

Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung

Version 1.0 vom 04.05.2016



Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien

Version 1.0

Das Manual „Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung“ wurde im Rahmen des vom Bundesministerium für Gesundheit (BMG) geförderten Projekts „ACQuire“ 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: RoB@cochrane.de.

Autoren

Schmucker C¹, Nothacker M², Rücker G^{1,3}, Muehle-Borowski C², Kopp I², Meerpohl JJ¹

Kontakte

¹Cochrane Deutschland

Universitätsklinikum Freiburg

Medizinische Fakultät

Berliner Allee 29

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

³Institut für Medizinische Biometrie und Statistik

Universitätsklinikum Freiburg

Medizinische Fakultät

Stefan-Meier-Strasse 26

D-79104 Freiburg

www.imbi.uni-freiburg.de

Wir bedanken uns bei Prof. Dr. Gerd Antes, Nico Gagelmann, Dipl. Soz. Wiss. Thomas Langer, PD Dr. Petra Lynen, Dr. Jost Schnell und Prof Dr. Karl Werdan für Durchsicht und Kommentierung.

Bitte wie folgt zitieren

Cochrane Deutschland, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement. „Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung“. 1. Auflage 2016.

Verfügbar: Cochrane Deutschland: <http://www.cochrane.de/de/rob-manual>; AWMF: <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html>.

Disclaimer

Die in diesem Manual verwendete männliche Form gilt für *Personen* 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/rob-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 Biasbewertung in Studien für die Leitlinienerstellung wichtig?	7
1.1 Ziele und Struktur des Manuals	8
2. Das Biasrisiko in klinischen Studien (interne Validität)	9
2.1 Biasrisiko <i>versus</i> externe Validität	9
2.2 Biasrisiko <i>versus</i> unzureichende Präzision der Ergebnisse	10
2.3 Biasrisiko <i>versus</i> Studienqualität	10
2.4 Biasrisiko <i>versus</i> Berichtsqualität	11
2.5 Biasrisiko im Kontext von GRADE (Bewertung der Qualität der Evidenz).....	12
3. Verschiedene Biasformen und deren Auswirkung.....	13
3.1 Wichtige Biasformen in randomisierten kontrollierten Studien	13
3.2 Wichtige Biasformen in nicht-randomisierten Studien	16
3.3 Auswirkung von Bias auf die Ergebnisse von klinischen Studien	19
4. Bewertungsinstrumente	20
5. Bewertung des Biasrisikos nach den Cochrane Methoden.....	22
5.1 Randomisierte kontrollierte Studien.....	22
5.1.1. Domänen der Bewertung.....	23
5.1.2. Dokumentation und Berichten des Verzerrungspotenzials	30
5.2 Nicht-randomisierte Studien	33
5.2.1 Vergleichende nicht-randomisierte Studien	33
5.2.1.1 Domänen der Bewertung (nach Acrobot-NRSI)	35
5.2.1.2 Dokumentation und Berichten des Verzerrungspotenzials	39
5.2.2 Nicht-vergleichende Studien.....	42
6. Bewertung des Biasrisikos nach der Newcastle Ottawa Skala (NOS)	43
6.1 Bewertung von Fall-Kontrollstudien	43
6.2 Bewertung von Kohortenstudien.....	45
6.3 Tabellenvorlage	47
7. Quellen	48
8. Weiterführende Informationen und Praxishilfen	51

Glossar

ACROBAT-NRSI	A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions
Allocation concealment	Verdeckte Zuteilung (Geheimhaltung der Behandlungsfolge)
Attrition (Abnahme)	Der Verlust von Teilnehmern während der Studiendurchführung.
Bias	Systematischer Fehler; Verzerrung
Carry-Over-Effekt	Ein Carry-Over-Effekt tritt häufig bei Cross-Over-Studien auf. Dabei kann die erste Therapie (Therapie A) Einfluss auf den weiteren Behandlungsverlauf (unter Therapie B) nehmen.
Confounder (Störgröße)	Confounding liegt vor, wenn ein Faktor (Confounder), der nicht direkt Gegenstand der Untersuchung ist, sowohl mit der Intervention/Exposition als auch mit der Zielgröße assoziiert ist und dadurch bei Aussagen über die Beziehung zwischen Intervention/Exposition und Zielgröße „Verwirrung“ stiftet. Häufige Confounder sind z.B. Alter, Geschlecht oder Nikotingenuss. Confounding lässt sich grundsätzlich durch ein entsprechendes Studiendesign (z.B. Randomisierung) minimieren oder durch die Anwendung bestimmter statistischer Verfahren bei der Analyse (z.B. Regression oder Stratifizierung) kontrollieren.
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.
Cross-Over-Studie	Studiendesign, in dem die zu vergleichenden Interventionen in den Vergleichsgruppen in zeitlicher Folge angewandt werden. Dabei erhält z.B. die eine Gruppe zunächst Therapie A, dann Therapie B, die andere Gruppe zuerst Therapie B und dann Therapie A.
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).
Drop-Outs	Studienteilnehmer, die während der eigentlichen Studienphase aus der Studie ausscheiden.
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]).
GCP	Good Clinical Practice
GEP	Good Epidemiological Practice
GRADE	Grading of Recommendations, Assessment, Development and Evaluation

Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien
Version 1.0

Indirektheit	Diskrepanz zwischen interessierender Patientenpopulation oder interessierender Intervention oder interessierenden Endpunkt(en) (outcomes) und der in der Studie untersuchten Population, Intervention oder Endpunkt(en).
Inkonsistenz	Vorliegen widersprüchlicher Studienergebnisse
Intention-to-Treat Auswertung (ITT Analyse)	Spezifische Auswertungsmethodik mit Berücksichtigung der Studienabbrecher und Patienten, die die Behandlungsgruppen wechselten; Behandlungsgruppenwechsler werden dabei der ursprünglich zugeteilten Gruppe zugeordnet.
Lost to Follow-Up	Studienteilnehmer, die die eigentliche Studienphase zwar abgeschlossen haben, jedoch nicht mehr zu den Nachsorgeterminen erscheinen.
Matching	Beim Matching wird für jeden Patienten bzw. Probanden aus einer bestimmten Gruppe (Gruppe A) ein „statistischer Zwilling“ für Gruppe B gesucht. Dieser sollte in verschiedenen, zuvor definierten Variablen (bzw. Störgrößen) den Patienten bzw. Probanden aus Gruppe A entsprechen (identisch sein oder zumindest möglichst ähnlich). Probanden werden vor der Gruppenzuordnung zu Paaren zusammengestellt, die auf einer oder mehreren Matchingvariablen möglichst ähnlich sind.
Minimierung	Randomisierungsverfahren, um auch bei kleinen Fallzahlen eine Gleichverteilung der Patientencharakteristika zu erreichen.
NOS	Newcastle Ottawa Skala
Per-Protokoll Auswertung	Auswertung, bei der nur die Personen eingeschlossen werden, die protokollgemäß behandelt wurden.
RCT	Randomised Controlled Trial; randomisierte kontrollierte Studie
RoB	Risk of Bias (Biasrisiko, Risiko für Bias), Risiko für einen systematischen Fehler, Verzerrungsrisiko
RoBANS	Risk of Bias Assessment Tool for Non-randomized Studies
SIGN	Scottish Intercollegiate Guidelines Network
Stratifizierung (Schichtung)	Analyse unter Berücksichtigung von Subgruppen (z.B. Geschlecht, Alter oder auch Zentrum).

Tabellen- und Abbildungsverzeichnis

Tabelle 1. Wesentliche Aspekte der internen und externen Validität in randomisierten Studien.....	9
Tabelle 2. Beispiel für eine Cochrane RoB Tabelle für eine einzelne randomisierte Studie.....	31
Tabelle 3. Beispiel für eine Cochrane RoB Tabelle für eine einzelne nicht-randomisierte Studie.....	40
Tabelle 4. Cochrane RoB Zusammenfassung für eine einzelne nicht-randomisierte Studie.....	41
Tabelle 5. NOS RoB Tabelle für nicht-randomisierte Studien.....	47
Abb. 1. Cochrane RoB Zusammenfassung für fiktive Studien	32
Abb. 2. Cochrane RoB Graph für fiktive Studien	32
Abb. 3. Screenshot der Equator Webseite	51

