz aus indirekten Vergleichen poolen, soll die Vermutung überprüft werden, ob durch die Hinzunahme der Evidenz aus indirekten Vergleichen die Präzision des Ergebnisses der direkten Evidenz erhöht wird¹⁰⁹.

2.3 Abgrenzung indirekter Vergleich – Cross design-Synthese

Definition der Cross design-Synthese

Die Cross design-Synthese⁹⁰ beschäftigt sich mit der Zusammenführung von Ergebnissen aus Studien unterschiedlicher Designs (z. B. aus RCT und Beobachtungsstudien)

Einsatzgebiete der Cross design-Synthese

Der Bedarf für diese Methode entstand aus der Problematik, dass sich Ergebnisse aus klinischen Studien nicht eins zu eins in die Praxis umsetzen lassen, weil in der Praxis andere Bedingungen herrschen, als sie der RCT simuliert hat.

Diese liegen z. B. in der Patientencharakteristik, die in der Praxis (häufig Alte und zur Hälfte Frauen) häufig nicht der des RCT (häufig Männer mittleren Alters) entsprechen¹. Routinedaten geben eher die Praxisrealität wieder. Allerdings sind Beobachtungsstudien mit Routinedaten keine randomisierten Vergleiche und den entsprechenden Biasformen ausgeliefert²¹⁰. Die Cross design-Synthese adjustiert deshalb hinsichtlich Patientencharakteristika und anderen erwarteten Biasquellen, die potenzielle Confounder darstellen⁹⁰. Die Gefahr von für Beobachtungsstudien typischen Biasformen ist auch in indirekten Vergleichen präsent und muss stets bedacht und diskutiert werden¹⁰⁹.

Abgrenzung von Cross design-Synthese und indirektem Vergleich

In der Adjustierung bezüglich Confoundern ähneln sich die Cross design-Synthese und der indirekte Vergleich mit Metaregression (vergleiche Kapitel 5.3.4 „Metaregression“). In der Methodentheorie besteht die weitere Parallele, dass die Cross design-Synthese und der MTC zu den sogenannten Multiparametersynthesen¹¹⁸ gezählt werden, da sie komplexere Datensätze als konventionelle Metaanalysen analysieren^{22, 229}. Nichtsdestotrotz sind die Cross design-Synthese und die Methoden für indirekte Vergleiche unterschiedlicher Natur in ihrer Intention und Anwendbarkeit. Die Intention der Cross design-Synthese ist es, die Generalisierbarkeit der Ergebnisse von RCT durch Hinzunahme von Evidenz aus nicht-randomisierten Studien zu erhöhen. Indirekte Vergleiche streben dagegen bisher nicht auf die Generalisierbarkeit ihrer Ergebnisse in der Praxis ab, sondern wollen Lücken an Therapievergleichen füllen, bei denen keine Head-to-head-Studien durchgeführt wurden. Indirekte Vergleiche werden also dort durchgeführt, wo Head-to-head-Studien fehlen, im Gegensatz zu Cross-Design-Analysen, die auf Head-to-head-RCT aufbauen und diese durch Evidenz aus Beobachtungsstudien ergänzen²².

Die Methoden für die Cross design-Synthese und die indirekten Vergleiche sind dementsprechend sehr unterschiedlich. Aus diesem Grund werden die Methodik und die Validität der Cross design-Synthese in diesem HTA über indirekte Vergleiche nicht behandelt. Für eine Übersicht über die Einsatzgebiete, Stärken und Schwächen von sowohl MTC als auch Cross design-Synthese sei auf den Review von Ades et al.²¹ verwiesen.

3 Forschungsfragen

Auf der Grundlage der Vorarbeiten, insbesondere der Arbeitsgruppe um Glenny et al.¹⁰⁹, gibt der vorliegende Bericht anhand von fünf Forschungsfragen einen Überblick über den aktuellen Kenntnis- und Erfahrungsstand mit Methoden für indirekte Therapievergleiche. Dabei wird im Rahmen einer systematischen Literaturanalyse auf indirekte Vergleiche in systematischen Reviews fokussiert, die Aussagen zur Wirksamkeit therapeutischer Verfahren machen und hierzu die Ergebnisse von RCT heranziehen.

1. Welche Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen existieren bisher und wie und unter welchen Voraussetzungen können sie eingesetzt werden?
2. Welche Methoden des indirekten Wirksamkeitsvergleichs therapeutischer Interventionen wurden wie häufig in bisher publizierten systematischen Reviews eingesetzt?
3. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen und führen sie zu den gleichen Schlussfolgerungen über Therapieeffektunterschiede?
4. Wie ist die Validität indirekter Vergleiche gegenüber den Ergebnissen von direkten Vergleichen zu beurteilen, wenn in den indirekten Vergleich zusätzlich Ergebnisse von Head-to-head-Studien eingeschlossen werden?
5. Lässt sich ein Goldstandard identifizieren, nach dem indirekte Vergleiche der Wirksamkeit von therapeutischen Interventionen vorgenommen werden sollten?

4 Methodik

4.1 Literaturrecherche

Die im Rahmen des Projekts durchzuführenden Literaturrecherchen haben drei Ziele:

1. Auffinden von Methodenpapieren und Methodenbeschreibungen zur Durchführung von indirekt vergleichenden Bewertungen der Wirksamkeit von Interventionen.
2. Auffinden von systematischen Übersichtsarbeiten, in denen sowohl indirekt als auch direkt vergleichende Wirksamkeitsbewertungen von Interventionen berichtet werden.
3. Auffinden von systematischen Übersichtsarbeiten, die ausschließlich indirekt vergleichende Bewertungen der Wirksamkeit von Interventionen vornehmen.

Hierzu werden unterschiedliche Recherchestrategien eingesetzt:

4.1.1 Recherche in elektronischen Datenbanken

Deutsches Institut für medizinische Dokumentation und Information (DIMDI)-Datenbanken

Die erste elektronische Recherche wird durch eine externe Expertin (Frau Edda Bhattacharjee) der Deutschen Agentur für Health Technology Assessment (DAHTA) in einer Auswahl von Literaturdatenbanken des DIMDI vorgenommen. Die recherchierten Datenbanken umfassen das medizinische Datenbankangebot des DIMDI. Der Recherchezeitraum wird, aufgrund des Vorliegens eines qualitativ hochwertigen systematischen Reviews zur Thematik von Glenny et al.¹⁰⁹, der die Publikationsjahre bis 1999 abdeckt, auf den Publikationszeitraum ab 1999 beschränkt. Die Recherchestrategie wird, in Anlehnung an die im vorliegenden Review dokumentierte Strategie, in vier terminologischen Modulen („indirekte Vergleiche“, „RCT“, „besondere Studiencharakteristika“ und „Metaanalyse“) konstruiert, die additiv miteinander verknüpft werden. Die Suche wird auf systematische Reviews (mit oder ohne Metaanalyse) über (überwiegend) RCT beschränkt, um damit die spätere Validitätsprüfung der Methoden zur Durchführung indirekter Vergleiche nicht durch zusätzliche Biasquellen aus dem Studienmaterial zu erschweren. Keine Einschränkung dagegen wird in der Art der therapeutischen Interventionen, die in den Reviews verglichen werden, vorgenommen. So werden auch Vergleiche z. B. zwischen einer Operationstechnik und einer Pharmakotherapie eingeschlossen, sofern ihre Wirksamkeit in RCT untersucht wird. Datenbanken und Suchstrategie sind in Tabelle 8 und Tabelle 12 dokumentiert.

ISI Web of Knowledge®

Die zweite elektronische Suche wird im ISI Web of Knowledge® (umfasst die Datenbanken Science Citation Index Expanded, Social Science Citation Index und Arts and Humanities Citation Index) vorgenommen. Ziel der Recherche ist das Auffinden von Referenzen, in denen drei relevante Schlüsselpublikationen^{55, 109, 217} zitiert werden. Datenbanken und Suchstrategie sind in Tabelle 9 und Tabelle 13 dokumentiert.

4.1.2 Handsuchen

Die Handsuchen zielen in erster Linie auf das Auffinden von Methodenpapieren und -beschreibungen zur Durchführung von indirekt vergleichenden Bewertungen der Wirksamkeit von Interventionen. Da vergleichende Wirksamkeits- und Nutzenbewertungen vor allem von HTA-Institutionen vorgenommen werden, bilden ihre Internetseiten die Ausgangsbasis für die Handsuchen. Dabei beschränkt sich die Handsuche auf Institutionen, die dem International network of agencies for Health Technology Assessment (INAHTA) angehören. Diese Institutionen werden mindestens zur Hälfte ihres Budgets aus öffentlich-rechtlichen Mitteln finanziert. Eine Liste der Institutionen und die Anzahl der gefundenen Papiere sind in Tabelle 10 und Tabelle 14 dokumentiert. Wegen des unterschiedlichen Aufbaus der Internetseiten lässt sich keine einheitliche Suchstrategie berichten.

Auf den Internetseiten der Cochrane Collaboration of Systematic Reviews (Cochrane Collaboration) wird ebenfalls eine Handsuche nach Methodenpapieren durchgeführt. Die Suchstrategien sind in Tabelle 11 sowie die Ergebnisse in Tabelle 15 und Tabelle 16 wiedergegeben.

Um Methodenpapiere und -beschreibungen, die in der medizinisch-wissenschaftlichen Literatur vor 1999 publiziert wurden, aufzufinden, werden die Referenzen einiger Schlüsselpublikationen zu der Thematik sowie im speziellen die Referenzen des Kapitels „Statistical methods for indirect comparisons“ des systematischen Review von Glenny et al.¹⁰⁹ einem Volltextscreening unterzogen. Die Ergebnisse sind in Tabelle 17 dokumentiert. Ebenfalls diesem Review werden alle systematischen Reviews, die sowohl direkte als auch indirekte Vergleiche therapeutischer Interventionen ermöglichen und vor 1999 publiziert wurden, entnommen. Eine Auflistung dieser 28 systematischen Reviews bietet Abschnitt 9.5 „Ergebnisse der Handsuchen“ im Anhang.

4.2 Literaturselektion

Die Auswahl der zu berücksichtigenden Publikationen erfolgt in zwei Schritten: Die erste Auswahl geschieht anhand der übersandten Kurzfassungen (Abstracts), die zweite Selektion anhand der Volltextpublikationen. Beide Schritte werden als Doppelbeurteilung von zwei Reviewern unabhängig voneinander durchgeführt. Diskrepanzen werden im direkten Austausch aufgelöst.

4.2.1 Abstractselektion

Ziel der ersten Selektion ist die Identifikation von potenziell relevanten Artikeln aus den Rechercheergebnissen. Dabei wird auf eine hohe Sensitivität geachtet, d. h. es werden nur solche Referenzen verworfen, die sicher auszuschließen sind. Für die erste Selektion, die auf Basis der Abstracts durchgeführt werden soll, werden folgende, sehr grobe Ein- und Ausschlusskriterien formuliert:

Einschlusskriterien der Abstractselektion

Referenzen werden im Volltext bestellt, wenn sie einem oder mehreren der folgenden Kriterien genügen:

- Die Publikation enthält methodische Ausführungen zum indirekten Vergleich der Wirksamkeit von Interventionen.
- Bei der Publikation handelt es sich um einen systematischen Review (mit oder ohne Metaanalyse), in dem die Ergebnisse von direkten (Head-to-head-Studien) als auch indirekten Vergleichen der Wirksamkeit therapeutischer oder umschriebener präventiver Interventionen berichtet werden.
- Bei der Publikation handelt es sich um einen systematischen Review (mit oder ohne Metaanalyse), indem die Ergebnisse von indirekten Vergleichen der Wirksamkeit therapeutischer Interventionen berichtet werden.

Ausschlusskriterien der Abstractselektion

Ausgeschlossen werden:

- Einzelstudien
- Übersichtsarbeiten, aus deren Kurzfassung nicht hervorgeht, dass es sich um einen systematischen Review mit oder ohne Metaanalyse handelt.
- Übersichtsarbeiten zu diagnostischen Testverfahren oder komplexen Interventionen.
- Publikationen, die ausschließlich als Abstract vorliegen.
- Mehrfachpublikationen mit identischem Inhalt.

In einem zweiten Selektionsschritt werden die Volltexte hinsichtlich ihrer Relevanz zur Beantwortung der Forschungsfragen gesichtet.

4.2.2 Volltextselektion

In der Volltextselektion sollen die zur Beantwortung der Forschungsfragen relevanten Publikationen von den irrelevanten getrennt werden. Damit eine Publikation als irrelevant ausgeschlossen werden kann, muss sie mindestens einer der folgenden acht Ausschlusskriterien zur Begründung zugeordnet werden können. Dabei handelt es sich um eine konkretisierte Version der Ausschlussgründe der Abstractselektion.

Im Anhang in Kapitel 9.6 „Nach Durchsicht im Volltext ausgeschlossene Literatur mit Ausschlussgrund“ wird die in der Volltextselektion ausgeschlossene Literatur, sortiert nach Ausschlussgründen, aufgelistet. Treffen mehr als zwei Ausschlussgründe auf eine Publikation zu, wird subjektiv der am besten passende Ausschlussgrund ausgewählt.

Ausschlusskriterien der Volltextselektion

1. Einzelstudien
2. Übersichtsarbeiten, Editorials und Kommentare, bei denen es sich nicht um systematische Übersichtsarbeiten handelt (Definition eines systematischen Reviews im Glossar).
3. Systematische Übersichtsarbeiten zu diagnostischen Testverfahren und komplexen Interventionen (z. B. im Rahmen umfangreicher Präventionskampagnen).
4. Übersichtsarbeiten, in denen nicht über die Durchführung und die Ergebnisse eines indirekten Vergleichs der Wirksamkeit therapeutischer Interventionen berichtet wird.
5. Publikationen, ausschließlich methodischen Inhalts, die keine Anleitung geben, wie indirekte Vergleiche therapeutischer Interventionen durchzuführen sind.
6. Mehrfachpublikationen mit identischem Inhalt.
7. Kosten-Nutzen-Betrachtungen, für die keine systematische Literaturrecherche durchgeführt wurde.
8. Der Volltext konnte nicht beschafft werden, da der Artikel nicht über „Subito“ bestellbar war, nur in einem Buch erschienen ist oder bisher nur der Abstract oder das Protokoll publiziert wurde.
9. Systematische Übersichtsarbeiten, die den indirekten Vergleich überwiegend auf Basis von nicht-randomisierten oder nicht-kontrollierten Studien durchführten.

Auswertungsstrategie

In der Volltextselektion werden fünf Typen von Publikationen für die weitergehenden Analysen selektiert:

Typ 1: Methodenpapiere

Mithilfe der Methodenpapiere soll die Grundlage zur Beantwortung der ersten Forschungsfrage gelegt werden. Dafür werden die verschiedenen Methoden identifiziert und den Publikationen Informationen über die Durchführung, Anwendungsmöglichkeiten und -beschränkungen sowie Stärken und Schwächen dieser Methoden entnommen. Ziel ist, für jede der möglichen Methoden eine möglichst vollständige Methodenbeschreibung mit den in Kapitel 9.8.1 „Checkliste zur Beschreibung und Charakterisierung von Methoden zur Durchführung von indirekten Vergleichen“ benannten Kriterien zu erhalten.

Typ 2: Systematische Reviews ohne Metaanalyse, die nur einen indirekten Vergleich vornehmen.

In diesen systematischen Reviews wird die Methode des narrativen indirekten Vergleichs angewendet. Sie werden zur Methodenbeschreibung herangezogen und zur Beantwortung der Frage nach der Häufigkeit der verschiedenen Methoden gezählt.

Typ 3: Systematische Reviews ohne Metaanalyse, die sowohl einen direkten als auch einen indirekten Vergleich vornehmen.

Wie die Reviews des Typs 2 werden diese Publikationen der Methode des narrativen indirekten Vergleichs zugeordnet. Zur Beurteilung der Validität dieser Methode werden die Ergebnisse des narrativen indirekten Vergleichs dem des direkten Vergleichs gegenübergestellt und die Anzahl der kongruenten bzw. nicht-kongruenten Schlussfolgerungen ermittelt (vergleiche Forschungsfrage 3).

Typ 4: Systematische Reviews mit Metaanalyse, die nur einen indirekten Vergleich vornehmen.

In dieser Gruppe kann gezählt werden, wie oft welche Methoden zur Durchführung von indirekten Vergleichen im Rahmen von Metaanalysen bisher zum Einsatz kamen. Zudem sollen durch Typ-1-Publikationen nicht identifizierte Methoden aufgefunden und dem Katalog von Methodenbeschreibungen zugeführt werden. Die Beschreibung von bereits bekannten Methoden soll vervollständigt und durch Varianten ergänzt werden.

Diese Ziele sollen durch die Extraktion von Informationen anhand des im Anhang in Kapitel 9.8.2 „Auswertungsbogen für systematische Reviews mit Metaanalyse(n), die nur einen indirekten Vergleich durchführten“ dargestellten Auswertungsbogens erreicht werden.

Typ 5: Systematische Reviews mit Metaanalyse, die sowohl einen direkten als auch indirekten Vergleich vornehmen.

Die für Publikationen des Typs 4 geltenden Auswertungsstrategien werden genauso auf die des Typs 5 angewendet. Zudem soll anhand der Typ-5-Publikationen die Validitätsprüfung der Methoden zur Durchführung indirekter Vergleiche im Verhältnis zu den Ergebnissen von direkten Vergleichen durchgeführt werden (vergleiche Forschungsfrage 3 und 4). Aus diesem Grund wird der Auswertungsbogen für Typ-4-Publikationen um Fragen nach Ergebnissen der Metaanalysen und der in sie eingeschlossenen Einzelstudien ergänzt (siehe Kapitel 9.8.3 „Auswertungsbogen für systematische Reviews mit Metaanalyse(n), die einen direkten und indirekten Vergleich durchführten“ im Anhang).

4.3 Beschreibung der indirekt vergleichenden Methoden

Die Beschreibung der Methoden erfolgt anhand der gefundenen Methodenpapiere und den Methodenteilen in Publikationen von Anwendungsbeispielen. Zunächst werden die relevanten Publikationen benannt, es folgen die Begriffsbestimmung und die Methodenbeschreibung mit grafischer Darstellung. Abschließend wird auf Varianten und Weiterentwicklungen der spezifischen Methoden eingegangen.

4.4 Häufigkeit des Einsatzes der verschiedenen Methoden für indirekte Vergleiche

Für die quantitative Erhebung, wie häufig die einzelnen Methoden zur Durchführung von indirekten Vergleichen eingesetzt werden, werden die Publikationstypen 2 bis 5 herangezogen. Da eine systematische Recherche nach systematischen Reviews, die indirekte Vergleiche durchführen, nur in den DIMDI-Datenbanken und im ISI Web of Knowledge[®] vorgenommen wird, beschränkt sich der Zeitraum, für den diese Erhebung durchgeführt werden kann auf den Recherchezeitraum vom 01.01.1999 bis 20.02.2008. Literatur, die über Handsuchen identifiziert wird, wird für die quantitative Erhebung nicht berücksichtigt.

4.5 Methodik der Validitätsprüfung

Als Goldstandard für den Vergleich von Therapieverfahren gilt die direkt vergleichende Head-to-head-Studie. Mithilfe der Validitätsprüfung soll festgestellt werden, ob die verschiedenen Methoden zur Durchführung indirekter Vergleiche zu den gleichen Ergebnissen kommen wie Head-to-head-Studien zur gleichen Fragestellung.

4.5.1 Indirekte Vergleiche ohne Metaanalysen

Die Ergebnisse narrativer indirekter Vergleiche werden anhand der Typ-3-Publikationen überprüft. Die Ergebnisinterpretation und Schlussfolgerungen der indirekten Vergleiche werden den Ergebnissen von Head-to-head-Studien gegenübergestellt und auf Kongruenz kontrolliert.

4.5.2 Indirekte Vergleiche mit metaanalytischen Methoden

4.5.2.1 Datensätze für die Validitätsprüfung

Zur Validitätsprüfung von metaanalytischen Methoden des indirekten Vergleichs werden Vergleiche therapeutischer Interventionen herangezogen, für die RCT-Daten zur Verfügung stehen, die sowohl einen direkten (Head-to-head) als auch einen indirekten Vergleich über einen gemeinsamen Komparator ermöglichen.

In den Typ-5-Publikationen führen die Autoren sowohl eine Metaanalyse zu der direkten Evidenz durch (sofern mehr als eine Head-to-head-Studie vorliegt) als auch eine oder mehrere Metaanalysen für die indirekten Vergleiche. Aus diesen Reviews werden Datensätze mit Ergebnissen direkter und indirekter Vergleiche entnommen. Dabei kann bei mehreren zu vergleichenden Therapieoptionen und unterschiedlichen erhobenen Zielgrößen auch mehr als ein Datensatz aus einem systematischen Review extrahiert werden. Die Datensätze werden nach der für den indirekten Vergleich gewählten Methode sortiert und nummeriert (siehe Tabelle 28 bis Tabelle 31). Die Validitätsprüfung erfolgt methodenspezifisch. Auf eine methodenübergreifende Prüfung wird aufgrund der großen Unterschiede in den indirekt vergleichenden Methoden verzichtet.

4.5.2.2 Berechnung der Diskrepanzen zwischen direktem und indirektem Vergleich (z-Statistik)

Für jede Methode des indirekten Vergleichs wird die Hypothese geprüft, dass sich die Ergebnisse des indirekten Vergleichs von denen des direkten Vergleichs nicht signifikant unterscheiden. Hierzu wird zunächst für jeden Datensatz die Differenz der Therapieeffektunterschiede zwischen direktem (θ_{Direkt}) und indirektem Vergleich (θ_{Indirekt}) gebildet; in der Folge Diskrepanz ($\theta_{\text{Diskrepanz}}$) genannt. Dies geschieht für dichotome und kontinuierliche Daten nach unterschiedlichen Formeln:

Kontinuierliche Daten:

$$\theta_{\text{Diskrepanz}} = \theta_{\text{Direkt}} - \theta_{\text{Indirekt}}$$

Dichotome Daten:

$$\theta_{\text{Diskrepanz}} = \ln \theta_{\text{Direkt}} - \ln \theta_{\text{Indirekt}}$$

Für die Berechnung der Standardfehler der Effektschätzer des direkten und indirekten Vergleichs und der Konfidenzintervalle der Diskrepanz wird auf die Formelsammlung im Anhang in Abschnitt 9.1 „Formeln für die Methodik der Validitätsprüfung“ verwiesen.

Um die Diskrepanzen aus verschiedenen Therapievergleichen, die unterschiedliche Effektmaße wählen und unterschiedliche Standardfehler aufweisen, miteinander vergleichbar zu machen, wird für jede Diskrepanz zwischen direktem und indirektem Vergleich die z-Statistik berechnet¹⁰⁹. Dabei handelt es sich um eine Teststatistik für den Gesamteffektunterschied zwischen zwei Verteilungen^{8, 71}. Unter der Nullhypothese, dass es keinen Therapieeffektunterschied gibt, folgt z einer Standardnormalverteilung⁸:

$$SE(\theta_{\text{Diskrepanz}}) = \sqrt{SE(\theta_{\text{Direkt}})^2 + SE(\theta_{\text{Indirekt}})^2}$$

$$z = \frac{\theta_{\text{Diskrepanz}}}{SE(\theta_{\text{Diskrepanz}})}$$

Auf diese Weise erhält jede Diskrepanz zwischen direktem und indirektem Vergleich einen zugehörigen z-Wert. Ein z-Wert von Null zeigt an, dass direkter und indirekter Vergleich den Therapieeffektunterschied zwischen Therapie A versus B auf exakt den gleichen Wert schätzen. Ein negativer z-Wert zeigt eine Überschätzung dieses Therapieeffektunterschieds durch den indirekten Vergleich an. Ein positiver z-Wert weist auf eine Unterschätzung dieses Therapieeffektunterschieds durch den indirekten Vergleich hin¹⁰⁹.

Eine statistisch signifikante Diskrepanz ($p = 0,05$) ist gegeben, wenn der z-Wert entweder größer als 1,96 oder kleiner als -1,96 ist, bzw. der Betrag des z-Wertes größer 1,96 ist¹⁰⁹.

4.5.2.3 Test auf systematische Über- oder Unterschätzung

Die Grundvoraussetzung für die Validität einer indirekt vergleichenden Methode ist, dass sie das Ergebnis des direkten Vergleichs weder systematisch über- noch unterschätzt.

Hierzu wird geprüft, ob sich die Verteilung der erhaltenen z-Werte wie eine Normalverteilung verhält, deren Mittelwert nicht statistisch signifikant vom Wert $z = 0$ abweicht.

Dafür wird zunächst für jede Methode des indirekten Vergleichs getrennt aus den z-Werten der einzelnen Datensätze der Mittelwert gebildet. Dieser darf auf 95-prozentigem Konfidenzniveau nicht vom Wert $z = 0$ abweichen, d. h. sein 95 %-Konfidenzintervall muss den Wert Null mit einschließen²³⁸. Zudem wird für jede Methodik des indirekten Vergleichs mit dem Kolmogorov-Smirnov-Test darauf getestet, dass die z-Werte um diesen Mittelwert normalverteilt sind. Dabei darf auf 95-prozentigem Konfidenzniveau die Nullhypothese, dass eine Normalverteilung vorliegt, nicht widerlegt werden.

4.5.2.4 Quantifizierung des Ausmaßes der Diskrepanz

Um auch einer hohen Streuung und stark vom Mittelwert abweichenden Werten gerecht zu werden, wird für jede Methode neben dem Mittelwert des z-Wertes auch der Mittelwert des Betrags des z-Wertes berechnet. Bei der Mittelwertberechnung des z-Wertes können sich hohe z-Werte mit negativem Vorzeichen mit z-Werten mit positivem Vorzeichen ausgleichen. Dies ist bei der Mittelwertberechnung der Beträge ausgeschlossen. Je größer der Mittelwert der Beträge der z-Werte ist, desto größer ist die mittlere Diskrepanz zwischen direktem und indirektem Vergleich. Diese Größe ist auch geeignet, um die Validität der verschiedenen indirekt vergleichenden Methoden einander gegenüber zu stellen.

4.5.2.5 Häufigkeit von statistisch signifikant diskrepanten Datensätzen

Eine weitere interessierende Größe ist die Häufigkeit von statistisch signifikant diskrepanten Ergebnissen für die einzelnen indirekt vergleichenden Methoden.

Die einzelnen Diskrepanzen werden mit ihren 95 %-Konfidenzintervallen in Forest-plot-Diagrammen dargestellt. Hierdurch wird auf den ersten Blick ersichtlich, wie viele Konfidenzintervalle den Wert 0 nicht mit einschließen und damit auf eine statistisch signifikante Über- oder Unterschätzung des Therapieeffekts auf dem 95 %-Konfidenzniveau durch den indirekten Vergleich hinweisen.

Es wird der Anteil der auf dem 95 %-Konfidenzniveau statistisch signifikant diskrepanten Datensätze an allen für die Validitätsprüfung einer Methode herangezogenen Datensätzen berechnet. Der Berechnung der 95 %-Konfidenzintervalle für diese Anteile wird die folgende Formel zugrunde gelegt⁴⁹:

$$95\%KI = p \pm 1,96 \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Der Wert p beschreibt die Wahrscheinlichkeit für das Eintreten eines Ereignisses - in diesem Fall den Anteil der statistisch signifikant diskrepanten Datensätze. Der Wert n entspricht der Stichprobengröße, die in diesem Fall die Anzahl der für die Validitätsprüfung einer indirekt vergleichenden Methode herangezogenen Datensätze darstellt.

Geringe Stichprobengrößen der eingeschlossenen Studien führen zu geringer Präzision der Effektschätzer, sowohl beim direkten als auch beim indirekten Vergleich, was sich in einem weiten Konfidenzintervall für die Diskrepanz widerspiegelt. Bei geringer Präzision können auch große Ergebnisunterschiede zwischen den Ergebnissen des direkten und indirekten Vergleichs keine statistische Signifikanz in der Diskrepanz ($|\bar{z}| > 1,96$) erreichen. An der Darstellung der Diskrepanzen in Forest-Plots kann visuell beurteilt werden, ob geringe Präzision einen eher seltenen oder häufigen Grund für nicht-signifikante Diskrepanz darstellt.

4.5.2.6 Subgruppenanalyse

Wenn die Ergebnisse von direkten Vergleichen mit in den indirekten Vergleich eingerechnet werden (z. B. bei dreiarmigen Head-to-head-Studien mit dem gemeinsamen Komparator als drittem Arm oder in MTC), sind die Stichproben indirekter Vergleich und direkter Vergleich nicht mehr als statistisch unabhängig voneinander zu betrachten, da sie teilweise auf dieselben Daten zurückgreifen.

Inwieweit dies Einfluss auf die Ergebnisse der Validitätsprüfung hat, wird in einer Subgruppenanalyse geprüft. Diese dient auch der Beantwortung der Forschungsfrage 4 (siehe Kapitel 3 „Forschungsfragen“), die auf die Validitätsprüfung von indirekten Vergleichen abzielt, die Head-to-head-Studien mit einschließen.

Für die Subgruppenanalyse werden alle Analysen der Validitätsprüfung mit den Datensätzen wiederholt, bei denen definitiv keine Head-to-head-Studien in den indirekten Vergleich eingeschlossen wurden.

Die Ergebnisse der Subgruppenanalyse werden den Resultaten der Validitätsprüfung aller Datensätze (Hauptanalyse) gegenübergestellt.

4.6 Beurteilung der Präzision indirekt vergleichender Methoden

Eine mangelnde Präzision kann sich auf die Ergebnisse der Validitätsprüfung auswirken. Sind die Konfidenzintervalle um die Effektschätzer der indirekten Vergleiche sehr weit, so könnten sie mit den Konfidenzintervallen um die Effektschätzer des direkten Vergleichs überlappen, obwohl die Punktschätzer weit auseinander liegen. Die Diskrepanz zwischen direktem und indirektem Vergleich wäre in solchen Fällen zwar groß, aber nicht statistisch zu sichern.

Um solche Fälle erkennen zu können, wird die Konfidenzintervallweite des Effektschätzers des indirekten Vergleichs ins Verhältnis zu dem des direkten Vergleichs gesetzt:

$$X(\%) = \left(\frac{\text{ObereGrenze}_{\text{indirekt}} - \text{UntereGrenze}_{\text{indirekt}}}{\text{ObereGrenze}_{\text{Direkt}} - \text{UntereGrenze}_{\text{Direkt}}} - 1 \right) \times 100$$

Dichotome Daten gehen dabei als logarithmierte Werte in diese Formel ein. X steht für die prozentuale Zu- bzw. Abnahme der Weite des Konfidenzintervalls des Therapieeffektunterschieds aus dem indirekten Vergleich in Relation zur derjenigen aus dem direkten Vergleich. Ein negativer Wert für X, wie z. B. -70 % bedeutet, dass die Weite des Konfidenzintervalls vom indirekten Vergleich um 70 % geringer ist als die des direkten. Ein positiver Wert von X von z. B. 70 % bedeutet, dass das Konfidenzintervall des indirekten Vergleichs um 70 % weiter ist als das Konfidenzintervall des direkten.

Abschließend wird diese Stichprobe genutzt, eine Theorie von Glenny et al.¹⁰⁹ empirisch zu prüfen, die besagt, dass indirekte Vergleiche mit einer vierfach größeren Anzahl eingeschlossener Studien als beim direkten Vergleich, die gleiche Präzision erreichen, wie der direkte Vergleich.

4.7 Überprüfung der Kongruenz in den Schlussfolgerungen von direktem und indirektem Vergleich

Zuletzt wird erhoben, in wie vielen Fällen der direkte und der indirekte Vergleich zu den gleichen Schlussfolgerungen kommen.

Werden zwei Therapieoptionen miteinander verglichen, bestehen fünf Möglichkeiten, wie die Frage nach kongruenten Schlussfolgerungen von direktem und indirektem Vergleich beantwortet werden kann:

1. Ja, die gleiche Therapieoption ist sowohl im indirekten als auch im direkten Vergleich signifikant überlegen.
2. Ja, die Therapieoptionen zeigen sich sowohl im direkten als auch im indirekten Vergleich gleich wirksam (Es liegt kein signifikanter Therapieeffektunterschied vor).
3. Nein, der Therapieeffektunterschied im indirekten Vergleich ist nicht signifikant und im direkten Vergleich signifikant.
4. Nein, der Therapieeffektunterschied im indirekten Vergleich ist signifikant und im direkten Vergleich nicht signifikant.
5. Nein, im direkten Vergleich zeigt sich die eine Therapie als signifikant überlegen und im indirekten Vergleich die andere.

In der Stichprobe wird das Verhältnis der Anzahl der mit „Ja“ beantworteten Fragen zu der Anzahl der mit „Nein“ beantworteten Fragen gebildet.

5 Ergebnisse

5.1 Ergebnisse der Literaturrecherchen und Literaturselektion

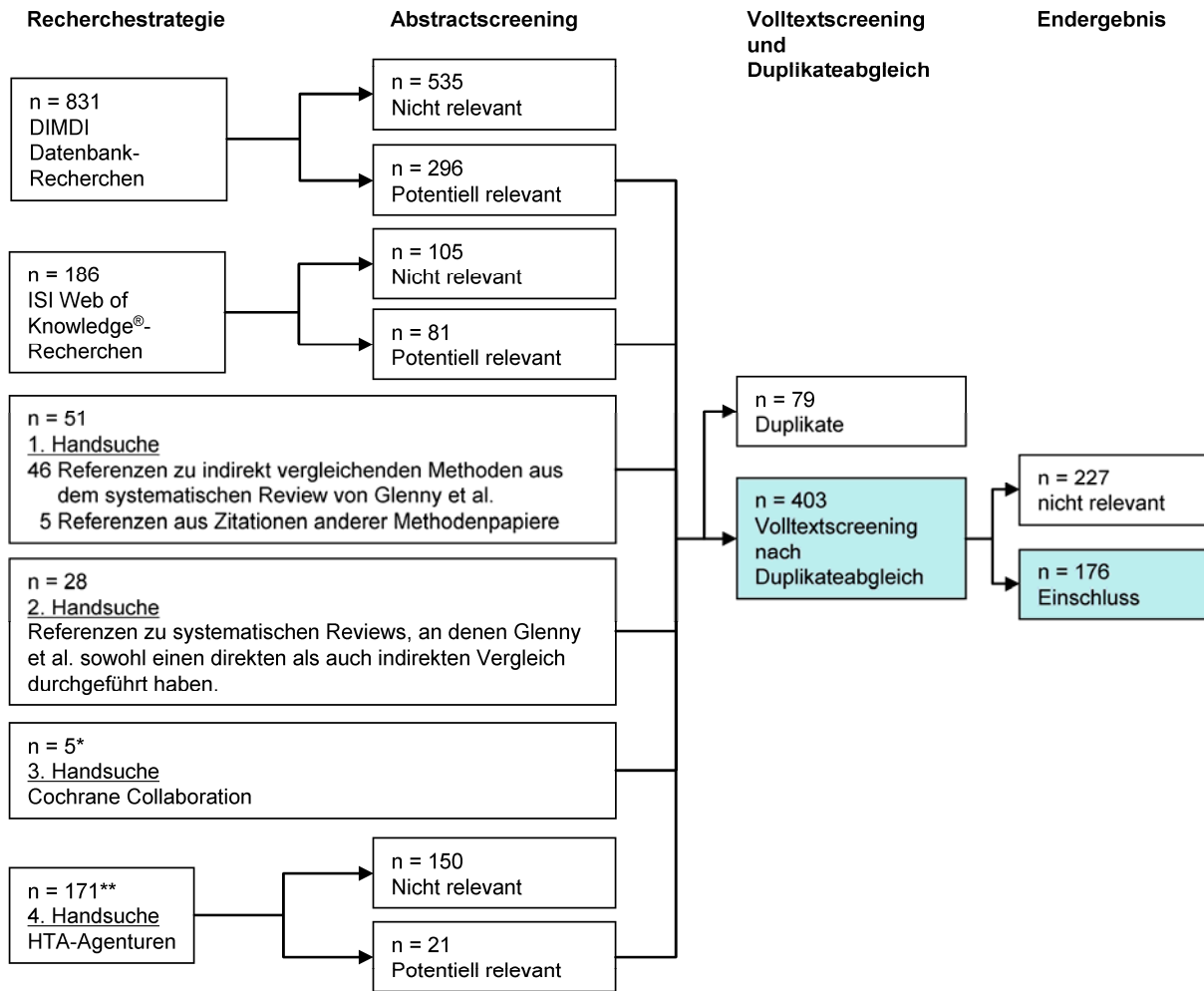
Der Ablauf und die Ergebnisse der Literaturrecherchen sowie der Handsuchen sind in Abbildung 3 dargestellt. Nach der ersten Literaturselektion auf Abstractebene werden alle in den DIMDI-Datenbanken und ISI Web of Knowledge[®] gefundenen, potenziell relevanten Referenzen dem Volltextscreening zugeführt. In das Volltextscreening werden zudem Publikationen, die über die vier Handsuchen gefunden wurden, übernommen. In der ersten Handsuche werden von den Zitationen des Kapitels zur Beschreibung der Methoden indirekter Vergleiche im Review von Glenny et al.¹⁰⁹ alle 46 Zitationen dem Volltextscreening unterzogen. Insgesamt liefert diese Handsuche 15 relevante Methodenpapiere (siehe Tabelle 17), die wiederum zu fünf weiteren Methodenpapieren aus ihren Zitationen führen, die auch für relevant befunden werden (siehe Tabelle 18).

In der zweiten Handsuche werden aus dem systematischen Review von Glenny et al.¹⁰⁹, an den dieser Bericht anknüpft, die 28 Referenzen zu systematischen Reviews, mit deren Daten Glenny et al.¹⁰⁹ sowohl einen direkten als auch indirekten Vergleich vorgenommen haben, entnommen. Da sich die Literatursuche von Glenny et al.¹⁰⁹ auf 1966 bis 1999 beschränkte, gab es nur minimale Überschneidungen (drei Reviews) (siehe Tabelle 19).

Die dritte Handsuche auf den Internetseiten der Cochrane Collaboration liefert fünf potenziell relevante Publikationen, die sich bis auf das Handbuch⁸ im Volltextscreening als irrelevant für die zu bearbeitenden Fragestellungen herausstellen (siehe Tabelle 16).

Durch die vierte Handsuche auf den Internetseiten der recherchierten HTA-Institutionen (Tabelle 10) werden 21 methodische Volltexte gefunden, die ebenfalls einem zweiten Screening unterzogen werden. Im Ergebnis enthält nur ein Workshopbericht zur Überarbeitung des Methodenpapiers des National Institute for Health and Clinical Excellence (NICE) Ausführungen zu indirekten Vergleichen²²⁶ (siehe Tabelle 14). Auf Basis dieses Workshopberichts wurden kurz vor Fertigstellung dieses HTA-Berichts indirekt vergleichende Methoden in den Guide to the Technology Appraisal Process aufgenommen¹⁵. Auf seine Ausführungen wird im Diskussionskapitel eingegangen. Das gleiche gilt auch für einen kürzlich veröffentlichten Entwurf für eine neue Version des Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews der amerikanischen HTA-Institution Agency for Healthcare Research and Quality (AHRQ)⁶.

Nach Zusammenführung der Ergebnisse der Handsuchen und der Suchen in digitalen Datenbanken wurden nach Duplikatabgleich 403 Volltexte gelesen und bewertet. Von diesen 403 Volltexten werden 176 Publikationen für relevant befunden und insgesamt 227 ausgeschlossen. Eine Auflistung der Referenzen der im Volltextscreening ein- und ausgeschlossenen Publikationen, sortiert nach den im Methodikkapitel festgelegten Ein- bzw. Ausschlussgründen wird im Anhang in Tabelle 20 und Tabelle 21 gegeben.



* Anzahl der für diesen Review potentiell relevanten Methodenpapiere

** Zwei HTA-Agenturen (AHRQ, NICE) bieten auf Ihren Internetseiten eine Vielzahl von Methodenpapieren an. Hier wird auf eine Zählung der Papiere verzichtet und nur gezielt nach Publikationen zu indirekten Vergleichen gesucht.

Abbildung 3: Ergebnisse der Literaturrecherchen und Handsuchen

AHRQ = Agency for Healthcare Research and Quality. DIMDI = Deutsches Institut für Medizinische Dokumentation und Information. HTA = Health Technology Assessment. NICE = National Institute for Health and Clinical Excellence.

Entsprechend den unterschiedlichen Forschungsfragen werden die 176 eingeschlossenen Publikationen in die im Methodikkapitel beschriebenen fünf Typen unterschieden (siehe Abbildung 4).

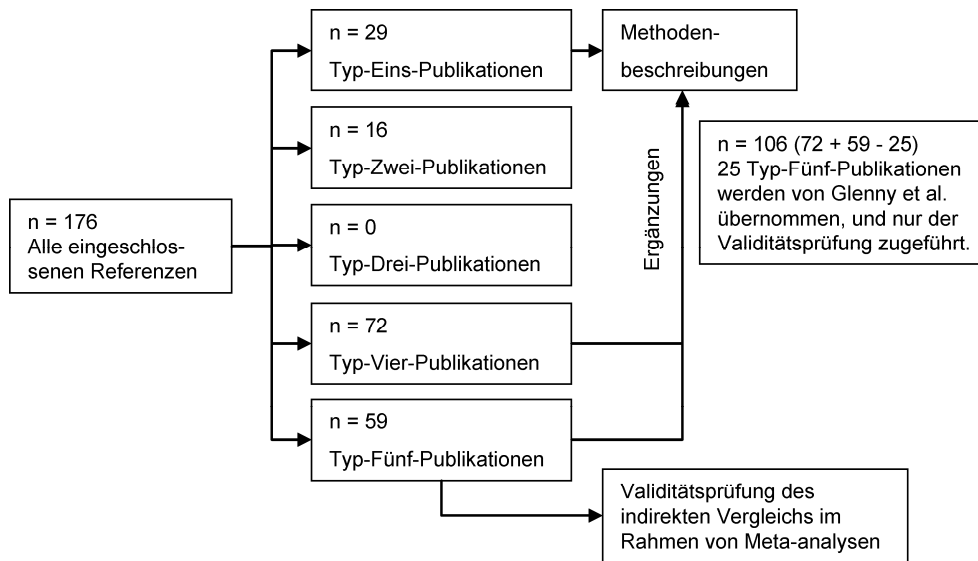


Abbildung 4: Verwendungszwecke der gesamten eingeschlossenen Literatur

29 Methodenpapiere (Typ 1) dienen als Basis für die Methodenbeschreibungen zur Durchführung indirekter Vergleiche (siehe Kapitel 5.3 „Beschreibung der indirekt vergleichenden Methoden“). Ergänzt werden die Methodenbeschreibungen durch Informationen aus Methodenanwendungen in 106 Publikationen. Diese Zahl setzt sich aus den Typ-4- (nur indirekter Vergleich durchgeführt) und Typ-5-Publikationen (indirekter und direkter Vergleich durchgeführt) zusammen minus der Anzahl an Publikationen, die Glenny et al.¹⁰⁹ bereits ausgewertet haben. Diese 106 Veröffentlichungen werden auch für die Erhebung der relativen Häufigkeit des Einsatzes der verschiedenen Methoden indirekter Vergleiche verwendet. Dieses wird in Abbildung 14 in Kapitel 5.4 „Relative Häufigkeit der indirekt vergleichenden Methoden“ dargestellt.

Für die Validitätsprüfung indirekter Vergleiche ohne Metaanalysen werden leider keine Typ-3-Publikationen (indirekter und direkter Vergleich ohne Metaanalysen) gefunden. Für die Validitätsprüfung indirekter Vergleiche, die im Rahmen von Metaanalysen eingesetzt werden, können 59 Typ-5-Publikationen herangezogen werden (siehe Kapitel 5.5 „Validitätsprüfung indirekt vergleichender Methoden“). Aus diesem Grund fokussiert der vorliegende HTA-Bericht in der Folge auf Methoden indirekter Verfahren mit metaanalytischen Methoden.

5.2 Allgemeine Annahmen und Voraussetzungen für indirekte Vergleiche

Die Durchsicht der Literatur ergibt, dass allen Methoden für indirekte Vergleiche die gleiche Annahme zugrunde liegt. Es wird davon ausgegangen, dass die Variabilität zwischen den Studienergebnissen durch den Zufall bedingt ist. In anderen Worten: Es darf keine signifikante Heterogenität zwischen den Studien bestehen. Da die Einhaltung dieser Voraussetzung so essenziell für die Validität der indirekten Vergleiche ist, wird das Thema Heterogenität im folgenden Kapitel ausführlich besprochen.

5.2.1 Heterogenität

Definition

Heterogenität beschreibt die Variabilität der in mehreren Studien ermittelten Therapieeffekte, die über die zufallsbedingte Variabilität hinausgeht²³⁵.

Arten und Ursachen von Heterogenität

Häufig wird zwischen Heterogenität in (within-study heterogeneity) und zwischen (between-study heterogeneity) Studien unterschieden. Der Unterschied liegt darin, dass in der within-study heterogeneity die Variabilität des Therapieeffekts auf Patientenebene und bei der between-study heterogeneity auf der Studienebene stattfindet.

Eine andere Unterscheidungsform von Heterogenitätsarten zielt auf die Ursachen der Heterogenität ab. Sie unterscheidet in methodische und klinische Heterogenität (siehe Glossar von Woolacott et al.²⁶⁴). Methodische Heterogenität resultiert aus Unterschieden im Studiendesign (Variation in der Qualität des technischen Designs oder der Durchführung der Studie, wie z. B. bei der Verblindung²⁰⁸). Die klinische Heterogenität zielt auf Unterschiede in Schlüsselcharakteristika der Probanden (u. a. Schwere der Erkrankung oder demografische Faktoren), Interventionen (unterschiedliche Studienprotokolle und Therapieregime, wie z. B. bei der Dosierung und der Therapiedauer²⁰⁸) oder Therapieeffektschätzer (z. B. Skalen auf Basis unterschiedlicher Fragebögen) ab.

Heterogenitätstests

Für die Validität von metaanalytischen Ergebniszusammenfassungen ist es unverzichtbar, dass die eingeschlossenen Studien homogen sind^{115, 154}.

Zur Prüfung, ob diese Voraussetzung gegeben ist, wurden sogenannte Heterogenitätstests entwickelt. Ein Heterogenitätstest untersucht die Nullhypothese, dass die zwischen den Studien beobachtete Streuung der Effektparameter allein durch Zufallseffekte erklärbar ist. Der übliche Test ist Cochranes Q-Test⁷³. Signifikante Heterogenität wird angenommen, wenn die Nullhypothese mit einer Wahrscheinlichkeit von unter 5 % ($p < 0,05$) bzw. unter 10 % ($p < 0,1$)⁹⁸ zutreffend ist. Die in der Statistik unübliche Schwelle von 0,1 wurde eingeführt, um die geringe Power des Tests zu berücksichtigen. Vorliegende Heterogenität kann der Test besonders schlecht erkennen, wenn Metaanalysen wenige Studien einschließen.

Der eigentlich interessante Parameter ist jedoch das Ausmaß, in dem die Heterogenität die metaanalytisch zusammengefassten Ergebnisse beeinflusst¹²¹. Zu seiner Bestimmung wurde der Q-Test zum sogenannten I^2 -Test weiterentwickelt, der Heterogenität in und zwischen den Studien nicht nur qualitativ, sondern auch quantitativ erheben kann¹⁷³. Es handelt sich hierbei um eine Standardmethode, die in jedem systematischen Review der Cochrane Collaboration angewendet wird⁸. I^2 kann nur Werte von 0 % bis 100 % annehmen. Sie entsprechen dem prozentualen Anteil der beobachteten Varianz in dem Studiensatz, der auf Heterogenität und nicht auf zufällige Abweichungen zurückgeführt werden kann. Eine Bewertung der Prozentangaben wird von Higgins et al.¹²¹ vorgeschlagen:

$I^2 < 25$ %:	Geringe Heterogenität
25 % < $I^2 < 75$ %:	Moderate Heterogenität (ab $I^2 = 50$ % bedeutsame Heterogenität)
$I^2 > 75$ %:	Hohe Heterogenität

Heterogenität in Random und Fixed effects-Modellen

In der Methodik jeder Metaanalyse gibt es stets die Auswahl, ob ein Random oder ein Fixed effects-Modell angewendet werden soll. Unterscheidungskriterium zwischen Random und Fixed effects-Modellen ist die Art, wie die Studienvariabilität in der Metaanalyse berücksichtigt werden soll.

In einem Fixed effects-Modell wird die Variabilität zwischen den Studienergebnissen auf den Wert Null festgelegt¹⁶². Dies entspricht der Annahme, dass keine between-study heterogeneity vorliegt und der wahre Therapieeffekt in allen Studien konstant ist. Die within-study heterogeneity (Variabilität des Therapieeffekts auf Patientenebene) dagegen wird im Fixed effects-Ansatz berücksichtigt, da die Varianz der Effektschätzer der Einzelstudien im Konfidenzintervall des zusammengefassten Gesamteffektschätzers abgebildet wird.

Random effects-Modelle berücksichtigen dagegen sowohl die within- als auch die between-study heterogeneity. Ein Random effects-Modell erlaubt Variabilität zwischen den Studienergebnissen, indem es annimmt, dass alle Studien einen unterschiedlichen wahren Effekt messen und ihre Studienergebnisse normalverteilte Abweichungen von einem wahren durchschnittlichen Gesamteffekt darstellen⁸⁶.

Random effects-Modelle liefern deshalb weniger präzise Effektschätzer als Fixed effects-Modelle^{98, 109}. Die breiteren Konfidenzintervalle spiegeln die größere Unsicherheit des gepoolten Therapieeffekts einer heterogenen Studienlage wider. Liegt geringe oder gar keine Heterogenität vor, sind die Resultate von Fixed und Random effects-Modellen nahezu identisch⁹⁸.

Die Entscheidung, ob ein Random oder ein Fixed effects-Modell für eine Metaanalyse verwendet werden soll kann in sogenannten Zwei-Schritt-Analysen getroffen werden. Als erstes wird ein Heterogenitätstest durchgeführt. Zeigt dieser signifikante Heterogenität an, wird ein Random effects-Modell verwendet. So kann die große Unsicherheit des gepoolten Effektschätzers in einem weiten Konfidenzintervall ausgedrückt werden. Liegt keine oder geringe Heterogenität vor, wird ein Fixed effects-Modell verwendet, das eine präzise Schätzung des gepoolten Therapieeffekts liefert.

5.2.2 Umgang mit Heterogenität in systematischen Reviews mit indirekten Vergleichen

Herangezogene Datensätze

In allen Typ-4- (nur indirekter Vergleich durchgeführt) und Typ-5-Publikationen (indirekter und direkter Vergleich durchgeführt) wird der Umgang mit Heterogenität betrachtet (mit Ausnahme der Publikationen, die dem Review von Glenny et al.¹⁰⁹ entnommen werden).

Heterogenitätstests

Tabelle 23 zeigt, ob und welche Heterogenitätstests in indirekten Vergleichen verwendet werden. Insgesamt wird in 80 der 106 Reviews über die Durchführung eines oder mehrerer Heterogenitätstests berichtet (75 %). Am häufigsten werden Chi²-Tests verwendet, zu denen auch Cochranes Q-Test zählt. 51 der 80 Autoren(teams), die statistische Tests anwenden, wählen entweder einen Chi²-Test oder Q-Test (64 %). Dabei wird häufig auf die von den folgenden Autoren wichtiger Methodenpapiere für ihre indirekt vergleichenden Methoden maßgeschneiderten Chi²-Tests zurückgegriffen: Bucher et al.⁵⁵, Lumley¹⁶⁰ und Lu et al.¹⁵⁷.

Auch der I²-Test gehört mit 29 Nennungen (36 %) noch zu den häufig verwendeten Tests. Selten verwendet werden: Der Breslow-Day-Test (vier Nennungen), die Inkonsistenzmessung nach Lu et al.¹⁵⁷ oder Lumley¹⁶⁰ (drei Nennungen), der Comfied-Gard-Test (zwei Nennungen), Box' Variante des Barlett-Tests, der H-Test, der Kruskal-Wallis-Test, der Riley-Day-Test und die Varianzanalyse ANOVA-1 (je eine Nennung).

Heterogenität in Random und Fixed effects-Modellen

Die Entscheidung für oder gegen ein Random effects-Modell in einem indirekten Vergleich ist eine Entscheidung, ob die Heterogenität zwischen den Studien (between-study heterogeneity) berücksichtigt werden soll¹⁰⁹.

In 90 der 106 ausgewerteten systematischen Reviews mit indirekten Vergleichen wird berichtet, ob Random oder Fixed effects-Modelle genutzt werden (siehe Tabelle 22). Von diesen 90 Reviews wird ein Random effects-Modell in 46 Publikationen gewählt (51 %). Dagegen entscheiden sich 27 Autoren(teams) für den Fixed effects-Ansatz (30 %). Sechs Autoren(teams) (7 %) legen fest, sowohl ein Fixed als auch ein Random effects-Modell anzuwenden, um die Wahl des Metaanalyseinstruments als Fehlerquelle ausschließen zu können.

Ein Zwei-Schritt-Verfahren, das erst nach einem Heterogenitätstest die Entscheidung zwischen Random und Fixed effects-Modell trifft, wählen elf Autoren(teams) (12 %).

Heterogenitätsaufklärung

Aus den analysierten Reviews lassen sich fünf alternative Verfahren für den Umgang mit heterogenen Studienlagen dokumentieren (siehe Tabelle 23). Diese werden in Abbildung 5 schematisch dargestellt:

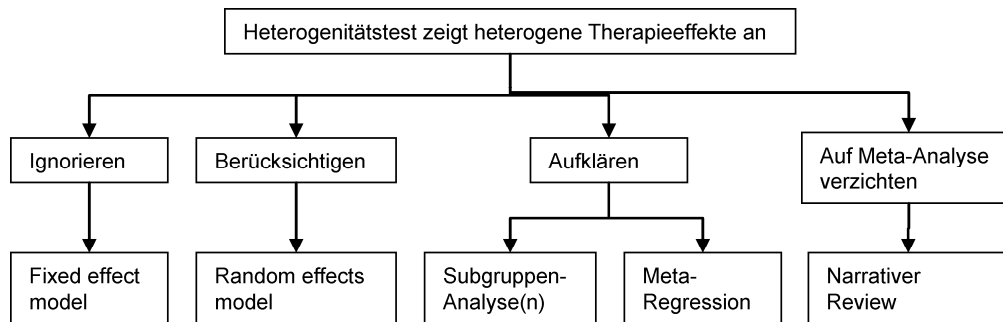


Abbildung 5: Wahlmöglichkeiten bei heterogenen Therapieeffekten

Wird Heterogenität statistisch nachgewiesen, entscheiden sich einige Reviewer dafür, die Daten nicht im Ganzen zu poolen, sondern stattdessen eine deskriptive Übersicht zu erstellen. Andere Autoren begründen mit einer tabellarischen Gegenüberstellung der Patientencharakteristika in den verschiedenen Studien, dass der Heterogenität keine bedeutsame Variabilität zwischen den Studienteilnehmern zugrunde liegt. In der Folge ignorieren sie die Heterogenität in der Metaanalyse durch die Wahl eines Fixed effects-Modells oder berücksichtigen diese durch ein Random effects-Modell.

Einige Reviews wenden Subgruppenanalysen oder Metaregressionen an, um die Gründe der Heterogenität aufzuklären.

In Subgruppenanalysen kann untersucht werden, durch welche Störfaktoren der Gesamttherapieeffekt besonders stark beeinflusst oder verzerrt wird⁸.

Metaregressionen können für potenzielle Störfaktoren adjustieren¹⁷⁸. Weicht das adjustierte Ergebnis der Metaregression vom Ergebnis der Metaanalyse nicht bedeutsam ab, wird auf einen geringen Einfluss des in der Metaregression berücksichtigten Störfaktors auf das Gesamtergebnis geschlossen.

Um die Studien zu identifizieren, die in besonderem Ausmaß zur Heterogenität beitragen, werden in einigen Reviews grafische Verfahren, wie der Funnel-Plot, Forest-Plot oder L'Abbè-Plot, eingesetzt.

5.3 Beschreibung der indirekt vergleichenden Methoden

Die Methodenbeschreibungen basieren in erster Linie auf Angaben, die die Erstautoren einer Methode in ihrem Methodenpapier ausführen. Ergänzt werden diese Beschreibungen durch Angaben von Autoren, die die Methode weiterentwickelten oder in Form einer Übersichtsarbeit über sie berichteten. Alle Angaben werden mithilfe der Checkliste aus Abschnitt 9.8.1 „Checkliste zur Beschreibung und Charakterisierung von Methoden zur Durchführung von indirekten Vergleichen“ im Anhang systematisch aus den vorliegenden Publikationen extrahiert.

Die Methodenbeschreibungen werden ergänzt um Informationen aus den Methodikkapiteln der Typ-4- und Typ-5-Publikationen (systematische Reviews, in denen indirekte Vergleiche durchgeführt wurden).

Die Bewertung der Methoden erfolgt auf der Basis der Ergebnisse der Validitätsprüfung in der Diskussion in Kapitel 6.5.2 „Wie ist die Validität indirekter Vergleiche gegenüber der von direkten Vergleichen zu beurteilen?“.

5.3.1 Nicht-adjustierter indirekter Vergleich

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Nicht vorhanden

Übersichtsarbeiten: Song et al.²¹⁸, Glenny et al.¹⁰⁹ und Gartlehner et al.¹⁰⁴

Methodenanwendungen: Sanchez-Ramoz et al.²⁰⁶ und Bottomley et al.⁵⁰

Weitere Methodenanwendungen werden in den Unterkapiteln genannt.

Begriffsbestimmung

Für die Methode, die im Folgenden als nicht-adjustierter indirekter Vergleich bezeichnet wird, hat sich noch keine Nomenklatur durchgesetzt. 2000 von Song et al.²¹⁸ als „The simple indirect method“ bezeichnet, werden 2005 in der Publikation von Glenny et al.¹⁰⁹, in der auch Song Co-Autor war, synonym die Begriffe „unadjusted indirect method“ und „naive method“ verwendet. Auch der Anfang 2008 veröffentlichte Review von Gartlehner et al.¹⁰⁴ verwendet den Begriff „unadjusted indirect comparison“.

Methodenbeschreibung

In Gartlehners Übersichtsarbeit wird das Vorgehen bei einem nicht-adjustierten indirekten Vergleich anschaulich erklärt. Diese Erläuterungen werden im Folgenden anhand eines fiktiven Beispiels (siehe Abbildung 6) wiedergegeben: Für den nicht-adjustierten indirekten Vergleich A versus B werden die Ergebnisse aller Studien, die einen aktiven Verumarm mit A besitzen in einer Metaanalyse zu einem Gesamtergebnis θ_A zusammengefasst. Das gleiche wird in einer zweiten Metaanalyse für alle Studien mit einem Verumarm B durchgeführt, um θ_B zu erhalten. Zur Veranschaulichung wird in Abbildung 6 ein Szenario entworfen, indem insgesamt zwölf Studien für den indirekten Vergleich von A versus B zur Verfügung stehen. Dabei handelt es sich je um drei zweiarmlige mit Therapieoption X kontrollierte Studien mit A (Studie 1 bis 3) und drei mit B (Studie 8 bis 10). Der Kontrollarm X stellt den gemeinsamen Komparator dar, über den der indirekte Vergleich durchgeführt wird. In den meisten Fällen handelt es sich um Placebo oder Standardtherapie, aber im Prinzip sind alle alternativen Therapieoptionen bei gegebener Indikation als gemeinsamer Komparator möglich. Die Studie 4 stellt eine dreiarmlige Head-to-head-Studie der Form A versus B versus X dar und die Studie 5 eine zweiarmlige Head-to-head-Studie der Form A versus B. In den Studien 6, 7, 11 und 12 wurden A bzw. B gegen andere Therapieoptionen als den gemeinsamen Komparator, hier als C und D bezeichnet, getestet.

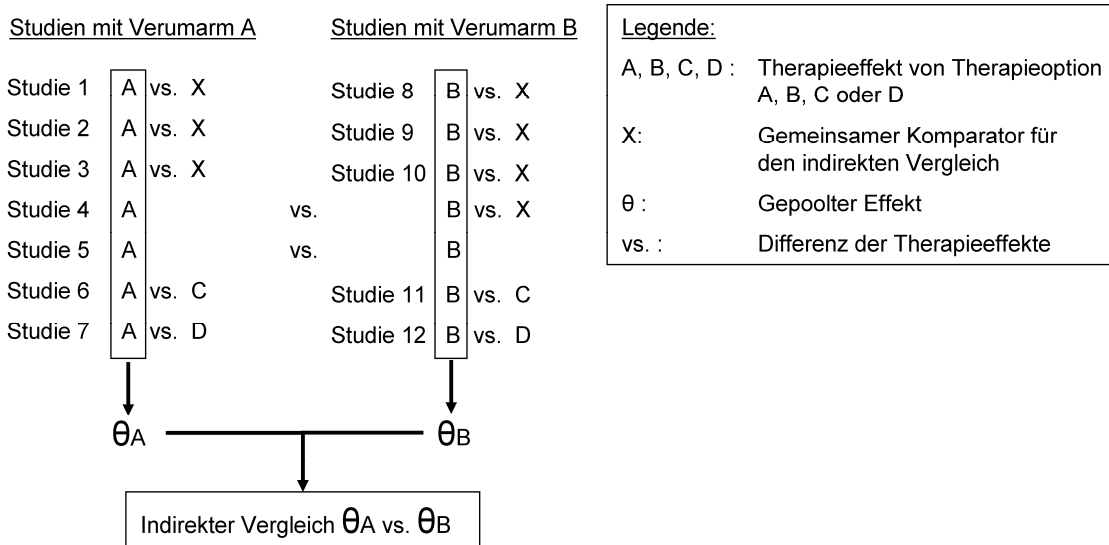


Abbildung 6: Nicht-adjustierter indirekter Vergleich

(Modifiziert nach Gartlehner et al.¹⁰⁴)

In diesem fiktiven Beispiel werden im nicht-adjustierten indirekten Vergleich alle verfügbaren Studien, die einen Verumarm mit A haben herangezogen, um ein zusammengefasstes Ergebnis für A (θ_A) zu erhalten und alle Studien mit einem Verumarm mit B, um ein zusammengefasstes Ergebnis θ_B zu erhalten.

Die Berechnung von θ_A und θ_B ist bei dichotomen und kontinuierlichen Ergebnismaßen unterschiedlich. Zunächst wird die Berechnung von dichotomen Daten allgemeingültig für relative Risiken (RR), Odds ratios (OR) und Hazard ratios (HR) vorgestellt:

Es wird die Anzahl der Ereignisse in allen Studienarmen mit A addiert und die Probandenzahlen, die mit A exponiert wurden, ebenfalls addiert. Anschließend werden diese zwei Summen in ein Verhältnis zueinander gesetzt²⁰⁶:

Für θ_B kann das entsprechende Verhältnis mit den Ergebnissen aus den Therapiearmen mit B der Einzelstudien gebildet werden.

Bei kontinuierlichen Daten wird, um θ_A zu erhalten, ein Mittelwert der in den Einzelstudien ermittelten Ergebnismaße nach folgender Formel gebildet⁵⁰:

$$\theta_A = \frac{\sum (\theta_{Ai} \cdot Y_i)}{\sum Y_i}$$

Dabei stellt θ_{Ai} das Ergebnis von A in einer Einzelstudie i und Y_i die Probandenzahl dieser Einzelstudie in dem Arm, der Therapie A bekam, dar. Für θ_B gilt die entsprechende Formel.

Unabhängig davon, ob dichotome oder kontinuierliche Daten verwendet werden, verfährt der nicht-adjustierte indirekte Vergleich so, als ob sich alle Probanden der Einzelstudien in der gleichen Studie befunden hätten¹⁰⁹. Ob es sich bei den zusammengefassten Studien um mit Komparator X kontrollierte Studien oder Head-to-head-Studien handelt, ist deshalb beim nicht-adjustierten indirekten Vergleich nicht von Relevanz, da die Ereignisse aus den Komparatorarmen bei der Berechnung des Gesamteffektschätzers nicht berücksichtigt werden. Aus dieser zentralen Eigenschaft rührt auch die Namensgebung des nicht-adjustierten indirekten Vergleichs, was eine Abkürzung der Bezeichnung „nicht bezüglich der Ergebnisse aus den Kontrollgruppen adjustierter indirekter Vergleich“ darstellt.

Für den Vergleich der zusammengefassten Ergebnisse θ_A und θ_B gibt es im Anschluss vier Möglichkeiten: Den indirekten Vergleich über ein Gesamtergebnis $\theta_{A \text{ versus } B}$, über einen statistischen Test, durch Vergleich der Konfidenzintervalle oder ein narratives Gegenüberstellen der Therapieeffekte von A und B.

1. Indirekter Vergleich über einen Gesamteffektschätzer $\theta_{A \text{ versus } B}$

Methodenanwendungen: Berry et al.⁴³, Bottomley et al.⁵⁰, Capstick et al.⁶¹, Medicare Services Advisory Committee³ und Sanchez-Ramos et al.²⁰⁶

Der Vergleich über die zusammengefassten Ergebnisse wird bei kontinuierlichen Daten durch einfache Subtraktion von θ_A minus θ_B erhalten⁵⁰. Bei dichotomen Daten wird das Gesamtergebnis $\theta_{A \text{ versus } B}$ über die Bildung eines Ratios θ_A / θ_B gebildet²⁰⁶. Diese Ratio kann ein RR, OR oder ein HR sein.

2. Indirekter Vergleich über einen statistischen Test

Methodenanwendungen: Adelman et al.¹⁹, Bakker et al.³³, Chen et al.⁶³, Einarson et al.⁹⁶ und Zarembski et al.²⁶⁸

Der statistische Test testet die Nullhypothese, dass sich die zusammengefassten Ergebnisse für A und B nicht unterscheiden bzw. der gleichen Grundgesamtheit entspringen.

3. Indirekter Vergleich durch Vergleich der Konfidenzintervalle

Methodenanwendungen: Chiba et al.⁶⁵ und van Dongen et al.²⁵¹

Es wird überprüft, ob sich die Konfidenzintervalle von θ_A und θ_B überschneiden. Überschneiden sie sich nicht, zeigt dies einen signifikanten Therapieeffektunterschied an.

4. Narratives Gegenüberstellen der Therapieeffekte von A und B

Methodenanwendungen: Coyle et al.⁷⁸ und van der Valk et al.²⁵⁰

Es wird logisch argumentiert, dass wenn das zusammengefasste Ergebnis von A einen größeren/kleineren absoluten Wert (je nach Definition eines positiven Therapieeffekts) aufweist als dasjenige von B, A im Vergleich wirksamer/weniger wirksam sein muss als B.

5.3.2 Adjustierter indirekter Vergleich

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Bucher et al.⁵⁵

Übersichtsarbeiten: Glenny et al.¹⁰⁹ und Gartlehner et al.¹⁰⁴

Methodenanwendungen: Siehe Unterkapitel

Begriffsbestimmung

Der adjustierte indirekte Vergleich subtrahiert (kontinuierliche Daten) oder dividiert (dichotome Daten) die Ergebnisse der Komparatorarme und Verumarme der Einzelstudien und adjustiert auf diese Weise vor dem Einsatz metaanalytischer Verfahren bezüglich der Ergebnisse aus den Kontrollgruppen¹⁰⁹.

Methodenbeschreibung

An dem gleichen fiktiven Beispiel wie in Kapitel 5.3.1 „Nicht-adjustierter indirekter Vergleich“ wird in Abbildung 7 gezeigt, wie ein adjustierter indirekter Vergleich vorgehen würde, um A und B über den gemeinsamen Komparator X zu vergleichen.

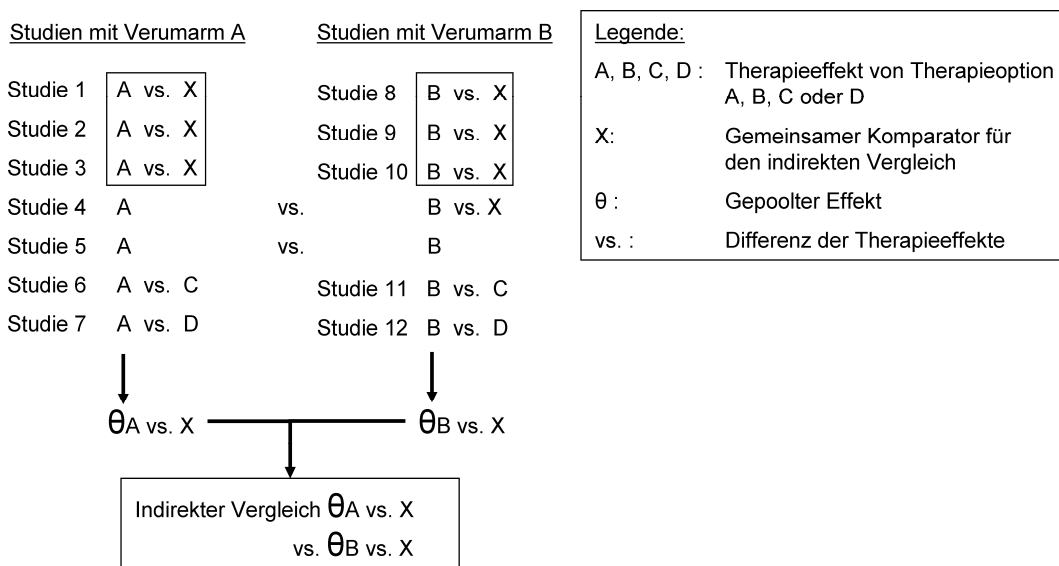


Abbildung 7: Adjustierter indirekter Vergleich

(Modifiziert nach Gartlehner et al.¹⁰⁴)

Es können nur die Studien für den adjustierten indirekten Vergleich herangezogen werden, die A oder B gegen den gemeinsamen Komparator testen. Studien, die nur gegen andere Komparatoren testen, können nicht eingeschlossen werden (Studie 6, 7, 11 und 12). Head-to-head-Studien, die A gegen B direkt testen, können ebenfalls nicht eingeschlossen werden (Studie 4 und 5).

Charakteristikum des adjustierten indirekten Vergleichs ist, dass in beide Metaanalysen die Effektschätzer des Vergleichs von Verum- mit Kontrollgruppe eingehen^{55, 104}. Die Ergebnisse der beiden Metaanalysen sind die gepoolten Effektschätzer, die in Abbildung 7 mit θ_A versus X und θ_B versus X bezeichnet sind. Für den Vergleich dieser beiden Therapieeffektunterschiede gibt es, wie beim nicht-adjustierten indirekten Vergleich, vier Möglichkeiten:

1. Über einen Gesamteffektschätzer θ_A versus B
2. Über einen statistischen Test
3. Durch Vergleich der Konfidenzintervalle
4. Narratives Gegenüberstellen der Therapieeffekte

In den folgenden Unterkapiteln werden diese vier Möglichkeiten für den adjustierten indirekten Vergleich, belegt an Literaturbeispielen, vorgestellt.

1. Indirekter Vergleich mithilfe eines Gesamteffektschätzers

Methodenpapier(e): Bucher et al.⁵⁵ und Goadsby et al.¹¹¹

Übersichtsarbeiten: Glenny et al.¹⁰⁹ und Gartlehner et al.¹⁰⁴

Methodenanwendungen: U. a. Testa et al.²³², Lee et al.¹⁵¹, McLeod et al.¹⁷⁰, Collins et al.⁷⁴, Buscemi et al.⁵⁸, Vandermeer et al.²⁵⁴ und Packer et al.¹⁸³

Viele systematische Reviews mit indirekten Vergleichen berufen sich in der Methodenbeschreibung für die Datenanalyse auf die Publikation von Bucher et al.⁵⁵ (27 identifizierte Reviews zitieren Bucher et al.⁵⁵) und in weiteren sieben Reviews kann die Verwendung dieser Methode eindeutig erkannt werden (siehe Tabelle 21 im Anhang). Bucher et al.⁵⁵ geben eine Anleitung zur Berechnung eines Gesamteffektschätzers (θ_{AB}) für den indirekten Vergleich aus den Ergebnissen zweier Metaanalysen.

Die Ergebnisse der Metaanalysen für A versus X (θ_{AX}) und B versus X (θ_{BX}) werden durch Subtraktion zu einem Gesamteffektschätzer für den indirekten Vergleich (θ_{AB}) zusammengefasst.

$$\ln\theta_{AB} = \ln\theta_{AX} - \ln\theta_{BX} \quad (1)$$

Die Varianz des Gesamteffektschätzers für den indirekten Vergleich darf durch Addition der Varianzen der beiden Effektschätzer aus den Metaanalysen berechnet werden, da unterschiedliche Studien in die beiden Metaanalysen eingegangen und ihre Ergebnisse somit als statistisch unabhängig zu betrachten sind^{55, 232}.

$$\text{Var}(\ln\theta_{AB}) = \text{Var}(\ln\theta_{AX}) + \text{Var}(\ln\theta_{BX}) \quad (2)$$

$\text{Var}(\ln\theta_{AX})$ und $\text{Var}(\ln\theta_{BX})$ können mit den Ergebnissen der Metaanalysen für A versus X und B versus X mit den für die entsprechenden Effektschätzer bekannten Standardmethoden^{83, 177, 197} berechnet werden.

Des Konfidenzintervall des Gesamteffektschätzers für den indirekten Vergleich kann im anschließenden Schritt leicht berechnet werden.

$$95\% \text{-KI}(\ln\theta_{AB}) = \ln\theta_{AB} \pm 1,96 \sqrt{\text{Var}(\ln\theta_{AB})} \quad (3)$$

Durch Potenzierung des logarithmierten Effektschätzers sowie der Konfidenzintervallober- und untergrenze ($\theta_{AB} = e^{\ln\theta_{AB}}$) erhält man dann den Gesamteffektschätzer für den indirekten Vergleich mit seinem entsprechenden Konfidenzintervall²³².

Effektmaße, die für θ in diesen Formeln geeignet sind, sind das OR⁵⁵, das RR^{54, 151, 170} und das HR⁷⁴. Für kontinuierliche Daten gelten die gleichen Formeln, nur dass nicht mit logarithmierten Effektschätzern und logarithmierten Varianzen gerechnet werden muss^{58, 170, 183, 254}.

Für die absolute Risikoreduktion und die Number-needed-to-treat (NNT) als Effektschätzer haben Goadsby et al.¹¹¹ die vorgestellten Formeln leicht abgewandelt. Somit stehen Berechnungsformeln für alle wichtigen Effektmaße in der Wirksamkeitsbetrachtung therapeutischer Interventionen zur Verfügung.

