

Bewertung des Verzerrungsrisikos von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung

Version 1.0 vom 08.05.2017



Bewertung von systematischen Übersichtsarbeiten

Version 1.0

Das Manual „Bewertung von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung“ wurde im Rahmen des vom Bundesministerium für Gesundheit (BMG) geförderten Projekts „ZEIG“ von Cochrane Deutschland in Kooperation mit dem Institut für Medizinisches Wissensmanagement der AWMF (AWMF-IMWi) erstellt. Das Manual ist zunächst für 3 Jahre gültig. Kommentare sind ausdrücklich erwünscht und können gerichtet werden an: sec@cochrane.de.

Autoren

Schmucker C¹, Nothacker M², Möhler R¹, Kopp I², Meerpohl JJ¹

Kontakte

¹Cochrane Deutschland

Universitätsklinikum Freiburg

Medizinische Fakultät

Breisacher Str. 153

D-79110 Freiburg

www.cochrane.de

²Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement (AWMF-IMWi)

Karl von Frisch Str. 1

Philipps Universität

D-35043 Marburg

www.awmf.org/leitlinien/awmf-imwi.html

Wir bedanken uns bei Gerta Rücker und Harriet Sommer für die Kommentierung des Abschnitts zu den statistischen Verfahren vom Metaanalysen und bei Prof. Dr. Stephan Herpertz, Prof. Dr. Susanne Krege und PD Dr. Tobias Weberschock für Durchsicht und Kommentierung des Manuals.

Bitte wie folgt zitieren

Cochrane Deutschland, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement. „Bewertung von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung“. 1. Auflage 2017.

Verfügbar: **Cochrane Deutschland:** <http://www.cochrane.de/de/review-bewertung-manual>

AWMF: <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html>

Disclaimer

Die in diesem Manual verwendete männliche Form gilt für *Personen* bzw. *Patienten* beiderlei Geschlechts.

Das Werk ist in allen seinen Teilen urheberrechtlich geschützt. Die vorliegenden Texte dürfen für den persönlichen Gebrauch (gemäß § 53 UrhG) in einer EDV-Anlage gespeichert und (in inhaltlich unveränderter Form) ausgedruckt werden. Bitte beachten Sie, dass nur das unter <http://www.cochrane.de/de/review-bewertung-manual> und <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html> verfügbare Dokument gültig ist. Verweise ("links") aus anderen Dokumenten des World Wide Web auf das Manual unter den vorstehenden Adressen sind ohne weiteres zulässig und erwünscht, für eine entsprechende Mitteilung sind wir jedoch dankbar. Jede darüber hinausgehende, insbesondere kommerzielle, Verwertung bedarf der schriftlichen Zustimmung der angegebenen Urheber und/oder Inhabern von Verwertungsrechten.

Inhaltsverzeichnis

Glossar.....	4
Tabellen- und Abbildungsverzeichnis	6
1. Einleitung: Warum ist die Berücksichtigung von systematischen Übersichtsarbeiten für die Leitlinienerstellung wichtig?	7
1.1 Einordnung: systematische Übersichtsarbeiten	7
1.2 Weitere Formen aggregierter Evidenz	8
1.3 Systematische Übersichtsarbeiten auf Basis von Primärstudien	8
1.4 Systematische Übersichtsarbeiten als Basis für Leitlinien	9
1.5 Ziele und Struktur des Manuals.....	10
2. Anforderungen an systematische Übersichtsarbeiten.....	11
2.1 Berichtsqualität	11
2.2 Methodische Qualität	12
2.3 Verzerrungsrisiko (Biasrisiko).....	12
2.4 Übertragbarkeit der Ergebnisse.....	13
2.5 Anforderungen an Metaanalysen.....	13
3. Bewertungsinstrumente	14
3.1 Allgemeines.....	14
3.2 AMSTAR: Assessment of Multiple Systematic Reviews.....	15
3.2.1. AMSTAR Checkliste (leicht modifizierte deutsche Übersetzung unter Berücksichtigung gegenwärtiger Standards)	16
3.2.2. Dokumentation der AMSTAR Bewertung	17
3.3 ROBIS: Risk Of Bias In Systematic reviews.....	18
3.3.1 Übertragbarkeit bzw. Relevanz der systematischen Übersichtsarbeit (1. Phase).....	18
3.3.2 Identifikation von Verzerrungen - Biasrisiko (2. Phase).....	18
3.3.3 Gesamteinschätzung des Verzerrungsrisikos (3. Phase).....	19
3.3.4 ROBIS Instrument (Phase 2 und Phase 3)	19
3.3.5 Dokumentation und Berichten der ROBIS Bewertung (Phase 2 und Phase 3)	23
4. Quellen	24
5. Weiterführende Informationen und Praxishilfen	27
5.1 Erläuterungen zu den statistische Verfahren von Metaanalysen	27
5.2 Erläuterungen zum Disseminationsbias	30
5.2 Leitlinien zur Verbesserung der Berichterstattung	32
5.3 PRISMA Statement – deutsche Übersetzung.....	33

Glossar

AMSTAR	Assessment of Multiple SysTEmAtic Reviews
Bias	Systematischer Fehler; Verzerrung
CASP	Critical Appraisal Skills Programme
CAT	Critical appraisal tool
CERqual	Confidence in the Evidence from Reviews of Qualitative research
CONSORT	CONsolidated Standards Of Reporting Trials; Das CONSORT Statement gibt Autoren Empfehlungen für die Erstellung von Publikationen oder Berichten von randomisierten kontrollierten Studien in Form einer Checkliste.
DART	Documentation Appraisal Review Tool
Dissemination bias (Disseminationsbias)	Dissemination bias ist der Oberbegriff zu einer ganzen Reihe von Biastypen, die auf eine selektive Verbreitung von Studienergebnissen zurückgehen. Dazu gehören: Publication bias (Publikationsbias; Studien mit negativen Ergebnissen werden seltener oder gar nicht publiziert), language bias (Sprachbias; Studien mit negativen Ergebnissen werden eher in nationalsprachlichen Zeitschriften publiziert), time-lag bias (Studien mit negativen Ergebnissen werden zeitverzögert publiziert), citation bias (Zitationsbias; Studien mit negativen Ergebnissen werden seltener zitiert).
EbM	Evidenzbasierte Medizin
Effektmaß	Maßzahl, um die Stärke eines Effekts zu quantifizieren. Gebräuchliche Effektmaße für dichotome Endpunkte sind das relative Risiko (RR) oder die Odds Ratio (OR), gebräuchliche Effektmaße für kontinuierliche Endpunkte sind in Einzelstudien die standardisierte mittlere Differenz (SMD) und in Metaanalysen die gewichtete mittlere Differenz (weighted mean difference [WMD]).
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HEBW	Health Evidence Bulletins - Wales: Critical Appraisal of a systematic review
HTA	Health Technology Assessment

Bewertung von systematischen Übersichtsarbeiten
Version 1.0

Indirektheit	Diskrepanz zwischen interessierender Patientenpopulation oder interessierender Intervention oder interessierenden Endpunkten und der in der Forschungsarbeit untersuchten Population, Intervention oder Endpunkten
Inkonsistenz	Vorliegen widersprüchlicher Studienergebnisse
MECIR	Methodological Expectations of Cochrane Intervention Reviews
MOOSE	Metaanalysis of ObservatiOnal Studies in Epidemiology
OQAW	Overview Quality Assessment Questionnaire
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-P	Preferred Reporting Items for Systematic reviews and Meta-Analyses Protocols
QUOROM	QUality Of Reporting Of MEtaanalyses
RAMESES	Realist And MEta-narrative Evidence Syntheses: Evolving Standards
R-Amstar	Revised Assessment of Multiple SysTemAtic Reviews
RCT	Randomised Controlled Trial; randomisierte kontrollierte Studie
ROBIS	Risk Of Bias In Systematic reviews
SIGN	Scottish Intercollegiate Guidelines Network
Small study effects	Tendenz, dass kleine Studien einen höheren Nutzen als große Studien zeigen; dieser Effekt kann jedoch das Resultat eines hohen Biasrisikos in kleinen Studien oder von Disseminationsbias.
SURE	Specialist Unit for Review Evidence

Tabellen- und Abbildungsverzeichnis

Tabelle 1. Auswahl verschiedener Bewertungsinstrumente für systematische Übersichtsarbeiten	14
Tabelle 2. Tabellarische Darstellung der ROBIS Bewertung	23
Tabelle 3. PRISMA Statement (deutsche Übersetzung).	33
Abb. 1. Funnel Plot	31
Abb. 2. Screenshot der Equator Webseite.....	32

1. Einleitung: Warum ist die Berücksichtigung von systematischen Übersichtsarbeiten für die Leitlinienerstellung wichtig?

1.1 Einordnung: systematische Übersichtsarbeiten

Auf die Evidenz einzelner wissenschaftlicher Studien (Primärstudien) zu vertrauen, kann irreführend sein. Lediglich eine oder nur selektiv ausgewählte Studien isoliert von anderen, ähnlichen Studien zu derselben Fragestellung zu betrachten, kann dazu führen, dass Therapien mehr oder weniger nützlich erscheinen, als sie es tatsächlich sind. Das Ziel von hochwertigen systematischen Übersichtsarbeiten ist es, die Evidenz nicht selektiv, sondern systematisch zu identifizieren,¹ methodisch zu bewerten (Biasbewertung)² und deskriptiv oder metaanalytisch zusammenfassen. Eine Metaanalyse erhöht die Aussagekraft für einen Effekt gegenüber dem Ergebnis aus einer Einzelstudie.³ Nicht jede systematische Übersichtsarbeit beinhaltet jedoch eine Metaanalyse (insbesondere bei ausgeprägter Heterogenität ist eine Metaanalyse nicht sinnvoll).