1. Einleitung: Warum ist die Biasbewertung in Studien für die Leitlinienerstellung wichtig?

Sowohl praktisch tätige Ärzte als auch Angehörige der Gesundheitsfachberufe treffen täglich eine Vielzahl von medizinischen Entscheidungen. Diese Entscheidungen basieren überwiegend auf dem im Studium und in der Ausbildung Erlernten und der persönlichen Erfahrung. Es ist jedoch wichtig, dass bei Entscheidungen im Gesundheitswesen darüber hinaus Patientenpräferenzen und die wissenschaftliche Evidenz, die zu Nutzen und Schaden einer Intervention vorliegt, berücksichtigt werden. Die Evidenzbasierte Medizin (EbM) hat zum Ziel, dass Behandlungsentscheidungen für den einzelnen Patienten auf der Basis der individuellen Erfahrung des Arztes unter Berücksichtigung der besten verfügbaren Evidenz in Abwägung der Wünsche und Vorstellungen des Patienten getroffen werden. Die EbM beinhaltet dabei: (1) die systematische Suche nach relevanter Evidenz (Studien) in der medizinischen Literatur¹, (2) die kritische Beurteilung der internen Validität der Evidenz (Bewertung systematischer Fehler, Bewertung des Biasrisikos [**Risk of Bias**, im Folgenden abgekürzt durch **RoB**]), (3) die Bewertung der Größe der beobachteten Effekte und deren klinischer Relevanz sowie (4) die Anwendung dieser Evidenz auf den individuellen Patienten. Ohne Verständnis der methodischen Grundlagen von klinischen Studien (insbesondere im Hinblick auf eine unverzerrte Auswahl und Bewertung der Evidenzbasis) ist die Gefahr von Fehleinschätzungen der vorhandenen Evidenz hoch und kann suboptimales Management bis hin zu Behandlungsfehlern zur Konsequenz haben.

Im vorliegenden Manual – aufbauend auf einer deutschen Übersetzung des RoB Tools aus dem Cochrane Handbook for Systematic Reviews of Interventions² – wird das Potential für eine systematische Verzerrung (RoB) in (klinischen) Studien und seine Bedeutung für die Erstellung von systematischen Übersichtsarbeiten und darauf aufbauenden Leitlinien und somit auch für das klinische Handeln erläutert.

Die Gesamtheit der Studien, die die Einschlusskriterien für eine systematische Übersichtsarbeit und konsekutiv auch Leitlinie erfüllen, müssen für die Erstellung berücksichtigt werden. Dabei ist es wichtig, dass die identifizierten Studien kritisch bewertet werden. Nur so lässt sich feststellen, wie vertrauenswürdig die Ergebnisse der systematischen Übersichtsarbeit (mit oder ohne statistische Datenauswertung in Form einer Metaanalyse) beziehungsweise der darauf basierenden Leitlinienempfehlungen sind.

1.1 Ziele und Struktur des Manuals

Das Manual „Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien, Version 1.0“ richtet sich vor allem an Leitlinienersteller und Leitlinienberater mit dem Ziel, sie über eine qualifizierte Literaturbewertung im Hinblick auf systematische Verzerrungen von Studien zu informieren und bei gegebener praktischer Anleitung (zum Beispiel im Rahmen eines Leitlinienseminars) zu einer eigenständigen RoB Bewertung zu befähigen.

Die Grundlagen dieses Manuals bilden international anerkannte Standards für die Beschreibung und Bewertung von Bias in klinischen Studien.

Das Manual ergänzt und vertieft das AWMF-Regelwerk zur Erstellung von Leitlinien, im Besonderen das Kapitel „Recherche, Auswahl und methodische Bewertung von Literatur und deren Aufbereitung, Planung und Organisation“.¹ Das Manual besteht aus den Kernkapiteln: (2.) Abgrenzung des RoB gegenüber anderen Qualitätsaspekten klinischer Studien, (3.) wichtige RoB Formen und (4.-6.) die RoB Bewertung einschliesslich der Darstellung verschiedener Instrumente zur RoB Bewertung.