2. Indirekter Vergleich mithilfe eines statistischen Tests

Methodenpapier(e): Cochrane Handbuch Kapitel 16.6.2 und 9.6.3.1⁸

Übersichtsarbeiten: Keine

Methodenanwendungen: Rice und Stead¹⁹², Antithrombotic Trialists' Collaboration²⁸, Small et al.²¹⁴ und Otto et al.¹⁸²

Werden nach der Durchführung der zwei Metaanalysen die Effektgrößen $\theta_{A \text{ versus } X}$ und $\theta_{B \text{ versus } X}$ erhalten, gibt es neben der Interpretation der Konfidenzintervalle auch die Möglichkeit, die Überlegenheit bzw. Nicht-Unterlegenheit einer Therapie im Vergleich zu der anderen über einen statistischen Test zu belegen.

Das Handbuch der Cochrane Collaboration empfiehlt für den indirekten Vergleich die Q-Statistik zum Vergleich von Subgruppen zu verwenden⁸. Dieser, auf einer χ^2 -Verteilung beruhende Test, der mit der Review Manager 5-Software ausgeführt werden kann⁸⁴, überprüft die Nullhypothese, dass keine Unterschiede in Therapieeffekten zwischen den beiden Subgruppen (in diesem Fall den zwei

Metaanalysen) bestehen. Verwendet wird dieses Verfahren in einem Cochrane Review von Rice und Stead¹⁹² und in einem Review der Antithrombotic Trialists' Collaboration²⁸.

Einige Beispiele aus der Literatur finden sich auch für die Verwendung des t-Tests nach Student: Small et al.²¹⁴ und Otto et al.¹⁸².

3. Indirekter Vergleich durch Vergleich von Konfidenzintervallen

Methodenpapier(e): Cochrane Handbuch⁸

Übersichtsarbeiten: Keine

Methodenanwendungen: U. a. Büttner et al.⁵⁹

In diesem Ansatz werden die Effektschätzer $\theta_{A \text{ versus } X}$ und $\theta_{B \text{ versus } X}$ in zwei getrennten Metaanalysen mit ihren entsprechenden Konfidenzintervallen berechnet. Es wird dann eine Therapie der anderen für statistisch signifikant überlegen erklärt, wenn sich die Konfidenzintervalle um die Einzelergebnisse nicht überschneiden^{8, 59}.

4. Narrativer indirekter Vergleich

Methodenpapier(e): Keine

Übersichtsarbeiten: Keine

Methodenanwendungen: Messerli et al.¹⁷¹ und Law et al.¹⁴⁸

Der narrative Ansatz soll am Beispiel des systematischen Reviews von Messerli et al.¹⁷¹ geschildert werden. Die Autoren führen zwei Sätze von Metaanalysen durch: Erstens die Bestimmung der Effektivität von Diuretika versus Placebo bei älteren Hypertonikern hinsichtlich unterschiedlicher kardiovaskulärer Therapieeffekte und zweitens die Bestimmung der Effektivität von Betablockern versus Placebo bei älteren Hypertonikern hinsichtlich der gleichen kardiovaskulären Therapieeffekte. In den Metaanalysen wurde für wichtige Therapieeffekte (wie z. B. Mortalität) eine signifikante Überlegenheit von Diuretika gegenüber Placebo nachgewiesen. Im Vergleich Betablocker versus Placebo hingegen wurde kein statistisch signifikanter Effektunterschied ermittelt. Aus diesen Befunden schließen die Autoren argumentativ auf eine Therapieüberlegenheit von Diuretika gegenüber Betablockern ohne dabei die Konfidenzintervalle der Effektschätzer der beiden Metaanalysen zu berücksichtigen.

5.3.3 Weiterentwicklung der adjustierten indirekten Vergleiche: Einschluss von Head-to-head-Studien

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Song et al.²¹⁸

Übersichtsarbeiten: Keine

Methodenanwendungen: Vandermeer et al.²⁵⁴

Methodenbeschreibung

In der Publikation von Bucher et al.⁵⁵ wird der Einschluss von Head-to-head-Studien in einen indirekten Vergleich aus zwei Gründen abgelehnt:

Erstens weil einige Head-to-head-Studien keinen Arm mit dem gemeinsamen Komparator für den indirekten Vergleich enthalten (im Beispiel in Abbildung 7 die Studie 5) und zweitens weil die Ergebnisse der Head-to-head-Studien mit Komparatorarm (im Beispiel in Abbildung 7 die Studie 4) in Metaanalysen eingehen würden. Damit ist die Vorbedingung der Unabhängigkeit für die Anwendung von Formel (2) nicht mehr gegeben¹⁵⁷. Weiterhin würden die Ergebnisse der Komparatorgruppe zweifach in die Metaanalysen eingehen³⁵, was dem Grundsatz widerspricht, dass jeder Patient in einer Metaanalyse nur einmal gezählt werden sollte.

Dennoch finden sich unter den systematischen Reviews mit indirekten Vergleichen drei Beispiele, die komparatorkontrollierte Head-to-head-Studien in ihren adjustierten indirekten Vergleich einschließen^{48, 146, 231}, ohne das erkenntlich wird, ob die Ergebnisse des Komparatorarmes in beide Metaanalysen eingehen und somit in der Gesamtanalyse zweimal verwendet werden.

Eine Lösung dieser Problematik wird von Song et al.²¹⁸ vorgeschlagen. Dabei wird zunächst ohne Berücksichtigung von Head-to-head-Studien ein adjustierter indirekter Vergleich mit Berechnung eines Gesamteffektschätzers (θ_{indirekt}) durchgeführt. Die Ergebnisse der verfügbaren Head-to-head-Studien werden mit konventionellen Metaanalysetechniken ebenfalls zu einem Gesamteffektschätzer (θ_{direkt}) zusammengefasst. Im Anschluss werden die beiden Gesamteffektschätzer aus dem indirekten und direkten Vergleich mit ihrer inversen Varianz gewichtet (ω_{indirekt} bzw. ω_{direkt}) und zu einem kombinierten Gesamteffektschätzer ($\theta_{\text{kombiniert}}$) analog dieser Formeln zusammengefasst:

$$\theta_{\text{kombiniert}} = \frac{\omega_{\text{indirekt}} \cdot \theta_{\text{indirekt}} + \omega_{\text{direkt}} \cdot \theta_{\text{direkt}}}{\omega_{\text{indirekt}} + \omega_{\text{direkt}}}$$

$$\omega_{\text{indirekt}} = \frac{1}{\text{Var}(\theta_{\text{indirekt}})} \quad \text{und} \quad \omega_{\text{direkt}} = \frac{1}{\text{Var}(\theta_{\text{direkt}})}$$

Am bereits eingeführten Beispiel aus zwölf Studien stellt sich dieser Ablauf schematisch wie folgt dar (siehe Abbildung 8):

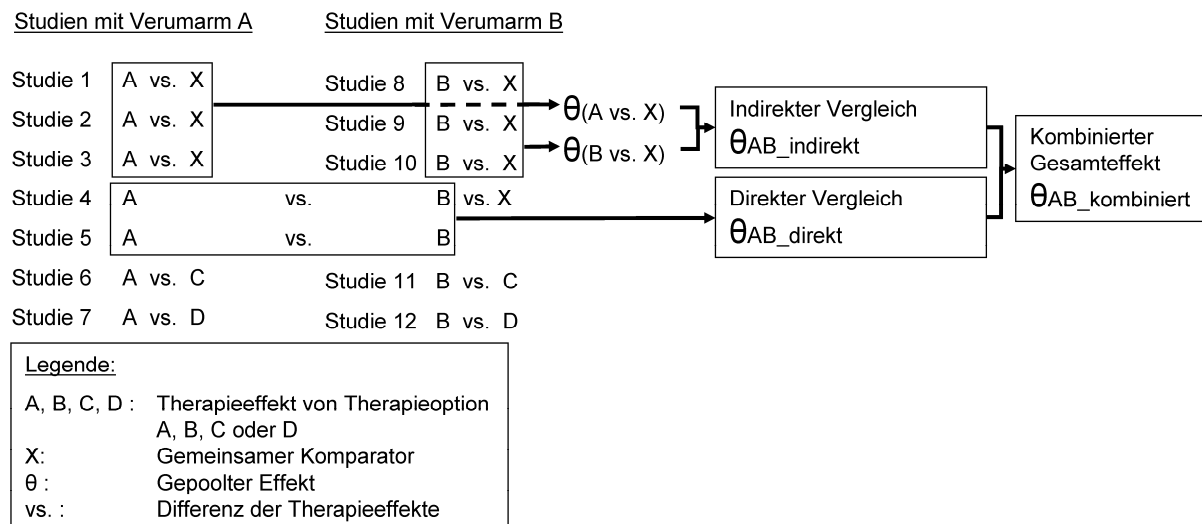


Abbildung 8: Adjustierter indirekter Vergleich mit Einschluss von kontrollierten Head-to-head-Studien

5.3.4 Metaregression

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Thompson et al. mit zwei Publikationen^{236, 237}

Übersichtsarbeiten: Glenny et al.¹⁰⁹ und Gartlehner et al.¹⁰⁴

Methodenanwendungen: U. a. Stettler et al.²²³, Mudge et al.¹⁷⁹, Kearney et al.¹³⁸, Peterson et al.¹⁸⁴ und Indolfi et al.¹³²

Begriffsbestimmung (Definition nach Thompson et al.²³⁶)

In Regressionsanalysen wird der quantitative Einfluss einer oder mehrerer erklärender Variablen auf eine Therapieeffektvariable untersucht. In Metaanalysen von Therapiestudien werden sogenannte Metaregressionen eingesetzt, um den Einfluss von Studiencharakteristika (erklärende Variablen) auf einen Therapieeffekt (Therapieeffektvariable) zu quantifizieren²³⁶.

Dieses zur Heterogenitätsaufklärung eingesetzte Verfahren soll die Größe des Gesamteffekts zu einem oder mehreren Studiencharakteristikum/a in Beziehung setzen²³⁶.

Methodenbeschreibung

Die Metaregression kann zur Durchführung eines indirekten Vergleichs genutzt werden, indem in der Regressionsgleichung die in den Einzelstudien ermittelten Therapieeffekte als die erklärenden Variablen und der gepoolte Gesamtherapieeffekt als Therapieeffektvariable festgelegt werden¹⁰⁹.

Die Regressionsgleichung dient als Modell, das die Wirklichkeit der Studienlage so gut wie möglich abbilden soll. Sie bringt die erklärenden Variablen mit der Therapieeffektvariablen (Gesamtherapieeffekt Y) in einen mathematischen Zusammenhang. Für kontinuierliche Daten wird ein linearer Zusammenhang zwischen den erklärenden und der Therapieeffektvariablen angenommen¹⁴.

Für dichotome Daten gilt prinzipiell die gleiche Regressionsgleichung. Einzige Ausnahme ist, dass alle Terme, die Schätzungen von Therapieeffekten beschreiben, logarithmiert in die Gleichung eingehen^{14, 179}. Dabei wird meistens mit der Logit-Funktion eine Transformation der natürlichen Logarithmusfunktion eingesetzt, die z. B. dem Odds ermöglicht einen unbeschränkten Wertebereich anzunehmen. Dies ist aus mathematischen Gründen notwendig, da normalerweise der Wertebereich für das Odds von Null bis unendlich beschränkt ist. Aufgrund der großen Bedeutung der Logarithmusfunktion in der Berechnung, wird die Metaregression mit dichotomen Effektschätzern auch als logistische (Meta-) Regression bezeichnet^{104, 237}.

Weitere Details der Modellierung der Regressionsgleichungen werden anhand eines Beispiels vorgestellt und in Abbildung 9 schematisch visualisiert: In dem Vergleich von Therapie A und Therapie B über den gemeinsamen Komparator X werden jeweils für Therapie A und Therapie B Regressionsgleichungen nach dem gleichen Muster aufgestellt. Das Prinzip soll exemplarisch für die Therapie A erklärt werden. Der gepoolte Gesamtherapieeffektunterschied zwischen A und X setzt sich aus den in den Einzelstudien 1, 2, ..., bis in Studie k beobachteten Effektunterschieden von A und X ($\theta_{A \text{ versus } X \text{ in Studie } 1}$ bis $\theta_{A \text{ versus } X \text{ in Studie } k}$) zusammen.

Die Effektunterschiede der Einzelstudien werden in einem Random oder Fixed effects-Modell gewichtet (in Abbildung 9 nicht dargestellt) und mit Korrekturfaktoren (b_1 bis b_k) multipliziert. Die b -Werte sind Abweichungsfaktoren zwischen den in der Metaregression modellierten und beobachteten Werten. Ergänzt wird zudem mit ϵ eine Zufallskomponente, in die Messfehler und nicht beobachtete Einflüsse auf Y eingehen. Für ϵ wird in der Regel der Wert Null angenommen.

Durch mathematische Operationen, auf die an dieser Stelle nicht genauer eingegangen werden soll, wird angestrebt, die Korrekturfaktoren (b_1 bis b_k) so zu optimieren, dass mithilfe der Regressionsgleichung der gepoolte Therapieeffektunterschied möglichst gut abgeschätzt werden kann. Auf diese Weise lassen sich Schätzungen der gepoolten Therapieeffektunterschiede von A versus X und B versus X ermitteln.

Meta-Regression für kontinuierliche Daten

$$b_0 + b_1\theta_{Avs.X \text{ in Studie } 1} + b_2\theta_{Avs.X \text{ in Studie } 2} + \dots + b_k\theta_{Avs.X \text{ in Studie } k} + \varepsilon = Y(\theta_{Avs.X})$$

$$b_0 + b_1\theta_{Bvs.X \text{ in Studie } 1} + b_2\theta_{Bvs.X \text{ in Studie } 2} + \dots + b_k\theta_{Bvs.X \text{ in Studie } k} + \varepsilon = Y(\theta_{Bvs.X})$$

Indirekter Vergleich durch:
 $Y_{Avs.X} \text{ vs. } Y_{Bvs.X}$

Meta-Regression für dichotome Daten (Logistische (Meta-)Regression)

$$b_0 + b_1\theta_{Avs.X \text{ in Studie } 1} + b_2\theta_{Avs.X \text{ in Studie } 2} + \dots + b_k\theta_{Avs.X \text{ in Studie } k} + \varepsilon = \text{Logit}(Y_{1/0} | \theta_{Avs.X})$$

$$b_0 + b_1\theta_{Bvs.X \text{ in Studie } 1} + b_2\theta_{Bvs.X \text{ in Studie } 2} + \dots + b_k\theta_{Bvs.X \text{ in Studie } k} + \varepsilon = \text{Logit}(Y_{1/0} | \theta_{Bvs.X})$$

Nach mathematischer Umformung indirekter Vergleich durch:
 $Y_{Avs.X} \text{ vs. } Y_{Bvs.X}$

Legende:

- A, B : Therapieeffekt von Therapieoption A oder B
- X: Gemeinsamer Komparator für den indirekten Vergleich
- vs. : Differenz der Therapieeffekte
- 1, 2, ..., k: Indices, die Einflüsseffekte aus Studie 1, 2... bis Studie k anzeigen
- Y: Funktion des gepoolten Gesamtherapieeffekts: „Abhängige Variable“
- θ: Therapieeffektunterschied einer beliebigen Studie: „Unabhängige Variable“ (Bei dichotomen Daten steht θ für den logarithmierten Therapieeffektunterschied)
- ε: Zufallskomponente, in die Messfehler und nicht beobachtete Einflüsse auf Y eingehen.
- b₀, b₁...b_k : Abweichungsfaktoren zwischen modellierten und beobachteten Werten
- 1/0: Index, der anzeigt, dass es sich um einen binären Outcome handelt, bei dem das Risiko oder die Wahrscheinlichkeit, dass ein Ereignis eintritt (1) oder nicht eintritt (0), ermittelt wird.

Abbildung 9: Indirekter Vergleich durch Metaregression

Erklärt wurde bisher, wie die gepoolten Therapieeffektunterschiede $Y_{A \text{ versus } X}$ und $Y_{B \text{ versus } X}$ erhalten werden können. Für den Vergleich von A versus B können wieder die vier Möglichkeiten Gesamteffektschätzer, statistischer Test, Vergleich der Konfidenzintervalle und narrativer Vergleich eingesetzt werden.

Für den Vergleich von A versus B über einen Gesamteffektschätzer wird dieser bei kontinuierlichen und dichotomen Daten unterschiedlich berechnet.

Bei kontinuierlichen Daten stellt die Differenz dieser beiden Therapieeffektunterschiede $Y_{A \text{ versus } X}$ und $Y_{B \text{ versus } X}$ das Ergebnis des indirekten Vergleichs von A und B über X dar¹⁸⁴.

Bei dichotomen Daten werden $Y_{A \text{ versus } X}$ und $Y_{B \text{ versus } X}$ zunächst durch mathematische Umformung wieder in OR oder andere entsprechende dichotome Therapieeffekte überführt¹⁷⁹. Im Anschluss wird der Vergleich von A und B durch die Bildung eines Ratios vorgenommen, indem der erhaltene Wert für $Y_{A \text{ versus } X}$ durch den von $Y_{B \text{ versus } X}$ dividiert wird^{138, 184, 223}.

Während es für den indirekten Vergleich mittels Metaregression Beispiele für den Vergleich mittels Gesamteffektschätzer^{138, 184, 223} und Konfidenzintervallvergleich¹³² gibt, werden für den Vergleich über einen statistischen Test und das narrative Verfahren keine Beispiele gefunden.

5.3.5 Weiterentwicklung der Metaregression: Aufnahme von Kovariaten in die Metaregression

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Keine

Übersichtsarbeiten: Van Houwelingen et al.²⁵²

Methodenanwendungen: U. a. Indolfi et al.¹³², Petersen et al.¹⁸⁴ und Mitte et al.¹⁷²

Methodenbeschreibung

In der Regressionsgleichung werden alle Summanden, die Terme für die erklärende Variable auf Studienebene darstellen, um Faktoren für Kovariatoreinflüsse ergänzt. Diese Faktoren beinhalten den Einfluss von Kovariaten auf den Therapieeffektunterschied der jeweiligen Einzelstudien.

Auf diese Weise kann in der Metaregression hinsichtlich potenzieller Confounder adjustiert werden. In der Publikation von van Houwelingen et al.²⁵² kann die Modellierung solcher komplexer Regressionsmodelle mit SAS nachvollzogen werden.

In den Literaturbeispielen mit indirekten Vergleichen wird häufig von der Adjustierung hinsichtlich Confoundern Gebrauch gemacht. In 12 von 17 Publikationen mit Metaregression wird mindestens eine Kovariate in die Regressionsgleichung mit aufgenommen (z. B. Indolfi et al.¹³², Petersen et al.¹⁸⁴ oder Mitte et al.¹⁷²).

5.3.6 Weiterentwicklung der Metaregression: Einschluss von kontrollierten Head-to-head-Studien (Mixed model)

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Hasselblad¹¹⁶, Brown et al.⁵², Glenny et al.¹⁰⁹ und Turner et al.²⁴⁶

Übersichtsarbeiten: Keine

Methodenanwendungen: Ballesteros³⁵, Eckert et al. mit zwei Publikationen^{91, 92} und Farré et al.⁹⁹

Methodenbeschreibung

Die geschilderte Metaregressionstechnik für den indirekten Vergleich kann nur zweiarmige Studien einschließen, die einen Therapiearm mit einer Therapieoption von Interesse (Therapie A oder B) und einen Therapiearm mit dem gemeinsamen Komparator X aufweisen. Head-to-head-Studien mit drei Armen (Therapieoption A, B und X) können nicht eingeschlossen werden. Der Grund hierfür ist, dass zwei Metaregressionen für den indirekten Vergleich durchgeführt werden müssen (siehe Abbildung 9). Eine dreiarmige Studie könnte man nicht in beide Metaregressionen einbeziehen, ohne die Probanden aus dem Kontrollarm in beiden Metaanalysen und somit in der Gesamtanalyse zweimal zu berücksichtigen. In der Folge wären die Ergebnisse aus den beiden Metaregressionen nicht mehr statistisch unabhängig voneinander, da sie teilweise auf das gleiche Datenmaterial zurückgreifen. Sind die Ergebnisse der beiden Metaregressionen nicht statistisch unabhängig, können Korrelationen zwischen ihren Ergebnissen auftreten, die zu einer Verzerrung des Gesamtergebnisses des indirekten Vergleichs führen können¹²².

Hasselblad¹¹⁶ entwickelte deshalb eine Methode zur Aufnahme von verschiedenen Therapien in eine Regressionsgleichung, ohne dass zwei Metaregressionen erforderlich sind. Er bezeichnet dieses Verfahren als multiple logistische Regression. Es kann sowohl als Fixed als auch als Random effects-Modell eingesetzt werden und eignet sich nicht nur für dichotome, sondern auch für kontinuierliche Daten.

Eine Weiterentwicklung des Random effects-Modells von Hasselblads multipler logistischer Regression nehmen Glenny et al.¹⁰⁹ mithilfe von Erkenntnissen aus einer Publikation von Brown et al.⁵² vor. Ihre Methode, für die sie einen SAS- und STATA-Code entwerfen (siehe Glenny et al.¹⁰⁹ Appendix 7), nennen sie „Mixed model“.

In der Literatur zu indirekten Vergleichen wird das sogenannte Mixed model bereits in drei systematischen Übersichtsarbeiten verwendet^{35, 91, 92}.

Eine weitere, dem Mixed model ähnliche Methode zum Einschluss von mehr als zweiarmigen Head-to-head-Studien, bzw. unterschiedlichen Dosen der untersuchten Medikation, verwenden Farré et al.⁹⁹ für ihren indirekten Vergleich und zitieren ein Methodenpapier von Turner et al.²⁴⁶. Dieses gibt darüber hinaus eine Anleitung, wie eine logistische Regression auf der Basis von individuellen Patientendaten berechnet werden kann.

5.3.7 Weiterentwicklung der Metaregression: Metaregression mit dem Bayes'schen Theorem

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Smith, Spiegelhalter und Thomas²¹⁵

Übersichtsarbeiten: Keine

Methodenanwendungen: Geddes et al.¹⁰⁷ und Robinson et al.¹⁹⁸

Methodenbeschreibung

Auch das von Smith, Spiegelhalter und Thomas²¹⁵ entwickelte Modell zur Berechnung einer Metaregression kann für die Durchführung eines indirekten Vergleichs eingesetzt werden^{107, 198}.

Das Modell ist dem beschriebenen Regressionsmodell ähnlich, nutzt jedoch die Bayes'sche Statistik. Eine kurze Einführung in die Bayes'sche Statistik wird im Anhang in Abschnitt 9.2 „Kurze Einführung in die Bayes'sche Statistik“ gegeben. Für nähere Informationen sei auf das Methodenpapier von Smith, Spiegelhalter und Thomas²¹⁵ verwiesen. Kurzgefasst nutzt eine Bayes'sche Analyse das Bayes'sche Theorem, um eine Priorverteilung eines unbekanntem Effekts durch Verarbeitung von Ergebnissen einer oder mehrerer Studien in eine Posteriorverteilung zu überführen. Als Priorverteilung wird das Vorwissen, das man über den Effekt vor Durchführung der Studie bzw. der Metaanalyse hat, bezeichnet und als Posteriorverteilung das Resultat der Bayes'schen Analyse.

Anwendungen dieser Methode als indirekte Vergleiche geben die Publikationen von Geddes et al.¹⁰⁷ und Robinson et al.¹⁹⁸. Geddes et al. liefern zudem ein Beispiel, wie bei der Metaregression mit Bayes'schem Theorem bezüglich Kovariaten, die den Therapieeffekt beeinflussen, adjustiert werden kann¹⁰⁷.

5.3.8 Mixed treatment comparison (MTC)

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Lumley¹⁶⁰, Lu und Ades¹⁵⁷⁻¹⁵⁹, Eddy⁹³⁻⁹⁵; Higgins 1996¹²², Gleser und Olkin¹¹⁰, Hirotsu und Yamada¹²⁶, Dominici⁸⁹, Hasselblad¹¹⁶ und Ades^{20, 21, 23}

Übersichtsarbeiten: Salanti²⁰⁴, Sutton et al.^{226, 229}, Spiegelhalter et al.²²¹ und Hasselblad¹¹⁸

Methodenanwendungen: Kyrgiou et al.¹⁴², Psaty et al.¹⁸⁹, Stettler et al.²²⁴, Jansen et al.¹³³, Lam et al.¹⁴³ und Nixon et al.¹⁸⁰

Einleitung und Begriffsdefinition

Mixed treatment analysis steht als Sammelbegriff für indirekt vergleichende Methoden, die mehr als zwei Therapieoptionen simultan miteinander vergleichen können. Der Begriff Mixed treatment analysis für indirekt vergleichende Methoden mit dieser Eigenschaft wird in der Literatur nicht durchgängig verwendet. Die Cochrane Collaboration gebraucht den Begriff Multiple Treatment Meta-Analysis⁸ und wichtige Methodenpapiere von Lumley et al.¹⁶⁰ den Ausdruck Network Meta-Analysis. Diese Termini können synonym mit dem von uns benutzten Begriff des MTC verwendet werden⁸. Die Bezeichnung MTC wird gewählt, da sich mit dem NICE aus Großbritannien eine erste HTA-Institution intensiv mit dieser Methodik beschäftigt und diesen Begriff verwendet²²⁶.

Als deutscher Begriff für den MTC wird die Übersetzung von Lumleys Bezeichnung „Network Meta-Analysis“ vom Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) vorgeschlagen. In einer kurz vor der Fertigstellung dieses HTA-Berichts veröffentlichten IQWiG-Publikation wird ein MTC eingesetzt und als Netzwerk Metaanalyse bezeichnet¹³.

Einsatzgebiete des MTC

In Forschungsfeldern, in denen viele Therapieoptionen miteinander konkurrieren, ist eine wichtige Frage, wie diese in eine Rangfolge nach ihrer Wirksamkeit gebracht werden können, um die beste Alternative auszuwählen²⁰⁴. Dies hat zu der Entwicklung des MTC geführt. Er bietet weiterhin die Möglichkeit, über den Einschluss aller zu einem Therapievergleich zur Verfügung stehenden RCT eine präzisere Schätzung des Gesamteffektes zu erreichen.

Dies soll an dem bisher verwendeten Beispiel aus zwölf Studien demonstriert werden (siehe Abbildung 10).

Studien mit Verumarm A			Studien mit Verumarm B		
Studie 1	A vs. X		Studie 8	B vs. X	
Studie 2	A vs. X		Studie 9	B vs. X	
Studie 3	A vs. X		Studie 10	B vs. X	
Studie 4	A	vs.	B vs. X		
Studie 5	A	vs.	B		
Studie 6	A vs. C		Studie 11	B vs. C	
Studie 7	A vs. D		Studie 12	B vs. D	

Abbildung 10: Beispiel für zwölf Studien mit Indizien zur Wirksamkeit der Interventionen A und B

Zur Erinnerung: Der indirekte Vergleich nach Bucher et al.⁵⁵ kann nur die Studien 1 bis 3 und 8 bis 9 einschließen. Die Head-to-head-Studien (Studie 4 und 5) können über die Methodenerweiterung von Song²¹⁸ zusätzlich berücksichtigt werden. Keine der bisher vorgestellten Methoden zur Durchführung indirekter Vergleiche ist aber in der Lage, die Studien 6, 7, 11 und 12 einzuschließen. Dass diese Studien auch wichtige Informationen für den Wirksamkeitsvergleich von A und B geben können, wird deutlich, wenn man das Beispiel in Form eines Evidenznetzwerks (siehe Kapitel 2.1.2 „Geometrie von Evidenznetzwerken“) darstellt (siehe Abbildung 11).

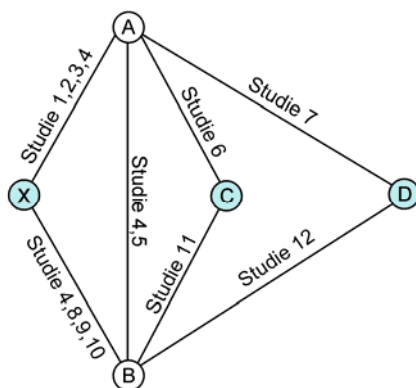


Abbildung 11: Darstellung des Beispiels als Evidenznetzwerk

Nicht nur über X als gemeinsamen Komparator, sondern auch über C und D kann der indirekte Vergleich von A und B durchgeführt werden. In diesem Beispiel gibt es insgesamt vier Möglichkeiten den Therapieeffektunterschied von A und B zu ermitteln:

1. Direkt über die zwei Head-to-head-Studien
2. Indirekt über X als gemeinsamen Komparator
3. Indirekt über C als gemeinsamen Komparator
4. Indirekt über D als gemeinsamen Komparator.

Der MTC geht weiter als diese vier Ansätze, indem er die gesamte verfügbare Evidenz aus Studien berücksichtigt¹⁵⁷. In dem Beispiel fasst er deshalb die Evidenz aus allen vier Möglichkeiten simultan zu einem Ergebnis des Therapieeffektvergleichs von A und B zusammen. Auf diese Weise kann er die Aussagekraft des direkten Vergleichs, der eventuell aufgrund zu geringer Fallzahlen ein weites Konfidenzintervall aufweist, durch Hinzunahme von Evidenz aus indirekten Vergleichen erhöhen¹⁵⁷.

Darüber hinaus kann er auch noch die Wirksamkeit von A, B, C, D und X in eine Rangfolge bringen, mit einer Aussage, welche Therapie mit einer bestimmten Wahrscheinlichkeit die wirksamste ist¹⁵⁷.

Eine weitere Anwendungsmöglichkeit des MTC ist es, Lücken im Evidenznetzwerk zu schließen¹⁵⁷. Ein Evidenznetzwerk ist erst vollständig, wenn alle Therapieoptionen Head-to-head in Studien miteinander verglichen wurden und eine polygonale Geometrie resultiert (siehe Kapitel 2.1.2 „Geometrie von Evidenznetzwerken“). In unserem Beispiel fehlen z. B. Head-to-head-Studien für den Vergleich von C und D. Ein MTC mit diesen zwölf Studien könnte hier einen Wert für den Therapieeffektunterschied von C und D schätzen. Dieser Wert kann bei der Berechnung der nötigen Fallzahl helfen, wenn eine Head-to-head-Studie mit C versus D geplant wird. Bei einem Fehlen von Head-to-head-Studien in einem Evidenznetzwerk können MTC somit wertvolle Hilfestellung für die Planung des Designs einer solchen Head-to-head-Studie liefern¹⁵⁷ oder sein mögliches Ergebnis abschätzen, falls seine Durchführung aus ethischen oder anderen Gründen nicht möglich ist¹⁶⁰.

Der MTC nach Lu und Ades¹⁵⁷ kann in allen Evidenznetzwerken durchgeführt werden²². Dies schließt sehr komplexe reale Netzwerke genauso mit ein, wie Netzwerke, die auf polygonalen, sternförmigen oder linearen Strukturen beruhen²⁰⁵ (siehe Abbildung 2). Einzige Voraussetzung ist, dass keine Unterbrechung des Netzwerks wie in Abbildung 1f vorliegt²³.

Vorläufer des MTC

Eddy publizierte zwischen 1989 und 1992 mit der Confidence profile-Methode⁹³⁻⁹⁵ eine Pionierarbeit für die Anwendung von Bayes'schen Verfahren in der evidenzbasierten Medizin²²¹. Die Confidence profile-Methode eignet sich zur Durchführung indirekter Vergleiche⁹⁴ und kann bezüglich Confoundern adjustieren^{95, 118}. Sie wurde dennoch unter Statistikern und Epidemiologen nicht sehr bekannt²²¹. Zusammen mit der richtungweisenden Publikation von Higgins und Whitehead „Borrowing strength from external trials in meta-analysis“¹²² sind die Ideen der Confidence profile-Methode in den von Lumley¹⁶⁰ sowie Lu und Ades¹⁵⁷ publizierten Verfahren des MTC^{157, 229} aufgegangen. In der Praxis werden fast ausschließlich die beiden letztgenannten Publikationen von Reviewautoren in ihren Methodikkapiteln zitiert, wenn indirekte Vergleiche multipler Therapieoptionen mit Bayes'schen Verfahren vorgenommen werden (siehe Tabelle 26). Die Verfahren von Lumley und Lu und Ades sind sich sehr ähnlich. Der wichtigste Unterschied besteht darin, dass Lumleys Modell keine Studien mit mehr als zwei Therapiearmen einschließen kann^{157, 160}.

Andere Methoden, die auch zu den MTC gehören, sich aber nicht durchgesetzt haben, sind die von Gleser und Olkin¹¹⁰, Hirotsu und Yamada¹²⁶, Dominici⁸⁹ und Hasselblad¹¹⁶.

In der Folge wird der MTC nach Lu und Ades¹⁵⁷ vorgestellt, der weitestgehend der Network Meta-Analysis von Lumley et al.¹⁶⁰ entspricht. Die in diesen beiden Methodenpapieren beschriebenen Verfahren werden in zehn der zwölf systematischen Reviews, die einen MTC durchführen, eingesetzt (siehe Tabelle 26).

Die Methodentheorie der MTC nach Lu und Ades¹⁵⁷

Lu und Ades nehmen das Modell von Smith, Spiegelhalter und Thomas²¹⁵, das unter dem Abschnitt „Metaregression mit Bayes' Theorem“ besprochen wird, auf und erweitern es von einem Modell für Vergleiche von zwei Therapieoptionen zu einem Modell mit multiplen Therapieoptionen. Der MTC ist somit eng verwandt mit der Metaregression.

Die Modellierung soll an einem Beispiel aus vier Therapieoptionen A, B, C und D erklärt werden. Eine Therapie wird als Referenztherapie festgelegt, wie z. B. Therapie A. Relativ zum Therapieeffekt von A werden die sogenannten Basisparameter dAB, dAC und dAD erhoben. Dabei ist dAB die Differenz des Therapieeffekts von A zum Therapieeffekt von B, dAC die Differenz zum Therapieeffekt von C und dAD die Differenz zum Therapieeffekt von D. Im Gegensatz zu den Basis-Parametern haben die sogenannten funktionalen Parameter keinen direkten Bezug zu A. Es sind die Übrigen: dBC, dBD und dCD. Sie heißen funktionale Parameter, da sie Funktionen der Basisparameter nach den Formeln

$$dBC = dAB - dAC$$

$$dBD = dAB - dAD \text{ und}$$

$$dCD = dAC - dAD \text{ darstellen.}$$

Voraussetzung für diese Gleichungen ist ein konsistentes Netzwerk (siehe Abschnitt „Die Konsistenzannahme“).

Im Rahmen des MTC hat die Evidenz aus den funktionalen Parametern über diese Formeln Einfluss auf die Basisparameter. Dabei folgt der Bayes'sche Ansatz einem iterativem Konzept, das in vielen Schritten die Evidenz verarbeitet (siehe Kapitel 9.2 „Einführung in die Bayes'sche Statistik“). Ausgangspunkt ist die Priorverteilung. Für einen MTC müssen Priorverteilungen für mindestens die folgenden Parameter festgelegt werden: Die Therapieeffektschätzer für die Einzelstudien und ihre Varianzen, die gepoolten Therapieeffektschätzer für die Basisparameter sowie für die Varianz in und zwischen den Studien einen Heterogenitätsparameter¹⁵⁷.

Dabei wird empfohlen non-informative (vague) Priorverteilungen zu benutzen. Non-informative Priorverteilungen setzen im Unterschied zu informativen Priorverteilungen kein Vorwissen über Therapieeffekte ein, sodass die Studienergebnisse in den Endergebnissen die Priorannahmen vollständig dominieren können^{60, 143}. Diesem Ansatz folgen sieben von zwölf Autoren der systematischen Reviews mit MTC und ein weiterer wählt teilweise informative und non-informative Priorverteilungen. Die restlichen vier Autorentams machen in dieser Frage keine Angaben in ihrer Publikation (siehe Tabelle 26). Von den acht Autoren, die Angaben zu der Wahl ihrer Priorverteilungen machen, führen drei auch Sensitivitätsanalysen durch, die die Annahmen für die Priorverteilungen überprüfen.

Durch den Heterogenitätsparameter spielt schon bei der Wahl der Priorverteilungen die Entscheidung für Fixed oder Random effects eine Rolle. Wird von einem homogenen Datensatz ausgegangen, kann der Heterogenitätsparameter gleich Null gesetzt werden. Es gibt Methodenpapiere, die eine solche Fixed effects-Variante vorstellen^{20, 21} und dabei auf die Confidence profile-Methode aufbauen.

Wird von einem heterogenen Datensatz ausgegangen, kann dem Heterogenitätsparameter ermöglicht werden, die Variation zwischen (between-trial heterogeneity) und in (instudy heterogeneity) den Studien auszudrücken^{157, 221} (Random effects-Modell), was sich in geringerer Präzision der Effektschätzer in der Posteriorverteilung auswirken kann¹⁵⁹.

Da aber von Heterogenität in einem MTC ausgegangen werden kann²¹, wird in den meisten Fällen mit Random effects-Modellen eine passendere Modellierung des Datensatzes erreicht. In den zwölf Reviews mit MTC werden ohne Ausnahme Random effects-Modelle verwendet. Erwähnenswert ist in diesem Zusammenhang auch, dass Woolacott et al.²⁶⁴ sowohl ein Random als auch Fixed effects-Modell durchführen und sich nach einem Test auf Geeignetheit der Modelle für das Random effects-Modell entscheiden. Lu und Ades¹⁵⁷ haben für unterschiedliche Szenarien Fixed und Random effects-Modelle für MTC entwickelt. Es handelt sich um:

1. Ein Fixed effects-Modell
2. Ein Random effects-Modell, das Korrelationen durch mehrarmige Studien ignoriert
3. Ein Random effects-Modell für zwei- und dreiarmlige Studien
4. Ein Random effects-Modell für mehrarmige Studien.

Sie geben eine Einleitung²³ zu ihrer Verwendung und die zugehörigen WinBugs-Codes auf der Homepage www.bris.ac.uk/cobm/research/mpes/mixed-treatment-comparisons.html bekannt. WinBugs ist eine kostenfreie Software, die in der Berechnung von Bayes'schen Modellen in der medizinischen Forschung breite Verwendung findet. Neun der zwölf Autoren verwenden sie für ihren MTC. Für die Modellierung ist es wichtig, ob Studien mit mehr als zwei Therapiearmen mit in die Analyse aufgenommen werden sollen, da bei mehrarmigen Studien Korrelation auftreten kann¹²². Korrelation kann immer dann auftreten, wenn Parameter nicht von statistisch unabhängigen Datensätzen stammen. Dies ist bei mehrarmigen Studien der Fall, da die verschiedenen Arme mit aktiven Therapien den gleichen Placeboarm als Kontrollgruppe haben. Das Random effects-Modell für mehrarmige Studien berücksichtigt die mögliche Korrelation durch mehrarmige Studien²³. Auch wenn bei kleinen oder wenigen mehrarmigen Studien kein großer Einfluss der Korrelation auf das Endergebnis zu erwarten ist, so empfehlen Ades et al.²³ die Korrelation dennoch immer mit im Modell zu berücksichtigen.

Die besondere Fähigkeit des MTC, die Wahrscheinlichkeit berichten zu können, dass eine Therapie unter den untersuchten die Wirksamste ist, wird mithilfe der Markov-Ketten Monte-Carlo-Simulation¹⁰⁸ in WinBugs berechnet. Die Priorannahme ist, dass alle Therapien mit der gleichen Wahrscheinlichkeit die Wirksamsten sind¹⁴².

Durch die Studienergebnisse kann sich diese Wahrscheinlichkeit in der Posteriorverteilung für eine Therapie mehr oder weniger deutlich von den Wahrscheinlichkeiten für die anderen Therapien absetzen. So ermittelten z. B. Golfinopoulos et al.¹¹² beim fortgeschrittenen Kolonkarzinom eine Wahrscheinlichkeit von 65 %, dass von den zwölf betrachteten Therapieregimen eine fluorouracil-basierte Irinotecan plus Bevacizumab-Therapie die Krankheitsprogression am Wirksamsten verzögert. Die anderen betrachteten Therapieregime sind mit deutlich geringerer Wahrscheinlichkeit oder einer Wahrscheinlichkeit von 0 % die wirksamste Therapieoption.

Für Therapieeffekte, die zu unterschiedlichen Nachbeobachtungszeiten erhoben wurden, bietet eine Publikation von Lu und Ades eine Methodenerweiterung¹⁵⁸.

Die Konsistenzannahme

Neben der Heterogenität in (within-study heterogeneity) und zwischen den Studien (between-study heterogeneity) kommt im MTC noch eine dritte Ebene der Heterogenität hinzu, genannt Inkonsistenz^{159, 234} (Synonym zu Lumleys Begriff Inkohärenz¹⁶⁰). Inkonsistenz ist definiert als die Diskrepanz zwischen dem Ergebnis aus einem direkten und einem oder mehreren indirekten Vergleichen für den gleichen Wirksamkeitsvergleich, die nicht mehr durch Zufallsfehler erklärbar ist¹⁵⁹. Der MTC baut auf der Konsistenzannahme auf. Das heißt, er nimmt an, dass keine (Fixed effects-Modell) oder eine um den Wert Null normalverteilte (Random effects-Modell) Inkonsistenz in dem zu analysierenden Evidenznetzwerk vorliegt. In anderen Worten bedeutet dies, dass er davon ausgeht, dass die direkten und indirekten Vergleiche im Evidenznetzwerk die gleichen Effekte messen.

Inkonsistenz ist nur in geschlossenen Schleifen von Evidenznetzwerken messbar¹⁶⁰, da in diesen Schleifen immer die Möglichkeit besteht, zwei beliebige Therapieoptionen sowohl über einen direkten als auch über einen indirekten Vergleich einander gegenüber zu stellen. Dies soll an einem Beispiel aus drei Therapieoptionen, die in ihrer Darstellung eines Evidenznetzwerks eine Schleife bilden, verdeutlicht werden (siehe Abbildung 12). In der Realität weist ein Evidenznetzwerk meistens mehrere solcher Schleifen auf, die auch mehr als nur drei Therapien umfassen können²⁰⁵.

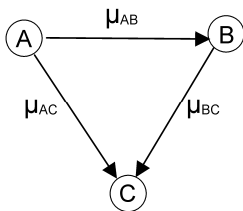


Abbildung 12: Evidenznetzwerk aus drei Therapieoptionen

Angenommen, es gibt die Therapieoptionen A, B und C und die Therapieeffektunterschiede μ_{AB} , μ_{BC} und μ_{AC} aus drei paarweise vergleichenden Metaanalysen. Unter Anwendung der Konsistenzannahme kann aus den Ergebnissen von zwei Metaanalysen der Effekt der dritten bestimmt werden^{159, 160, 204}. Dies verdeutlichen folgende drei Gleichungen, die durch Umformung ineinander überführbar sind:

$$\mu_{AC} = \mu_{AB} + \mu_{BC}$$

$$\mu_{BC} = \mu_{AC} - \mu_{AB}$$

$$\mu_{AB} = \mu_{AC} - \mu_{BC}$$

Die Terme links vom Gleichheitszeichen stellen direkte Vergleiche dar und die Terme rechts der Gleichheitszeichen indirekte Vergleiche, die den gleichen Effekt schätzen, wie der direkte Vergleich. Weichen in der Realität die Ergebnisse des indirekten Vergleichs von dem des direkten Vergleichs ab, weil nicht die gleichen Studienbedingungen herrschen, müssen Fehlerterme in diese Gleichungen aufgenommen werden:

$$\mu_{AC} = \mu_{AB} + \mu_{BC} - \Phi_1$$

$$\mu_{BC} = \mu_{AC} - \mu_{AB} + \Phi_1$$

$$\mu_{AB} = \mu_{AC} - \mu_{BC} + \Phi_1$$

Φ_1 ist der Fehlerterm, der die Inkonsistenz in dieser Schleife beinhaltet. Weist ein Evidenznetzwerk mehrere Schleifen auf, so gibt es auch mehrere Fehlerterme Φ_i . Ein Random effects-Modell für die Inkonsistenz geht davon aus, dass die Fehlerterme Φ_i um den Wert Null herum normalverteilt sind. Die Standardabweichung der Verteilung der Fehlerterme stellt das Ausmaß der Inkonsistenz des gesamten Evidenznetzwerks dar. Sie lässt sich quantifizieren und als sogenannter Inkonsistenzfaktor ausdrücken. Über den Inkonsistenzfaktor kann die Inkonsistenz in die Bayes'sche Modellierung mit aufgenommen werden. Publikationen von Lu et al.¹⁵⁹ und Lumley et al.¹⁶⁰ bieten Methoden hierfür. Eine Gegenüberstellung der zwei Verfahren bietet Salanti²⁰⁴. Vergleichbar ist die Aufnahme der Inkonsistenz in den MTC über den Inkonsistenzfaktor mit der Aufnahme von Heterogenität in eine konventionelle Metaanalyse durch ein Random effects-Modell. Beide führen zu einem Verlust an Präzision, was die Heterogenität bzw. Inkonsistenz des Datensatzes widerspiegelt.

Liegt moderate Inkonsistenz und/oder Heterogenität vor, gibt es verschiedene Möglichkeiten mit dem MTC weiter zu verfahren. Einige Möglichkeiten wurden in Kapitel 5.2.2 „Umgang mit Heterogenität in systematischen Reviews mit indirekten Vergleichen“ bereits vorgestellt. In den zwölf systematischen Reviews, die MTC durchführen, finden sich weitere Alternativen, die kurz in Tabelle 1 und ausführlicher im Anhang in Tabelle 27 dargelegt werden.

Tabelle 1: Möglichkeiten bei Inkonsistenz und/oder Heterogenität mit einem MTC zu verfahren

Möglichkeiten	Ausgeführt in
Verzicht auf das Poolen der Daten und dem MTC	Methodenpapier von Lumley et al. ¹⁶⁰
Ausschluss von Studien, die die Inkonsistenz maßgeblich verursachen	Methodenpapier von Lu und Ades ¹⁵⁹
Berücksichtigung eines Inkonsistenzfaktors in der Modellierung ^{159, 160}	- MTC von Psaty et al. ¹⁸⁹ und Kyrgiou et al. ¹⁴² nach dem Methodenpapier von Lumley et al. ¹⁶⁰ - MTC von Stettler et al. ²²⁴ nach dem Methodenpapier von Lu und Ades ¹⁵⁹
Nach Kovariaten stratifizierte Analyse	MTC von Jansen et al. ¹³³ , Lam et al. ¹⁴³ und Stettler et al. ²²⁴
Multivariate Analyse mit Adjustierung bezüglich Kovariaten	MTC von Jansen et al. ¹³³ , Stettler et al. ²²⁴ und Nixon et al. ¹⁸⁰ . Nixons Publikation weist einen umfangreichen Methodenteil auf.

MTC = Mixed treatment comparison.

5.3.9 Sonstige Methoden

Die Literaturrecherche identifizierte fünf systematische Reviews, die Methoden für indirekte Vergleiche verwandten, die sich nicht den bisher vorgestellten Methoden zuordnen lassen. Da es sich in der Tat um Einzelfälle handelt, werden die benutzten Methoden nur kurz beschrieben und nicht ausführlich diskutiert.

Brophy und Lawrence⁵¹

Brophy und Lawrence stellen drei Bayes'sche Methoden vor, wovon zwei indirekte Vergleiche darstellen. Die eine indirekt vergleichende Methode nutzt als Informationen für die Priorverteilung die Ergebnisse einer Head-to-head-Studie und verarbeitet sie zusammen mit den Ergebnissen aus einem indirekten Vergleich zur Likelihood-Funktion. Hier ergibt sich, dass das Ergebnis einer Head-to-head-Studie der Arzneistoffe Tirofiban und Abciximab durch Evidenz aus placebokontrollierten Studien mit diesen Arzneistoffen gestärkt wird. Das heißt, dass durch zusätzliche Evidenz aus dem indirekten Vergleich die Präzision des Effektschätzers des direkten Vergleichs erhöht wird. Dies könnte auch durch einen MTC erreicht werden, allerdings fehlen bei Brophy und Lawrence Angaben, dass sie sich auf diese Methodik beziehen.

Die andere indirekt vergleichende Methode ähnelt einem adjustierten indirekten Vergleich nach Bucher et al.⁵⁵ (siehe Kapitel 5.3.2 „Adjustierter indirekter Vergleich“), der allerdings mit Bayes'schen Verfahren durchgeführt wird. In diesem Beispiel werden jeweils Abciximab versus Tirofiban und Abciximab versus Eptifibatid paarweise über Placebo als gemeinsamem Komparator verglichen. Eine Head-to-head-Studie wird in dieser Methodik nicht berücksichtigt. Angaben über die Wahl der Parameter für die Wahl der Priorverteilung fehlen.

Ob die Randomisation der RCT in den zwei geschilderten indirekt vergleichenden Methoden bewahrt wird, ist aus den von Brophy und Lawrence gegebenen Methodenbeschreibungen nicht eindeutig ersichtlich.

Thijs et al.²³⁴

Thijs et al. verwenden einen multivariaten Ansatz einer Metaregression und berufen sich auf das Methodenpapier von van Houwelingen et al.²⁵². Van Houwelingens Modell ermöglicht es, die verschiedenen Therapieoptionen in einer multiplen Metaregression als Kovariaten zu berücksichtigen und verzichtet dabei auf ein Bayes'sches Verfahren. Das Modell kann, wie von Thijs et al. durchgeführt, sowohl als Fixed als auch als Random effects-Modell in SAS mit dem PROC MIXED- Modul berechnet werden. Da das von Thijs et al. analysierte Evidenznetzwerk aus vier Thrombozytenaggregationshemmerregimen und Placebo konsistent ist (nach Lumley¹⁶⁰ berechnet), und auch keine signifikante Heterogenität vorliegt, berichtet Thijs nur die Ergebnisse des Fixed effects-Modells. Trotz des hohen Komplexitätsgrads des Netzwerks kommt Thijs ohne Bayes'sche Verfahren aus. Dies ist der Hauptunterschied zu dem in Kapitel 5.3.8 „Mixed treatment comparison (MTC)“ beschriebenen MTC. Letztendlich vergleichen jedoch auch Thijs et al. multiple Interventionen und bringen sie in eine Wirksamkeitsrangfolge. Zur genaueren Abgrenzung der von Thijs verwendeten multiplen Regression von dem MTC nach Lu und Ades¹⁵⁷ bzw. Lumley et al.¹⁶⁰ fehlen genauere Angaben. Auch, ob von einer Bewahrung des Randomisationseffekts ausgegangen werden kann, ist aufgrund fehlender Angaben nicht sicher.

Tudur Smith et al.²⁴⁰

Tudur Smith et al. stellen den bisher einzigen indirekten Vergleich vor, der ausschließlich mit individuellen Patientendaten durchgeführt wird. Dies erlaubt es Tudur Smith et al. eine Fixed effects multiple Metaregression auf Patienten- und nicht Studienebene durchzuführen²⁴⁴. Ähnlich wie bei Thijs et al.²³⁴ modellieren Tudur Smith et al. die Therapieoptionen als Kovariate. Zusätzlich wird noch der Confounder Publikationsjahr als Kovariate berücksichtigt. Hauptunterschied zum MTC ist, dass Tudur Smith et al. aufgrund der individuellen Patientendaten auf einen Bayes'schen Ansatz verzichten können. Dennoch bewahrt diese Methode, gleich den MTC, die Randomisation und kann die Therapieoptionen in eine Rangfolge nach ihrer Wirksamkeit bringen. Im Appendix publizieren die Autoren die für ihr Modell benutzten Formeln. Die verwendete Methodik ähnelt dem MTC in hohem Maß und kann, wie auch von den Autoren selbst, als MTC bezeichnet werden.

Hind et al.¹²⁵ und **Moore et al.**¹⁷⁶

Die in den systematischen Reviews von Hind et al.¹²⁵ und Moore et al.¹⁷⁶ eingesetzte Methode ähnelt dem adjustierten indirekten Vergleich, schlägt aber ein abweichendes Vorgehen zur Berücksichtigung der Ergebnisse aus den Kontrollgruppen vor. Im Gegensatz zum Ansatz von Bucher et al.⁵⁵ werden hier die Differenzen der Effektgrößen zwischen Verum- und Kontrollgruppe nicht auf der Einzelstudienebene gebildet, sondern analog dem nicht-adjustierten indirekten Vergleich aus den einzelnen Studienarmen.

Zur besseren Abgrenzbarkeit vom adjustierten und nicht-adjustierten indirekten Vergleich wird in Abbildung 13 an dem bereits verwendeten Beispiel aus zwölf Studien demonstriert, wie ein indirekter Vergleich mit der von Moore et al. und Hind et al. benutzten Methodik aussehen würde.

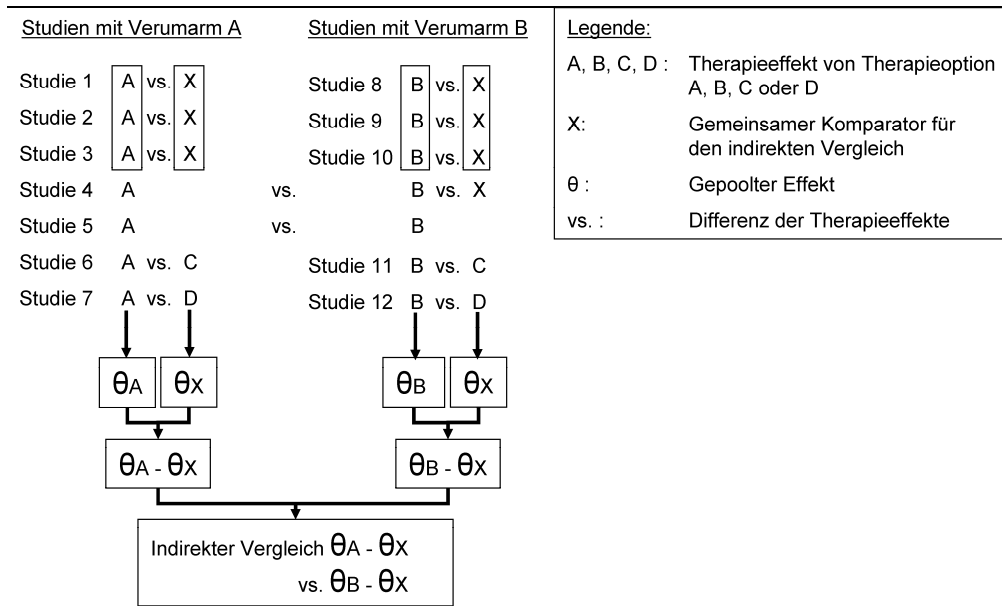


Abbildung 13: Indirekter Vergleich nach Moore et al. und Hind et al.

5.3.10 Methoden ohne Metaanalysen

5.3.10.1 Indirekter Vergleich von Einzelstudien

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Keine

Übersichtsarbeiten: Keine

Methodenanwendungen: Boland et al.⁴⁷, Jones et al.¹³⁶, Swift et al.²³⁰, Hofmann et al.¹²⁸ und Kristensen et al.¹⁴¹

Begriffsbestimmung

Sind jeweils nur eine Studie für den Vergleich A versus X und eine Studie für den Vergleich B versus X verfügbar, so erübrigt sich die Notwendigkeit Metaanalysen durchzuführen und der indirekte Vergleich A versus B kann mit den Ergebnissen dieser beiden Studien durchgeführt werden.

Methodenbeschreibung

Zur Anwendung kommen die in Abschnitt 5.3.1 „Nicht-adjustierter indirekter Vergleich“ besprochenen nicht-adjustierenden indirekten Vergleiche (Beispiel bei Boland et al.⁴⁷) und die in Abschnitt 5.3.2 „Adjustierter indirekter Vergleich“ besprochenen adjustierten indirekten Vergleiche (Beispiele bei Jones et al.¹³⁶, Swift et al.²³⁰ und Hofmann et al.¹²⁸).

Da nur die Ergebnisse von Einzelstudien miteinander verglichen werden, ist eine Metaanalyse nicht notwendig, um für den nicht-adjustierten indirekten Vergleich die Effektschätzer θ_A und θ_B und für den adjustierten indirekten Vergleich die Effektschätzer $\theta_{A \text{ versus } X}$ und $\theta_{B \text{ versus } X}$ zu bekommen.

Der Therapieeffektunterschied zwischen A und B wird aus diesen jeweiligen zwei Effektschätzern quantitativ oder qualitativ wieder über eine der vier bereits vorgestellten Möglichkeiten für einen indirekten Vergleich berechnen:

1. Über einen Gesamteffektschätzer $\theta_{A \text{ versus } B}$
2. Über einen statistischen Test
3. Über Vergleich der Konfidenzintervalle
4. Über ein narratives Gegenüberstellen der Therapieeffekte von A und B.

5.3.10.2 Deskriptive, nicht-quantitative indirekte Vergleiche (narrativ)

Literaturquellen für die Methodenbeschreibung

Methodenpapier(e): Keine

Übersichtsarbeiten: Keine

Methodenanwendungen: Berry et al.⁴⁴, Verma et al.²⁵⁵ und Yabroff et al.²⁶⁶

Begriffsbestimmung

Liegen mehr als eine Studie pro zu vergleichender Intervention vor, aber aus bestimmten Gründen wird auf die Durchführung von Metaanalysen verzichtet, nehmen einige Autoren eine deskriptiv beschreibende Zusammenfassung der Studienlage vor, ohne Studienergebnisse zu poolen.

Methodenbeschreibung

Dabei werden die einzelnen Studienergebnisse wertend einander gegenübergestellt ohne sie zusammenzufassen. Eine andere Möglichkeit ist es, sie teilweise zusammenzufassen, indem aus den Ergebnissen mehrerer ähnlicher Studien, Effektintervalle (kleinster bis größter Effektschätzer dieser Studiengruppe) gebildet werden²⁶⁶. In beiden Fällen werden die Studienergebnisse jedoch nicht wie bei Metaanalysen zu einem Effektschätzer gepoolt.

Der indirekte Vergleich erfolgt deskriptiv, indem die Ergebnisse der Studien mit A versus X denen der Studien mit B versus X beschreibend gegenübergestellt werden.

Auch für deskriptive indirekte Vergleiche gibt es adjustierte (für die Effekte in den Kontrollgruppen) oder nicht-adjustierte Varianten: Berry et al.⁴⁴ und Verma et al.²⁵⁵ vergleichen die absoluten Therapieeffekte der Verumarme der einzelnen Studien narrativ miteinander, ohne die Ereignisse in den entsprechenden Kontrollgruppen zu berücksichtigen.

Yabroff et al.²⁶⁶ nehmen einen adjustierten deskriptiven indirekten Vergleich vor, indem sie die Ereignisse in den Kontroll- von denen in den Verumgruppen subtrahieren. Diese Differenzen werden anschließend beschreibend miteinander verglichen.

5.4 Relative Häufigkeit der indirekt vergleichenden Methoden

In dem ungefähr neunjährigen Zeitraum von Januar 1999 bis Februar 2008, den die für diesen HTA durchgeführte systematische Literaturrecherche abdeckt, werden in 106 systematischen Übersichtsarbeiten 108 indirekte Vergleiche vorgenommen. Die Differenz erklärt sich durch den Review von Vandermeer et al.²⁵⁴, in dem drei verschiedene Methoden des indirekten Vergleichs für die Fragestellung verwendet werden (adjustierter indirekter Vergleich nach Bucher et al.⁵⁵, adjustierter indirekter Vergleich nach Song et al.²¹⁸ und MTC).

Die mit Abstand am häufigsten verwendete Methodenform ist der adjustierte indirekte Vergleich mit 60 von 108 zwischen 1999 und 2008 publizierten Beispielen (siehe Abbildung 14).

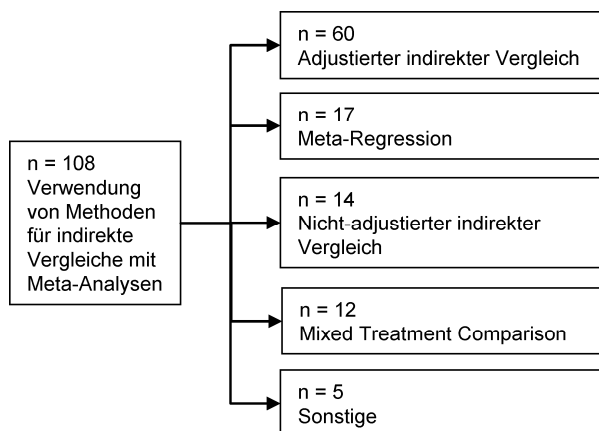


Abbildung 14: Anzahl der Einsätze der verschiedenen Methoden zur Durchführung indirekter Vergleiche mit Metaanalyse(n) in systematischen Reviews, publiziert im Zeitraum von Januar 1999 bis Februar 2008

Er wird damit deutlich häufiger als der vielfach kritisierte^{104, 109, 217} nicht-adjustierte indirekte Vergleich verwendet (14 Beispiele). Etwas häufiger als ein nicht-adjustierter indirekter Vergleich werden Meta-regressionen für indirekte Vergleiche eingesetzt und etwas seltener MTC. Die sonstigen Methoden, die sich nicht in die vier wichtigsten Methodengruppen einordnen lassen, spielen mit fünf von 108 Methodeneinsätzen (5 %) eine untergeordnete Rolle.

Entwicklung auf der Zeitachse

Abbildung 15 stellt die Häufigkeiten der Publikationen mit indirekten Vergleichen für den Zeitraum Januar 2000 bis Dezember 2007 dar. 1999 wird nicht mit abgebildet, da in diesem Jahr nur eine Publikation mit indirektem Vergleich erschienen ist, und 2008 nicht, weil das Jahr bei Abschluss der Literaturrecherche noch nicht vollendet war. Insgesamt werden 94 von den 108 systematischen Reviews aus Abbildung 14 übernommen.

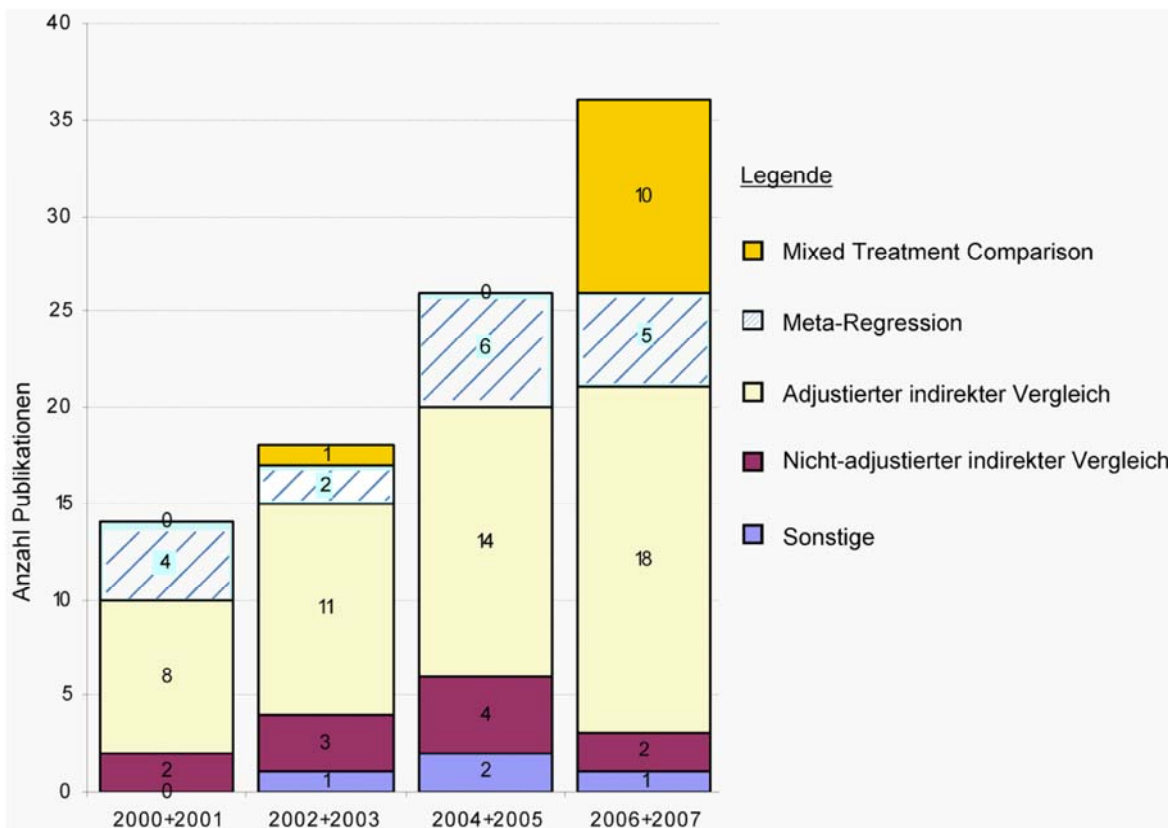


Abbildung 15: Häufigkeit von Publikationen mit indirekten Vergleichen von 2000 bis 2007

Insgesamt ist ein Anstieg bei der Gesamtanzahl an Publikationen mit indirekten Vergleichen über die Zeit zu erkennen.

Dieser ist vor allem auf die zunehmende Verwendung von adjustierten indirekten Vergleichen und MTC zurückzuführen, während die Anzahl der Einsätze der anderen Methoden größtenteils konstant bleiben.

Bei den MTC ist zusätzlich zu berücksichtigen, dass die wichtigsten Methodenpapiere erst 2002¹⁶⁰ und 2004¹⁵⁷ erschienen sind. Betrachtet man nur den Publikationszeitraum von Januar 2006 bis Dezember 2007 liegt der MTC in der Häufigkeit der Methodenanwendungen auf dem zweiten Platz:

1. Adjustierter indirekter Vergleich (n = 18)
2. MTC (n = 10)
3. Metaregression (n = 5)
4. Nicht-adjustierter indirekter Vergleich (n = 2)

5. Sonstige (n = 1)

Erstellung des Wirksamkeitsvergleichs

Als sekundäres Methodencharakteristikum kann die Vorgehensweise bei der Erstellung des Wirksamkeitsvergleichs beschrieben werden. Dabei spielen narrative indirekte Vergleiche, statistische Tests oder Konfidenzintervallvergleiche im Gegensatz zur Berechnung von Gesamteffektschätzern eine untergeordnete Rolle (siehe Tabelle 2).

Tabelle 2: Anzahl der Einsätze der verschiedenen Methoden zur Durchführung indirekter Vergleiche mit Metaanalyse(n) in systematischen Reviews, publiziert im Zeitraum von Januar 1999 bis Februar 2008

Methoden des indirekten Vergleichs	Anzahl der Methodeinsätze	Indirekter Vergleich über: (Sekundäres Methodencharakteristikum)	Anzahl der Einsätze dieser sekundären Charakteristika
Nicht-adjustierter indirekter Vergleich	14	Gesamteffektschätzer	5
		Statistischer Test	5
		Konfidenzintervalle	2
		Narrativ	2
Adjustierter indirekter Vergleich	60	Gesamteffektschätzer (hauptsächlich nach Bucher et al. ⁵⁵)	36
		Statistischer Test	8
		Konfidenzintervalle	14
		Narrativ	2
Metaregression	17	Metaregression mit Gesamteffektschätzer	15
		Metaregression mit Bayes'schem Verfahren und Gesamteffektschätzer	2
Mixed treatment comparison	12	Methode nach Lu und Ades ¹⁵⁷ mit Gesamteffektschätzer	8
		Methode nach Lumley ¹⁶⁰ mit Gesamteffektschätzer	2
		Nach anderer Methodik für eine Mixed treatment comparison mit Gesamteffektschätzer	2
Sonstige Methoden	5	Hind 2003 ¹²⁵ , Moore 2005 ¹⁷⁶ , Brophy 2005 ⁵¹ , Smith 2007 ²⁴⁰ und Thijs 2008 ²³⁴ (alle berechneten Gesamteffektschätzer)	5
Summe	108		108

5.5 Validitätsprüfung indirekt vergleichender Methoden

5.5.1 Datensätze für die Validitätsprüfung

Insgesamt können 250 Datensätze mit Ergebnissen eines direkten und indirekten Vergleichs aus 59 systematischen Reviews extrahiert werden. Die Metaanalyseergebnisse von Eckert et al.⁹¹ und der Antithrombotic Trialists' Collaboration²⁸ müssen allerdings von der Validitätsprüfung ausgeschlossen werden. Bei Eckert et al. lassen sich aus den Effektschätzern keine Standardfehler rekonstruieren und bei der Arbeit der Antithrombotic Trialists' Collaboration bleibt unklar, ob Daten aus Head-to-head-Studien in den indirekten Vergleich eingebracht wurden. Der Datenpool für die Gegenüberstellungen von Ergebnissen von direkten und indirekten Vergleichen zur Validitätsprüfung umfasst somit 248 Datensätze.

Unter den 59 systematischen Reviews befinden sich keine Typ-3-Publikationen, die zur Validitätsprüfung von indirekt vergleichenden Methoden ohne Metaanalysen eingesetzt werden können. Die Angaben in diesem Kapitel beziehen sich daher nur auf metaanalytische Methoden des indirekten Vergleichs.

5.5.2 Methodenspezifische Validitätsprüfung

Prüfkriterien

Für jede indirekt vergleichende Methode werden, wie in Kapitel 4.5.2 „Indirekte Vergleiche mit metaanalytischen Methoden“ beschrieben, drei Prüfkriterien berechnet:

1. Test auf systematische Über- oder Unterschätzung der im direkten Vergleich (Goldstandard) ermittelten Ergebnisse (Mittelwert des z-Wertes $\bar{z} = 0$ und Test auf Normalverteilung (Kolmogorov-Smirnov-Test))
2. Mittleres Ausmaß der Diskrepanz (Mittelwert der Beträge der z-Werte $|\bar{z}|$)
3. Anteil der statistisch signifikanten diskrepanten Datensätze (Anteil der Datensätze mit $|\bar{z}| > 1,96$ pro Methode).

Darstellung der Diskrepanzen in Forest-Plots

Sortiert nach der verwendeten Methodik sowie nach Effektmaß (dichotom, kontinuierlich) werden die Diskrepanzen zwischen den Ergebnissen des direkten und indirekten Vergleichs in Forest-Plots dargestellt. Signifikante Diskrepanz zeigt sowohl ein Betrag des z-Wertes an, der größer als 1,96 ist, als auch ein Konfidenzintervall der Diskrepanz, das den Wert $z = 0$ nicht mit einschließt. Das Konfidenzintervall gibt den Bereich an, in dem der wahre Wert einer Messung mit einer bestimmten Sicherheit erwartet werden kann (hier 95 %-Konfidenzintervall)¹⁰.

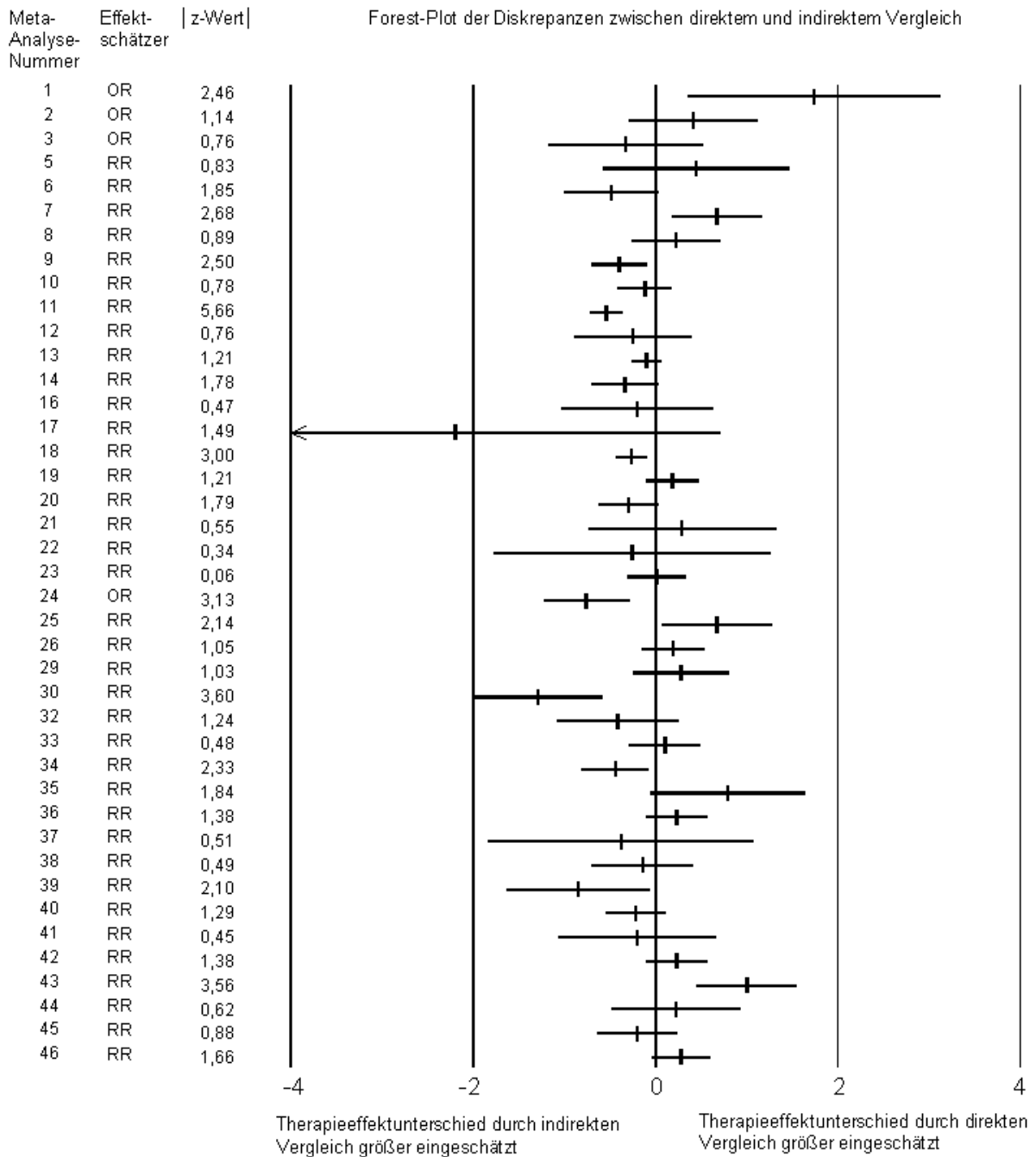
Aus Gründen der Übersichtlichkeit werden in den Abbildungen keine weiteren Informationen zu den Metaanalysen dargestellt. Diese können unter den angegebenen Nummern den Tabellen im Anhang entnommen werden. Tabelle 28 weist Informationen zum Autor und Erscheinungsjahr des Reviews, dem untersuchten Patientenkollektiv, den verglichenen Therapieoptionen, dem gemeinsamen Komparator für den indirekten Vergleich sowie den betrachteten Therapieeffekten auf. Tabelle 29 enthält für die indirekten Vergleiche die Gesamtanzahl der eingeschlossenen Studien und Probanden, den geschätzten Therapieeffektunterschied und die Information, ob Random oder Fixed effects-Modelle für die Metaanalyse gewählt werden. Tabelle 30 zeigt die entsprechenden Charakteristika für die direkten Vergleiche.