Im Gegensatz zu systematischen Übersichtsarbeiten bieten narrative Übersichtsarbeiten meist einen breiten Überblick zu einem bestimmten Thema und haben häufig einen edukativen Charakter. Die Auswahl der berücksichtigten Literatur erfolgt subjektiv und unsystematisch, was die Gefahr einer erheblichen Verzerrung bei den Schlussfolgerungen birgt.

Auch Health Technology Assessments (HTA Berichte) basieren auf systematischen Übersichtsarbeiten. In HTA Berichten wird normalerweise versucht, eine klinische und/oder eine für Patienten relevante Frage im Hinblick auf Wirksamkeit, Sicherheit und Kosten unter Berücksichtigung von sozialen, rechtlichen und ethischen Aspekten zu beantworten.⁴ HTA Berichte sollen dabei primär als Entscheidungshilfe bei gesundheitspolitischen Fragestellungen eines spezifischen Gesundheitssystems dienen und sind deshalb selten auf die Versorgungssituation anderer Länder übertragbar.

Zunehmend werden auch systematische Übersichtsarbeiten erstellt, die sich anstelle von Evidenz aus Primärstudien auf aggregierte Evidenz aus bereits existierenden systematischen Übersichtsarbeiten beziehen. Die Nomenklatur dieser Arbeiten variiert, sie werden meist als Overviews, aber auch als Meta-Reviews, Overviews of Reviews oder Umbrella Reviews bezeichnet.^{5,6} Ziel solcher Overviews ist, eine ressourcensparende Evidenzbasis in der Gesundheitsversorgung zu schaffen.⁶ Häufig liegen jedoch zwischen der interessierenden Fragestellung und der Fragestellung der identifizierten aggregierten Evidenz Differenzen vor (scope mismatch). In solchen Fällen muß dann trotzdem auf die primären Publikationen aggregierter Evidenz zurückgegriffen werden.⁷ Anforderungen, die an

systematische Übersichtsarbeiten gestellt werden, um als Evidenzbasis herangezogen werden zu können, sind in *Kapitel 2 Anforderungen an systematische Übersichtsarbeiten* dargestellt.

1.2 Weitere Formen aggregierter Evidenz

Weitere neuere Publikationsformen aggregierter Evidenz sind zum Beispiel Evidence Maps oder Scoping Reviews.^{8 9} Evidence Maps oder Scoping Reviews bilden meist die Evidenz zu breiten medizinischen Fragestellungen in ihrer Quantität mit entsprechenden Studiencharakteristika systematisch in Tabellenform ab. Scoping Reviews liefern zudem eine deskriptive Zusammenfassung der Literatur. Eine kritische Bewertung der eingeschlossenen Studien erfolgt dabei in der Regel nicht. Durch die quantitative Darstellung der vorhandenen Forschung erlauben diese Formen der aggregierten Evidenz vor allem das Erkennen von Lücken im Wissenspool. Diese Publikationsformen können prinzipiell auch in einem Leitlinienentwicklungsprozess dem Aufzeigen von Wissenslücken und von künftigen Forschungsbedarf dienen, sie sind aber nicht als Grundlage für Leitlinienempfehlungen geeignet. Daher werden sie in diesem Manual nicht weiter behandelt.

Noch in der Entwicklung sind auch Realist Reviews. Bei dieser Publikationsform liegt der Fokus auf komplexen Interventionen. Neben quantitativer wird auch qualitative Forschung bei der Erstellung dieser Reviews berücksichtigt.^{10 11} Für diese Evidenzaufbereitung liegen noch keine Maßstäbe (Instrumente) vor, mit Hilfe derer sich die Qualität solcher Arbeiten zuverlässig bewerten lässt. Erste ‚Standards‘ für die Berichterstattung von Realist Reviews wurden jedoch schon veröffentlicht (RAMESES, Realist And MEta-narrative Evidence Syntheses: Evolving Standards).^{12 13}

1.3 Systematische Übersichtsarbeiten auf Basis von Primärstudien

Im Falle von systematischen Übersichten über die Wirkungen medizinischer Interventionen haben randomisierte kontrollierte Studien (RCTs) aufgrund ihres Designs das geringste Verzerrungsrisiko. RCTs werden in der klinischen Forschung als Goldstandard für die Evaluation von Wirksamkeit und Sicherheit einer (neuen) Therapie angesehen. Sie werden zur Untersuchung von patientenrelevanten Forschungsfragestellungen eingesetzt und stellen in der Arzneimittelentwicklung die Grundlage für Zulassungsentscheidungen der Behörden dar. Bei anderen Fragestellungen wie etwa zur Ätiologie, oder Prognose, oder bei der qualitativen Beschreibung von Patientenerfahrungen besteht die geeignete Evidenzbasis für eine systematische Übersichtsarbeit aus Primärstudien eines anderen Typus.¹⁴

1.4 Systematische Übersichtsarbeiten als Basis für Leitlinien

Gemäß dem Regelwerk der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF-Regelwerk) wird der folgende iterative, hierarchische Rechercheprozess für die Erstellung von Leitlinien empfohlen: *(i)* Berücksichtigung von Leitlinien, *(ii)* Berücksichtigung von aggregierter Evidenz (systematische Übersichtsarbeiten), *(iii)* Berücksichtigung von Primärliteratur.^{1 15} D.h., nach der Berücksichtigung von bereits vorhandenen Leitlinien, wird aus pragmatischen (ressourcensparenden) Gründen empfohlen, zunächst gezielt nach systematischen Übersichtsarbeiten zu suchen. Kann die Evidenz nicht durch aktuelle, hochwertige systematischen Übersichtsarbeiten abgebildet werden, erfolgt die Berücksichtigung von Primärstudien.

Werden Handlungsempfehlungen für Leitlinien formuliert, soll die Gesamtheit der gefundenen Evidenz für einen Endpunkt (Therapieziel/Outcome) bewertet werden. Dafür bietet sich die Bewertung nach GRADE (Grading of Recommendations Assessment, Development and Evaluation) an.¹⁶ Neben der Bewertung des Verzerrungsrisikos der Forschungsarbeit fließen in eine umfassende Bewertung im Hinblick auf das Vertrauen in den Effektschätzer auch Aspekte wie eine unzureichende Präzision der Ergebnisse (u.a. weites Konfidenzintervall), Inkonsistenz (Vorliegen widersprüchlicher Ergebnisse), Indirektheit (Diskrepanz zwischen interessierender Patientenpopulation oder interessierender Intervention oder interessierenden Endpunkten und der in der Forschungsarbeit untersuchten Population, Intervention oder Endpunkten) und Disseminationsbias ein.¹⁶ Werden systematische Übersichtsarbeiten für die Leitlinienerstellung herangezogen, ist deshalb eine Bewertung des Verzerrungsrisikos, sowie im vorliegenden Manual beschrieben, notwendig.

1.5 Ziele und Struktur des Manuals

Das Manual „Bewertung von systematischen Übersichtsarbeiten, Version 1.0“ richtet sich vor allem an Leitlinienersteller und Leitlinienberater mit dem Ziel, sie über eine qualifizierte Bewertung im Hinblick auf systematische Verzerrungen von systematischen Übersichtsarbeiten zu informieren und bei gegebener praktischer Anleitung (zum Beispiel im Rahmen eines Leitlinienseminars) zu einer eigenständigen Bewertung zu befähigen. Die Grundlagen dieses Manuals bilden dabei international anerkannte Standards für die Bewertung von systematischen Übersichtsarbeiten ab.

Das Manual ergänzt und vertieft das AWMF-Regelwerk zur Erstellung von Leitlinien (speziell das Kapitel „Recherche, Auswahl und methodische Bewertung von Literatur“).¹⁵

Das Manual besteht aus den Kernkapiteln: (2.) Anforderungen an systematische Übersichtsarbeiten und (3.) Bewertungsinstrumente, um das Verzerrungsrisiko einer systematischen Übersichtsarbeiten abzuschätzen.

Das Manual soll kontinuierlich fortgeschrieben werden. Kommentare sind daher ausdrücklich erwünscht und können an sec@cochrane.de gerichtet werden.

2. Anforderungen an systematische Übersichtsarbeiten

- 2.1 Berichtsqualität
- 2.2 Methodische Qualität
- 2.3 Verzerrungsrisiko (Biasrisiko)
- 2.4 Übertragbarkeit der Ergebnisse
- 2.5 Anforderungen an Metaanalysen

3. Bewertungsinstrumente

- 3.1 Allgemeines
- 3.2 AMSTAR:
Assessment of Multiple
SysTemAtic Reviews
- 3.3 ROBIS:
Risk Of Bias In Systematic reviews

2. Anforderungen an systematische Übersichtsarbeiten

2.1 Berichtsqualität

Mängel in der Berichtsqualität (Berichterstattung) einer systematischen Übersichtsarbeit dürfen nicht als gleichbedeutend mit einem hohen Verzerrungspotential angesehen werden. Die Berichtsqualität umfasst Aspekte wie Vollständigkeit, Detailliertheit, Objektivität und Nachvollziehbarkeit einer wissenschaftlichen Arbeit. Zur klaren Abgrenzung zwischen Mängeln in der Berichtsqualität und anderen (methodischen) Schwächen wurde unter anderem das PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Statement entwickelt.¹⁷ Dabei handelt es sich um eine Orientierungshilfe, um die Berichterstattung von systematischen Übersichtsarbeiten und Metaanalysen von Studien, die Interventionen im Gesundheitswesen bewerten und zu verbessern. Das PRISMA Statement ist eine Weiterentwicklung der 1999 veröffentlichten QUOROM (QUality Of Reporting Of MEtaanalyses)¹⁸ Empfehlung und liegt auch in deutscher Übersetzung vor (*Kapitel 5.3 Prisma Statement – deutsche Übersetzung*).¹⁹

Neben dem für systematische Übersichtsarbeiten und Metaanalysen allgemein gültigen PRISMA Statement liegen spezifische Empfehlungen zur Verbesserung der Berichterstattung von Metaanalysen basierend auf Beobachtungsstudien (MOOSE, Metaanalysis of ObservatiOnal Studies in Epidemiology)²⁰ oder von Protokollen zu systematischen Übersichtsarbeiten (PRISMA-P, Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) vor.²¹ Diese und andere Orientierungshilfen zur Verbesserung der Berichtsqualität von systematischen Übersichten und Metaanalysen sind auch auf den Internetseiten des EQUATOR Netzwerks abrufbar: <http://www.equator-network.org/> (*Kapitel 5.2 Leitlinien zur Verbesserung der Berichterstattung*).