Das Manual soll kontinuierlich fortgeschrieben werden. Kommentare sind daher ausdrücklich erwünscht und können gerichtet werden an:

RoB@cochrane.de

2. Das Biasrisiko in klinischen Studien (interne Validität)

- 2.1 Biasrisiko *versus* externe Validität
- 2.2 Biasrisiko *versus* unzureichende Präzision der Ergebnisse
- 2.3 Biasrisiko *versus* Studienqualität
- 2.4 Biasrisiko *versus* Berichtsqualität
- 2.5 Biasrisiko im Kontext von GRADE

3. Verschiedene Biasformen und deren Auswirkung

- 3.1 Wichtige Biasformen in randomisierten kontrollierten Studien
- 3.2 Wichtige Biasformen in nicht-randomisierten Studien
- 3.3 Auswirkung von Bias auf die Ergebnisse von klinischen Studien

4. Bewertungsinstrumente

5. Bewertung des Biasrisikos nach den Cochrane Methoden

- 5.1 Randomisierte kontrollierte Studien
- 5.2 Nicht-randomisierte Studien
 - 5.2.1 Vergleichende nicht-randomisierte Studien
 - 5.2.2 Nicht-vergleichende Studien

6. Bewertung des Biasrisikos nach der Newcastle Ottawa Skala (NOS)

2. Das Biasrisiko in klinischen Studien (interne Validität)

2.1 Biasrisiko *versus* externe Validität

Ein wesentlicher Aspekt einer jeden Studie stellt die Validität, d.h. die Gültigkeit der Studienergebnisse, dar.³ Bei der Validität wird dabei zwischen interner und externer Studienvolidität unterschieden (Tabelle 1):

(i) Die *interne Validität* lässt eine Aussage darüber zu, inwieweit man dem gemessenen Effekt vertrauen kann. Sie hängt von der Durchführung, Auswertung und Berichterstattung der Studie ab und wird über die RoB Bewertung bestimmt. Bias in klinischen Studien kann sowohl zu einer Über- als auch Unterschätzung der Wirksamkeit und/oder Risiken einer Maßnahme oder Exposition führen.

(ii) Die *externe Validität* hingegen bezeichnet die Generalisierbarkeit oder Übertragbarkeit der Untersuchungsergebnisse und hängt damit von der Fragestellung, den Ein- und Ausschlusskriterien und dem Setting der Studie ab. Sie gibt an, ob Studienresultate auf andere Personen, Situationen und/oder Zeitpunkte übertragen werden können.

Tabelle 1. Wesentliche Aspekte der internen und externen Validität in randomisierten kontrollierten Studien (Quelle: modifiziert nach Jüni et al.³)

Interne Validität: Ausmaß, in dem der beobachtete Effekt in einer Studie nicht durch systematische Fehler verzerrt wurde				
selection bias	performance bias	detection bias	attrition bias	reporting bias
Verzerrung durch Unterschiede in den Patientencharakteristika zwischen den Studiengruppen	Verzerrung durch Unterschiede in der Behandlung; abgesehen von der untersuchten Intervention	Verzerrte Erfassung von Endpunkten	Verzerrung durch Unterschiede in der Anzahl und den Gründen für fehlende Daten zwischen den Studiengruppen	Verzerrung durch selektives Berichten von positiven Ergebnissen
Externe Validität: Ausmaß, in dem eine Verallgemeinerung der Studienergebnisse möglich ist				
Patienten		Behandlungsplan		Setting
Alter, Geschlecht, Schweregrad, (bio-psycho-soziale) Risikofaktoren, Ko-Morbidität		Dosierung, Häufigkeit und Art der Verabreichung, Art des Präparats, Begleitbehandlungen		Versorgungsstufe (primär, sekundär, tertiär), Erfahrung und Spezialisierung des Leistungserbringers

2.2 Biasrisiko *versus* unzureichende Präzision der Ergebnisse

Eine unzureichende Präzision von Ergebnissen ist auf kleine Fallzahlen beziehungsweise auf eine geringe Anzahl an Ereignissen (Events), aber nicht auf systematische Fehler (RoB) zurückzuführen und muss von Bias unterschieden werden. Die Präzision des Effektschätzers wird dabei durch das Konfidenzintervall angegeben. In einer Metaanalyse spiegelt sich die Präzision einer Studie im jeweiligen ‚Gewicht‘ der Studie wider. Studien mit präziseren Ergebnissen (d.h. Studien mit hohen Fall- und Eventzahlen) bekommen dabei ein größeres Gewicht als Studien mit weiten Konfidenzintervallen (d.h. Studien mit kleinen Fall- und/oder Eventzahlen).⁴

2.3 Biasrisiko *versus* Studienqualität

Grundsätzlich soll in einer klinischen Studie von der Planung über die Durchführung bis zur Auswertung und Publikation nach einem standardisierten Konzept vorgegangen werden, um eine hohe Studienqualität zu gewährleisten. Neben einem Votum der Ethikkommission gehört dazu vor allem ein Studienprotokoll, in dem alle wichtigen Methoden und Vorgehensweisen prospektiv beschrieben werden. Die wichtigsten Qualitätsstandards für (randomisierte) Studien stellen die Grundprinzipien der „Good Clinical Practice“ (GCP)^{5 6} und für epidemiologische Studien die Empfehlungen zur Sicherung „Guter Epidemiologischer Praxis“ (GEP)⁷ dar.