5.5.2.1 Nicht-adjustierter indirekter Vergleich

47 Datensätze (siehe Tabelle 29) aus 28 systematischen Reviews (siehe Tabelle 28) stehen für die Validitätsprüfung der nicht-adjustierten indirekten Vergleiche zur Verfügung. Im Mittel weisen die Datensätze mit nicht-adjustierten indirekten Vergleichen einen z-Wert von $\bar{z} = -0,08$ (95 %-KI: -0,73 bis 0,58) auf, $p = 0,200$ (Kolmogorov-Smirnov-Test). Damit kann kein systematisches Über- oder Unterschätzen des Ergebnisunterschieds durch den nicht-adjustierten indirekten Vergleich festgestellt werden.

Allerdings liegt in der Verteilung eine große Streuung vor ($\sigma = 2,22$), die eine Betrachtung der Diskrepanzen auf der Ebene der einzelnen Datensätze erforderlich macht (siehe Abbildung 16). Die Ergebnisse werden getrennt nach dichotomen und kontinuierlichen Effektmaßen dargestellt, da unterschiedliche Skalierungen für die Diskrepanz zwischen direktem und indirektem Vergleich gewählt werden müssen.

Dichotome Effektschätzer



Kontinuierliche Effektschätzer

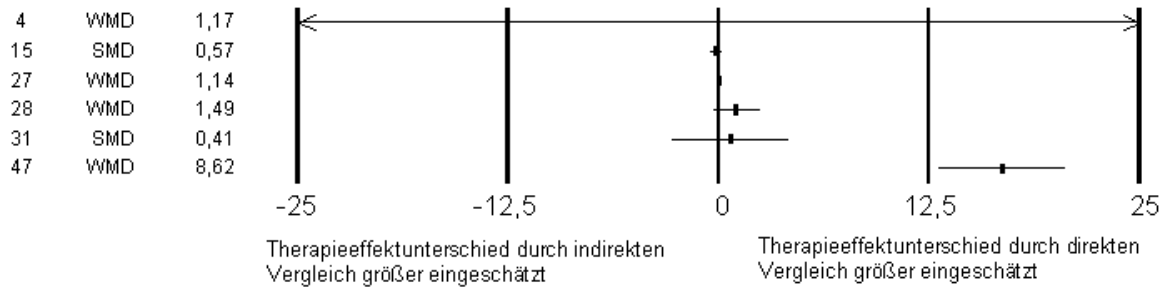


Abbildung 16: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von nicht-adjustierten indirekten Verfahren

OR = Odds ratio. RR = Relatives Risiko. SMD = Standardisierte mittlere Differenz. WMD = Gewichtete mittlere Differenz.

Der mittlere Betrag des z-Wertes der 47 Datensätze ist $|\bar{z}| = 1,63$ (95 %-KI: 1,20 bis 2,07), der Signifikanz-Cut-off-Wert von $|\bar{z}| = 1,96$ ist im Konfidenzintervall eingeschlossen.

Folgerichtig werden in der detaillierten Betrachtung viele statistisch signifikant diskrepante Datensätze ermittelt. In zwölf der 47 Datensätze (25,5 %; 95 %-KI: 13,1 % bis 38 %) werden statistisch signifikante Diskrepanzen zwischen den Ergebnissen der direkten und indirekten Vergleiche gefunden (Metaanalysenummern (Nr.) 1, 7, 9, 11, 18, 24, 25, 30, 34, 39, 43 und 47).

5.5.2.2 Adjustierter indirekter Vergleich

116 Datensätze (siehe Tabelle 29) aus 45 systematischen Reviews (siehe Tabelle 28) können für die Validitätsprüfung des adjustierten indirekten Vergleichs herangezogen werden.

Im Mittel weisen die Datensätze mit adjustierten indirekten Vergleichen einen z-Wert von $\bar{z} = -0,18$ (95 %-KI: -0,05 bis 0,40) auf, $p = 0,200$ (Kolmogorov-Smirnov-Test). Damit kann kein systematisches Über- oder Unterschätzen des Therapieeffektunterschieds durch den adjustierten indirekten Vergleich festgestellt werden. Die Standardabweichung der Verteilung der z-Werte ist verglichen mit der beim nicht-adjustierten indirekten Vergleich gering ($\sigma = 1,21$).

Eine Betrachtung der Diskrepanzen zwischen direktem und adjustiertem indirekten Vergleich in den einzelnen Datensätzen nehmen Abbildung 17, Abbildung 18 und Abbildung 19 vor. In Abbildung 17 und Abbildung 18 werden die Datensätze mit dichotomen Effektmaßen dargestellt, Abbildung 19 ergänzt die Datensätze mit kontinuierlichen Effektmaßen.

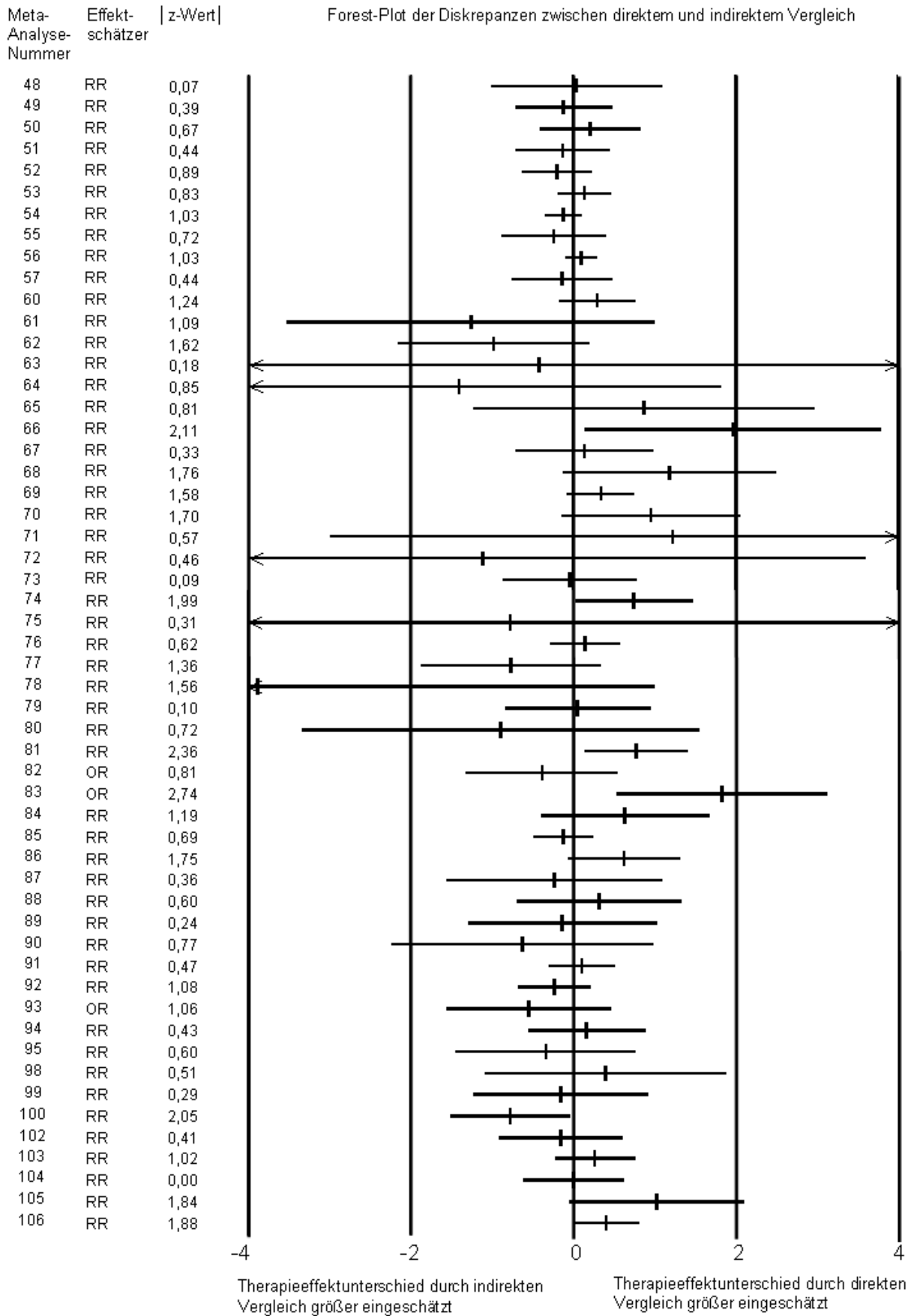


Abbildung 17: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und dichotomen Daten_Teil 1

OR = Odds ratio. RR = Relatives Risiko.

Dichotome Effektschätzer

Meta-Analyse-Nummer	Effektschätzer	z-Wert
107	RR	0,18
108	RR	0,13
109	RR	1,65
110	RR	0,00
111	RR	0,05
112	RR	1,88
113	OR	0,65
115	OR	3,04
116	OR	3,03
117	OR	1,03
118	OR	2,35
119	OR	3,73
120	OR	0,06
121	RR	1,39
122	RR	0,17
123	RR	0,17
124	RR	1,00
141	OR	1,09
142	RR	1,03
144	RR	0,49
155	OR	1,89
156	OR	0,08
157	OR	0,20
161	RR	0,59
164	RR	2,17

Forest-Plot der Diskrepanzen zwischen direktem und indirektem Vergleich

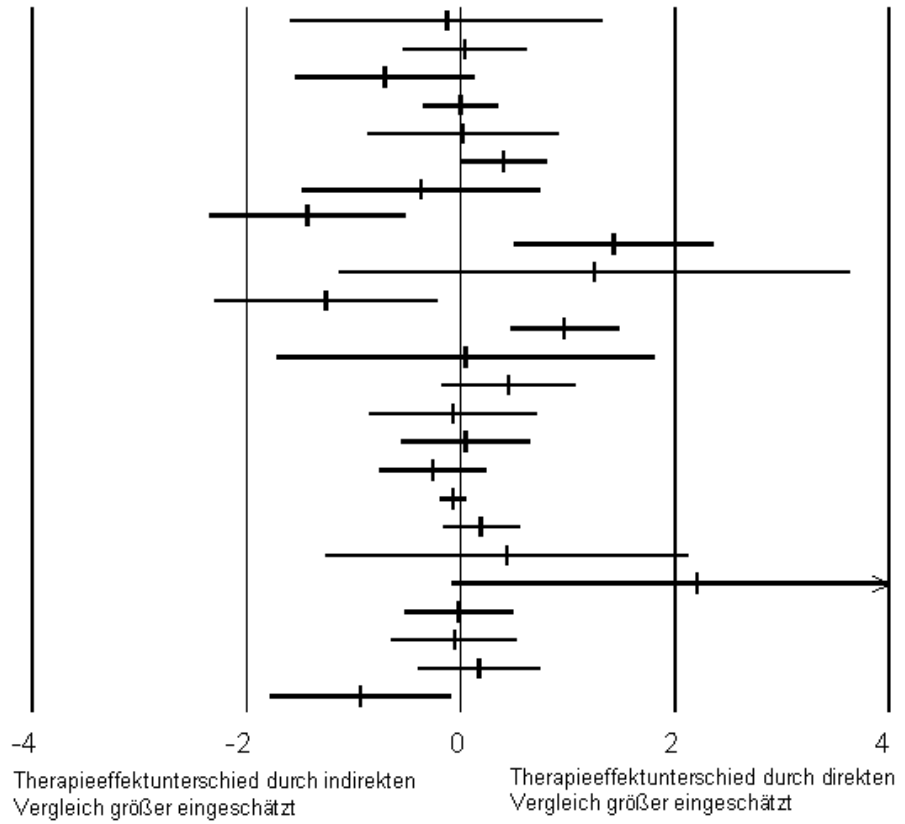


Abbildung 18: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und dichotomen Daten_Teil 2

OR = Odds ratio. RR = Relatives Risiko.

Kontinuierliche Effektschätzer

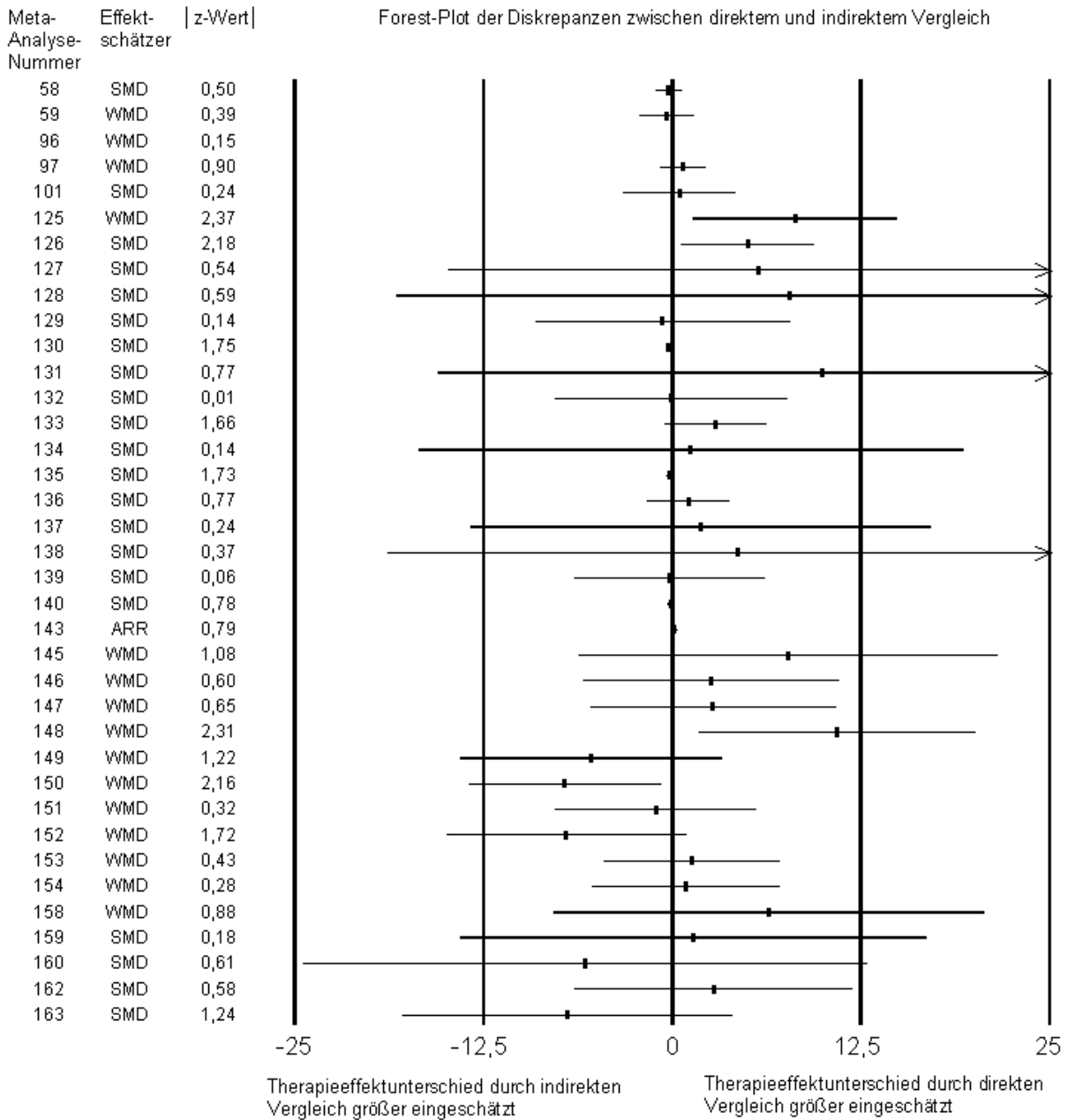


Abbildung 19: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von adjustierten indirekten Verfahren und kontinuierlichen Daten

SMD = Standardisierte mittlere Differenz. WMD = Gewichtete mittlere Differenz.

Alle 116 Datensätzen aus den drei Abbildungen zusammengefasst, weisen beim Betrag des z-Wertes einen Mittelwert von $|\bar{z}| = 0,95$ (95 %-KI: 0,80 bis 1,09) auf. Er liegt damit deutlich unter der Signifikanzschwelle von $|\bar{z}| = 1,96$. Dennoch treten einige extreme Werte auf, deren Diskrepanz zwischen direktem und indirektem Vergleich statistische Signifikanz erreicht. Dies betrifft die folgenden 14 Datensätze (Anteil: 12,1 %; 95 %-KI: 6,1 % bis 18 %): Metaanalysenummer 66, 74, 81, 83, 100, 115, 116, 118, 119, 125, 126, 148, 150 und 164.

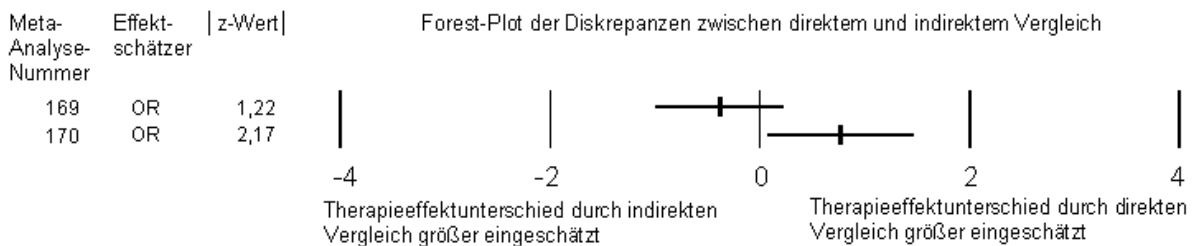
5.5.2.3 Metaregression

Für die Validitätsprüfung der Metaregression stehen nur sechs Datensätze (siehe Tabelle 29) aus zwei systematischen Reviews (siehe Tabelle 28) zur Verfügung.

Im Mittel weisen die Datensätze mit Metaregressionen einen z-Wert von $\bar{z} = 0,46$ (95 %-KI: -0,82 bis 1,75) auf, $p = 0,200$ (Kolmogorov-Smirnov-Test). Damit kann kein systematisches Über- oder Unterschätzen des Therapieeffektunterschieds durch den indirekten Vergleich mittels Metaregression festgestellt werden. Die Präzision dieser Schätzung ist trotz einer geringen Standardabweichung ($\sigma = 1,22$) aufgrund der kleinen Stichprobe aber sehr gering.

Im Forest-Plot (Abbildung 20) wird ein statistisch signifikanter diskrepanter Datensatz erkennbar.

Dichotome Effektschätzer



Kontinuierliche Effektschätzer

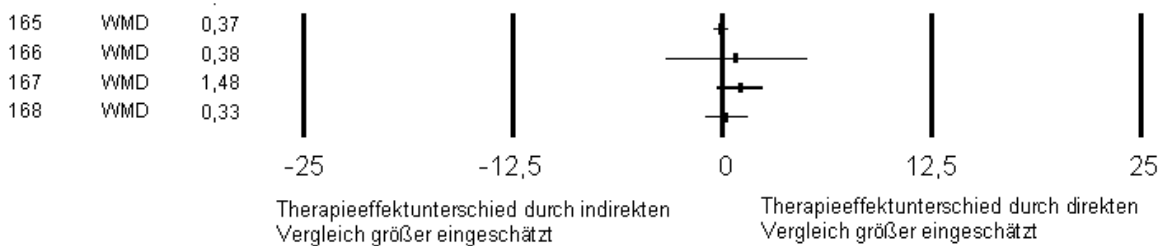


Abbildung 20: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von Metaregressionen für den indirekten Vergleich

OR = Odds ratio. WMD = Gewichtete mittlere Differenz.

Der Mittelwert von $|z|$ lässt sich bei sechs Datensätzen nur unpräzise abschätzen. Er beträgt $|\bar{z}| = 0,99$ (95 %-KI: 0,20 bis 1,79), der Cut-off-Wert für statistisch signifikante Diskrepanz ($z = 1,96$) ist nicht im Konfidenzintervall enthalten. Dennoch existiert mit dem Metaanalysenpaar Nummer 170 ein Datensatz, der eine statistisch signifikante Diskrepanz zwischen direktem und indirektem Vergleich aufweist (Anteil: 16,7 %; 95 %-KI: -13,2 % bis 46,5 %).

5.5.2.4 Mixed treatment comparison (MTC)

Bei den MTC können 57 Datensätze (siehe Tabelle 29) aus sieben systematischen Reviews (siehe Tabelle 28) für die Validitätsprüfung herangezogen werden.

Im Mittel weisen die Datensätze mit MTC einen z-Wert von $\bar{z} = 0,20$ (95 %-KI: (-0,004 bis 0,401) auf, $p = 0,059$ (Kolmogorov-Smirnov-Test). Bei hoher Präzision und geringer Streuung ($\sigma = 0,76$) der Schätzung ist er knapp nicht vom Nullwert statistisch signifikant unterschiedlich.

Die sehr geringe Diskrepanz zwischen den Ergebnissen eines MTC und den entsprechenden direkten Vergleichen lassen sich visuell am Forest-Plot der 57 Diskrepanzen in Abbildung 21 und Abbildung 22 erkennen. Abbildung 21 stellt die Diskrepanzen zwischen direktem und indirektem Vergleich dar, die dichotome Effektmaße verwenden. Abbildung 22 ergänzt sechs weitere Datensätze mit Diskrepanzen, die kontinuierliche Effektmaße wählen.

Dichotome Effektschätzer

Meta- | Effekt- | z-Wert |
Analyse- | schätzer |
Nummer

Forest-Plot der Diskrepanzen zwischen direktem und indirektem Vergleich

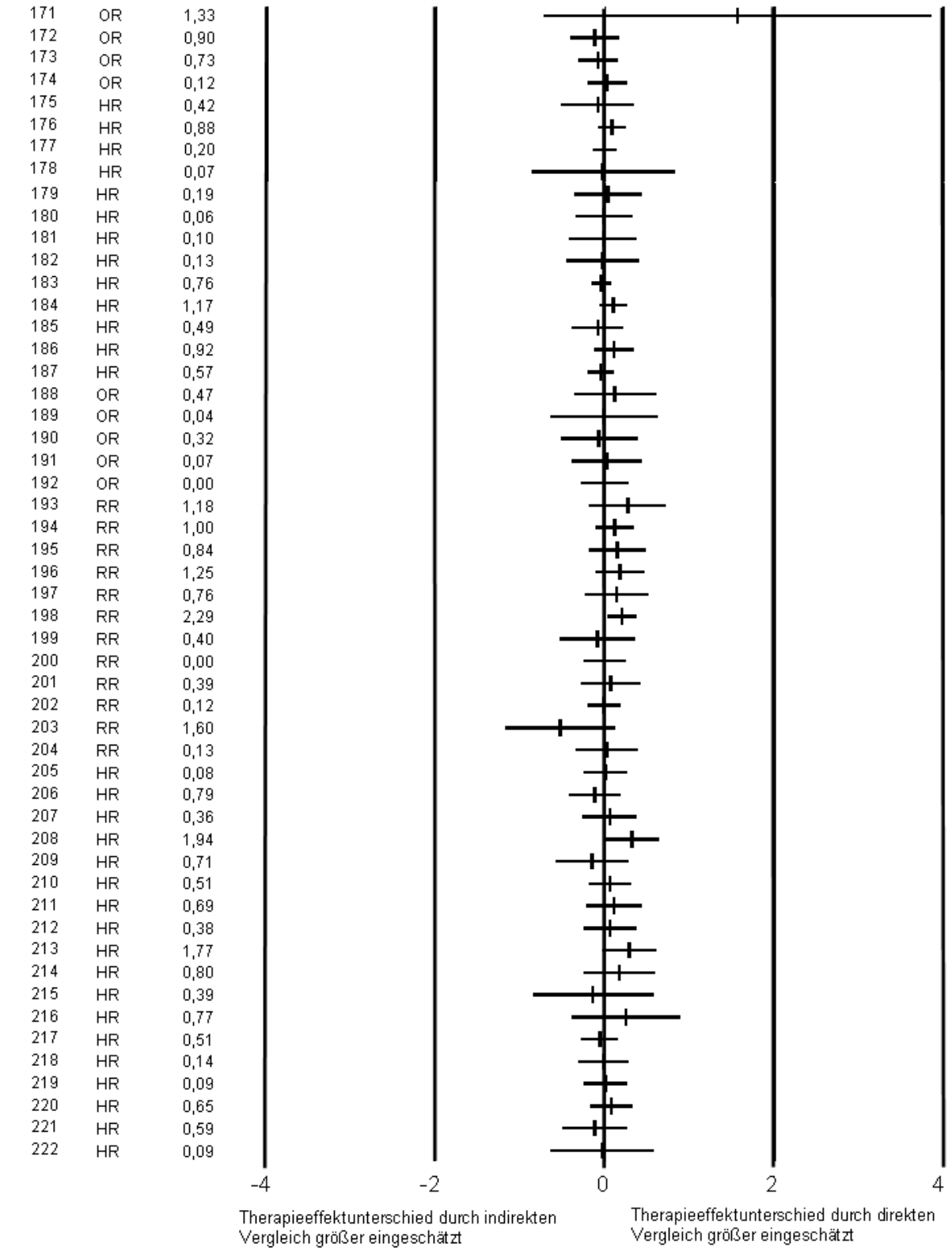


Abbildung 21: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von MTC und dichotomen Daten

HR = Hazard ratio. OR = Odds ratio. RR = Relatives Risiko.

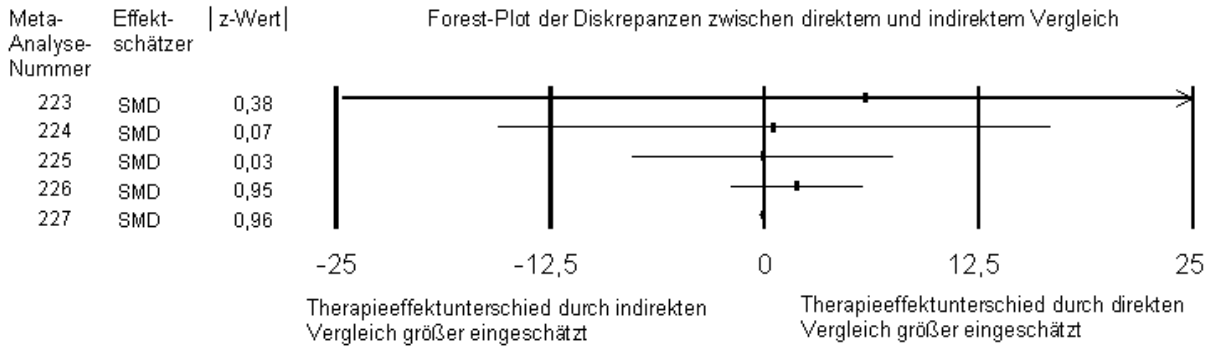
Kontinuierliche Effektschätzer

Abbildung 22: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von MTC und kontinuierlichen Daten

MTC = Mixed treatment comparison. SMD = Standardisierte mittlere Differenz.

Zunächst fällt auf, dass die Diskrepanzen eine sehr hohe Präzision aufweisen. Dies ist darauf zurückzuführen, dass in den MTC eine sehr große Studienanzahl (in der Regel auch die Head-to-head-Studien) eingeschlossen werden kann. Der Standardfehler der Diskrepanz, der sich aus den Standardfehlern des direkten und indirekten Vergleichs zusammensetzt, wird deshalb bei einem kleinen Standardfehler des indirekten Vergleichs hauptsächlich durch den Standardfehler des direkten bestimmt. Dies ist die Begründung für das in Abbildung 21 aus dem Rahmen fallende weite Konfidenzintervall von Metaanalysenummer 171. Für den indirekten Vergleich werden 22 Studien zusammengefasst, während für den direkten Vergleich nur eine Studie ausgewertet werden kann. Der Standardfehler der Schätzung des Therapieeffektunterschiedes im direkten Vergleich ($SE_{dir} = 1,17$) liegt deshalb sehr viel höher als der des indirekten Vergleichs ($SE_{indir} = 0,11$).

Der Mittelwert des Betrags der z-Werte weist mit $|\bar{z}| = 0,59$ (95 %-KI: 0,45 bis 0,73) einen vergleichsweise niedrigen Wert auf, der den Cut-off-Wert für statistisch signifikante Diskrepanz von $|\bar{z}| = 1,96$ deutlich unterschreitet.

Statistische signifikante Diskrepanzen zwischen direktem Vergleich und MTC werden dementsprechend selten in dieser Stichprobe von 57 Gegenüberstellungen gefunden. Nur ein Datensatz (Anteil: 1,8 %; 95 %-KI: -1,7 % bis 5,2 %), im Forest-Plot als Metaanalysenummer 198 erkenntlich, weist signifikante Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs auf. Ebenfalls hohe Diskrepanzen, die aber nicht statistisch signifikant sind, weisen die Metaanalysenummern 171, 203, 208 und 213 auf.

5.5.2.5 Sonstige Methoden für indirekte Vergleiche

In Abbildung 23 sind die Diskrepanzen für die „sonstigen Methoden“ dargestellt. Hierbei werden die Methoden nach Brophy et al. (Metaanalysenummer 228), Tudur Smith et al. (Metaanalysenummer 229 bis 245) und Thijs et al. (Metaanalysenummer 246 bis 249) einer Validitätsprüfung unterzogen. Für die Methoden nach Moore et al. und Hind et al. liegen keine systematischen Reviews vor, die auch einen direkten Vergleich durchgeführt haben. Eine Validitätsprüfung ist daher nicht möglich.

Aufgrund der Unterschiedlichkeit der unter dem Punkt „Sonstige Methoden“ zusammengefassten Ansätze wird bewusst auf die Durchführung eines Tests auf Normalverteilung und einer Mittelwertberechnung beim z-Wert und Betrag des z-Wertes verzichtet.

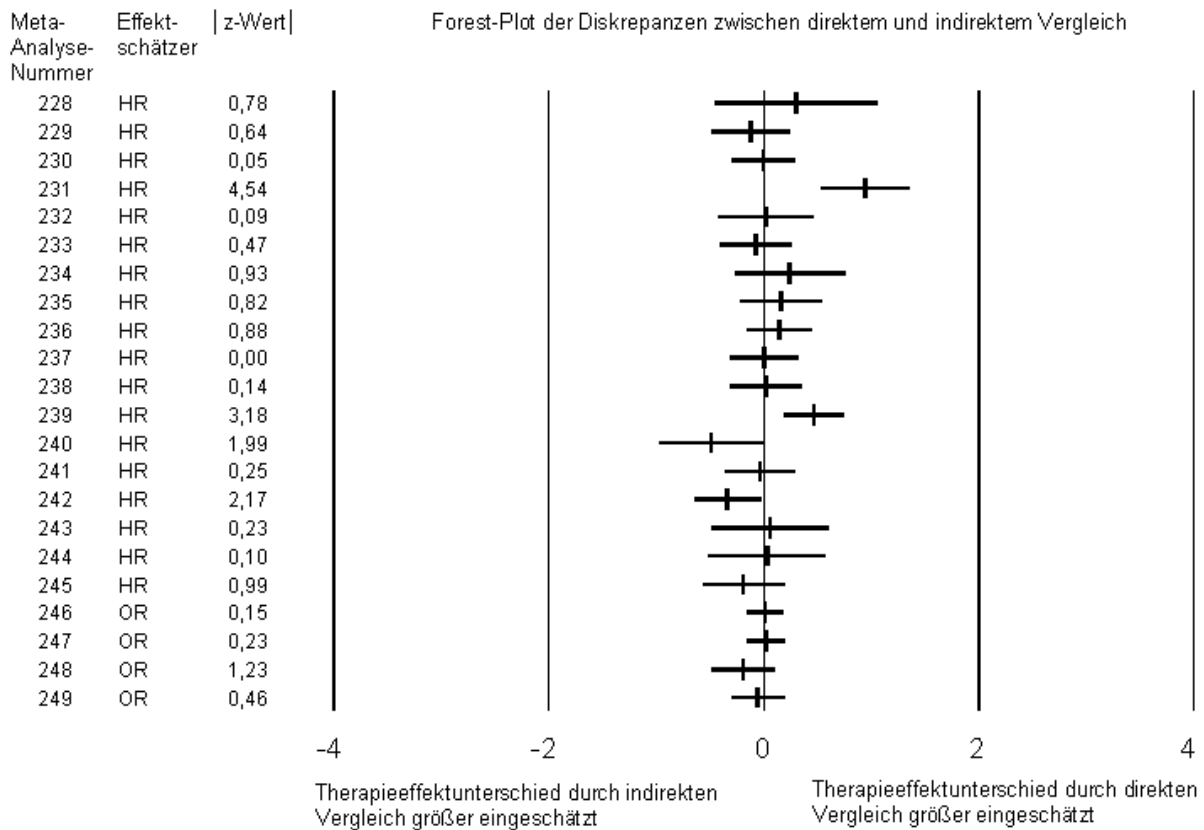


Abbildung 23: Diskrepanz zwischen direktem und indirektem Vergleich bei der Verwendung von sonstigen Methoden (in gefundenen Beispielen wurden nur dichotome Daten benutzt)

Die Diskrepanzen bei Metaanalysenummer 231, 239, 240 und 242 erreichen statistische Signifikanz. Diese Datensätze entstammen alle der Methodik nach Tudur Smith et al. Die Methoden nach Brophy et al. und Thijs et al. weisen keine signifikanten Diskrepanzen zwischen direktem und indirektem Vergleich auf. Allerdings steht für beide Verfahren nur je ein systematischer Review für die Validitätsprüfung zur Verfügung.

Eine tabellarische Zusammenfassung der Ergebnisse der Hauptanalysen befindet sich im Anhang (Abschnitt 9.9.5 „Übersichtstabellen zu den Ergebnissen der Validitätsprüfung (Hauptanalyse, Typ-5-Publikationen)“; Tabelle 32 bis 34).

5.5.3 Subgruppenanalyse

Die Stichproben indirekter Vergleich und direkter Vergleich sind nicht als unabhängig voneinander zu betrachten, wenn sie teilweise auf dieselben Daten zurückgreifen (z. B. durch Einschluss von Head-to-head-Studien mit gemeinsamem Komparator). Inwieweit dies zu anderen Ergebnissen in der Validitätsprüfung führen kann, wird in dieser Subgruppenanalyse evaluiert.

5.5.3.1 Datensätze für die Subgruppenanalyse

Von den 248 in der Hauptanalyse zur Validitätsprüfung herangezogenen Datensätzen werden die 148 Datensätze in die Subgruppenanalyse eingeschlossen, die keine Head-to-head-Studien in den indirekten Vergleich mit aufnehmen.

Die Datensätze verteilen sich wie folgt auf die verschiedenen methodischen Ansätze (Tabelle 3):

Tabelle 3: Übersicht über die in die Haupt- und Subgruppenanalyse eingeschlossenen Datensätze

Methoden des indirekten Vergleichs	Hauptanalyse (n)	Subgruppenanalyse (n)	Differenz der Stichprobengrößen n (%)
Nicht-adjustierter indirekter Vergleich	47	47	0 (0 %)
Adjustierter indirekter Vergleich	116	84	32 (28 %)
Metaregression	6	4	2 (33 %)
Mixed treatment comparison	57	12	45 (79 %)
Sonstige	22	1	21 (95 %)
Gesamt	248	148	100 (40 %)

Der Ausschluss von Datensätzen, in denen Head-to-head-Studien in den indirekten Vergleich mit aufgenommen wurden, betrifft somit vor allem den MTC und die sonstigen Methoden, während die nicht-adjustierten indirekten Vergleiche gar nicht betroffen sind.

Die 148 Datensätze werden den gleichen Validitätsprüfungen wie in der Hauptanalyse, nach der für den indirekten Vergleich gewählten Methodik stratifiziert, unterzogen.

Somit sind die Ergebnisse von Haupt- und Subgruppenanalyse bei den nicht-adjustierten indirekten Vergleichen identisch. Sie werden an dieser Stelle wiederholt aufgeführt um sie den Ergebnissen der anderen indirekt vergleichenden Methoden auch in der Subgruppenanalyse gegenüberzustellen.

5.5.3.2 Test auf systematische Über- oder Unterschätzung

Beim Test auf Über- oder Unterschätzung der Ergebnisse des direkten Vergleichs durch den indirekten Vergleich werden keine signifikanten Ergebnisse gefunden (Tabelle 4).

Tabelle 4: Test auf signifikante Über- oder Unterschätzung des Ergebnisses des direkten Vergleichs durch den indirekten Vergleich (Subgruppenanalyse)

Methoden	n für die Validitätsprüfung	Durchschnittliche Diskrepanz zwischen direktem und indirektem Vergleich \bar{z}	Streuung σ	Test auf Normalverteilung*
Nicht-adjustierter indirekter Vergleich	47	$\bar{z} = -0,08$ (95 %-KI: -0,73 bis 0,58)	$\sigma = 2,22$	$p = 0,200$
Adjustierter indirekter Vergleich	84	$\bar{z} = 0,21$ (95 %-KI: -0,05 bis 0,48)	$\sigma = 1,21$	$p = 0,200$
Metaregression	4	$\bar{z} = 0,45$ (95 %-KI: -0,76 bis 1,67)	$\sigma = 0,76$	$p = 0,200$
Mixed treatment comparison	12	$\bar{z} = 0,48$ (95 %-KI: -0,15 bis 1,10)	$\sigma = 0,98$	$p = 0,200$

* Kolmogorov-Smirnov-Test auf statistisch signifikante Abweichung von der Normalverteilung; berechnet mit SPSS 16.0; Signifikanzniveau $p = 0,05$.
KI = Konfidenzintervall.

Die mittleren z-Werte sind nicht vom Wert $\bar{z} = 0$ signifikant unterschiedlich und normalverteilt (Kolmogorov-Smirnov-Test). Somit kann auch in der Subgruppenanalyse keine systematische Über- oder Unterschätzung des Therapieeffektunterschieds durch eine indirekt vergleichende Methode festgestellt werden.

5.5.3.3 Ausmaß der durchschnittlichen Diskrepanz

Beim Ausmaß der durchschnittlichen Diskrepanz wird die Höhe des Betrags des z-Wertes ($|\bar{z}|$) für die einzelnen Methoden mit entsprechendem Konfidenzintervall berechnet (Tabelle 5).

Tabelle 5: Ausmaß der durchschnittlichen Diskrepanz zwischen direktem und indirektem Vergleich (Subgruppenanalyse)

Methoden	n für die Validitätsprüfung	Durchschnittliche Diskrepanz zwischen direktem und indirektem Vergleich $ \bar{z} $	Streuung σ
Nicht-adjustierter indirekter Vergleich	47	$ \bar{z} = 1,63$ (95 %-KI: 1,20 bis 2,07)	1,49
Adjustierter indirekter Vergleich	84	$ \bar{z} = 0,96$ (95 %-KI: 0,80 bis 1,12)	0,75
Metaregression	4	$ \bar{z} = 0,64$ (95 %-KI: 0,00 bis 1,53)	0,56
Mixed treatment comparison	12	$ \bar{z} = 0,83$ (95 %-KI: 0,40 bis 1,26)	0,68

KI = Konfidenzintervall.

Um die Streuung und die Stichprobengröße mit in die Validitätsprüfung einzubeziehen, wird die obere Grenze des Konfidenzintervalls des mittleren z-Wertes betrachtet. Die Breite des Konfidenzintervalls hängt von der Stichprobengröße und der Streuung der Werte um den Mittelwert ab. Bis zu der oberen Grenze des Konfidenzintervalls kann der wahre Wert des mittleren z-Wertes mit 95-prozentiger Sicherheit auch erwartet werden.

Den größten Wert für die obere Grenze des Konfidenzintervalls von $|\bar{z}|$ besitzt der nicht-adjustierte indirekte Vergleich. Mit $|\bar{z}| = 2,07$ liegt der Wert sogar oberhalb der Marke von $|\bar{z}| = 1,96$ für statistisch signifikante Diskrepanz zwischen direktem und indirektem Vergleich. Die oberen Konfidenzintervallgrenzen der Mittelwerte der Beträge der z-Werte der anderen Methoden liegen unterhalb des Werts von $|\bar{z}| = 1,96$. Es folgen die Metaregression mit $|\bar{z}| = 1,53$, der MTC mit $|\bar{z}| = 1,26$ und der adjustierte indirekte Vergleich mit $|\bar{z}| = 1,12$. Somit tritt in dieser Stichprobe auf dem 95 %-Konfidenzniveau beim adjustierten indirekten Vergleich die geringste maximale durchschnittliche Diskrepanz zwischen direktem und indirektem Vergleich auf.

5.5.3.4 Anteil der statistisch signifikant diskrepanten Datensätze bei den verschiedenen Methoden

Stratifiziert nach der verwendeten Methode für den indirekten Vergleich zeigen sich auch in der Subgruppenanalyse große Unterschiede in der Häufigkeit der statistisch signifikanten Diskrepanzen zwischen direktem und indirektem Vergleich (siehe Tabelle 6).

Tabelle 6: Anteil der statistisch signifikant diskrepanten Datensätze (Subgruppenanalyse)

Methoden	n für die Validitätsprüfung	Statistisch signifikant diskrepante Datensätze	Prozentualer Anteil statistisch signifikanter Datensätze in der Stichprobe der Methode
Nicht-adjustierter indirekter Vergleich	47	12	25,5 % [95 %-KI: 13,1 % bis 38,0 %]
Adjustierter indirekter Vergleich	84	10	11,9 % [95 %-KI: 5,0 % bis 18,8 %]
Metaregression	4	0	0,0 % [95 %-KI: 0,0 % bis 0,0 %]
Mixed treatment comparison	12	1	8,3 % [95 %-KI: -7,3 % bis 24,0 %]
Sonstige	1	0	0,0 % [95 %-KI: 0,0 % bis 0,0 %]
Gesamt	148	23	15,5 % [95 %-KI: 9,7 % bis 21,4 %]

KI = Konfidenzintervall.

Nicht-adjustierende indirekte Vergleiche weisen mit 25,5 % die meisten diskrepanten Ergebnisse auf. Ergebnisse der adjustierenden indirekten Vergleiche und MTC sind mit 11,9 % und 8,3 % der Datensätze deutlich seltener diskrepant zu denen der direkten Vergleiche. Die relative Häufigkeitsausgabe zur Validität der MTC basiert allerdings nur auf einem von zwölf Datensätzen. Wenige Fälle gab es auch für die Metaregressionen (n = 4) und die „sonstigen Methoden“ (n = 1). Es kann auf dem Zufall beruhen, dass für diese Methoden keine diskrepanten Ergebnisse gefunden werden.

5.5.3.5 Ergebnisgegenüberstellung von Hauptanalyse und Subgruppenanalyse und Zusammenfassung der Validitätsprüfung

Insgesamt drei Prüfkriterien werden für die Validität der indirekt vergleichenden Methoden erhoben:

1. Testung auf systematische Über- oder Unterschätzung: Bei der Testung auf systematische Über- oder Unterschätzung stellt sich sowohl in der Hauptanalyse als auch in der Subgruppenanalyse heraus, dass keine der untersuchten indirekt vergleichenden Methoden systematisch die Ergebnisse des direkten Vergleichs über- oder unterschätzt.

2. Mittleres Ausmaß der Diskrepanz: Dennoch tritt Diskrepanz zwischen direktem und indirektem Vergleich auf, die mit Abstand am größten bei den nicht-adjustierten indirekten Vergleichen ist. Die durchschnittlichen Diskrepanzen liegen bei den adjustierten indirekten Vergleichen, den Metaregressionen und den MTC auf einem vergleichbaren Niveau - mit Vorteilen für den MTC, wenn er Head-to-head-Studien einschließt (Hauptanalyse).

3. Anteil der statistisch signifikanten diskrepanten Datensätze: Es muss allerdings berücksichtigt werden, dass die Streuung um die Mittelwerte der durchschnittlichen Diskrepanz bei allen indirekt vergleichenden Methoden hoch ist. Besonders hoch ist sie bei den nicht-adjustierten indirekten Vergleichen. Dies bedingt, dass viele extreme Werte mit statistisch signifikanter Diskrepanz gefunden werden (25,5 %). Beim adjustierten indirekten Vergleich werden sowohl in der Hauptanalyse als auch in der Subgruppenanalyse ca. 12 % signifikant diskrepante Datensätze gefunden. Bei den MTC liegt diese Quote in der Hauptanalyse bei 1,8 %. In der Subgruppenanalyse wird ein signifikant diskrepanter Datensatz von zwölf gefunden.

Für die Metaregressionen sind die Stichproben in Haupt- und Subgruppenanalyse sehr klein ($n = 6$ bzw. $n = 4$). Folgerichtig sind die Konfidenzintervalle um die Anteile statistisch signifikant diskrepanter Ergebnisse zwischen direktem und indirektem Vergleich weit. Dies gilt auch für die sonstigen Methoden, für die jeweils nur ein bis zwei systematische Reviews vorliegen.

5.6 Präzision indirekt vergleichender Methoden

5.6.1 Für die Präzisionsbetrachtung herangezogene Datensätze

Grundlage sind die gleichen 248 Datensätze, die für die Validitätsprüfung herangezogen werden. Von diesen werden all diejenigen ausgeschlossen, die für die direkten und indirekten Vergleiche unterschiedliche (Random oder Fixed effects) Metaanalysemodelle wählen. Da Fixed effects-Modelle zu präziseren Schätzungen führen als Random effects-Modelle, würde der Einsatz unterschiedlicher Modelle die Präzisionsbetrachtung verzerren. Vom Ausschluss betroffen sind insgesamt 50 Datensätze, sodass 198 für die Präzisionsbetrachtung verbleiben. Diese 50 Datensätze entstammen den systematischen Reviews von Sanchez-Ramos et al.²⁰⁶ (Metaanalysennummer 1 bis 3), Wilhelmus et al.²⁶¹ (Metaanalysennummer 114 bis 116 und 118) und Glenny et al.¹⁰⁹ (siehe Tabelle 28 und Tabelle 29).

5.6.2 Theoretischer Ansatz

Im theoretischen Ansatz soll festgestellt werden, wie viele Studien ein indirekter Vergleich einschließen muss, um ein Ergebnis von gleicher Präzision wie ein direkter Vergleich zu erreichen. Glenny et al.¹⁰⁹ führen den mathematischen Nachweis, dass viermal so viele gleichgroße Studien für den indirekten Ansatz gebraucht werden, wie für den direkten Vergleich, um die gleiche Präzision zu erreichen. Diese Relation sei näherungsweise wahr, wenn der Therapieeffektunterschied aus Studien mit unterschiedlichen Stichprobengrößen gepoolt werde. Neben den geforderten gleichgroßen Studien liegen weitere Einschränkungen vor, damit diese 4:1-Relation gilt¹⁰⁹:

1. Die metaanalytische Zusammenfassung der Einzelstudienresultate erfolgt in einem Fixed effects-Modell mit Studiengewichtung nach dem Prinzip der inversen Varianz.
2. Alle Effektschätzer der Einzelstudien haben den gleichen Standardfehler und die Heterogenität zwischen den einzelnen Studien ist gleich groß.
3. Der indirekte Vergleich wird als adjustierter indirekter Vergleich durchgeführt.

Da es unwahrscheinlich ist, dass alle diese Annahmen in der Realität zeitgleich angetroffen werden können, wird im folgenden Abschnitt der theoretische Ansatz verlassen und auf der rein empirischen Ebene versucht, sich dieser Frage zu nähern.

5.6.3 Empirischer Ansatz

Im empirischen Ansatz wird die Präzision der Effektschätzer aus direkten und indirekten Vergleichen einander gegenüber gestellt. Hierzu werden die im ersten Abschnitt dieses Kapitels beschriebenen 198 Datensätze herangezogen. Die Inhomogenität der Stichprobe hinsichtlich der unterschiedlich verwendeten Metaanalysemethoden (Fixed und Random effects-Modelle) und Methoden für den indirekten Vergleich nicht weiter berücksichtigt. Für jeden indirekten Vergleich wird nach der Formel aus Abschnitt 4.6 „Beurteilung der Präzision indirekt vergleichender Methoden“ berechnet, in welchem Verhältnis die Weite des Konfidenzintervalls des indirekten Vergleichs zur Weite des direkten Vergleichs steht. Der Quotient wird als prozentualer Anteil ausgedrückt. Negative Werte zeigen ein engeres, positive Werte ein weiteres Konfidenzintervall des indirekten im Vergleich zum direkten Vergleich an.

Abbildung 24 gibt die Häufigkeitsverteilung der Konfidenzintervallverengungen und -ausweitungen in der Stichprobe wieder:

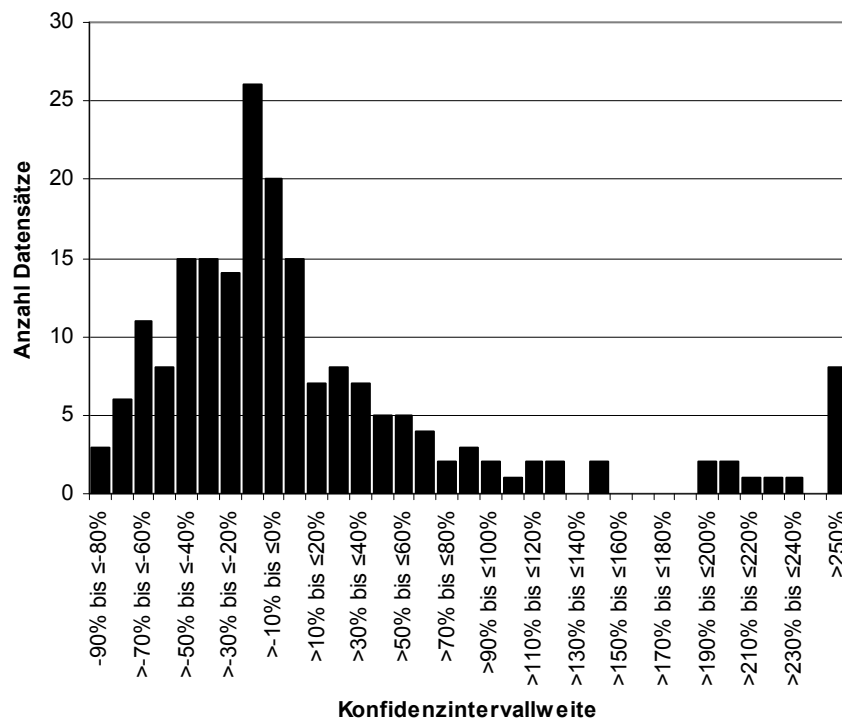


Abbildung 24: Häufigkeitsverteilung der prozentualen Zu- bzw. Abnahme der Konfidenzintervallweite im indirekten Vergleich im Verhältnis zum direkten Vergleich in einer Stichprobe von 198 Datensätzen mit direktem und indirektem Vergleich

Es resultiert eine schiefe Verteilung, da eine Verengung des Konfidenzintervalls nur um maximal 100 % möglich ist, aber eine Ausweitung auch über 100 %. Eine Ausweitung um 100 % entspräche einer Verdoppelung der Weite des Konfidenzintervalls; eine Verdreifachung und mehr sind auch möglich. Aufgrund der Schiefe der Verteilung wird nicht ihr Mittelwert, sondern ihr Median berechnet. Im Median tritt im indirekten Vergleich ein um 9 % (gerundet) engeres Konfidenzintervall auf als im direkten Vergleich (25. Perzentil: -34 %; 50. Perzentil (Median): -9 %; 75. Perzentil: 30 %). In dieser Stichprobe weist in einer Mehrzahl der Datensätze (59,6 %) der indirekte Vergleich eine höhere Präzision auf als der direkte Vergleich.

Die Anzahl der Studien, die in den direkten und indirekten Vergleich eingeschlossen werden, ist in 194 der 198 Datensätze nachvollziehbar dokumentiert (siehe Tabelle 29). Im Median beinhalten diese indirekten Vergleiche exakt sechsmal so viele Studien wie die direkten (25. Perzentil: 4; 75. Perzentil: 13). Durch Ausschluss der vier Datensätze, in denen die eingeschlossenen Studienzahlen nicht transparent sind, verändert sich der Median der Konfidenzintervallweitenänderung im indirekten Vergleich nicht: Er beträgt -9 % (25. Perzentil: -34 %; 75. Perzentil: 27 %).

In der untersuchten Stichprobe ergibt sich somit, bei einem Verhältnis an eingeschlossenen Studien von sechs zu eins zwischen indirektem und direktem Vergleich, ein geringfügig engeres Konfidenzintervall beim indirekten Vergleich. Dieses empirische Ergebnis liegt in der Nähe des durch Glenny et al.¹⁰⁹ theoretisch ermittelten Verhältnisses von vier zu eins bei der Studienanzahl, um durch einen indirekten Vergleich die gleiche Präzision zu erhalten, wie durch einen direkten Vergleich.

5.6.4 Einfluss der Präzision indirekter Vergleiche auf die Validitätsprüfung

Wie bereits im Methodikkapitel erläutert, ist für die Validitätsprüfung wichtig, dass neben den direkt vergleichenden Methoden auch die indirekt vergleichenden Verfahren präzise Effektschätzer liefern. Anderenfalls können die Diskrepanzen zwischen direktem und indirektem Vergleich zwar numerisch groß sein, ihr Unterschied ist aber nicht auf einem akzeptablen statistischen Signifikanzniveau abzusichern.

Diese Problematik trifft nur auf die Gegenüberstellungen zu, bei denen eine der beiden Vergleichsmethoden einen deutlich unpräziseren Effektschätzer liefert als die andere. Exemplarisch werden die Datensätze betrachtet, in denen das Konfidenzintervall des indirekten Vergleichs um $\geq 50\%$ weiter ist als dasjenige des direkten.

Dies ist in 38 der 198 Datensätze (19,2 %) der Fall (siehe Abbildung 24). Aber auch in diesen 38 Fällen werden in der Validitätsprüfung diskrepante Datensätze erkannt: Insgesamt sechs Datensätze weisen statistisch signifikante Diskrepanz zwischen direktem und indirektem Vergleich auf (Metaanalysenummer 66, 81, 83, 125, 164 und 226). Damit ist der Anteil an diskrepanten Datensätzen in diesen 38 Fällen mit 15,8 % (95 %-KI: 4,2 % bis 27,4 %) nicht geringer als in der Haupt- (12,9 %; 95 %-KI: 8,7 % bis 17,1 %) oder Subgruppenanalyse (15,5 %; 95 %-KI: 9,7 % bis 21,4 %).

5.7 Kongruenz in den Schlussfolgerungen von direktem und indirektem Vergleich

Alle systematischen Reviews, die sowohl den direkten als auch den indirekten Vergleich durchgeführt haben, werden daraufhin überprüft, ob die beiden Verfahren zu gleichlautenden Schlussfolgerungen kommen (siehe Tabelle 31).

Die Frage nach kongruenten Schlussfolgerungen hat fünf Antwortmöglichkeiten (siehe Tabelle 7).

Tabelle 7: Anzahl der kongruenten Schlussfolgerungen aus direktem und indirektem Vergleich

Möglichkeit	Kommen der direkte und der indirekte Vergleich zu der gleichen Schlussfolgerung?	Anzahl Gegenüberstellungen	Prozentualer Anteil
1	Ja, die gleiche Therapieoption wird sowohl im indirekten als auch im direkten Vergleich als signifikant überlegen erkannt.	54	21,8 % (95 %-KI: 16,6 % bis 26,9 %)
2	Ja, die Therapieoptionen werden sowohl im direkten als auch im indirekten Vergleich als gleich wirksam eingestuft (es wird kein signifikanter Therapieeffektunterschied gefunden).	122	49,2 % (95 %-KI: 43,0 % bis 55,4 %)
3	Nein, der Therapieeffektunterschied wird im indirekten Vergleich als nicht signifikant und im direkten Vergleich signifikant beurteilt.	39	15,7 % (95 %-KI: 11,2 % bis 20,3 %)

Fortsetzung Tabelle 7: Anzahl der kongruenten Schlussfolgerungen aus direktem und indirektem Vergleich

4	Nein, der Therapieeffektunterschied im indirekten Vergleich als signifikant und im direkten Vergleich nicht signifikant beurteilt.	28	11,3 % (95 %-KI: 7,4 % bis 15,2 %)
5	Nein, im indirekten Vergleich wird eine andere Therapieoption als signifikant überlegen beurteilt als im direkten Vergleich.	5	2,0 % (95 %-KI: 0,3 % bis 3,8 %)

KI = Konfidenzintervall.

In knapp der Hälfte der 248 Gegenüberstellungen wird sowohl im direkten als auch im indirekten Vergleich kein signifikanter Therapieeffektunterschied festgestellt (49,2 %; 95 %-KI: 43,0 % bis 55,4 %). Am zweithäufigsten wird sowohl vom indirekten als auch vom direkten Vergleich die gleiche Therapieoption als signifikant überlegen erkannt (21,8 %; 95 %-KI: 16,6 % bis 26,9 %). Addiert man die Fälle, in denen der direkte und der indirekte Vergleich Übereinstimmung in ihren Schlussfolgerungen erzielen, errechnet sich ein Anteil von 71 % (95 %-KI: 65,3 % bis 76,6 %). Das Pendant zu diesem Anteil ist der Anteil, in dem der direkte und indirekte Vergleich nicht zu den gleichen Schlussfolgerungen kommen: 29 % (95 %-KI: 23,4 % bis 34,7 %). Ins Verhältnis gesetzt werden 2,45-mal mehr kongruente Schlussfolgerungen zwischen direktem und indirektem Vergleich angetroffen als nicht-kongruente.

Für die Nicht-Übereinstimmung in den Schlussfolgerungen gibt es drei Varianten:

Erstens, der indirekte Vergleich kann einen signifikanten Therapieeffektunterschied dort erkennen, wo ihn der direkte Vergleich nicht erkennt (11,3 %; 95 %-KI: 7,4 % bis 15,2 %). In diesen Fällen weist der Effektschätzer des indirekten Vergleichs häufig eine höhere Präzision auf als der des direkten. Dies kann daran belegt werden, dass der indirekte Vergleich in diesen 28 Gegenüberstellungen im Median zehnmal mehr Studien einschließt als der direkte. Im Übrigen beruhen die direkten Vergleiche in zwölf Fällen auf Ergebnissen aus nur einer Studie und in weiteren sieben auf den Ergebnissen von zwei oder drei Studien.

Zweitens, der indirekte Vergleich kann einen signifikanten Therapieeffektunterschied dort nicht erkennen, wo ihn der direkte Vergleich erkennt (15,7 %; 95 %-KI: 11,2 % bis 20,3 %). In vielen dieser 39 Fälle weist der indirekte Vergleich eine zu niedrige Power auf. Dies kann daran erkannt werden, dass in diesen 39 Fällen nur 4,3-mal so viele Studien in den direkten Vergleich eingeschlossen werden wie in den direkten. Dies liegt deutlich unter dem Durchschnittswert der Gesamtstichprobe, in der der indirekte Vergleich im Median eine sechsfach größere Anzahl Studien einschließt als der direkte.

Drittens, im indirekten Vergleich zeigt sich eine andere Therapieoption als signifikant überlegen als im direkten (2 %; 95 %-KI: 0,3 % bis 3,8 %). Bei diesen fünf Datensätzen (Metaanalysenummer 43, 47, 83, 231 und 239 siehe Tabelle 28) wird auch eine statistisch signifikante Diskrepanz zwischen direktem und indirektem Vergleich beobachtet, deren mögliche Ursachen in Kapitel 6.5.1 „Kommen indirekter und direkter Vergleich zu identischen Schlussfolgerungen?“ diskutiert werden.

6 Diskussion und Beantwortung der Forschungsfragen

6.1 Literatur für die Methodenbeschreibung und -bewertung

Die systematische Literaturrecherche nach Methodenpapieren und -anwendungen gestaltete sich als schwierig, da es einerseits keine MeSH-Terminologie (Medical subject heading) zu indirekt vergleichenden Methoden gibt und andererseits eine unspezifische methodische Terminologie sehr viele falsch positive Treffer (d. h. thematisch irrelevante Referenzen) liefert. Aus diesem Grund wurde diesem HTA die Literaturrecherche aus dem systematischen Review von Glenny et al.¹⁰⁹ zugrunde gelegt und nur wenig angepasst.

In der Literaturselektion für die Validitätsprüfung wurde der Schwerpunkt auf systematische Reviews gelegt, die ausschließlich RCT eingeschlossen haben. Diese stellen, wenn methodisch korrekt durchgeführt, wegen ihrer geringen Biasanfälligkeit den höchsten Evidenzgrad für die Bewertung der Wirksamkeit von therapeutischen Verfahren dar^{5, 7, 12}. Auf diese Weise soll sichergestellt werden, dass in der Validitätsprüfung festgestellte Diskrepanzen zwischen direkt und indirekt vergleichenden Methoden nicht auf studiendesignbedingte Einflüsse zurückzuführen sind. Systematische Reviews wurden als solche akzeptiert, wenn zum Auffinden der einzuschließenden RCT eine systematische und umfassende Literatursuche dokumentiert worden ist.

Das Qualitätskriterium „Berücksichtigung von Heterogenität“ ist Gegenstand ausführlicher Analysen in diesem HTA. Weitere methodische und inhaltliche Qualitätskriterien der Reviews wurden nicht beurteilt, da die zentralen Fragestellungen des vorliegenden HTA ausschließlich die Gegenüberstellung von direkt und indirekt vergleichenden Methoden betreffen.

Die Anzahl an gefundenen Methodenpapieren und -anwendungen ist ausreichend, um die Forschungsfragen dieses HTA weitgehend zu beantworten. Ein Mangel an Methodenpapieren, wie bei den Methoden ohne Metaanalysen oder nicht-adjustierten indirekten Vergleichen, kann durch Informationen aus den Methodikabschnitten in Anwendungsbeispielen ausgeglichen werden. Die Anzahl der geeigneten Publikationen für die Validitätsprüfung ist im Vergleich zum Bericht von Glenny et al.¹⁰⁹ deutlich höher, dennoch fehlen nach wie vor ausreichend geeignete Beispiele für Metaregressionen, MTC und die sonstigen Methoden, um deren Validität abschließend beurteilen zu können.

6.2 Validität von indirekten Vergleichen

6.2.1 Grundvoraussetzung: Systematischer Review

Eine Grundvoraussetzung für die Bewertung der Wirksamkeit von therapeutischen Verfahren ist, dass die beste verfügbare Evidenz vollständig erfasst wird. Dies erfordert die Durchführung eines systematischen Reviews, der alle verfügbaren Studien aufsucht, kritisch bewertet und zusammenführt. Dies gilt unabhängig davon, ob für den Wirksamkeitsvergleich ein direktes oder indirektes Verfahren gewählt wird.

Die Methoden zur Erstellung qualitativ hochwertiger systematischer Reviews sind einer Vielzahl einschlägiger Methodenpapiere, Lehr- und Handbüchern zu entnehmen (u. a. Cochrane Handbuch⁸).

Wird ein systematischer Review als Element eines HTA konzipiert, steht zu Beginn des Prozesses die Formulierung der präzisen Forschungsfragen aus dem versorgungsrelevanten Kontext heraus. Bei der Bewertung der Wirksamkeit werden in diesem Zusammenhang regelmäßig angesprochen: Die interessierende Patientenklientel, die Interventionen von Interesse, die relevanten Zielgrößen (inklusive der zeitlichen Dimension) sowie das relevante Setting. Danach erfolgen Informationsrecherche und -charakterisierung sowie Planung der Informationssynthesen. Vor allem wenn indirekte Vergleiche geplant sind, bietet es sich an dieser Stelle des Assessment an eine kontextbezogene Charakterisierung der Evidenzlage aus sowohl methodischer als auch klinischer Sicht vorzunehmen. Ein Leitfaden hierzu könnte in Anlehnung an die von der GRADE Working Group³⁰ vorgeschlagenen Kriterien entworfen werden:

- Publikations-/Sprach-/Retrieval-Bias (systematische Verzerrungen der Ergebnisse von Übersichtsarbeiten, die durch unvollständiges Auffinden der relevanten Studienergebnisse entstehen): Wie wahrscheinlich ist es, dass die vorhandene Evidenz vollständig erfasst wurde?
- Studiendesign: Die im vorliegenden HTA betrachteten Methoden erfordern Evidenz aus RCT.
- Methodische Qualität der RCT: Aspekte wie Adäquatheit der Randomisationsmethode, Verblindung und die Vollständigkeit der Nachbeobachtung.
- Konsistenz der Studienergebnisse: Z. B. mithilfe eines Heterogenitätstests (vergleiche Kapitel 5.2.1 „Heterogenität“).
- Direktheit wird hier wie folgt verstanden: Wie vergleichbar sind die Studien aus klinischer Sicht (Probanden, Interventionen, Zielgrößen, Beobachtungsdauern) und in wie weit sind die Studien geeignet, die Forschungsfragen für das Gesamtprojekt zu beantworten?

Die Prüfung formaler Voraussetzungen für Metaanalysen bzw. indirekte Vergleiche schließt sich somit an die klinische, methodische und kontextbezogene Bewertung der Evidenzlage an. Nur erstere ist Gegenstand des vorliegenden Berichts.

6.2.2 Formale Voraussetzungen für indirekte Vergleiche

Homogenität als Voraussetzung für das Poolen von Ergebnissen unterschiedlicher Studien

Grundvoraussetzung für die Durchführung von indirekten Vergleichen ist die Homogenität der zusammenfassenden Studien. Die Annahme, auf der alle indirekten Vergleiche basieren, ist, dass allen Einzelstudienergebnissen ein gemeinsamer wahrer Therapieeffekt für die jeweiligen Therapien von Interesse zugrunde liegt^{60, 109}. Die Einzelstudienergebnisse weichen nicht (Fixed-effects-Annahme) oder nur durch Zufallsfehler bedingt (Random-effects-Annahme) von diesem wahren Therapieeffekt ab^{109, 134, 227}. Dies ist nur gegeben, wenn alle zusammenfassenden Studien sich nicht hinsichtlich wichtiger Patientencharakteristika, Complieraten, Studienprotokolle, methodischer Qualität und Auswertungsmethoden unterscheiden^{167, 185}.

Damit sind für die Durchführung von indirekten Vergleichen keine weiteren Grundvoraussetzungen zu erfüllen als für konventionelle Metaanalysen⁶⁰.

Heterogenität in der untersuchten Stichprobe und ihre Folgen

In der Praxis ist diese Grundvoraussetzung selten gegeben¹⁰⁴. Aus methodischen und klinischen Unterschieden in den Studien resultiert Heterogenität, die die gepoolten Ergebnisse verzerren kann.

Die Autoren von 20 der 33 (61 %) in der vorliegenden Literaturrecherche gefundenen systematischen Reviews mit direkten und indirekten Vergleichen berichten statistisch signifikante Heterogenität (siehe Tabelle 24, Eckert et al. ausgeschlossen). In nur vier Publikationen (12 %) ist keine statistisch signifikante Heterogenität nachweisbar, neun Publikationen (27 %) führen keinen statistischen Test durch oder publizieren sein Ergebnis nicht.

In den 20 Publikationen, in denen signifikante Heterogenität berichtet wird, weisen neun (45 %) mindestens einen Datensatz auf, dessen indirekter Vergleich ein signifikant diskrepantes Ergebnis in der Gegenüberstellung mit dem direkten Vergleich aufweist. Dahingegen ist unter den vier Publikationen mit homogenem Datensatz in nur einem systematischen Review ein Datensatz diskrepant (Anteil: 25 %).

Diese Zahlen geben Hinweise, dass erstens viele Autorengruppen indirekte Vergleiche vornehmen, obwohl sie Heterogenität in ihrem Studienpool festgestellt haben, zweitens die Wahrscheinlichkeit in einem indirekten Vergleich ein Ergebnis zu erzielen, dass von dem eines direkten Vergleichs deutlich abweicht, vermutlich um ein Vielfaches höher ist, wenn die Studienlage statistisch signifikante Heterogenität aufweist und drittens, dass Heterogenität nicht zwangsläufig zu verzerrten Ergebnissen von indirekten Vergleichen führen muss.

Bias als Folge von Heterogenität

Indirekte Vergleiche können unabhängig von der gewählten Methode alle Biasformen, die auch in Beobachtungsstudien auftreten, betreffen^{6, 8}. Hiermit muss vor allem dann gerechnet werden, wenn die zusammengefassten Studien unterschiedliche Patientencharakteristika, Complianceraten, Studienprotokolle, methodische Qualität und Auswertungsmethoden und somit Heterogenität aufweisen^{167, 185}. Die wichtigsten Biasformen sind der Selektionsbias und das Confounding. Zudem besteht Biaspotential bei indirekten Vergleichen, die mit einer nicht-aktiven Vergleichsgruppe als gemeinsamem Komparator durchgeführt werden.

1. Selektionsbias

Die Probanden werden in RCT zwar in den Einzelstudien randomisiert, die Randomisation findet aber nicht auf der Metaebene statt. Das heißt, dass die einzelnen Patienten nicht über alle Studien zufällig verteilt werden. Ein Selektionsbias kann eintreten, wenn einzelne Studien andere Einschlusskriterien als andere aufweisen und dadurch bestimmte Patientengruppen im gepoolten Ergebnis überrepräsentiert werden.

2. Confounding

Ist das überrepräsentierte Patientencharakteristikum mit dem Therapieeffekt assoziiert, kann es auch einen Confounder darstellen³². Confounder täuschen einen Therapieeffekt vor, der nicht ursächlich durch die Therapie, sondern durch den Confounder verursacht wird. Typische Confounder sind z. B. Alter, Geschlecht oder Raucherstatus.

3. Bias durch unterschiedliche nicht-aktive Vergleichsgruppen

Für indirekte Vergleiche wird häufig Placebo als gemeinsamer Komparator gewählt. Hierbei muss beachtet werden, dass ein Placebo auch einen therapeutischen Effekt besitzt, der durch viele Faktoren, wie z. B. Form, Farbe, Geschmack, Geruch und Applikationsart bei unterschiedlichen Patienten verschieden ausgeprägt ausfällt¹⁴⁰.

Ein Bias kann durch die Placebogruppen in indirekten Vergleichen, wie auch in konventionellen Metaanalysen dann auftreten, wenn in den Vergleichsgruppen der zusammenfassenden Studien unterschiedliche Effektstärken auftreten und somit keine Homogenität zwischen den Studien besteht⁶.

Analog kann auch die statistisch signifikante Diskrepanz in einem Datensatz von Psaty et al.¹⁸⁹ (Metaanalysenummer 198) erklärt werden. Für den indirekten Vergleich (ein MTC) werden Studienarme in denen mit Placebo, Standardtherapie oder „Keine Therapie“ behandelt wurde, zu dem gemeinsamen Komparator zusammengefasst. Der direkte Vergleich dagegen wird in einem Megatrial (n = 6.083) mit Placebogruppe vorgenommen.

Erkennung bedeutsamer Heterogenität

Statistisch erfasste Heterogenität sollte durch methodische oder klinische Unterschiedlichkeit in den Studien erklärt werden können⁸. Eine solche Betrachtung der Unterschiedlichkeit der zu poolenden Studien ist außerdem hilfreich, da Heterogenitätstests in vielen Fällen (vor allem bei wenigen Studien) nicht sensitiv genug sind, bedeutsame Heterogenität zu erkennen¹⁰⁵. Weitere Methoden zur Heterogenitätsaufklärung, wie Subgruppenanalysen oder Metaregressionen auszuschöpfen, werden ebenfalls empfohlen⁸.

Die Frage, wann bedeutsame Heterogenität vorliegt, kann nicht pauschal beantwortet werden. Ein erstes Indiz kann der I²-Test liefern, wenn er eine Heterogenität über 50 % anzeigt, die nicht durch Zufallsfehler erklärt werden kann¹²⁰. Es gibt bisher allerdings keinen Konsens, was als Schwellenwert für bedeutsame Heterogenität anzusehen ist⁶. Zur Vorbereitung der vergleichenden Analysen müsste ein systematischer Review mit indirektem Vergleich eine umfangreiche Heterogenitätsdiskussion führen. Standards für solch eine Heterogenitätserfassung und -diskussion sind noch nicht vorhanden⁶⁰. Es ist deshalb wichtig, dass sich der Leser des Reviews selbst ein Bild davon machen kann, ob es aus klinischer und methodischer Sicht sinnvoll war, die Studien zu poolen sowie einen indirekten Vergleich durchzuführen. Hierfür kommt der transparenten Darstellung aller in einen Review eingeschlossenen Studien eine besondere Bedeutung zu. Evidenztabelle sollten verfügbar sein, die alle relevanten Angaben zu den inhaltlichen und methodischen Studiencharakteristika enthalten.

Umgang mit Heterogenität

Die Wahl von Random effects-Modellen⁸⁶ kann aufgrund von Untersuchungen von Glenny et al.¹⁰⁹ bereits als ausreichend evaluierter Ansatz angesehen werden, um mit Heterogenität in indirekten Vergleichen umzugehen. Random effects-Modelle reduzieren zwar nicht die Heterogenität im Datensatz und somit auch nicht ihren potenziell ergebnisverzerrenden Effekt, aber sie bilden sie in weiten Konfidenzintervallen für den Schätzer des Therapieeffektunterschiedes ab. Es resultiert somit eine konservativere Schätzung als bei Fixed effects-Modellen. Nach Glenny et al.¹⁰⁹ liefern Fixed effects-Modelle zu präzise Effektschätzer, die unzureichend die Unsicherheiten, die Ergebnisse von indirekten Vergleichen betreffen, abbilden.

Diese Problematik tritt nur dann nicht auf, wenn die Heterogenität zwischen den Studien Null ist, da in diesem Fall die Random und Fixed effects-Modelle identische Ergebnisse liefern¹⁸¹.

Andere Wege mit Heterogenität in indirekten Vergleichen umzugehen stehen zwar zur Verfügung (Aufnahme von Kovariaten in Metaregressionen¹¹⁶ oder Aufnahme eines Inkonsistenzfaktors in den MTC¹⁵⁹), allerdings fehlt es noch an Beweisen, dass eine Verzerrung des Ergebnisses des indirekten Vergleichs durch Heterogenität durch ihren Einsatz verhindert werden kann.

6.2.3 Diskussion methodenspezifischer Ursachen für systematische Verzerrungen

Neben den im vorangegangenen Kapitel diskutierten Biasformen, die in allen indirekt vergleichenden Methoden auftreten können, werden in diesem Kapitel die Biasursachen diskutiert, die einen methodenspezifischen Hintergrund aufweisen. Zum Abschluss der Diskussion wird der Stellenwert jeder indirekt vergleichenden Methode für den Einsatz im Rahmen eines HTA eingeschätzt.

Nicht-adjustierter indirekter Vergleich

Das führende Kennzeichen des nicht-adjustierten indirekten Vergleichs ist, dass nur die Ergebnisse der Verumgruppen der Studien gepoolt werden^{104, 109}. Die Ergebnisse der Kontrollgruppen werden nicht in der Analyse berücksichtigt. Deshalb wird von einem Bruch der Randomisation gesprochen^{109, 217}. Die Strukturgleichheit der Stichprobe, die bei der zufälligen Verteilung der Probanden auf die Verum- und Kontrollgruppe entsteht, geht bei diesem Verfahren verloren⁸². Effekte von Confoundern werden nicht mehr dadurch abgeschwächt, dass sie sich in der Verum- wie in der Vergleichsgruppe, gleich stark auswirken und bei der Subtraktion der Therapieeffekte beider Gruppen eliminiert werden⁸².