Ist die Berichterstattung nicht vollständig beziehungsweise unklar, besteht immer die Option, mit den Autoren der Übersichtsarbeit Kontakt aufzunehmen. Oft lassen sich dadurch fehlende Aspekte, die für eine valide Bewertung der systematischen Übersichtsarbeit ausschlaggebend sind, klären.

2.2 Methodische Qualität

Grundsätzlich soll bei der Erstellung einer systematischen Übersichtsarbeit von der Planung über die Durchführung bis zur Auswertung und Publikation nach einem standardisierten Konzept vorgegangen werden (methodische Qualität). Dazu gehört insbesondere die Entwicklung eines Protokolls, in dem bereits *a priori* beschrieben wird, wie bei der Erstellung der systematischen Übersichtsarbeit vorgegangen werden soll. Anstelle eines publizierten Protokolls kann auch eine *a priori* Registrierung von systematischen Übersichtsarbeiten, z.B. in der Datenbank PROSPERO (<http://www.crd.york.ac.uk/PROSPERO>), kostenfrei erfolgen.

Methodische Standards zur Durchführung von systematischen Übersichtsarbeiten sind den MECIR (Methodological Expectations of Cochrane Intervention Reviews) Richtlinien²² oder dem Cochrane Handbuch²³ zu entnehmen. Die MECIR Richtlinien sind eng mit dem Cochrane Handbuch verknüpft und beinhalten neben Details zur standardisierten Planung und Durchführung auch Empfehlungen zur transparenten Berichterstattung von systematischen Übersichtsarbeiten zu therapeutischen Fragestellungen.

2.3 Verzerrungsrisiko (Biasrisiko)

Auch systematische Übersichtsarbeiten können einem Verzerrungsrisiko unterliegen. Verzerrungen können insbesondere bei der Studiauswahl, der Datenextraktion, der Biasbewertung der Primärstudien und der Datensynthese auftreten. Es liegt in den Händen der Autoren der systematischen Übersichtsarbeit, potentielle Mängel der Primärstudien transparent darzustellen, diese bei der Ergebnisinterpretation adäquat zu berücksichtigen und eine selektive Berichterstattung zu vermeiden.

Damit eine systematische Übersichtsarbeit Berücksichtigung in der Leitlinienentwicklung finden kann, sollte sie nur ein geringes Verzerrungsrisiko aufweisen und die Evidenzlage vollständig, transparent und nachvollziehbar abbilden.^{24 25} Das Verzerrungsrisiko einer systematischen Übersichtsarbeit lässt sich mit verschiedenen Instrumenten bewerten (*Kapitel 3 Bewertungsinstrumente*).

2.4 Übertragbarkeit der Ergebnisse

Neben einer adäquaten Vorgehensweise bei der Erstellung der systematischen Übersichtsarbeit, sollte auch die Relevanz beziehungsweise die Übertragbarkeit der Evidenz aus einer systematischen Übersichtsarbeit für die Leitlinienentwicklung berücksichtigt werden. Folgende Faktoren sollten bei der Untersuchung der Übertragbarkeit der Ergebnisse (externen Validität) aus einer systematischen Übersichtsarbeit eine besondere Berücksichtigung finden:

- Studienpopulation (darunter Geschlecht, Alter, Komorbidität und Erkrankungsrisiko zu Beginn der Studie)
- Angemessenheit der beschriebenen Interventionen und Vergleichsinterventionen
- Anwendbarkeit und die wahrscheinliche Akzeptanz der Intervention in der Form, in der sie in den Primärstudien untersucht wurde^{26 27}
- medizinischer Kontext (darunter die medizinischen Versorgungsbereiche und Leistungserbringer)
- Patientenrelevanz der Endpunkte

2.5 Anforderungen an Metaanalysen

Metaanalysen unterliegen bestimmten statistischen Anforderungen. Zum Beispiel soll die Wahl des Modells (Fixed Effect Modell oder Random Effects Modell) bereits bei der Erstellung des Protokolls, also *a priori* erfolgen. Auch ist es wichtig, die Ergebnisse der Metaanalyse mit entsprechenden Tests auf statistische Heterogenität zu prüfen. Mit diesen Heterogenitätstests kann festgestellt werden, ob die Unterschiede zwischen den Studien größer sind als zufallsbedingt zu erwarten wäre. Die Durchführung einer Metaanalyse aus heterogenen Studien kann problematisch sein. Bei vorliegender Heterogenität sind deshalb zusätzliche Untersuchungen, wie zum Beispiel Subgruppenanalysen oder Sensitivitätsanalysen, notwendig, um Erklärungen dafür zu finden. Eine zusammenfassende Darstellung der statistischen Anforderungen an Metaanalysen findet sich in *Kapitel 5.1 Erläuterungen zu den statistischen Verfahren von Metaanalysen*.

3. Bewertungsinstrumente

3.1 Allgemeines

Es stehen verschiedene Instrumente zur Bewertung von systematischen Übersichtsarbeiten zur Verfügung (auch „critical appraisal tools [CAT]“ genannt). Der überwiegende Teil der Instrumente beschreibt dabei in Form einer Checkliste, welche Aspekte man bei einer systematischen Übersichtsarbeit überprüfen sollte. Die zur Verfügung stehenden Checklisten sind jedoch häufig nicht validiert, sehr umfangreich, und bewerten neben dem eigentlichen Verzerrungsrisiko der systematischen Übersichtsarbeit auch die Berichtsqualität.

Tabelle 1 verweist auf eine Auswahl verschiedener Bewertungsinstrumente für systematische Übersichtsarbeiten, die auf Webseiten führender Institutionen der Evidenzbasierten Medizin oder in den Referenzlisten bereits bekannter Bewertungsinstrumente, wie zum Beispiel AMSTAR (Assessment of Multiple SysTEMatic Reviews),^{5 28-31} identifiziert wurden. Ausreichend validiert sind dabei nur das AMSTAR und das bereits 1991 veröffentlichte OQAO (Overview Quality Assessment Questionnaire)^{32 33} Instrument. Auf die AMSTAR Checkliste, welche eine weiterentwickelte Version des OQAO darstellt und auf das im Jahr 2016 pilotierte und veröffentlichte ROBIS (Risk of Bias in Systematic Reviews)³⁴ Instrument wird im Weiteren detaillierter eingegangen.

Tabelle 1. Auswahl verschiedener Bewertungsinstrumente für systematische Übersichtsarbeiten.

Bezeichnung	Veröffentlicht (Jahr)	Land	Link zum Instrument
Risk of Bias in Systematic Reviews (ROBIS) ³⁴	2016	UK	http://www.bristol.ac.uk/media-library/sites/social-community-medicine/robis/robis.pdf
Health Evidence Quality Assessment Tool	2016	Kanada	https://www.healthevidence.org/documents/our-appraisal-tools/QA_Tool&Dictionary_10Nov16.pdf
Documentation Appraisal Review Tool (DART)	2015	USA	http://www.wjgnet.com/2308-3840/coretip/v3/i3/142.htm
An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual) ³⁵	2015	-	-
Quality Assessment of Systematic Reviews and Meta-Analyses	2014	US	https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/sr_ma
Critical Appraisal of a Metaanalysis or Systematic Review	2014	Niederland	https://www.cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-SR-or-MA-july-2014.pdf
Specialist Unit for Review Evidence (SURE)	2013	UK	http://www.cardiff.ac.uk/_data/assets/pdf_file/0005/212756/SURE_SR-Checklist_2015-update.pdf
Critical Appraisal Skills Programme (CASP)	2013	UK	http://media.wix.com/ugd/dded87_a02ff2e3445f4952992d5a96ca562576.pdf

Quality Criteria Checklist: Review Article	2012	USA	http://www.adaevidencelibrary.com/files/Docs/2012_Jan_EA_Manual.pdf
Revised-AMSTAR (R-AMSTAR). ³⁶	2010	US	-
A Measurement Tool to Assess systematic Reviews (AMSTAR) ^{5 28-31}	2007	Kanada	https://amstar.ca/Amstar_Checklist.php
Overview Quality Assessment Questionnaire (OQAQ) ^{32 33}	1991	Kanada	-
Critical Appraisal Sheet	k.A.	UK	http://www.cebm.net/critical-appraisal/
Health Evidence Bulletins - Wales: Critical Appraisal of a systematic review (HEBW)	k.A.	UK	http://hebw.cf.ac.uk/projectmethod/appendix5.htm
Critical Appraisal Checklist for Systematic Reviews and Research Syntheses	k.A.	Australien	http://joannabriggs.org/assets/docs/critical-appraisal-tools/JBI_Critical_Appraisal-Checklist_for_Systematic_Reviews.pdf
Assessing the Credibility of the Systematic Review Process	k.A.	UK	http://guides.mcclibrary.duke.edu/ebm/appraise
Critical Appraisal Checklist for a systematic review	k.A.	UK	http://www.gla.ac.uk/media/media_64047_en.pdf
Systematic review (of therapy) worksheet	k.A.	Kanada	http://ktclearinghouse.ca/cebm/teaching/worksheets/sr
Methodology Checklist: Systematic Reviews and Meta-analyses - Scottish Intercollegiate Guidelines Network (SIGN)	k.A.	UK	http://www.sign.ac.uk/methodology/checklists.html
Reviews and Meta-Analyses Checklist	k.A.	UK	http://bestbets.org/ca/pdf/review.pdf

k.A.: keine Angabe.