Im Gegensatz zur Studienqualität adressiert die RoB Bewertung das Vertrauen in die im Rahmen der Studie generierten Effektschätzer. RoB wird von der Qualität der Studien zwar maßgeblich beeinflusst, die Bewertung des Verzerrungspotentials ist jedoch nicht gleichbedeutend mit einer Qualitätseinschätzung von Studien. Demzufolge kann Bias in Studien auftreten, die methodisch adäquat durchgeführt wurden. Zum Beispiel ist es in der Chirurgie häufig nicht möglich, Studienteilnehmer und/oder -personal im Hinblick auf die Intervention und/oder Endpunkterhebung zu verblinden. Obwohl solche Studien nach bestmöglichen Standards durchgeführt wurden, können einzelne Endpunkte – bedingt durch die fehlende Verblindung – ein hohes RoB haben. Andererseits führen nicht alle methodischen Mängel zu verzerrten Studienergebnissen. Zum Beispiel wirkt sich eine fehlende Fallzahlplanung (*siehe Abschnitt 2.2: Biasrisiko versus unzureichende Präzision der Ergebnisse*), das Fehlen eines Ethikvotums, eines Studienprotokolls oder einer Studienregistrierung nicht auf die interne Validität einer Studie aus.

2.4 Biasrisiko *versus* Berichtsqualität

Die interne Validität einer klinischen Studie muss klar von der Berichtsqualität abgegrenzt werden.⁸ Die Berichtsqualität umfasst Aspekte wie Vollständigkeit, Detailliertheit, Objektivität und Nachvollziehbarkeit eines Studienberichtes (in der Regel einer Publikation in einer wissenschaftlichen Fachzeitschrift). Studien mit geringem Verzerrungspotential können durchaus eine geringe Berichtsqualität aufweisen, wenn zum Beispiel wichtige Details zu methodischen Aspekten wie Randomisierung oder Verblindung fehlen. Auf der anderen Seite können Studien mit hohem Verzerrungspotential, zum Beispiel durch fehlende Geheimhaltung der Behandlungsfolge (Allocation Concealment), fehlende Verblindung oder hohem Verlust an Patienten bei der Nachbeobachtung (Loss to follow-up) eine hohe Berichtsqualität aufweisen, sofern diese Aspekte explizit beschrieben und mögliche, resultierende Limitationen im Artikel diskutiert wurden. Eine valide Bewertung des Verzerrungspotentials einer Studie setzt in der Regel eine ausreichende Berichtsqualität voraus.

Zur klareren Abgrenzung zwischen Mängeln in der Berichtsqualität und tatsächlichem RoB wurde von Herausgebern wissenschaftlicher Zeitschriften, klinischen Forschern, Epidemiologen und Methodikern zu Beginn der 1990er Jahre eine Initiative zur Verbesserung der Berichtsqualität von Publikationen zu randomisierten kontrollierten Studien ins Leben gerufen. Das Ergebnis war das CONSORT (**CON**solidated **S**tandards **Of** **R**eporting **T**rials) -Statement, eine ‚Orientierungshilfe‘, um die Berichterstattung von randomisierten kontrollierten Studien zu verbessern. Das CONSORT-Statement enthält 25 Aspekte, die in Publikationen zu randomisierten Studien enthalten sein sollten.⁹ Eine Überarbeitung des CONSORT-Statements erfolgte zuletzt im Jahr 2010.¹⁰ Neben wichtigen Aspekten zur Studienmethodik und der Ergebnisdarstellung wird im CONSORT-Statement ein Flussdiagramm gefordert, das die Anzahl der Patienten (einschließlich fehlender Daten) von Beginn bis Ende einer Studie abbildet. Wenige Jahre nach Veröffentlichung des CONSORT-Statements verbesserte sich in drei hochrangigen Zeitschriften, die die obligatorische Verwendung der ‚Checkliste‘ vorsehen (British Medical Journal, Journal of the American Medical Association, The Lancet), die Berichtsqualität von randomisierten kontrollierten Studien signifikant.¹¹ In der Zwischenzeit wurden weitere Leitlinien zur Verbesserung der Berichterstattung für andere Studientypen (als randomisierte kontrollierte Studien) sowohl für Autoren als auch für Herausgeber von Zeitschriften und Gutachter publiziert. Sie sind auf den Internetseiten des EQUATOR Netzwerks abrufbar: <http://www.equator-network.org/> (siehe Abschnitt 8: Weiterführende Informationen und Praxishilfen).