Das Handbuch der Cochrane Collaboration⁸ beurteilt die nicht-adjustierten indirekten Vergleiche wie folgt: „Ein Ansatz, der niemals genutzt werden sollte, ist der direkte Vergleich von einzelnen relevanten Therapiearmen der Studien (...). Dieser Vergleich ignoriert die potenziellen Vorteile der Randomisation und leidet unter den gleichen (üblicherweise extremen) systematischen Fehlern, wie ein Vergleich von unabhängigen Kohortenstudien.“ (Zitat aus dem Englischen).

Die Einschätzung der Cochrane Collaboration wird von anderen Experten geteilt^{6, 82, 104, 109, 217} und in der Validitätsprüfung im vorliegenden HTA bestätigt, so dass von der Durchführung nicht-adjustierter indirekter Vergleiche grundsätzlich abgeraten wird.

Adjustierter indirekter Vergleich

Der adjustierte indirekte Vergleich berücksichtigt die Ergebnisse in den Kontrollgruppen der Einzelstudien. Auf diese Weise wird der Effekt der Randomisation – die Strukturgleichheit – in den Einzelstudien bewahrt^{82, 109}.

Werden die Ergebnisse der RCT zu einem Gesamteffektschätzer (θ_A versus X oder θ_B versus X) zusammengefasst, stellt dieser allerdings kein Ergebnis dar, dass aus zwei perfekt randomisierten Gruppen entstand. Der Grund hierfür ist, dass die Probanden zwar innerhalb der Studien randomisiert, aber nicht zwischen den Studien zufällig verteilt werden. Bei z. B. unterschiedlichen Ein- und Ausschlusskriterien der Einzelstudien können bestimmte Patientencharakteristika mit verzerrendem Effekt das Gesamtergebnis beeinflussen²¹⁷. Es sei darauf hingewiesen, dass dies auch in jeder konventionellen Metaanalyse der Fall sein kann¹⁰⁴.

Die einzige Besonderheit im adjustierten indirekten Vergleich ist, dass er die Ergebnisse aus zwei Metaanalysen (A versus X und B versus X) zusammenbringt²²⁷. Nicht nur die Studien innerhalb einer Metaanalyse, sondern auch in beiden Metaanalysen müssen sich methodisch und von den Charakteristika ihrer Patienten her ähnlich sein¹⁰⁴. Dies sollte in einer umfangreichen Heterogenitätsbetrachtung inklusive statistischer Tests auf Heterogenität überprüft werden²²⁷. Liegt bedeutsame Heterogenität vor, kann sie die Ergebnisse der Metaanalysen und damit auch die des adjustierten indirekten Vergleichs verzerren¹⁰⁹.

Es ist anzunehmen, dass ein adjustierter indirekter Vergleich theoretisch, bei einer perfekt homogenen Studienlage, die gleichen Ergebnisse wie ein direkter Vergleich, der die gleiche Anzahl Patienten in eine große Head-to-head-Studie einschließt, liefern würde^{55, 109}. Da diese perfekte Homogenität in der Praxis meistens nicht existiert, werden die Ergebnisse von adjustierten indirekten Vergleichen von denen direkter Vergleiche mehr oder weniger deutlich abweichen (siehe Kapitel 5.5.2.2 „Adjustierter indirekter Vergleich“)^{109, 217, 227}.

Indirekter Vergleich mittels Metaregression

Im Prinzip entspricht das in der Methodenbeschreibung dargestellte Vorgehen bei der Metaregression einem hinsichtlich der Ergebnisse in den Kontrollgruppen adjustierten indirekten Vergleich¹⁰⁹. Durch Subtraktion der Ereignisse in den Kontroll- von denen in den Verumgruppen auf Einzelstudien Ebene werden die Ergebnisse in den Kontrollgruppen, wie beim adjustierten indirekten Vergleich beschrieben, berücksichtigt. Der Randomisationseffekt auf Studienebene bleibt erhalten, da Patienten mit bestimmten Charakteristika (= potenziellen Confoundern) zufällig auf die Verum- und Kontrollgruppenanteile verteilt bleiben.

Glenny et al.¹⁰⁹ führen in Simulationsstudien den Beweis, dass die Ergebnisse von indirekten Vergleichen mittels Metaregressionen, die keine Kovariaten einschließen, denen von adjustierten indirekten Vergleichen weitgehend entsprechen. Aus diesen Ergebnissen lässt sich ableiten, dass für die Metaregression das gleiche Prinzip gelten kann wie für adjustierte indirekte Vergleiche: Bei vorhandener, bedeutsamer Heterogenität in der Studienlage sollte kein indirekter Vergleich durchgeführt werden, da Heterogenität zwischen den Studienergebnissen den gepoolten Gesamteffektschätzer verzerren.

Ein nicht-adjustiertes Vorgehen wäre auch mit Metaregressionen denkbar, wenn für die Therapieeffektschätzer in der Regressionsgleichung nicht die Differenz der Ereignisse aus Verum- und Kontrollgruppe, sondern nur die Ereignisse der Verumgruppe eingesetzt werden (vergleiche „Arm-level logistic regression model“ in Berlin et al.⁴¹). Für ein solches nicht-adjustiertes Vorgehen beim indirekten Vergleich mit Metaregression werden aber keine Anwendungsbeispiele gefunden. Literaturbeispiele für eine adjustiert indirekt vergleichende Metaregression gibt es dagegen einige. Sehr deutlich erkennbar ist die Adjustierung hinsichtlich der Kontrollgruppenereignisse z. B. in Mudge et al.¹⁷⁹, Petersen et al.¹⁸⁴ und Stettler et al.²²⁴.

Zu den beim adjustierten indirekten Vergleich genannten besonderen Charakteristika mit Einfluss auf die Validität der Methode kommen bei der Metaregression noch drei weitere wichtige Punkte hinzu:

1. Studienanzahl und -größe

Das Ergebnis einer Metaregression kann vor allem dann vom wahren Wert abweichen, wenn wenige Studien eingeschlossen werden, selbst wenn diese groß sind²³⁶. Das Handbuch der Cochrane Collaboration empfiehlt daher Metaregressionen nicht durchzuführen, wenn weniger als zehn Studien gepoolt werden⁸. Bei einer logistischen (Meta-)Regression kann zudem das Ergebnis vom wahren Wert abweichen, wenn die eingeschlossenen Studien klein sind²³⁶. Eine Gewichtung der Studien ist deshalb besonders wichtig, damit große, präzise Studien einen größeren Einfluss auf das Gesamtergebnis bekommen⁸.

2. Aufnahme von Kovariaten in die Regressionsgleichung(en)

Der große Vorteil des indirekten Vergleichs mittels Metaregression gegenüber dem adjustierten indirekten Vergleich ist die Möglichkeit hinsichtlich Kovariaten, die Heterogenitätsgründe darstellen, adjustieren zu können¹⁰⁹. Allerdings lässt sich hierdurch das Biasrisiko auch nicht vollständig eliminieren.

Der Grund hierfür ist, dass theoretisch alle Variablen, die zur klinisch relevanten Heterogenität beitragen, in das Metaregressionsmodell aufgenommen werden müssten²³⁶. Dabei ist mit statistischen Verfahren kaum feststellbar, ob alle relevanten Kovariaten berücksichtigt werden. Die Problematik liegt in der Kolinearität, auch Multilinearität genannt^{14, 40, 41, 178}: Einzelne Kovariaten, die nur gering zur Heterogenität beitragen und erst durch ihr Zusammenwirken zu statistisch signifikanter Heterogenität führen, können schwer identifiziert werden. Eine vollständige Sicherheit, dass auf den Therapieeffekt keine weiteren als die identifizierten Kovariaten als Confounder wirken, gibt es nicht¹⁷⁹.

3. Aggregationsbias

In der Literatur zu Metaregressionen wird gefordert, wenn möglich individuelle Patientendaten zu verwenden²³⁶, um einen Aggregationsbias zu vermeiden. Ein Aggregationsbias kann entstehen, wenn die erklärenden Variablen in der Regressionsgleichung Durchschnittswerte von Kovariaten darstellen. Wird z. B. nur das mittlere Alter der Probanden in den zu poolenden Studiendaten angegeben, so geben diese keine Information über die Streuung des Alters in den jeweiligen Einzelstudien. Ist der Therapieeffekt mit dem Alter linear assoziiert, kann es möglich sein, dass dies auf der Ebene der aggregierten Studiendaten nicht mehr zu beobachten ist²³⁶. Auch das Gegenteil kann eintreffen, nämlich dass keine lineare Assoziation zwischen dem Therapieeffekt und der Kovariate besteht, aber die aggregierten Studiendaten dieses vortäuschen²³⁶. Da häufig keine individuellen Patientendaten verfügbar sind, können Metaregressionen, die Confounder einschließen, einem Aggregationsbias unterliegen.

Liegt keine bedeutsame Heterogenität vor, besteht keine Notwendigkeit Metaregressionen zum Poolen der Daten einzusetzen, denn es kann auf konventionelle Metaanalysetechniken im Rahmen eines adjustierten indirekten Vergleichs ausgewichen werden. Nur wenn erklärable Heterogenität in dem zusammenzufassenden Studiensatz vorliegt, kann eine Metaregression nicht durch einen adjustierten indirekten Vergleich gleichwertig ersetzt werden. Es gilt aber zu beachten, dass in solchen Fällen die Grundvoraussetzung für einen indirekten Vergleich – eine homogene Studienlage – nicht vorliegt und damit die statistische Zusammenfassung der Daten prinzipiell nicht indiziert ist.

MTC

Der Grundannahme des MTC sind keine weiteren Annahmen zugrunde gelegt als bei adjustierten indirekten Vergleichen²²⁷: Alle zusammenzufassenden Studien müssen vergleichbar sein hinsichtlich ihrer Methodik und ihrer klinischen Charakteristika.

Durch den MTC wird die Randomisation auf Ebene der Einzelstudien bewahrt, allerdings können sich die Patientencharakteristika zwischen den Einzelstudien unterscheiden⁸. Dieser und andere Heterogenitätsauslöser können auch im MTC zu einer erheblichen Diskrepanz zwischen den Ergebnissen eines direkten und indirekten Vergleichs führen^{60, 67}, die im MTC Inkonsistenz genannt wird.

Es sind Methoden verfügbar, die bezüglich dieser Inkonsistenz(en) in einem Evidenznetzwerk adjustieren können^{159, 160}. Dabei gibt es jedoch noch keinen Konsens, ob der Einschluss von Inkonsistenzfaktoren regelhaft im MTC berücksichtigt werden sollte, oder ob die entsprechenden Methoden lediglich dazu eingesetzt werden sollten, die Lokalisation von Inkonsistenz in einem Evidenznetzwerk zu ermöglichen²⁰⁴. Diese Lokalisation von besonders hoher Inkonsistenz in einem Evidenznetzwerk ist möglich, da sich die Diskrepanzen zwischen direkten und indirekten Vergleichen für alle geschlossenen Schleifen in einem Evidenznetzwerk getrennt erheben lassen¹⁵⁹. Hohe Inkonsistenz in bestimmten Schleifen kann dann über statistische Tests auf statistische Signifikanz überprüft werden. Die Erfassung von Inkonsistenz in nicht-kreisförmigen Strukturen, wie sternförmigen oder linearen Strukturen, ist dagegen nicht möglich^{160, 204}. Zudem wird, wie es für statistische Tests auf Heterogenität in Metaanalysen bekannt ist, auch für statistische Tests auf Inkonsistenz eine geringe Power vermutet²⁰⁴. Das bedeutet, dass erstens, meistens nicht das gesamte Netzwerk auf Inkonsistenz überprüft werden kann und zweitens, dass, wenn die Tests keine Inkonsistenz anzeigen, noch nicht gesichert ist, dass Konsistenz wirklich vorliegt. Wird statistische Signifikanz dagegen erreicht, sollte die Konsistenzannahme widerrufen werden¹⁵⁹ und ein Verzicht auf das Poolen der Daten ist angezeigt¹⁶⁰.

Die Problematik, die gesamte Heterogenität und Inkonsistenz in einem Evidenznetzwerk zu erfassen führt auch zu der Frage, wie groß das Evidenznetzwerk sein sollte, um einen Therapievergleich mit einem MTC durchzuführen.

Ein sehr weites Evidenznetzwerk kann prinzipiell zwei Effekte haben: Zum einen kann es bei hoher Heterogenität der zusätzlich herangezogenen Studien im Verhältnis zu den Basisparametern dazu kommen, dass die Ergebnisse sich weit von denen der Basisparameter entfernen²¹⁷. Zum anderen können die zusätzlich herangezogenen Studien bei hoher Homogenität mit den Basisparametern auch die Präzision der gewünschten Ergebnisse des MTC erhöhen^{60, 227}. Eine grundsätzliche Empfehlung nach welchen Kriterien die Weite des Evidenznetzwerks, das theoretisch beliebig erweitert werden kann, bestimmt werden sollte, muss in Zukunft gefunden werden²²⁷.

Dies betrifft auch die Wahl der Priorverteilungen, die uneinheitlich in den identifizierten Anwendungsbeispielen gehandhabt wird (siehe Tabelle 26). In den identifizierten Beispielen werden nur non-informative Priorverteilungen gewählt oder keine Angaben zu ihnen gemacht.

Aus Reviews, die die gewählten Prioren nicht veröffentlichen, kann ein Leser keine Schlüsse ziehen, da die Wahl der Prioren Einfluss auf das Posteriorergebnis hat. Nur ein non-informativer Prior hat keinen Einfluss auf die Posteriorverteilung⁸. Auch für den Leser schwer nachvollziehbar ist die Validität der Datenauswertung einer MTC. Was in WinBUGS zwischen Prior- und Posteriorverteilung geschieht, sollte so anschaulich wie möglich (z. B. durch Publikation und Erläuterung des WinBugs-Codes) publiziert werden, damit sich ein Leser ein dezidiertes Bild von der Qualität eines Reviews mit einer MTC machen kann.

Da sich der MTC derzeit noch in Entwicklung befindet, kann das Verfahren derzeit noch nicht für den Einsatz unter Routinebedingungen empfohlen werden. Es ist zu erwarten, dass weitere Anwendungsbeispiele mit der Möglichkeit der Validitätsprüfung anhand der Ergebnisse von Head-to-head-Studien den Stellenwert des MTC in absehbarer Zeit weiter klären werden.

Sonstige Methoden

Für keine der unter dem Punkt „Sonstige Methoden“ zusammengefassten Verfahren lässt sich anhand der Angaben in den Publikationen die methodische Vorgehensweise hinsichtlich der Aufrechterhaltung der Randomisation in den Einzelstudien eindeutig nachvollziehen. Daher muss ihr Stellenwert für ein HTA derzeit als nicht beurteilbar eingeschätzt werden.

Methoden ohne Metaanalysen

Theoretisch gelten für indirekte Vergleiche ohne Metaanalysen die gleichen Voraussetzungen wie für metaanalytische Methoden: Die zusammenzufassenden Studien müssen homogen sein.

Beim Vergleich von Einzelstudien werden die Prinzipien des nicht-adjustierten (wenn Kontrollgruppen nicht berücksichtigt werden) und adjustierten indirekten Vergleichs (wenn Ergebnisse aus den Kontrollgruppen berücksichtigt werden) eingesetzt. Infolgedessen treffen die Angaben, die zu ihrer Validität in Gegenüberstellung zu direkten Vergleichen gemacht wurden, auch auf den Vergleich von Einzelstudien zu. Da Heterogenität bereits zwischen zwei Studien für den indirekten Vergleich auftreten kann, gilt auch für die Validität von indirekten Vergleichen von Einzelstudien die Voraussetzung, dass eine homogene Studienlage vorliegen muss.

Narrative indirekte Vergleiche, die mehr als zwei Studien miteinander vergleichen, sollten sehr vorsichtig interpretiert werden, da der Grund für die narrative Zusammenfassung häufig bedeutsame Heterogenität ist.

6.3 Zu Forschungsfrage 1 - Methoden für indirekte Vergleiche

6.3.1 Welche Methoden zur Durchführung indirekter Vergleiche therapeutischer Interventionen existieren bisher?

Bei der Benennung von Methoden für indirekte Vergleiche sind zwei Arbeitsschritte zu unterscheiden:

1. Die Zusammenfassung der verfügbaren Evidenz, einzeln für die interessierenden Therapieoptionen.
2. Der eigentliche vergleichende Schritt, in dem die zusammengefassten Daten für die interessierenden Therapieoptionen miteinander verglichen werden.

Zum ersten Punkt sind metaanalytisch vorgehende Methoden von rein narrativen Ergebniszusammenfassungen abzugrenzen. Die narrative Methode steht als einzige auch als Option zur Verfügung, wenn die Ergebnisse von mehr als zwei Studien miteinander verglichen werden sollen, aber auf eine Metaanalyse verzichtet werden soll oder muss. Die metaanalytischen Methoden lassen sich fast alle (103 von 108 systematischen Reviews, in denen indirekte Vergleiche durchgeführt werden) den vier großen Methodengruppen nicht-adjustierter indirekter Vergleich, adjustierter indirekter Vergleich, Metaregression und MTC zuordnen. Mit Ausnahme des MTC, sind diese Methoden aber nur Mittel zum Zweck, um gepoolte Effektschätzer für die Wirksamkeit der Therapieoptionen von Interesse zu erlangen.

Zum zweiten Punkt, dem vergleichenden Schritt, stehen ebenfalls vier Methoden zur Verfügung: Bildung eines Gesamteffektschätzers für die komparative Wirksamkeit durch Subtraktion oder Division der gepoolten Effektschätzer, Vergleich der Konfidenzintervalle auf Überlappung, statistischer Test auf Unterschiedlichkeit der Studienergebnisse oder ein narrativer Vergleich.

Die Bildung eines Gesamteffektschätzers für den Wirksamkeitsvergleich der Therapieverfahren ist dabei die einzige Möglichkeit, eine quantitative Aussage über die komparative Wirksamkeit der Therapien zu machen (z. B. über ein OR) und wird wohl aus diesem Grund auch am häufigsten gewählt (Anteil: 69,4 %). Die anderen drei Möglichkeiten liefern nur die qualitative Aussage, ob eine Therapie wirksamer ist als die andere, oder ob sie sich in ihrer Wirksamkeit nicht statistisch signifikant unterscheiden.

6.3.2 Wie können sie eingesetzt werden?

Die Grundlage für die im Rahmen des vorliegenden HTA betrachteten Methoden bilden RCT, in denen die zu vergleichenden Therapieverfahren gegen andere Komparatoren, wie z. B. Placebo oder Standardtherapie getestet werden.

Im Rahmen des vorliegenden HTA werden keine indirekten Vergleiche, die auch Daten aus nicht-randomisierten Studien einschließen, berücksichtigt. Hierdurch soll einerseits eine klare Abgrenzung zu den Cross design-Synthesen erreicht und andererseits sollen zusätzliche Biasquellen durch methodisch weniger hochwertige Studien für die Validitätsprüfung ausgeschlossen werden.

Die hier vorgestellten Methoden sind somit nur geeignet die komparative Wirksamkeit von Therapieverfahren unter Studien- (efficacy) und nicht unter Alltagsbedingungen (effectiveness) zu erheben.

Wie indirekte Vergleiche im Einzelfall eingesetzt werden können, ist in den Methodenbeschreibungen in Kapitel 5.3 „Beschreibung der indirekt vergleichenden Methoden“ ausgeführt.

Von einfacher Punkt- und Strichrechnung (nicht-adjustierter indirekter Vergleich), über konventionelle Metaanalysen (adjustierter indirekter Vergleich) und anspruchsvolleren Regressionen (Metaregression) zur Bayes'schen Statistik (MTC), steigen die Anforderungen an mathematischen Kenntnissen für den Anwender. Software, die diese Arbeit erleichtern kann, steht jedoch zur Verfügung: U. a. Review Manager[®] für den adjustierten indirekten Vergleich, STATA[®] für die Metaregression und WinBugs[®] für den MTC.

6.3.3 Unter welchen Voraussetzungen können sie eingesetzt werden?

Wie auch für konventionelle Metaanalysen ist die Grundvoraussetzung für indirekte Vergleiche mit metaanalytischen Methoden die Homogenität des zusammenzufassenden Studienpools⁶⁰. Allerdings ist die homogene Studienlage nicht nur für Studien innerhalb einer Therapieoption zu fordern, sondern für alle zu vergleichenden Therapieoptionen. Die Annahme, die allen indirekten Vergleichen dabei zugrunde gelegt wird ist, dass alle Einzelstudienergebnisse auf einem gemeinsamen wahren Therapieeffekt für die jeweiligen Therapien von Interesse basieren^{60, 109}. Die Einzelstudienergebnisse weichen nicht (Fixed effects-Annahme), oder nur durch Zufallsfehler bedingt (Random effects-Annahme), von diesem wahren Therapieeffekt ab^{109, 134, 227}. Dies kann nur gegeben sein, wenn alle Studien exakt nach derselben Methodik durchgeführt werden und Probandenkollektive mit identischen Patientencharakteristika einschließen. In der Praxis sind diese Voraussetzungen in den seltensten Fällen gegeben¹⁰⁴. Durch diese methodischen und klinischen Unterschiede resultiert Heterogenität, die die gepoolten Studienergebnisse verzerren kann.

In allen gefundenen Methodenpapieren und systematischen Reviews zu indirekten Vergleichen raten die Autoren von der Durchführung indirekter Vergleiche ab, wenn der zusammenfassende Datensatz bedeutsame Heterogenität aufweist.

6.4 Zu Forschungsfrage 2 - Häufigkeit der verschiedenen indirekt vergleichenden Methoden

Die Suchstrategie in der Literaturrecherche, mit ihrem starken Fokus auf RCT und Metaanalyse bedingt, dass diese Forschungsfrage nur für indirekt vergleichende Verfahren beantwortet werden kann, die mit metaanalytischen Methoden arbeiten.

Die Auswertungen zeigen, dass der adjustierte indirekte Vergleiche das mit Abstand am häufigsten eingesetzte Verfahren ist, die nicht-adjustierte indirekte Methode weist im untersuchten Zeitraum einen untergeordneten Stellenwert auf. Es ist anzunehmen, dass die Diskussionen in der publizierten Literatur, die das Biaspotenzial von nicht-adjustierten Verfahren hervorheben und von ihrer Durchführung abraten^{82, 104, 109, 217}, überzeugen konnten.

Angesichts der seit 1999 kontinuierlich ansteigenden Anzahl von systematischen Reviews mit indirekten Vergleichen ist zu vermuten, dass dieser Trend anhalten und in den folgenden Jahren vor allem durch den neu entwickelten MTC getragen wird. Für eine Kosten-Effektivitäts-Analyse wurde ein MTC bereits in einem NHS-Bericht eingesetzt²⁶⁴. Des Weiteren findet er Erwähnung in der neuen Fassung des britischen Guide to the methods of technology appraisal¹⁵. Der Einsatz von MTC in Kosten-Effektivitäts-Analysen ist zudem ein Feld, in dem aktuell intensiv geforscht, diskutiert und publiziert wird^{21, 134, 227}.

6.5 Zu Forschungsfrage 3 - Validität indirekter Vergleiche

6.5.1 Kommen indirekter und direkter Vergleich zu identischen Schlussfolgerungen?

In der untersuchten Stichprobe beträgt der Anteil an kongruenten Schlussfolgerungen zwischen direktem und indirektem Vergleich 71 % (95 %-KI: 65,3 % bis 76,6 %). In ca. drei Viertel der Fälle wird somit die Schlussfolgerung eines direkten Vergleichs durch einen indirekten bestätigt und umgekehrt.

In den meisten nicht übereinstimmenden Fällen sind die Unterschiede bei den Schlussfolgerungen durch eine mangelnde Präzision der Schätzung entweder im direkten oder im indirekten Vergleich erklärbar, die das Erkennen von Therapieeffektunterschieden mit einer akzeptablen statistischen Sicherheit verhindert. Zum einen tritt der klassische Fall auf, dass nur wenige kleinere Head-to-head-Studien ohne signifikanten Therapieeffektunterschied vorliegen, während der indirekte Vergleich auf eine große Studienanzahl zurückgreifen kann, die es ermöglicht, Effekte mit hoher Präzision zu schätzen und statistische Signifikanz zu erreichen. Zum anderen benötigen indirekte Vergleiche ca. sechsmal so viele Studien mit vergleichbarer Probandenzahl, um die gleiche Präzision zu erreichen wie eine einzige direkt vergleichende Studie.

Der gefürchtete Fall, dass direkter und indirekter Vergleich jeweils das andere Therapieverfahren als signifikant überlegen erachten, kann nur in fünf Fällen (Anteil 2 %; 95 %-KI: 0,3 % bis 3,8 %) beobachtet werden.

In vier dieser fünf Datensätze könnte dieses diskrepante Ergebnis durch hohe, statistisch signifikante Heterogenität im direkten und/oder indirekten Vergleich erklärt werden. In Metaanalyse 47 und 83 liegt statistisch signifikante Heterogenität sowohl im direkten als auch im indirekten Vergleich vor (siehe Tabelle 25). In den Datensätzen 231 und 239 wird das unter „Sonstige Methoden“ gefasste Verfahren von Tudur Smith et al. für den indirekten Vergleich eingesetzt und die Heterogenität im Studiensatz des indirekten Vergleichs nicht erhoben. Da jedoch der direkte Vergleich eine hohe Heterogenität aufweist, ließe sich die Diskrepanz zwischen direktem und indirektem Vergleich über dessen Heterogenität erklären. Beim fünften Datensatz (Metaanalysenummer 43) muss die Aussagekraft der einzigen für den direkten Vergleich zugrundeliegenden Head-to-head-Studie als gering eingeschätzt werden, da sich der indirekte Vergleich auf die Ergebnisse von 22 RCT stützen kann.

Es lassen sich also Anhaltspunkte dafür finden, dass wenn sowohl der indirekte als auch der direkte Vergleich ihre Voraussetzungen zur Durchführung (homogene Studienlage) einhalten und sie eine für signifikante Ergebnisse ausreichende Anzahl Studien einschließen, mit hoher Wahrscheinlichkeit zu übereinstimmenden Schlussfolgerungen kommen.

6.5.2 Wie ist die Validität indirekter Vergleiche gegenüber der von direkten Vergleichen zu beurteilen?

Für die Validitätsprüfung der indirekt vergleichenden Methoden werden in diesem HTA jedoch nicht die Schlussfolgerungen des direkten und indirekten Vergleichs, sondern die ungedeuteten Ergebnisse herangezogen, da sie statistisch besser auswertbar sind. Letztendlich weist eine geringe Diskrepanz in den Ergebnissen des direkten und indirekten Vergleichs aber darauf hin, dass beide auch zu den gleichen Schlussfolgerungen kommen.

Es werden zunächst die Ergebnisse der Haupt- sowie der Subgruppenanalyse kurz zusammenfassend diskutiert und anschließend die Datensätze, die statistisch signifikante Diskrepanz zum direkten Vergleich aufwiesen, näher betrachtet.

Ein diskrepantes Ergebnis zwischen direktem und indirektem Vergleich kann dadurch hervorgerufen werden, dass die Voraussetzungen für die Durchführung von indirekten Vergleichen nicht ausreichend beachtet werden^{109, 217, 218}. Als notwendige Voraussetzung für die Durchführung einer Metaanalyse wird eine ausreichend hohe klinische und methodische Homogenität aller in die Metaanalyse aufgenommenen Studien festgelegt^{8, 227}. Die klinische und methodische Homogenität der Studien spiegelt sich in Ergebnissen von Heterogenitätstests als statistische Heterogenität wider. Deshalb werden in den Fällen, in denen der direkte und indirekte Vergleich zu signifikant unterschiedlichen Ergebnissen führen, die entsprechenden Reviews daraufhin überprüft, ob statistisch signifikante Heterogenität ($p < 0,10$) vorliegt. Sollte signifikante Heterogenität in der Studienlage entweder des direkten oder des indirekten Vergleichs vorliegen, kann die Diskrepanz nicht nur auf methodischen Schwächen, sondern auch auf die heterogene Studienlage zurückgeführt werden^{104, 109}.

Im Rahmen des vorliegenden HTA wurde mit den Angaben der jeweiligen Reviewautoren gearbeitet. Für Übersichtsarbeiten, die keinen Heterogenitätstest durchgeführt oder berichtet haben, kann daher die Heterogenität oder Homogenität des zugrundeliegenden Studienpools nicht beurteilt werden.

6.5.2.1 Validität nicht-adjustierter indirekter Vergleiche

Zwar werden durch den nicht-adjustierten indirekten Vergleich die Ergebnisse eines direkten Vergleichs in unseren Daten nicht systematisch über- oder unterschätzt, allerdings ist die mittlere Abweichung der Ergebnisse der indirekten von denen der direkten Vergleiche und ihre Streuung ausgesprochen hoch.

Die folglich hohe Anzahl von festgestellten statistisch signifikanten Diskrepanzen zwischen direktem und indirektem Vergleich (25,5 %; 95 %-KI: 13,1 % bis 38 %) spricht dafür, dass sich mit nicht-adjustierten indirekten Vergleichen nicht verlässlich die Ergebnisse von direkt vergleichenden RCT vorhersagen lassen.

In der analysierten Stichprobe schließt kein Review Daten aus Head-to-head-Studien in den indirekten Vergleich ein (vergleiche Abbildung 6). Somit ist die Anzahl der analysierten Datensätze in der Hauptanalyse (alle Datensätze) und Subgruppenanalyse (nur die Datensätze, bei denen keine Head-to-head-Studien in den indirekten Vergleich eingeschlossen werden) gleich groß und ihre Ergebnisse sind identisch.

Lagen für den direkten und nicht-adjustierten indirekten Vergleich in den diskrepanten Datensätzen die Voraussetzungen zur Durchführung einer Metaanalyse vor?

Zwölf Metaanalysedatensätze aus unterschiedlichen Reviews liefern diskrepante Ergebnisse für den direkten und indirekten Vergleich. Bei zwei dieser Reviews sind die Voraussetzungen für mindestens eine Metaanalyse im direkten oder indirekten Vergleich nicht gegeben (siehe Tabelle 25). Signifikante Heterogenität tritt im indirekten Vergleich bei Silagy et al.²¹³ auf. Sowohl im indirekten als auch im direkten Vergleich liegt signifikante Heterogenität bei Zhang et al.²⁶⁹ vor.

Bei diesen beiden Reviews kann die Diskrepanz zwischen direktem und indirektem Vergleich somit auch in der hohen Heterogenität zwischen den gepoolten Studien begründet liegen.

Keine Heterogenität liegt dagegen in den Reviews von Rostom et al.²⁰¹ und Sanchez-Ramoz et al.²⁰⁶ vor. In diesen beiden Reviews ist die Wahrscheinlichkeit groß, dass die Methodik des nicht-adjustierten indirekten Vergleichs ursächlich dafür verantwortlich ist, dass ihre Ergebnisse von denen der direkten Vergleiche signifikant abweichen.

In den anderen acht Reviews mit statistisch signifikant diskrepanten direkten und nicht-adjustierten indirekten Vergleichen ist aufgrund fehlender Daten die Heterogenität der Studienlage nicht beurteilbar (siehe Tabelle 25).

6.5.2.2 Validität adjustierter indirekter Vergleiche

Die Validitätsprüfung des adjustierten indirekten Vergleichs führt sowohl in der Hauptanalyse als auch in der Subgruppenanalyse zu nahezu identischen Ergebnissen. Nur wenige Datensätze (32) haben auch Head-to-head-Studien in den adjustierten indirekten Vergleich eingeschlossen. Dies wurde entweder durch die Methodenerweiterung nach Song et al.²¹⁸ erreicht (siehe Kapitel 5.3.3 „Weiterentwicklung der adjustierten indirekten Vergleiche: Einschluss von Head-to-head-Studien“) oder indem die Ergebnisse kontrollierter Head-to-head-Studien - entgegen der Methodentheorie - in den indirekten Vergleich eingebracht wurden. Von dreiarmigen Head-to-head-Studien (A versus B versus X) wurden dabei entweder nur zwei Arme berücksichtigt und in die jeweilige Metaanalyse für A versus X oder B versus X eingeschlossen, oder die Ergebnisse der Kontrollgruppe X wurden doppelt gezählt (Gefahr des Korrelationsbias), indem die Ergebnisse sowohl in die A versus X- als auch in die B versus X-Metaanalyse eingeschlossen wurden.

Durch den adjustierten indirekten Vergleich werden in den vorliegenden Daten die Ergebnisse eines direkten Vergleichs nicht systematisch über- oder unterschätzt und die mittlere Abweichung des Ergebnisses des indirekten Vergleichs vom direkten Vergleich und ihre Streuung befinden sich auf einem Niveau, das nicht viele extreme Werte mit statistisch signifikanter Diskrepanz erwarten lässt.

Lagen für den direkten und adjustierten indirekten Vergleich in den diskrepanten Datensätzen die Voraussetzungen zur Durchführung einer Metaanalyse vor?

14 diskrepante Metaanalysedatensätze werden bei den adjustierenden indirekten Vergleichen ermittelt. Sie werden in insgesamt zehn systematischen Reviews publiziert (siehe Tabelle 25).

Bei sechs der Datensätze sind die Voraussetzungen für mindestens eine Metaanalyse im direkten oder indirekten Vergleich nicht gegeben (siehe Tabelle 25). Signifikante Heterogenität tritt nur im indirekten Vergleich bei Metaanalysenummer 115, 119 und 164, nur im direkten Vergleich bei Nummer 83 sowie sowohl im indirekten als auch im direkten Vergleich bei Nummer 116 und 125 auf. Bei diesen sechs Datensätzen kann die Diskrepanz zwischen direktem und indirektem Vergleich auch durch die hohe Heterogenität in und zwischen den gepoolten Studien begründet sein.

Keine Heterogenität liegt dagegen in den Metaanalysenummern 100, 118 und 148 vor. In diesen drei Datensätzen ist die Wahrscheinlichkeit groß, dass die Methodik des adjustierten indirekten Vergleichs ursächlich dafür verantwortlich ist, dass ihre Ergebnisse von denen der direkten Vergleiche signifikant abweichen.

In den anderen fünf Reviews mit statistisch signifikant diskrepanten direkten und adjustierten indirekten Vergleichen wurden nicht ausreichend Ergebnisse von Heterogenitätstests zur Beurteilung der Validität ihrer Metaanalysen publiziert (siehe Tabelle 25).

Die untersuchten 45 systematischen Reviews stellen somit eine Stichprobe dar, in der der adjustierte indirekte Vergleich in einem breiten Spektrum unterschiedlich heterogener Datensätze eingesetzt wurde. Die Einhaltung der Voraussetzung für den adjustierten indirekten Vergleich, eine homogene Studienlage, trifft nur in wenigen Fällen zu.

Unter diesen Voraussetzungen zeigt der adjustierte indirekte Vergleich mit nur ca. 12 % statistisch signifikant diskrepanten Ergebnissen in Gegenüberstellung zum direkten Vergleich eine hohe Ergebnisnisonformität mit dem direkten Vergleich.

6.5.2.3 Validität von Metaregressionen für indirekte Vergleiche

Für die Validitätsprüfung von Metaregressionen steht nur eine sehr begrenzte Anzahl von Datensätzen zur Verfügung. Zwei der sechs Datensätze werden im Gegensatz zur Hauptanalyse nicht in die Subgruppenanalyse eingeschlossen, da sie durch das Mixed model-Verfahren kontrollierte Head-to-head-Studien mit in die Metaregression aufnahmen (vergleiche Kapitel 5.3.5 „Weiterentwicklung der Metaregression: Aufnahme von Kovariaten in die Metaregression“). Die Ergebnisse der Hauptanalyse und Subgruppenanalyse sind aufgrund der kleinen Stichprobe als vorläufig zu betrachten.

In der analysierten Stichprobe werden die Ergebnisse direkter Vergleiche durch die korrespondierenden indirekten Gegenüberstellungen mittels Metaregression nicht systematisch über- oder unterschätzt. Zusätzlich liegen die oberen Konfidenzintervallgrenzen der durchschnittlichen Diskrepanz zwischen direktem und indirektem Vergleich in Haupt- und Subgruppenanalyse unter dem Wert für statistisch signifikante Diskrepanz. Dies spricht für eine hohe Validität der Metaregressionen für indirekte Vergleiche, da einerseits die geringe Fallzahl zu einem weiten Konfidenzintervall führt, dieses aber dennoch nicht die Signifikanzgrenze von 1,96 einschließt.

Die dritte erhobene Validitätskennziffer - der Anteil der statistisch signifikant diskrepanten Datensätze - hat angesichts der geringen Fallzahl keine Aussagekraft. Bei einem 95 %-Konfidenzintervall mit den Grenzen -13,2 % und 46,5 % lässt sich die Validität indirekter Vergleiche mit Metaregression nicht beurteilen.

Lagen für den direkten und indirekten Vergleich durch Metaregression in den diskrepanten Datensätzen die Voraussetzungen zur Durchführung einer Metaanalyse vor?

Bei den Metaregressionen wird nur ein statistisch signifikant diskrepanter Datensatz ermittelt. Er entstammt dem systematischen Review von Ballesteros et al.³⁵. Der Datensatz weist sowohl im direkten als auch im indirekten Vergleich keine signifikante Heterogenität auf (siehe Tabelle 25).

Deshalb ist die Wahrscheinlichkeit groß, dass in diesem Datensatz die Ursache dafür, dass die Ergebnisse von direktem und indirektem Vergleich diskrepanz sind, in der Methodik des indirekten Vergleichs liegt.

6.5.2.4 Validität von Mixed treatment comparison (MTC)

Durch Aufnahme von Head-to-head-Studien in den indirekten Vergleich kann das Gesamtergebnis des indirekten Vergleichs in Richtung des direkten Vergleichs beeinflusst werden. Deutliche Diskrepanz zwischen den Ergebnissen des indirekten und direkten Vergleichs wird unwahrscheinlicher.

Da es zu den Charakteristika des MTC gehört, vorhandene Head-to-head-Studien einzuschließen, werden nur wenige Beispiele für die Subgruppenanalyse (indirekte Vergleiche ohne Einschluss von Head-to-head-Studien) gefunden.

Folglich ist der MTC die einzige Methode, bei der sich die Ergebnisse aus Haupt- und Subgruppenanalyse bedeutsam unterscheiden. Einigkeit herrscht zwar in der Feststellung, dass die Ergebnisse eines direkten Vergleichs nicht systematisch über- oder unterschätzt werden, aber bei den übrigen Validitätsprüfungen gehen die Ergebnisse auseinander.

Die durchschnittliche Diskrepanz, die zwischen direktem und indirektem Vergleich auftritt, liegt in der Haupt- deutlich unter der in der Subgruppenanalyse. Dies ist dadurch erklärbar, dass durch Aufnahme der Head-to-head-Studien in den MTC das Gesamtergebnis des MTC in Richtung des direkten Vergleichs beeinflusst wird. Für die Validitätsprüfung des MTC bietet deshalb das Ergebnis der Subgruppenanalyse die höhere Aussagekraft. Während der MTC in der Hauptanalyse gegenüber dem adjustierten indirekten Vergleich sogar statistisch signifikant die Methode mit geringerer durchschnittlicher Diskrepanz darstellt, so liegen sie in der Subgruppenanalyse auf ähnlichem Niveau. Damit gilt, wie auch für den adjustierten indirekten Vergleich festgestellt, dass beim MTC nicht viele extreme Werte mit statistisch signifikanter Diskrepanz zu erwarten sind.

In Haupt- und Subgruppenanalyse ist es dann auch jeweils nur ein indirekter Vergleich (Psaty et al.), der statistisch signifikante Diskrepanz zum Ergebnis des direkten Vergleichs aufweist. Für die Hauptanalyse bedeutet dies einen Anteil von diskrepanten Datensätzen von 1,8 % und für die Subgruppenanalyse von 8,3 %.

Aus diesem Grund bezieht sich die Schlussfolgerung, dass die Ergebnisse des MTC gute Übereinstimmung mit den Ergebnissen von direkten Vergleichen zeigen, derzeit auf MTC, die direkt vergleichende Studien beinhalten. Für MTC, die keine Head-to-head-Studien einschließen, ist die Stichprobe zu gering, um verallgemeinerbare Aussagen über ihre Validität zu treffen.

Lagen für den direkten und indirekten Vergleich durch MTC in dem diskrepanten Datensatz die Voraussetzungen zur Durchführung einer Metaanalyse vor?

Der einzige im MTC gefundene diskrepante Datensatz (Metaanalysenummer 198) entstammt dem systematischen Review von Psaty et al.¹⁸⁹ (siehe Tabelle 25).

In diesem Review bleibt das Ausmaß an Heterogenität im Studienpool für den indirekten Vergleich unklar, da kein Heterogenitätstest durchgeführt wurde (siehe Tabelle 25). Für den direkten Vergleich liegt nur eine Head-to-head-Studie vor. Diese ist zudem mit 6.083 eingeschlossenen Probanden sehr groß, so dass eine hohe Ergebnissicherheit im direkten Vergleich angenommen werden kann.

Aufgrund fehlender Informationen zu der Heterogenität im indirekten Vergleich lässt sich somit keine Aussage machen, ob die Diskrepanz zwischen direktem und indirektem Vergleich auf die Methodik des MTC oder eher auf eine heterogene Studienlage zurückgeführt werden kann.

Eine hohe Heterogenität im MTC ist aber wahrscheinlich, wenn man berücksichtigt, dass Psaty et al.¹⁸⁹ als gemeinsamen Komparator Studienarme mit Placebo, Standardtherapie oder „Keine Therapie“ zusammengefasst haben. Dies kann zu dem in Kapitel 6.2.2 „Formale Voraussetzungen für indirekte Vergleiche“ unter dem Abschnitt „Biasformen als Folge von Heterogenität, die in allen indirekten Vergleichen potenziell auftreten können“ beschriebenen Bias durch unterschiedliche nicht-aktive Vergleichsgruppen führen.

6.5.2.5 Validität von sonstigen Methoden für indirekte Vergleiche

Bei den sonstigen Methoden können keine aussagekräftigen Ergebnisse erwartet werden, da nur drei Reviews zu den drei verschiedenen methodischen Ansätzen für die Validitätsprüfung zur Verfügung stehen. Aus diesem Grund werden nur die Diskrepanzen in den einzelnen Datensätzen dargestellt und auf weitere Analysen verzichtet.

Lagen für den direkten und mit sonstigen Methoden durchgeführten indirekten Vergleich in den diskrepanten Datensätzen die Voraussetzungen zur Durchführung einer Metaanalyse vor?

Alle vier statistisch signifikant diskrepanten Datensätze entstammen dem systematischen Review von Tudur Smith et al.²⁴⁰, der mit individuellen Patientendaten arbeitet. Die Autoren berichten für keinen der indirekten Vergleiche die Heterogenität der zugrundeliegenden Studien (siehe Tabelle 25). Bei den konventionellen Metaanalysen für die direkten Vergleiche wurden dagegen Heterogenitätstests durchgeführt. Signifikante Heterogenität liegt in den Metaanalysen der direkten Vergleiche der Datensätze 231, 239 und 240 vor, während der Test in Nummer 241 keine Heterogenität findet.

Die Diskrepanz zwischen direktem und indirektem Vergleich könnte somit in drei der vier Datensätze auch auf die Nicht-Einhaltung der Voraussetzung für die Metaanalyse des direkten Vergleichs zurückgeführt werden und muss nicht zwangsläufig mit der von Tudur Smith et al. verwendeten Methodik zusammenhängen.

6.5.2.6 Validität von indirekten Vergleichen ohne Metaanalysen

Für die Validitätsprüfung von indirekt vergleichenden Methoden ohne Metaanalysen werden keine geeigneten systematischen Reviews gefunden, was vor allem darauf zurückzuführen ist, dass in der Literaturrecherche keine hohe Sensitivität bei systematischen Reviews ohne Metaanalysen angestrebt wurde.

6.5.3 Limitationen der Validitätsprüfung

1. Empirie

Für die durchgeführte Validitätsprüfung wird angestrebt alle geeigneten, bisher in der medizinisch wissenschaftlichen Literatur publizierten systematischen Reviews, die sowohl einen direkten als auch einen indirekten Vergleich (unter Verwendung metaanalytischer Methoden) durchgeführt haben, einzuschließen. Auf Basis dieser Daten ist die Validitätsprüfung rein empirisch beschreibend und lässt sich nur begrenzt verallgemeinern.

Ein kausaler Beweis der Validität einer indirekt vergleichenden Methode könnte über mathematisch modellierte Beispiele erfolgen. Allerdings bleibt auch diesem Fall die Relevanz der Ergebnisse für die Praxis unklar, da sie nur für die besonderen Annahmen der Modellbeispiele eine hohe Aussagekraft hätten^{104, 109}. Die Realität der unendlich vielen verschiedenen Studienlagen, in denen indirekte Vergleiche eingesetzt werden können, lässt sich nicht simulieren¹⁰⁴. Aus diesem Grund soll mit dieser empirischen Validitätsprüfung ein Abbild der Validität von indirekten Vergleichen in ihrer praktischen Verwendung gegeben werden.

Man könnte zudem vermuten, dass die empirische Validitätsprüfung einem Publikationsbias unterliegen könnte, wenn Ergebnisse von indirekten Vergleichen, die deutlich von denen direkt vergleichender Studien abweichen, überproportional häufig nicht publiziert werden. Der Einfluss eines solchen Publikationsbias ist in der untersuchten Stichprobe aber vermutlich gering, da bei allen indirekt vergleichenden Methoden fast gleich viele Reviews identifiziert werden, in denen der indirekte den direkten Vergleich über- oder unterschätzt. Auch deutliche Über- und Unterschätzungen, die statistische Signifikanz erreichen, wurden in nicht geringer Zahl durch die Literaturrecherche identifiziert.

2. Validität des Komparators „Direkter Vergleich“

Bei der Betrachtung der Heterogenität in den statistisch signifikant diskrepanten Datensätzen kann in einigen Fällen vermutet werden, dass die beobachtete Diskrepanz durch eine hohe Heterogenität in der Metaanalyse des direkten Vergleichs bedingt ist. Für solche Fälle postulieren Song et al.²¹⁹, dass die Validität eines indirekten Vergleichs sogar höher einzuschätzen ist als diejenige des direkten Vergleichs, da eine hohe methodische Qualität des direkten Vergleichs nicht immer gegeben ist^{104, 217}.

Auch eine kleine Stichprobe beim direkten Vergleich, der zufällig eine nicht-repräsentative Stichprobe gezogen haben kann (z. B. bei nur einer Head-to-head-Studie mit wenigen Probanden), kann ursächlich für die Diskrepanz zum Ergebnis eines indirekten Vergleichs sein, wenn sich dieser auf eine große Anzahl Studien mit entsprechend großer Probandenzahl stützen kann^{185, 219}.

Nicht nur Zufallsfehler und Heterogenität im indirekten Vergleich, sondern auch dergleichen im direkten Vergleich tragen somit dazu bei, dass die Ergebnisse von direktem und indirektem Vergleich nicht deckungsgleich sind^{109, 219}.

3. Präzision der Effektschätzer des direkten und indirekten Vergleichs

Extreme Diskrepanzen zwischen den Punktschätzern für direkte und indirekte Vergleiche können häufig nicht mit statistischer Signifikanz (z. B. auf dem 5 % Niveau) nachgewiesen werden, wenn einer der Effektschätzer (entweder vom direkten oder vom indirekten Vergleich) eine geringe Präzision aufweist. Dies ist in der untersuchten Stichprobe, wie in Kapitel 5.6.4 „Einfluss der Präzision indirekter Vergleiche auf die Validitätsprüfung“ gezeigt, nur bei wenigen Datensätzen der Fall. Im Validitätsprüfungskapitel werden die Diskrepanzen zusätzlich in Forest-Plots dargestellt. An ihnen ist auf den ersten Blick erkennbar, dass nur wenige erhebliche Diskrepanzen zwischen den Ergebnissen von direktem und indirektem Vergleich in dieser Stichprobe aufgrund geringer Präzision das Niveau statistische Signifikanz nicht erreichen.

6.6 Zu Forschungsfrage 4 - Welchen Einfluss hat es auf die Validität, wenn Head-to-head-Studien mit in den indirekten Vergleich aufgenommen werden?

Nach der Definition eines indirekten Vergleichs (siehe Glossar) ist der Einschluss von Head-to-head-Studien zunächst einmal nicht vorgesehen¹⁵.

Da es jedoch das Ziel eines systematischen Reviews ist, die gesamte verfügbare, geeignete Evidenz zu einer Fragestellung abzubilden¹⁷⁴, sollten neben dem indirekten Vergleich, wenn verfügbar, auch die Ergebnisse von direkt vergleichenden Studien in die Beantwortung der Forschungsfrage mit einfließen^{60, 104}. Ein nicht-signifikantes Ergebnis eines direkten Vergleichs kann durch Hinzunahme von Evidenz aus einem indirekten Vergleich durch Erhöhung der Präzision des Effektschätzers Signifikanz erreichen^{60, 217-219}.

Die Cochrane Collaboration empfiehlt dagegen in ihrem aktuellen Methodenpapier den direkten und indirekten Vergleich getrennt voneinander durchzuführen, darzustellen und zu interpretieren⁸. Bei Inkonsistenz zwischen den Ergebnissen ist den direkten Vergleichen eine höhere Aussagekraft einzuräumen⁸.

Diesen Empfehlungen schließt sich auch das im Juni 2008 erneuerte Methodenpapier des NICE an¹⁵, geht aber noch einen Schritt weiter, indem es auch die Möglichkeit zulässt, den direkten und indirekten Vergleich simultan durchzuführen und zu einem Gesamteffektschätzer zusammenzufassen: „Wenn Head-to-head-Studien existieren, kann Evidenz von MTC präsentiert werden, wenn angenommen wird, dass sie Informationen ergänzt, die nicht aus dem Head-to-head-Vergleich verfügbar sind¹⁵“ (Zitat aus dem Englischen übersetzt).

Neben den MTC werden auch für die anderen drei indirekt vergleichenden Methoden Ansätze berichtet, die den Einschluss von Head-to-head-Studien vorsehen: In einen nicht-adjustierten indirekten Vergleich können die einzelnen Arme der Head-to-head-Studien aufgenommen werden, da die Ergebnisse der Kontrollgruppen nicht berücksichtigt werden (siehe Kapitel 5.3.1 „Nicht-adjustierter indirekter Vergleich“). Ein Ergebnis eines adjustierten indirekten Vergleichs kann über die Inverse-Varianz-Methode von Song et al.²¹⁸ mit dem Ergebnis eines direkten Vergleichs zusammengefasst werden (siehe Kapitel 5.3.2 „Adjustierter indirekter Vergleich“). Metaregressionen können durch die Methodenerweiterung von Hasselblad et al.¹¹⁶ um die Ergebnisse von Head-to-head-Studien, die zusätzlich einen Studienarm mit dem gemeinsamen Komparator mitführen, ergänzt werden (siehe Kapitel 5.3.6 „Weiterentwicklung der Metaregression: Einschluss von kontrollierten Head-to-head-Studien (Mixed model)“).

Der Review von Gartlehner und Moore¹⁰⁴ von 2008 bezieht zu dieser Thematik Stellung, indem er feststellt: „Wenn die Populationen ähnlich erscheinen, dann können Analysen die Ergebnisse von direkten und indirekten Ansätzen kombinieren, unabhängig von Unterschieden in den Ergebnissen.“ (Zitat aus dem Englischen übersetzt). Eine empirische Validitätsprüfung von indirekten Vergleichen, die Ergebnisse von Head-to-head-Studien beinhalten, kann im vorliegenden HTA nur für den MTC vorgenommen werden, da für die anderen genannten Ansätze bisher nicht ausreichend Anwendungsbeispiele publiziert wurden. Es zeigt sich, dass durch die Hinzunahme von Head-to-head-Studien zu einem MTC in keinem von den 47 Fällen das Ergebnis des MTC signifikant vom Ergebnis des direkten Vergleichs abweicht. Zudem liegt eine sehr geringe durchschnittliche Diskrepanz zwischen den Ergebnissen der direkten und indirekten Vergleiche vor (siehe Kapitel 5.5.2.4 „Mixed treatment comparison (MTC)“). Für die bisher publizierten Beispiele gilt daher, dass die MTC, die Head-to-head-Studien mit einschließt, eine ähnlich hohe Validität aufweist, wie Head-to-head-Studien allein. Damit stützen die hier vorgelegten Ergebnisse die Entscheidung des NICE, diese Methode, auch wenn bereits Head-to-head-Studien vorliegen, zum zusätzlichen Informationsgewinn einzusetzen⁶⁰.

6.7 Zu Forschungsfrage 5 - Lässt sich ein Goldstandard für indirekte Vergleiche identifizieren?

Die Frage, welche indirekt vergleichende Methode zu präferieren ist, lässt sich nicht pauschal beantworten. Dies muss von Fall zu Fall entschieden werden. Vier Kriterien sind bei der Auswahl einer indirekt vergleichenden Methode zu berücksichtigen:

1. Validität der Methode

Es sollte eine indirekt vergleichende Methode ausgewählt werden, deren Ergebnisse keine unerklärlichen Diskrepanzen zu denen direkt vergleichender Studien aufweisen. In den vorliegenden Daten ist dieses Kriterium lediglich für die nicht-adjustierten indirekten Vergleiche grundsätzlich nicht erfüllt. Die anderen indirekt vergleichenden Methoden können valide Ergebnisse liefern, wenn ihre Voraussetzung der homogenen Studienlage eingehalten wird. Diese Beobachtungen werden auch gestützt durch die Daten und Interpretationen anderer Arbeitsgruppen^{8, 82, 104, 109, 217, 227}.

2. Anzahl der zu vergleichenden Therapien

Die Anzahl der zu vergleichenden Therapien bedingt auch die Methodenauswahl²²⁷. Sollen nur zwei Therapieverfahren indirekt miteinander verglichen werden, können primär der adjustierte indirekte Vergleich und die Metaregression (falls ausreichend Studien zur Verfügung stehen) eingesetzt werden. Ein MTC ist auch möglich, jedoch ist der Aufwand für ihre Anfertigung viel höher als bei den anderen beiden genannten Methoden und es sollte genau begründet werden, wie weit man das Netzwerk der RCT um den Vergleich der zwei Therapien spannt²²⁷.

Sollen mehr als zwei Therapien in eine Rangfolge hinsichtlich ihrer Wirksamkeit gebracht werden, ist nur der MTC einsetzbar^{8, 60, 159}.

3. Wenn Head-to-head-Studien eingeschlossen werden sollen

Sollen Head-to-head-Studien in den indirekten Vergleich eingeschlossen werden, ist dies am elegantesten mit einem MTC durchführbar, in der Informationen aus allen Armen der Head-to-head-Studien verarbeitet werden können²²⁷.

Der adjustierte indirekte Vergleich nach Song et al.²¹⁸ kann nur die zwei Arme mit den interessierenden Therapieoptionen einer Head-to-head-Studie einschließen.

Metaregressionen dagegen können über die Methode nach Hasselblad et al.¹¹⁶ nur dreiarmige Head-to-head-Studien, die den gemeinsamen Komparator als dritten Arm aufweisen, einschließen. Potenziell vorliegende zweiarmige Head-to-head-Studien können nicht berücksichtigt werden.

4. Bei heterogenem Datensatz

Bei bedeutsamer Heterogenität darf kein indirekter Vergleich durchgeführt werden¹⁰⁹.

Liegt unbedeutende Heterogenität vor, sollte ein Random statt eines Fixed effects-Modells eingesetzt werden, um diese Heterogenität zwischen den Studien durch eine konservative Schätzung zu berücksichtigen^{8, 109}.

Liegt ein weitgehend homogener Studiensatz vor, können sowohl Fixed als auch Random effects-Modelle eingesetzt werden, da sie zu ähnlichen Ergebnissen führen¹⁸¹. Für alle in diesem HTA beschriebenen Methoden für indirekte Vergleiche stehen sowohl Fixed als auch Random effects-Modelle zur Verfügung (siehe entsprechende Methodenpapierzitate in Kapitel 5.3 „Beschreibung der indirekt vergleichenden Methoden“).

6.8 Status quo der Implementation von indirekten Vergleichen in Methodenpapieren von HTA-Institutionen

Aus HTA-Institutionen liegen zwei aktuelle Methodenpapiere vor, die sich mit der Thematik der indirekten Vergleiche befassen (veröffentlicht nach Redaktionsschluss für die systematischen Literaturrecherchen).

Das erste stammt von der Agency for Healthcare Research and Quality (AHRQ) aus den USA und wurde im April 2008 als Entwurf⁶ veröffentlicht (Informationsstand vom 01.07.2009). Der Stellenwert, der indirekt vergleichenden Methoden zugeordnet wird, lässt sich aus dem folgenden Zitat ableiten:

„Nichtsdestotrotz ist die Evidenz von Head-to-head-Studien immer den Ergebnissen von adjustierten indirekten Vergleichen von placebo- oder aktiv-kontrollierten Studien vorzuziehen, um einen Wirksamkeits- und Schadensvergleich zu evaluieren“ (Zitat aus dem Englischen übersetzt)⁶.

Das NICE (Großbritannien) hat im Juni 2008 indirekte Vergleiche (inklusive MTC) als Methode im Rahmen systematischer Reviews bzw. HTA bereits in sein offizielles Methodenpapier¹⁵ aufgenommen. Dort heißt es unter anderem: „Es gibt Situationen, in denen Daten von Head-to-head-RCT zu Interventionen (und/oder Komparatoren) nicht verfügbar sind. In diesen Fällen sollten indirekte Vergleiche berücksichtigt werden¹⁵“ (Zitat aus dem Englischen übersetzt). Zudem gibt das NICE-Methodenpapier differenzierte Empfehlungen, unter welchen Umständen welche indirekt vergleichende Methode eingesetzt werden sollte. Sie werden an verschiedenen Stellen des vorliegenden HTA zitiert (u. a. Kapitel 2.1.1 „Einsatzszenarien“).

Auch das deutsche IQWiG hat 2008 erstmals einen indirekten Vergleich zur Nutzenbewertung therapeutischer Verfahren eingesetzt. In dem am 09.09.2008 erschienenen Vorbericht „Vergleichende Nutzenbewertung verschiedener antihypertensiver Wirkstoffgruppen als Therapie der ersten Wahl bei Patienten mit essentieller Hypertonie“ wurde ein MTC im Rahmen einer Nutzenbewertung durchgeführt¹³. Im Methodenpapier des IQWiG (Version vom 27.05.2008) findet sich allerdings (noch) keine Beschreibung von indirekt vergleichenden Methoden.

6.9 Ergänzungen zum aktuellen Forschungsstand zu indirekten Therapievergleichen durch diesen HTA-Bericht

Die Basis des vorliegenden HTA-Berichts stellt die Aktualisierung des HTA-Berichts von Glenny et al.¹⁰⁹ dar, dessen systematische Literaturrecherche nur Literatur bis 1999 abdeckt. Alle anderen Übersichtsarbeiten zu Methoden indirekter Therapievergleiche (z. B. Gartlehner et al.) weisen keine systematische Literaturrecherche auf und geben keinen vollständigen Überblick über alle indirekt vergleichende Methoden.

In diesem HTA sind alle 2008 verfügbaren Methoden ausführlich beschrieben. Es wird angestrebt sie so verständlich wie möglich zu erklären, damit Personen, die systematische Reviews durchführen, einen Überblick über die zur Verfügung stehenden Verfahren gewinnen können. Eine vollständige Auflistung der weiterführenden Literatur kann ein Reviewer der Literaturübersicht zu Beginn jedes Methodenbeschreibungskapitels entnehmen.

Dies schließt auch die Methodengruppe des MTC mit ein, die Glenny et al.¹⁰⁹ noch nicht besprechen konnten, da ihr erstes richtungsweisendes Methodenpapier erst 2002 (Lumley¹⁶⁰) erschien.

Der HTA-Bericht von Glenny et al.¹⁰⁹ war auch bisher die einzige publizierte Literaturübersicht, die eine empirische Validitätsprüfung der Methoden enthält (publiziert in Glenny et al.¹⁰⁹ und Song et al.²¹⁷). Jedoch wurde diese empirische Validitätsprüfung nur für den adjustierten und nicht-adjustierten indirekten Vergleich durchgeführt. Diese können nun auf eine breitere Literaturbasis gestellt werden. Es wurden auch Konfidenzintervalle berechnet, mit deren Hilfe es besser abschätzbar ist, mit welchen Größenordnungen an signifikant diskrepanten Ergebnissen bei den verschiedenen indirekt vergleichenden Methoden zu rechnen ist. In der Validitätsprüfung auf durchschnittliche Diskrepanz überschneiden sich diese Konfidenzintervalle von adjustiertem und nicht-adjustiertem indirektem Vergleich nicht, sodass erstmals auch statistisch signifikant die Überlegenheit der Adjustierung bzgl. der Ergebnisse in den Kontrollgruppen nachgewiesen werden konnte (siehe Kapitel 5.5.3.3 „Ausmaß der durchschnittlichen Diskrepanz“).

Auch für die Hypothese von Glenny et al.¹⁰⁹, dass indirekte Vergleiche die Ergebnisse direkter Vergleiche nicht systematisch über- oder unterschätzen, konnte ein statistisches Testverfahren entwickelt werden, dass diese Hypothese durch empirische Evidenz stützt (siehe Kapitel 4.5.2.3 „Test auf systematische Über- oder Unterschätzung“).

Die indirekt vergleichenden Methoden Metaregression, MTC und die Methode nach Tudur Smith et al. werden erstmals einer empirischen Validitätsprüfung unterzogen. Aufgrund der bisher begrenzten Anzahl an publizierten Anwendungen dieser Methoden, bedarf es aber einer Wiederholung dieser Analysen in der Zukunft, wenn ausreichend Beispiele für aussagekräftige Ergebnisse zur Verfügung stehen.

Zuletzt konnte durch das Kapitel zur Präzisionsbetrachtung indirekter Vergleiche empirisch die Faustregel bestätigt werden, dass ein indirekter Vergleich viermal so viele (möglichst gleichgroße) Studien einschließen muss, um gleich präzise Ergebnisse zu erhalten, wie ein direkter Vergleich.

6.10 Forschungsbedarf

Vergleichende Validitätsprüfung an einem Beispiel

Mit Ausnahme der Publikation von Glenny et al.¹⁰⁹ stehen für die vorliegende Validitätsprüfung ausschließlich systematische Reviews zur Verfügung, in denen jeweils eine Methode des indirekten Vergleichs eingesetzt wurde. Für eine direkt vergleichende Validitätsprüfung der verschiedenen Methoden müssten alle an dem gleichen Studiensatz angewendet werden, zu dem auch als Referenzgröße eine Metaanalyse von mehreren homogenen Head-to-head-Studien mit sehr hoher interner Validität zur Verfügung steht. An einem solchen exemplarischen Beispiel ließen sich die Ergebnisse der verschiedenen indirekt vergleichenden Methoden einander gegenüberstellen und an dem Ergebnis der Metaanalyse der Head-to-head-Studien messen. Dieses Vorgehen könnte auch Aufschlüsse über die Validität der Methoden geben, für die bisher die Anzahl der Anwendungsbeispiele zur Validitätsprüfung nicht ausreicht. Dies betrifft vor allem die Metaregression und einen MTC, der vorhandene Head-to-head-Studien nicht mit einschließt.

Vergleichende Validitätsprüfung adjustierter versus nicht-adjustierter indirekter Vergleich

Des Weiteren könnte beispielhaft an einem oder mehreren systematischen Reviews mit adjustierten indirekten Vergleichen durch Wiederholung der Metaanalysen mit der Methodik des nicht-adjustierten indirekten Vergleichs getestet werden, unter welchen Bedingungen (Heterogenität) durch einen nicht-adjustierten indirekten Vergleich vergleichbare Ergebnisse erzielt werden können.

Da die Adjustierung auch so interpretiert werden kann, dass sie das Ausmaß der Heterogenität zwischen den Studien deutlich vermindert, besteht theoretisch die Möglichkeit, dass auch durch nicht-adjustierte indirekte Vergleiche weniger diskrepante Ergebnisse erreicht werden, wenn hier strenge Einschlusskriterien für die Studien aufgestellt werden. Diese sollten eine vergleichbare Patientenpopulation hinsichtlich Indikationsstellung, Krankheitsstadium, Begleiterkrankungen etc. und eine vergleichbare Methodik der Studien hinsichtlich Erhebungsinstrumenten, Zielgröße, Beobachtungszeitraum, erlaubten Begleittherapien etc. sicherstellen. Es ist zu überprüfen, ob die Ergebnisse solch strengerer nicht-adjustierter indirekter Vergleiche nicht ähnlich selten deutlich diskrepant zu den Ergebnissen eines direkten Vergleichs wären, wie adjustierte indirekte Vergleiche.

Bei welchem Ausprägungsniveau von Heterogenität kann ein indirekter Vergleich noch valide Ergebnisse liefern?

Den aktuellen Status quo in der methodischen Forschung zur Heterogenitätserhebung beschreiben Caldwell et al.⁶⁰ wie folgt: „Es ist unwahrscheinlich, dass die Annahmen hinter Vergleichen multipler Therapien statistisch nachweisbar sind und es erscheint begründet sich auf den Rat klinischer Experten und Epidemiologen, sowohl für multiple Therapievergleiche als auch für konventionelle paarweise Metaanalysen, zu verlassen“ (Zitat, übersetzt aus dem Englischen).

Da ein indirekter Vergleich nur valide sein kann, wenn er bei einer homogenen Studienlage angewendet wird, muss zunächst das Ausmaß der zulässigen Heterogenität definiert werden. Des Weiteren ist wichtig zu klären, welche Heterogenitätstests und welche weiteren Methoden zugelassen werden können, um den Zustand der Abwesenheit von bedeutsamer Heterogenität zu ermitteln. Dafür geeignet sein könnte der von der Cochrane Collaboration verwendete I^2 -Test (siehe Kapitel 5.2.1 „Heterogenität“). Eine Fragestellung, die hierfür geklärt werden müsste, ist, ob es eine Korrelation zwischen dem I^2 -Wert und der Validität einer Metaanalyse gibt. Hierzu könnte die Übereinstimmung der Ergebnisse der Metaanalyse mit denen einer großen Head-to-head-Studie (im Fall eines indirekten Vergleichs) geprüft werden. Gegebenenfalls lassen sich über diese Zusammenhänge Grenzwerte für tolerable statistische Heterogenität identifizieren. Eventuell könnte so auch die Vermutung von Higgins empirisch überprüft werden, dass bei einem I^2 -Wert $> 50\%$ bedeutsame Heterogenität vorliegen könnte¹²¹.

Methodik des MTC

Die Methoden der Heterogenitäts- und Inkonsistenzmessung in MTC sind ungleich komplizierter und ihre methodische Entwicklung ist noch nicht abgeschlossen¹⁵⁷. Ein weiterer Aspekt des MTC für den nach Sutton et al.²²⁹ noch Forschungsbedarf besteht, ist der bei Bayes'schen Modellierungen übliche Fit-of-model-Check. Dabei handelt es sich um einen Test, ob die Modellierung zu der tatsächlich vorliegenden Datenlage im Evidenznetzwerk passt²²². Bisher werden solche Fit-of-model-Checks noch nicht durchgängig in MTC verwendet, obwohl sie zu ihrer Qualitätssicherung ein geeignetes Werkzeug darstellen könnten²²⁹.

Verbindung eines indirekten Vergleichs mit einer Cross design-Synthese

Forschungsbedarf auf methodischer Ebene besteht ebenfalls, wenn in indirekten Vergleichen über RCT hinaus andere Studientypen eingeschlossen werden sollen. Hierbei kann auf Arbeiten aus der Literatur der Cross design-Synthese⁹⁰ und der Confidence profile-Methode⁹³⁻⁹⁵ zurückgegriffen werden. Die AHRQ stellt allerdings in ihrem aktuellen Methodenpapier fest: „Momentan sprechen wir uns gegen eine Kombination aus randomisierten und nicht-randomisierten Studien, die Daten aus Beobachtungsstudien enthalten, aus“. Statistische Methoden, die behandeln, wie Beobachtungsdaten eingeschlossen werden können, wurden bisher noch nicht ausreichend entwickelt. Forschungsbedarf besteht darin herauszufinden, ob die Kombination aus randomisierten und nicht-randomisierten Studien zweckmäßig ist, und wenn ja, unter welchen Bedingungen (Zitat, übersetzt aus dem Englischen)⁶.

Für den genannten Ansatz gibt es jedoch, unserer Kenntnis nach, bisher keine Anwendungsbeispiele, in denen auch indirekte Vergleiche durchgeführt wurden (mit Ausnahme der von dem Entwickler der Confidence profile-Methode selbst durchgeführten).

Ebenfalls bisher nur theoretisch betrachtet wurde die Möglichkeit die Ergebnisse randomisierter kontrollierter und nicht-randomisierter unkontrollierter Studien durch einen nicht-adjustierten indirekten Vergleich zusammenzufassen¹⁰⁹. Da dieser randomisations- und kontrollgruppenunabhängig ist, kann er sowohl Beobachtungsstudien als auch RCT einschließen.

Die Validität dieser potenziellen Erweiterungen der Methoden für indirekte Vergleiche bedürfen aber vor ihrer Verwendung zusätzlich noch einer kritischen Validitätsprüfung.

Berichtsqualität

Zum Abschluss wird noch Forschungsbedarf zur Verbesserung der Berichtsqualität von indirekten Vergleichen angemerkt. Dies betrifft vor allem die Darstellung der Heterogenitätserfassung in den Studiensätzen, mit denen ein indirekter Vergleich durchgeführt wird, da diese häufig unzureichend berichtet wurde. Die Entwicklung einer Checkliste für Autoren, die systematische Reviews mit indirekten Vergleichen durchführen, ähnlich dem QUORUM (Quality of reporting of metaanalyses)- oder CONSORT (Consolidated standards for reporting of trials)-Statement¹⁷⁴, könnte eine sinnvolle Maßnahme sein²²⁷. Diese könnte dann auch von HTA-Institutionen dazu eingesetzt werden, die Qualität von Publikationen indirekter Vergleiche zu beurteilen. Dem Leser von indirekten Vergleichen gäbe sie gleichzeitig einen Leitfaden für die kritische Bewertung der Publikation.

Einordnung in Evidenzhierarchie

Evidenzhierarchien, auf deren Basis Entscheidungen in der evidenzbasierten Medizin beruhen, sind umfangreich vorhanden. Die Frage ist, wo indirekte Vergleiche von Therapieverfahren eingeordnet werden müssten²²⁷. Ein erster, nicht mehr die aktuelle Erkenntnislage widerspiegelnder Vorschlag von McAlister et al.¹⁶⁷ von 1999, ordnet nur den adjustierten indirekten Vergleich ein. Er wird auf Level 2 zwischen u. a. einem Head-to-head-RCT (Level 1) und einem Vergleich auf Basis von nicht-randomisierten Studien (Level 4) angesiedelt. Eine Aktualisierung der Evidenzhierarchie auf Basis aktueller Evidenz zu der Validität indirekter Vergleiche sollte vorgenommen und dabei die Unterschiedlichkeit der verschiedenen Methoden berücksichtigt werden. Dabei könnte an den Überlegungen der GRADE Working Group³⁰ angeknüpft werden.

7 Empfehlungen und Schlussfolgerung

7.1 Einleitung - Methoden, die für die Durchführung indirekter Vergleiche zur Verfügung stehen

Es bleibt festzustellen, dass für indirekte Vergleiche eine Reihe von Methoden zur Verfügung steht, die sich hinsichtlich ihrer Möglichkeiten, die Evidenz verschiedener Studienlagen zusammenzufassen, unterscheiden. Die Auswahl einer geeigneten Methode hat sich der Anzahl der zu vergleichenden Therapieoptionen sowie an Qualität und Quantität der verfügbaren Studien zu orientieren.

Bei den Methoden für indirekte Vergleiche, die metaanalytisch vorgehen, lassen sich fast alle gefundenen Beispiele (95 %) den vier Methodengruppen - Nicht-adjustierter indirekter Vergleich, adjustierter indirekter Vergleich, Metaregression und MTC - zuordnen. Hervorzuheben ist, dass die adjustierten indirekten Vergleiche die mit Abstand am häufigsten eingesetzten Verfahren sind, während die anderen Verfahren deutlich seltener verwendet werden. Seit 2006 ist aber auch ein Anstieg der Publikationshäufigkeit bei den MTC zu verzeichnen.

7.2 Voraussetzungen für einen indirekten Vergleich

Die einzige Grundvoraussetzung, die für die Durchführung von indirekten Vergleichen erfüllt sein muss, ist die Homogenität der Studien, deren Ergebnisse zusammengefasst werden sollen. Im Gegensatz zu konventionellen Metaanalysen, die Homogenität innerhalb eines Studienpools zu einer Therapieoption erfordern, ist es für indirekte Vergleiche erforderlich, dass auch die Studien zu den unterschiedlichen zu vergleichenden Optionen homogen sind. In der Praxis sind diese Voraussetzungen nur selten gegeben¹⁰⁴.

7.3 Empfehlungen zu den verschiedenen indirekt vergleichenden Methoden

Empfehlung zu nicht-adjustierten indirekten Vergleichen

In der empirischen Überprüfung zeigt sich, dass vor allem die Ergebnisse nicht-adjustierter indirekter Vergleiche häufig von denen direkter Therapievergleiche abweichen. Dies bestätigt die Empfehlungen aller gefundener Methodenpapiere, die sich mit nicht-adjustierten indirekten Vergleichen auseinandersetzen und von ihrer Durchführung abraten^{8, 82, 104, 109, 217}.

Empfehlung zu adjustierten indirekten Vergleichen

Es ist anzunehmen, dass ein adjustierter indirekter Vergleich bei einer perfekt homogener Studienlage zu den gleichen Ergebnissen wie eine Metaanalyse von direkt vergleichenden Studien kommt^{55, 109}. In der empirischen Validitätsprüfung zeigt der adjustierte indirekte Vergleich, in einem breiten Spektrum von Beispielen mit unterschiedlich ausgeprägter Heterogenität eingesetzt, verhältnismäßig wenig Diskrepanz zu einem korrespondierenden direkten Vergleich. Sein Einsatz im Rahmen von systematischen Reviews, z. B. als Grundlage für ein HTA, scheint somit vertretbar, wenn zu einem Therapievergleich keine Head-to-head-Studien zur Verfügung stehen.

Empfehlung zu indirekten Vergleichen mit Metaregressionen

Liegt keine bedeutsame Heterogenität vor, besteht keine Notwendigkeit Metaregressionen zum Poolen von Studien einzusetzen, da auf die besser evaluierten adjustierten indirekten Vergleiche ausgewichen werden kann. Nur wenn erklärbare Heterogenität in dem zu poolenden Studiensatz vorliegt, kann eine Metaregression, die bezüglich Kovariaten adjustieren kann, nicht durch einen adjustierten indirekten Vergleich gleichwertig ersetzt werden. In solchen Fällen sollten sich Autoren bewusst sein, dass die Voraussetzung eines indirekten Vergleichs – einen homogene Studienlage – nicht vorliegt und es in Erwägung gezogen werden sollte, das Poolen der Daten nicht durchzuführen.