3.2 AMSTAR: Assessment of Multiple Systematic Reviews

Ein validiertes und das bis dato am häufigsten verwendete Instrument zur Bewertung von systematischen Übersichtsarbeiten ist die im Jahr 2007 publizierte AMSTAR Checkliste.^{5 28-31} AMSTAR stellt eine methodisch weiterentwickelte Version des im Jahr 1991 von Oxman *et al.* vorgestellten OQAW (Overview Quality Assessment Questionnaire) Instruments^{32 33} und der Checkliste von Sacks *et al.*³⁷ dar. Das Instrument, das von einer kanadischen Arbeitsgruppe entwickelt wurde und auch als Online Tool zur Verfügung steht, umfasst insgesamt 11 Leitfragen. Für die einzelnen Leitfragen stehen dabei die Antwortmöglichkeiten JA, NEIN, UNKLAR und NICHT ANWENDBAR zur Verfügung. Ein *cut-off* für eine Gesamteinschätzung ist nicht vorgesehen.

Laut Webseite wird AMSTAR überarbeitet (NEW-AMSTAR; https://amstar.ca/Amstar_Checklist.php, Stand 03/2017). Außerdem wird an einem Instrument zur Bewertung von systematischen Übersichtsarbeiten auf Basis nicht randomisierter Studien gearbeitet (NSR-AMSTAR).

Im Jahr 2010 veröffentlichte zudem eine andere, US-amerikanische Arbeitsgruppe eine revidierte Version des AMSTAR Instruments (Revised-AMSTAR, R-AMSTAR).³⁶ AMSTAR und R-AMSTAR beinhaltet zwar dieselben Leitfragen, unterscheiden sich aber dadurch, dass R-AMSTAR Punktezahlen (Scores von 1 bis 4) für die einzelnen Leitfragen vergibt. Ziel von R-AMSTAR ist, eine

bessere Quantifizierung im Hinblick auf die Qualität der einzelnen systematischen Übersichtsarbeiten zu erhalten. Im Gegensatz zum AMSTAR Instrument, das eine hohe Interrater-Reliabilität aufweist, wurde die revidierte Form (R-AMSTAR) - auch im Hinblick auf die implizite Wichtung der Punktezahlen - noch nicht ausreichend validiert.³⁸

3.2.1. AMSTAR Checkliste

Deutsche (leicht modifizierte) Übersetzung der AMSTAR Checkliste unter Berücksichtigung gegenwärtiger Standards:

1. Wurde die Übersichtsarbeit *a priori* geplant/definiert?

JA NEIN UNKLAR NICHT ANWENDBAR

*Für ein JA ist es erforderlich, dass sich die Autoren der Übersichtsarbeit auf ein Protokoll oder auf vorabdefinierte/*a priori* veröffentlichte Forschungsziele beziehen.*

2. Wurde die Studienauswahl und Datenextraktion von 2 Personen unabhängig von einander ausgeführt?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA ist es erforderlich, dass 2 Personen die Studienauswahlen und Datenextraktion unabhängig voneinander ausführen. Im Anschluss müssen die Ergebnisse konsentiert werden.

3. Wurde eine umfassende und systematische Literatursuche durchgeführt?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA müssen mindestens 2 elektronische Datenbanken durchsucht und die elektronische Suche muss mindestens durch eine zusätzliche Suchmaßnahme, wie z.B. einer Suche nach grauer Literatur oder Kontaktaufnahme mit dem Hersteller, ergänzt worden sein. Weiterführende Informationen zur systematischen Literatursuche finden sich auch im Manual ‚Systematische Literaturrecherche‘.^{1}

4. Wurden unpublizierte Studiendaten und graue Literatur in der Übersichtsarbeit berücksichtigt?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA muss nach unpublizierten Daten oder grauer Literatur gesucht worden sein.

5. Wurden die Referenzen der ein- und ausgeschlossenen Studien in der Übersichtsarbeit angegeben?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA müssen die Referenzangaben vollständig gelistet sein und elektronische LINKS abrufbar.

6. Wurden die Studiencharakteristika (Charakteristika der Patienten, Intervention(en) und Endpunkte) der eingeschlossenen Studien in Tabellenform oder ausführlich in Textform angegeben?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA müssen für die Fragestellung der Übersichtsarbeit alle als wichtig erachteten Studiencharakteristika im Hinblick auf die Patienten, Intervention, Vergleichsbehandlung und Endpunkte entweder in Tabellenform oder deskriptiv in Textform, für jede der eingeschlossenen Studien, vorliegen

7. Wurde das Verzerrungsrisiko der eingeschlossenen Primärstudien nach etablierten Methoden bewertet?

JA NEIN UNKLAR NICHT ANWENDBAR

*Für ein JA sollte das Verzerrungsrisiko idealerweise nach einem Komponentensystem (z.B. mit dem

„Cochrane Risk of Bias“ Instrument) bewertet worden sein. Jedoch kann lt. AMSTAR Autoren auch eine Checkliste oder Skala dafür herangezogen werden. Bei Verwendung solcher Instrumente muss jedoch ersichtlich sein, für welche Studien ein ‚hohes‘ beziehungsweise ‚niedriges‘ Biasrisiko vorliegt. Weiterführende Informationen zur Biasbewertung von klinischen Studien finden sich auch im Manual ‚Bewertung des Biasrisikos in klinischen Studien‘.^{2*}

8. Wurde das Verzerrungsrisiko der eingeschlossenen Studien in der Ergebnisinterpretation der Übersichtsarbeit berücksichtigt?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA ist es erforderlich, dass das Biasrisiko endpunktabhängig bei der Ergebnisinterpretation berücksichtigt wurde. Falls das Verzerrungsrisiko der eingeschlossenen Primärstudien nicht nach etablierten Methoden bewertet wurde (Frage 7), kann hier grundsätzlich keine JA Einstufung erfolgen.

9. Wurden die Studienergebnisse statistisch adäquat ausgewertet?

JA NEIN UNKLAR NICHT ANWENDBAR

*An dieser Stelle müssen statistische Aspekte in Betracht gezogen werden. D.h., liegen gepoolte Ergebnisse vor, dann muss ein Heterogenitätstest eingesetzt worden sein. Heterogene Ergebnisse müssen zudem weiter untersucht werden. Ausführliche Details zu den statistischen Aspekten von Metaanalysen finden sich unter *Kapitel 5.1 Erläuterungen zu den statistischen Verfahren von Metaanalysen, Absatz: Wie teste ich meine Ergebnisse auf statistische Heterogenität? und Absatz: Was mache ich bei heterogenen Ergebnissen?**

10. Wurde potentieller Publikationsbias adressiert?

JA NEIN UNKLAR NICHT ANWENDBAR

*Für ein JA müssen graphische oder statische Methoden zur Untersuchung von Publikationsbias eingesetzt worden sein. Um Publikationsbias (Disseminationsbias) jedoch annäherungsweise (adäquat) zu bewerten, müssen mindestens 10 Primärstudien vorliegen, denn dann erst machen graphische oder statistische Untersuchungen überhaupt Sinn. Weitere Informationen zur Untersuchung von Disseminationsbias finden sich unter *Kapitel 5.2 Erläuterungen zum Disseminationsbias.**

11. Wurden potentielle Interessenkonflikte adressiert?

JA NEIN UNKLAR NICHT ANWENDBAR

Für ein JA müssen potentielle Interessenkonflikte und Angaben zur Finanzierung sowohl für die systematische Übersichtsarbeit selbst als auch für jede der eingeschlossenen Primärstudien nachvollziehbar berichtet worden sein.

UNKLAR: wenn aufgrund mangelnder Berichtsqualität keine Aussage möglich ist

NICHT ANWENDBAR: wenn für die systematische Übersichtsarbeit nicht zutreffend.

Gesamteinschätzung:

Im welchen Rahmen eine Gesamteinschätzung der systematischen Übersichtsarbeit erfolgen soll, wird von den Autoren des AMSTAR Instruments nicht adressiert.

3.2.2. Dokumentation der AMSTAR Bewertung

Die AMSTAR Bewertungen können auf der Webseite der Entwickler online vorgenommen und als PDF Dokument abgespeichert werden (https://amstar.ca/Amstar_Checklist.php).

3.3 ROBIS: Risk Of Bias In Systematic reviews

Ein weiteres, neueres, jedoch noch nicht (ausreichend) validiertes Instrument ist ROBIS.³⁴ Im Gegensatz zu bereits existierenden Instrumenten adressiert ROBIS ausschließlich das Verzerrungsrisiko von systematischen Übersichtsarbeiten (Leitfrage 5 und 6 des AMSTAR Instruments adressieren beispielsweise in erster Linie die Berichtsqualität). ROBIS basiert genauso wie das Cochrane Instrument zur Bewertung von RCTs³ ²³ oder non-RCTs³⁹ auf einem Komponentensystem. Das bedeutet, dass das Verzerrungsrisiko für verschiedene Komponenten (Domänen) separat eingeschätzt und dargestellt wird. Die Bewertung von systematischen Übersichtsarbeiten mit dem ROBIS Instrument erfolgt in drei Phasen: (i) Einschätzung der Relevanz (optional); (ii) Identifikation potentieller Verzerrungen, die bei der Erstellung der systematischen Übersichtsarbeit aufgetreten sind und (iii) Gesamteinschätzung des Verzerrungsrisikos. Eine übersetzte Version des ROBIS Instruments findet sich in *Kapitel 3.3.4 ROBIS Instrument*.