✓ Werden in einer Publikation Standards der Berichterstattung nicht eingehalten, so kann dies ein Indikator für ein potentielles RoB sein, muss es aber nicht sein. Es ist wichtig, dass Review- und Leitlinienautoren angeben, ob Studienaspekte unzureichend oder nicht berichtet wurden und ob eine konkrete Einschätzung des Verzerrungspotenzials erfolgen konnte. Eine Option bei unklarer Berichterstattung stellt die Kontaktaufnahme mit den Studienautoren dar. Oft lassen sich dadurch fehlende, beziehungsweise unklare Studienaspekte, die für eine valide Bewertung des Biasrisikos ausschlaggebend sind, klären.

2.5 Biasrisiko im Kontext von GRADE (Bewertung der Qualität der Evidenz)

GRADE (Grading of Recommendations, Assessment, Development and Evaluation) bietet ein System, um die Qualität der identifizierten Evidenz zu einer Fragestellung einzuschätzen.¹² Beurteilungen von GRADE beziehen sich nicht auf einzelne Studien, sondern auf die Gesamtheit der gefundenen Evidenz für einen Endpunkt (Therapieziel/Outcome), d.h. die Evidenz wird jeweils bezogen auf einen Endpunkt bewertet. Für die Studien, die in diesen Evidenzkörper eingehen, wird eine RoB Bewertung vorausgesetzt. Bei der Bewertung der Qualität der gesamten Evidenz, wie sie in GRADE erfolgt, wird somit zunächst das (herkömmliche) Risiko der Studie für Bias (also die interne Validität) beurteilt. Zusätzlich werden weitere Aspekte beurteilt. Ein Evidenzkörper, mit zum Beispiel einer Anzahl adäquat geplanter und gut durchgeführter Studien, mag mit einem geringen RoB behaftet sein. Trotzdem kann das Vertrauen in die Effektschätzer durch andere Faktoren wie unzureichende Präzision (weites Konfidenzintervall), Inkonsistenz (Vorliegen widersprüchlicher Studienergebnisse), Indirektheit (Diskrepanz zwischen interessierender Patientenpopulation oder interessierender Intervention oder interessierenden Endpunkten und der in der Studie untersuchten Population, Intervention oder Endpunkten) und publication bias (das Nicht-Publizieren von Studien, z.B. mit negativen Ergebnissen), beeinträchtigt sein. Alle diese Faktoren fließen in die Bewertung nach GRADE ein.

3. Verschiedene Biasformen und deren Auswirkung

In der Literatur werden unzählige Formen von Bias, die innerhalb klinischer Studien auftreten können, beschrieben. Im Folgenden wird sich ausschließlich auf die Biasformen bezogen, die für Leitlinienautoren, insbesondere auf Grund ihrer Häufigkeit, eine besondere Relevanz aufweisen.¹³

3.1 Wichtige Biasformen in randomisierten kontrollierten Studien

Bias vor Behandlungszuteilung

Selection bias (allocation bias; Selektionsbias): Selection bias ist als Verzerrung durch Unterschiede in den Patientencharakteristika zwischen den Studiengruppen definiert. Wenn das Studienpersonal, insbesondere das Personal, das die Patienten zuteilt, vorab weiß oder vorhersagen kann, welcher Gruppe der nächste Patient zugeteilt wird, können zum Beispiel bestimmte Patienten verzögert in die Studie eingeschlossen werden, um in die andere Gruppe zu kommen, oder ganz von der Teilnahme an der Studie abgehalten werden.

Um selection bias zu vermeiden, werden die Patienten den Behandlungsgruppen zufällig zugeteilt (randomisiert). Für eine erfolgreiche Randomisierung ist vor allem die Wahl einer adäquaten Randomisierungssequenz und die Geheimhaltung und Unvorhersehbarkeit der Gruppenzuteilung (verdeckte Gruppenzuteilung, englisch: allocation concealment) ausschlaggebend.