Aufgrund dieser theoretischen Betrachtungen und der zu geringen Anzahl von Beispielen in der Validitätsprüfung, kann zum jetzigen Zeitpunkt die Durchführung von indirekten Vergleichen in HTA mit Metaregressionen nicht abschließend beurteilt werden.

Empfehlung zu MTC

MTC, die vorliegende Head-to-head-Studien zu den Therapieoptionen von Interesse mit einschließen, wird eine ähnlich hohe Validität wie direkt vergleichenden Studien eingeräumt. Für MTC allerdings, die keine solchen Head-to-head-Studien mit einschließen, ist die Stichprobe trotz vielversprechender Ergebnisse in der Validitätsprüfung noch zu gering, um abschließende Aussagen über ihre Validität treffen zu können.

Da sich der MTC derzeit noch in Entwicklung befindet, kann das Verfahren momentan noch nicht für den Einsatz unter Routinebedingungen empfohlen werden. Es ist zu erwarten, dass weitere Anwendungsbeispiele mit der Möglichkeit der Validitätsprüfung anhand der Ergebnisse von Head-to-head-Studien den Stellenwert des MTC in absehbarer Zeit weiter klären werden.

Empfehlung zu sonstigen Methoden für indirekte Vergleiche

Für keine der unter dem Punkt „Sonstige“ zusammengefassten Methoden lässt sich anhand der Angaben in den Publikationen die methodische Vorgehensweise hinsichtlich der Aufrechterhaltung der Randomisation in den Einzelstudien eindeutig nachvollziehen. Zudem stehen nur sehr wenige Datensätze für die Validitätsprüfung zur Verfügung. Daher muss ihr Stellenwert für ein HTA derzeit als nicht beurteilbar eingeschätzt werden.

Empfehlungen zu Methoden ohne Metaanalysen

Theoretisch gelten für indirekte Vergleiche ohne Metaanalysen die gleichen Voraussetzungen wie für indirekte Vergleiche mit metaanalytischen Methoden: Die zusammenzufassenden Studien müssen homogen sein.

In dem vorliegenden HTA werden jedoch nicht ausreichend geeignete systematische Reviews gefunden, um die indirekt vergleichenden Methoden ohne Metaanalysen einer Validitätsprüfung zu unterziehen. Aus diesem Grund können diese Methoden nicht beurteilt werden.

7.4 Schlussfolgerung

Indirekte Vergleiche therapeutischer Interventionen können theoretisch bei Einhaltung ihrer Voraussetzung – Homogenität in den zusammenzufassenden Studien – zu den gleichen Ergebnissen kommen wie Metaanalysen von direkt vergleichenden Studien.

Diese Schlussfolgerung impliziert, dass bei Vorliegen von bedeutsamer Heterogenität in der zusammenzufassenden Studiengrundlage, auf die Durchführung von indirekten Vergleichen verzichtet werden sollte, da die Gefahr der Ergebnisverzerrung hoch ist.

In der Praxis ist eine perfekt homogene Studienlage aber selten anzutreffen. Zu den prioritären Forschungszielen gehört daher die Klärung, bei welchem Ausmaß von Heterogenität (und Inkonsistenz), es aus Validitätsgründen noch vertretbar ist, indirekte Vergleiche zur Beantwortung von Forschungsfragen heranzuziehen.

8 Literaturverzeichnis

1. N. N. Cross design synthesis: A new strategy for studying medical outcomes? *Lancet* 1992; 340(8825): 944-946.
2. N. N. International Conference on Harmonisation Homepage © 1998. ICH Harmonised Tripartite Guideline. Choice of control group and related issues in clinical trials. www.ich.org/LOB/media/MEDIA385.pdf (15.07.2008).
3. N. N. Medicare Services Advisory Committee Homepage © 2000. Samarium-153-lexidronam for bone pain due to skeletal metastases. [www.msac.gov.au/internet/msac/publishing.nsf/Content/1016-1/\\$FILE/msac1016.pdf](http://www.msac.gov.au/internet/msac/publishing.nsf/Content/1016-1/$FILE/msac1016.pdf) (20.02.2008).
4. N. N. Cochrane Collaboration Homepage © 2005. Cochrane Glossary. www.cochrane.org/resources/handbook/glossary.pdf (16.07.2008).
5. N. N. Gemeinsamer Bundesausschuss Homepage © 2006. BAnz. 2005 Nr. 242. Verfahrensordnung des Gemeinsamen Bundesausschusses. www.g-ba.de/downloads/62-492-83/VerfO_2006-04-18.pdf (18.06.2008).
6. N. N. Agency for Healthcare Research and Quality Homepage © 2007. Draft: Methods reference guide for effectiveness and comparative effectiveness reviews. Effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf (12.02.2009).
7. N. N. National Institute for Health and Clinical Excellence Homepage © 2007. The Guidelines Manual. www.nice.org.uk/niceMedia/pdf/GuidelinesManualAllChapters.pdf (12.02.2009).
8. N. N. Cochrane Database of Systematic Reviews Homepage © 2008. Handbook for systematic reviews of interventions. Version 5.0.0. www.cochrane-handbook.org (01.12.2008).
9. N. N. Deutsches Institut für medizinische Dokumentation und Information Homepage © 2008. HTA beim DIMDI. www.dimdi.de/static/de/hta/basisinfo_de.pdf (11.11.2008).
10. N. N. Deutsches Netzwerk Evidenzbasierte Medizin e.V. Homepage © 2008. EbM-Glossar. www.ebm-netzwerk.de/grundlagen/images/glossar_060920.pdf (18.06.2008).
11. N. N. Institut für Biometrie der Medizinischen Hochschule Hannover Homepage © 2008. Glossar. www.mh-hannover.de/institute/biometrie/JUMBO/bio/glossar.html#T (12.11.2008).
12. N. N. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen Homepage © 2008. Allgemeine Methoden Vol. 3.0. www.iqwig.de/download/IQWIG_Methoden_Version_3_0.pdf (12.02.2009).
13. N. N. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen Homepage © 2008. Vergleichende Nutzenbewertung verschiedener antihypertensiver Wirkstoffgruppen als Therapie der ersten Wahl bei Patienten mit essentieller Hypertonie. www.iqwig.de/download/A05-09_Vorbericht_Version_2_Antihypertensive_Wirkstoffgruppen_als_Therapie_der_ersten_Wahl.pdf (12.02.2009).
14. N. N. Leibniz-Rechenzentrum München Homepage © 2008. Internet-Lexikon der Methoden der empirischen Sozialforschung. www.lrz-muenchen.de/~wlm/ein_voll.htm (07.08.2008).
15. N. N. National Institute of Health and Clinical Excellence Homepage © 2008. Guide to the methods of technology appraisal. www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (07.10.2008).
16. N. N. US American National Library of Medicine Homepage © 2008. MeSH. www.ncbi.nlm.nih.gov/sites/entrez?db=mesh (29.11.2008).
17. Abou-Setta AM. Firm embryo transfer catheters for assisted reproduction: A systematic review and meta-analysis using direct and adjusted indirect comparisons. *Reprod Biomed Online* 2006; 12(2): 191-198.

18. Abou-Setta AM. What is the best site for embryo deposition? A systematic review and meta-analysis using direct and adjusted indirect comparisons. *Reprod Biomed Online* 2007; 14(5): 611-619.
19. Adelman JU, Belsey J. Meta-analysis of oral triptan therapy for migraine: Number needed to treat and relative cost to achieve relief within 2 hours. *J Manag Care Pharm* 2003; 9(1): 45-52.
20. Ades AE. A chain of evidence with mixed comparisons: Models for multi-parameter synthesis and consistency of evidence. *Stat Med* 2003; 22(19): 2995-3016.
21. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu GB. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006; 24(1): 1-19.
22. Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *J R Stat Soc Ser A Stat Soc* 2006; 169: 5-35.
23. Ades AE, Welton N, Lu G. University of Bristol Homepage © 2007. Introduction to Mixed Treatment Comparison.
www.bris.ac.uk/cobm/docs/intro%20to%20mtc.doc (14.08.2008).
24. Anderson S, Hauck WW. The transitivity of bioequivalence testing: Potential for drift. *Int J Clin Pharmacol Ther* 1996; 34(9): 369-374.
25. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ*. 1994; 308(6921): 81-106.
26. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy II: Maintenance of vascular graft or arterial patency by antiplatelet therapy. *BMJ*. 1994; 308(6922): 159-168.
27. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy III: Reduction in venous thrombosis and pulmonary embolism by antiplatelet prophylaxis among surgical and medical patients. *BMJ*. 1994; 308(6923): 235-246.
28. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002; 324: 71-86.
29. Armitage P, Berry G. The sampling error of a difference. *Statistical methods in medical research*. 2 ed. Oxford, 1987, 88-90.
30. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Williams HJW, Zaza S. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454): 1490-1494.
31. Ausejo M, Saenz A, Pham B, Kellner JD, Johnson DW, Moher D, Klassen TP. Glucocorticoids for croup. *Cochrane Database Syst Rev* 2000; (2): CD001955.
32. Baker SG, Kramer BS. The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol* 2002; 2:13.
33. Bakker A, van Balkom AJ, Spinhoven P, Blaauw BM, van Dyck R. Follow-up on the treatment of panic disorder with or without agoraphobia: A quantitative review. *J Nerv Ment Dis* 1998; 186(7): 414-419.
34. Balk EM, Lau J, Goudas LC, Jordan HS, Kupelnick B, Kim LU, Karas RH. Effects of statins on nonlipid serum markers associated with cardiovascular disease. *Annals of Internal Medicine* 2003; 139(8): 670-682.

35. Ballesteros J. Orphan comparisons and indirect meta-analysis: A case study on antidepressant efficacy in dysthymia comparing tricyclic antidepressants, selective serotonin reuptake inhibitors, and monoamine oxidase inhibitors by using general linear models. *J Clin Psychopharmacol* 2005; 25(2): 127-131.
36. Barden J, Edwards JE, Mcquay HJ, Wiffen PJ, Moore RA. Relative efficacy of oral analgesics after third molar extraction. *Br Dent J* 2004; 197(7): 407-411.
37. Begg CB, Pilote L. A model for incorporating historical controls into a meta-analysis. *Biometrics* 1991; 47(3): 899-906.
38. Bekkering GE, Abou-Setta AM, Kleijnen J. The application of quantitative methods for identifying and exploring the presence of bias in systematic reviews: PDE-5 inhibitors for erectile dysfunction. *Int J Impot Res* 2008; 20(3): 264-277.
39. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Stat Med* 1996; 15(5): 537-557.
40. Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *J Curr Clin Trials* 1994; Dokumentnummer 134.
41. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. *Stat Med* 2002; 21(3): 371-387.
42. Berner MM, Kriston L, Harms A. Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administering the International Index of Erectile Function in broad-spectrum populations. *Int J Impot Res* 2006; 18(3): 229-235.
43. Berry C, Norrie J, McMurray JJ. Ximelagatran compared with warfarin for the prevention of systemic embolism and stroke. An imputed placebo analysis. *Cardiovasc Drugs Ther* 2005; 19(2): 149-151.
44. Berry J. Are all aromatase inhibitors the same? A review of controlled clinical trials in breast cancer. *Clin Ther* 2005; 27(11): 1671-1684.
45. Bhandari M, Guyatt GH, Swiontkowski MF, Schemitsch EH. Treatment of open fractures of the shaft of the tibia - A systematic overview and meta-analysis. *J Bone Joint Surg* 2001; 83-B(1): 62-68.
46. Biondi-Zoccai GG, Agostoni P, Abbate A, Testa L, Burzotta F, Lotrionte M, Crea F, Biasucci LM, Vetovec GW, Colombo A. Adjusted indirect comparison of intracoronary drug-eluting stents: Evidence from a metaanalysis of randomized bare-metal-stent-controlled trials. *Int J Cardiol* 2005; 100(1): 119-123.
47. Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica MR, Walley T, Dickson R. Early thrombolysis for the treatment of acute myocardial infarction: A systematic review and economic evaluation. *Health Technol Assess* 2003; 7(15).
48. Boonen S, Lips P, Bouillon R, Bischoff-Ferrari HA, Vanderschueren D, Haentjens P. Need for additional calcium to reduce the risk of hip fracture with vitamin D supplementation: Evidence from a comparative metaanalysis of randomized controlled trials. *J Clin Endocrinol Metab* 2007; 92(4): 1415-1423.
49. Bortz J, *Statistik für Human- und Sozialwissenschaftler*, 6 ed. Berlin, 2005.
50. Bottomley JM, Auland ME, Morais J, Boyd G, Douglas WS. Cost-effectiveness of the two-compound formulation calcipotriol and betamethasone dipropionate compared with commonly used topical treatments in the management of moderately severe plaque psoriasis in Scotland. *Curr Med Res Opin* 2007; 23(8): 1887-1901.
51. Brophy JM, Joseph L. Medical decision making with incomplete evidence - Choosing a platelet glycoprotein IIb/IIIa receptor inhibitor for percutaneous coronary interventions. *Med Decis Making* 2005; 25(2): 222-228.
52. Brown H, Prescott R: *Applied mixed models in medicine*. Chichester, 1999.

53. Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, Rostom A, Symmons D. A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: A systematic review with economic modelling. *Health Technol Assess* 2006; 10(38).
54. Bucher HC, Griffith L, Guyatt GH, Opravil M. Meta-analysis of prophylactic treatments against *Pneumocystis carinii* pneumonia and toxoplasma encephalitis in HIV-infected patients. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; 15(2): 104-114.
55. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.
56. Buchner T, Dohner H, Ehninger G, Ganser A, Hasford J. Up-front randomization and common standard arm: A proposal for comparing AML treatment strategies between different studies. *Leuk Res* 2002; 26(12): 1073-1075.
57. Buchner T, Dohner H, Ehninger G, Ganser A, Niederwieser D, Hasford J. Cross-trial networking in AML: A step forward rather than corner cutting. *Leuk Res* 2004; 28(6): 649-650.
58. Buscemi N, Vandermeer B, Friesen C, Bialy L, Tubman M, Ospina M, Klassen TP, Witmans M. The efficacy and safety of drug treatments for chronic insomnia in adults: A meta-analysis of RCTs. *J Gen Intern Med* 2007; 22(9): 1335-1350.
59. Buttner M, Walder B, von Elm E, Tramer MR. Is low-dose haloperidol a useful antiemetic: A meta-analysis of published and unpublished randomized trials. *Anesthesiology* 2004; 101(6): 1454-1463.
60. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331(7521): 897-900.
61. Capstick T, Henry MT. Efficacy of thrombolytic agents in the treatment of pulmonary embolism. *Eur Res J* 2005; 26(5): 864-874.
62. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005; 365(9465): 1159-1162.
63. Chen JT, Wesley R, Shamburek RD, Pucino F, Csako G. Meta-analysis of natural therapies for hyperlipidemia: Plant sterols and stanols versus policosanols. *Pharmacotherapy* 2005; 25(2): 171-183.
64. Cheng L, Gulmezoglu AM, Ezcurra E, Van Look PF. Interventions for emergency contraception. *Cochrane Database Syst Rev* 2000; (2): CD001324.
65. Chiba N, De Gara CJ, Wilkinson JM, Hunt RH. Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: A meta-analysis. *Gastroenterol* 1997; 112(6): 1798-1810.
66. Chou R, Fu R, Huffman LH, Korthuis PT. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: Discrepancies between direct and indirect meta-analyses. *Lancet* 2006; 368(9546): 1503-1515.
67. Chou R, Fu RW. Validity of indirect comparisons in meta-analysis - Authors' reply. *Lancet* 2007; 369(9558): 271.
68. Clark W, Jobanputra P, Barton P, Burls A. The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: A systematic review and economic analysis. *Health Technol Assess* 2004; 8(18).
69. Clarke M. Individual patient data meta-analyses compared with meta-analyses based on aggregate data. *Cochrane Database Syst Rev* 2008.
70. Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? *BMJ*. 1994; 309(6960): 1007-1010.
71. Clauß C, Finze FR, Partzsch L: Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner - Grundlagen, 5 ed. Frankfurt, 2004.

72. Clegg A, Scott DA, Hewitson P, Sidhu M, Waugh N. Clinical and cost effectiveness of paclitaxel, docetaxel, gemcitabine, and vinorelbine in non-small cell lung cancer: A systematic review. *Thorax* 2002; 57(1): 20-28.
73. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; 10: 101-129.
74. Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K, Birtle A, Palmer S, Riemsma R. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. *Health Technol Assess* 2007.
75. Collins SL, Moore RA, McQuay HJ, Wiffen PJ, Edwards JE. Single dose oral ibuprofen and diclofenac for postoperative pain. *Cochrane Database Syst Rev* 2000; (2): CD001548.
76. Coomarasamy A, Knox EM, Gee H, Song F, Khan KS. Effectiveness of nifedipine versus atosiban for tocolysis in preterm labour: A meta-analysis with an indirect comparison of randomised trials. *BJOG* 2003; 110(12): 1045-1049.
77. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Arch Intern Med* 2006; 166(12): 1269-1275.
78. Coyle D, Hadj TA, Murphy G, Perras C, Skidmore B, Boucher M, Huserreau D. CADTH Homepage © 2006. Teriparatide and bisphosphonates for treatment of osteoporosis in women: A clinical and economic analysis (28.02.2008).
79. Cranney A, Wells G, Willan A, Griffith L, Zytaruk N, Robinson V, Black D, Adachi J, Shea B, Tugwell P, Guyatt G. Meta-analyses of therapies for postmenopausal osteoporosis II. Meta-analysis of alendronate for the treatment of postmenopausal women. *Endocr Rev* 2002; 23(4): 508-516.
80. Davey SG, Egger M. Meta-analyses of randomised controlled trials. *Lancet* 1997; 350(9085): 1182.
81. Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C. Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: A systematic review and economic model. *Health Technol Assess* 2006; 10(44).
82. Daya S. Adjusted indirect comparison of competing interventions. *Evid Based Obstet Gynecol* 2004; (6): 103-104.
83. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G and Altman DG (Eds). *Systematic reviews in health care: Meta-analysis in context*. 2 ed. London, 2001, pp 285-312.
84. Deeks JJ, Higgins JPT. *Cochrane Database Syst Rev Homepage* © 2007. Statistical algorithms in Review Manager 5. www.cochrane.org/resources/handbook/Statistical_Methods_in_RevMan5.pdf (18.02.2009).
85. Delaney BC, Innes MA, Deeks J, Wilson S, Cooner MK, Moayyedi P, Oakes R, Hobbs FD, Forman D. Initial management strategies for dyspepsia. *Cochrane Database Syst Rev* 2001; (3): CD001961.
86. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
87. Di Mario F, Battaglia G, Leandro G, Grasso G, Vianello F, Vigneri S. Short-term treatment of gastric ulcer. A meta-analytical evaluation of blind trials. *Dig Dis Sci* 1996; 41(6): 1108-1131.
88. Djulbegovic B. New treatments compared to established treatments in randomized trials. *Cochrane Database Syst Rev* 2008.

89. Dominici F, Parmigiani G, Wolpert RL, Hasselblad V. Meta-analysis of migraine headache treatments: Combining information from heterogeneous designs. *J Am Stat Assoc* 1999; (94): 16-28.
90. Droitcour J, Silberman G, Chelimsky E. Cross design synthesis: A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *Int J Technol Assess Health Care* 1993; 9(3): 440-449.
91. Eckert L, Falissard B. Using meta-regression in performing indirect comparisons: Comparing escitalopram with venlafaxine XR. *Curr Med Res Opin* 2006; 22(11): 2313-2321.
92. Eckert L, Lançon C. Duloxetine compared with fluoxetine and venlafaxine: Use of meta-regression analysis for indirect comparisons. *BMC Psychiatry* 2006; 6: 30.
93. Eddy DM. The confidence profile method: A Bayesian method for assessing health technologies. *Oper Res* 1989; 37(2): 210-228.
94. Eddy DM, Hasselblad V, Schachter R: *Meta-analysis by the confidence profile method*. Boston, 1992.
95. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence. The confidence profile method. *Int J Technol Assess Health Care* 1990; 6(1): 31-55.
96. Einarson TR, Kulin NA, Tingey D, Iskedjian M. Meta-analysis of the effect of latanoprost and brimonidine on intraocular pressure in the treatment of glaucoma. *Clin Ther* 2000; 22(12): 1502-1515.
97. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: A network meta-analysis. *Lancet* 2007; 369(9557): 201-207.
98. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. *Stat Med* 2000; 19(13): 1707-1728.
99. Farre M, Mas A, Torrens M, Moreno V, Cami J. Retention rate and illicit opioid use during methadone maintenance interventions: A meta-analysis. *Drug Alcohol Depend* 2002; 65(3): 283-290.
100. Ferrari MD, Goadsby PJ, Roon K, I, Lipton RB. Triptans (serotonin, 5-HT_{1B/1D} agonists) in migraine: Detailed results and methods of a meta-analysis of 53 trials. *Cephalalgia* 2002; 22(8): 633-658.
101. Fisher LD, Gent M, Buller HR. Active-control trials: How would a new agent compare with placebo? A method illustrated clopidogrel, aspirin, and placebo. *Am Heart J* 2001; 141(1): 26-32.
102. Gamble CL, Williamson PR, Marson AG. Lamotrigine versus carbamazepine monotherapy for epilepsy. *Cochrane Database Syst Rev* 2006; (1): CD001031.
103. Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. The comparative efficacy and safety of biologics for the treatment of rheumatoid arthritis: A systematic review and metaanalysis. *J Rheumatol* 2006; 33(12): 2398-2408.
104. Gartlehner G, Moore CG. Direct versus indirect comparisons: A summary of the evidence. *Int J Technol Assess Health Care* 2008; 24(2): 170-177.
105. Gavaghan DJ, Moore RA, McQuay HJ. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain* 2000; 85(3): 415-424.
106. Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: Clinical experience using a structured follow up programme in southern Sweden. *Ann Rheum Dis* 2002; 61(9): 793-798.
107. Geddes J, Freemantle N, Harrison P, Bebbington P. Atypical antipsychotics in the treatment of schizophrenia: Systematic overview and meta-regression analysis. *BMJ* 2000; 321: 1371-1376.

108. Gilks WR, Richardson S, Spiegelhalter DJ: Markov chain monte carlo in practice. London, 1996.
109. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood A. Indirect comparisons of competing interventions. *Health Technol Assess* 2005; 9(26).
110. Gleser LJ, Olkin I. Meta-analysis for 2 x 2 tables with multiple treatment groups. In: Stangl DK and Berry DA (Eds). *Meta-analysis in medicine and health policy*. New York, 2001, pp 179-189.
111. Goadsby PJ. The scientific basis of medication choice in symptomatic migraine treatment. *Can J Neurol Sci* 1999; 26(Suppl 3): 20-26.
112. Golfopoulos V, Salanti G, Pavlidis N, Ioannidis JPA. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: A meta-analysis. *Lancet Oncol* 2007; 8(10): 898-911.
113. Habib AS, El-Moalem HE, Gan TJ. The efficacy of the 5-HT₃ receptor antagonists combined with droperidol for PONV prophylaxis is similar to their combination with dexamethasone: A meta-analysis of randomized controlled trials. *Can J Anaesth* 2004; 51(4): 311-319.
114. Handoll HH, Farrar MJ, McBirnie J, Tytherleigh-Strong G, Milne AA, Gillespie WJ. Heparin, low molecular weight heparin and physical methods for preventing deep vein thrombosis and pulmonary embolism following surgery for hip fractures. *Cochrane Database Syst Rev* 2002; (4): CD000305.
115. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323(7308): 334-336.
116. Hasselblad V. Meta-analysis of multitreatment studies. *Med Decis Making* 1998; 18(1): 37-43.
117. Hasselblad V, Kong DF, Hasselblad V, Kong DF. Statistical methods for comparison to placebo in active-control trials. *Drug Inform J* 2001; 35: 435-449.
118. Hasselblad V, McCrory DC. Meta-analytic tools for medical decision making: A practical guide. *Med Decis Making* 1995; 15(1): 81-96.
119. Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. *Drug Inform J* 1999; 33: 109-118.
120. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21(11): 1539-1558.
121. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
122. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15(24): 2733-2749.
123. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001; 20(15): 2219-2241.
124. Hills RK, Richards SM, Wheatley K. Corner cutting compromises clinical trials: The inherent problems with up-front randomisation and a common standard arm. *Leuk Res* 2003; 27(12): 1071-1073.
125. Hind D, Calvert N, McWilliams R, Davidson A, Paisley S, Beverley C, Thomas S. Ultrasonic locating devices for central venous cannulation: Meta-analysis. *BMJ* 2003; 327(7411): 361.
126. Hirotsu C, Yamada Y. Estimating odds ratios through the connected comparative experiments. *Communication in Statistics - Theory and Methods* 1999; 28: 905-929.

127. Hochberg MC, Tracy JK, Hawkins-Holt M, Flores RH. Comparison of the efficacy of the tumour necrosis factor alpha blocking agents adalimumab, etanercept, and infliximab when added to methotrexate in patients with active rheumatoid arthritis. *Ann Rheum Dis* 2003; 62: 13-16.
128. Hofmann D, Hecker M, Volp A. Efficacy of dry extract of ivy leaves in children with bronchial asthma: A review of randomized controlled trials. *Phytomedicine* 2003; 10(2-3): 213-220.
129. Horn J, Limburg M. Calcium antagonists for acute ischemic stroke. *Cochrane Database Syst Rev* 2000; (4): CD001928.
130. Horn J, Limburg M. Calcium antagonists for ischemic stroke: A systematic review. *Stroke* 2001; 32(2): 570-576.
131. Ijsselmuiden AJ, Tangelder GJ, Cotton JM, Vaijifdar B, Kiemeneij F, Slagboom T, Wieken R, Serruys PW, Laarman GJ. Direct coronary stenting compared with stenting after predilatation is feasible, safe, and more cost-effective in selected patients: Evidence to date indicating similar late outcomes. *Int J Cardiovasc Intervent* 2003; 5(3): 143-150.
132. Indolfi C, Pavia M, Angelillo IF. Drug-eluting stents versus bare metal stents in percutaneous coronary interventions (a meta-analysis). *Am J Cardiol* 2005; 95(10): 1146-1152.
133. Jansen JP. Self-monitoring of glucose in type 2 diabetes mellitus: A Bayesian meta-analysis of direct and indirect comparisons. *Curr Med Res Opin* 2006; 22(4): 671-681.
134. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: An introduction to mixed treatment comparisons. *Value Health* 2008; 11(5): 956-964.
135. Jones G, Halbert J, Crotty M, Shanahan EM, Batterham M, Ahern M. The effect of treatment on radiological progression in rheumatoid arthritis: A systematic review of randomized placebo-controlled trials. *Rheumatology* 2003; 42(1): 6-13.
136. Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, Sudlow C, Henderson R, Hawkins N, Riemsma R. Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: A systematic review and economic evaluation. *Health Technol Assess* 2004; 8(38): 1-210.
137. Jung C, Engelmann E, Borner K, Offermann G. Preemptive oral ganciclovir therapy versus prophylaxis to prevent symptomatic cytomegalovirus infection after kidney transplantation. *Transplant Proc* 2001; 33(7-8): 3621-3623.
138. Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclooxygenase-2-inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis: Meta-analysis of randomised trials. *BMJ* 2006; 332: 1302.
139. Khan KS, ter Ried G, Glanville J, Sowden AJ, Kleijnen J. University of York Homepage © 2001. CRD Report 4.
www.york.ac.uk/inst/crd/report4.htm (20.04.2008).
140. Kleijnen J, de Craen AJ, Van Everdingen J, Krol L. Placebo effect in double-blind clinical trials: A review of interactions with medications. *Lancet* 1994; 344(8933): 1347-1349.
141. Kristensen LE, Christensen R, Bliddal H, Geborek P, Danneskiold-Samsøe B, Saxne T. The number needed to treat for adalimumab, etanercept, and infliximab based on ACR50 response in three randomized controlled trials on established rheumatoid arthritis: A systematic literature review. *Scand J Rheumatol* 2007; 36(6): 411-417.
142. Kyrgiou M, Salanti G, Pavlidis N, Paraskevaidis E, Ioannidis JP. Survival benefits with diverse chemotherapy regimens for ovarian cancer: Meta-analysis of multiple treatments. *J Natl Cancer Inst* 2006; 98(22): 1655-1663.
143. Lam SKH, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. *BMJ* 2007; 335(7626): 925-928.

144. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002; 55(1): 86-94.
145. Lancaster T, Stead LF. Physician advice for smoking cessation. *Cochrane Database Syst Rev* 2008; 1: CD000165.
146. Lange B, Bachert C, Lange B, Bachert C. Die topische intranasale Therapie der allergischen Rhinitis: Eine Wirksamkeits- und Kostenanalyse. *Allergologie* 2003; 26(5):177-195.
147. Law MR, Wald NJ, Morris JK, Jordan RE. Value of low dose combination treatment with blood pressure lowering drugs: Analysis of 354 randomised trials. *BMJ* 2003; 326(7404): 1427-1431.
148. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: Systematic review and meta-analysis. *BMJ* 2003; 326: 1423-1427.
149. Lee C, Hunsche E, Balshaw R, Kong SX, Schnitzer TJ. Need for common internal controls when assessing the relative efficacy of pharmacologic agents using a meta-analytic approach: Case study of cyclooxygenase 2-selective inhibitors for the treatment of osteoarthritis. *Arthritis Rheum* 2005; 53(4): 510-518.
150. Lee YH, Ji JD, Song GG. Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk. *Rheumatol Int* 2007; 27(5): 477-482.
151. Lee YH, Woo JH, Rho YH, Choi SJ, Ji JD, Song GG. Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis. *Rheumatol Int* 2008; 28(6): 553-559.
152. Leucht S, Pitschel-Walz G, Engel RR, Kissling W. Amisulpride, an unusual 'atypical' antipsychotic: A meta-analysis of randomized controlled trials. *Am J Psychiatry* 2002; 159(2): 180-190.
153. Li Z, Begg CB. Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *J Am Stat Assoc* 1994; 89: 1523-1527.
154. Liberati A, Buzzetti R, Grilli R, Magrini N, Minozzi S. Which guidelines can we trust? Assessing strength of evidence behind recommendations for clinical practice. *West J Med* 2001; 174(4): 262-265.
155. Lim E. *BMJ Homepage* © 2008. First principles or evidence based critique? bmj.bmjournals.com/cgi/eletters/327/7427/1309#46233 (13.04.2008).
156. Lim E, Ali Z, Ali A, Routledge T, Edmonds L, Altman DG, Large S. Indirect comparison meta-analysis of aspirin therapy after coronary surgery. *BMJ* 2003; 327(7427): 1309-1311.
157. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.
158. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* 2007; 26(20): 3681-3699.
159. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; 101(474): 447-459.
160. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002; 21(16): 2313-2324.
161. Mainland D: The treatment of clinical and laboratory data. Edinburgh, 1938.
162. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; 22(4): 719-748.
163. Marshall JK, Irvine EJ. Rectal corticosteroids versus alternative treatments in ulcerative colitis: A meta-analysis. *Gut* 1997; 40(6): 775-781.

164. Marson AG, Williamson PR, Hutton JL, Clough HE, Chadwick DW. Carbamazepine versus valproate monotherapy for epilepsy. *Cochrane Database Syst Rev* 2000; 3: CD001030.
165. Mason L, Moore RA, Edwards JE, Derry S, McQuay HJ. Topical NSAIDs for acute pain: A meta-analysis. *BMC Fam Pract* 2004; 5: 10.
166. Massel D. Primary angioplasty in acute myocardial infarction: Hypothetical estimate of superiority over aspirin or untreated controls. *Am J Med* 2005; 118(2): 113-122.
167. McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' guides to the medical literature - XIX. Applying clinical trial results - B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999; 282(14): 1371-1377.
168. McAlister FA, Stewart S, Ferrua S, McMurray JJ. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: A systematic review of randomized trials. *J Am Coll Cardiol* 2004; 44(4): 810-819.
169. McIntosh HM, Olliaro P. Artemisinin derivatives for treating uncomplicated malaria. *Cochrane Database Syst Rev* 2000; 2: CD000256.
170. McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, Hill RA, Jones A, Mujica MR, Walley T. Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: A systematic review and economic evaluation. *Health Technol Assess* 2007; 11(28): 1-158.
171. Messerli FH, Grossman E, Goldbourt U. Are beta-blockers efficacious as first-line therapy for hypertension in the elderly? A systematic review. *JAMA* 1998; 279(23): 1903-1907.
172. Mitte K, Noack P, Steil R, Hautzinger M. A meta-analytic review of the efficacy of drug treatment in generalized anxiety disorder. *J Clin Psychopharmacol* 2005; 25(2): 141-150.
173. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Stat Med* 2006; 25(24): 4321-4333.
174. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM Statement. *Onkologie* 2000; 23(6): 597-602.
175. Moore RA, Collins S, Carroll D, McQuay HJ. Paracetamol with and without codeine in acute pain: A quantitative systematic review. *Pain* 1997; 70(2-3): 193-201.
176. Moore RA, Derry S, McQuay HJ. Indirect comparison of interventions using published randomised trials: Systematic review of PDE-5 inhibitors for erectile dysfunction. *BMC Urology* 2005; 5: 18.
177. Morris JA, Gardner MJ. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *BMJ* 1988; 296(6632): 1313-1316.
178. Morton SC, Adams JL, Suttrop MJ, Shekelle PG. AHRQ Homepage © 2004. Meta-regression approaches: What, Why, When and How? www.ahrq.gov (20.02.2008); AHRQ Publication No. 04-0033 (Technical Review 8).
179. Mudge MA, Davey PJ, Coleman KA, Montgomery W, Croker VS, Mullen K, Castle DJ. A comparison of olanzapine versus risperidone for the treatment of schizophrenia: A meta-analysis of randomised clinical trials. *Int J Psychiatry Clin Pract* 2005; 9(1): 3-15.
180. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* 2007; 26(6): 1237-1254.
181. Otoul C, Arrigo C, van Rijckevorsel K, French JA. Meta-analysis and indirect comparisons of levetiracetam with other second-generation antiepileptic drugs in partial epilepsy. *Clin Neuropharmacol* 2005; 28(2): 72-78.
182. Otto MW, Tuby KS, Gould RA, McLean RY, Pollack MH. An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. *Am J Psychiatry* 2001; 158(12): 1989-1992.

183. Packer M, Antonopoulos GV, Berlin JA, Chittams J, Konstam MA, Udelson JE. Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: Results of a meta-analysis. *Am Heart J* 2001; 141(6): 899-907.
184. Peterson K, McDonagh MS, Fu R. Comparative benefits and harms of competing medications for adults with attention-deficit hyperactivity disorder: A systematic review and indirect comparison meta-analysis. *Psychopharmacology (Berl)* 2007.
185. Phillips A. Trial and error: Cross-trial comparisons of antiretroviral regimens. *AIDS* 2003; 17(4): 619-623.
186. Po AL, Zhang WY. Systematic overview of co-proxamol to assess analgesic effects of addition of dextropropoxyphene to paracetamol. *BMJ* 1997; 315(7122): 1565-1571.
187. Po AL, Zhang WY. Analgesic efficacy of ibuprofen alone and in combination with codeine or caffeine in post-surgical pain: A meta-analysis. *Eur J Clin Pharmacol* 1998; 53(5): 303-311.
188. Poynard T, Leroy V, Cohard M, Thevenot T, Mathurin P, Opolon P, Zarski JP. Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: Effects of dose and duration. *Hepatology*. 1996; 24(4): 778-789.
189. Psaty BN, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, Weiss NS. Health outcomes associated with various antihypertensive therapies used as first-line agents - A network meta-analysis. *JAMA* 2003; 289(19): 2534-2544.
190. Quan A, Chavanu K, Merkel J. A review of the efficacy of fixed-dose combinations olmesartan medoxomil/hydrochlorothiazide and amlodipine besylate/benazepril in factorial design studies. *Am J Cardiovasc Drugs* 2006; 6(2): 103-113.
191. Raghunathan TE. Pooling controls from different studies. *Stat Med* 1991; 10(9): 1417-1426.
192. Rice VH, Stead LF. Nursing interventions for smoking cessation. *Cochrane Database Syst Rev* 2000; 2: CD001188.
193. Richardson J, Smith JE, McCall G, Pilkington K. Hypnosis for procedure-related pain and distress in pediatric cancer patients: A systematic review of effectiveness and methodology related to hypnosis interventions. *Database of Abstracts of Reviews of Effectiveness* 2007; 4: DA20060861.
194. Richy F, Dukas L, Schacht E. Differential effects of D-hormone analogs and native vitamin D on the risk of falls: A comparative meta-analysis. *Calcif Tissue Int* 2008; 82(2): 102-107.
195. Richy F, Schacht E, Bruyere O, Ethgen O, Gourlay M, Reginster JY. Vitamin D analogs versus native vitamin D in preventing bone loss and osteoporosis-related fractures: A comparative meta-analysis. *Calcif Tissue Int* 2005; 76(3): 176-186.
196. Robenshtok E, Gafter-Gvili A, Goldberg E, Weinberger M, Yeshurun M, Leibovici L, Paul M. Antifungal prophylaxis in cancer patients after chemotherapy or hematopoietic stem-cell transplantation: Systematic review and meta-analysis. *J Clin Oncol* 2007; 25(34): 5471-5489.
197. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. *Am J Epidemiol* 1986; 124(5): 719-723.
198. Robinson JG, Smith B, Maheshwari N, Schrott H. Pleiotropic effects of statins: Benefit beyond cholesterol reduction? A meta-regression analysis. *J Am Coll Cardiol* 2005; 46(10): 1855-1862.
199. Roddy E, Zhang W, Doherty M. Aerobic walking or strengthening exercise for osteoarthritis of the knee: A systematic review. *Ann Rheum Dis* 2005; 64(4): 544-548.
200. Roozen HG, De Waart R, Van den Brink W. Efficacy and tolerability of naltrexone in the treatment of alcohol dependence: Oral versus injectable delivery. *Eur Addict Res* 2007; 13(4): 201-206.

201. Rostom A, Wells G, Tugwell P, Welch V, Dube C, McGowan J. Prevention of chronic NSAID induced upper gastrointestinal toxicity. *Cochrane Database Syst Rev* 2000; 3: CD002296.
202. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?". *Lancet* 2005; 365(9453): 82-93.
203. Rouder JN, Lu J. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychon Bull Rev* 2005; 12(4): 573-604.
204. Salanti G, Higgins J, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17(3): 279-301
205. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Annals of Internal Medicine* 2008; 148(7): 544-553.
206. Sanchez-Ramos L, Kaunitz AM, Delke I. Labor induction with 25 microg versus 50 microg intravaginal misoprostol: A systematic review. *Obstet Gynecol* 2002; 99(1): 145-151.
207. Sauriol L, Laporta M, Edwardes MD, Deslandes M, Ricard N, Suissa S. Meta-analysis comparing newer antipsychotic drugs for the treatment of schizophrenia: Evaluating the indirect approach. *Clin Ther* 2001; 23(6): 942-956.
208. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998; 17(17): 1923-1942.
209. Senn S. Active control equivalence trials. *Statistical issues in drug development*. Chichester, 2002, pp 207-217.
210. Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, Brophy J, Rossignol M. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166(10): 1203-1209.
211. Siegel JP. Equivalence and non-inferiority trials. *Am Heart J* 2000; 139: 166-170.
212. Silagy C, Lancaster T, Stead L, Mant D, Fowler G. Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev* 2001; 3: CD000146.
213. Silagy C, Stead LF. Physician advice for smoking cessation. *Cochrane Database Syst Rev* 2001; 2: CD000165.
214. Small LN, Lau J, Snyderman DR. Preventing post-organ transplantation cytomegalovirus disease with ganciclovir: A meta-analysis comparing prophylactic and preemptive therapies. *Clin Infect Dis* 2006; 43(7): 869-880.
215. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: A comparative study. *Stat Med* 1995; 14(24): 2685-2699.
216. Song F, Altman DG, Glenny A, Eastwood AJ, Deeks JJ. Adjusted indirect comparison for estimating relative effects of competing healthcare interventions. *Cochrane Database Syst Rev* 2007; 4: MR000020.
217. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analyses. *BMJ* 2003; 326(7387): 472-475.
218. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000; 21(5): 488-497.
219. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol* 2008; 61(5): 455-463.
220. Soo S, Moayyedi P, Deeks J, Delaney B, Lewis M, Forman D. Psychological interventions for non-ulcer dyspepsia. *Cochrane Database Syst Rev* 2004; 3: CD002301.

221. Spiegelhalter DJ, Abrams KR, Myles JP. Indirect comparison studies. In: Senn S and Barnett V (Eds). Bayesian approaches to clinical trials and health-care evaluation. Chichester, 2004, pp 282-285.
222. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research. An introduction to bayesian methods in health technology assessment. *BMJ* 1999; 319(7208): 508-512.
223. Stettler C, Allemann S, Egger M, Windecker S, Meier B, Diem P. Efficacy of drug eluting stents in patients with and without diabetes mellitus: Indirect comparison of controlled trials. *Heart* 2006; 92(5): 650-657.
224. Stettler C, Wandel S, Allemann S, Kastrati A, Morice MC, Schömig A, Pfisterer ME, Stone GW, Leon MB, de Lezo JS, Goy JJ, Park SJ, Sabaté M, Suttrop MJ, Kelbaek H, Spaulding C, Menichelli M, Vermeersch P, Dirksen MT, Cervinka P, Petronio AS, Nordmann AJ, Diem P, Meier B, Zwahlen M, Reichenbach S, Trelle S, Windecker S, Jüni P. Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis. *Lancet* 2007; 370(9591): 937-948.
225. Stier W: Empirische Forschungsmethoden. Berlin, 1999.
226. Sutton A, Ades AE, Abrams K, Cooper NJ. National Institute of Health and Clinical Excellence Homepage © 2007. Briefing paper for methods review workshop on evidence synthesis (indirect and mixed treatment comparisons). www.nice.org.uk (07.03.2008).
227. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*. 2008; 26(9): 753-767.
228. Sutton AJ, Abrams KR, Jones DR, Sheldon TA: Methods fo meta-analysis in medical research. Chichester, 2000.
229. Sutton AJ, Higgins JP. Recent developments in meta-analysis. *Stat Med* 2008; 27(5): 625-650.
230. Swift JA, Conway P, Purdie DW. A cost-utility analysis of low-dose hormone replacement therapy in postmenopausal women with an intact uterus. *Curr Med Res Opin* 2005; 21(12): 2051-2061.
231. Testa L, van Gaal WJ, Biondi-Zoccai GG, Abbate A, Agostoni P, Bhindi R, Banning AP. Repeat thrombolysis or conservative therapy vs. rescue percutaneous coronary intervention for failed thrombolysis: Systematic review and meta-analysis. *QJM* 2008; 101(5): 387-395.
232. Testa L, Zoccai GB, Porto I, Trotta G, Agostoni P, Andreotti F, Crea F. Adjusted indirect meta-analysis of aspirin plus warfarin at international normalized ratios 2 to 3 versus aspirin plus clopidogrel after acute coronary syndromes. *Am J Cardiol* 2007; 99(12): 1637-1642.
233. The RECORD Trial Group. Oral vitamin D3 and calcium for secondary prevention of low-trauma fractures in elderly people (randomised evaluation of calcium or vitamin D, RECORD): A randomised placebo-controlled trial. *Lancet* 2005; 365(9471): 1621-1628.
234. Thijs V, Lemmens R, Fieuws S. Network meta-analysis: Simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. *Eur Heart J* 2008; 29(9): 1086-1092.
235. Thompson SG. Why and how sources of heterogeneity should be investigated. In: Egger M, Davey Smith G and Altman DG (Eds). *Systematic reviews in health care: Meta-analysis in context*. 2 ed. London, 2001, pp 157-175.
236. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21(11): 1559-1573.
237. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat Med* 1999; 18(20): 2693-2708.
238. Trampisch H, Windeler J, Ehle B: *Medizinische Statistik*, 2 ed. Berlin, 2000.

239. Trindade E, Menon D. Selective serotonin reuptake inhibitors (SSRIs) for major depression. Part I. Evaluation of the clinical literature. Canadian Coordinating Office for Health Technology Assessment; CCOHTA Report 1997; 3E: 45
240. Tudur Smith C, Marson AG, Chadwick DW, Williamson PR. Multiple treatment comparisons in epilepsy monotherapy trials. *Trials* 2007; 8(34): 1-10.
241. Tudur Smith C, Marson AG, Clough HE, Williamson PR. Carbamazepine versus phenytoin monotherapy for epilepsy. *Cochrane Database Syst Rev* 2002; (2): CD001911.
242. Tudur Smith C, Marson AG, Williamson PR. Phenytoin versus valproate monotherapy for partial onset seizures and generalized onset tonic-clonic seizures. *Cochrane Database Syst Rev* 2001; (4): CD001769.
243. Tudur Smith C, Marson AG, Williamson PR. Carbamazepine versus phenobarbitone monotherapy for epilepsy. *Cochrane Database Syst Rev* 2003; (1): CD001904.
244. Tudur Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 2005; 24(9): 1307-1319.
245. Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K. Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. DAHTA 2003.
246. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000; 19(24): 3417-3432.
247. Vale L, Grant A, McCormack K, Scott NW, EU Hernia Trialists Collaboration. Cost-effectiveness of alternative methods of surgical repair of inguinal hernia. *Int J Technol Assess Health Care* 2004; 20(2): 192-200.
248. Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC. A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease. *Database of Abstracts of Reviews of Effectiveness* 2004; (4): DA20028437.
249. Van der Heijden JF, Prins MH, Buller HR. For the initial treatment of venous thromboembolism: Are all low-molecular-weight heparin compounds the same? *Thromb Res* 2000; 100(2): 121-130.
250. Van der Valk R, Webers CA, Schouten JS, Zeegers MP, Hendrikse F, Prins MH. Intraocular pressure-lowering effects of all commonly used glaucoma drugs: A meta-analysis of randomized clinical trials. *Ophthalmology* 2005; 112(7): 1177-1185.
251. Van Dongen CJ, Vink R, Hutten BA, Buller HR, Prins MH. The incidence of recurrent venous thromboembolism after treatment with vitamin K antagonists in relation to time since first event: A meta-analysis. *Arch Intern Med* 2003; 163(11): 1285-1293.
252. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Stat Med* 2002; 21(4): 589-624.
253. Van Pinxteren B, Numans ME, Bonis PA, Lau J. Short-term treatment with proton pump inhibitors, H2-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopy negative reflux disease. *Cochrane Database Syst Rev* 2000; (2): CD002095.
254. Vandermeer BW, Buscemi N, Liang Y, Witmans M. Comparison of meta-analytic results of indirect, direct, and combined comparisons of drugs for chronic insomnia in adults: A case study. *Med Care* 2007; 45(10): 166-172.
255. Verma S, Trudeau M, Pritchard K, Oliver T, Breast Cancer Disease Site Group. Cancer Care Ontario Practice Guidelines Initiative 2003 Homepage © 2003. The role of the taxanes in the management of metastatic breast cancer. www.cancercare.on.ca/pdf/pebc1-3f.pdf (20.02.2008).

256. Vestergaard P, Jorgensen NR, Mosekilde L, Schwarz P. Effects of parathyroid hormone alone or in combination with antiresorptive therapy on bone mineral density and fracture risk - A meta-analysis. *Osteoporos Int* 2007; 18(1): 45-57.
257. Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. Circular instead of hierarchical: Methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006; 6:29.
258. Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A. Taxanes for the adjuvant treatment of early breast cancer: Systematic review and economic evaluation. *Health Technol Assess* 2007; 11(40): 1-144.
259. Wehren LE, Hosking D, Hochberg MC. Putting evidence-based medicine into clinical practice: Comparing anti-resorptive agents for the treatment of osteoporosis. *Curr Med Res Opin* 2004; 20(4): 525-531.
260. Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, Mason A, Golder S, O'Meara S, Sculpher M, Drummond M, Forbes C. Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: A systematic review and economic evaluation. *Health Technol Assess* 2005; 9(15).
261. Wilhelmus KR. The treatment of herpes simplex virus epithelial keratitis. *Trans Am Ophthalmol Soc* 2000; 98: 505-532.
262. Wilhelmus KR. Therapeutic interventions for herpes simplex virus epithelial keratitis. *Cochrane Database Syst Rev* 2008; (1): CD002898.
263. Wilson SH, Bell MR, Rihal CS, Bailey KR, Holmes DR, Berger PB. Infarct artery reocclusion after primary angioplasty, stent placement, and thrombolytic therapy for acute myocardial infarction. *Am Heart J* 2001; 141(5): 704-710.
264. Woolacott NF, Hawkins N, Mason A, Kainth A, Khadjesari Z, Vergel YB, Misso K, Light K, Chalmers R, Sculpher M, Riemsma R. Etanercept and efalizumab for the treatment of psoriasis: A systematic review. *Health Technol Assess* 2006; 10(46): 1-233.
265. Wu P, Wilson K, Dimoulas P, Mills EJ. Effectiveness of smoking cessation therapies: A systematic review and meta-analysis. *BMC Public Health* 2006; 6: 300.
266. Yabroff KR, Kerner JF, Mandelblatt JS. Effectiveness of interventions to improve follow-up after abnormal cervical cancer screening. *Prev Med* 2000; 31(4): 429-439.
267. Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chene G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: Indirect comparison of controlled trials. *BMJ* 2004; 328: 249-256
268. Zaremski DG, Nolan PE, Slack MK, Caruso AC. Treatment of resistant atrial fibrillation: A meta-analysis comparing amiodarone and flecainide. *Arch Intern Med* 1995; 155(17): 1885-1891.
269. Zhang WY, Po AL. Analgesic efficacy of paracetamol and its combination with codeine and caffeine in surgical pain - A meta-analysis. *J Clin Pharm Ther* 1996; 21(4): 261-282.
270. Zhang WY, Po AL. Do codeine and caffeine enhance the analgesic effect of aspirin: A systematic overview. *J Clin Pharm Ther* 1997; 22(2): 79-97.
271. Zhang WY, Po AL. Efficacy of minor analgesics in primary dysmenorrhoea: A systematic review. *Br J Obstet Gynaecol* 1998; 105(7): 780-789.
272. Zhou Z, Rahme E, Pilote L. Are statins created equal: Evidence from randomized trials of pravastatin, simvastatin and atorvastatin for cardiovascular disease prevention. *Am Heart J* 2006; 151(2): 273-281.

9 Anhang

9.1 Formeln für die Methodik der Validitätsprüfung

Kontinuierliche Daten

Berechnung der Diskrepanz:

$$\theta_{\text{Diskrepanz}} = \theta_{\text{Direkt}} - \theta_{\text{Indirekt}}$$

Berechnung des Standardfehlers des Effektschätzers des direkten Vergleichs aus seinen Konfidenzintervallgrenzen:

$$SE(\theta_{\text{Direkt}}) = \frac{(\text{ObereGrenze} - \theta_{\text{Direkt}}) / 1,96 + (\theta_{\text{Direkt}} - \text{UntereGrenze}) / 1,96}{2}$$

Der Standardfehler des Effektschätzers des indirekten Vergleichs wird nach einer entsprechenden Formel berechnet:

$$SE(\theta_{\text{Indirekt}}) = \frac{(\text{ObereGrenze} - \theta_{\text{Indirekt}}) / 1,96 + (\theta_{\text{Indirekt}} - \text{UntereGrenze}) / 1,96}{2}$$

Berechnung der Konfidenzintervallgrenzen für die Diskrepanz:

$$95\%CI(\theta_{\text{Diskrepanz}}) = \theta_{\text{Diskrepanz}} \pm 1,96 \cdot \sqrt{SE(\theta_{\text{Direkt}})^2 + SE(\theta_{\text{Indirekt}})^2}$$

Dichotome Daten

Berechnung der Diskrepanz:

$$\theta_{\text{Diskrepanz}} = \ln \theta_{\text{Direkt}} - \ln \theta_{\text{Indirekt}}$$

Dabei stellt $\theta_{\text{Diskrepanz}}$ auch eine logarithmierte Größe dar. Sie wird aber nicht weiter umgeformt, um eine bessere Vergleichbarkeit mit den kontinuierlichen Daten zu gewährleisten. Bei kontinuierlichen Daten ist der Therapieeffektunterschied „Null“, wenn sich die Therapien nicht in ihrer Wirksamkeit unterscheiden. Bei dichotomen Daten wäre der Therapieeffektunterschied dagegen Eins. Da der natürliche Logarithmus von „Eins“ gleich „Null“ ist, ist es für die bessere Vergleichbarkeit hilfreich, die Therapieeffektunterschiede der dichotomen Daten logarithmiert darzustellen.

Berechnung des Standardfehlers des Effektschätzers des direkten Vergleichs aus seinen Konfidenzintervallgrenzen:

$$SE(\ln \theta_{\text{Direkt}}) = \frac{(\ln \text{ObereGrenze} - \ln \theta_{\text{Direkt}}) / 1,96 + (\ln \theta_{\text{Direkt}} - \ln \text{UntereGrenze}) / 1,96}{2}$$

Der Standardfehler des Effektschätzers des indirekten Vergleichs wird nach einer entsprechenden Formel berechnet:

$$SE(\ln \theta_{\text{Indirekt}}) = \frac{(\ln \text{ObereGrenze} - \ln \theta_{\text{Indirekt}}) / 1,96 + (\ln \theta_{\text{Indirekt}} - \ln \text{UntereGrenze}) / 1,96}{2}$$

Berechnung der Konfidenzintervallgrenzen für die Diskrepanz:

$$95\%CI(\theta_{\text{Diskrepanz}}) = \theta_{\text{Diskrepanz}} \pm 1,96 \cdot \sqrt{SE(\ln \theta_{\text{Direkt}})^2 + SE(\ln \theta_{\text{Indirekt}})^2}$$

9.2 Kurze Einführung in die Bayes'sche Statistik

Die Bayes'sche Statistik fußt auf einer anderen Philosophie als die konventionellen Statistik mit den bekannten Signifikanztests und Konfidenzintervallen. Man kann den Bayes'schen Ansatz so verstehen, dass es ein formalisierter Lernprozess durch neue Experimente ist²²¹. Eine Bayes'sche Analyse nutzt das Bayes'sche Theorem, um eine Priorverteilung eines unbekanntem Effekts (z. B. einer OR) durch Verarbeitung von Ergebnissen einer oder mehrerer Studien in eine Posteriorverteilung zu überführen. Als Priorverteilung wird das Vorwissen, das man über den Effekt vor Durchführung der Studie bzw. die Metaanalyse hat, bezeichnet. Die Posteriorverteilung ist das Resultat der Bayes'schen Analyse, das den Wissensfortschritt durch die durchgeführte Studie bzw.

Metaanalyse im Vergleich zur Priorverteilung wiedergibt. Die Priorverteilung kann auf externer Evidenz (z. B. nicht-randomisierte Studien⁹⁵, Expertenmeinung²²² oder historische Metaanalysen¹²²) beruhen oder es wird angenommen, dass nichts über den Effekt bekannt ist (non-informativer Ansatz). Im Rahmen der Metaanalyse werden die Ergebnisse der Einzelstudien in einer Likelihood-Funktion zusammengefasst⁸. Die Priorverteilung repräsentiert somit eine Zusammenfassung der externen Evidenz mit Relevanz für die Metaanalyse und die Likelihood-Funktion die Evidenz der Studien²²². Durch die Multiplikation der Priorverteilung mit der Likelihood-Funktion wird die Posteriorverteilung generiert, aus der statistische Schlussfolgerungen abgeleitet werden können²²². Sie kann als Punktschätzer mit „Credibility interval“ (Vertrauensintervall), das das Bayes'sche Äquivalent zu einem Konfidenzintervall darstellt, berichtet werden⁴. Es gibt jedoch Unterschiede in der Interpretation von Konfidenz- und Vertrauensintervallen²²².

Eine große Debatte in der Bayes'schen Statistik ist die Wahl der Parameter für die Priorverteilung²²². Besonders kontrovers ist es, objektive Studiendaten mit subjektiven Annahmen zu verbinden⁸. Die Cochrane Collaboration empfiehlt daher den non-informativen Ansatz⁸, bei dem kein Vorwissen über den Therapieeffektunterschied und seine Varianz in die Priorverteilung mit einfließt. Es wird aber auch anerkannt, dass es Fälle gibt, in denen es sehr hilfreich ist, externe Evidenz in der Priorverteilung zu berücksichtigen. Die getroffenen Annahmen sollten dann jedoch in Sensitivitätsanalysen überprüft werden⁸. Dies ist wichtig, da die gleichen Daten zu unterschiedlichen Ergebnissen führen können, wenn unterschiedliche Priorverteilungen genutzt werden^{8, 122}.

9.3 Durchsuchte Datenbanken

Tabelle 8: Über das DIMDI recherchierte elektronische Datenbanken

Kürzel	Datenbankname	Datenbankhersteller
INAHTA	NHS-CRD-HTA (INAHTA); Health Technology Assessment (HTA) Database	NHS CRD 2008
DAHTA	DAHTA-Datenbank	Bundesministerium für Gesundheit
NHSEED	NHS Economic Evaluation Database	NHS EED 2003
CDAR94	NHS-CRD-DARE; Database of Abstracts of Reviews of Effectiveness	Cochrane Collaboration
CDSR93	Cochrane Library - CDSR; Cochrane Database of Systematic Reviews	Cochrane Collaboration
ME00	MEDLINE; MEDical Literature Analysis and Retrieval System OnLINE	NLM
EM00	EMBASE; Excerpta Medica DataBASE	2008 Elsevier B. V.
CB85	AMED; Allied and Complementary MEDicine Database	THE BRITISH LIBRARY 2008
BA00	BIOSIS Previews	The Thomson Corporation
MK77	MEDIKAT	ZBMED
CCTR93	Cochrane Library - Central; Cochrane Central Register of Controlled Trials	Cochrane Collaboration
GA03	gms; German Medical Science	gms
SM78	SOMED; SOzialMEDizin	LOEGD 2002
CV72	CAB Abstracts	CAB
II98	ISTPB + ISTEP/ISSHP; Index to Scientific and Technical Proceedings and Books/Index to Social Sciences and Humanities Proceedings	The Thomson Corporation
ED93	ETHMED; ETHik in der MEDizin	IDEM 2006
AZ72	GLOBAL Health	CAB
AR96	Deutsches Ärzteblatt	DAEB
EA08	EMBASE Alert	2008 Elsevier B. V.
IS00;	SciSearch; Current Content/SciSearch (Science Citation Index)	The Thomson Corporation
CC00	CCMed; Current Contents Medizin	ZBMED
IN00	Social SciSearch; Current Contents/Social SciSearch (Social Science Citation Index)	The Thomson Corporation

Fortsetzung Tabelle 8: Über das DIMDI recherchierte elektronische Datenbanken

KR03	Karger-Verlagsdatenbank	Karger-Verlag
KL97	Kluwer-Verlagsdatenbank	Kluwer Academic Publishers
SP97	Springer-Verlagsdatenbank	Springer-Verlag
SPPP	Springer-Verlagsdatenbank-PrePrint	Springer-Verlag
TV01	Thieme-Verlagsdatenbank	Thieme-Verlag

DIMDI = Deutsches Institut für Medizinische Dokumentation und Information. NLM = National library of medicine. ZBMED = Zentralbibliothek für Medizin.

Tabelle 9: Im ISI Web of Knowledge® enthaltene Datenbanken

SCI-EXPANDED	Science Citation Index Expanded
SSCI	Social Sciences Citation Index
A & HCI	Arts & Humanities Citation Index

Tabelle 10: HTA-Institutionen (Internetseiten für die Handsuche)

Kürzel, Land	Name	Internetadresse
AETMIS, Kanada	Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé	www.aetmis.gouv.qc.ca
AETS, Spanien	Agencia de Evaluación de Tecnologías Sanitarias	www.isciii.es/htdocs/en/investigacion/Agencia_quees.jsp
AETSA, Spanien	Andalusian Agency for Health Technology Assessment	www.juntadeandalucia.es/salud/aetsa
AHRQ, USA	Agency for Healthcare Research and Quality	www.ahrq.gov
AHTA, Australien	Adelaide Health Technology Assessment	www.public-health.adelaide.edu.au/consult/health_techn_assess.html
AHTAPol, Polen	Agency for Health Technology Assessment in Poland	www.aotm.gov.pl
ASERNIP-S, Australien	Australian Safety and Efficacy Register of New Interventional Procedures -Surgical	www.surgeons.org/asernip-s
AVALIA-T, Spanien	Galician Agency for Health Technology Assessment	avalia-t.sergas.es
CADTH, Kanada	Canadian Agency for Drugs and Technologies in Health	www.cadth.ca
CAHTA, Spanien	Catalan Agency for Health Technology Assessment and Research	www.aatrm.net
CEDIT, Frankreich	Comité d'Évaluation et de Diffusion des Innovations Technologiques	cedit.aphp.fr
CENETEC, Mexiko	Centro Nacional de Excelencia Tecnológica en Salud Reforma	www.cenetec.gob.mx
CMT, Schweden	Center for Medical Technology Assessment	www.cmt.liu.se
CRD, UK	Centre for Reviews and Dissemination	www.york.ac.uk/inst/crd/
CVZ, Niederlande	College voor Zorgverzekeringen	www.cvz.nl
DACEHTA, Dänemark	Danish Centre for Evaluation and Health Technology Assessment	www.dacehta.dk
DAHTA @DIMDI, Deutschland	German Agency for HTA at the German Institute for Medical Documentation and Information	www.dimdi.de
DECIT-CGATS, Brasilien	Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia	portal.saude.gov.br/portal/saude/area.cfm?id_area=1026
DSI, Dänemark	Danish Institute for Health Services Research	www.dsi.dk
FinOHTA, Finnland	Finnish Office for Health Care Technology Assessment	www.stakes.fi/finohta
GR, Niederlande	Gezondheidsraad	www.gr.nl
HAS, Frankreich	Haute Autorité de Santé	www.has-sante.fr

Fortsetzung Tabelle 10: HTA-Institutionen (Internetseiten für die Handsuche)

HSAC, Neuseeland	Health Services Assessment Collaboration	www.hsci.canterbury.ac.nz/hnac/about.shtml
HunHTA, Ungarn	Health Economics and Technology Assessment Research Centre	Keine Homepage verfügbar
IAHS, GB	Institute of Applied Health Sciences	Keine Homepage verfügbar
ICTAHC, Israel	Israel Center for Technology Assessment in Health Care	www.health.gov.il/english/pages_e/default.asp?pageid=28&parented=15&catid=13&maincat=2
IECS, Argentinien	Institute for Clinical Effectiveness and Health Policy	www.iecs.org.ar
IHE, Kanada	Institute of Health Economics	www.ihe.ca
IMSS, Mexiko	Mexican Institute of Social Security	www.imss.gob.mx/imss/ims_s_sitios/dpm/informacion/tecnologia/principal.htm
INAHTA, International	International network of agencies for Health Technology Assessment	www.inahta.org
IQWiG, Deutschland	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	www.iqwig.de
KCE, Belgien	Belgian Federal Health Care Knowledge Centre	kce.fgov.be
LBI of HTA, Österreich	Ludwig Boltzmann Institut für Health Technology Assessment	hta.lbg.ac.at
MAS, Kanada	Medical Advisory Secretariat	www.health.gov.on.ca/english/providers/program/mas/mas_about.html
MSAC, Australien	Medicare Services Advisory Committee	www.msac.gov.au
MTU-SFOPH, Schweiz	Medical Technology Unit - Swiss Federal Office of Public Health	www.snhta.ch
NCCHTA, GB	National Coordinating Centre for Health Technology Assessment	www.hta.ac.uk
NHS QIS, GB	Quality Improvement Scotland	www.nhshealthquality.org
NHSC, GB	National Horizon Scanning Centre	www.pcpoh.bham.ac.uk/publichealth/horizon
NOKC, Norwegen	Norwegian Knowledge Centre for Health Services	www.nokc.no
NZHTA, Neuseeland	New Zealand Health Technology Assessment	nzhta.chmeds.ac.nz
OSTEBA, Spanien	Basque Office for Health Technology Assessment	www.osanet.euskadi.net/osteaba/es
SBU, Schweden	Swedish Council on Technology Assessment in Health Care	www.sbu.se
UETS, Spanien	Unidad de evaluación Tecnologías Sanitarias	www.madrid.org
VATAP, USA	VA Technology Assessment Program	www.va.gov/vatap
VSMTVA, Lettland	Health Statistics and Medical Technologies State Agency	www.vsmstva.gov.lv
ZonMw, Niederlande	The Medical and Health Research Council of The Netherlands	www.zonmw.nl

HTA = Health Technology Assessment. GB = Großbritannien. USA = Vereinigte Staaten von Amerika.

Tabelle 11: Cochrane Collaboration (Internetseiten für die Handsuche)

1. Suchort	Methodology Review Group	www.cochrane.org/reviews/en/topics/33.html
2. Suchort	Training Resources	www.cochrane.org/resources/revpro.htm

9.4 Suchstrategien

9.4.1 Suchstrategien in digitalen Datenbanken

DIMDI-Datenbanken, Recherchezeitraum: 1999 bis 2008

Tabelle 12: Suchstrategie, DIMDI-Datenbanken, 20.02.2008

Schritt	Treffer	Suchterminus
1	38010257	INAHTA; DAHTA; NHSEED; CDAR94; CDSR93; ME00; EM00; CB85; BA00; MK77; CCTR93; GA03; SM78; CV72; I198; ED93; AZ72; AR96; EA08; IS00; CC00; IN00; KR03; KL97; SP97; SPPP; TV01
2	1430	(DIRECT AND INDIRECT AND COMPARISON?)/SAME SENT
3	66	(DIRECT AND ADJUSTED AND INDIRECT AND COMPARISON?)/SAME SENT
4	2787	(DIRECT AND INDIRECT AND ESTIMATE?)/SAME SENT
5	335	(DIRECT? AND RANDOMIZED AND COMPARISON?)/SAME SENT
6	878	INDIRECT COMPARISON?
7	87	ADJUSTED INDIRECT COMPARISON?
8	224	(MIXED AND TREATMENT AND COMPARISON?)/SAME SENT
9	286	ACTIVE-CONTROL-TRIAL#
10	87	COMPETING INTERVENTION#
11	5	HEAD# %%#HEAD
12	5293	2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11
13	4590	12 AND PY>=1999
14	4428	13 AND LA=(ENGL OR GERM)
15	2315	check duplicates: unique in s=14
16	5232	2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9
17	5232	S=16
18	3017	check duplicates: unique in s=17
19	561	18 AND RANDOMI %ED CONTROLLED TRIAL#
20	104	18 AND RANDOMI %ED CLINICAL TRIAL#
21	0	18 AND RANDOMISIERTE KONTROLLIERTE KLINISCHE STUDIE#
22	87	18 AND RANDOMISIERTE KONTROLLIERTE STUDIE#
23	1	18 AND RANDOMISIERTE KLINISCHE STUDIE#
24	81	18 AND RCT
25	147	18 AND RCT
26	15	18 AND RANDOM ALLOCATION?
27	106	18 AND RANDOM? ?, ALLOCAT?
28	2	18 AND RANDOMISIERT? ? VERSUCH?
29	94	18 AND RANDOMISIERT? ? STUDIE?
30	643	18 AND RANDOMI %ED? ? STUD?
31	821	18 AND RANDOMI %ED? ? TRIAL?
32	324	18 AND RANDOMI %ED? ? CLINICAL? ? TRIAL?
33	270	18 AND RANDOMI %ED? ? CLINICAL? ? STUD?
34	916	19 OR 20 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33
35	1386	COMPET? ## TECHNOLOG?
36	407	COMPET? ## INTERVENTION#
37	554	COMPET? ## THERAP?
38	11371	TREATMENT? ## ARM?
39	139126	TREATMENT? ## GROUP?
40	30258	RANDOMI? ## GROUP?
41	13256	RANDOMI? ## COMPARISON?

Schritt	Treffer	Suchterminus
42	105	THERAPEUTIC # # ARM#
43	0	4#LIMB# STUD?
44	1	FOUR LIMB# STUD?
45	295	TRIAL # # ARM
46	7784	TRIAL # # DESIGN
47	464	TRIAL # # ARMS
48	4	PREVENTIVE # # ARM
49	0	PREVENTATIVE # # ARM
50	5	MULTIPLE ARM STUDY
51	0	MULTIPLE ARMS STUDY
52	2	MULTIPLE ARM STUDIES
53	0	MULTIPLE ARMS STUDIES
54	149	MULTIPLE ARM#
55	193	MULTI ARM#
56	331	MULTI # # ARM#
57	298	MULTIPLE # # ARM#
58	2152	THREE ARM# OR 3#ARM#
59	277	THREE LIMB# OR 3#LIMB#
60	1208	FOUR ARM# OR 4#ARM#
61	1201	FOUR LIMB# OR 4#LIMB#
62	16147552	TRIAL# OR STUD### OR RANDOM?
63	2428	58 OR 59
64	2035	63 AND 62
65	2409	60 OR 61
66	1555	65 AND 62
67	201852	34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42 OR 44 OR 45 OR 46 OR 47 OR 48 OR 50 OR 52 OR 54 OR 55 OR 56 OR 57 OR 64 OR 66
68	844	15 AND 67
69	844	check duplicates: unique in s=68
70	119454	META-ANALYS## OR METAANALYS##
71	236	69 AND 70
72	608	69 NOT 71

Tabelle 13: Suchstrategie, ISI Web of Knowledge®, 28.2.2008

Schritt	Treffer	Suchterminus		
		Cited Author	Cited Work	Year
1	21	Glenny A	HTA	2005
2	106	Song F	BMJ	2003
3	88	Bucher H	J Clin Epi	1997
4	163	#1 OR #2 OR #3		

Am 30.05.2008 wurde die in Tabelle 13 geschilderte Suchstrategie wiederholt und lieferte insgesamt 23 weitere Treffer.

9.4.2 Suchstrategie der Handsuchen

Internetseiten der HTA-Institutionen

Es werden alle Internetpräsentationen der in der INAHTA zusammengeschlossenen HTA-Institutionen durchsucht und die Institutionen mit der Frage nach Methodenpapieren angeschrieben. Der Aufbau der Internetseiten der HTA-Institutionen ist zu heterogen, um eine einheitliche Suchstrategie zu berichten.

Cochrane Collaboration

1. Suchort: „Methodology Review Group“:

Alle Protokolle and Reviews werden in einer Titelselektion auf das Vorkommen der Schlagwörter „Comparison, Compare, Metaanalysis“ oder „Aggregation of data“ hin durchsucht und beim Vorkommen mindestens eines dieser Schlagworte deren Volltext gelesen. Protokolle, die bisher keine Ergebnisse publizieren, werden im Rahmen des Volltextscreenings ausgeschlossen.

2. Suchort: „Training Resources“:

Es wird nach Trainingsmaterialien gesucht, die potenziell Methodenbeschreibungen für die Durchführung von Metaanalysen enthalten könnten und deren Volltexte auf Anleitungen zur Durchführung von indirekten Vergleichen hin untersucht.

Zitationssuche Methodenpapiere

Die Zitationssuche wird im Kapitel 4: Statistical methods for indirect comparisons, Seite 19 bis 24 des Reviews von Glenny et al.¹⁰⁹ durchgeführt, indem alle Volltexte bestellt und auf Relevanz überprüft werden. Ergänzungen durch Zitationen einiger anderer wichtiger und aktueller Reviews und Methodenpapiere wurden auch mit aufgenommen.

9.5 Ergebnisse der Handsuchen

Tabelle 14: Ergebnisse der Handsuchen auf den Internetseiten der HTA-Institutionen

Kürzel, Land	Potenziell relevante Methodenpapiere	Potenziell relevant nach erster Durchsicht	Relevant nach zweiter Durchsicht
AETMIS, Kanada	-	-	-
AETS, Spanien	1*	1*	0
AETSA, Spanien	-	-	-
AHRQ, USA	Multiple	0	0
AHTA, Australien	-	-	-
AHTAPol, Polen*	-	-	-
ASERNIP-S, Australien	1	0	0
AVALIA-T, Spanien*	-	-	-
CADTH, Kanada	8	1	0
CAHTA, Spanien	1*	1*	0
CEDIT, Frankreich	-	-	-
CENETEC, Mexiko*	-	-	-
CMT, Schweden	-	-	-
CRD, GB	4	1	0
CVZ, Niederlande	-	-	-
DACEHTA, Dänemark	2	2	0
DAHTA @DIMDI, Deutschland	1	0	0
DECIT-CGATS, Brasilien*	-	-	-
DSI, Dänemark	2	0	0
FinOHTA, Finnland	4	0	0
GR, Niederlande	1	0	0
HAS, Frankreich	7*	1*	0
HSAC, Neuseeland**	-	-	-
HunHTA, Ungarn	1	0	0
IAHS, GB	-	-	-
ICTAHC, Israel*	-	-	-
IECS, Argentinien	-	-	-
IHE, Kanada	3	0	0
IMSS, Mexiko*	-	-	-
INAHTA, International	36****	1	0
IQWiG, Deutschland	2	2	0

Fortsetzung Tabelle 14: Ergebnisse der Handsuchen auf den Internetseiten der HTA-Institutionen

Kürzel, Land	Potenziell relevante Methodenpapiere	Potenziell relevant nach erster Durchsicht	Relevant nach zweiter Durchsicht
KCE, Belgien	3	0	0
LBI of HTA, Österreich	2	1	0
MAS, Kanada	-	-	-
MSAC, Australien	2	-	-
MTU-SFOPH, Schweiz	1	1	0***
NCCHTA, GB	61	1	0
NHS QIS, GB	-	-	-
NHSC, GB	-	-	-
NICE, GB	Multiple	3	1****
NOKC, Norwegen	1*	1*	0**
NZHTA, Neuseeland	-	-	-
OSTEBA, Spanien	6*	1*	0***
SBU, Schweden	1	1	0***
UETS, Spanien*	-	-	-
VATAP, USA	1	1	0
VSMTVA, Lettland	-	-	-
ZonMw, Niederlande	-	-	-
Welt-Gesundheits-Organisation (WHO)	1	1	0
Cochrane Collaboration	25	5	1

* Internetseite/Publikation nicht-englisch- oder deutschsprachig.

** Ausschluss, da Sprache nicht Deutsch, Englisch, Französisch oder Spanisch.

*** Ausschluss, da in den letzten zehn Jahren nicht überarbeitet.

**** Davon 35 auch von den genannten Mitgliedsorganisationen veröffentlicht.

***** Veröffentlichung steht unmittelbar bevor.

HTA = Health Technology Assessment. GB = Großbritannien. USA = Vereinigte Staaten von Amerika.