3.3.1 Übertragbarkeit bzw. Relevanz der systematischen Übersichtsarbeit (1. Phase)

Die erste (optionale) Phase adressiert die Relevanz der Arbeit, d.h., ob die Anwendbarkeit der Ergebnisse der systematischen Übersichtsarbeit auf die jeweilige Fragestellung (Zielfrage) gegeben ist. Werden systematische Übersichtsarbeiten zur Entwicklung einer Leitlinie herangezogen, sollte die Fragestellung der jeweiligen identifizierten Übersichtsarbeit mit der Zielfragestellung in der Leitlinie gut übereinstimmen. Der Abgleich soll dabei anhand des PICO Schemas, welches eine Definition der Patientenpopulation, Intervention, des Comparator (Vergleichsbehandlung) und des Outcomes (Endpunkt/Zielgröße) erfordert, erfolgen.

3.3.2 Identifikation von Verzerrungen - Biasrisiko (2. Phase)

In der zweiten Phase wird das Verzerrungsrisiko der systematischen Übersichtsarbeit eingeschätzt. Diese Phase setzt sich aus den folgenden vier Domänen zusammen, die unterschiedliche Aspekte der Durchführung einer systematischen Übersichtsarbeit bewerten:

- Domäne 1: Einschlusskriterien der Studien
- Domäne 2: Identifikation und Auswahl der Studien
- Domäne 3: Datenextraktion und Bewertung des Verzerrungsrisikos der Studien
- Domäne 4: Datensynthese und Ergebnisse

Jede Domäne beinhaltet unterschiedliche Leitfragen („signalling questions“, siehe *Kapitel 3.3.4 ROBIS Instrument*). Es stehen dabei die Antwortmöglichkeiten JA (geringes Biasrisiko), WAHRSCHEINLICH JA, WAHRSCHEINLICH NEIN, NEIN (hohes Biasrisiko) und KEINE INFORMATION zur Verfügung. Nach Beantwortung der Leitfragen wird für jede Domäne eine Einschätzung getroffen. Das Verzerrungsrisiko für jede Domäne wird dann zusammenfassend unter Berücksichtigung der Einschätzung für die verschiedenen Leitfragen als *gering, hoch* oder *unklar* eingestuft. Transparent werden die Bewertungen dadurch, dass die Entscheidung für jede Domäne begründet werden soll.

3.3.3 Gesamteinschätzung des Verzerrungsrisikos (3. Phase)

In der dritten Phase erfolgt auf Basis von drei Leitfragen die Gesamteinschätzung des Verzerrungsrisikos der systematischen Übersichtsarbeit: Die *erste Leitfrage* adressiert dabei die potentiellen Verzerrungsrisiken aus Phase 2. Die *zweite Leitfrage* adressiert die klinische Relevanz der eingeschlossenen Studien im Hinblick auf die Fragestellung der systematischen Übersichtsarbeit und innerhalb der *dritten Leitfrage* wird bewertet, ob die Autoren der systematischen Übersichtsarbeit alle Ergebnisse transparent berichten (d.h., ob selektive Berichterstattung ausgeschlossen werden kann).

3.3.4 ROBIS Instrument (Phase 2 und Phase 3)

Phase 2 (Domänen und Leitfragen)

Domäne 1: Einschlusskriterien der Studien

- Erfolgte die Erstellung der systematischen Übersicht nach vordefinierten Zielen und Einschlusskriterien?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH Nicht NEIN KEINE INFORMATION

- Passten die Einschlusskriterien zur Fragestellung der systematischen Übersichtsarbeit (im Hinblick auf die Patientenpopulation [z.B. Alter, Indikation, Diagnose], Setting der Studie, Intervention/Exposition [Dosierung, Häufigkeit])?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

- Wurden die Einschlusskriterien eindeutig definiert und im Detail dargestellt / erlaubt die Beschreibung der Einschlusskriterien eine Replizierung der systematischen Übersichtsarbeit?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Bewertung von systematischen Übersichtsarbeiten Version 1.0

- Wurden bestehende Einschränkungen bei den Einschlusskriterien (z.B. im Hinblick auf das Studiendesign, den Publikationszeitraum, der Fallzahl oder der Endpunktauswahl) angemessen ausgewählt?
 - Falls Einschränkungen vorlagen, diese detailliert beschrieben und als sinnvoll erachtet wurden
=> JA oder WAHRSCHEINLICH JA
 - Falls keine Einschränkungen beschrieben wurden, die Einschlusskriterien aber ansonsten detailliert berichtet wurden, kann angenommen werden, dass keine Einschränkungen vorlagen
=> JA
 - Falls Einschränkungen vorlagen, diese aber ungenügend beschrieben wurden
=> WAHRSCHEINLICH NEIN oder NEIN

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

- Wurden bestehende Einschränkungen bei den Einschlusskriterien im Hinblick auf die Informationsquelle (z.B. Publikationsstatus oder die Sprache) angemessen ausgewählt?
 - Falls Einschränkungen vorlagen, diese detailliert beschrieben und als sinnvoll erachtet wurden
=> JA oder WAHRSCHEINLICH JA
 - Falls keine Einschränkungen beschrieben wurden, die Einschlusskriterien aber ansonsten detailliert berichtet wurden, kann angenommen werden, dass keine Einschränkungen vorlagen
=> JA
 - Falls Einschränkungen vorlagen, diese aber ungenügend beschrieben oder nicht angemessen waren => WAHRSCHEINLICH NEIN oder NEIN

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Bedenken im Hinblick auf die Einschlusskriterien (Verzerrungsrisiko):

- Gering* Hoch** Unklar***

Domäne 2: Identifikation und Auswahl der Studien - weiterführende Informationen zur systematischen Literatursuche finden sich im Manual „Systematische Literaturrecherche“¹

- Wurden eine angemessene Anzahl an Datenbanken bzw. elektronischen Quellen durchsucht (z.B. Medline, Embase, verschiedene Datenbanken der Cochrane Library)?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden zusätzlich zur elektronischen Suche weitere Quellen durchsucht (z.B. Suche nach unpublizierten Daten, graue Literatur, Suche in den Referenzlisten bereits vorhandener Reviews)?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurde eine systematische Literatursuche mit adäquaten Suchbegriffen durchgeführt (z.B. Textwortsuche und Suche mit Schlagwörtern (Medical Subject Headings, bzw. MeSH Terms)?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden adäquate Einschränkungen im Hinblick auf Suchzeitraum oder Sprache angewandt?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden Maßnahmen ergriffen, um Fehler bei der Studienausswahl zu minimieren (2 unabhg. Reviewer)?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Bedenken im Hinblick auf die Identifikation und Auswahl der Studien (Verzerrungsrisiko):

- Gering* Hoch** Unklar***

Domäne 3: Datenextraktion und Bewertung der Studien

- Wurden Maßnahmen ergriffen, um Fehler bei Datenerhebung zu minimieren (2 unabhg. Reviewer)?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden Studiendaten detailliert genug beschrieben, um Reviewautoren und Lesern eine Ergebnisinterpretation zu erlauben?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden alle relevanten Daten, die für die Synthese wichtig sind, extrahiert?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurde das Verzerrungsrisiko der Primärstudien nach standardisierten Methoden bewertet (z.B. mit dem ‚Cochrane Risk of Bias‘ Instrument)?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden Maßnahmen ergriffen, um Fehler bei der Bewertung des Verzerrungsrisikos zu minimieren (2 unabhg. Reviewer)?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Bedenken im Hinblick auf die Datenextraktion und Bewertung der Studien (Verzerrungsrisiko):

- Gering* Hoch** Unklar***

Domäne 4: Datensynthese und Ergebnisse - Weiterführende Informationen zu den statistischen Verfahren von Metaanalysen und Disseminationsbias finden sich unter Kapitel 5.1 und 5.2.

- Wurden in die Datensynthese die Ergebnisse aus allen relevanten Studien aufgenommen?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurden alle *a priori* festgelegten Analysen durchgeführt oder Abweichungen davon erläutert?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Wurde die Datensynthese im Hinblick auf die Fragestellung, potentiellen Divergenzen zwischen den Studientypen und Endpunkten adäquat durchgeführt?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Lagen homogene Studienergebnisse vor? Falls nicht, wurden sie in der Analyse adäquat berücksichtigt?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Lagen robuste Ergebnisse vor? Wurden dementsprechende statistische Tests durchgeführt, z.B. eine Funnel Plot oder Sensitivitätsanalyse?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION
- Lag ein niedriges Verzerrungsrisiko der Primärstudien vor? Falls nicht, wurden adäquate Analysen durchgeführt, um das Verzerrungsrisiko zu berücksichtigen?
 JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Bedenken im Hinblick auf die Datensynthese und Ergebnisse (Verzerrungsrisiko):

- Gering* Hoch** Unklar***

*Gering: Falls die Leitfragen der entsprechenden Domäne mit JA / WAHRSCHEINLICH JA beantwortet wurden.
**Hoch: Falls eine der Leitfragen mit NEIN / WAHRSCHEINLICH NEIN beantwortet wurde.
***Unklar: Falls nicht genügend Informationen vorliegen, um eine Einschätzung abzugeben.

Phase 3 (Leitfragen und Gesamtschätzung)

i) Leitfragen

Leitfrage I:* Wurden die potentiellen Verzerrungsrisiken aus Phase 2 in der Ergebnisinterpretation der entsprechenden systematischen Übersichtsarbeit adressiert?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Leitfrage II:** Adressieren die in die systematische Übersichtsarbeit eingeschlossenen Studien die Fragestellung, d.h., ist die klinische Relevanz der eingeschlossenen Studien im Hinblick auf die Fragestellung der systematischen Übersichtsarbeit gegeben?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

Leitfrage III:*** Kann selektive Berichterstattung innerhalb der systematischen Übersichtsarbeit ausgeschlossen werden?