Bias nach Behandlungszuteilung

Performance bias: Performance bias wird durch unterschiedliche Rahmenbedingungen (zusätzlich zur Intervention) und ‚Abweichungen‘ von der eigentlichen Intervention zwischen den Studienarmen hervorgerufen. Dieser Bias entsteht in erster Linie durch mangelnde Verblindung, indem die betreuenden Personen einer Gruppe von Patienten eine bessere medizinische Versorgung zukommen lassen (häufig in Form einer Begleittherapie). Nur eine Verblindung kann Unterschiede in begleitenden Maßnahmen zwischen den zu vergleichenden Gruppen vermeiden.

Detection bias (observer bias): Das Studienpersonal neigt häufig dazu, Beobachtungen bewusst oder unbewusst den eigenen Erwartungen anzupassen. Detection bias kann insbesondere auftreten, wenn Studienpersonal und/oder -teilnehmer bei einer subjektiven Endpunkterhebung bzw. -bewertung über die Gruppenzugehörigkeit informiert sind. Dieser Bias kann ebenfalls durch Verblindung vermieden werden.

Attrition bias: Attrition (Abnahme) entsteht durch systematische Unterschiede in der Anzahl und den Ursachen fehlender Daten zwischen den Behandlungsgruppen. Fehlende Daten bei der Endpunkterhebung können zu einer Verzerrung des Therapieeffektes führen, insbesondere wenn diese nicht „zufällig“ fehlen. Fehlende Daten können dabei sowohl durch ‚Ausschluss‘ oder ‚Attrition‘ hervorgerufen werden: (i) Bei ‚Ausschluss‘ liegen die Daten der Endpunktbewertung des jeweiligen Patienten vor, werden bei der Datenanalyse jedoch nicht mit berücksichtigt (bewusster Ausschluss). (ii) Bei ‚Attrition‘ liegen die Patientendaten der Endpunktbewertung nicht vor. Zum Beispiel, wenn Studienteilnehmer noch während der eigentlichen Studienphase aus der Studie ausscheiden (Drop-Outs), wenn Studienteilnehmer die eigentliche Studienphase zwar abgeschlossen haben und lediglich nicht mehr zu den Nachsorgeterminen erscheinen (Lost to Follow-Up) oder wenn Patienten die erforderlichen Daten in Fragebögen etc. nicht eintragen.

Problematisch ist es vor allem, wenn die Anzahl der fehlenden Daten zwischen den Behandlungsgruppen ungleich verteilt ist. Diese Ungleichverteilung hat häufig Ursachen, die im Zusammenhang mit der Intervention stehen und Einfluss auf den Endpunkt ausüben. Als Beispiele seien hier Studienabbrecher und/oder Gruppenwechsler aufgrund mangelnder Wirksamkeit oder unerwünschter Wirkungen genannt. Hingegen führen fehlende Daten von Patienten, die aus persönlichen Gründen, wie zum Beispiel einem Umzug, nicht weiter an der Studie teilnehmen können, in der Regel nicht zu einem erhöhten RoB.

Eine Möglichkeit, attrition bias entgegenzuwirken, ist durch eine angemessene statistische Auswertung der Daten gegeben. Standard ist eine sogenannte *Intention-to-Treat-Auswertung (ITT-Analyse)*, bei der auch Studienabbrecher und/oder Gruppenwechsler mit berücksichtigt werden.¹⁴

Reporting bias (selektives Berichten von Endpunkten): Reporting bias wird durch systematische Unterschiede zwischen selektiv berichteten und nicht-berichteten Ergebnissen einer Studie hervorgerufen. Häufig werden in einer Studie nachträglich ausgewählte positive und signifikante Studienendpunkte (insbesondere selektive Meßzeitpunkte und/oder auch Subgruppenergebnisse) unter Vernachlässigung von negativen Endpunktergebnissen berichtet. Reporting bias kann deshalb zu einer Überschätzung der Wirksamkeit von Interventionen führen. Reporting bias wird häufig auch als „within-study publication bias“ bezeichnet und trägt zusammen mit der Tatsache, dass publizierte Studien eine verzerrte Auswahl aller durchgeführten Studien darstellen (publication bias oder auch dissemination bias)^{15 16}, am wesentlichsten zur Ergebnisverzerrung in Metaanalysen bei.¹⁷