Tabelle 15: Ergebnisse der Handsuchen nach Methodenpapieren in der Cochrane Collaboration

Suchort	Potenziell relevant	Relevant nach Volltextscreening
Methodology Review Group	3	0
Training Resources	2	1

Tabelle 16: Details über die im Volltextscreening ein- bzw. ausgeschlossenen Publikationen bei der Suche in der Cochrane Collaboration

Erstautor, Jahr des letzten Updates	Titel	Ausschluss mit Grund	Nicht beschaffbar mit Grund
Song 2007 ²¹⁶	Adjusted indirect comparison for estimating relative effects of competing healthcare interventions	Ausschluss nach Grund 8	Protokoll; Review noch nicht publiziert
Clarke 2008 ⁶⁹	Individual patient data Metaanalyses compared with Metaanalyses based on aggregate data	Ausschluss nach Grund 8	Protokoll; Review noch nicht publiziert
Djulgovic 2008 ⁸⁸	New treatments compared to established treatments in randomized trials	Ausschluss nach Grund 8	Protokoll; Review noch nicht publiziert
Higgins 2008 ⁸	Handbook for systematic Reviews of Interventions Version 5.0.0	Einschluss	
Khan 2001 ¹³⁹	CRD Report 4 (2nd edition)	Ausschluss nach Grund 5	

Tabelle 17: Ergebnisse der Referenzsuche im Kapitel „Statistical methods for indirect comparisons“ des systematischen Reviews von Glenny et al.¹⁰⁹ (S. 18-25)

Erstautor, Jahr	Einschluss mit Publikationstyp	Ausschluss mit Grund	Nicht beschaffbar mit Grund
Bucher 1997 ⁵⁵	5		
Berkey 1996 ³⁹		5	
Dominici 1999 ⁸⁹	1		
Hasselblad 1998 ¹¹⁶	1		
Higgins 1996 ¹²²	5		
Hirotsu 1999 ¹²⁶	1		
Gleser 2001 ¹¹⁰	1		
Thompson 2001 ²³⁵		5	
Deeks 2001 ⁸³		5	
Thompson 2002 ²³⁶	1		
Anderson 1996 ²⁴		2	
Hauck 1999 ¹¹⁹		5	
Mainland 1938 ¹⁶¹		8	Buch
Armitage 1987 ²⁹		8	Buch
Eddy 1992 ⁹⁴	1		
Hasselblad 2001 ¹¹⁷		4	
Packer 2001 ¹⁸³	5		
Fisher 2001 ¹⁰¹		4	
Baker 2002 ³²		5	
Clarke 1994 ⁷⁰		4	
Higgins 2001 ¹²³		5	
Smith 2005 ²⁴⁴		5	
Brown 1999 ⁵²		8	Buch
Ades 2003 ²⁰	1		
Sutton 2000 ²²⁸		8	Buch
Thompson 1999 ²³⁷		5	
Lim 2003 ¹⁵⁶	4		
Lim 2004 ¹⁵⁵		2	
Spiegelhalter 2004 ²²¹	1		
Phillips 2003 ¹⁸⁵		5	
Senn 2002 ²⁰⁹		8	Buch
Siegel 2000 ²¹¹		5	
ICH Harmonised Tripartite Guideline 1998 ²		5	
Engels 2000 ⁹⁸		5	
Song 2000 ²¹⁸	1		
Lumley 2002 ¹⁶⁰	1		
Chan 2005 ⁶²		2	
Lambert 2002 ¹⁴⁴		5	
Berlin 2002 ⁴¹		5	
Turner 2000 ²⁴⁶	1		
Begg 1991 ³⁷		5	
Li 1994 ¹⁵³		5	
Raghunathan 1991 ¹⁹¹		5	
Büchner 2002 ⁵⁶		2	
Büchner 2004 ⁵⁷		5	
Hills 2003 ¹²⁴		2	

Insgesamt werden somit 15 der 46 zitierten Publikationen als relevante Veröffentlichungen über die Durchführung von indirekten Vergleichen befunden. Hinzu kommen noch fünf weitere Publikationen, die aus den Referenzen anderer Methodenpapiere oder Reviews stammen (siehe Tabelle 18).

Tabelle 18: Übersicht über die weiteren Methodenpapiere, die aus den Referenzen anderer Reviews oder bereits gefundener Methodenpapiere exzerpiert wurden.

Nummer	Autor, Jahr	Einschluss nach Publikationstyp
1	Eddy 1990 ⁹⁵	1
2	Wilby 2005 ²⁶⁰	4
3	Glenny 2005 ¹⁰⁹	1
4	Eddy 1989 ⁹³	1
5	Van Houwelingen ²⁵²	1

Zudem werden die Metaanalysen der 28 in Glennys Review eingeschlossenen Übersichtsarbeiten übernommen, mit denen Glenny et al.¹⁰⁹ sowohl einen direkten als auch indirekten Vergleich durchführen konnte. Damit wird die Zeitspanne der Suche nach Typ-5-Publikationen ab 1999 um die Studienergebnisse der Jahre 1966 bis 1999 ergänzt (Überschneidungszeitraum: Januar 1999 bis Februar 2001). Drei Übersichtsarbeiten aus diesem Überschneidungszeitraum werden in der eigenen Literaturrecherche gefunden (Packer et al.¹⁸³, Sauriol et al.²⁰⁷ und Chiba et al.⁶⁵), sodass aus Glennys Literaturrecherche letztendlich 25 systematische Reviews ergänzt werden.

Tabelle 19: Systematische Übersichtsarbeiten, denen Glenny et al.¹⁰⁹ Metaanalysen entnahmen, um mit ihnen sowohl direkte als auch indirekte Vergleiche durchzuführen

Nummer	Erstgenannter Autor
1	Antiplatelet Trialists' Collaboration 1994 ²⁶
2	Antiplatelet Trialists' Collaboration 1994 ²⁵
3	Antiplatelet Trialists' Collaboration 1994 ²⁷
4	Ausejo 2000 ³¹
5	Bucher 1997 ⁵⁵
6	Cheng 2000 ⁶⁴
7	Chiba 1997 ⁶⁵
8	Collins 2000 ⁷⁵
9	Delaney 2001 ⁸⁵
10	Di Mario 1996 ⁸⁷
11	Handoll 2002 ¹¹⁴
12	Higgins 1996 ¹²²
13	Horn 2001 ¹³⁰
14	Marshall 1997 ¹⁶³
15	McIntosh 2000 ¹⁶⁹
16	Moore 1997 ¹⁷⁵
17	Packer 2001 ¹⁸³
18	Po 1997 ¹⁸⁶
19	Poynard 1996 ¹⁸⁸
20	Rostom 2000 ²⁰¹
21	Sauriol 2001 ²⁰⁷
22	Silagy 2001 ²¹³
23	Silagy 2001 ²¹²
24	Soo 2004 ²²⁰
25	Trindade 1997 ²³⁹
26	Van Pinxteren 2000 ²⁵³
27	Zhang 1996 ²⁶⁹
28	Zhang 1998 ²⁷¹

9.6 Nach Durchsicht im Volltext ausgeschlossene Literatur mit Ausschlussgrund

Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

Nr.	Bibliographie
1. Ausschlussgrund: Einzelstudien	
1	Cutrer EM, Goadsby PJ, Ferrari MD, Lipton RB, Dodick DW, McCrory D, Williams P. Priorities for triptan treatment attributes and the implications for selecting an oral triptan for acute migraine: A study of US primary care physicians (the TRIPSTAR project). Clin Ther 2004; 26(9): 1533-1545.
2	Halpern MT, Covert DW, Robin AL. Projected impact of travoprost versus both timolol and latanoprost on visual field deficit progression and costs among black glaucoma subjects. Trans Am Ophthal Soc 2002; 100: 109-117.
3	Noyes K, Dick AW, Holloway RG, The Parkinson Study Group. Pramipexole and levodopa in early Parkinson's disease: Dynamic changes in cost-effectiveness. Pharmacoeconomics 2005; 23(12): 1257-1270.
4	Noyes K, Dick AW, Holloway RG. Pramipexole vs. levodopa as initial treatment for Parkinson's disease: A randomized clinical-economic trial. Med Decis Making 2004; 24: 472-485.
2. Ausschlussgrund: Übersichtsarbeiten, Editorials und Kommentare, bei denen es sich nicht um systematische Übersichtsarbeiten handelt	
1	N. N. Insulin in type 2 diabetes: A useful alternative despite limited assessment based on surrogate endpoints. Prescrire international 2005; 14(79): 187-193.
2	Anderson S, Hauck WW. The transitivity of bioequivalence testing: Potential for drift. Int J Clin Pharmacol Ther 1996; 34(9): 369-374.
3	Anstey AV, Kragballe K. Retrospective assessment of PASI 50 and PASI 75 attainment with a calcipotriol/betamethasone dipropionate ointment. Int J Dermatol 2006; 45(8): 970-975.
4	Bonuccelli U. Comparing dopamine agonists in Parkinson's disease. Curr Opin Neurol 2003; 16(Suppl 1): 13-19.
5	Buchner T, Dohner H, Ehninger G, Ganser A, Hasford J. Up-front randomization and common standard arm: A proposal for comparing AML treatment strategies between different studies. Leuk Res 2002; 26(12): 1073-1075.
6	Buyse M, Burzykowski T, Parmar M, Torri V, Omura G, Colombo N, Williams C, Conte P, Vermorken J. Using the expected survival to explain differences between the results of randomized trials: A case in advanced ovarian cancer. J Clin Oncol 2003; 21(9): 1682-1687.
7	Buzdar AU. A summary of second-line randomized studies of aromatase inhibitors. J Steroid Biochem Mol Biol 2001; 79(1-5): 109-114.
8	Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. BMJ 2005; 331(7521): 897-900.
9	Carey TS, Williams JW, Jr., Melvin C, Oldham J, Goodman F. Best practices: Comparing medication treatments in mental health: Drug class reviews and policy challenges. Psychiatr Serv 2007; 58(6): 746-748.
10	Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005; 365(9465): 1159-1162.
11	Chou R, Fu RW. Validity of indirect comparisons in meta-analysis - Authors' reply. Lancet 2007; 369(9558): 271
12	Cipriani A, Furukawa TA, Churchill R, Barbui C. Validity of indirect comparisons in meta-analysis. Lancet 2007; 369(9558): 270-271.
13	Clavarezza M, Del ML, Venturini M. Taxane-containing chemotherapy in the treatment of early breast cancer patients. Ann Oncol 2006; 17(Suppl 7): 22-26.
14	Cosman F, Borges J, Diaz CM. Clinical evaluation of novel bisphosphonate dosing regimens in osteoporosis: The role of comparative studies and implications for future studies. Clin Ther 2007; 29(6): 1116-1127.
15	Cundiff DK. Anticoagulants for nonvalvular atrial fibrillation (NVAf) - Drug review. Med Gen Med 2003; 5(1).
16	Daya S. Adjusted indirect comparison of competing interventions. Evid Based Obstet Gynecol 2004; 6: 103-104.
17	De Mey C. Alpha1-blocker therapy for lower urinary tract symptoms suggestive of benign prostatic obstruction: What are the relevant differences in randomised controlled trials? Eur Urol 2000; 38 (Suppl 1): 25-39.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

18	Etminan M, Carleton B, Rochon PA. Quantifying adverse drug events - Are systematic reviews the answer? <i>Drug Saf</i> 2004; 27(11): 757-761.
19	Feldman SR, Garton R, Averett W, Balkrishnan R, Vallee J. Strategy to manage the treatment of severe psoriasis: Considerations of efficacy, safety and cost. <i>Exp Opin Pharmacother</i> 2003; 4(9): 1525-1533.
20	Hankey GJ, Eikelboom JW. Cyclooxygenase-2-inhibitors: Are they really atherothrombotic, and if not, why not? <i>Stroke</i> 2003; 34(11): 2736-2740.
21	Hills RK, Richards SM, Wheatley K. Corner cutting compromises clinical trials: The inherent problems with up-front randomisation and a common standard arm. <i>Leuk Res</i> 2003; 27(12): 1071-1073.
22	Ioannidis JPA. Indirect comparisons: The mesh and mess of clinical trials. <i>Lancet</i> 2006; 368(9546): 1470-1472.
23	Leroy V, Sakarovitch C, Cortina-Borja M, McIntyre J, Coovadia H, Dabis F, Newell ML, Saba J, Gray G, Ndugwa C, Kilewo C, Massawe A, Kituuka P, Okong P, Grulich A, von Briesen H, Goudsmit J, Biberfeld G, Haverkamp G, Weverling GJ, Lange JM. Is there a difference in the efficacy of peripartum antiretroviral regimens in reducing mother-to-child transmission of HIV in Africa? <i>AIDS</i> 2005; 19(16): 1865-1875.
24	Lim E. BMJ Homepage © 2008. First principles or evidence based critique? bmj.bmjournals.com/cgi/eletters/327/7427/1309#46233 (13.04.2008).
25	Lufkin EG, Sarkar S, Kulkarni PM, Ciaccia AV, Siddhanti S, Stock J, Plouffe L. Antiresorptive treatment of postmenopausal osteoporosis: Review of randomized clinical studies and rationale for the evista alendronate comparison (EVA) trial. <i>Curr Med Res Opin</i> 2004; 20(3): 351-357.
26	Lundgren JD, Danoff-Burg S, Anderson DA. Cognitive-behavioral therapy for bulimia nervosa: An empirical analysis of clinical significance. <i>DARE</i> 2007; 4: DA20055049.
27	Lundgren JD, Phillips AN. Indirect comparisons: A novel approach to assessing the effect of anti-HIV drugs. <i>BMJ</i> 2004; 328(7434): 253.
28	Malone DC. Using indirect comparisons in pharmaco-economic studies - time for implementation. <i>Clin Ther</i> 2007; 29(11): 2454-2455.
29	Matas AJ. Resolved: In minimizing kidney transplant immunosuppression, steroids should go before calcineurin inhibitors. <i>J Am Soc Nephrol</i> 2007; 18(12): 3026-3028.
30	Michel P, Merle V, Chiron A, Ducrotte P, Paillet B, Hecketsweiler P, Czernichow P, Colin R. Postoperative management of stage II/III colon cancer: A decision analysis. <i>Gastroenterology</i> 1999; 117(4): 784-793.
31	Moore RA, McQuay HJ. Single-patient data meta-analysis of 3453 postoperative patients: Oral tramadol versus placebo, codeine and combination analgesics. <i>Pain</i> 1997; 69(3): 287-294.
32	Schellinger PD, Juttler E, Myeding-Lamade UK, Schwark C. The value of platelet inhibitors in the secondary prophylaxis of stroke - A review. <i>Fortschr Neurol Psychiatr Grenzgeb</i> 2004; 72(N5): 270-281.
33	Skjoldt NM, Rowe BH. Evidence-based emergency medicine/systematic review abstract. The role of leukotriene receptor antagonists in asthma care. <i>Ann Emerg Med</i> 2008; 51(5): 663-665.
34	Waters JS, O'Brien ME. The case for the introduction of new chemotherapy agents in the treatment of advanced non small cell lung cancer in the wake of the findings of the National Institute of Clinical Excellence (NICE). <i>Br J Cancer</i> 2002; 87(5): 481-490.
35	Wehren LE, Hosking D, Hochberg MC. Comment on putting evidence-based medicine into clinical practice: Comparing anti-resorptive agents for the treatment of osteoporosis - Authors' reply. <i>Curr Med Res Opin</i> 2004; 20(11): 1822-1824.
3. Ausschlussgrund: Systematische Übersichtsarbeiten zu diagnostischen Testverfahren und komplexen Interventionen	
1	Burnett MG, Stein SC, Sonnad SS, Zager EL. Cost-effectiveness of intraoperative imaging in carotid endarterectomy. <i>Neurosurgery</i> 2005; 57(3): 478-484.
2	Clark AM, Hartling L, Vandermeer B, Lissel SL, McAlister FA. Secondary prevention programmes for coronary heart disease: A Metaregression showing the merits of shorter, generalist, primary care-based interventions. <i>Eur J Cardiovasc Prev Rehabil</i> 2007; 14(4): 538-546.
3	Guisse JM, Palda V, Westhoff C, Chan BK, Helfand M, Lieu TA. The effectiveness of primary care-based interventions to promote breastfeeding: Evidence review and meta-analysis for the U.S. Preventive Services Task Force. <i>Ann Fam Med</i> 2003; 1(2): 70-78.
4	Issakidis C, Sanderson K, Corry J, Andrews G, Lapsley H. Modelling the population cost-effectiveness of current and evidence-based optimal treatment for anxiety disorders. <i>Psychol Med</i> 2004; 34(1): 19-35.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

4. Ausschlussgrund: Übersichtsarbeiten, in denen nicht über die Ergebnisse eines indirekten Vergleichs der Wirksamkeit therapeutischer Intervention berichtet wird	
1	N. N. Angioplasty and stenting of the cervical carotid artery with distal embolic protection of the cerebral circulation. Chicago IL: Blue Cross Blue Shield Association (BCBS) 2005
2	N. N. Imiquimod: Basal cell carcinoma: Inferior to other treatments. Prescrire Int 2006; 130-131.
3	N. N. Pegaptanib: New drug. In macular degeneration: Too many risks for too little benefit. Prescrire Int 2006; 15(84): 127-129.
4	N. N. Zonisamide: New drug. No advantage in refractory partial epilepsy. Prescrire Int 2007; 16(89): 95-97.
5	Akl EA, Karmath G, Yosucio V, Kim SY, Barba M, Sperati F, Cook D, Schunemann HJ. Anticoagulation for thrombosis prophylaxis in cancer patients with central venous catheters. Cochrane Database Syst Rev 2007; (3).
6	Attia J, Ray JG, Cook DJ, Douketis J, Ginsberg JS, Geerts WH. Deep vein thrombosis and its prevention in critically ill adults. DARE 2002; 4: DA20018181.
7	Babcock JC, Green CE, Robie C. Does batterers' treatment work: A metaanalytic review of domestic violence treatment. Clin Psychol Rev 2004; 23(8): 1023-1053.
8	Bergh J, Jönsson PE, Glimelius B, Nygren P. A systematic overview of chemotherapy effects in breast cancer. Acta Oncol 2001; 40(2-3): 253-281.
9	Bucher HC, Griffith L, Guyatt GH, Opravil M. Meta-analysis of prophylactic treatments against pneumocystis carinii pneumonia and toxoplasma encephalitis in HIV-infected patients. J Acquir Immun Defic Syndr Hum Retrovirol 1997; 15(2): 104-114.
10	Clarke JA, Tulder MW, Blomberg SE, Vet HC, Heijden GJ, Bronfort G, Bouter LM. Traction for low-back pain with or without sciatica. Cochrane Database of Systematic Reviews 2007; 4: CD003010.
11	Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer. Health Technol Assess 2001; 5(32): 1-195.
12	Cranney A, Guyatt G, Griffith L, Wells G, Tugwell P, Rosen C. Summary of metaanalyses of therapies for postmenopausal osteoporosis. Endocr Rev 2002; 23(4): 570-578.
13	Cuncins-Hearn A, Saunders C, Walsh D, Borg M, Buckingham J, Frizelle F, Maddern G. A systematic review of intraoperative radiotherapy in early stage breast cancer. ASERNIP-S Report Nr. 27, Adelaide, South Australia: ASERNIP-S, 2002.
14	Dean BB, Gano AD, Knight K, Ofman JJ, Fass R. Effectiveness of proton pump inhibitors in nonerosive reflux disease. Clin Gastroenterol Hepatol 2004; 2(8): 656-664.
15	Dolder CR, Lacro JP, Leckband S, Jeste DV. Interventions to improve antipsychotic medication adherence: Review of recent literature. J Clin Psychopharmacol 2003; 23(4): 389-399.
16	Dolovich LR, Ginsberg JS, Douketis JD, Holbrook AM, Cheah G. A meta-analysis comparing low-molecular-weight heparins with unfractionated heparin in the treatment of venous thromboembolism: Examining some unanswered questions regarding location of treatment, product type and dosing frequency. Arch Int Medi 2000; 160(2): 181-188.
17	Dranitsaris G, Leung P, Mather J, Oza A. Cost-utility analysis of second-line hormonal therapy in advanced breast cancer: A comparison of two aromatase inhibitors to megestrol acetate. Anticancer Drugs 2000; 11(7): 591-601.
18	Etminan M, Gill S, Samii A. Comparison of the risk of adverse events with pramipexole and ropinirole in patients with Parkinson's disease: A meta-analysis. Drug Saf 2003; 26(6): 439-444.
19	Fisher LD, Gent M, Buller HR. Active-control trials: How would a new agent compare with placebo? A method illustrated clopidogrel, aspirin, and placebo. Am Heart J 2001; 141(1): 26-32.
20	Furukawa TA, Watanabe N, Churchill R. Psychotherapy plus antidepressant for panic disorder with or without agoraphobia - Systematic review. Br J Psychiatry 2006; 188: 305-312.
21	Gisbert JP, Gonzalez L, Calvet X, Roque M, Gabriel R, Pajares JM. Helicobacter pylori eradication: Proton pump inhibitor versus ranitidine bismuth citrate plus two antibiotics for 1 week. A meta-analysis of efficacy. Aliment Pharmacol Ther 2000; 14(9): 1141-1150.
22	Hasselblad V, Kong DF, Hasselblad V, Kong DF. Statistical methods for comparison to placebo in active-control trials 209. Drug Inform J 2001; 435-449.
23	Hillier SL, Hollohan V. Vestibular rehabilitation for unilateral peripheral vestibular dysfunction. Cochrane Database of Syst Rev 2007; (2007 Issue 4): CD005397.
24	Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen. Stem cell transplantation in adults with acute lymphoblastic leukaemia (ALL) or acute myeloid leukaemia (AML). Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2007.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

25	King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, Golder S, Taylor E, Drummond M, Riemsma R. A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. <i>Health Technol Assess</i> 2006; 10(23): 1-162.
26	Kotalik J, Yu E, Markman BR, Gagliardi A, Evans WK. Cancer Care Ontario Practice Guidelines Initiative Homepage © 2004. Prophylactic cranial irradiation in small cell lung cancer. www.cancercare.on.ca/pdf/pebc7-13-2f.pdf (20.02.2008).
27	Lotan Y, Gettman MT, Roehrborn CG, Cadeddu JA, Pearle MS. Management of ureteral calculi: A cost comparison and decision making analysis. <i>J Urol</i> 2002; 167(4): 1621-1629.
28	Lubomski LH, Magaziner J, Sprintz M, Kempen J, Reeves SW, Robinson KA, Bass EB. Anesthesia management during cataract surgery. Rockville, MD, USA: Agency for Healthcare Research and Quality 2001
29	MacFadyen CA, Acuin JM, Gamble C. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations. <i>Cochrane Database of Syst Rev</i> 2005; (4): CD004618.
30	Martin CW, WCB Evidence Practice Group. Artificial cervical and lumbar disc implants: A review of the literature. <i>Work Safe BC</i> 2005
31	McQuay HJ, Moore RA. Dose-response in direct comparisons of different doses of aspirin, ibuprofen and paracetamol (acetaminophen) in analgesic studies. <i>Br J Clin Pharmacol</i> 2007; 63(3): 271-278.
32	Molloy D, Kaloo PD, Cooper M, Nguyen T, V. Laparoscopic entry: A literature review and analysis of techniques and complications of primary port entry. <i>Aust N Z J Obstet Gynaecol</i> 2002; 42(3 Suppl 2): 246-254.
33	Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. <i>JAMA</i> 2001; 286(8): 954-959.
34	Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P. The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: Systematic review and economic evaluation. <i>Health Technol Assess</i> 2006; 10(41): 1-204.
35	Playford EG, Webster AC, Sorell TC, Craig JC. Antifungal agents for preventing fungal infections in solid organ transplant recipients. <i>Cochrane Database of Syst Rev</i> 2004(3): CD004291.
36	Radke PW, Kaiser A, Frost C, Sigwart U. Outcome after treatment of coronary in-stent restenosis: Results from a systematic review using meta-analysis techniques. <i>Eur Heart J</i> 2003; 24(3): 266-273.
37	Richy F, Bruyere O, Ethgen O, Cucherat M, Honroin Y, Reginster J. Structural and symptomatic efficacy of glucosamine and chondroitin in knee osteoarthritis: A comprehensive meta-analysis. <i>Arch Intern Med</i> 2003; 163(13): 1514-1522.
38	Rodgers A, Walker N, Schug S, McKee A, Kehlet H, van ZA, 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. <i>BMJ</i> 2000; 321: 1493-1497.
39	Rostom A, Goldkind L, Laine L. Nonsteroidal anti-inflammatory drugs and hepatic toxicity: A systematic review of randomized controlled trials in arthritis patients. <i>Clin Gastroenterol Hepatol</i> 2005; 3(5): 489-498.
40	Sampaio C, Costa J, Ferreira JJ. Clinical comparability of marketed formulations of botulinum toxin. <i>Mov Disord</i> 2004; 19(Suppl 8): 129-136.
41	Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published Metaanalyses. <i>BMJ</i> 2003; 326(7387)
42	Soomro GM, Altman D, Rajagopal S, Oakley-Browne M. Selective serotonin re-uptake inhibitors (SSRIs) versus placebo for obsessive compulsive disorder (OCD). <i>Cochrane Database Syst Rev</i> 2008; (1): CD001765
43	Sprague DA, Loewen PS, Raymond CB. Selection of atypical antipsychotics for the management of schizophrenia. <i>Ann Pharmacother</i> 2004; 38(2): 313-319.
44	Tomiak E, Verma S, Levine M, Pritchard K, Sawka C, Breast Cancer Disease Site Group. Use of capecitabine in state IV breast cancer: An evidence summary. <i>Curr Oncol</i> 2000; 7(2): 84-90.
45	Torrens M, Fonseca F, Mateu G, Farre M. Efficacy of antidepressants in substance use disorders with and without comorbid depression: A systematic review and meta-analysis. <i>Drug Alcohol Depend</i> 2005; 78(1): 1-22.
46	Ward S, Kaltenthaler E, Cowen J, Brewer N. Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: Systematic review and economic evaluation. 2003.
47	Woods SW, Gueorguieva RV. Control group bias in randomized atypical antipsychotic medication trials for schizophrenia. <i>Arch Gen Psychiatry</i> 2005; 96: 1-970.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

48	Zhang WY, Po AL, Dua HS, Zuara-Blanco A. Meta-analysis of randomised controlled trials comparing latanoprost with timolol in the treatment of patients with open angle glaucoma or ocular hypertension. <i>Br J Ophthalmol</i> 2001; 85(8): 983-990.
5. Ausschlussgrund: Publikationen, ausschließlich methodischen Inhalts, die keine Anleitung geben, wie indirekte Vergleiche therapeutischer Interventionen durchzuführen sind oder eine Methode zur Durchführung indirekter Vergleiche validieren oder bewerten.	
1	Alioum A, Dabis F, Quae-Merchadou L, Haverkamp G, Hudgens M, Hughes J, Karon J, Leroy V, Newell ML, Richardson B, Weverling GJ. Estimating the efficacy of interventions to prevent mother-to-child transmission of HIV in breast-feeding populations: Development of a consensus methodology. <i>Stat Med</i> 2001; 20(23): 3539-3556.
2	Antman EM, Ferguson JJ. Should evidence-based proof of efficacy as defined for a specific therapeutic agent be extrapolated to encompass a therapeutic class of agents? <i>Circulation</i> 2003; 108(21): 2604-2607.
3	Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Williams HJW, Zaza S. Grading quality of evidence and strength of recommendations. <i>BMJ</i> 2004; 328(7454): 1490-1494.
4	Baker SG, Kramer BS. The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? <i>BMC Med Res Methodol</i> 2002; 2:13.
5	Bansback N, Maetzel A, Drummond M, Anis A, Marra C, Conway P, Boers M, Tugyvell P, Boonen A. Considerations and preliminary proposals for defining a reference case for economic evaluations in ankylosing spondylitis. <i>J Rheumatol</i> 2007; 34(5): 1178-1183.
6	Begg CB, Pilote L. A model for incorporating historical controls into a meta-analysis. <i>Biometrics</i> 1991; 47(3): 899-906.
7	Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. <i>Stat Med</i> 1996; 15(5): 537-557.
8	Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data Metaregressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. <i>Stat Med</i> 2002; 21(3): 371-387.
9	Beydoun A, Kutluay E. Conversion to monotherapy - Clinical trials in patients with refractory partial seizures. <i>Neurology</i> 2003; 60(11): 13-25.
10	Bien CG, Elger CE. Monotherapy trials in antiepileptic drugs: Are modified presurgical studies a way out of the dilemma? <i>Epilepsy Res</i> 2001; 44(N1): 1-5.
11	Bouvenot G. Quantification of pharmacological progress by the French national health authorities. <i>Bull Acad Natl Med</i> 2006; 190(4-5): 893-903.
12	Brody BA, Dickey N, Ellenberg SS, Heaney RP, Levine RJ, O'Brien RL, Purtilo RB, Weijer C. Is the use of placebo controls ethically permissible in clinical trials of agents intended to reduce fractures in osteoporosis? <i>J Bone Miner Res</i> 2003; 18(6): 1105-1109.
13	Buchner T, Dohner H, Ehninger G, Ganser A, Niederwieser D, Hasford J. Cross-trial networking in AML: A step forward rather than corner cutting. <i>Leuk Res</i> 2004; 28(6): 649-650.
14	Califf RM. Benefit assessment of therapeutic products: The centers for education and research on therapeutics. <i>Pharmacoepidemiol Drug Saf</i> 2007; 16(N1): 5-16.
15	Chow SC, Shao J. On non-inferiority margin and statistical tests in active control trials. <i>Stat Med</i> 2006; 25(7): 1101-1113.
16	Clarke MJ, Stewart AL. Obtaining data from randomised controlled trials: How much do we need for reliable and informative Metaanalyses? <i>BMJ</i> 1994; 309(6960): 1007-1010.
17	Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, and Altman DG (Eds). <i>Systematic reviews in health care: Meta-analysis in context</i> . 2 ed. London, 2001, pp 285-312.
18	Drummond M, Sculpher M. Common methodological flaws in economic evaluations. <i>Med Care</i> 2005; 43(7): 5-14.
19	Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 Metaanalyses. <i>Stat Med</i> 2000; 19(13): 1707-1728.
20	Gardiner JC, Huebner M, Jetton J, Bradley CJ. On parametric confidence intervals for the cost-effectiveness ratio. <i>Biometrical Journal</i> 2001; 43(3): 283-296.
21	Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. <i>Drug Inform J</i> 1999; 33: 109-118.
22	Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. <i>Stat Med</i> 2001; 20(15): 2219-2241.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

23	Hung HMJ, Wang SJ, O'Neill R. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. <i>Biometrical Journal</i> 2005; 47(1): 28-36.
24	Hung HMJ, Wang SJ, O'Neill R. Issues with statistical risks for testing methods in noninferiority trial without a placebo arm. <i>J Biopharm Stat</i> 2007; 17(2): 201-213.
25	N. N. International Conference on Harmonisation Homepage © 1998. ICH Harmonised Tripartite Guideline. Choice of control group and related issues in clinical trials. www.ich.org/LOB/media/MEDIA385.pdf (15.07.2008).
26	Kaul S., Diamond GA. Good enough: A primer on the analysis and interpretation of noninferiority trials. <i>Ann Inter Med</i> 2006; 145(1): 62-69.
27	Kaul S, Diamond GA. Making sense of noninferiority: A clinical and statistical perspective on its application to cardiovascular clinical trials. <i>Prog Cardiovasc Dis</i> 2007; 49(4): 284-299.
28	Khan KS, ter Ried G, Glanville J, Sowden AJ, Kleijnen J. University of York Homepage © 2001. CRD Report 4. www.york.ac.uk/inst/crd/report4.htm (20.04.2008).
29	Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level Covariates in Metaregression with individual patient data meta-analysis. <i>J Clin Epidemiol</i> 2002; 55(1): 86-94.
30	Li Z, Begg CB. Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. <i>J Am Stat Assoc</i> 1994; 89: 1523-1527.
31	Manca A, Willan AR. 'Lost in translation' - Accounting for between-country differences in the analysis of multinational cost-effectiveness data. <i>Pharmacoeconomics</i> 2006; 24(11): 1101-1119.
32	Mason JM, Mason AR. The generalisability of pharmacoeconomic studies. Issues and challenges ahead. <i>Pharmacoeconomics</i> 2006; 24(10): 937-945.
33	McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' guides to the medical literature - XIX. Applying clinical trial results - B. Guidelines for determining whether a drug is exerting (more than) a class effect. <i>JAMA</i> 1999; 282(14): 1371-1377.
34	Phillips A. Trial and error: Cross-trial comparisons of antiretroviral regimens. <i>AIDS</i> 2003; 17(4): 619-623.
35	Raghunathan TE. Pooling controls from different studies. <i>Stat Med</i> 1991; 10(9): 1417-1426.
36	Sanchez OD, Latorre FP, Blanco MP, Bosch PV. Therapeutic equivalence: Concept and evidence levels. <i>Med Clin (Barc)</i> 2007; 129(19): 736-745.
37	Santaguida PL, Helfand M, Raina P. Challenges in systematic reviews that evaluate drug efficacy or effectiveness. <i>Ann Intern Med</i> 2005; 142(12): 1066-1072.
38	Schneeweiss S. Developments in post-marketing comparative effectiveness research. <i>Clin Pharmacol Ther</i> 2007; 82(2): 143-156.
39	Siegel JP. Equivalence and non-inferiority trials. <i>Am Heart J</i> 2000; 139: 166-170.
40	Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. <i>Stat Med</i> 2005; 24(9): 1307-1319.
41	Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. <i>Stat Sci</i> 2004; 19(1): 156-174.
42	Sutton AJ, Higgins JP. Recent developments in meta-analysis. <i>Stat Med</i> 2008; 27(5): 625-650.
43	Thompson SG. Why and how sources of heterogeneity should be investigated. In: Egger M, Davey Smith G, and Altman DG (Eds). <i>Systematic reviews in health care: Meta-analysis in context</i> . 2 ed. London, 2001, pp 157-175.
44	Tosteson AN, Jönsson B, Grima DT, O'Brien BJ, Black DM, Adachi JD. Challenges for model-based economic evaluations of postmenopausal osteoporosis interventions. <i>Osteoporos Int</i> 2001; 12(10): 849-857.
45	Willan AR. Analysis, sample size and power for estimating incremental net health benefit from clinical trial data. <i>Control Clin Trials</i> 2001; 22(3): 228-237.
46-65	Hinzu kommen unter diesem Ausschussgrund die 20 ausgeschlossenen Methodenpapiere, die bei der Handsuche auf den Internetseiten der HTA-Institutionen gefunden wurden (siehe Tabelle 14)
6. Ausschlussgrund: Mehrfachpublikationen mit identischem Inhalt	
1	Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu GB. Bayesian methods for evidence synthesis in cost-effectiveness analysis. <i>Pharmacoeconomics</i> 2006; 24(1): 1-19.
2	Ferrari MD. Current perspectives on effective migraine treatments: Are small clinical differences important for patients? <i>Drugs Today</i> 2003; 39: 37-41.
3	Guyatt GH, Cranney A, Griffith L, Walter S, Krolicki N, Favus M, Rosen C. Summary of Metaanalyses of therapies for postmenopausal osteoporosis and the relationship between bone density and fractures. <i>Endocrinol Metab Clin North Am</i> 2002; 31(3): 659-679

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

4	Nixon R, Bansback N, Brennan A. The efficacy of inhibiting tumour necrosis factor alpha and interleukin 1 in patients with rheumatoid arthritis: A meta-analysis and adjusted indirect comparisons. <i>Rheumatology (Oxford)</i> 2007; 46(7): 1140-1147
5	Playford EG, Webster AC, Sorrell TC, Craig JC. Systematic review and meta-analysis of antifungal agents for preventing fungal infections in liver transplant recipients. <i>Eur J Clin Microbiol Infect Dis</i> 2006; 25(9): 549-561
6	Wehren LE, Hosking D, Hochberg MC, Melton M, Ross PD. Adjusted indirect comparisons to assess the relative efficacy of osteoporosis treatments. <i>Osteoporosis Int</i> 2003; 14: 75-76
7. Ausschlussgrund: Kosten-Nutzen-Betrachtungen, für die keine systematische Literaturrecherche durchgeführt wurde	
1	Bachmann MO. Effectiveness and cost effectiveness of early and late prevention of HIV/AIDS progression with antiretrovirals or antibiotics in Southern African adults. <i>AIDS Care</i> 2006; 18(2): 109-120.
2	Bansback NJ, Brennan A, Ghatnekar O. Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden. <i>Ann Rheum Dis</i> 2005; 64(7): 995-1002.
3	Bjorvatn A, Kristiansen F. Fondaparinux sodium compared with enoxaparin sodium: A cost-effectiveness analysis. <i>Am J Cardiovasc Drugs</i> 2005; 5(2): 121-130.
4	Borghgi J, Guest JF. Economic impact of using mirtazapine compared to amitriptyline and fluoxetine in the treatment of moderate and severe depression in the UK. <i>Eur Psychiatry</i> 2000; 15(6): 378-387.
5	Botteman M, Barghout V, Stephens J, Hay J, Brandman J, Aapro M. Cost effectiveness of bisphosphonates in the management of breast cancer patients with bone metastases. <i>Ann Oncol</i> 2006; 17(7): 1072-1082.
6	Brown MCJ, Van Loon JMT, Guest JF, Brown MCJ, Van Loon JMT, Guest JF. Cost-effectiveness of mirtazapine relative to amitriptyline in the treatment of moderate and severe depression in France. <i>Eur Psychiatry</i> 1999; 197-208.
7	Calvert NW, Burch SP, Fu AZ, Reeves P, Thompson TR. The cost-effectiveness of lamotrigine in the maintenance treatment of adults with bipolar I disorder. <i>J Manag Care Pharm</i> 2006; 12(4): 8322-330.
8	Chan PS, Vijan S, Morady F, Oral H. Cost-effectiveness of radiofrequency catheter ablation for atrial fibrillation. <i>J Am Coll Cardiol</i> 2006; 47(12): 2513-2520.
9	Choi HK, Seeger JD, Kuntz KM. A cost effectiveness analysis of treatment options for methotrexate-naive rheumatoid arthritis. <i>J Rheumatol</i> 2002; 29(6): 1156-1165.
10	Cowper PA, DeLong ER, Whellan DJ, LaPointe NM, Califf RM. Economic effects of beta-blocker therapy in patients with heart failure. <i>Am J Med</i> 2004; 116(2): 104-111.
11	Dewilde S, Turk F, Tambour M, Sandstrom T. The economic value of anti-IgE in severe persistent, IgE-mediated (allergic) asthma patients: Adaptation of INNOVATE to Sweden. <i>Curr Med Res Opin</i> 2006; 22(9): 1765-1776.
12	Dominguez-Gil A, Martin I, Garcia VM, Del CA, Diaz S, Sanchez C. Economic evaluation of voriconazole versus caspofungin for the treatment of invasive aspergillosis in Spain. <i>Clin Drug Investig</i> 2007; 27(3): 197-205.
13	Eggington S, Tappenden P, Pandor A, Paisley S, Saunders M, Seymour M, Sutcliffe P, Chilcott J. Cost-effectiveness of oxaliplatin and capecitabine in the adjuvant treatment of stage III colon cancer. <i>Br J Cancer</i> 2006; 95(9): 1195-1201.
14	Ess SM, Schaad UB, Gervais A, Pinosch S, Szucs TD. Cost-effectiveness of a pneumococcal conjugate immunisation program for infants in Switzerland. <i>Vaccine</i> 2003; 21(23): 3273-3281.
15	Freedberg KA, Losina E, Weinstein MC, Paltiel AD, Cohen CJ, Seage GR, Craven DE, Zhang H, Kimmel AD, Goldie SJ. The cost effectiveness of combination antiretroviral therapy for HIV disease. <i>N Engl J Med</i> 2001; 344(11): 824-831.
16	Frighetto L, Loewen PS, Dolman J, Marra CA. Cost-effectiveness of prophylactic dolasetron or droperidol vs. rescue therapy in the prevention of PONV in ambulatory gynecologic surgery. <i>Can J Anaesth</i> 1999; 46(6): 536-543.
17	Goeree R, Blackhouse G, Adachi J. Cost-effectiveness of alternative treatments for women with osteoporosis in Canada. <i>Curr Med Res Opin</i> 2006; 22(7): 1425-1436.
18	Heaney DC, Shorvon SD, Sander JW, Boon P, Komarek V, Marusic P, Dravet C, Peerucca E, Majkowski J, Lopes LJ, Arroyo S, Tomson T, Ried S, van Donselaar C, Eskazan E, Peeters P, Carita P, Hung I, Myon E, Taieb C. Cost minimization analysis of antiepileptic drugs in newly diagnosed epilepsy in 12 European countries. <i>Epilepsia</i> 2000; 41(Suppl 5): 37-44.
19	Hur C, Nishioka NS, Gazelle GS. Cost-effectiveness of aspirin chemoprevention for Barrett's esophagus. <i>J Natl Cancer Inst</i> 2004; 96(4): 316-325.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

20	Jansen JP, Meis JF, Blijlevens NM, van't Wout JW. Economic evaluation of voriconazole in the treatment of invasive aspergillosis in the Netherlands. <i>Curr Med Res Opin</i> 2005; 21(10): 1535-1546.
21	Kongsakon R, Leelahanj T, Price N, Birinyi-Strachan L, Davey P. Cost analysis of the treatment of schizophrenia in Thailand: A simulation model comparing olanzapine, risperidone, quetiapine, ziprasidone and haloperidol. <i>J Med Assoc Thai</i> 2005; 88(9): 1267-1277.
22	Malone DC, Tran TT, Poordad FF. Cost-efficacy analysis of peginterferon alfa-2b plus ribavirin compared with peginterferon alfa-2a plus ribavirin for the treatment of chronic hepatitis C. <i>J Manag Care Pharm</i> 2005; 11(8): 687-694.
23	Marchetti M, Cavallo MC, Annoni E, Gerzeli S. Cost-utility of inhaled corticosteroids in patients with moderate-to-severe asthma. <i>Expert Rev Pharmacoeconomics Outcomes Res</i> 2004; 43(5): 549-564.
24	Messori A, Bosi A, Bacci S, Laszlo D, Trippoli S, Locatelli F, Van Lint MT, Di Bartolomeo P, Amici A. Retrospective survival analysis and cost-effectiveness evaluation of second allogeneic bone marrow transplantation in patients with acute leukemia. <i>Bone Marrow Transplant</i> 1999; 23(5): 489-495.
25	Mohr PE, Neumann PJ, Franco SJ, Marainen J, Lockridge R, Ting G. The case for daily dialysis: Its impact on costs and quality of life. <i>Am J Kidney Dis</i> 2001; 37(4): 777-789.
26	Russo MW, Zacks SL, Sandler RS, Brown RS. Cost-effectiveness analysis of transjugular intrahepatic portosystemic shunt (TIPS) versus endoscopic therapy for the prevention of recurrent esophageal variceal bleeding. <i>Hepatology</i> 2000; 31(2): 358-363.
27	Scanlon E, Karlsmark T, Leaper DJ, Carter K, Poulsen PB, Hart-Hansen K, Hahn TW. Cost-effective faster wound healing with a sustained silver-releasing foam dressing in delayed healing leg ulcers: A health-economic analysis. <i>Int Wound J</i> 2005; 2(2): 150-160.
28	Schnitzler MA, Woodward RS, Lowell JA, Singer GG, Brennan DC. Ten-year cost effectiveness of alternative immunosuppression regimens in cadaveric renal transplantation. <i>Transplant Proc</i> 1999; 31(3B Suppl S): 19-21.
29	Shanahan MD, Doran CM, Digiusto E, Bell J, Lintzeris N, White J, Ali R, Saunders JB, Mattick RP, Gilmour S. A cost-effectiveness analysis of heroin detoxification methods in the Australian National Evaluation of Pharmacotherapies for Opioid Dependence (NEPOD). <i>Addict Behav</i> 2006; 31(3): 371-387.
30	Simon J, Gray A, Du L. Cost-effectiveness of prophylactic magnesium sulphate for 9996 women with pre-eclampsia from 33 countries: Economic evaluation of the Magpie Trial. <i>Int J Obst Gynaecol</i> 2006; 113(2): 144-151.
31	Song F, Raftery J, Aveyard P, Hyde C, Barton P, Woolacott N. Cost-effectiveness of pharmacological interventions for smoking cessation: A literature review and a decision analytic analysis. <i>Med Decis Making</i> 2002; 22 (Suppl): 26-37.
32	Starling N, Tilden D, White J, Cunningham D. Cost-effectiveness analysis of cetuximab/irinotecan vs active/best supportive care for the treatment of metastatic colorectal cancer patients who have failed previous chemotherapy treatment. <i>Br J Cancer</i> 2007; 96(2): 206-212.
33	Tarride JE, Gordon A, Vera-Llonch M, Dukes E, Rousseau C. Cost-effectiveness of pregabalin for the management of neuropathic pain associated with diabetic peripheral neuropathy and postherpetic neuralgia: A Canadian perspective. <i>Clin Ther</i> 2006; 28(11): 1922-1934.
34	Taylor RJ, Taylor RS. Spinal cord stimulation for failed back surgery syndrome: A decision-analytic model and cost-effectiveness analysis. <i>Int J Technol Assess Health Care</i> 2005; 21(3): 351-358.
35	UK700 Group. Cost-effectiveness of intensive vs. standard case management for severe psychotic illness. <i>Br J Psychiatry</i> 2000; 176: 537-543.
36	VA-HIT Study Group. Cost-effectiveness of gemfibrozil for coronary heart disease patients with low levels of high-density lipoprotein cholesterol. <i>Arch Intern Med</i> 2002; 162: 177-182.
37	Vale L, Grant A, McCormack K, Scott NW, EU Hernia Trialists Collaboration. Cost-effectiveness of alternative methods of surgical repair of inguinal hernia. <i>Int J Technol Assess Health Care</i> 2004; 20(2): 192-200.
38	Van Baardewijk M, Vis PM, Einarson TR. Cost effectiveness of duloxetine compared with venlafaxine-XR in the treatment of major depressive disorder. <i>Curr Med Res Opin</i> 2005; 21(8): 1271-1279.
39	Verma S, Rocchi A. Economic evaluation of antiaromatase agents in the second-line treatment of metastatic breast cancer. <i>Support Care Cancer</i> 2003; 11(11): 728-734.
40	Wang PS, Ganz DA, Benner JS, Glynn RJ, Avorn J. Should clozapine continue to be restricted to third-line status for schizophrenia: A decision-analytic model. <i>J Ment Health Policy Econ</i> 2004; 7: 77-85.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

8. Ausschlussgrund: Der Volltext konnte nicht beschafft werden, da der Artikel nicht über Subito bestellbar war, nur in einem Buch erschienen ist oder bisher nur der Abstract oder das Protokoll publiziert wurde	
1	Armitage P, Berry G. The sampling error of a difference. Stat Methods Med Res. 2 ed. Oxford, 1987, pp 88-90.
2	Belsey J. Reconciling effectiveness and tolerability in oral triptan therapy: A quantitative approach to decision making in migraine management. J Clin Res 2001; 4: 105-125.
3	Belsey JD. The clinical and financial impact of oral triptans: An updated meta-analysis. J Med Econ 2002; 5: 79-89.
4	Clarke M. Individual patient data meta-analyses compared with meta-analyses based on aggregate data. Cochrane Database Syst Rev 2008
5	Djulbegovic B. New treatments compared to established treatments in randomized trials. Cochrane Database Syst Rev 2008.
6	Eisenberg MJ, Yavin D, Filion KB, Belisle P, Joseph L, Gervais A, O'Loughlin J, Paradis G, Rinfret S, Pilote L. A hierarchical Bayesian meta-analysis of randomized controlled trials of smoking cessation pharmacotherapies. Circulation 2006; 114(18, Suppl S): 855.
7	Hansen RA, Gaynes BN, Gartlehner G, Moore CG, Tiwari R, Lohr KN. Efficacy and tolerability of second-generation antidepressants in social anxiety disorder. Int Clin Psychopharmacol 2008; 23(3): 170-179.
8	Jansen JP, Crawford B. A Bayesian approach to predict effectiveness of newly introduced drugs in daily practice based on the relation between efficacy and effectiveness of competing interventions. Value Health 2006; 9(3): 62.
9	Mainland D: The treatment of clinical and laboratory data. Edinburgh, 1938.
10	Norris SL, Carson S, Roberts C. Comparative effectiveness of pioglitazone and rosiglitazone in type 2 diabetes, prediabetes, and the metabolic syndrome: A meta-analysis. Curr Diabetes Rev 2007; 3(2): 127-140.
11	Senn S. Active control equivalence trials. Statistical issues in drug development. Chichester, 2002, pp 207-217.
12	Song F, Altman DG, Glenny A, Eastwood AJ, Deeks JJ. Adjusted indirect comparison for estimating relative effects of competing healthcare interventions. Cochrane Database Syst Rev 2007; (4): MR000020.
13	Sutton AJ, Abrams KR, Jones DR, Sheldon TA: Methods fo meta-analysis in medical research. Chichester, 2000.
14	Swart AM, Burdett S, Ledermann J, Mook P, Parmar MK. Why i. p. therapy cannot yet be considered as a standard of care for the first-line treatment of ovarian cancer: A systematic review. Ann Oncol 2007.
9. Ausschlussgrund: Systematische Übersichtsarbeiten, die den indirekten Vergleich überwiegend auf Basis von nicht-randomisierten oder nicht-kontrollierten Studien durchführten	
1	Anderson R, Dyer M, Garside R, Mealing S, Pitt M, Price A, Rogers G, Somerville M, Stein K. The effectiveness and cost-effectiveness of carmustine implants and temozolomide for the treatment of newly diagnosed high-grade glioma: A systematic review and economic evaluation. Health Technol Assess 2007; 11(45)
2	Blue Cross Blue Shield Association. Metal-on-metal total hip resurfacing. Chicago IL: Blue Cross Blue Shield Association (BCBS) 2007
3	Caro JJ, Huybrechts KF, Green TC. Estimates of the effect on hepatic iron of oral deferiprone compared with subcutaneous desferrioxamine for treatment of iron overload in thalassemia major: A systematic review. BMC Blood Disord 2002; 2:4.
4	De Gans K, Nieuwkamp DJ, Rinkel GJ, Algra A. Timing of aneurysm surgery in subarachnoid hemorrhage: A systematic review of the literature. Neurosurgery 2002; 50(2): 336-340.
5	Hazel SJ. Systematic review of intraoperative ablation for the treatment of atrial fibrillation. North Adelaide, S.Australia, Australia: Royal Australasian College of Surgeons, Australian Safety and Efficacy Register of New Interventional Procedures (ASERNIP)-Surgical 2004
6	Jones SC. Relative thromboembolic risks associated with COX-2 inhibitors. Ann Pharmacother 2005; 39(7-8): 1249-1259.
7	Kanjeekal S, Chambers A, Fung MF, Verma S. Systemic therapy for advanced uterine sarcoma: A systematic review of the literature. Gynecol Oncol 2005; 97(2): 624-637.
8	Kwon BK, Hilibrand AS, Malloy K, Savas PE. A critical analysis of the literature regarding surgical approach and outcome for adult low-grade isthmic spondylolisthesis. DARE 2007; (4): DA20053922.
9	Le Saux N, Howard A, Barrowman NJ, Gaboury I, Sampson M, Moher D. Shorter courses of parenteral antibiotic therapy do not appear to influence response rates for children with acute hematogenous osteomyelitis: A systematic review. BMC Infect Dis 2002; 2:16.

Fortsetzung Tabelle 20: Übersicht über die Anzahl an ausgeschlossener Literatur nach Ausschlussgründen

10	Pasquina P, Tramer MR, Walder B. Prophylactic respiratory physiotherapy after cardiac surgery: Systematic review. <i>BMJ</i> 2003; 327(7428): 1379.
11	Taylor RS, Taylor RJ, Fritzell P. Balloon kyphoplasty and vertebroplasty for vertebral compression fractures: A comparative systematic review of efficacy and safety. <i>Spine</i> 2006; 31(23): 2747-2755.
Insgesamt in Volltextselektion ausgeschlossen: 227 Publikationen	

9.7 Nach Durchsicht im Volltext eingeschlossene Literatur**Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs**

Autor, Jahr	Methodengruppe des indirekten Vergleichs
Typ-1-Publikationen	
Cochrane Handbuch ⁸ 2008	Adjustierter indirekter Vergleich
Ades et al. ²⁰ 2003	Mixed treatment comparison
Ades et al. ²² 2006	Mixed treatment comparison
Ades et al. ²³ 2007	Mixed treatment comparison
Brown et al. ⁵² 1999	Metaregression (Mixed model)
Eddy ⁹³ 1989	Mixed treatment comparison
Eddy et al. ⁹⁵ 1990	Mixed treatment comparison
Eddy et al. ⁹⁴ 1992	Mixed treatment comparison
Gartlehner et al. ¹⁰⁴ 2008	Nicht-adjustierter indirekter Vergleich, Adjustierter indirekter Vergleich, Metaregression
Glenny et al. ¹⁰⁹ 2005	Nicht-adjustierter indirekter Vergleich, Adjustierter indirekter Vergleich, Metaregression, Metaregression (Mixed model)
Gleser et al. ¹¹⁰ 2001	Mixed treatment comparison
Goadsby et al. ¹¹¹ 1999	Adjustierter indirekter Vergleich
Hasselblad et al. ¹¹⁸ 1995	Mixed treatment comparison
Hasselblad et al. ¹¹⁶ 1998	Metaregression (Mixed model)
Hirotsu et al. ¹²⁶ 1999	Mixed treatment comparison
Lu und Ades ¹⁵⁷ 2004	Mixed treatment comparison
Lu et al. ¹⁵⁸ 2007	Mixed treatment comparison
Lu und Ades ¹⁵⁹ 2006	Mixed treatment comparison
Lumley ¹⁶⁰ 2002	Mixed treatment comparison
Salanti et al. ²⁰⁴ 2008	Mixed treatment comparison
Salanti et al. ²⁰⁵ 2008	Mixed treatment comparison
Smith et al. ²¹⁵ 1995	Metaregression mit Bayes'schem Theorem
Song et al. ²¹⁸ 2000	Nicht-adjustierter indirekter Vergleich, adjustierter indirekter Vergleich
Spiegelhalter et al. ²²¹ 2004	Mixed treatment comparison
Sutton et al. ²²⁶ 2007	Mixed treatment comparison
Thompson und Sharp ²³⁷ 1999	Metaregression
Thompson und Higgins ²³⁶ 2002	Metaregression
Turner et al. ²⁴⁶ 2000	Metaregression (Mixed model)
Van Houwelingen et al. ²⁵² 2002	Metaregression
Zusätzlich zur Methodenbeschreibung, und somit als Typ-1-Publikation, verwendet werden einige systematische Reviews, die auch Anwendungsbeispiele enthalten und deshalb bei den Typ-4- oder Typ-5-Publikationen aufgelistet sind. Es handelt sich um:	
- Bucher et al. ⁵⁵ (Adjustierter indirekter Vergleich) -> Bei den Typ-4-Publikationen	
- Dominici et al. ⁸⁹ (Mixed treatment comparison) -> Bei den Typ-4-Publikationen	
- Higgins et al. ¹²² (Mixed treatment comparison) -> Bei den Typ-5-Publikationen	

Fortsetzung Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs

Typ-2-Publikationen	
Abou-Setta ¹⁷ 2006	Adjustierter indirekter Vergleich zweier Einzelstudien
Berry ⁴⁴ 2005	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Boland et al. ⁴⁷ 2003	Nicht-adjustierter indirekter Vergleich zweier Einzelstudien
Clegg et al. ⁷² 2002	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Hofmann et al. ¹²⁸ 2003	Adjustierter indirekter Vergleich zweier Einzelstudien
Ijsselmuiden et al. ¹³¹ 2003	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Jones et al. ¹³⁶ 2004	Adjustierter indirekter Vergleich zweier Einzelstudien
Kristensen et al. ¹⁴¹ 2007	Adjustierter und nicht-adjustierter indirekter Vergleich zweier Einzelstudien
Quan et al. ¹⁹⁰ 2006	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Richardson et al. ¹⁹³ 2007	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Roozen et al. ²⁰⁰ 2007	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Swift et al. ²³⁰ 2005	Adjustierter indirekter Vergleich zweier Einzelstudien
Vale et al. ²⁴⁸ 2004	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Verma et al. ²⁵⁵ 2003	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Ward et al. ²⁵⁸ 2007	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Yabroff et al. ²⁶⁶ 2000	Narrativer indirekter Vergleich ohne Metaanalyse(n)
Typ-3-Publikationen	
Keine	
Typ-4-Publikationen	
Abou-Setta ¹⁸ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Adelman und Belsey ¹⁹ 2003	Nicht-adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Bakker et al. ³³ 1998	Nicht-adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Balk et al. ³⁴ 2003	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Barden et al. ³⁶ 2004	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Berner et al. ⁴² 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Berry et al. ⁴³ 2005	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer
Bhandari et al. ⁴⁵ 2001	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Biondi-Zoccai et al. ⁴⁶ 2005	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Publikation von Song 2003 zitiert, was der Methode nach Bucher entspricht)
Bottomley et al. ⁵⁰ 2007	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer
Buscemi et al. ⁵⁸ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Büttner et al. ⁵⁹ 2004	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Capstick et al. ⁶¹ 2005	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer
Chen et al. ⁶³ 2005	Nicht-adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Clark et al. ⁶⁸ 2004	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methodenberatung durch F. Song, was der Methode nach Bucher et al. ⁵⁵ entspricht)
Collins et al. ⁷⁴ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Coomarasamy et al. ⁷⁶ 2003	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Cooper et al. ⁷⁷ 2006	Mixed treatment comparison
Coyle et al. ⁷⁸ 2006	Nicht-adjustierter indirekter Vergleich narrativ

Fortsetzung Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs

Typ-4-Publikationen	
Davies et al. ⁸¹ 2006	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Dominici et al. ⁸⁹ 1999	Mixed treatment comparison
Eckert et al. ⁹² 2006	Metaregression (Mixed model)
Einarson et al. ⁹⁶ 2000	Nicht-adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Farré et al. ⁹⁹	Metaregression
Geddes et al. ¹⁰⁷ 2000	Metaregression (mit Bayes'schem Verfahren)
Habib et al. ¹¹³ 2004	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Hind et al. ¹²⁵ 2003	"Sonstige" Methode nach Moore et al. und Hind et al.
Hochberg et al. ¹²⁷ 2003	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Indolfi et al. ¹³² 2005	Metaregression
Jansen et al. ¹³³ 2006	Mixed treatment comparison
Jones et al. ¹³⁵ 2003	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Kahn et al. ¹³⁹ 2001	Metaregression
Kearney et al. ¹³⁸ 2006	Metaregression
Lancaster und Stead ¹⁴⁵ 2008	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Law et al. ¹⁴⁸ 2003	Adjustierter indirekter Vergleich narrativ
Law et al. ¹⁴⁷ 2003	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Lee et al. ¹⁴⁹ 2005	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Lee et al. ¹⁵⁰ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Lee et al. ¹⁵¹ 2008	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Leucht et al. ¹⁵² 2002	Metaregression
Lim et al. ¹⁵⁶ 2003	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Aus der Publikation wird ersichtlich, dass die verwendete Methode der nach Bucher et al. ⁵⁵ entspricht)
Massel et al. ¹⁶⁶ 2005	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Aus der Publikation wird ersichtlich, dass die verwendete Methode der nach Bucher et al. ⁵⁵ entspricht)
McAlister et al. ¹⁶⁸ 2004	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Publikation von Song 2003 zitiert, was der Methode nach Bucher et al. ⁵⁵ entspricht)
McLeod et al. ¹⁷⁰ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Publikation von Song 2003 zitiert, was der Methode nach Bucher et al. ⁵⁵ entspricht)
Medicare Services Advisory Committee ³ 2000	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer
Messerli et al. ¹⁷¹ 1998	Adjustierter indirekter Vergleich narrativ
Mitte et al. ¹⁷² 2005	Metaregression
Moore et al. ¹⁷⁶ 2005	"Sonstige" Methode nach Moore et al. und Hind et al.
Nixon et al. ¹⁸⁰ 2007	Mixed treatment comparison
Otoul et al. ¹⁸¹ 2005	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher)
Otto et al. ¹⁸² 2001	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Peterson et al. ¹⁸⁴ 2007	Metaregression
Rice et al. ¹⁹² 2000	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Richy et al. ¹⁹⁴ 2008	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests

Fortsetzung Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs

Robinson et al. ¹⁹⁸ 2005	Metaregression (mit Bayes'schem Verfahren)
Roddy et al. ¹⁹⁹ 2005	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Stettler et al. ²²³ 2006	Metaregression
Testa et al. ²³² 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Aus der Publikation wird ersichtlich, dass die verwendete Methode der nach Bucher entspricht)
Testa et al. ²³¹ 2008	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Publikation von Song 2003 zitiert, was der Methode nach Bucher et al. ⁵⁵ entspricht)
Turner et al. ²⁴⁵ 2003	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Van der Heijden et al. ²⁴⁹ 2000	Metaregression
Van der Valk et al. ²⁵⁰ 2005	Nicht-adjustierter indirekter Vergleich narrativ
Van Dongen et al. ²⁵¹ 2003	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Vestergaard et al. ²⁵⁶ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Wehren et al. ²⁵⁹ 2004	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Wilby et al. ²⁶⁰ 2005	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Wilhelmus et al. ²⁶² 2008	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Wilson et al. ²⁶³ 2001	Metaregression
Woolacott et al. ²⁶⁴ 2006	Mixed treatment comparison
Yazdanpanah et al. ²⁶⁷ 2004	Metaregression
Zaremski et al. ²⁶⁸ 1995	Nicht-adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Zhou et al. ²⁷² 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Typ-5-Publikationen	
Antiplatelet Trialists' Collaboration ²⁶ 1994	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Antiplatelet Trialists' Collaboration ²⁵ 1994	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Antiplatelet Trialists' Collaboration ²⁷ 1994	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Antithrombotic Trialists' Collaboration ²⁸ 2002	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Ausejo et al. ³¹ 2000	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Ballesteros ³⁵ 2005	Metaregression (Mixed model)
Bekkering et al. ³⁸ 2008	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Boonen et al. ⁴⁸ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Brophy und Lawrence ⁵¹ 2005	„Sonstige“ Methode nach Brophy und Lawrence
Brown et al. ⁵³ 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Bucher et al. ⁵⁵ 1997	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Cheng et al. ⁶⁴ 2000	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Chiba et al. ⁶⁵ 1997	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen und adjustierter indirekter Vergleich berechnet durch Glenny et al. ¹⁰⁹)

Fortsetzung Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs

Chou et al. ⁶⁶ 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Collins et al. ⁷⁵ 2000	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵) und nicht-adjustierter indirekter Vergleich berechnet durch Glenny et al. ¹⁰⁹)
Delaney et al. ⁸⁵ 2001	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Di Mario et al. ⁸⁷ 1996	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Eckert und Falissard ⁹¹ 2006	Metaregression (Mixed model)
Elliott und Meyer ⁹⁷ 2007	Mixed treatment comparison
Ferrari et al. ¹⁰⁰ 2002	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Gartlehner et al. ¹⁰³ 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Golfinopoulos et al. ¹¹² 2007	Mixed treatment comparison
Handoll et al. ¹¹⁴ 2002	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Higgins und Whitehead ¹²² 1996	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Horn und Limburg ¹²⁹ 2000	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Horn und Limburg ¹³⁰ 2001	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Kyrgiou et al. ¹⁴² 2006	Mixed treatment comparison
Lam und Owen ¹⁴³ 2007	Mixed treatment comparison
Lange et al. ¹⁴⁶ 2003	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher)
Marshall und Irvine ¹⁶³ 1997	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Mason et al. ¹⁶⁵ 2004	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
McIntosh et al. ¹⁶⁹ 2000	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Moore et al. ¹⁷⁵ 1997	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Mudge et al. ¹⁷⁹ 2005	Metaregression
Packer et al. ¹⁸³ 2001	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher) und nicht-adjustierter indirekter Vergleich berechnet durch Glenny et al. ¹⁰⁹)
Po und Zhang ¹⁸⁷ 1998	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Po und Zhang ¹⁸⁶ 1997	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Poynard et al. ¹⁸⁸ 1996	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Psaty et al. ¹⁸⁹ 2003	Mixed treatment comparison
Richy et al. ¹⁹⁵ 2005	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Robenshtok et al. ¹⁹⁶ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher)
Rostom et al. ²⁰¹ 2000	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Sanchez-Ramos et al. ²⁰⁶ 2002	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer
Sauriol et al. ²⁰⁷ 2001	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher) und nicht-adjustierter indirekter Vergleich berechnet durch Glenny et al. ¹⁰⁹)

Fortsetzung Tabelle 21: Eingeschlossene Literatur, sortiert von Typ-1- bis Typ-5-Publikationen mit Zuordnung zu der in der jeweiligen Publikation behandelten oder eingesetzten Methode des indirekten Vergleichs

Silagy et al. ²¹² 2001	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Silagy und Stead ²¹³ 2001	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Small et al. ²¹⁴ 2006	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests
Smith et al. ²⁴⁰ 2007	„Sonstige“ Methode nach Tudur Smith et al.
Soo et al. ²²⁰ 2004	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Stettler et al. ²²⁴ 2007	Mixed treatment comparison
Thijs et al. ²³⁴ 2008	„Sonstige“ Methode nach Thijs et al.
Trindade und Menon ²³⁹ 1997	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Van Pinxteren et al. ²⁵³ 2000	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Vandermeer et al. ²⁵⁴ 2007	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Einmal nach der Methode nach Bucher et al. ⁵⁵ und einmal nach der Methode nach Song et al. 2000 ²¹⁸) und Mixed treatment comparison
Wilhelmus et al. ²⁶¹ 2000	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Wu et al. ²⁶⁵ 2006	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer (Methode nach Bucher et al. ⁵⁵)
Zhang und Po ²⁶⁹ 1996	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)
Zhang und Po ²⁷⁰ 1997	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen
Zhang und Po ²⁷¹ 1998	Adjustierter und nicht-adjustierter indirekter Vergleich (Berechnet durch Glenny et al. ¹⁰⁹)

9.8 Checklisten

9.8.1 Checkliste zur Beschreibung und Charakterisierung von Methoden zur Durchführung von indirekten Vergleichen

Name der Methode	
Erstautor der Methode und Jahr der Veröffentlichung	
Anwendungsgebiete	
Anwendungsvoraussetzungen	
Annahmen	
Effektmaß, mit der der Vergleich durchgeführt wird	
Statistische Standardmethoden, die eingesetzt werden	
Stärken	
Schwächen	
Biasquellen	
Was hat diese Methode an Neuem ergänzt?	
Definition der Variablen	
Rechenweg	

9.8.2 Auswertungsbogen für systematische Reviews mit Metaanalyse(n), die nur einen indirekten Vergleich durchführten

Allgemeines			
Ref-ID			
Erstautor, Jahr			
Effektmaß(e)			
Art des Effektmaßes, mit dem der indirekte Vergleich durchgeführt wird	Diskret	Kontinuierlich	Time-to-event

Methodik				
Methodengruppe				
Random oder Fixed effects-Modell?	Random	Fixed	Beides durchgeführt	
Neue Aspekte der Methode? (Wenn ja, bitte unten weiter ausführen)			Ja	Nein
Adjustierung bzgl. der Kontrollgruppe?			Ja	Nein
Adjustierung bzgl. möglichen Confoundern?			Ja	Nein
Gewichtung der Einzelstudien?	Inverse Varianz	Andere:	Ja	Nein
Multivariate Analyse?			Ja	Nein
Wie wurde die Homogenität der Daten getestet? Wenn ja, mit welcher Methode?			Ja	Nein

Neue Aspekte der Methode

Bemerkungen

9.8.3 Auswertungsbogen für systematische Reviews mit Metaanalyse, die einen direkten und indirekten Vergleich durchführten

Allgemeines			
Ref-ID			
Erstautor, Jahr			
Effektmaß(e)			
Art der Therapieeffektvariablen für indirekten Vergleich	Diskret	Kontinuierlich	Time-to-event
Indirekter Vergleich			
Kurzbeschreibung der Probanden:			
Intervention A versus P	A: versus P:	Intervention B versus P	B: versus P:
Gesamtanzahl Studien		Gesamtanzahl Studien	
Gesamtanzahl Probanden		Gesamtanzahl Probanden	
Anzahl Studien nicht-RCT		Anzahl Studien nicht-RCT	
Einschluss der Head-to-head-Studien in den indirekten Vergleich?	Ja	Nein	
Gepoolter Effekt mit Konfidenzintervall A versus P Fixed effects-Modell		Gepoolter Effekt mit Konfidenzintervall B versus P Fixed effects-Modell	
Gepoolter Effekt mit Konfidenzintervall A versus P Random effects-Modell		Gepoolter Effekt mit Konfidenzintervall B versus P Random effects-Modell	
Therapieeffekt indirekter Vergleich mit Konfidenzintervall Fixed effects			
Therapieeffekt indirekter Vergleich mit Konfidenzintervall Random effects			
Direkter Vergleich			
Kurzbeschreibung der Probanden:			
Intervention A versus B	A: versus B:		
Gesamtanzahl Studien			
Gesamtanzahl Probanden			
Anzahl Studien nicht-RCT			
Therapieeffekt direkter Vergleich mit Konfidenzintervall Fixed effects			
Therapieeffekt direkter Vergleich mit Konfidenzintervall Random effects			
Überschneiden sich die Konfidenzintervalle der Ergebnisse des direkten und des indirekten Vergleichs?			Ja Nein
Erreicht der indirekte Vergleich auch mit Signifikanz die gleiche Schlussfolgerung, wie der direkte Vergleich?			

Methodik			
Methodengruppe			
Neue Aspekte der Methode? (Wenn ja, bitte auf Rückseite weiter ausführen)			Nein
Gewichtung der Einzelstudien?	Inverse Varianz	Andere:	
Adjustierung bzgl. möglichen Confoundern?		Nein	
Multivariate Analyse?		Nein	
Wie wurde die Homogenität der Daten getestet?			
Neue Aspekte der Methode			

Kommentare oder Anmerkungen

9.9 Ergebnistabellen und -abbildungen

9.9.1 Methodische Daten der systematischen Reviews mit indirekten Vergleichen

Tabelle 22: Methodische Daten systematischer Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen); geordnet nach der verwendeten Methode

Nummer	Erstgenannter Autor	Jahr	Effektschätzer	Art der Daten	Random oder Fixed effects-Modelle?	Adjustiert bzgl. Kontrollgruppe?
Nicht-adjustierter indirekter Vergleich						
1	Berry ⁴³	2005	OR	Binär	Fixed effects	Nein
2	Bottomley	2007	WMD	Kontinuierlich	Nicht berichtet	Nein
3	Capstick	2005	RR, NNT	Binär	Random effects	Nein
4	Medicare Services Advisory Committee	2000	OR, ARR	Binär	Random effects	Nein
5	Sanchez-Ramos	2002	OR, WMD	Binär, Kontinuierlich	Zwei-Schritt-Analyse	Nein
6	Adelman	2003	NNT	Binär	Nicht berichtet	Nein
7	Bakker	1998	SMD (Cohen's d)	Kontinuierlich	Nicht berichtet	Nein
8	Chen	2005	WMD	Kontinuierlich	Nicht berichtet	Nein
9	Einarson	2000	SMD (Glass's Δ)	Kontinuierlich	Random effects	Nein
10	Zaremski	1995	ARR	Binär	Fixed effects	Nein
11	Chiba	1997	Heilungsrate	Binär	Nicht berichtet	Nein
12	van Dongen	2003	Inzidenzrate	Binär	Nicht berichtet	Nein
13	Coyle	2006	RR	Binär	Fixed effects	Nein
14	van der Valk	2005	MD	Kontinuierlich	Random effects	Nein
Adjustierter indirekter Vergleich						
1	Vandermeer	2007	SMD	Binär	Random effects	Ja
2	Boonen	2007	RR, NNT	Binär	Beides berechnet	Ja
3	Lee	2007	RR	Binär	Nicht berichtet	Ja
4	Gartlehner	2006	RR	Binär	Random effects	Ja
5	Vestergaard	2007	RR	Binär	Zwei-Schritt-Analyse	Ja
6	Zhou	2006	RR	Binär	Random effects	Ja

Fortsetzung Tabelle 22: Methodische Daten systematischer Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen); geordnet nach der verwendeten Methode

Adjustierter indirekter Vergleich						
7	Wehren ²⁵⁹	2004	RR	Binär	Random effects	Ja
8	Buscemi	2007	RR, WMD	Binär, Kontinuierlich	Random effects	Ja
9	Hochberg	2003	RR	Binär	Fixed effects	Ja
10	Law ¹⁴⁷	2003	WMD	Kontinuierlich	Random effects	Ja
11	Chou	2006	OR	Binär	Random effects	Ja
12	Brown	2006	RR	Binär	Random effects	Ja
13	Packer	2001	WMD	Kontinuierlich	Beides berechnet	Ja
14	Bhandari	2001	RR	Binär	Random effects	Ja
15	Berner	2006	WMD	Kontinuierlich	Fixed effects	Ja
16	Otoul	2005	OR	Binär	Beides berechnet	Ja
17	Coomarasamy	2003	OR	Binär	Random effects	Ja
18	Abou-Setta	2007	OR	Binär	Fixed effects	Ja
19	Wilhelmus	2000	OR	Binär	Zwei-Schritt-Analyse	Ja
20	Collins	2007	HR	Time-to-event	Beides berechnet	Ja
21	Lange	2003	RR, SMD	Binär, Kontinuierlich	Zwei-Schritt-Analyse	Ja
22	Sauriol	2001	SMD	Kontinuierlich	Zwei-Schritt-Analyse	Ja
23	Wu	2006	OR	Binär	Random effects	Ja
24	Lee	2008	RR	Binär	Random effects	Ja
25	Robenshtok	2007	RR	Binär	Zwei-Schritt-Analyse	Ja
26	Wilhelmus	2008	OR	Binär	Fixed effects	Ja
27	Bekkering	2008	OR, WMD	Binär, Kontinuierlich	Random effects	Ja
28	Clark	2004	RR	Binär	Fixed effects	Ja
29	McAliste	2004	RR	Binär	Nicht berichtet	Ja
30	Biondi-Zoccai	2005	OR	Binär	Random effects	Ja
31	McLeod	2007	RR, WMD	Binär, Kontinuierlich	Zwei-Schritt-Analyse	Ja
32	Testa	2008	OR	Binär	Beides berechnet	Ja
33	Testa	2007	OR, NNT	Binär	Zwei-Schritt-Analyse	Ja
34	Massel	2005	OR	Binär	Fixed effects	Ja
35	Lim	2003	RR	Binär	Nicht berichtet	Ja
36	Vandermeer	2007	SMD	Binär	Random effects	Ja
37	Rice	2000	RR	Binär	Fixed effects	Ja
38	Small	2006	RR	Binär	Random effects	Ja
39	Lee	2005	SMD (Cohen's d)	Kontinuierlich	Random effects	Ja
40	Otto	2001	SMD (Glass's Δ), MD	Kontinuierlich	Fixed effects	Ja
41	Richy	2005	RR, SMD (Cohen's d)	Binär, Kontinuierlich	Zwei-Schritt-Analyse	Ja
42	Mason	2004	RR, NNT	Binär	Fixed effects	Ja
43	Antithrombotic Trialists' Collaboration	2002	OR	Binär	Nicht berichtet	Ja
44	Richy	2008	RR	Binär	Fixed effects	Ja
45	Turner	2003	NNT	Time-to-event	Random effects	Ja
46	Zhang	1997	RR, NNT, SMD	Binär, Kontinuierlich	Random effects	Ja
47	Habib	2004	RR, NNT	Binär	Random effects	Ja
48	Büttner	2004	OR, RR, NNT	Binär	Fixed effects	Ja
49	Barden	2004	RR, NNT	Binär	Fixed effects	Ja
50	Davies	2006	RR, WMD	Binär, Kontinuierlich	Random effects	Ja
51	Lancaster	2008	OR	Binär	Fixed effects	Ja
52	Balk	2003	MD	Kontinuierlich	Random effects	Ja

Fortsetzung Tabelle 22: Methodische Daten systematischer Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen); geordnet nach der verwendeten Methode

Adjustierter indirekter Vergleich						
53	Ferrari	2002	ARR, WMD	Kontinuierlich	Random effects	Ja
54	Roddy	2005	WMD	Kontinuierlich	Beides berechnet	Ja
55	Horn	2000	OR	Binär	Fixed effects	Ja
56	Jones	2003	OR, SMD	Binär, Kontinuierlich	Fixed effects	Ja
57	Wilby	2005	RR, HR	Binär, Time-to-event	Fixed effects	Ja
58	Po	1998	WMD	Kontinuierlich	Zwei-Schritt-Analyse	Ja
59	Law ¹⁴⁸	2003	WMD	Kontinuierlich	Nicht berichtet	Ja
60	Messerli	1998	OR	Binär	Fixed effects	Ja
Metaregression						
1	Mitte	2005	SMD (Hedges' g)	Binär	Fixed effects	Unklar
2	van der Heijden	2000	OR	Binär	Fixed effects	Ja
3	Kearney	2006	RR	Binär	Fixed effects	Ja
4	Leucht	2002	Korrelationskoeffizient	Kontinuierlich	Random effects	Unklar
5	Mudge	2005	OR, WMD	Binär, Kontinuierlich	Random effects	Ja
6	Peterson	2007	RR	Binär	Random effects	Ja
7	Khan	2000	ARR	Kontinuierlich	Random effects	Unklar
8	Yazdanpanah	2004	OR	Binär	Random effects	Unklar
9	Stettler	2006	Inzidenzratenratio	Binär	Random effects	Ja
10	Indolfi	2005	RR	Binär	Random effects	Ja
11	Wilson	2001	OR	Binär	Nicht berichtet	Ja
12	Farré	2002	OR	Binär	Random effects	Ja
13	Geddes	2000	SMD (Hedge's g)	Kontinuierlich	Fixed effects	Ja
14	Robinson	2005	RR	Binär	Random effects	Ja
15	Eckert ⁹¹	2006	OR, SMD (Hedge's g)	Binär, Kontinuierlich	Random effects	Ja
16	Eckert ⁹²	2006	SMD (Hedge's g)	Kontinuierlich	Random effects	Ja
17	Ballesteros	2005	OR, RR, ARR	Binär	Nicht berichtet	Ja
MTC						
1	Vandermeer	2007	ARR, SMD	Kontinuierlich	Random effects	Ja
2	Golfiopoulous	2007	HR	Time-to-event	Nicht berichtet	Ja
3	Lam	2007	OR	Binär	Random effects	Ja
4	Jansen ¹³³	2006	WMD	Kontinuierlich	Random effects	Ja
5	Psaty	2003	RR	Binär	Random effects	Ja
6	Kyrgiou	2006	HR	Time-to-event	Random effects	Ja
7	Cooper	2006	RR	Binär	Random effects	Ja
8	Elliott	2007	OR	Binär	Random effects	Ja
9	Nixon ¹⁸⁰	2007	OR	Binär	Random effects	Ja
10	Stettler	2007	RR, HR	Binär	Random effects	Ja
11	Woolacott	2006	RR, WMD	Binär, Kontinuierlich	Fixed effects	Ja
12	Dominici	1999	SMD	Kontinuierlich	Random effects	Ja
Sonstige Methoden						
1	Hind	2003	RR	Binär	Nicht berichtet	unklar
2	Moore	2005	RR, NNT, WMD	Binär, Kontinuierlich	Fixed effects	unklar
3	Brophy	2005	OR	Binär	Nicht berichtet	unklar
4	Tudur Smith	2007	HR	Time-to-event	Fixed effects	unklar
5	Thijs	2008	OR	Binär	Zwei-Schritt-Analyse	unklar

ARR = Absolute Risikoreduktion. HR = Hazard ratio. MD = Mittelwertsdifferenz. MTC = Mixed treatment comparison. NNT = Number-needed-to-treat. OR = Odds ratio. RR = Relatives Risiko. SMD = Standardisierte mittlere Differenz. WMD = Gewichtete mittlere Differenz.