JA WAHRSCHEINLICH JA WAHRSCHEINLICH NEIN NEIN KEINE INFORMATION

* Leitfrage I

JA/WAHRSCHEINLICH JA: Wenn in keiner Domäne aus Phase 2 ein Verzerrungsrisiko vorliegt *oder* wenn für eine oder für mehrere Domänen zwar ein Verzerrungsrisiko vorliegt, das Verzerrungsrisiko dabei aber adäquat in der Ergebnisinterpretation der systematischen Übersichtsarbeit berücksichtigt wurde.

WAHRSCHEINLICH NEIN/NEIN: Wenn für eine oder für mehrere Domänen aus Phase 2 ein Verzerrungsrisiko vorliegt, dass Verzerrungsrisiko jedoch nicht adäquat in der Ergebnisinterpretation der systematischen Übersichtsarbeit berücksichtigt wurde.

** Leitfrage II

JA/WAHRSCHEINLICH JA: Wenn die klinische Relevanz gegeben ist *oder* wenn die klinische Relevanz zwar nicht gegeben, dieses ‚Problem‘ jedoch adäquat in der Ergebnisinterpretation der systematischen Übersichtsarbeit berücksichtigt wurde.

WAHRSCHEINLICH NEIN/NEIN: Wenn die klinische Relevanz nicht gegeben ist *und* dieses ‚Problem‘ auch nicht in der Ergebnisinterpretation berücksichtigt wurde.

*** Leitfrage III

JA/WAHRSCHEINLICH JA: Wenn eine selektive Berichterstattung ausgeschlossen werden kann.

WAHRSCHEINLICH NEIN/NEIN: Wenn von den Autoren der systematischen Übersichtsarbeit in erster Linie Endpunkte mit positivem Ergebnis berichtet wurden.

ii) Aus den drei Leitfragen resultiert die Gesamtschätzung des Verzerrungsrisikos[§]

Gering Hoch Unklar

[§] Gesamtschätzung des Verzerrungsrisikos

Niedrig: Wenn alle drei Leitfragen mit JA beantwortet wurden

Hoch: Wenn eine der drei Leitfragen mit NEIN beantwortet wurde

Unklar: Wenn unzureichende Infos für eine abschließende Bewertung vorliegen.

3.3.5 Dokumentation und Berichten der ROBIS Bewertung (Phase 2 und Phase 3)

Die Entwickler des ROBIS Instruments empfehlen, die Bewertung der einzelnen Domänen für jede systematische Übersichtsarbeit tabellarisch darzustellen (Tabelle 2). Weiterführende Angaben zur tabellarischen und einer möglichen graphischen Darstellung des Verzerrungspotenzials sind unter folgendem Link zu finden: <http://www.bristol.ac.uk/social-community-medicine/projects/robis/resources/>. Neben der tabellarischen Darstellung sollte die Leitlinie die (möglicherweise) bestehende Ergebnisverzerrung auch narrativ beschreiben und diskutieren.

Tabelle 2. Tabellarische Darstellung der ROBIS Bewertung (adaptiert aus folgender Quelle: <http://www.bristol.ac.uk/social-community-medicine/projects/robis/resources/>).

Review	Phase 2				Phase 3
	1. STUDY ELIGIBILITY CRITERIA	2. IDENTIFICATION AND SELECTION OF STUDIES	3. DATA COLLECTION AND STUDY APPRAISAL	4. SYNTHESIS AND FINDINGS	RISK OF BIAS IN THE REVIEW
1	😊	😊	😊	😊	😞*
2	😊	😊	😊	😊	😊
3	😞	😞	😊	😊	😞
4	😞	😞	😊	😊	😊
5	😞	?	😊	😊	😞
6	😞	?	😊	😊	😊
7	😞	😊	😊	😊	😞
8	😞	😊	😊	😊	😞
9	😞	😊	😊	😊	😞
10	😞	😊	😊	😞	😞
11	😊	😊	😊	😊	😊

😊 = low risk; 😞 = high risk; ? = unclear risk

*Die Gesamteinschätzung des Verzerrungsrisikos (Phase 3) basiert auf drei Leitfragen (*Kapitel 3.3.3 Gesamteinschätzung des Verzerrungsrisikos [3. Phase]*). D.h., auch wenn für alle Domänen aus Phase 2 ein niedriges Verzerrungsrisiko vorliegt, kann ein hohes Gesamtverzerrungsrisiko aufgrund von fehlender klinischer Relevanz oder selektiver Berichterstattung vorliegen.

4. Quellen

1. Deutsches Cochrane Zentrum, AdWMF, Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin. Manual Systematische Literaturrecherche für die Erstellung von Leitlinien. Secondary Manual Systematische Literaturrecherche für die Erstellung von Leitlinien 1. Auflage 2013. Verfügbar: <http://www.cochrane.de/de/manual-systematische-literaturrecherche>, <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html>, <http://www.aezq.de/aezq/publikationen/kooperation>, DOI: 10.6094/UNIFR/2013/2, <http://www.freidok.uni-freiburg.de/volltexte/9020/>, Zugriff: 07.03.2017.
2. Deutsches Cochrane Zentrum, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement. Manual: Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien. 1. Auflage 2016. Verfügbar: <http://www.cochrane.de/de/rob-manual>, Zugriff: 07.03.2017.
3. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
4. Draborg E, Gyrd-Hansen D, Poulsen PB, et al. International comparison of the definition and the practical application of health technology assessment. *Int J Technol Assess Health Care* 2005;21(1):89-95.
5. Hartling L, Chisholm A, Thomson D, et al. A descriptive analysis of overviews of reviews published between 2000 and 2011. *PloS One* 2012;7(11):e49667.
6. Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. *Res Synth Methods* 2017.
7. Caird J, Sutcliffe K, Kwan I, et al. Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach? *Evidence & Policy* 2015;11(1):81-97.
8. Schmucker C, Motschall E, Antes G, et al. Methods of evidence mapping. A systematic review. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 2013;56(10):1390-7.
9. Miake-Lye IM, Hempel S, Shanman R, et al. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic reviews* 2016;5:28.
10. Pawson R, Greenhalgh T, Harvey G, et al. Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10 Suppl 1:21-34.
11. Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research. *Implementation science: IS* 2012;7:33.
12. Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses. *BMC Medicine* 2013;11:21.
13. Wong G, Westhorp G, Manzano A, et al. RAMESES II reporting standards for realist evaluations. *BMC Medicine* 2016;14(1):96.
14. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;328(7430):39-41.
15. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)-Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien“ 1. Auflage 2012. Secondary Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)-Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien“ 1. Auflage 2012. <http://www.awmf.org/leitlinien/awmf-regelwerk.html>, Zugriff: 07.03.2017.
16. Langer G, Meerpohl JJ, Perleth M, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. *Z Evid Fortbild Qual Gesundhwes* 2012;106(5):357-68.
17. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews

- and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* 2009;6(7):e1000100.
18. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354(9193):1896-900.
 19. Ziegler A, Antes G, IR K. Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *Dtsch Med Wochenschr* 2011;136 e9–e15.
 20. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group*. *JAMA* 2000;283(15):2008-12.
 21. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
 22. Higgins J, Lasserson T, Chandler J, et al. *Methodological Expectations of Cochrane Intervention Reviews (MECIR): methodological standard for the conduct of new cochrane intervention reviews*. Cochrane: London, 2016.
 23. Higgins J, Green S. *Cochrane Handbook of Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. *Secondary Cochrane Handbook of Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]* 2011. <http://handbook.cochrane.org/>, Zugriff: 07.03.2017.
 24. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490.
 25. Whitlock EP, Lin JS, Chou R, et al. Using existing systematic reviews in complex systematic reviews. *Ann Intern Med* 2008;148(10):776-82.
 26. Dans AL, Dans LF, Guyatt GH, et al. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *Evidence-Based Medicine Working Group*. *JAMA* 1998;279(7):545-9.
 27. Bonell C, Oakley A, Hargreaves J, et al. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006;333(7563):346-9.
 28. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PloS One* 2007;2(12):e1350.
 29. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
 30. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62(10):1013-20.
 31. Pieper D, Buechter R, Jerinic P, et al. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol* 2012;65(12):1267-73.
 32. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991;44(1):91-8.
 33. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44(11):1271-8.
 34. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.
 35. Lewin S, Glenton C, Munthe-Kaas H, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Medicine* 2015;12(10):e1001895.
 36. Kung J, Chiappelli F, Cajulis OO, et al. *From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic*

- Reviews (R-AMSTAR) for Grading of Clinical Relevance. Open Dent J 2010;4:84-91.
37. Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. N Engl J Med 1987;316(8):450-5.
 38. Pieper D, Buechter RB, Li L, et al. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol 2015;68(5):574-83.
 39. Sterne JA, Higgings JPT, Reeves BC, et al. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014
Secondary A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0., 24 September 2014
<http://www.riskofbias.info>, Zugriff: 07.03.2017.
 40. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-58.

5. Weiterführende Informationen und Praxishilfen

5.1 Erläuterungen zu den statistische Verfahren von Metaanalysen

Was ist ein Forestplot?

- Häufigste Abbildungsart zur zusammenfassenden Darstellung der einzelnen Studienergebnisse einer Metaanalyse
- Für jede Studie wird der geschätzte Therapieeffekt mit zugehörigem Konfidenzintervall dargestellt
- Gibt für jede Studie Auskunft über die (1) Stärke des Therapieeffekts = Lage des Effektschätzers, (2) Unsicherheit = Breite des Konfidenzintervalls, (3) Informationsgehalt = Gewicht der Studie/Dicke des Markerpunktes

Was ist ein gemeinsamer Schätzer?

Ein gewichteter Mittelwert der geschätzten Therapieeffekte über alle Einzelstudien hinweg.

Gibt es bei Metaanalysen unterschiedliche Methoden?