9.9.2 Heterogenitätsbetrachtung in systematischen Reviews mit indirekten Vergleichen

Tabelle 23: Verwendete Verfahren zur Ermittlung von Heterogenität in systematischen Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen)

Erstautor	Jahr	Wurde ein Verfahren berichtet, das die Homogenität der Daten untersucht?	Werden ein oder mehrere statistische Tests eingesetzt?	Wird der Einsatz weiterer Verfahren berichtet?
Abou-Setta	2007	Ja	I ² -Test	Nein
Adelman	2003	Nein	Nicht berichtet	Nein
Antithrombotic Trialists' Collaboration	2002	Ja	Chi ² -Test	Nein
Bakker	1998	Ja	Kruskal-Wallis-Test	Nein
Balk	2003	Ja	Ja, Test nicht benannt	Nein
Ballesteros	2005	Ja	Chi ² -Test	Nein
Barden	2004	Nein	Nicht berichtet	Nein
Bekkering	2008	Ja	Chi ² -Test, I ² -Test	Forest-Plot
Berner	2006	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse
Berry ⁴³	2005	Nein	Nicht berichtet	Nein
Bhandari	2001	Ja	Breslow-Day-Test	Nein
Biondi-Zoccai	2005	Ja	I ² -Test	Nein
Boonen	2007	Ja	Q-Test, I ² -Test	Sensitivitätsanalyse, Funnel-Plot
Bottomley	2007	Ja	Nicht berichtet	Beurteilung der Vergleichbarkeit der Studien (Klinische und methodische Homogenität)
Brophy	2005	Nein	Nicht berichtet	Nein
Brown	2006	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse, visueller Vergleich der Konfidenzintervalle
Buscemi	2007	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse
Büttner	2004	Ja	Ja, Test nicht benannt	Nein
Capstick	2005	Ja	Chi ² -Test	Sensitivitätsanalyse
Chen	2005	Ja	Chi ² -Test, Breslow-Day-Test	Nein
Chiba	1997	Nein	Nicht berichtet	Nein
Chou	2006	Ja	Q-Test, I ² -Test	Nein
Clark	2004	Ja	Chi ² -Test	Sensitivitätsanalyse
Collins	2007	Ja	Chi ² -Test	Beurteilung der Vergleichbarkeit der Studien (Klinische und methodische Homogenität)
Coomarasamy	2003	Ja	Chi ² -Test	Nein
Cooper	2006	Nein	Nicht berichtet	Sensitivitätsanalyse
Coyle	2006	Nein	Nicht berichtet	Nein
Davies	2006	Ja	Q-Test	Nein
Dominici	1999	Nein	Nicht berichtet	Nein
Eckert ⁹¹	2006	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse
Eckert ⁹²	2006	Ja	Chi ² -Test	Nein
Einarson	2000	Ja	Chi ² -Test, Box's Variante des Barlett-Tests	Nein

Fortsetzung Tabelle 23: Verwendete Verfahren zur Ermittlung von Heterogenität in systematischen Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen)

Erstautor	Jahr	Wurde ein Verfahren berichtet, das die Homogenität der Daten untersucht?	Werden ein oder mehrere statistische Tests eingesetzt?	Wird der Einsatz weiterer Verfahren berichtet?
Elliott	2007	Ja	Riley-Day-Test, Inkonsistenztest nach Lumley	Sensitivitätsanalyse
Farré	2002	Ja	I ² -Test	Nein
Ferrari	2002	Ja	Chi ² -Test	Nein
Gartlehner	2006	Ja	I ² -Test	Nein
Geddes	2000	Ja	Nicht berichtet	Sensitivitätsanalyse
Golfinopoulos	2007	Ja	I ² -Test	Nein
Habib	2004	Ja	Ja, Test nicht benannt	Nein
Hind	2003	Nein	Nicht berichtet	Nein
Hochberg	2003	Ja	Chi ² -Test	Nein
Horn	2000	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse
Indolfi	2005	Ja	Nicht berichtet	Sensitivitätsanalyse
Jansen ¹³³	2006	Ja	Nicht berichtet	Beurteilung der Vergleichbarkeit der Studien (Methodische Homogenität)
Jones	2003	Ja	Q-Test	Nein
Kearney	2006	Ja	Chi ² -Test	Nein
Khan	2000	Ja	Q-Test	Nein
Kyrgiou	2006	Ja	I ² -Test	Sensitivitätsanalyse
Lam	2007	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse, L'Abbè-Plot
Lancaster	2008	Ja	I ² -Test	Nein
Lange	2003	Ja	Chi ² -Test	Nein
Law ¹⁴⁸	2003	Nein	Nicht berichtet	Nein
Law ¹⁴⁷	2003	Nein	Nicht berichtet	Nein
Lee	2007	Nein	Nicht berichtet	Nein
Lee	2005	Ja	Ja, Test nicht benannt	Sensitivitätsanalyse
Lee	2008	Ja	Q-Test	Nein
Leucht	2002	Ja	Chi ² -Test, I ² -Test	L'Abbè-Plot, Sensitivitätsanalyse
Lim	2003	Nein	Nicht berichtet	Nein
Mason	2004	Ja	Chi ² -Test, I ² -Test	L'Abbè-Plot, Sensitivitätsanalyse
Massel	2005	Ja	Chi ² -Test, I ² -Test	Nein
McAlister	2004	Ja	Q-Test	Sensitivitätsanalyse
McLeod	2007	Ja	Chi ² -Test, I ² -Test	Nein
Medical Services Advisory Committee	2000	Ja	Ja, Test nicht benannt	Nein
Messerli	1998	Ja	Comfied-Gard-Test	Nein
Mitte	2005	Nein	Nicht berichtet	Nein
Moore	2005	Ja	Nicht berichtet	Grafische Methode
Mudge	2005	Nein	Nicht berichtet	Nein
Nixon ¹⁸⁰	2007	Ja	Ja, Test nicht benannt	Nein
Otoul	2005	Ja	Breslow-Day-Test	Nein
Otto	2001	Ja	Nicht berichtet	Sensitivitätsanalyse
Packer	2001	Ja	Chi ² -Test	Sensitivitätsanalyse

Fortsetzung Tabelle 23: Verwendete Verfahren zur Ermittlung von Heterogenität in systematischen Reviews mit indirekten Vergleichen (Typ-4- und Typ-5-Publikationen)

Erstautor	Jahr	Wurde ein Verfahren berichtet, das die Homogenität der Daten untersucht?	Werden ein oder mehrere statistische Tests eingesetzt?	Wird der Einsatz weiterer Verfahren berichtet?
Peterson	2007	Ja	Chi ² -Test	Nein
Po	1998	Ja	Chi ² -Test	Nein
Psaty	2003	Ja	Inkonsistenztest nach Lumley	Nein
Rice	2000	Ja	I ² -Test	Nein
Richy	2005	Ja	Ja, Test nicht benannt	Nein
Richy	2008	Ja	Varianzanalyse (ANOVA-1)	Sensitivitätsanalyse
Robenshtok	2007	Ja	Chi ² -Test, I ² -Test	Sensitivitätsanalyse, Metaregression
Robinson	2005	Ja	Ja, Test nicht benannt	Nein
Roddy	2005	Ja	Q-Test	Nein
Sanchez-Ramos	2002	Ja	Breslow-Day-Test	L'Abbè-Plot, Sensitivitätsanalyse
Sauriol	2001	Ja	Q-Test	Nein
Small	2006	Nein	Nicht berichtet	Nein
Smith	2007	Ja	Chi ² -Test, I ² -Test	Visueller Vergleich der Konfidenzintervalle
Stettler	2006	Ja	I ² -Test, Between-Study-Varianztest (nicht näher bezeichnet), Inkonsistenztest nach Lu und Ades	Sensitivitätsanalyse, Test auf Geeignetheit des Modells mit Residuale-Varianz-Methode
Stettler	2007	Nein	Nicht berichtet	Nein
Testa	2007	Ja	Chi ² -Test, I ² -Test	Nein
Testa	2008	Ja	Q-Test, I ² -Test	Nein
Thijs	2008	Ja	I ² -Test und H-Test	Nein
Turner	2003	Nein	Nicht berichtet	Nein
Van der Heijden	2000	Ja	Comfierd-Gard-Test	Nein
Van der Valk	2005	Nein	Nicht berichtet	Nein
Van Dongen	2003	Ja	Chi ² -Test	Nein
Vandermeer	2007	Ja	Q-Test	Nein
Vestergaard	2007	Ja	Chi ² -Test	Funnel-Plots
Wehren ²⁵⁹ und Cranney ⁷⁹	2004	Ja	Chi ² -Test	Nein
Wilby	2005	Ja	Q-Test	Nein
Wilhelmus	2000	Ja	Chi ² -Test	Nein
Wilhelmus	2008	Ja	Chi ² -Test, I ² -Test	Nein
Wilson	2001	Ja	Ja, Test nicht benannt	Sensitivitätsanalyse
Woolacott	2006	Ja	Chi ² -Test	Nein
Wu	2006	Ja	I ² -Test	Metaregression
Yazdanpanah	2004	Ja	Chi ² -Test	Nein
Zarembski	1995	Nein	Nicht berichtet	Nein
Zhang	1997	Ja	Chi ² -Test	Nein
Zhou	2006	Ja	Chi ² -Test	L'Abbè-Plot

ANOVA = Analysis of variance

Tabelle 24: Umgang mit vorliegender Heterogenität in den Typ-5-Publikationen

Erstautor	Jahr	Lag statistisch signifikante Heterogenität vor?	Random oder Fixed effects-Modelle?	Ausmaß der Heterogenität diskutiert?	Wie wurde mit Heterogenität umgegangen?
Antithrombotic Trialists' Collaboration	2002	Ja	Nicht berichtet	Ja	Das Vorliegen von Heterogenität war erwünscht, da durch signifikante Heterogenität signifikante Therapieeffektunterschiede zwischen Populationen oder Interventionen aufgedeckt werden sollten.
Ballesteros.	2005	Nein	Nicht berichtet	Nein	Keine besonderen Maßnahmen berichtet
Bekkering	2008	Ja	Random effects	Ja	Es wurden Metaregressionen durchgeführt, um die Heterogenität zu erklären.
Boonen	2007	Nein	Beides berechnet	Nein	Es lag keine Heterogenität vor
Brophy	2005	Heterogenität nicht untersucht	Nicht berichtet	Nein	Heterogenität nicht untersucht
Brown	2006	Ja	Random effects	Ja	Es wurden Metaregressionen durchgeführt, um die Heterogenität zu erklären.
Chiba	1997	Heterogenität nicht untersucht	Nicht berichtet	Nein	Heterogenität nicht untersucht
Chou	2006	Ja	Random effects	Ja	Es wurden Sensitivitätsanalysen und Metaregressionen durchgeführt, um die Heterogenität zu erklären.
Eckert ⁹¹	2006	Ja	Random effects	Nein	Keine besonderen Maßnahmen berichtet
Elliott	2007	Ergebnis des Heterogenitätstests nicht berichtet	Random effects	Ja	Die Heterogenität wurde über Sensitivitätsanalysen aufgeklärt.
Ferrari	2002	Ja	Random effects	Nein	Es wurden Sensitivitätsanalysen für Ausreißerstudien durchgeführt, aber keine Maßnahmen zur Reduktion der Heterogenität ergriffen, sondern darauf verwiesen, dass ein Random effects-Modell verwendet worden sei.
Gartlehner	2006	Ergebnis des Heterogenitätstests nicht berichtet	Random effects	Nein	Keine besonderen Maßnahmen berichtet
Golfinopoulos	2007	Ja	Nicht berichtet	Ja	Keine besonderen Maßnahmen berichtet
Horn	2000	Ja	Fixed effects	Ja	Es wurden Sensitivitätsanalysen für Ausreißerstudien durchgeführt, aber keine Maßnahmen zur Reduktion der Heterogenität ergriffen
Kyrgiou	2006	Ja	Random effects	Ja	Subgruppenanalyse für Therapien erster und zweiter Wahl durchgeführt.

Fortsetzung Tabelle 24: Umgang mit vorliegender Heterogenität in den Typ-5-Publikationen

Erstautor	Jahr	Lag statistisch signifikante Heterogenität vor?	Random oder Fixed effects-Modelle?	Ausmaß der Heterogenität diskutiert?	Wie wurde mit Heterogenität umgegangen?
Lam	2007	Ergebnis des Heterogenitätstests nicht berichtet	Random effects	Nein	Es wurde eine Subgruppenanalyse für Patienten mit schwerer Herzinsuffizienz durchgeführt und die Analyse stratifiziert nach den Interventionen in den Kontrollgruppen (Pharmakotherapie oder implantierbarer Herzschrittmacher).
Lange	2003	Ja	Zwei-Schritt-Analyse	Ja	In den Therapievergleichen, in denen signifikante Heterogenität vorlag, wurde auf den indirekten Vergleich verzichtet (Cromone versus Glucocorticoide und Cromone versus Antihistaminika)
Mason	2004	Ergebnis des Heterogenitätstests nicht berichtet	Fixed effects	Nein	Keine besonderen Maßnahmen berichtet
Mudge	2005	Heterogenität nicht untersucht	Random effects	Nein	Heterogenität nicht untersucht
Packer	2001	Nein	Beides berechnet	Nein	Es lag keine Heterogenität vor
Po	1998	Ja	Zwei-Schritt-Analyse	Nein	Keine besonderen Maßnahmen berichtet
Psaty	2003	Ja	Random effects	Ja	Keine besonderen Maßnahmen berichtet
Richy	2005	Ja	Zwei-Schritt-Analyse	Nein	Es wurde a priori festgelegt, dass wenn die Heterogenität nur durch eine Studie verursacht werden würde, diese auszuschließen und ein Fixed effects-Modell zu verwenden. Sollte die Heterogenität durch zwei oder mehr Studien verursacht werden, soll ein Random effects-Modell eingesetzt werden. Das erstere war der Fall.
Robenshtok	2007	Ja	Zwei-Schritt-Analyse	Nein	Die Heterogenität wurde über Sensitivitätsanalysen für Studienqualitäts- und Populationscharakteristika aufgeklärt. Wahl des Random effects-Modells wird mit vorliegender Heterogenität begründet.
Sanchez-Ramos	2002	Ja	Zwei-Schritt-Analyse	Ja	Random effects-Modell aufgrund vorliegender Heterogenität im direkten Vergleich eingesetzt.
Sauriol	2001	Ja	Zwei-Schritt-Analyse	Ja	Aufgrund der vorliegenden Heterogenität wurde deutlich empfohlen, die Ergebnisse mit Vorsicht zu interpretieren.
Small	2006	Heterogenität nicht untersucht	Random effects	Nein	Heterogenität nicht untersucht

Fortsetzung Tabelle 24: Umgang mit vorliegender Heterogenität in den Typ-5-Publikationen

Erstautor	Jahr	Lag statistisch signifikante Heterogenität vor?	Random oder Fixed effects-Modelle?	Ausmaß der Heterogenität diskutiert?	Wie wurde mit Heterogenität umgegangen?
Smith	2007	Ja	Fixed effects	Ja	Es wurden Sensitivitätsanalysen für Ausreißerstudien durchgeführt, aber keine Maßnahmen zur Reduktion der Heterogenität ergriffen. Es wurden die Ergebnisse des indirekten Vergleichs mit den Ergebnissen von Head-to-head-Studien verglichen.
Stettler	2007	Nein	Random effects	Ja	Es lag keine Heterogenität vor nachdem bezüglich der unterschiedlichen Nachbeobachtungszeiten adjustiert und nach Dauer bis zur Stent-Thrombose und der Co-Diagnose Diabetes mellitus stratifiziert untersucht wurde.
Thijs	2008	Ja	Zwei-Schritt-Analyse	Ja	Es wurden die Richtung des Effekts in den einzelnen Studien betrachtet, die Robustheit des Modells getestet und eine Sensitivitätsanalyse durchgeführt.
Vandermeer	2007	Ergebnis des Heterogenitätstests nicht berichtet	Random effects	Nein	Keine besonderen Maßnahmen berichtet
Wilhelmus	2000	Ja	Zwei-Schritt-Analyse	Ja	Bei hoher Heterogenität sollte der indirekte Vergleich nicht durchgeführt werden.
Wu	2006	Ja	Random effects	Nein	Die Heterogenität wurde über Sensitivitätsanalysen aufgeklärt. Es wurden keine Maßnahmen zur Reduktion der Heterogenität berichtet.
Zhang	1997	Ja	Random effects	Nein	Keine besonderen Maßnahmen berichtet

Tabelle 25: Ausmaß der statistischen Heterogenität in Reviews (Typ-5-Publikationen) mit signifikanter Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs

Metaanalyse- nummer (Review, dem sie entstammt)	Indirekter Vergleich			Direkter Vergleich	
	Heterogenität in Studien mit A versus X	Heterogenität in Studien mit B versus X	Gesamtbeurteilung der Heterogenität im indirekten Vergleich	Heterogenität in Studien mit A versus B	Gesamtbeurteilung der Heterogenität im direkten Vergleich
Diskrepanze Datensätze mit nicht-adjustiertem indirektem Vergleich					
1 (Sanchez-Ramos ²⁰⁶)	Nur eine Studie	Nur eine Studie	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Im Fixed effect-Modell nach Mantel-Haenszel liegt signifikante Heterogenität vor (Ergebnis des Chi ² -Tests nicht berichtet).	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
7 (Antiplatelet Trialists' Collaboration 1994 ²⁶ , Glenny et al. ¹⁰⁹)	Nicht exakt für den Vergleich Hochdosis-Aspirin (A) versus mittelhochdosiertes Aspirin (B) berichtet. Zwischen zwei Aspirinkategorien ist Heterogenität nicht signifikant (Chi ² = 0,4)		=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nur eine Studie	=> Es liegt keine signifikante Heterogenität im direkten Vergleich vor.
9 (Antiplatelet Trialists' Collaboration 1994 ²⁶ , Glenny et al. ¹⁰⁹)	Nicht exakt für den Vergleich Aspirin + Dipyramidol (A) versus Aspirin (B) berichtet. Zwischen Aspirin und allen fünf anderen Medikamenten ist Heterogenität nicht signifikant (Chi ² = 1,6)		=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nicht berichtet. Einzige Angabe: Es gibt geringe signifikante Heterogenität (Chi ² bei 2 df = 7,6; p = 0,02) zwischen den Effekten einer Thrombozytenaggregationstherapie in den drei verschiedenen Patientenkategorien.	=> Ausmaß der Heterogenität im direkten Vergleich unklar
11 (Antiplatelet Trialists' Collaboration 1994 ²⁵ , Glenny et al. ¹⁰⁹)	„Die formalen statistischen Vergleiche stellen fest, dass keine signifikante Heterogenität zwischen den protektiven Effekten der sieben verschiedenen Arzneistoffkombinationen vorliegt“. Zwischen den drei Aspirin-Kategorien ist Heterogenität nicht signifikant (Chi ² = 2,7)		=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nicht berichtet	=> Ausmaß der Heterogenität im indirekten Vergleich unklar
18 (Chiba ⁶⁵ , Glenny et al. ¹⁰⁹)	Heterogenität statistisch nicht erfasst.	Heterogenität statistisch nicht erfasst	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Heterogenität statistisch nicht erfasst.	=> Ausmaß der Heterogenität im direkten Vergleich unklar
24 (Marshall ¹⁶³ , Glenny et al. ¹⁰⁹)	Ergebnis des Heterogenitätstests (Breslow-Day-Test) nicht berichtet und auch nicht diskutiert		=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Ergebnis des Heterogenitätstests (Breslow-Day-Test) nicht berichtet und auch nicht diskutiert	=> Ausmaß der Heterogenität im direkten Vergleich unklar
25 (McIntosh ¹⁶⁹ , Glenny et al. ¹⁰⁹)	Nur eine Studie (n = 49)	Nur eine Studie (n = 86)	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 39)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
30 (Rostom ²⁰¹ , Glenny et al. ¹⁰⁹)	Chi ² = 0,61; df = 2; p = 0,74; I ² = 0 %	Chi ² = 0,39; df = 2; p = 0,83; I ² = 0 %	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n=425)	=> Es liegt keine Heterogenität im direkten Vergleich vor.

Fortsetzung Tabelle 25: Ausmaß der statistischen Heterogenität in Reviews (Typ-5-Publikationen) mit signifikanter Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs

34 (Silagy 2001 ²¹³ ; Glenny)	Chi ² = 7,13; df = 5; p = 0,21; I ² = 29,9 %	Chi ² = 25,96; df = 17; p = 0,08*; I ² = 34,5 %	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Chi ² = 1,72; df = 4; p = 0,79; I ² = 0,0 %	=> Es liegt keine signifikante Heterogenität im direkten Vergleich vor.
39 (Trindade ²³⁹ ; Glenny)	Heterogenität nicht statistisch erfasst.	Heterogenität nicht statistisch erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nur eine Studie (n = 97)	=> Es liegt keine signifikante Heterogenität im direkten Vergleich vor.
43 (Zhang 1998 ²⁷¹ ; Glenny et al. ¹⁰⁹)	Keine signifikante Heterogenität. Zwar wird das Ergebnis des Q-Tests nicht berichtet, aber es wird ein Fixed effects-Modell verwendet und dies sollte nur angewendet werden, wenn keine signifikante Heterogenität vorliegt.	Keine signifikante Heterogenität. Zwar wird das Ergebnis des Q-Tests nicht berichtet, aber es wird ein Fixed effects-Modell verwendet und dies sollte nur angewendet werden, wenn keine signifikante Heterogenität vorliegt.	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 44)	=> Es liegt keine signifikante Heterogenität im direkten Vergleich vor.
47 (Zhang 1996 ²⁶⁹ ; Glenny et al. ¹⁰⁹)	Chi ² = 29,94; df = 12; p ≤ 0,01*	Chi ² = 49,54; df = 36; p > 0,05	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Chi ² = 30,0; df = 12; p ≤ 0,01*	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
Diskrepante Datensätze mit adjustiertem indirekten Vergleich					
66 (Brown ⁵³)	Nicht berechnet	Nicht berechnet	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nur eine Studie (n = 287)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
74 (Brown ⁵³)	Nicht berechnet	Nicht berechnet	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nur eine Studie (n = 287)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
81 (Chiba ⁶⁵ ; Glenny et al. ¹⁰⁹)	Heterogenität statistisch nicht erfasst.	Heterogenität statistisch nicht erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Heterogenität statistisch nicht erfasst.	=> Ausmaß der Heterogenität im direkten Vergleich unklar
83 (Chou ⁶⁶)	p = 0,02*; I ² = 63,5 %	p < 0,0005*; I ² = 85,8 %	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Chi ² = 18,0; df = 11; p = 0,08*; I ² = 38,9 %	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
100 (Rostom ²⁰¹ ; Glenny et al. ¹⁰⁹)	Chi ² = 0,61; df = 2; p = 0,74; I ² = 0 %	Chi ² = 0,39; df = 2; p = 0,83; I ² = 0 %	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 425)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
115 (Wilhelmus ²⁶¹)	p = 0,06*	p = 0,03*	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	P = 0,16	=> Es liegt keine Heterogenität im direkten Vergleich vor.

Fortsetzung Tabelle 25: Ausmaß der statistischen Heterogenität in Reviews (Typ-5-Publikationen) mit signifikanter Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs

116 (Wilhelmus ²⁶¹)	p = 0,06*	p = 0,16	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	p = 0,03*	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
118 (Wilhelmus ²⁶¹)	p = 0,46	p = 0,15	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	p = 0,16	=> Es liegt keine Heterogenität im direkten Vergleich vor.
119 (Wu ²⁶⁵)	p = 0,01*; I ² = 53,6 %	p < 0,001*; I ² = 57,6 %	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 60)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
125 (Zhang 1996 ²⁶⁹ , Glennly et al. ¹⁰⁹)	Chi ² = 29,94; df = 12; p ≤ 0,01*	Chi ² = 49,54; df = 36; p > 0,05	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Chi ² = 30,0; df = 12; p ≤ 0,01*	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
126 (Vandermeer ²⁵⁴)	Nicht berichtet	Nicht berichtet	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nicht berichtet	=> Ausmaß der Heterogenität im direkten Vergleich unklar
148 (Ferrari ¹⁰⁰)	p > 0,10	p > 0,10	=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 551)	=> Es liegt keine Heterogenität im direkten Vergleich vor
150 (Ferrari ¹⁰⁰)	p > 0,10	p < 0,10*	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Nicht berichtet	=> Ausmaß der Heterogenität im direkten Vergleich unklar
164 (Zhang ²⁷⁰)	Nur eine Studie (n = 141)	Chi ² = 35,28; p ≤ 0,05*	=> Es liegt signifikante Heterogenität im indirekten Vergleich vor.	Nur eine Studie (n = 134)	=> Es liegt keine Heterogenität im direkten Vergleich vor.
Diskrepanze Datensätze mit indirekten Vergleich durch Metaregression					
170 (Ballesteros ³⁵)	Für den gesamten Studiensatz des systematischen Reviews: Chi ² = 10,62; df = 9; p = 0,93		=> Es liegt keine signifikante Heterogenität im indirekten Vergleich vor.	Chi ² = 0,39; df = 1; p = 0,53	=> Es liegt keine Heterogenität im direkten Vergleich vor.
Diskrepanze Datensätze mit indirekten Vergleich durch eine Mixed treatment comparison					
198 (Psaty et al. ¹⁸⁹)	Im gesamten Evidenznetzwerk liegt für den untersuchten Therapieeffekts signifikante Heterogenität vor (p < 0,001). Ergebnis eines Heterogenitätstests für einzelne Therapieeffektvergleiche der Arzneistoffgruppen werden nicht berichtet.		=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Nur eine Studie (n = 6.083)	=> Es liegt keine Heterogenität im direkten Vergleich vor.

Fortsetzung Tabelle 25: Ausmaß der statistischen Heterogenität in Reviews (Typ-5-Publikationen) mit signifikanter Diskrepanz zwischen den Ergebnissen des direkten und indirekten Vergleichs

Diskrepanz Datensätze mit indirekten Vergleich durch sonstige Methoden				
231 (Tudur Smith ^{240, 243})	Heterogenität bzw. Inkonsistenz statistisch nicht erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Chi ² = 7,76; df = 3; p = 0,05*; I ² = 61,3 %	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
239 (Tudur Smith ²⁴⁰ und Marson ¹⁶⁴)	Heterogenität bzw. Inkonsistenz statistisch nicht erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Chi ² = 15,79; df = 4; p = 0,003*; I ² = 74,7 %	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
240 (Tudur Smith ²⁴⁰ und Marson ¹⁶⁴)	Heterogenität bzw. Inkonsistenz statistisch nicht erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Chi ² = 7,80; df = 3; p = 0,05*; I ² = 61,3 %	=> Es liegt signifikante Heterogenität im direkten Vergleich vor.
242 (Tudur Smith ²⁴⁰ und Marson ¹⁶⁴)	Heterogenität bzw. Inkonsistenz statistisch nicht erfasst.	=> Ausmaß der Heterogenität im indirekten Vergleich unklar	Chi ² = 2,58; df = 3; p = 0,46; I ² = 0,0 %	=> Es liegt keine signifikante Heterogenität im direkten Vergleich vor.

* Signifikante Heterogenität für $p < 0,10$.
df = Anzahl der Freiheitsgrade.

9.9.3 Publikationen mit Mixed treatment comparison (MTC)

Tabelle 26: Zusätzlich abgefragte Charakteristika in Typ-4- und Typ-5-Publikationen mit einem MTC

Erstautor	Zitiertes Methodenpapier	Random oder Fixed effects-Modell	Mehr als zweiarmige Studien eingeschlossen?	Wahl der Priorverteilungen	Sensitivitätsanalyse für Priorverteilungen	Benutzte Software	WinBugs-Code publiziert?	Ranking nach den Wahrscheinlichkeiten, wirksamste der Therapieoptionen zu sein, durchgeführt?
Cooper ⁷⁷	Lu ¹⁵⁷	Random	Ja	Nicht-informativ	Nein	WinBugs	Ja	Ja
Golfinopoulos ¹¹²	Salanti ^{204*}	Random	Ja	n. b.	n. b.	WinBugs	Nein	Ja
Jansen ¹³³	Lu ¹⁵⁷	Random	Ja	Nicht-informativ	Nein	WinBugs	Nein	Ja
Kyrgiou ¹⁴²	Lu ¹⁵⁷	Random	Ja	n. b.	n. b.	WinBugs	Nein	Ja
Lam ¹⁴³	Lu ¹⁵⁷ , Higgins ¹²²	Random	Ja	Nicht-informativ	Ja	WinBugs	Nein	Ja
Nixon ¹⁸⁰	Lu ¹⁵⁷	Random	Ja	Nicht-informativ	Nein	WinBugs	Ja	Nein
Stettler ²²⁴	Lu ¹⁵⁷	Random	Ja	Nicht-informativ	Nein	WinBugs	Ja	Nein
Vandermeer ²⁵⁴	Lu ¹⁵⁷	Random	Ja	Nicht-informativ	Ja	WinBugs	Nein	Ja
Woolcott ²⁶⁴	Higgins ¹²²	Beides durchgeführt**	Ja	Nicht-informativ	Nein	WinBugs	Ja	Nein
Elliott ⁹⁷	Lumley ¹⁶⁰	Random	Ja, aber n. b. wie.	n. b.	n. b.	R	Nein	Nein
Psaty ¹⁸⁹	Lumley ¹⁶⁰	Random	Nein***	n. b.	n. b.	n. b.	Nein	Nein
Dominici ⁸⁹	n. b.	Random	Nein	Nicht-informativ für Mittelwerte und informativ für Varianzen	Ja	n. b.	Nein	Ja

* Hauptsächlich ein Review über Lu¹⁵⁷.

** Das Random effects-Modell setzt sich in der Prüfung auf Abweichungen der Residuen als das passendere Modell für den Datensatz durch.

*** Alle mehr als zweiarmigen Studien ausgeschlossen, da diese nicht in Lumleys Modell eingeschlossen werden können.

MTC = Mixed treatment comparison. n. b. = Nicht berichtet. R: Statistiksoftware

Tabelle 27: Erhebung und Umgang mit Heterogenität und Inkonsistenz in Typ-4- und Typ-5-Publikationen mit MTC

Erstautor	Erhebung der Heterogenität?	Liegt hohe/statistisch signifikante Heterogenität vor?	Erhebung der Inkonsistenz?	Lag hohe/signifikante Inkonsistenz vor?	Berücksichtigung der Inkonsistenz in der Modellierung?	Maßnahmen zur Identifikation oder Reduzierung von Heterogenität und/oder Inkonsistenz in der Analyse?
Cooper ⁷⁷	n. b.	n. b.	Ja, Geeignetheit des Modells für die Daten mit Abweichungen der Residuen überprüft.	Nein, aber es gibt kleine Anzeichen („little Evidenz“) für Inkonsistenz.	Nein	Sensitivitätsanalyse für zwei Studien, die nicht Placebo, sondern „keine Therapie“ als Kontrollintervention aufweisen.
Golfinopoulos ¹¹²	Ja, mit I ² -Test.	Ja, hohe Heterogenität (I ² > 50 %) lag in fünf paarweisen Vergleichen vor.	Nein	n. b.	Nein	Keine
Jansen ¹³³	Ja, allerdings nicht mit einem statistischen Test, sondern durch klinische Beurteilung	Ja, klinische Heterogenität liegt vor.	Nein	n. b.	Ja, durch Adjustierung bezüglich des potenziellen Confunders Hb _{A1c} -Durchschnittswert	<ol style="list-style-type: none"> 1. Nach definierten Patientencharakteristika und Interventionen wird der Einschluss der Studien in die Analyse festgelegt. 2. Stratifizierte Analyse für Patienten, die Insulin spritzen und für Patienten, die kein Insulin spritzen 3. Adjustierung bezüglich eines Qualitätsscores der eingeschlossenen Studien und bezüglich des HbA1c-Grundwertes der Probanden.
Kyrgiou ¹⁴²	Ja, mit I ² -Test	Ja, hohe Heterogenität (I ² > 50 %) lag in einem paarweisen Vergleichen vor.	Ja, nach Lumley ¹⁶⁰	Ja, aber nur in einem Kreis des Netzwerks liegt Inkonsistenz vor.	Ja, durch Einschluss eines Inkonsistenzfaktors	Subgruppenanalyse für Therapien 1. oder 2. Wahl
Lam ¹⁴³	Ja, mit Chi ² - und I ² -Test. Zusätzlich werden ein L'Abbé-Plot und a priori sechs Sensitivitätsanalysen für potenzielle Confounder durchgeführt.	Nein, die Ergebnisse zeigen sich in den Sensitivitätsanalysen robust und es liegt keine statistisch signifikante oder hohe (I ² > 50 %) Heterogenität vor.	Nein	n. b.	Nein	<ol style="list-style-type: none"> 1. Subgruppenanalyse für Patienten mit schwerer Herzinsuffizienz. 2. Stratifizierte Analyse für Studien mit der Kontrollgruppe „Pharmakotherapie“ und der Kontrollgruppe „implantierbarer Herzschrittmacher“

Fortsetzung Tabelle 27: Erhebung und Umgang mit Heterogenität und Inkonsistenz in Typ-4- und Typ-5-Publikationen mit MTC

Nixon ¹⁸⁰	Ja, sie wird in der Posteriorverteilung als Zwischen-Studien-Varianz ausgedrückt.	Ja, Heterogenität liegt vor. Sie kann aber durch Adjustierung bezüglich der zwei Confounder um 61 % reduziert werden.	Nein	n. b.	Nein	Adjustierung bezüglich Krankheitsdauer und Grad der Behinderung mit Metaregressionsmethoden
Stettler ²²⁴	Ja, sie wird in der Posteriorverteilung als Zwischen-Studien-Varianz ausgedrückt. Zusätzlich I ² -Test durchgeführt.	Nein, die Heterogenität ist nur bei einem Outcome und einem paarweisen Vergleich moderat erhöht und sonst immer niedrig (Tau ² < 0,04).	Ja, nach Lu ¹⁵⁹ . Zusätzlich Geeignetheit des Modells mit den Abweichungen der Residuen geprüft.	Nein, keine signifikante Inkonsistenz festgestellt. Das Model eignet sich gut, mit Ausnahme für einen Outcome.	Ja, durch Einschluss eines Inkonsistenz-faktors	1. Adjustierung bezüglich der unterschiedlichen Nachbeobachtungszeiten. 2. Stratifizierungen nach Dauer bis zur Stent-Thrombose und ob Patienten Diabetes hatten. 3. Sensitivitätsanalysen für Verstrebungsdicke („strut thickness“) und Typ der Stentplattform sowie Kriterien zur methodischen Qualität der Studien.
Vandermeer ²⁵⁴	Ja, mit Q-Test.	n. b.	n. b.	n. b.	Nein	Keine
Woolcott ²⁶⁴	Ja, mit Chi ² -Test.	Nein, Heterogenität nicht statistisch signifikant	Nein	n. b.	Nein	Keine
Elliott ⁹⁷	Ja, mit Riley-Day-Test.	n. b.	Ja, nach Lumley ¹⁶⁰	Nein, als gering bezeichnet	n. b.	Keine
Psaty ¹⁸⁹	Ja, aber nicht berichtet, welcher Test genutzt wurde	Unklar, da nur für gesamtes Evidenznetzwerk und nicht für die einzelnen Therapieeffektvergleiche der Arzneistoffgruppen berichtet.	Ja, nach Lumley ¹⁶⁰	Nein, als gering bezeichnet	Ja, durch Einschluss eines Inkonsistenz-faktors	Keine
Dominici ⁸⁹	n. b.	N. b.	n. b.	n. b.	n. b.	Keine

HbA1c = Hämoglobin A1c. MTC = Mixed treatment comparison. n. b. = Nicht berichtet.

9.9.4 Gegenüberstellung der Metaanalyseergebnisse des direkten und indirekten Vergleichs aus den Typ-5-Publikationen

Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Erstautor	Jahr	Patientenkollektiv	Meta-analyse-nummer	Therapieoption A	Therapieoption B	Gemeinsamer Komparator X für den indirekten Vergleich	Zielgröße
Sanchez-Ramos ²⁰⁶	2002	Schwangere, die eine medikamentöse Geburtseinleitung benötigen	1	25 µg vaginal Misoprostol	50 µg Misoprostol	Intracervikales Dinoprostongel	Niederkunft nach einer Einzeldosis
			2				
			3				
			4				
Antiplatelet Trialists' Collaboration ²⁶ ; Glenny et al. ¹⁰⁹	1994	Patienten mit einem erhöhten Risiko für eine vaskuläre Okklusion (z. B. Koronar- oder Beinarterien Bypass-Transplantation oder -angioplastie)	5	Ticlopidin	Aspirin	Kontrolle	Kardiovaskuläre Ereignisse
			6	Aspirin + Dipyridamol	Sulfinpyrazon		
			7	Hochdosis-Aspirin	Mittelhochdosiertes Aspirin		
			8	Sulfinpyrazon	Aspirin		
			9	Aspirin + Dipyridamol	Aspirin		
Antiplatelet Trialists' Collaboration ²⁵ ; Glenny et al. ¹⁰⁹	1994	Patienten mit einem erhöhten Risiko für okklusive vaskuläre Erkrankungen (z. B. Koronare Herzkrankheit, Schlaganfall, PVK)	10	Sulfinpyrazon	Aspirin	Kontrolle	Kardiovaskuläre Ereignisse
			11	Hochdosis-Aspirin	Mittelhochdosiertes Aspirin		
			12	Ticlopidin	Aspirin		
			13	Aspirin + Dipyridamol	Aspirin		
Antiplatelet Trialists' Collaboration ²⁷ ; Glenny et al. ¹⁰⁹	1994	Hochrisikopatienten und Patienten auf einer chirurgischen Station	14	Aspirin + Dipyridamol	Aspirin	Kontrolle	Tiefe Beinvenenthrombose
Ausejo ³¹ ; Glenny et al. ¹⁰⁹	2000	Kinder mit Krupp	15	Budesonid	Dexamethason	Placebo	Verbesserung im Krupp-Schweregrad-Score

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Bucher ⁵⁴ ; Glenn et al. ¹⁰⁹	1997	HIV-infizierte Patienten	16	Trimethoprim-Sulfamethoxazol	Dapson + Pyrimethamin oder Dapson allein	Aerosolisiertes Pentamidin	Pneumocystitis carinii-Pneumonie
Cheng ⁶⁴ ; Glenn et al. ¹⁰⁹	2000	Frauen mit Wunsch nach Notfallkontrazeption	17	Levonogestrel	Mifepistone	Yuzpe	Anzahl Schwangerschaften
Chiba ⁶⁵ ; Glenn et al. ¹⁰⁹	1997	Patienten mit gastroösophagealer Refluxkrankheit	18	H ₂ RA	PPI	Placebo	Heilungsrate
Collins ⁷⁵ ; Glenn et al. ¹⁰⁹	2000	Patienten mit postoperativem Schmerz	19	Ibuprofen 400 mg	Ibuprofen 200 mg	Kontrolle	50 prozentige Schmerzstillung
Delaney ⁸⁵ ; Glenn et al. ¹⁰⁹	2001	Patienten mit Dyspepsie	20	PPI	H ₂ RA	Alginate/Antazida	Allgemeine Verbesserung
Handoll ¹¹⁴ ; Glenn et al. ¹⁰⁹	2002	Patienten, die eine Hüftfraktur operiert bekommen	21	Niedermolekulares Heparin	Unfraktioniertes Heparin	Placebo	Tiefe Beinvenenthrombose
Higgins ¹²² ; Glenn et al. ¹⁰⁹	1996	Patienten mit Leberzirrhose und Gastroösophagealen Varizen	22	Betablocker	Sklerotherapie	Kontrolle	Erste Blutung
Horn ¹³⁰ ; Glenn et al. ¹⁰⁹	2001	Patienten mit akutem ischämischem Schlaganfall	23	Mimodipin 240 mg	Mimodipin 120 mg	Kontrolle	Tod oder Pflegefall
Marshall ¹⁶³ ; Glenn et al. ¹⁰⁹	1997	Patienten mit Colitis Ulzerosa	24	5-Aminosalicylsäure	Rektale Glukocorticoide	Budesonid	Endoskopisch bestätigte Remission
McIntosh ¹⁶⁹ ; Glenn et al. ¹⁰⁹	2000	Patienten mit unkomplizierter Malaria	25	Artemisinin	Artesunat	Chinin	Parasitenfreiheit am 28. Tag
Moore ¹⁷⁵ ; Glenn et al. ¹⁰⁹	1997	Patienten mit starken postoperativen Schmerzen	26	Paracetamol + Codein	Paracetamol	Placebo	Mindestens 50 prozentige Schmerzlinderung
Packer ¹⁸³ ; Glenn et al. ¹⁰⁹	2001	Patienten mit Herzinsuffizienz	27	Carvedilol	Metoprolol	Placebo	Änderung in der linksventrikulären mittleren Auswurffraktion
Po ¹⁸⁶ ; Glenn et al. ¹⁰⁹	1997	Patienten mit postoperativem Schmerz	28	Paracetamol + Dexamethason	Paracetamol	Placebo	Differenz in der Schmerzintensität
Poynard ¹⁸⁸ ; Glenn et al. ¹⁰⁹	1996	Patienten mit viraler Hepatitis C	29	Zwölfmonatige Interferontherapie	Sechsmonatige Interferontherapie	Kontrolle	Nachhaltige Alanin-Transaminase-Anspruchrate
Rostom ²⁰¹ ; Glenn et al. ¹⁰⁹	2000	Patienten, die länger als drei Wochen NSAR eingenommen haben	30	PPI	H ₂ RA	Placebo	Endoskopisch nachgewiesene Geschwüre
Sauriol ²⁰⁷ ; Glenn et al. ¹⁰⁹	2001	Patienten mit Schizophrenie	31	Olanzapin	Risperidon	Placebo	Verbesserung auf einer psychiatrischen Skala

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Silagy ²¹² ; Glenny et al. ¹⁰⁹	2001	Patienten mit Nikotinersatztherapie	32	Nikotinpflaster für 24 Stunden	Nikotinpflaster für 16 Stunden	Kontrolle	Abstinenz
			33	Nikotinpflaster tragende Patienten	Keine Nikotinpflaster tragende Patienten		
Silagy ²¹³ ; Glenny et al. ¹⁰⁹	2001	Raucher	34	Mehr als eine Konsultation	Eine Konsultation	Kontrolle	Abstinenz
Soo ²²⁰ ; Glenny et al. ¹⁰⁹	2004	Patienten mit Dyspepsie aber ohne Ulzerationen	35	H ₂ RA	Sucralfat	Placebo	Verbesserung der allgemeinen Symptomatik
			36	Prokinetika	H ₂ RA	Placebo	
Trindade ²³⁹ ; Glenny et al. ¹⁰⁹	1997	Patienten mit schwerer Depression	37	Fluoxetin	Fluvoxamin	Kontrolle	Therapieabbruch aufgrund von Unwirksamkeit
			38	Paroxetin	Fluvoxamin		
			39	Sertralin	Fluvoxamin		
			40	Fluoxetin	Paroxetin		
			41	Fluoxetin	Sertralin		
Van Pinxteren ²⁵³ ; Glenny et al. ¹⁰⁹	2000	Patienten mit Sodbrennen	42	PPI	H ₂ RA	Placebo	Abklingen des Sodbrennens
Zhang ²⁷¹ ; Glenny et al. ¹⁰⁹	1998	Patienten mit Dysmenorrhoe	43	Naproxen	Mefenaminsäure	Placebo	Zumindest moderate Schmerzlinderung
			44	Naproxen	Aspirin		
			45	Ibuprofen	Aspirin		
			46	Naproxen	Ibuprofen		
Zhang ²⁶⁹ ; Glenny et al. ¹⁰⁹	1996	Patienten mit postoperativem Schmerz	47	Paracetamol + Codein	Paracetamol	Placebo	Differenz in der Schmerzintensität
Antiplatelet Trialists' Collaboration ²⁶ ; Glenny et al. ¹⁰⁹	1994	Patienten mit einem erhöhten Risiko für eine vaskuläre Okklusion (z. B. Koronar- oder Beinarterien-Bypasstransplantation oder -angioplastie)	48	Ticlopidin	Aspirin	Kontrolle	Kardiovaskuläre Ereignisse
			49	Aspirin + Dipyridamol	Sulfinpyrazon		
			50	Hochdosis-Aspirin	Mittelhoch-dosiertes Aspirin		
			51	Sulfinpyrazon	Aspirin		
			52	Aspirin + Dipyridamol	Aspirin		

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Antiplatelet Trialists' Collaboration ²⁵ ; Glenny et al. ¹⁰⁹	1994	Patienten mit einem erhöhten Risiko für okklusive vaskuläre Erkrankungen (z. B. Koronare Herzkrankheit, Schlaganfall, periphere arterielle Verschlusskrankheit)	53	Sulfinpyrazon	Aspirin	Kontrolle	Kardiovaskuläre Ereignisse
			54	Hochdosis-Aspirin	Mittelhoch-dosiertes Aspirin		
			55	Ticlopidin	Aspirin		
			56	Aspirin + Dipyridamol	Aspirin		
Antiplatelet Trialists' Collaboration ²⁷ ; Glenny et al. ¹⁰⁹	1994	Hochrisikopatienten und Patienten auf einer chirurgischen Station	57	Aspirin + Dipyridamol	Aspirin	Kontrolle	Tiefe Beinvenenthrombose
Ausejo ³¹ ; Glenny et al. ¹⁰⁹	2000	Kinder mit Krupp	58	Budesonid	Dexamethason	Placebo	Verbesserung im Krupp-Schweregrad-Score
Bekkering ³⁸	2008	Patienten mit erektiler Dysfunktion	59	Sildenafil	Vardenafil	Placebo	Index für erektile Funktion
Boonen ⁴⁸ ; The Record Trial Group ²³³	2007	Postmenopausale Frauen und/oder Männer über 50 Jahre	60	Vitamin D3 + Calcium	Vitamin D3	Placebo/Keine Therapie	Hüftbruch

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Brown ⁵³	2006	Patienten mit Langzeit-NSAR-Therapie	61	Misoprostol	COX-2-preferierende NSAR	Placebo	Symptomatisches Geschwür
			62	Misoprostol	COX-2-preferierende NSAR		Gastrointestinale Symptome
			63	H ₂ RA	Misoprostol		Symptomatisches Geschwür
			64	COX-2-selektiver Inhibitor	COX-2-preferierende NSAR		Gravierende gastrointestinale Komplikation
			65	PPI	COX-2-selektiver Inhibitor		Gravierende gastrointestinale Komplikation
			66	PPI	COX-2-selektiver Inhibitor		Symptomatisches Geschwür
			67	PPI	COX-2-selektiver Inhibitor		Gastrointestinale Symptome
			68	Misoprostol	COX-2-preferierende NSAR		Gravierende gastrointestinale Komplikation
			69	Misoprostol	COX-2-selektiver Inhibitor		Gastrointestinale Symptome
			70	H ₂ RA	Misoprostol		Endoskopisch diagnostiziertes Geschwür
			71	PPI	Misoprostol		Gravierende gastrointestinale Komplikation
			72	Misoprostol	COX-2-preferierende NSAR		Anämie
			73	PPI	Misoprostol		Endoskopisch diagnostiziertes Geschwür
			74	H ₂ RA	PPI		Endoskopisch diagnostiziertes Geschwür
			75	H ₂ RA	PPI		Gravierende gastrointestinale Komplikation
			76	H ₂ RA	Misoprostol		Gastrointestinale Symptome
			77	Misoprostol	COX-2-preferierende NSAR		Endoskopisch diagnostiziertes Geschwür
			78	H ₂ RA	PPI		Symptomatisches Geschwür

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Bucher ⁵⁴ ; Glenny et al. ¹⁰⁹	1997	HIV-infizierte Patienten	79	Trimethoprim-Sulfamethoxazol	Dapson + Pyrimethamin oder Dapson allein	Aerosolisiertes Pentamidin	Pneumocystitis carinii-Pneumonie
Cheng ⁶⁴ ; Glenny et al. ¹⁰⁹	2000	Frauen mit Wunsch nach Notfallkontrazeption	80	Levonogestrel	Mifepistone	Yuzpe	Anzahl Schwangerschaften
Chiba ⁶⁵ ; Glenny et al. ¹⁰⁹	1997	Patienten mit gastroösophagealer Refluxkrankheit	81	H ₂ RA	PPI	Placebo	Heilungsrate
Chou ⁶⁶ ; Glenny et al. ¹⁰⁹	2006	HIV-infizierte Patienten	82	NNRTI-basierte HAART	Proteaseinhibitor-basierte HAART	Kombinations-therapie mit zwei NRTI	Tod oder Krankheitsfortschritt
			83				Unterdrückung der Virenvermehrung
Collins ⁷⁵ ; Glenny et al. ¹⁰⁹	2000	Patienten mit postoperativem Schmerz	84	Ibuprofen 400 mg	Ibuprofen 200 mg	Kontrolle	50 prozentige Schmerzstillung
Delaney ⁸⁵ ; Glenny et al. ¹⁰⁹	2001	Patienten mit Dyspepsie	85	PPI	H ₂ RA	Alginate/Antazida	Allgemeine Verbesserung
Di Mario ⁸⁷ ; Glenny et al. ¹⁰⁹	1996	Patienten mit bisher unbehandeltem Magengeschwür	86	Cimetidin	Ranitidin	Placebo	Endoskopisch bestätigte Abheilung
Gartlehner ¹⁰³ ; Geborek ¹⁰⁶	2006	Symptomatische Patienten mit Rheumatoider Arthritis, die bereits Methotrexat bekommen	87	Etanercept	Infliximab	Placebo	ACR50-Ansprechrage
			88				ACR20-Ansprechrage
Handoll ¹¹⁴ ; Glenny et al. ¹⁰⁹	2002	Patienten, die eine Hüftfraktur operiert bekommen	89	Niedermolekulares Heparin	Unfraktioniertes Heparin	Placebo	Tiefe Beinvenenthrombose
Higgins ¹²² ; Glenny et al. ¹⁰⁹	1996	Patienten mit Leberzirrhose und gastroösophagealen Varizen	90	Betablocker	Sklerotherapie	Kontrolle	Erste Blutung
Horn ¹³⁰ ; Glenny et al. ¹⁰⁹	2001	Patienten mit akutem ischämischen Schlaganfall	91	Mimodipin 240 mg	Mimodipin 120 mg	Kontrolle	Tod oder Pflegefall
Lange ¹⁴⁶	2003	Patienten mit allergischer Rhinitis	92	Topische Glukocorticoide	Topische Antihistaminika	Placebo	Therapieversagen
Marshall ¹⁶³ ; Glenny et al. ¹⁰⁹	1997	Patienten mit Colitis Ulzerosa	93	5-Aminosalicylsäure	Rektale Glukocorticoide	Budesonid	Endoskopisch bestätigte Remission
McIntos ¹⁶⁹ h; Glenny et al. ¹⁰⁹	2000	Patienten mit unkomplizierter Malaria	94	Artemisinin	Artesunat	Chinin	Parasitenfreiheit am 18. Tag
Moore ¹⁷⁵ ; Glenny et al. ¹⁰⁹	1997	Patienten mit starken postoperativen Schmerzen	95	Paracetamol + Codein	Paracetamol	Placebo	Mindestens 50-prozentige Schmerzlinderung
Packer ¹⁸³ ; Glenny et al. ¹⁰⁹	2001	Patienten mit Herzinsuffizienz	96	Carvedilol	Metoprolol	Placebo	Änderung in der linksventrikulären mittleren Auswurffraktion
Po ¹⁸⁶ ; Glenny et al. ¹⁰⁹	1997	Patienten mit postoperativem Schmerz	97	Paracetamol + Dexamethason	Paracetamol	Placebo	Differenz in der Schmerzintensität

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Poynard ¹⁸⁸ , Glenn et al. ¹⁰⁹	1996	Patienten mit viraler Hepatitis C	98	Zwölfmonatige Interferontherapie	Sechsmonatige Interferontherapie	Kontrolle	Nachhaltige Alanin- Transaminase-Anspruchrate
Robenshtok ¹⁹⁶	2007	Krebspatienten nach Chemotherapie oder Stammzellentransplantation	99	Fluconazol	Itraconazol orale Suspension	Placebo/Keine Therapie/Nicht- systemische Antimykotika	Invasive Pilzinfektionen
Rostom ²⁰¹ , Glenn et al. ¹⁰⁹	2000	Patienten, die länger als drei Wochen NSAR eingenommen haben	100	PPI	H ₂ RA	Placebo	Endoskopisch nachgewiesene Geschwüre
Sauriol ²⁰⁷ , Glenn et al. ¹⁰⁹	2001	Patienten mit Schizophrenie	101	Olanzapin	Risperidon	Placebo	Verbesserung auf einer psychiatrischen Skala
Silagy ²¹² , Glenn et al. ¹⁰⁹	2001	Patienten mit Nikotinersatztherapie	102	Nikotinpflaster für 24 Stunden	Nikotinpflaster für 16 Stunden	Kontrolle	Abstinenz
			103	Nikotinpflaster tragende Patienten	Kein Nikotinpflaster tragende Patienten		
Silagy ²¹³ , Glenn et al. ¹⁰⁹	2001	Raucher	104	Mehr als eine Konsultation	Eine Konsultation	Kontrolle	Abstinenz
Soo ²²⁰ , Glenn et al. ¹⁰⁹	2004	Patienten mit Dyspepsie aber ohne Ulzerationen	105	H ₂ RA	Sucralfat	Placebo	Verbesserung der allgemeinen Symptomatik
			106	Prokinetika	H ₂ RA		
Trindade ²³⁹ , Glenn et al. ¹⁰⁹	1997	Patienten mit schwerer Depression	107	Fluoxetin	Fluvoxamin	Kontrolle	Therapieabbruch aufgrund von Unwirksamkeit
			108	Paroxetin	Fluvoxamin		
			109	Sertralin	Fluvoxamin		
			110	Fluoxetin	Paroxetin		
			111	Fluoxetin	Sertralin		
van Pinxteren ²⁵³ , Glenn et al. ¹⁰⁹	2000	Patienten mit Sodbrennen	112	PPI	H ₂ RA	Placebo	Abklingen des Sodbrennens
Wilhelmus ²⁶¹	2000	Patienten mit Keratitis (verursacht durch Herpes-simplex-Virus)	113	Trifluridin	Idoxuridin	Aciclovir	Heilung
			114	Aciclovir	Idoxuridin	Trifluridin	
			115	Vidarabin	Idoxuridin	Aciclovir	
			116	Aciclovir	Idoxuridin	Vidarabin	
			117	Trifluridin	Idoxuridin	Vidarabin	
			118	Vidarabin	Idoxuridin	Trifluridin	

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Wu ²⁶⁵	2006	Raucher	119	Bupropion	Nikotinersatz- therapie	Placebo	Nichtraucher noch nach drei Monaten nach dem Aufhören
			120	Nikotinersatz- therapie	Bupropion	Placebo	Nichtraucher noch nach einem Jahr nach dem Aufhören
Zhang ²⁷¹ ; Glenn et al. ¹⁰⁹	1998	Patienten mit Dysmenorrhoe	121	Naproxen	Mefenaminsäure	Placebo	Zumindest moderate Schmerzlinderung
			122	Naproxen	Aspirin		
			123	Ibuprofen	Aspirin		
			124	Naproxen	Ibuprofen		
Zhang ²⁶⁹ ; Glenn et al. ¹⁰⁹	1996	Patienten mit postoperativem Schmerz	125	Paracetamol + Codein	Paracetamol	Placebo	Differenz in der Schmerzintensität
Vandermeer ²⁵⁴	2007	Patienten mit chronischer Schlaflosigkeit	126	Benzodiazepine	Non- Benzodiazepine	Placebo	Schlafeffizienz
			127				Totale Schlafzeit
			128				Wachzustand nach dem Einschlafen
			129				Einschlaflatenzzeit
			130				Schlafqualität
			131				Wachzustand nach dem Einschlafen
			132				Einschlaflatenzzeit
			133				Schlafeffizienz
			134				Totale Schlafzeit
			135				Schlafqualität
			136				Schlafeffizienz
			137				Totale Schlafzeit
			138				Wachzustand nach dem Einschlafen
			139				Einschlaflatenzzeit
140	Schlafqualität						
Antithrombotic Trialists' Collaboration ²⁸	2002	Hochrisikopatienten für kardiovaskuläre Ereignisse	141	Anderer Thrombo- zytenaggrega- tionshemmer in Monotherapie	Aspirin	Kontrollinter- vention; nicht näher spezifiziert	Kardiovaskuläres Ereignis
Mason ¹⁶⁵	2004	Patienten mit akuten Schmerzen durch Zerrungen, Verrenkungen oder Sportunfälle	142	Piroxicam topisch	Indometacin topisch	Placebo	Pain

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Richy ¹⁹⁵	2005	Patienten mit Risiko für eine glucocorticoidinduzierter Osteoporose	143	Vitamin D-Analoga	Natürliches Vitamin D	Placebo	Prävention von Glukocortoid-induzierten Knochenbrüchen in der Wirbelsäule
Small ²¹⁴ ; Jung ¹³⁷	2006	Patienten nach Organtransplantation	144	Ganciclovir-Bevorzugung	Ganciclovir-Prophylaxe	Universelle Prophylaxe	Symptomatische Cytomegalievirus-Krankheit
Ferrari ¹⁰⁰	2002	Migränepatienten	145	Naratriptan 2,5 mg	Zolmitriptan 2,5 mg	Placebo	Langfristige Schmerzfreiheit
			146	Rizatriptan 10 mg	Naratriptan 2,5 mg		
			147	Rizatriptan 10 mg	Zolmitriptan 2,5 mg		
			148	Sumatriptan 100 mg	Rizatriptan 5 mg		
			149	Sumatriptan 100 mg	Eletriptan 20 mg		
			150	Sumatriptan 100 mg	Eletriptan 40 mg		
			151	Sumatriptan 100 mg	Naratriptan 2,5 mg		
			152	Sumatriptan 100 mg	Eletriptan 80 mg		
			153	Sumatriptan 100 mg	Rizatriptan 10 mg		
154	Sumatriptan 100 mg	Zolmitriptan 5 mg					
Horn ¹²⁹	2000	Patienten mit überlebtem ischämischem Schlaganfall	155	Orale Verabreichung des Calciumantagonisten	Intravenöse Verabreichung des Calciumantagonisten	Kontrolle	Negatives Ereignis
			156	Nimodipin 60 mg oral	Nimodipin 120 mg oral		
			157	Nimodipin 60 mg oral	Nimodipin 240 mg oral		
Po ¹⁸⁷	1998	Schmerzpatienten	158	Ibuprofen + Codein	Ibuprofen	Placebo	Schmerzlinderung

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Zhang ²⁷⁰	1997	Patienten mit postoperativem Schmerz	159	Acetylsalicylsäure + Koffein	Acetylsalicylsäure	Placebo	SPID %
			160	Acetylsalicylsäure + Koffein	Acetylsalicylsäure	Placebo	TOTPAR %
			161	Acetylsalicylsäure + Codein	Acetylsalicylsäure	Placebo	ResRR
			162	Acetylsalicylsäure + Codein	Acetylsalicylsäure	Placebo	SPID %
			163	Acetylsalicylsäure + Codein	Acetylsalicylsäure	Placebo	TOTPAR %
			164	Acetylsalicylsäure + Koffein	Acetylsalicylsäure	Placebo	ResRR
Mudge ¹⁷⁹	2005	Patienten mit Schizophrenie	165	Olanzapin	Risperidon	Haloperidol	PANSS-GPS-Veränderung
			166				PANSS-Gesamtveränderung
			167				PANSS positive Veränderung
			168				PANSS negative Veränderung
Ballesteros ³⁵	2005	Patienten mit depressiver Verstimmung	169	Trizyklische Antidepressiva	Selektive Serotoninwiederaufnahmehemmer	Placebo	50-prozentige Reduktion der depressiven Symptome im Vergleich zur Symptomatik bei Studienbeginn
			170	Trizyklische Antidepressiva	Monoaminoxidasehemmer		
Elliott ⁹⁷	2007	Patienten mit Bluthochdruck aber ohne Diabetes mellitus zu Studienbeginn	171	AT(II)-Rezeptorblocker	Diuretikum	Multiple	Entwicklung eines Diabetes Mellitus
			172	Betablocker	Diuretikum		
			173	Calciumkanalblocker	Diuretikum		
			174	ACE-Inhibitor	Diuretikum		

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Golfinopoulos ¹¹²	2007	Patienten mit wiederkehrendem oder metastasierendem Kolonkarzinom	175	FU	FU + Bevacizumab	Multiple	Krankheitsprogression
			176	FU	FU + Oxaliplatin	Multiple	Krankheitsprogression
			177	FU	FU + Irinotecan	Multiple	Krankheitsprogression
			178	FU	Oxaliplatin	Multiple	Mortalität
			179	FU	FU + Bevacizumab	Multiple	Mortalität
			180	FU	Irinotecan	Multiple	Mortalität
			181	FU	Oxaliplatin	Multiple	Krankheitsprogression
			182	FU	Irinotecan	Multiple	Krankheitsprogression
			183	FU	FU + Irinotecan	Multiple	Mortalität
			184	FU	FU + Oxaliplatin	Multiple	Mortalität
Kyrgiou ¹⁴²	2006	Patienten mit Ovarialkarzinom	185	Platin-Monotherapie	Non-Platin/Non- Texan-Monotherapie	Multiple	Mortalität
			186	Platinbasierte Kombinations- therapie	Non-Platin/Non- Texan-Mono- therapie		
			187	Non-Platin/Non- Texan-Kombi- nationstherapie	Non-Platin/Non- Texan-Mono- therapie		
Lam ¹⁴³	2007	Patienten mit Funktionsstörung des linken Ventrikels	188	ICD + Kardiale Resynchronisa- tionstherapie	Pharmakotherapie	Multiple	Mortalität
			189	ICD + Kardiale Resynchronisa- tionstherapie	ICD		
			190	ICD + Kardiale Resynchronisa- tionstherapie	Kardiale Resynchronisa- tionstherapie		
			191	Kardiale Resynchronisa- tionstherapie	Pharmakotherapie		
			192	ICD	Pharmakotherapie		

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Psaty ¹⁸⁹	2003	Hypertensive Patienten	193	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Kongestive Herzinsuffizienz
			194	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Gesamt mortalität
			195	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Schlaganfall
			196	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Koronare Herzkrankheit
			197	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Herz-Kreislauf-Erkrankungsmortalität
			198	Gering dosierte Diuretika	ACE-Hemmer	Placebo/Keine Therapie	Schweres Herz-Kreislauf-Ereignis
			199	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	Herz-Kreislauf-Erkrankungsmortalität
			200	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	Gesamt mortalität
			201	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	Schlaganfall
			202	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	Schweres Herz-Kreislauf-Ereignis
			203	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	kongestive Herzinsuffizienz
			204	Gering dosierte Diuretika	Calciumkanal-blocker	Placebo/Keine Therapie	Koronare Herzkrankheit

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Stettler ²²⁴	2007	Patienten mit Stent	205	PES	BMS	Multiple	Tod oder Herzinfarkt
			206	PES	BMS	Multiple	Mortalität
			207	PES	BMS	Multiple	Herzinfarkt
			208	PES	BMS	Multiple	Zielläsionsrevaskularisation
			209	PES	BMS	Multiple	Kardiale Mortalität
			210	SES	BMS	Multiple	Tod oder Herzinfarkt
			211	SES	BMS	Multiple	Mortalität
			212	SES	BMS	Multiple	Herzinfarkt
			213	SES	BMS	Multiple	Zielläsionsrevaskularisation
			214	SES	BMS	Multiple	Kardiale Mortalität
			215	PES	BMS	Multiple	Endgültige Stentthrombose
			216	SES	BMS	Multiple	Endgültige Stentthrombose
			217	SES	PES	Multiple	Tod oder Herzinfarkt
			218	SES	PES	Multiple	Mortalität
			219	SES	PES	Multiple	Herzinfarkt
220	SES	PES	Multiple	Zielläsionsrevaskularisation			
221	SES	PES	Multiple	Kardiale Mortalität			
222	SES	PES	Multiple	Endgültige Stentthrombose			
Vandermeer ²⁵⁴	2007	Patienten mit chronischer Schlaflosigkeit	223	Benzodiazepine	Non-Benzodiazepine	Placebo	Wachzustand nach dem Einschlafen
			224	Benzodiazepine	Non-Benzodiazepine	Placebo	Totale Schlafzeit
			225	Benzodiazepine	Non-Benzodiazepine	Placebo	Einschlaflatenzzeit
			226	Benzodiazepine	Non-Benzodiazepine	Placebo	Schlafeffizienz
			227	Benzodiazepine	Non-Benzodiazepine	Placebo	Schlafqualität
Brophy ⁵¹	2005	Patienten mit ischämischer Herzerkrankung	228	Abciximab + PCI	Tirofiban + PCI	Placebo	Multivariater Endpunkt aus Mortalität, Myokardinfarkt oder wiederholter Vaskularisation
Smith 2007 ²⁴⁰ , Gamble 2006 ¹⁰²	2007	Patienten mit fokalen epileptischen Krampfanfällen	229	Lamotrigin	Carbamazepin	Multiple	Zeit bis zum Therapieversagen
			230				Zeit bis zum ersten Krampfanfall

Fortsetzung Tabelle 28: Allgemeine Informationen zu den Metaanalysen aus den Typ-5-Publikationen

Tudur Smith ^{80, 243}	2007	Patienten mit fokalen epileptischen Krampfanfällen	231	Carbamazepin	Phenobarbital	Multiple	Zeit bis zum Therapieversagen
			232				Zeit bis zur zwölfmonatigen Remission
			233				Zeit bis zum ersten Krampfanfall
Tudur Smith ^{80, 241}	2007	Patienten mit fokalen epileptischen Krampfanfällen	234	Carbamazepin	Phenytoin	Multiple	Zeit bis zum Therapieversagen
			235				Zeit bis zur zwölfmonatigen Remission
			236				Zeit bis zum ersten Krampfanfall
Tudur Smith ⁸⁰ , Marson ¹⁶⁴	2007	Patienten mit fokalen epileptischen Krampfanfällen	237	Carbamazepin	Valproinsäure	Multiple	Zeit bis zum Therapieversagen
			238				Zeit bis zur zwölfmonatigen Remission
			239				Zeit bis zum ersten Krampfanfall
		Patienten mit generalisierten epileptischen Krampfanfällen	240	Carbamazepin	Valproinsäure	Multiple	Zeit bis zum Therapieversagen
			241				Zeit bis zur zwölfmonatigen Remission
			242				Zeit bis zum ersten Krampfanfall
Tudur Smith ^{80, 242}	2007	Patienten mit generalisierten epileptischen Krampfanfällen	243	Phenytoin	Valproinsäure	Multiple	Zeit bis zum Therapieversagen
			244				Zeit bis zur zwölfmonatigen Remission
			245				Zeit bis zum ersten Krampfanfall
Thijs ²³⁴	2008	Patienten nach überstandener Transitorischer ischämischer Attacke oder ischämischem Schlaganfall	246	Acetylsalicylsäure + Dipyridamol	Acetylsalicylsäure	Multiple	Schwerwiegendes vaskuläres Ereignis
			247	Thienopyridine + Acetylsalicylsäure	Thienopyridine		
			248	Thienopyridine	Acetylsalicylsäure		
			249	Thienopyridine + Acetylsalicylsäure	Acetylsalicylsäure		

ACE = Angiotensin converting enzyme (Angiotensin umsetzenden Enzym). ACR = Krankheitsaktivitätsscore des American College of Rheumatology. BMS = Arzneistofffreier Metallstent. COX = Cyclooxygenase. FU = 5-Fluorouracil + Leucovorin. GPS = General Psychopathology Scale (Allgemeine pathopsychologische Skala). H₂RA = Histamin-2-Rezeptorantagonist. HAART = Hochaktive antiretrovirale Therapie. HIV = Humanes Immunodefizienz Virus. ICD = Implantierbarer Cardioverter/Defibrillator. NRTI = Nukleosidischer Reverse-Transkriptase-Inhibitor. NNRTI = Nicht-nukleosidischer Reverse-Transkriptase-Inhibitor. NSAR = Nicht-steroidale Antirheumatika. PANSS = Positiv- und Negativsyndromskala. PCI = Perkutane koronare Interventionen. PES = Paclitaxel freisetzender Stent. PPI = Protonenpumpeninhibitor. ResRR = Response rate ratio (Ratio der Ansprechraten). SES = Sirolimus freisetzender Stent. SPID % = Sum of pain intensity difference (Summe der Schmerzintensitätendifferenzen). TOTPAR % = Total pain relief (Gesamtschmerzreduktion).

Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

Meta-analyse-nummer	Methodengruppe	Effekt-schätzer	Random oder Fixed effects?	Gesamtanzahl Studien, die in den indirekten Vergleich eingeschlossen werden	Anzahl Studien für den indirekten Vergleich, die nicht RCT sind	Gesamtanzahl Probanden, die in den indirekten Vergleich eingeschlossen werden (Verum- und Kontrollgruppen)	Schätzung des Therapieeffektunterschieds
1	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer	OR	Random effects	2	0	Nicht berichtet	0,10 (0,03 bis 0,38)
2	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer	OR	Random effects	2	0	Nicht berichtet	0,45 (0,25 bis 0,84)
3	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer	OR	Random effects	2	0	Nicht berichtet	2,68 (1,21 bis 5,91)
4	Nicht-adjustierter indirekter Vergleich über einen Gesamteffektschätzer	WMD	Random effects	2	0	Nicht berichtet	420 (202 bis 638)
5	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	25	0	4.032	0,75 (0,59 bis 0,96)
6	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	24	0	4.507	1,54 (1,13 bis 2,12)
7	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	12	0	2.482	0,59 (0,46 bis 0,77)
8	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	17	0	3.043	0,81 (0,58 bis 1,13)
9	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	21	0	3.881	1,53 (1,21 bis 1,92)
10	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	66	0	84.698	1,32 (1,17 bis 2,47)
11	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	49	0	76.025	1,64 (1,54 bis 1,74)
12	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	79	0	87.369	0,91 (0,81 bis 1,03)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

13	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	65	0	87.002	1,12 (1,03 bis 1,22)
14	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	20	0	2.062	0,94 (0,73 bis 1,2)
15	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Fixed effects	13	0	1.066	0,20 (0,02 bis 0,35)
16	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	14	0	2.744	0,55 (0,35 bis 0,87)
17	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	4	0	3.993	19,83 (1,21 bis 326,03)
18	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	13	0	3.185	0,73 (0,68 bis 0,78)
19	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	29	0	2.787	1,16 (1,00 bis 1,34)
20	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	3	0	1.441	0,86 (0,71 bis 1,05)
21	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	12	0	1.030	0,68 (0,44 bis 1,07)
22	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	24	0	2.231	0,69 (0,53 bis 0,91)
23	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	14	0	4.329	1,06 (0,79 bis 1,42)
24	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	OR	Fixed effects	7	0	617	1,13 (0,86 bis 1,49)
25	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	2	0	135	0,42 (0,26 bis 0,66)
26	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	12	0	743	1,03 (0,78 bis 1,35)
27	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Fixed effects	15	0	1.986	0,02 (0,01 bis 0,02)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

28	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Fixed effects	19	0	1.451	0,13 (-0,61 bis 0,88)
29	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	11	0	581	1,67 (1,15 bis 2,42)
30	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	9	0	1.960	1,02 (0,72 bis 1,44)
31	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Fixed effects	11	0	3.866	1,07 (-0,28 bis 2,42)
32	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	32	0	15.824	1,07 (0,97 bis 1,18)
33	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	29	0	13.518	0,88 (0,74 bis 1,05)
34	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	19	0	16.837	2,34 (2,02 bis 2,71)
35	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	10	0	1.471	1,24 (0,91 bis 1,70)
36	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	3	0	1.333	0,53 (0,40 bis 0,71)
37	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	86	0	7.621	1,14 (1,03 bis 1,26)
38	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	69	0	4.667	0,92 (0,81 bis 1,03)
39	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	32	0	2.911	0,94 (0,81 bis 1,09)
40	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	73	0	7.628	1,24 (1,13 bis 1,35)
41	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	44	0	6.044	1,10 (0,98 bis 1,23)
42	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	3	0	1.333	0,53 (0,40 bis 0,71)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

43	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	22	0	2.472	0,88 (0,80 bis 0,97)
44	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	22	0	2.205	1,83 (1,49 bis 2,23)
45	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	15	0	1.020	2,32 (1,84 bis 2,93)
46	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	27	0	2.472	0,82 (0,75 bis 0,89)
47	Nicht-adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Fixed effects	43	0	Nicht berichtet	-9,89 (-11,65 bis -8,13)
48	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	25	0	4.032	1,12 (0,80 bis 1,56)
49	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	24	0	4.507	1,06 (0,69 bis 1,65)
50	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	12	0	2.482	0,93 (0,58 bis 1,48)
51	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	17	0	3.043	1,15 (0,73 bis 1,80)
52	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	21	0	3.881	1,26 (0,85 bis 1,86)
53	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	66	0	84.698	1,02 (0,87 bis 1,20)
54	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	49	0	76.025	1,08 (0,94 bis 1,24)
55	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	79	0	87.369	0,90 (0,78 bis 1,04)
56	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	65	0	87.002	0,91 (0,80 bis 1,05)
57	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	20	0	2.062	0,77 (0,44 bis 1,33)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

58	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	SMD	Random effects	13	0	1.066	0,32 (-0,52 bis 1,16)
59	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	WMD	Random effects	48	0	11.052	0,01 (-0,65 bis 0,67)
60	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	10	0	54.597	0,75 (0,58 bis 0,96)
61	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	18	0	30.230	0,88 (0,41 bis 1,87)
62	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	35	0	25.632	1,33 (0,95 bis 1,86)
63	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	6	0	9.807	4,06 (0,16 bis 104,45)
64	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	30	0	44.179	0,90 (0,41 bis 1,81)
65	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	15	0	22.562	0,84 (0,12 bis 5,60)
66	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	13	0	22.065	0,18 (0,04 bis 0,91)
67	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	10	0	12.913	0,53 (0,30 bis 0,95)
68	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	29	0	34.232	0,93 (0,44 bis 1,97)
69	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	14	0	14.711	1,20 (0,85 bis 1,68)
70	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	30	0	7.829	1,67 (1,22 bis 2,28)
71	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	14	0	12.615	0,44 (0,03 bis 7,36)
72	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	5	0	1.140	8,87 (0,27 bis 293,96)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

73	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	24	0	7.440	1,12 (0,83 bis 1,51)
74	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	18	0	3.105	1,49 (1,08 bis 2,04)
75	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	8	0	2.002	0,72 (0,02 bis 33,22)
76	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	7	0	2.316	0,74 (0,49 bis 1,12)
77	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	24	0	2.049	0,80 (0,31 bis 2,11)
78	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	4	0	686	16,22 (0,46 bis 570,93)
79	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	14	0	2.744	0,43 (0,25 bis 0,75)
80	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	4	0	3.993	5,40 (0,53 bis 54,82)
81	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	13	0	3.185	0,26 (0,14 bis 0,48)
82	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	14	0	4.042	1,28 (0,56 bis 2,94)
83	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	14	0	4.042	0,26 (0,07 bis 0,91)
84	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	29	0	2.787	0,74 (0,27 bis 2,02)
85	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	3	0	1.441	0,73 (0,56 bis 0,96)
86	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	21	0	1.608	0,65 (0,35 bis 1,20)
87	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	9	0	1.855	1,39 (0,39 bis 4,93)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

88	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	9	0	1.855	1,00 (0,38 bis 2,66)
89	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	12	0	1.030	1,05 (0,25 bis 2,13)
90	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	24	0	2.231	1,00 (0,53 bis 1,89)
91	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	14	0	4.329	0,97 (0,68 bis 1,48)
92	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Fixed effects	14	0	1.868	0,74 (0,58 bis 0,94)
93	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	7	0	617	0,92 (0,36 bis 2,36)
94	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	2	0	135	0,70 (0,38 bis 1,28)
95	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	12	0	743	1,74 (0,59 bis 5,18)
96	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	WMD	Random effects	15	0	1.986	0,03 (0,01 bis 0,04)
97	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	WMD	Random effects	19	0	1.451	0,51 (-0,43 bis 1,45)
98	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	11	0	581	1,49 (0,35 bis 6,31)
99	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Nicht berichtet	27	0	5.602	0,68 (0,18 bis 1,19)
100	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	9	0	1.960	0,61 (0,40 bis 0,93)
101	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	SMD	Random effects	11	0	3.866	1,33 (-0,63 bis 3,29)
102	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	32	0	15.824	0,82 (0,56 bis 1,20)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

103	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	29	0	13.518	0,75 (0,53 bis 1,06)
104	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	19	0	16.837	1,51 (0,90 bis 2,56)
105	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	10	0	1.471	0,99 (0,47 bis 2,08)
106	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	3	0	1.333	0,45 (0,31 bis 0,66)
107	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	86	0	7.621	0,89 (0,77 bis 1,02)
108	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	69	0	4.667	0,77 (0,63 bis 0,93)
109	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	32	0	2.911	0,81 (0,60 bis 1,10)
110	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	73	0	7.628	1,00 (0,86 bis 1,17)
111	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	44	0	6.044	0,88 (0,70 bis 1,11)
112	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	3	0	1.333	0,45 (0,31 bis 0,66)
113	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Fixed effects	10	*	Nicht berichtet	4,51 (1,90 bis 10,71)
114	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	9	*	Nicht berichtet	3,15 (1,18 bis 8,40)
115	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	12	*	Nicht berichtet	5,01 (2,33 bis 10,77)
116	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	8	*	Nicht berichtet	1,09 (0,50 bis 2,38)
117	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Fixed effects	6	*	Nicht berichtet	0,89 (0,09 bis 8,63)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

118	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	5	*	Nicht berichtet	4,22 (1,69 bis 10,54)
119	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	53	0	24.364	1,01 (0,79 bis 1,29)
120	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	OR	Random effects	60	0	26.660	1,08 (0,68 bis 1,36)
121	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	22	0	2.472	1,53 (1,11 bis 2,12)
122	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	22	0	2.205	2,45 (1,65 bis 3,64)
123	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	15	0	1.020	1,80 (1,12 bis 2,89)
124	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	RR	Random effects	27	0	2.472	1,40 (0,94 bis 2,09)
125	Adjustierter indirekter Vergleich über einen Gesamteffektschätzer nach Bucher	WMD	Random effects	43	0	Nicht berichtet	-1,16 (-6,95 bis 4,64)
126	Adjustierter indirekter Vergleich nach Bucher (Ohne Einschluss von Head-to-head-trials)	SMD	Random effects	13	0	Nicht berichtet	-1,6 (-5,56 bis 2,36)
127	Adjustierter indirekter Vergleich nach Bucher (Ohne Einschluss von Head-to-head-Studien)	SMD	Random effects	31	0	Nicht berichtet	6,66 (-10,27 bis 23,59)
128	Adjustierter indirekter Vergleich nach Bucher (Ohne Einschluss von Head-to-head-Studien)	SMD	Random effects	16	0	Nicht berichtet	-8,27 (-25,57 bis 9,03)
129	Adjustierter indirekter Vergleich nach Bucher (Ohne Einschluss von Head-to-head-Studien)	SMD	Random effects	54	0	Nicht berichtet	2,2 (-4,64 bis 9,04)
130	Adjustierter indirekter Vergleich nach Bucher (Ohne Einschluss von Head-to-head-Studien)	SMD	Random effects	38	0	Nicht berichtet	0,35 (0,13 bis 0,57)
131	Adjustierter indirekter Vergleich (Mit Einschluss von drei-armigen Head-to-head-Studien mit A, B und Placebo)	SMD	Random effects	17	0	Nicht berichtet	-10,37 (-26,78 bis 5,84)
132	Adjustierter indirekter Vergleich (Mit Einschluss von drei-armigen Head-to-head-Studien mit A, B und Placebo)	SMD	Random effects	62	0	Nicht berichtet	1,63 (-4,35 bis 7,61)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

133	Adjustierter indirekter Vergleich (Mit Einschluss von dreiarmligen Head-to-head-Studien mit A, B und Placebo)	SMD	Random effects	16	0	Nicht berichtet	0,48 (-2,24 bis 3,20)
134	Adjustierter indirekter Vergleich (Mit Einschluss von dreiarmligen Head-to-head-Studien mit A, B und Placebo)	SMD	Random effects	37	0	Nicht berichtet	11,13 (-2,51 bis 24,77)
135	Adjustierter indirekter Vergleich (Mit Einschluss von dreiarmligen Head-to-head-Studien mit A, B und Placebo)	SMD	Random effects	45	0	Nicht berichtet	0,32 (0,14 bis 0,50)
136	Adjustierter indirekter Vergleich nach Song 2000 unter Einschluss aller verfügbaren Head-to-head-Studien	SMD	Random effects	16	0	Nicht berichtet	2,29 (0,44 bis 4,15)
137	Adjustierter indirekter Vergleich nach Song 2000 unter Einschluss aller verfügbaren Head-to-head-Studien	SMD	Random effects	39	0	Nicht berichtet	10,52 (0,88 bis 20,17)
138	Adjustierter indirekter Vergleich nach Song 2000 unter Einschluss aller verfügbaren Head-to-head-Studien	SMD	Random effects	19	0	Nicht berichtet	-4,82 (-17,74 bis 8,10)
139	Adjustierter indirekter Vergleich nach Song 2000 unter Einschluss aller verfügbaren Head-to-head-Studien	SMD	Random effects	65	0	Nicht berichtet	1,79 (-2,22 bis 5,81)
140	Adjustierter indirekter Vergleich nach Song 2000 unter Einschluss aller verfügbaren Head-to-head-Studien	SMD	Random effects	49	0	Nicht berichtet	0,19 (0,06 bis 0,32)
141	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests	OR	Nicht berichtet	166	0	83.204	0,99 (0,89 bis 1,09)
142	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests	RR	Fixed effects	6	0	957	1,08 (0,78 bis 1,49)
143	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests	ARR	Fixed effects	Nicht berichtet	0	Nicht berichtet	0,03 (-0,32 bis 0,25)
144	Adjustierter indirekter Vergleich mithilfe eines statistischen Tests	RR	Random effects	25	1	1.966	0,61 (0,30 bis 1,27)
145	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	8	0	3.343	-6,7 (-10,7 bis -3,23)
146	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	16	0	6.460	9,40 (6,39 bis 12,41)
147	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	14	0	5.757	6,30 (3,03 bis 9,57)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

148	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	17	0	5.701	1,10 (-4,28 bis 6,48)
149	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	12	0	3.414	9,40 (6,11 bis 12,69)
150	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	15	0	5.273	-0,90 (-3,28 bis 1,48)
151	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	14	0	4.447	4,10 (1,12 bis 7,08)
152	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	15	0	4.719	-5,00 (-7,69 bis -2,31)
153	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	20	0	6.861	-5,30 (-7,53 bis -3,07)
154	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	14	0	4.020	-1,90 (-4,97 bis 1,17)
155	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	OR	Fixed effects	Nicht berichtet	0	6.868	0,84 (0,67 bis 1,07)
156	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	OR	Fixed effects	Nicht berichtet	0	5.248	1,24 (0,86 bis 1,80)
157	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	OR	Fixed effects	Nicht berichtet	0	1.202	1,05 (0,66 bis 1,67)
158	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	WMD	Random effects	47	0	3.531	2,10 (-9,98 bis 14,18)
159	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Fixed effects	78	0	5.284	-2,71 (-13,82 bis 8,40)
160	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Fixed effects	48	0	3.763	6,00 (-6,98 bis 18,98)
161	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	22	0	1.541	1,00 (0,61 bis 1,63)
162	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Random effects	80	0	5.372	2,55 (-4,36 bis 9,46)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

163	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	SMD	Random effects	50	0	3.851	13,20 (5,53 bis 20,87)
164	Adjustierter indirekter Vergleich durch Betrachtung der Überlappung von Konfidenzintervallen	RR	Fixed effects	22	0	1.618	2,70 (1,26 bis 5,82)
165	Metaregression	WMD	Random effects	6	0	3.326	0,09 (-0,32 bis 0,51)
166	Metaregression	WMD	Random effects	8	0	4.592	-1,68 (-4,77 bis 1,40)
167	Metaregression	WMD	Random effects	8	0	4.592	-0,64 (-1,59 bis 0,30)
168	Metaregression	WMD	Random effects	8	0	4.592	-0,53 (-1,48 bis 0,42)
169	Logistische (Meta-)Regression (Mixed model)	OR	Nicht berichtet	9	0	1.359	1,19 (0,81 bis 1,75)
170	Logistische (Meta-)Regression (Mixed model)	OR	Nicht berichtet	8	0	898	0,67 (0,42 bis 1,06)
171	MTC	OR	Random effects	22	0	143.153	0,12 (0,01 bis 0,97)
172	MTC	OR	Random effects	22	0	143.153	1,03 (0,82 bis 1,29)
173	MTC	OR	Random effects	22	0	143.153	0,82 (0,70 bis 0,95)
174	MTC	OR	Random effects	22	0	143.153	0,66 (0,56 bis 0,77)
175	MTC	HR	Nicht berichtet	36	0	15.158	0,56 (0,41 bis 0,76)
176	MTC	HR	Nicht berichtet	36	0	15.158	0,64 (0,56 bis 0,73)
177	MTC	HR	Nicht berichtet	36	0	15.158	0,73 (0,65 bis 0,82)
178	MTC	HR	Nicht berichtet	34	0	13.875	1,00 (0,67 bis 1,54)
179	MTC	HR	Nicht berichtet	34	0	13.875	0,78 (0,60 bis 1,03)
180	MTC	HR	Nicht berichtet	34	0	13.875	1,00 (0,86 bis 1,17)
181	MTC	HR	Nicht berichtet	36	0	15.158	1,42 (1,08 bis 1,86)
182	MTC	HR	Nicht berichtet	36	0	15.158	1,07 (0,88 bis 1,30)
183	MTC	HR	Nicht berichtet	34	0	13.875	0,92 (0,84 bis 1,01)
184	MTC	HR	Nicht berichtet	34	0	13.875	0,87 (0,78 bis 0,98)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

185	MTC	HR	Random effects	60	0	16.478	0,68 (0,59 bis 0,78)
186	MTC	HR	Random effects	60	0	16.478	0,70 (0,62 bis 0,80)
187	MTC	HR	Random effects	60	0	16.478	0,87 (0,78 bis 0,97)
188	MTC	OR	Random effects	12	0	8.307	0,57 (0,40 bis 0,80)
189	MTC	OR	Random effects	12	0	8.307	0,82 (0,57 bis 1,18)
190	MTC	OR	Random effects	12	0	8.307	0,85 (0,60 bis 1,22)
191	MTC	OR	Random effects	12	0	8.307	0,66 (0,50 bis 0,89)
192	MTC	OR	Random effects	12	0	8.307	0,69 (0,55 bis 0,87)
193	MTC	RR	Random effects	14	0	Nicht berichtet	0,86 (0,63 bis 1,19)
194	MTC	RR	Random effects	22	0	Nicht berichtet	0,96 (0,84 bis 1,09)
195	MTC	RR	Random effects	24	0	Nicht berichtet	0,83 (0,67 bis 1,02)
196	MTC	RR	Random effects	24	0	Nicht berichtet	0,94 (0,76 bis 1,16)
197	MTC	RR	Random effects	25	0	Nicht berichtet	0,85 (0,69 bis 1,04)
198	MTC	RR	Random effects	28	0	Nicht berichtet	0,88 (0,77 bis 1,01)
199	MTC	RR	Random effects	26	0	Nicht berichtet	0,91 (0,72 bis 1,15)
200	MTC	RR	Random effects	23	0	Nicht berichtet	0,99 (0,88 bis 1,13)
201	MTC	RR	Random effects	25	0	Nicht berichtet	0,96 (0,78 bis 1,17)
202	MTC	RR	Random effects	29	0	Nicht berichtet	0,90 (0,78 bis 1,03)
203	MTC	RR	Random effects	15	0	Nicht berichtet	0,85 (0,68 bis 1,06)
204	MTC	RR	Random effects	25	0	Nicht berichtet	0,81 (0,66 bis 0,99)
205	MTC	HR	Random effects	38	0	18.023	1,00 (0,84 bis 1,23)
206	MTC	HR	Random effects	38	0	18.023	1,03 (0,84 bis 1,22)
207	MTC	HR	Random effects	37	0	17.962	1,00 (0,81 bis 1,23)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

208	MTC	HR	Random effects	37	0	17.712	0,42 (0,33 bis 0,53)
209	MTC	HR	Random effects	36	0	17.705	1,05 (0,80 bis 1,36)
210	MTC	HR	Random effects	38	0	18.023	0,92 (0,77 bis 1,08)
211	MTC	HR	Random effects	38	0	18.023	1,00 (0,82 bis 1,25)
212	MTC	HR	Random effects	37	0	17.962	0,81 (0,66 bis 0,97)
213	MTC	HR	Random effects	37	0	17.712	0,30 (0,24 bis 0,37)
214	MTC	HR	Random effects	36	0	17.705	1,02 (0,80 bis 1,31)
215	MTC	HR	Random effects	24	0	12.973	1,38 (0,96 bis 2,24)
216	MTC	HR	Random effects	24	0	12.973	1,00 (0,68 bis 1,63)
217	MTC	HR	Random effects	38	0	18.023	0,92 (0,79 bis 1,08)
218	MTC	HR	Random effects	38	0	18.023	0,96 (0,83 bis 1,24)
219	MTC	HR	Random effects	37	0	17.962	0,83 (0,71 bis 1,00)
220	MTC	HR	Random effects	37	0	17.712	0,70 (0,56 bis 0,84)
221	MTC	HR	Random effects	36	0	17.705	0,99 (0,74 bis 1,26)
222	MTC	HR	Random effects	24	0	12.973	0,71 (0,48 bis 1,13)
223	MTC	SMD	Random effects	19	0	Nicht berichtet	-3,91 (-19,8 bis 9,86)
224	MTC	SMD	Random effects	39	0	Nicht berichtet	9,98 (-0,53 bis 20,26)
225	MTC	SMD	Random effects	65	0	Nicht berichtet	1,74 (-3,33 bis 6,62)
226	MTC	SMD	Random effects	16	0	Nicht berichtet	1,50 (-1,12 bis 4,09)
227	MTC	SMD	Random effects	49	0	Nicht berichtet	0,22 (0,08 bis 0,37)
228	Sonstige Methoden	OR	Nicht berichtet	9	0	Nicht berichtet	0,71 (0,31 bis 1,38)
229	Sonstige Methoden	HR	Fixed effects	17	0	4.265	0,70 (0,58 bis 0,83)
230	Sonstige Methoden	HR	Fixed effects	19	0	3.959	1,29 (1,13 bis 1,48)

Fortsetzung Tabelle 29: Hintergrundinformationen zu den indirekt vergleichenden Metaanalysen aus den Typ-5-Publikationen

231	Sonstige Methoden	HR	Fixed effects	17	0	4.265	0,63 (0,48 bis 0,82)
232	Sonstige Methoden	HR	Fixed effects	14	0	3.526	1,01 (0,77 bis 1,31)
233	Sonstige Methoden	HR	Fixed effects	19	0	3.959	0,77 (0,61 bis 0,96)
234	Sonstige Methoden	HR	Fixed effects	17	0	4.265	0,76 (0,43 bis 1,02)
235	Sonstige Methoden	HR	Fixed effects	14	0	3.526	0,85 (0,59 bis 1,06)
236	Sonstige Methoden	HR	Fixed effects	19	0	3.959	0,96 (0,76 bis 1,12)
237	Sonstige Methoden	HR	Fixed effects	17	0	4.265	1,00 (0,76 bis 1,18)
238	Sonstige Methoden	HR	Fixed effects	14	0	3.526	0,80 (0,58 bis 0,99)
239	Sonstige Methoden	HR	Fixed effects	19	0	3.959	0,77 (0,59 bis 0,94)
240	Sonstige Methoden	HR	Fixed effects	17	0	1.552	1,45 (1,07 bis 1,96)
241	Sonstige Methoden	HR	Fixed effects	14	0	1.360	1,00 (0,81 bis 1,22)
242	Sonstige Methoden	HR	Fixed effects	19	0	1.765	1,21 (0,99 bis 1,47)
243	Sonstige Methoden	HR	Fixed effects	17	0	1.552	0,92 (0,72 bis 1,18)
244	Sonstige Methoden	HR	Fixed effects	14	0	1.360	1,03 (0,71 bis 1,51)
245	Sonstige Methoden	HR	Fixed effects	19	0	1.765	0,97 (0,77 bis 1,23)
246	Sonstige Methoden	OR	Fixed effects	24	0	42.688	0,78 (0,70 bis 0,87)
247	Sonstige Methoden	OR	Fixed effects	24	0	42.688	0,95 (0,84 bis 1,07)
248	Sonstige Methoden	OR	Fixed effects	24	0	42.688	0,93 (0,85 bis 1,02)
249	Sonstige Methoden	OR	Fixed effects	24	0	42.688	0,88 (0,77 bis 1,00)

*Im gesamten Review sind 69 von 99 kontrollierten Studien auch als RCT eingestuft worden.

ARR = Absolute Risikoreduktion. HR = Hazard ratio. MTC = Mixed treatment comparison. OR = Odds ratio. RCT = Randomisierte kontrollierte Studie. SMD = Standardisierte Mittelwertsdifferenz

Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

(Zugehörige Therapieeffektunterschiede sind im Forest-Plot unter den entsprechenden Metaanalysenummern dargestellt (siehe Kapitel 5.5.2 „Methodenspezifische Validitätsprüfung“))

Meta-analyse-nummer	Effektschätzer	Random oder Fixed effects?	Gesamtanzahl Studien, die in den direkten Vergleich eingeschlossen werden	Anzahl Studien für den direkten Vergleich, die nicht RCT sind	Gesamtanzahl Probanden, die in den direkten Vergleich eingeschlossen werden	Schätzung des Therapieeffektunterschiedes		
						Erwartungswert	Untere Konfidenzintervallgrenze	Obere Konfidenzintervallgrenze
1	OR	Fixed effects	5	0	933	0,57	0,32	0,99
2	OR	Fixed effects	5	0	933	0,68	0,47	0,98
3	OR	Fixed effects	5	0	933	1,93	1,44	2,59
4	WMD	Random effects	5	0	933	283,60	212,90	354,30
5	RR	Random effects	2	Nicht berichtet	82	1,16	0,43	3,16
6	RR	Nur eine Studie	1	Nicht berichtet	310	0,94	0,62	1,43
7	RR	Nur eine Studie	1	Nicht berichtet	309	1,15	0,76	1,74
8	RR	Random effects	2	Nicht berichtet	493	1,01	0,71	1,45
9	RR	Random effects	10	Nicht berichtet	2.531	1,03	0,84	1,27
10	RR	Random effects	4	Nicht berichtet	1.163	1,17	0,88	1,54
11	RR	Random effects	3	Nicht berichtet	2.425	0,96	0,81	1,15
12	RR	Random effects	3	Nicht berichtet	3.471	0,71	0,38	1,34
13	RR	Random effects	16	Nicht berichtet	5.669	1,01	0,87	1,16
14	RR	Random effects	9	Nicht berichtet	481	0,67	0,51	0,89
15	SMD	Nur eine Studie	1	Nicht berichtet	134	0,09	-0,25	0,43
16	RR	Random effects	8	Nicht berichtet	1.618	0,45	0,22	0,91
17	RR	Nur eine Studie	1	Nicht berichtet	1.276	2,19	1,00	4,77
18	RR	Random effects	13	Nicht berichtet	1.615	0,56	0,48	0,66
19	RR	Random effects	5	Nicht berichtet	401	1,39	1,08	1,79

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

20	RR	Random effects	3	Nicht berichtet	1.267	0,64	0,49	0,82
21	RR	Random effects	3	Nicht berichtet	247	0,91	0,36	2,31
22	RR	Random effects	2	Nicht berichtet	226	0,53	0,12	2,36
23	RR	Random effects	2	Nicht berichtet	681	1,07	0,94	1,22
24	OR	Random effects	7	Nicht berichtet	683	0,53	0,36	0,78
25	RR	Nur eine Studie	1	Nicht berichtet	39	0,82	0,55	1,22
26	RR	Random effects	11	Nicht berichtet	622	1,24	1,01	1,54
27	WMD	Random effects	4	Nicht berichtet	248	0,03	0,01	0,05
28	WMD	Random effects	3	Nicht berichtet	202	1,22	0,00	2,45
29	RR	Random effects	4	Nicht berichtet	505	2,20	1,52	3,17
30	RR	Nur eine Studie	1	Nicht berichtet	425	0,28	0,15	0,51
31	SMD	Nur eine Studie	1	Nicht berichtet	339	1,80	-1,43	5,03
32	RR	Nur eine Studie	1	Nicht berichtet	106	0,70	0,36	1,35
33	RR	Nur eine Studie	1	Nicht berichtet	124	0,97	0,68	1,38
34	RR	Random effects	5	Nicht berichtet	1.254	1,51	1,08	2,12
35	RR	Nur eine Studie	1	Nicht berichtet	100	2,74	1,25	6,02
36	RR	Random effects	3	Nicht berichtet	1.892	0,67	0,57	0,80
37	RR	Nur eine Studie	1	Nicht berichtet	100	0,78	0,18	3,31
38	RR	Nur eine Studie	1	Nicht berichtet	120	0,80	0,45	1,35
39	RR	Nur eine Studie	1	Nicht berichtet	97	0,40	0,18	0,86
40	RR	Random effects	5	Nicht berichtet	655	1,00	0,73	1,37
41	RR	Random effects	3	Nicht berichtet	539	0,90	0,38	2,14
42	RR	Random effects	3	Nicht berichtet	1.892	0,67	0,57	0,80
43	RR	Nur eine Studie	1	Nicht berichtet	44	2,40	1,39	4,13

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

44	RR	Nur eine Studie	1	Nicht berichtet	64	2,29	1,16	4,52
45	RR	Nur eine Studie	1	Nicht berichtet	86	1,90	1,30	2,77
46	RR	Random effects	3	Nicht berichtet	235	1,08	0,79	1,48
47	WMD	Random effects	13	Nicht berichtet	897	6,97	3,56	10,37
48	RR	Random effects	2	Nicht berichtet	82	1,16	0,43	3,16
49	RR	Nur eine Studie	1	Nicht berichtet	310	0,94	0,62	1,43
50	RR	Nur eine Studie	1	Nicht berichtet	309	1,15	0,76	1,74
51	RR	Random effects	2	Nicht berichtet	493	1,01	0,71	1,45
52	RR	Random effects	10	Nicht berichtet	2.531	1,03	0,84	1,27
53	RR	Random effects	4	Nicht berichtet	1.163	1,17	0,88	1,54
54	RR	Random effects	3	Nicht berichtet	2.425	0,96	0,81	1,15
55	RR	Random effects	3	Nicht berichtet	3.471	0,71	0,38	1,34
56	RR	Random effects	16	Nicht berichtet	5.669	1,01	0,87	1,16
57	RR	Random effects	9	Nicht berichtet	481	0,67	0,51	0,89
58	SMD	Nur eine Studie	1	Nicht berichtet	134	0,09	-0,25	0,43
59	WMD	Nur eine Studie	1	0	677	-0,36	-2,08	1,36
60	RR	Nur eine Studie	1	0	2.649	1,01	0,68	1,50
61	RR	Nur eine Studie	1	0	90	0,25	0,03	2,15
62	RR	Nur eine Studie	1	0	90	0,50	0,16	1,54
63	RR	Nur eine Studie	1	0	49	2,67	0,11	62,42
64	RR	Nur eine Studie	1	0	289	0,22	0,01	5,38
65	RR	Nur eine Studie	1	0	287	2,01	0,84	4,84
66	RR	Nur eine Studie	1	0	287	1,29	0,50	3,38
67	RR	Nur eine Studie	1	0	287	0,61	0,33	1,14

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

68	RR	Random effects	2	0	909	3,05	1,03	9,06
69	RR	Nur eine Studie	1	0	483	1,68	1,32	2,13
70	RR	Random effects	3	0	454	4,35	1,51	12,55
71	RR	Random effects	2	0	972	1,51	0,06	36,71
72	RR	Nur eine Studie	1	0	118	2,90	0,12	69,81
73	RR	Random effects	2	0	972	1,08	0,50	2,32
74	RR	Nur eine Studie	1	0	425	3,11	1,62	5,95
75	RR	Nur eine Studie	1	0	425	0,33	0,01	7,95
76	RR	Random effects	2	0	587	0,85	0,74	0,97
77	RR	Random effects	2	0	934	0,37	0,21	0,65
78	RR	Nur eine Studie	1	0	425	0,33	0,01	7,95
79	RR	Random effects	8	Nicht berichtet	1.618	0,45	0,22	0,91
80	RR	Nur eine Studie	1	Nicht berichtet	1.276	2,19	1,00	4,77
81	RR	Random effects	13	Nicht berichtet	1.615	0,56	0,48	0,66
82	OR	Random effects	12	0	3.337	0,87	0,56	1,35
83	OR	Random effects	12	0	3337	1,60	1,31	1,96
84	RR	Random effects	5	Nicht berichtet	401	1,39	1,08	1,79
85	RR	Random effects	3	Nicht berichtet	1.267	0,64	0,49	0,82
86	RR	Random effects	5	Nicht berichtet	636	1,21	0,88	1,67
87	RR	Nur eine Studie	1	0	212	1,09	0,75	1,58
88	RR	Nur eine Studie	1	0	212	1,36	1,03	1,79
89	RR	Random effects	3	Nicht berichtet	247	0,91	0,36	2,31
90	RR	Random effects	2	Nicht berichtet	226	0,53	0,12	2,36
91	RR	Random effects	2	Nicht berichtet	681	1,07	0,94	1,22

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

92	RR	Fixed effects	3	0	132	0,58	0,40	0,84
93	OR	Random effects	7	Nicht berichtet	682	0,53	0,36	0,78
94	RR	Nur eine Studie	1	Nicht berichtet	39	0,82	0,55	1,22
95	RR	Random effects	10	Nicht berichtet	622	1,24	1,01	1,54
96	WMD	Random effects	4	Nicht berichtet	248	0,03	0,01	0,05
97	WMD	Random effects	3	Nicht berichtet	202	1,22	0,00	2,45
98	RR	Random effects	4	Nicht berichtet	505	2,20	1,52	3,17
99	RR	Nicht berichtet	7	0 Nicht berichtet		0,58	0,34	0,98
100	RR	Nur eine Studie	1	Nicht berichtet	425	0,28	0,15	0,51
101	SMD	Nur eine Studie	1	Nicht berichtet	339	1,80	-1,43	5,03
102	RR	Nur eine Studie	1	Nicht berichtet	106	0,70	0,36	1,35
103	RR	Nur eine Studie	1	Nicht berichtet	124	0,97	0,68	1,38
104	RR	Random effects	5	Nicht berichtet	1.254	1,51	1,08	2,12
105	RR	Nur eine Studie	1	Nicht berichtet	100	2,74	1,25	6,02
106	RR	Random effects	3	Nicht berichtet	1.892	0,67	0,57	0,80
107	RR	Nur eine Studie	1	Nicht berichtet	100	0,78	0,18	3,31
108	RR	Nur eine Studie	1	Nicht berichtet	120	0,80	0,45	1,35
109	RR	Nur eine Studie	1	Nicht berichtet	97	0,40	0,18	0,86
110	RR	Random effects	5	Nicht berichtet	655	1,00	0,73	1,37
111	RR	Random effects	3	Nicht berichtet	539	0,90	0,38	2,14
112	RR	Random effects	3	Nicht berichtet	1.892	0,67	0,57	0,80
113	OR	Fixed effects	2	Nicht berichtet	172	3,12	1,55	6,29
114	OR	Fixed effects	7	Nicht berichtet	310	4,56	2,76	7,52
115	OR	Fixed effects	3	Nicht berichtet	243	1,20	0,72	2,00

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

116	OR	Fixed effects	7	Nicht berichtet	310	4,56	2,76	7,52
117	OR	Fixed effects	2	Nicht berichtet	172	3,12	1,55	6,29
118	OR	Fixed effects	3	Nicht berichtet	243	1,20	0,72	2,00
119	OR	Nur eine Studie	1	0	60	2,66	1,70	4,15
120	OR	Random effects	2	0	548	1,14	0,20	6,42
121	RR	Nur eine Studie	1	Nicht berichtet	44	2,40	1,39	4,13
122	RR	Nur eine Studie	1	Nicht berichtet	64	2,29	1,16	4,52
123	RR	Nur eine Studie	1	Nicht berichtet	86	1,90	1,30	2,77
124	RR	Random effects	3	Nicht berichtet	235	1,08	0,79	1,48
125	WMD	Random effects	13	Nicht berichtet	897	6,97	3,56	10,37
126	SMD	Random effects	3	0	Nicht berichtet	3,39	1,29	5,49
127	SMD	Random effects	8	0	Nicht berichtet	12,38	0,64	24,12
128	SMD	Random effects	3	0	Nicht berichtet	-0,46	-19,88	18,97
129	SMD	Random effects	11	0	Nicht berichtet	1,58	-3,38	6,54
130	SMD	Random effects	11	0	Nicht berichtet	0,11	-0,04	0,27
131	SMD	Random effects	3	0	Nicht berichtet	-0,46	-19,88	18,97
132	SMD	Random effects	11	0	Nicht berichtet	1,58	-3,38	6,54
133	SMD	Random effects	3	0	Nicht berichtet	3,39	1,29	5,49
134	SMD	Random effects	8	0	Nicht berichtet	12,38	0,64	24,12
135	SMD	Random effects	11	0	Nicht berichtet	0,11	-0,04	0,27
136	SMD	Random effects	3	0	Nicht berichtet	3,39	1,29	5,49
137	SMD	Random effects	8	0	Nicht berichtet	12,38	0,64	24,12
138	SMD	Random effects	3	0	Nicht berichtet	-0,46	-19,88	18,97
139	SMD	Random effects	11	0	Nicht berichtet	1,58	-3,38	6,54

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

140	SMD	Random effects	11	0	Nicht berichtet	0,11	-0,04	0,27
141	OR	Nicht berichtet	27	0	34.452	0,92	0,85	0,99
142	RR	Fixed effects	3	0	716	1,30	1,10	1,50
143	ARR	Nur eine Studie	1	0	Nicht berichtet	0,15	0,07	0,25
144	RR	Nur eine Studie	1	0	70	0,94	0,20	4,34
145	WMD	Nur eine Studie	1	0	154	1,00	-12,00	15,00
146	WMD	Nur eine Studie	1	0	414	12,00	4,00	20,00
147	WMD	Nur eine Studie	1	0	581	9,00	1,00	16,00
148	WMD	Nur eine Studie	1	0	551	12,00	4,00	19,00
149	WMD	Nur eine Studie	1	0	244	4,00	-4,00	12,00
150	WMD	Random effects	2	0	561	-8,00	-14,00	-2,00
151	WMD	Random effects	2	0	521	3,00	-3,00	9,00
152	WMD	Random effects	2	0	553	-12,00	-20,00	-5,00
153	WMD	Random effects	2	0	933	-4,00	-9,00	2,00
154	WMD	Nur eine Studie	1	0	989	-1,00	-6,00	5,00
155	OR	Nur eine Studie	1	0	143	7,71	0,79	75,35
156	OR	Nur eine Studie	1	0	533	1,22	0,86	1,73
157	OR	Nur eine Studie	1	0	533	0,99	0,69	1,41
158	WMD	Random effects	2	0	139	8,48	0,95	16,01
159	SMD	Nur eine Studie	1	0	134	-1,33	-11,96	9,30
160	SMD	Nur eine Studie	1	0	141	0,21	-13,32	13,74
161	RR	Nur eine Studie	1	0	64	1,19	0,87	1,62
162	SMD	Random effects	3	0	224	5,28	-0,87	11,43
163	SMD	Random effects	3	0	224	6,25	-1,58	14,08

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

164	RR	Nur eine Studie	1	0	134	1,06	0,74	1,52
165	WMD	Nur eine Studie	1	0	339	0,00	-0,23	0,23
166	WMD	Random effects	2	0	746	-0,86	-3,82	2,11
167	WMD	Random effects	2	0	746	0,39	-0,60	1,37
168	WMD	Random effects	2	0	746	-0,31	-1,23	0,61
169	OR	Nur eine Studie	1	0	391	0,81	0,50	1,32
170	OR	Nicht berichtet	2	0	377	1,44	0,86	2,41
171	OR	Nur eine Studie	1	0	392	0,57	0,46	0,72
172	OR	Random effects	2	0	8.752	0,90	0,75	1,09
173	OR	Random effects	2	0	13.069	0,75	0,62	0,90
174	OR	Random effects	2	0	13.676	0,67	0,56	0,80
175	HR	Random effects	2	0	Nicht berichtet	0,51	0,38	0,70
176	HR	Random effects	6	0	Nicht berichtet	0,69	0,62	0,76
177	HR	Random effects	9	0	Nicht berichtet	0,72	0,67	0,77
178	HR	Nur eine Studie	1	0	Nicht berichtet	0,97	0,46	2,03
179	HR	Random effects	2	0	Nicht berichtet	0,81	0,60	1,08
180	HR	Random effects	3	0	Nicht berichtet	0,99	0,73	1,34
181	HR	Nur eine Studie	1	0	Nicht berichtet	1,39	1,03	1,87
182	HR	Random effects	3	0	Nicht berichtet	1,04	0,71	1,53
183	HR	Random effects	9	0	Nicht berichtet	0,88	0,82	0,94
184	HR	Random effects	6	0	Nicht berichtet	0,96	0,86	1,09
185	HR	Nicht berichtet	3	0	Nicht berichtet	0,63	0,48	0,83
186	HR	Nicht berichtet	6	0	Nicht berichtet	0,78	0,64	0,94
187	HR	Nicht berichtet	11	0	Nicht berichtet	0,83	0,74	0,94

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

188	OR	Nur eine Studie	1	0	903	0,64	0,46	0,90
189	OR	Random effects	3	0	1.045	0,81	0,48	1,37
190	OR	Nur eine Studie	1	0	1.212	0,79	0,60	1,06
191	OR	Random effects	4	0	2.249	0,67	0,50	0,90
192	OR	Random effects	4	0	3.470	0,69	0,58	0,81
193	RR	Nur eine Studie	1	0	6.083	1,13	0,82	1,56
194	RR	Nur eine Studie	1	0	6.083	1,08	0,89	1,30
195	RR	Nur eine Studie	1	0	6.083	0,96	0,73	1,24
196	RR	Nur eine Studie	1	0	6.083	1,13	0,93	1,38
197	RR	Nur eine Studie	1	0	6.083	0,98	0,72	1,33
198	RR	Nur eine Studie	1	0	6.083	1,08	0,97	1,21
199	RR	Random effects	3	0	8.149	0,83	0,59	1,26
200	RR	Random effects	4	0	9.032	0,99	0,80	1,22
201	RR	Random effects	4	0	9.032	1,03	0,77	1,38
202	RR	Random effects	4	0	9.032	0,89	0,78	1,01
203	RR	Random effects	4	0	9.032	0,50	0,27	0,92
204	RR	Random effects	4	0	9.032	0,83	0,61	1,11
205	RR	Random effects	8	0	4.874	1,01	0,85	1,20
206	RR	Random effects	8	0	4.874	0,91	0,72	1,17
207	RR	Random effects	8	0	4.874	1,06	0,83	1,34
208	RR	Random effects	8	0	4.874	0,58	0,46	0,72
209	RR	Random effects	8	0	4.874	0,90	0,65	1,26
210	RR	Random effects	17	0	5.537	0,98	0,82	1,16
211	RR	Random effects	17	0	5.537	1,12	0,88	1,44

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

212	RR	Random effects	17	0	5.537	0,86	0,67	1,09
213	RR	Random effects	17	0	5.537	0,40	0,32	0,51
214	RR	Random effects	17	0	5.537	1,21	0,86	1,69
215	RR	Random effects	8	0	4.874	1,20	0,68	2,11
216	RR	Random effects	17	0	5.537	1,29	0,80	2,07
217	RR	Random effects	15	0	8.438	0,87	0,75	1,01
218	RR	Random effects	15	0	8.438	0,94	0,75	1,17
219	RR	Random effects	15	0	8.438	0,84	0,69	1,02
220	RR	Random effects	15	0	8.438	0,76	0,66	0,88
221	RR	Random effects	15	0	8.438	0,88	0,66	1,17
222	RR	Random effects	15	0	8.438	0,69	0,45	1,06
223	SMD	Random effects	3	0	Nicht berichtet	1,99	-28,02	25,24
224	SMD	Random effects	8	0	Nicht berichtet	10,53	-0,38	24,50
225	SMD	Random effects	11	0	Nicht berichtet	1,62	-4,81	6,88
226	SMD	Random effects	3	0	Nicht berichtet	3,40	0,39	6,24
227	SMD	Random effects	11	0	Nicht berichtet	0,11	-0,06	0,28
228	HR	Nur eine Studie	1	0	4.809	0,96	0,82	1,11
229	HR	Fixed effects	4	0	870	0,62	0,45	0,86
230	HR	Fixed effects	4	0	265	1,28	0,98	1,66
231	HR	Fixed effects	4	0	519	1,60	1,18	2,17
232	HR	Fixed effects	4	0	517	1,03	0,72	1,49
233	HR	Fixed effects	4	0	517	0,71	0,55	0,91
234	HR	Fixed effects	4	0	591	0,97	0,74	1,28

Fortsetzung Tabelle 30: Hintergrundinformationen zu den direkt vergleichenden Metaanalysen aus den Typ-5-Publikationen

235	HR	Fixed effects	3	0	551	1,00	0,78	1,29
236	HR	Fixed effects	3	0	551	1,10	0,87	1,39
237	HR	Fixed effects	5	0	813	1,00	0,79	1,26
238	HR	Fixed effects	5	0	830	0,82	0,67	1,00
239	HR	Fixed effects	5	0	830	1,22	1,04	1,44
240	HR	Fixed effects	4	0	382	0,89	0,61	1,29
241	HR	Fixed effects	4	0	395	0,96	0,75	1,24
242	HR	Fixed effects	4	0	395	0,86	0,68	1,09
243	HR	Fixed effects	5	0	341	0,98	0,60	1,58
244	HR	Fixed effects	4	0	270	1,06	0,71	1,57
245	HR	Fixed effects	5	0	395	0,80	0,59	1,08
246	OR	Fixed effects	6	0	7.795	0,79	0,70	0,90
247	OR	Fixed effects	2	0	7.869	0,97	0,84	1,10
248	OR	Fixed effects	4	0	11.632	0,77	0,58	1,03
249	OR	Fixed effects	2	0	13.798	0,83	0,67	1,03

HR = Hazard ratio. OR = Odds ratio. RCT = Randomisierte kontrollierte Studie. RR = Relatives Risiko. SMD = Standardisierte mittlere Differenz. WMD = Gewichtete mittlere Differenz.

Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

Meta-analyse-nummer	Zeigen der direkte und der indirekte Vergleich gleichermaßen einen signifikanten bzw. nicht-signifikanten Therapieeffektunterschied an?	Zu- bzw. Abnahme der Weite des Konfidenzintervalls des Therapieeffektunterschieds aus dem indirekten Vergleich in Relation zu derjenigen aus dem direkten Vergleich in Prozent (binäre Daten logarithmiert)	Faktor, um den die Anzahl der in den indirekten Vergleich eingeschlossenen Studien größer ist als die des direkten Vergleichs	Faktor, um den die Anzahl der in den indirekten Vergleich eingeschlossenen Probanden größer ist als die des direkten Vergleichs	z-Wert*	Anmerkungen zur Methodik des Vergleichs und eigenen Berechnungen
1	Ja, der Therapieeffektunterschied ist signifikant.	125 %	0,4	Nicht berechenbar aufgrund fehlender Daten	2,46*	In direktem und indirektem Vergleich Fixed und Random effects-Modelle unterschiedlich gewählt.
2	Ja, der Therapieeffektunterschied ist signifikant.	65 %	0,4	Nicht berechenbar aufgrund fehlender Daten	1,14	
3	Ja, der Therapieeffektunterschied ist signifikant.	170 %	0,4	Nicht berechenbar aufgrund fehlender Daten	-0,76	
4	Ja, der Therapieeffektunterschied ist signifikant.	208 %	0,4	Nicht berechenbar aufgrund fehlender Daten	-1,17	
5	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-76 %	12,5	49,2	0,83	
6	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-25 %	24,0	14,5	-1,85	
7	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-38 %	12,0	8,0	2,68*	
8	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-7 %	8,5	6,2	0,89	
9	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	12 %	2,1	1,5	-2,50*	
10	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-59 %	16,5	72,8	-0,78	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

11	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-65 %	16,3	31,4	-5,66*
12	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-81 %	26,3	25,2	-0,76
13	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-41 %	4,1	15,3	-1,21
14	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-11 %	2,2	4,3	-1,78
15	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-51 %	13,0	8,0	-0,57
16	Ja, der Therapieeffektunterschied ist signifikant.	-36 %	1,8	1,7	-0,47
17	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	258 %	4,0	3,1	-1,49
18	Ja, der Therapieeffektunterschied ist signifikant.	-57 %	1,0	2,0	-3,00*
19	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-42 %	5,8	7,0	1,21
20	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-24 %	1,0	1,1	-1,79
21	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-52 %	4,0	4,2	0,55
22	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-82 %	12,0	9,9	-0,34
23	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	125 %	7,0	6,4	0,06

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

24	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-29 %	1,0	0,9	-3,13*
25	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	17 %	2,0	3,5	2,14*
26	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	30 %	1,1	1,2	1,05
27	Ja, der Therapieeffektunterschied ist signifikant.	-82 %	3,8	8,0	1,14
28	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-39 %	6,3	7,2	1,49
29	Ja, der Therapieeffektunterschied ist signifikant.	1 %	2,8	1,2	1,03
30	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-43 %	9,0	4,6	-3,60*
31	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-58 %	11,0	11,4	0,41
32	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-85 %	32,0	149,3	-1,24
33	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-51 %	29,0	109,0	0,48
34	Ja, der Therapieeffektunterschied ist signifikant.	-56 %	3,8	13,4	-2,33*
35	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-60 %	10,0	14,7	1,84
36	Ja, der Therapieeffektunterschied ist signifikant.	69 %	1,0	0,7	1,38

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

37	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-93 %	86,0	76,2	-0,51	
38	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-78 %	69,0	38,9	-0,49	
39	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-81 %	32,0	30,0	-2,10*	
40	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-72 %	14,6	11,6	-1,29	
41	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-87 %	14,7	11,2	-0,45	
42	Ja, der Therapieeffektunterschied ist signifikant.	69 %	1,0	0,7	1,38	
43	Nein, im direkten Vergleich ist die eine Therapieoption signifikant überlegen und im indirekten Vergleich die andere.	-82 %	22,0	56,2	3,56*	
44	Ja, der Therapieeffektunterschied ist signifikant.	-70 %	22,0	34,5	0,62	
45	Ja, der Therapieeffektunterschied ist signifikant.	-39 %	15,0	11,9	-0,88	
46	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-73 %	9,0	10,5	1,66	
47	Nein, im direkten Vergleich ist die eine Therapieoption signifikant überlegen und im indirekten Vergleich die andere.	-48 %	3,3	Nicht berechenbar aufgrund fehlender Daten	8,62*	
48	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-67 %	12,5	49,2	0,07	
49	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	4 %	24,0	14,5	-0,39	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

50	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	13 %	12,0	8,0	0,67	
51	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	26 %	8,5	6,2	-0,44	
52	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	89 %	2,1	1,5	-0,89	
53	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-43 %	16,5	72,8	0,83	
54	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-21 %	16,3	31,4	-1,03	
55	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-77 %	26,3	25,2	-0,72	
56	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-5 %	4,1	15,3	1,03	
57	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	99 %	2,2	4,3	-0,44	
58	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	147 %	13,0	8,0	-0,50	
59	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-62 %	48,0	16,3	-0,39	Auf Nachfrage wurde berichtet, dass die Head-to-head-Studie eine Studie mit fünf Armen ist (Vardenafil 5 mg, 10 mg, 20 mg, Sildenafil 50 mg und Placebo). Die Arme Sildenafil 50 mg und Vardenafil 10 mg wurden nur für den direkten Vergleich benutzt sowie die Arme Vardenafil 20 mg und Placebo nur für den indirekten Vergleich. Da es keine Überschneidungen zwischen den für den direkten und indirekten Vergleich benutzten Patientenpopulationen gibt, kann die Studie berücksichtigt werden.
60	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-36 %	10,0	20,6	1,24	Eigene Berechnung des Therapieeffektunterschiedes im direkten Vergleich auf Basis der Head-to-head-Studie.

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

61	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-64 %	18,0	335,9	-1,09	Fragestellung des Reviews: Welche Co-Medikation oder Ersatzmedikation wirkt am stärksten protektiv gegen NSAR-induzierte gastrointestinale Komplikationen. Die Zielgrößen Tod, schwere kardiovaskuläre oder renale Komplikationen und Drop-out-Raten wurden nicht berücksichtigt. Sie waren in keinem aktiv- oder placebokontrolliertem Vergleich signifikant, was darauf schließen lässt, dass sie keine relevanten Zielgrößen sind. Drop-out-Raten können viele Ursachen haben, die nicht zwangsläufig auf die Wirksamkeit der Arzneistoffe schließen lassen.
62	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-70 %	35,0	284,8	-1,62	
63	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	2 %	6,0	200,1	-0,18	
64	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-76 %	30,0	152,9	-0,85	
65	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	119 %	15,0	78,6	0,81	
66	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	64 %	13,0	76,9	2,11*	
67	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-7 %	10,0	45,0	0,33	
68	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-31 %	14,5	37,7	1,76	
69	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	42 %	14,0	30,5	1,58	
70	Ja, der Therapieeffektunterschied ist signifikant.	-70 %	10,0	17,2	1,70	
71	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-14 %	7,0	13,0	0,57	
72	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	10 %	5,0	9,7	-0,46	
73	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-61 %	12,0	7,7	-0,09	
74	Ja, der Therapieeffektunterschied ist signifikant.	-51 %	18,0	7,3	1,99*	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

75	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	11 %	8,0	4,7	-0,31
76	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	205 %	3,5	3,9	0,62
77	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	70 %	12,0	2,2	-1,36
78	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	7 %	4,0	1,6	-1,56
79	Ja, der Therapieeffektunterschied ist signifikant.	-23 %	1,8	1,7	0,10
80	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	197 %	4,0	3,1	-0,72
81	Ja, der Therapieeffektunterschied ist signifikant.	287 %	1,0	2,0	2,36*
82	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	88 %	1,2	1,2	-0,81
83	Nein, im direkten Vergleich ist die eine Therapieoption signifikant überlegen und im indirekten Vergleich die andere.	537 %	1,2	1,2	2,74*
84	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	298 %	5,8	7,0	1,19
85	Ja, der Therapieeffektunterschied ist signifikant.	5 %	1,0	1,1	-0,69
86	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	92 %	4,2	2,5	1,75

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

87	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	240 %	9,0	8,8	-0,36	Eigene Berechnung des Therapieeffektunterschiedes im direkten Vergleich auf Basis der Head-to-head-Studie. Der direkte und indirekte Vergleich sind aufgrund unterschiedlicher Nachbeobachtungszeiten schlecht vergleichbar. Im indirekten Vergleich lag die mittlere Nachbeobachtungszeit bei ca. sieben Monaten (31,4 Wochen). Deshalb wurden aus den Nachbeobachtungszeiten der Head-to-Head-Studie die Werte für sechs Monate zur Gegenüberstellung ausgewählt.
88	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	252 %	9,0	8,8	0,60	
89	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-24 %	4,0	4,2	-0,24	
90	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-57 %	12,0	9,9	-0,77	
91	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	198 %	7,0	6,4	0,47	
92	Ja, der Therapieeffektunterschied ist signifikant.	-35 %	4,7	14,2	-1,08	
93	Ja, der Therapieeffektunterschied ist signifikant.	143 %	1,0	0,9	-1,06	
94	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	52 %	2,0	3,5	0,43	
95	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	415 %	1,2	1,2	-0,60	
96	Ja, der Therapieeffektunterschied ist signifikant.	-36 %	3,8	8,0	0,15	
97	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-23 %	6,3	7,2	0,90	
98	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	293 %	2,8	1,2	0,51	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

99	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	78 %	3,9	Nicht berechenbar aufgrund fehlender Daten	-0,29	
100	Ja, der Therapieeffektunterschied ist signifikant.	-31 %	9,0	4,6	-2,05*	
101	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-39 %	11,0	11,4	0,24	
102	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-42 %	32,0	149,3	-0,41	
103	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-2 %	29,0	109,0	1,02	
104	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	55 %	3,8	13,4	0,00	
105	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-5 %	10,0	14,7	1,84	
106	Ja, der Therapieeffektunterschied ist signifikant.	123 %	1,0	0,7	1,88	
107	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-90 %	86,0	76,2	-0,18	
108	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-65 %	69,0	38,9	0,13	
109	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-61 %	32,0	30,0	-1,65	
110	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-51 %	14,6	11,6	0,00	
111	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-73 %	14,7	11,2	0,05	
112	Ja, der Therapieeffektunterschied ist signifikant.	123 %	1,0	0,7	1,88	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

113	Ja, der Therapieeffektunterschied ist signifikant.	23 %	5,0	Nicht berechenbar aufgrund fehlender Daten	-0,65	
114	Ja, der Therapieeffektunterschied ist signifikant.	96 %	1,3	Nicht berechenbar aufgrund fehlender Daten	0,66	In direktem und indirektem Vergleich Fixed und Random effects-Modelle unterschiedlich gewählt.
115	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	50 %	4,0	Nicht berechenbar aufgrund fehlender Daten	-3,04*	In direktem und indirektem Vergleich Fixed und Random effects-Modelle unterschiedlich gewählt.
116	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	56 %	1,1	Nicht berechenbar aufgrund fehlender Daten	3,03*	In direktem und indirektem Vergleich Fixed und Random effects-Modelle unterschiedlich gewählt.
117	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	226 %	3,0	Nicht berechenbar aufgrund fehlender Daten	1,03	
118	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	79 %	1,7	Nicht berechenbar aufgrund fehlender Daten	-2,35*	In direktem und indirektem Vergleich Fixed und Random effects-Modelle unterschiedlich gewählt.
119	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-45 %	53,0	406,1	3,73*	
120	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-80 %	30,0	48,6	0,06	
121	Ja, der Therapieeffektunterschied ist signifikant.	-41 %	22,0	56,2	1,39	
122	Ja, der Therapieeffektunterschied ist signifikant.	-42 %	22,0	34,5	-0,17	
123	Ja, der Therapieeffektunterschied ist signifikant.	25 %	15,0	11,9	0,17	
124	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	27 %	9,0	10,5	-1,00	
125	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	70 %	3,3	Nicht berechenbar aufgrund fehlender Daten	2,37*	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

126	Ja, der Therapieeffektunterschied ist signifikant.	89 %	4,3	Nicht berechenbar aufgrund fehlender Daten	2,18*	
127	Ja, der Therapieeffektunterschied ist signifikant.	44 %	3,9	Nicht berechenbar aufgrund fehlender Daten	0,54	
128	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-11 %	5,3	Nicht berechenbar aufgrund fehlender Daten	0,59	
129	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	38 %	4,9	Nicht berechenbar aufgrund fehlender Daten	-0,14	
130	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	42 %	3,5	Nicht berechenbar aufgrund fehlender Daten	-1,75	
131	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-16 %	5,7	Nicht berechenbar aufgrund fehlender Daten	0,77	
132	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	21 %	5,6	Nicht berechenbar aufgrund fehlender Daten	-0,01	
133	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	30 %	5,3	Nicht berechenbar aufgrund fehlender Daten	1,66	
134	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	16 %	4,6	Nicht berechenbar aufgrund fehlender Daten	0,14	
135	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	16 %	4,1	Nicht berechenbar aufgrund fehlender Daten	-1,73	
136	Ja, der Therapieeffektunterschied ist signifikant.	-12 %	5,3	Nicht berechenbar aufgrund fehlender Daten	0,77	
137	Ja, der Therapieeffektunterschied ist signifikant.	-18 %	4,9	Nicht berechenbar aufgrund fehlender Daten	0,24	
138	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-33 %	6,3	Nicht berechenbar aufgrund fehlender Daten	0,37	
139	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-19 %	5,9	Nicht berechenbar aufgrund fehlender Daten	-0,06	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

140	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-16 %	4,5	Nicht berechenbar aufgrund fehlender Daten	-0,78	
141	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	33 %	6,1	2,4	-1,09	Zur Validitätsprüfung wird der Gesamteffektschätzer mit Konfidenzintervall für den indirekten Vergleich nach Bucher berechnet.
142	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	109 %	2,0	1,3	1,03	
143	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	211 %	Nicht berechenbar aufgrund fehlender Daten	Nicht berechenbar aufgrund fehlender Daten	0,79	Nur der harte Endpunkt Prävention von Glukocortoid-induzierten Knochenbrüchen wird in die Analyse aufgenommen und nicht die Surrogatparameter zur Knochendichte, da für diese keine Konfidenzintervalle berichtet werden.
144	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-53 %	25,0	28,1	0,49	Zur Validitätsprüfung wird der Gesamteffektschätzer mit Konfidenzintervall für den indirekten Vergleich nach Bucher berechnet.
145	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-74 %	8,0	21,7	1,08	Zur Validitätsprüfung wird der Gesamteffektschätzer mit Konfidenzintervall für den indirekten Vergleich nach Bucher berechnet. Aufgrund der Fülle an Therapievergleichen und Zielgrößen wurde nur die Zielgröße „Langfristige Schmerzfreiheit“ übernommen, da er der klinisch relevanteste Outcome bei der
146	Ja, der Therapieeffektunterschied ist signifikant.	-62 %	16,0	15,6	0,60	Wirksamkeitsmessung von Triptanen bei Migräne ist (eigene Aussage des Autors) und anstatt von Vergleichen verschiedener Triptane mit Sumatriptan der Dosen 100 mg, 50 mg, und 25 mg wird nur der Vergleich der verschiedenen Triptane mit Sumatriptan 100 mg übernommen. Durch diese zwei
147	Ja, der Therapieeffektunterschied ist signifikant.	-56 %	14,0	9,9	0,65	Maßnahmen wurde die Zahl der Metaanlaysen von 66 auf 10 verringert. Alle Angaben der Therapieeffekte sind in Prozent.
148	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-28 %	17,0	10,3	2,31*	
149	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-59 %	12,0	14,0	-1,22	
150	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-60 %	7,5	9,4	-2,16*	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

151	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-50 %	7,0	8,5	-0,32	Zur Validitätsprüfung wird der Gesamteffektschätzer mit Konfidenzintervall für den indirekten Vergleich nach Bucher berechnet.
152	Ja, der Therapieeffektunterschied ist signifikant.	-64 %	7,5	8,5	-1,72	
153	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-59 %	10,0	7,4	0,43	
154	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-44 %	14,0	4,1	0,28	
155	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-90 %	Nicht berechenbar aufgrund fehlender Daten	48,0	1,89	
156	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	5 %	Nicht berechenbar aufgrund fehlender Daten	9,8	-0,08	
157	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	30 %	Nicht berechenbar aufgrund fehlender Daten	2,3	-0,20	
158	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	60 %	23,5	25,4	0,88	
159	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	5 %	78,0	39,4	0,18	
160	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-4 %	48,0	26,7	-0,61	
161	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	57 %	22,0	24,1	0,59	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

162	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	12 %	26,7	24,0	0,58	
163	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-2 %	16,7	17,2	-1,24	
164	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	113 %	22,0	12,1	-2,17*	
165	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	80 %	6,0	9,8	-0,37	
166	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	4 %	4,0	6,2	0,38	
167	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-4 %	4,0	6,2	1,48	
168	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	3 %	4,0	6,2	0,33	
169	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-21 %	9,0	3,5	-1,22	Als Effektmaß werden neben dem Odds ratio auch das relative Risiko und die absolute Risikodifferenz berechnet. Da diese Effektmaße in einander umrechenbar sind und nicht zu anderen Resultaten führen können, werden nur die Ergebnisse für ein Effektmaß (Odds ratio) in die Analyse übernommen. Probanden aus den Placeboarmen der placebokontrollierten Head-to-head-Studien werden für den indirekten Vergleich doppelt gezählt. Dies ist in den Gesamprobandenzahlen berücksichtigt. Ohne Doppelzählung wären es 271 Probanden weniger im indirekten Vergleich.
170	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-10 %	4,0	2,4	2,17*	
171	Ja, der Therapieeffektunterschied ist signifikant.	921 %	22,0	365,2	1,33	
172	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	21 %	11,0	16,4	-0,90	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

173	Ja, der Therapieeffektunterschied ist signifikant.	-18 %	11,0	11,0	-0,73
174	Ja, der Therapieeffektunterschied ist signifikant.	-11 %	11,0	10,5	0,12
175	Ja, der Therapieeffektunterschied ist signifikant.	1 %	18,0	Nicht berechenbar aufgrund fehlender Daten	-0,42
176	Ja, der Therapieeffektunterschied ist signifikant.	30 %	6,0	Nicht berechenbar aufgrund fehlender Daten	0,88
177	Ja, der Therapieeffektunterschied ist signifikant.	67 %	4,0	Nicht berechenbar aufgrund fehlender Daten	-0,20
178	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-44 %	34,0	Nicht berechenbar aufgrund fehlender Daten	-0,07
179	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-8 %	17,0	Nicht berechenbar aufgrund fehlender Daten	0,19
180	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-49 %	11,3	Nicht berechenbar aufgrund fehlender Daten	-0,06
181	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-9 %	36,0	Nicht berechenbar aufgrund fehlender Daten	-0,10
182	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-49 %	12,0	Nicht berechenbar aufgrund fehlender Daten	-0,13
183	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	35 %	3,8	Nicht berechenbar aufgrund fehlender Daten	-0,76
184	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-4 %	5,7	Nicht berechenbar aufgrund fehlender Daten	1,17
185	Ja, der Therapieeffektunterschied ist signifikant.	-49 %	20,0	Nicht berechenbar aufgrund fehlender Daten	-0,49
186	Ja, der Therapieeffektunterschied ist signifikant.	-34 %	10,0	Nicht berechenbar aufgrund fehlender Daten	0,92
187	Ja, der Therapieeffektunterschied ist signifikant.	-9 %	5,5	Nicht berechenbar aufgrund fehlender Daten	-0,57

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

188	Ja, der Therapieeffektunterschied ist signifikant.	3 %	12,0	9,2	0,47	
189	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-31 %	4,0	7,9	-0,04	
190	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	25 %	12,0	6,9	-0,32	
191	Ja, der Therapieeffektunterschied ist signifikant.	-2 %	3,0	3,7	0,07	
192	Ja, der Therapieeffektunterschied ist signifikant.	37 %	3,0	2,4	0,00	
193	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-1 %	14,0	Nicht berechenbar aufgrund fehlender Daten	1,18	Die ALLHAT-Studie, die auch direkte Vergleiche vornimmt, wird vom Autor nicht in die Metaanalyse der Head-to-head-Studien eingeschlossen.
194	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-31 %	22,0	Nicht berechenbar aufgrund fehlender Daten	1,00	
195	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-21 %	24,0	Nicht berechenbar aufgrund fehlender Daten	0,84	
196	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	7 %	24,0	Nicht berechenbar aufgrund fehlender Daten	1,25	
197	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-33 %	25,0	Nicht berechenbar aufgrund fehlender Daten	0,76	
198	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	23 %	28,0	Nicht berechenbar aufgrund fehlender Daten	2,29*	
199	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-38 %	8,7	Nicht berechenbar aufgrund fehlender Daten	-0,40	
200	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-41 %	5,8	Nicht berechenbar aufgrund fehlender Daten	0,00	
201	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-31 %	6,3	Nicht berechenbar aufgrund fehlender Daten	0,39	
202	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	8 %	7,3	Nicht berechenbar aufgrund fehlender Daten	-0,12	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

203	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-64 %	3,8	Nicht berechenbar aufgrund fehlender Daten	-1,60	Unterschiedliche Effektmaße für direkten und indirekten Vergleich gewählt.
204	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-32 %	6,3	Nicht berechenbar aufgrund fehlender Daten	0,13	
205	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	11 %	4,8	3,7	0,08	
206	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-23 %	4,8	3,7	-0,79	
207	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-13 %	4,6	3,7	0,36	
208	Ja, der Therapieeffektunterschied ist signifikant.	6 %	4,6	3,6	1,94	
209	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-20 %	4,5	3,6	-0,71	
210	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-2 %	2,2	3,3	0,51	
211	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-14 %	2,2	3,3	0,69	
212	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-21 %	2,2	3,2	0,38	
213	Ja, der Therapieeffektunterschied ist signifikant.	-7 %	2,2	3,2	1,77	
214	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-27 %	2,1	3,2	0,80	
215	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-25 %	3,0	2,7	-0,39	
216	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-8 %	1,4	2,3	0,77	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

217	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	5 %	2,5	2,1	-0,51	
218	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-10 %	2,5	2,1	-0,14	
219	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-12 %	2,5	2,1	0,09	
220	Ja, der Therapieeffektunterschied ist signifikant.	41 %	2,5	2,1	0,65	
221	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-7 %	2,4	2,1	-0,59	
222	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	0 %	1,6	1,5	-0,09	
223	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-44 %	6,3	Nicht berechenbar aufgrund fehlender Daten	0,38	
224	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-16 %	4,9	Nicht berechenbar aufgrund fehlender Daten	0,07	
225	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-15 %	5,9	Nicht berechenbar aufgrund fehlender Daten	-0,03	
226	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	-11 %	5,3	Nicht berechenbar aufgrund fehlender Daten	0,95	
227	Nein, der Therapieeffektunterschied ist im indirekten Vergleich signifikant und im direkten Vergleich nicht signifikant.	-15 %	4,5	Nicht berechenbar aufgrund fehlender Daten	-0,96	
228	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	393 %	9,0	Nicht berechenbar aufgrund fehlender Daten	0,78	Unterschiedliche Effektmaße für direkten und indirekten Vergleich gewählt. Umrechnung des Ergebnisses des direkten Vergleichs auch auf die Ansicht Abiximab + PCI versus Tirofiban + PCI durch Kehrwertberechnung.
229	Ja, der Therapieeffektunterschied ist signifikant.	-45 %	4,3	4,9	-0,64	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

230	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-49 %	4,8	14,9	-0,05	
231	Nein, im direkten Vergleich ist die eine Therapieoption signifikant überlegen und im indirekten Vergleich die andere.	-12 %	4,3	8,2	4,54*	Umrechnung des Ergebnisses des indirekten Vergleichs auch auf die Ansicht Carbamazepin versus Phenobarbital durch Kehrwertberechnung.
232	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-27 %	3,5	6,8	0,09	
233	Ja, der Therapieeffektunterschied ist signifikant.	-10 %	4,8	7,7	-0,47	
234	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	58 %	4,3	7,2	0,93	Der direkte Vergleich enthält im Gegensatz zum indirekten sowohl Patienten mit fokalen als auch mit generalisierten Anfällen.
235	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	16 %	4,7	6,4	0,82	
236	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-17 %	6,3	7,2	0,88	
237	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-6 %	3,4	5,2	0,00	
238	Nein, der Therapieeffektunterschied ist im direkten Vergleich signifikant und im indirekten Vergleich nicht signifikant.	34 %	2,8	4,2	0,14	
239	Nein, im direkten Vergleich ist die eine Therapieoption signifikant überlegen und im indirekten Vergleich die andere.	43 %	3,8	4,8	3,18*	
240	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-19 %	4,3	4,1	-1,99*	
241	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-19 %	3,5	3,4	-0,25	
242	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-16 %	4,8	4,5	-2,17*	

Fortsetzung Tabelle 31: Weitere Parameter der Gegenüberstellung der Therapieeffektunterschiede zwischen direktem und indirektem Vergleich in den Metaanalysen der Typ-5-Publikationen

243	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-49 %	3,4	4,6	0,23
244	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-5 %	3,5	5,0	0,10
245	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-23 %	3,8	4,5	-0,99
246	Ja, der Therapieeffektunterschied ist signifikant.	-13 %	4,0	5,5	0,15
247	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-10 %	12,0	5,4	0,23
248	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-68 %	6,0	3,7	-1,23
249	Ja, es gibt keinen signifikanten Therapieeffektunterschied.	-39 %	12,0	3,1	-0,46

* Ein * zeigt einen signifikanten Wert für den z-Wert an, der die Zahl von -1,96 unter- oder die Zahl von 1,96 überschreitet.

ALLHAT = Antihypertensive and lipid-lowering treatment to prevent heart attack trial (Antihypertensive-und-lipidsenkende-Therapie-in-der-Prävention-von-Herzinfarkten-Studie). NSAR = Nicht-steroidale Antirheumatika. PCI = Perkutane koronare Interventionen.

9.9.5 Übersichtstabellen zu den Ergebnissen der Validitätsprüfung (Hauptanalyse, Typ-5-Publikationen)

Tabelle 32: Test auf signifikante Über- oder Unterschätzung des Ergebnisses des direkten Vergleichs durch den indirekten Vergleich (Hauptanalyse)

Methode	n für die Validitätsprüfung	Durchschnittliche Diskrepanz zwischen direktem und indirektem Vergleich \bar{z}	Streuung σ	Test auf Normalverteilung*
Nicht-adjustierter indirekter Vergleich	47	$\bar{z} = -0,08$ (95 %-KI: -0,73 bis 0,58)	$\sigma = 2,22$	$p = 0,200$
Adjustierter indirekter Vergleich	116	$\bar{z} = 0,18$ (95 %-KI: -0,05 bis 0,40)	$\sigma = 1,21$	$p = 0,200$
Metaregression	6	$\bar{z} = 0,46$ (95 %-KI: -0,82 bis 1,75)	$\sigma = 1,22$	$p = 0,200$
MTC	57	$\bar{z} = 0,20$ (95 %-KI: -0,004 bis 0,40)	$\sigma = 0,76$	$p = 0,059$

* Kolmogorov-Smirnov- Test auf statistisch signifikante Abweichung von der Normalverteilung; berechnet mit SPSS 16.0; Signifikanzniveau $p = 0,05$.
KI = Konfidenzintervall. MTC = Mixed treatment comparison.

Tabelle 33: Ausmaß der durchschnittlichen Diskrepanz zwischen direktem und indirektem Vergleich (Hauptanalyse)

Methode	n für die Validitätsprüfung	Durchschnittliche Diskrepanz zwischen direktem und indirektem Vergleich $ \bar{z} $	Streuung σ
Nicht-adjustierter indirekter Vergleich	47	$ \bar{z} = 1,63$ (95 %-KI: 1,20 bis 2,07)	$\sigma = 2,22$
Adjustierter indirekter Vergleich	116	$ \bar{z} = 0,95$ (95 %-KI: 0,80 bis 1,09)	$\sigma = 0,78$
Metaregression	6	$ \bar{z} = 0,99$ (95 %-KI: 0,20 bis 1,79)	$\sigma = 0,76$
Mixed treatment comparison	57	$ \bar{z} = 0,59$ (95 %-KI: 0,45 bis 0,73)	$\sigma = 0,52$

Tabelle 34: Anteil der statistisch signifikant diskrepanten Datensätze (Hauptanalyse)

Methode	n für die Validitätsprüfung	Statistisch signifikant diskrepante Datensätze	Prozentualer Anteil statistisch signifikanter Datensätze in der Stichprobe der Methode
Nicht-adjustierter indirekter Vergleich	47	12	25,5 % [95 %-KI: 13,1 % bis 38,0 %]
Adjustierter indirekter Vergleich	116	14	12,1 % [95 %-KI: 6,1 % bis 18,0 %]
Metaregression	6	1	16,7 % [95 %-KI: -13,2 % bis 46,5 %]
Mixed treatment comparison	57	1	1,8 % [95 %-KI: -1,7 % bis 5,2 %]
Sonstige	22	4	18,2 % [95 %-KI: 2,1 % bis 34,3 %]
Gesamt	248	32	12,9 % [95 %-KI: 8,7 % bis 17,1 %]

KI = Konfidenzintervall.

Die systematische Bewertung medizinischer Prozesse und Verfahren, *Health Technology Assessment* (HTA), ist mittlerweile integrierter Bestandteil der Gesundheitspolitik. HTA hat sich als wirksames Mittel zur Sicherung der Qualität und Wirtschaftlichkeit im deutschen Gesundheitswesen etabliert.

Seit Einrichtung der Deutschen Agentur für HTA des DIMDI (DAHTA@DIMDI) im Jahr 2000 gehören die Entwicklung und Bereitstellung von Informationssystemen, speziellen Datenbanken und HTA-Berichten zu den Aufgaben des DIMDI.

Im Rahmen der Forschungsförderung beauftragt das DIMDI qualifizierte Wissenschaftler mit der Erstellung von HTA-Berichten, die Aussagen machen zu Nutzen, Risiko, Kosten und Auswirkungen medizinischer Verfahren und Technologien mit Bezug zur gesundheitlichen Versorgung der Bevölkerung. Dabei fallen unter den Begriff Technologie sowohl Medikamente als auch Instrumente, Geräte, Prozeduren, Verfahren sowie Organisationsstrukturen. Vorrang haben dabei Themen, für die gesundheitspolitischer Entscheidungsbedarf besteht.