Es gibt zwei bekannte Methoden der Metaanalyse, das Fixed Effect Modell und das Random Effects Modell. Sie unterscheiden sich in der Art der Gewichtung der Einzelstudien.

Was ist der Unterschied zwischen Fixed Effect Modell und Random Effects Modell?

Annahmen für das Fixed Effect Modell:

- allen Studien liegt, bis auf Zufallsschwankungen, der gleiche Therapieeffekt zugrunde
- Unterschiede in den Therapieeffekten sind ausschließlich auf Zufallsstreuung zurückzuführen
- es gibt dafür mehrere Gewichtungsmethoden

Annahmen für das Random Effects Modell:

- Therapieeffekt zwischen den Einzelstudien variiert natürlicherweise; Variation nach fest vorgegebener Verteilung (meistens Normalverteilung)

Wieso nimmt man nicht immer das Random Effects Modell?

Im Random Effects Modell erhalten kleinere Studien ein größeres Gewicht als im Fixed Effekt Modell, somit kommt es zu einer Vereinheitlichung der Gewichtung, was das Random Effects Modell anfälliger für „small study effects“ macht.

Inwiefern unterscheiden sich die Ergebnisse zwischen Fixed Effect und Random Effects Modell?

Wenn wenig oder keine statistische Heterogenität vorliegt, unterscheiden sich die Ergebnisse zwischen den beiden Modellen wenig oder gar nicht. Das Fixed Effect Modell kann somit als Spezialfall des Random Effects Modell angesehen werden für den Fall, dass keine Heterogenität vorliegt. Wenn (klinische) Heterogenität vermutet wird, sollte das Random Effects Modell gewählt werden, das Fixed Effect Modell würde ein zu enges Konfidenzintervall liefern. Große Unterschiede

zwischen Fixed Effect Modell und Random Effects Modell können darauf hinweisen, dass sich die kleinen Studien systematisch von den großen unterscheiden.

Die Wahl des Modells erfolgt bei der Erstellung des Protokolls, nicht auf der Basis der späteren Ergebnisse: „The choice between a fixed effect and a random effects meta-analysis should never be made on the basis of a statistical test of heterogeneity“²³

Welche Arten von Heterogenität gibt es?

- Klinische Heterogenität: verursacht durch Unterschiede in den Baseline-Patientencharakteristika (ggf. auch durch unterschiedliche Interventionen, die bei der Auswertung im Hinblick auf z.B. Dosierung oder Zeitdauer, zusammengefasst wurden)
- Statistische Heterogenität: auf der Endpunktskala quantifizierbar, statistische Heterogenität kann u.a. mit Cochran's Q, I^2 oder τ^2 gemessen werden (siehe unten)
- Heterogenität verursacht durch unterschiedliche Studientypen („design-related“)
- Heterogenität verursacht durch „small study effects“

Wie testet man Ergebnisse auf statistische Heterogenität?

Mit Hilfe der statistischen Heterogenität kann ich entscheiden, ob es überhaupt Sinn macht, die einzelnen Studienergebnisse in einem gemeinsamen Schätzer zusammenzufassen. Folgende Heterogenitätstest können u.a. dabei angewendet werden:

Cochran's Q:

- Q=0, falls der Therapieeffekt in allen Studien gleich
- Q untersucht nur, ob statistisch signifikante Heterogenität vorhanden ist oder nicht (JA/Nein)
- Q macht keine Angabe über das Ausmaß an Heterogenität, dafür ist I^2 besser

Higgins' und Thompson's I^2 :⁴⁰

- $I^2=0\%$, d.h., Variabilität zwischen den Studien ausschließlich durch Zufallsschwankungen erklärbar
- $I^2=100\%$, d.h., Variabilität zwischen den Studien ausschließlich durch Unterschiede zwischen den Studien erklärbar
- I^2 misst den Überlappungsgrad der Konfidenzintervalle
- I^2 wächst mit Präzision (ist also abhängig von Studiengröße), bei sehr großen Studien, die gleichzeitig ein enges Konfidenzintervalle aufweisen, geht I^2 zwangsläufig gegen 100%; das bedeutet jedoch nicht unbedingt einen klinisch relevanten Unterschied

Interstudienvarianz τ^2 :

- τ^2 beschreibt Variabilität zwischen den Studien, hängt aber nicht wie I^2 von der Studiengröße ab
- dafür gibt es in der Literatur mehrere Schätzer, der nach DerSimonian und Laird (1986) gilt als der am meisten verwendete
- die Wurzel von τ^2 , also τ , ist in der gleichen Einheit wie der Endpunkt, somit lässt sich Heterogenität quantifizieren; es sollte jedoch vorher festgelegt werden, ab wann eine bedeutsame Heterogenität auf der Endpunktskala vorliegt

Was mache man bei (sehr) heterogenen Ergebnissen?

Das Random Effects Modell kann Heterogenität beschreiben, aber nicht erklären. Zur Erklärung von Heterogenität sind zusätzliche Untersuchungen notwendig. Subgruppenanalysen und Metaregressionen können helfen, die statistische Heterogenität zu erklären. Solche Analysen müssen vorab spezifiziert werden. Heterogenität kann ggf. auch klinisch erklärbar sein, bei nicht erklärbarer erheblicher Heterogenität sollte auf eine Metaanalyse verzichtet werden.

Subgruppenanalyse

Es werden Subgruppen von Studien nach vorab spezifizierten dichotomen oder nominalen klinischen Faktoren (wie z.B. Geschlecht, Altersgruppen, Komorbidität, sofern diese in den Primärstudien getrennt berichtet werden) gebildet. Diese Subgruppen können eventuell Unterschiede in der Effektgröße erklären.

Metaregression

Eine Metaregression ist eine Verallgemeinerung der Subgruppenanalyse, bei der auch kontinuierliche Kovariablen und Effekte mehrerer Kovariablen jeglichen Typs simultan untersucht werden können. Eine Metaregression sollte nur durchgeführt werden, wenn mindestens 10 Studien in die Metaanalyse eingeschlossen wurden. Es wird allerdings davor gewarnt, Kovariablen, die auf Patientenebene definiert sind, in einer Subgruppenanalyse oder Metaregressionsanalyse zu untersuchen, da es dabei zu einem Bias durch Confounding kommen kann.⁴⁰

Sensitivitätsanalyse

Im Unterschied zu Subgruppenanalyse und Metaregression, bei denen es um klinische Heterogenität geht, ist das Ziel einer Sensitivitätsanalyse, zu untersuchen, ob der Ausschluss einzelner Studien aus der Analyse (z.B. aufgrund eines hohen Verzerrungsrisikos) einen Einfluss auf das Ergebnis (Effektgröße) hat.

5.2 Erläuterungen zum Disseminationsbias

Was ist ein Disseminationsbias?

Disseminationsbias (dissemination bias) ist der Oberbegriff zu einer ganzen Reihe von Biastypen, die auf eine selektive Verbreitung von Studienergebnissen zurückgehen. Dazu gehören:

- publication bias (Publikationsbias; Studien mit negativen Ergebnissen werden seltener oder gar nicht publiziert)
- language bias (Sprachbias; Studien mit negativen Ergebnissen werden eher in nationalsprachlichen Zeitschriften publiziert)
- time-lag bias (Studien mit negativen Ergebnissen werden zeitverzögert publiziert)
- citation bias (Zitationsbias; Studien mit negativen Ergebnissen werden seltener zitiert)

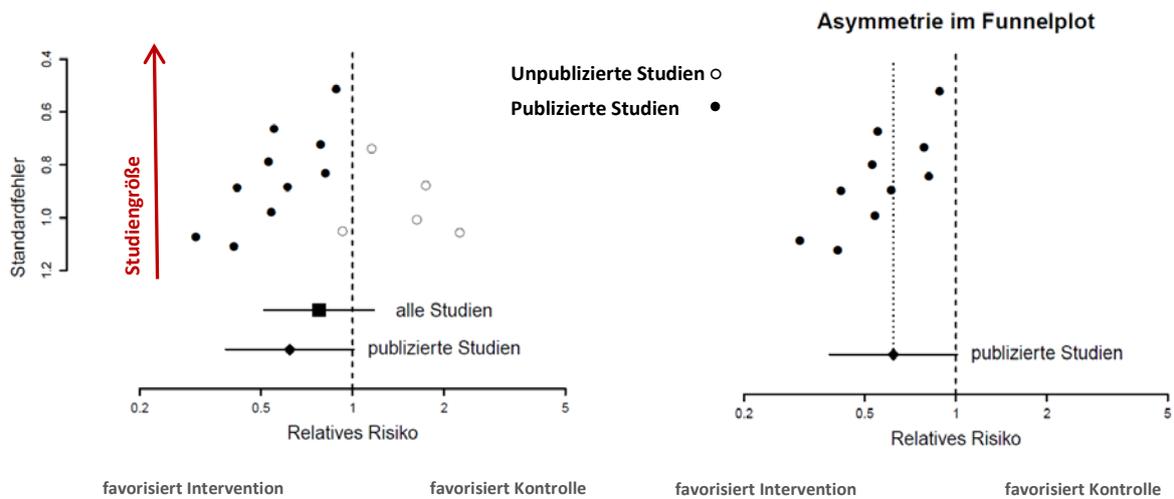
Wie kann man einen Disseminationsbias untersuchen?

Mit Hilfe von graphischen und/oder statistischen Tests lässt sich das Vorhandensein von Disseminationsbias ‚annäherungsweise‘ untersuchen.

Funnelplot (graphischer Test)

Der Funnelplot ist ein Scatterplot, bei dem ein Maß für die Studiengröße, zum Beispiel der Standardfehler in umgekehrter Richtung, gegen die Effektgröße (z.B. ein relatives Risiko) abgetragen wird. Werden kleine Studien mit geringen oder unerwünschten Effekten nicht publiziert, so erkennt man das an einem Funnelplot an der durch ihr Fehlen entstehenden Asymmetrie (in der rechten Abbildung fehlen die weißen Kreise).

Die Untersuchung von Disseminationsbias mit Hilfe eines Funnelplots macht jedoch erst ab einer Anzahl von mindestens 10 eingeschlossenen Studien in die Metaanalyse Sinn: „As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the metaanalysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry.“²³



Effekt bei publizierten Studien im Vergleich zu allen Studien stärker ausgeprägt

Abb. 1. Funnel Plot.

Mögliche Erklärungen für die Asymmetrie im Funnelplot:

- Klinische Heterogenität zwischen den Studien verursacht durch Unterschiede im Baseline-Risiko bei (stark) selektierte Patienten in kleinen Studien
- Sonstige Heterogenität: hohes Verzerrungsrisiko in kleinen Studien, was eine Überschätzung der Ergebnisse zur Folge haben kann („small study effects“)
- Selektive Berichterstattung: Wahl der Zielgröße nach Effektstärke
- Zufall

5.2 Leitlinien zur Verbesserung der Berichterstattung

Leitlinien zur Verbesserung der Berichterstattung verschiedener Studientypen einschließlich systematischer Übersichtsarbeiten sind auf der Webseite des EQUATOR Netzwerks abrufbar: <http://www.equator-network.org/>.

equator network Enhancing the **QUALITY and Transparency Of health Research** EQUATOR resources in Portuguese | Spanish

Home Library Toolkits Courses & events News Blog Librarian Network About us Contact

Your one-stop-shop for writing and publishing high-impact health research
find reporting guidelines | improve your writing | join our courses | run your own training course | enhance your peer review | implement guidelines

Library for health research reporting
The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.

- Search for reporting guidelines
- Not sure which reporting guideline to use?
- Reporting guidelines under development
- Visit the library for more resources

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions	Other
Observational studies	STROBE	Extensions	Other
Systematic reviews	PRISMA	Extensions	Other
Case reports	CARE	Extensions	Other
Qualitative research	SRQR	COREQ	Other
Diagnostic / prognostic studies	STARD	TRIPOD	Other
Quality improvement studies	SQUIRE		Other
Economic evaluations	CHEERS		Other
Animal pre-clinical studies	ARRIVE		Other
Study protocols	SPIRIT	PRISMA-P	Other

[See all 346 reporting guidelines](#)

EQUATOR Oncology Project
CANCER RESEARCH UK
resources supporting complete, accurate and transparent research
EQUATOR Oncology Project
Funder research

Toolkits EQUATOR highlights News

Abb. 2. Screenshot der Equator Webseite (<http://www.equator-network.org/>).

5.3 PRISMA Statement – deutsche Übersetzung

Leitlinie zur Verbesserung der Berichtsqualität von systematischen Übersichtsarbeiten oder Meta-Analysen (deutsche Übersetzung aus: Ziegler A, Antes G, IR K. Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. Dtsch Med Wochenschr 2011;136 e9–e15).¹⁹

Tabelle 3. PRISMA Statement (deutsche Übersetzung).

Publikationsabschnitt		Zusammenfassung
Titel		
Titel	1	Identifikation als Bericht einer systematischen Übersicht, Meta-Analyse oder beidem.
Zusammenfassung		
Strukturierte Zusammenfassung	2	Strukturierte Zusammenfassung mit den Stichworten (sofern geeignet): Hintergrund; Ziele; Datenquellen; Auswahlkriterien der Studien, Teilnehmer und Interventionen; Bewertung der Studie und Methoden der Synthese; Ergebnisse; Einschränkungen; Schlussfolgerungen und Implikation der wichtigsten Ergebnisse; Registrierungsnummer der systematischen Übersicht.
Einleitung		
Hintergrund & Rationale	3	Wissenschaftlicher Hintergrund und Begründung der Studie.
Ziele	4	Präzise Angabe der Fragestellungen mit Bezug auf Teilnehmer, Interventionen, Vergleiche, Zielkriterien und Studiendesign (engl.: participants, interventions, comparisons, outcomes, and study design; PICOS).
Methoden		
Protokoll und Registrierung	5	Existiert ein Studienprotokoll für die Übersichtsarbeit? Wenn ja, wo kann es gefunden bzw. wie kann es bezogen werden (z.B. Webseite); wenn verfügbar: Informationen zur Registrierung einschließlich Angabe der Registrierungsnummer.
Auswahlkriterien	6	Merkmale der Studien (z. B., PICOS, Dauer der Nachbeobachtung) und der Berichte (z. B., Zeitraum der Studien, Sprache, Publikationsstatus), die als Auswahlkriterien verwendet wurden, mit Begründung.
Informationsquellen	7	Beschreibung aller Informationsquellen (z. B., Datenbanken mit Zeitpunkten der Berichterstattung, Kontakt mit Autoren von Studien, um zusätzliche Studien zu identifizieren), die bei der Suche verwendet wurden einschließlich des letzten Suchdatums.
Suche	8	Beschreibung der vollständigen elektronischen Suchstrategie für mindestens eine Datenbank, einschließlich gewählter Limitierungen, so dass die Suche repliziert werden könnte.
Auswahl der Studien	9	Beschreibung des Auswahlprozesses von Studien (das heißt Vorauswahl, Eignung, Einschluss in die systematische Übersicht und, falls zutreffend, in die Meta-Analyse).

Bewertung von systematischen Übersichtsarbeiten

Version 1.0

Prozess der Datengewinnung	10	Beschreibung der Methode der Datenextraktion aus Berichten (z.B. Erhebungsbogen, unabhängig, doppelt) und alle Prozesse, um Daten von Untersuchern zu erhalten und zu bestätigen.
Datendetails	11	Aufzählung und Definition aller Variablen, nach denen gesucht wurde (z.B., PICOS, Finanzierungsquellen) sowie Annahmen und durchgeführte Vereinfachungen.
Risiko der Verzerrung in den einzelnen Studien	12	Methoden zur Beurteilung des Risikos von Verzerrungen der einzelnen Studien (einschließlich der Angabe, ob dieses auf der Studienebene oder für das Zielkriterium durchgeführt wurde) und wie diese Information bei der Datensynthese berücksichtigt wurde.
Effektschätzer	13	Wichtigste Effektschätzer (z. B. relatives Risiko, Mittelwertdifferenz).
Synthese der Ergebnisse	14	Beschreibung der Methoden zum Umgang mit den Daten und der Kombination der Ergebnisse der Studien; falls diese berechnet wurden, einschließlich Maßzahlen zur Homogenität der Ergebnisse (z. B. I^2) für jede Meta-Analyse.
Risiko der Verzerrung über Studien hinweg	15	Beschreibung der Beurteilung des Risikos von Verzerrungen, die die kumulative Evidenz beeinflussen könnten (z. B. Publikationsverzerrung, selektives Berichten innerhalb von Studien).
Zusätzliche Analysen	16	Methoden für zusätzliche Analysen (z.B. Sensitivitätsanalysen, Subgruppenanalysen, Meta-Regression) mit Beschreibung, welche vorab spezifiziert waren.

Ergebnisse

Auswahl der Studien	17	Anzahl der Studien, die in die Vorauswahl aufgenommen, auf Eignung geprüft und in die Übersicht eingeschlossen wurden, mit Begründung für Ausschluss in jeder Stufe, idealerweise unter Verwendung eines Flussdiagramms.
Studienmerkmale	18	Für jede Studie Darstellung der Merkmale, nach denen Daten extrahiert wurden (z. B. Fallzahl, PICOS, Nachbeobachtungszeitraum), Literaturstelle der Studie.
Risiko der Verzerrung innerhalb der Studien	19	Daten zum Risiko von Verzerrungen innerhalb jeder Studie und, falls verfügbar, eine Beurteilung der Güte der Zielkriterien (siehe Item 12).
Ergebnisse der einzelnen Studien	20	Für jede Studie Darstellung aller Endpunkte (Wirksamkeit und Nebenwirkungen): (a) einfache zusammenfassende Daten für jede Interventionsgruppe, (b) Effektschätzer und Konfidenzintervalle, idealerweise mit Forest Plot.
Ergebnissynthese	21	Darstellung der Meta-Analyse, einschließlich Konfidenzintervalle und Heterogenitätsmaße.
Risiko von Verzerrungen über Studien hinweg	22	Darstellung der Ergebnisse zur Beurteilung des Risikos von Verzerrungen über alle Studien hinweg (siehe Item 15).
Zusätzliche Analysen	23	Präsentation der Ergebnisse der zusätzlichen Analysen, falls durchgeführt (z. B. Sensitivitäts- oder Subgruppenanalysen, Meta-Regression [siehe Item 16]).

Diskussion

Zusammenfassung der Evidenz	24	Zusammenfassung der Hauptergebnisse einschließlich der Stärke der Evidenz für jedes Hauptzielkriterium; Relevanz für Zielgruppen (z. B. Gesundheitsdienstleister, Anwender, politische Entscheidungsträger).
-----------------------------	----	--

Bewertung von systematischen Übersichtsarbeiten
Version 1.0

Einschränkungen	25	Diskutiere Einschränkungen der Studie auf Studienebene und auf Ebene der Zielkriterien (z. B. Risiko von Verzerrungen) sowie auf Ebene der Übersicht (z. B. unvollständiges Auffinden der identifizierten Forschung, Verzerrung des Berichts).
Schlussfolgerungen	26	Interpretation der Ergebnisse unter Berücksichtigung des Stands der Forschung und Schlussfolgerungen für weitere Forschung.

Finanzielle Unterstützung

Finanzielle Unterstützung	27	Quellen der finanziellen Unterstützung sowie andere Unterstützung (z. B. zur Verfügung stellen von Daten); Funktion der Geldgeber für die systematische Übersicht.
---------------------------	----	--