✓ Verblindung während der Behandlung und/oder der Endpunkterhebung und verdeckte Gruppenzuteilung während der Randomisierung beinhalten unterschiedliche Aspekte und verhindern verschiedene Biasformen. Die verdeckte Zuteilung zur Behandlungsgruppe erfolgt vor der Behandlungsphase, um selection bias zu vermeiden (d.h. das Studienpersonal, das den Patienten in die jeweilige Behandlungsgruppe einteilt, muss bezüglich der Gruppenzuteilung verblindet sein). Unabhängig davon erfolgt die Verblindung von Studienpersonal und/oder Patienten bezüglich der unterschiedlichen Behandlungen, mit dem Ziel performance bias (während der Behandlung) und detection bias (bei der Endpunkterhebung) entgegenzuwirken.

✓ Die ITT-Analyse wird dabei als primäre, konservative Auswertung empfohlen, um den Therapieeffekt einer Studie zu bewerten.¹⁸ Diese Analyse erfordert, dass Patienten der Gruppe zugeordnet werden, in die sie ursprünglich randomisiert wurden – unabhängig davon, welche Intervention sie tatsächlich erhielten. Außerdem fordert die ITT-Analyse im strengen Sinne, dass Endpunktdaten für alle Patienten berücksichtigt werden. Eine Möglichkeit, um fehlende Daten zu ersetzen, ist dabei die Anwendung von Imputationsverfahren, d.h. das Ersetzen fehlender Daten in der Analyse.^{19 20} Es existiert eine Vielzahl von Imputationsverfahren. Man kann dabei „grob“ zwischen der singulären und der multiplen Imputation unterscheiden. Bei der ersteren wird jeder fehlende Wert durch jeweils einen bestimmten Schätzwert ersetzt, während bei der anspruchsvolleren multiplen Imputation für jeden fehlenden Wert mehrere Werte geschätzt werden. Erfolgt keine Imputation, dann handelt es sich um eine ‚modifizierte‘ ITT-Analyse, da nicht für alle Patienten Endpunktdaten vorliegen oder berechnet wurden. Eine inadäquate alleinige Anwendung des ITT-Prinzips in Nicht-Unterlegenheits- oder Äquivalenzstudien kann fälschlicherweise eine Äquivalenz vortäuschen.²¹ Somit ist insbesondere bei der Bewertung von Äquivalenzstudien Vorsicht geboten.

3.2 Wichtige Biasformen in nicht-randomisierten Studien

Bias vor Behandlungszuteilung

Selection bias: Da bei nicht-randomisierten Studien keine zufällige Zuteilung der Teilnehmer auf die Interventions- und Kontrollgruppe erfolgt, sind systematische Unterschiede zwischen den Gruppen, ausgelöst durch messbare bekannte und/oder unbekannte Störfaktoren, wahrscheinlicher als bei randomisierten Studien. Diese Störfaktoren beinhalten sowohl Patientencharakteristika als auch äußere Faktoren und können sowohl mit der Intervention (oder Exposition) als auch dem Endpunkt in Beziehung stehen. Wenn zum Beispiel die Patienten einer Behandlungsgruppe, in der Therapie A eingesetzt wird, jünger sind als die Patienten der Gruppe, in der Therapie B eingesetzt wird, so ist schwer zu ermitteln, inwieweit die Ergebnisse der Gruppe A nun auf die Therapie oder das geringere Alter zurückzuführen sind. Das Patientenalter würde somit eine messbare Störgröße darstellen. Bekannte und in der Studie erhobene Störgrößen können durch die Anwendung geeigneter statistischer Verfahren berücksichtigt werden (Adjustierung für Störgrößen zum Beispiel durch Stratifizierung oder Regressionsmodelle). Zur Verringerung des Risikos für das Auftreten von Störgrößen ist eine konsekutive Rekrutierung, d.h. aller im Einschlusszeitraum die Einschlusskriterien erfüllenden Patienten/Probanden wichtig. Zudem ermöglicht eine genaue Erhebung und Berichterstattung der relevanten Patientencharakteristika und (Risiko-)faktoren eine Einschätzung des selection bias.

Sonderformen des selection bias: Neben messbaren kann in nicht-randomisierten Studien auch durch den Ausschluss von nicht-messbaren Störgrößen selection bias verursacht werden. Als Beispiel sei hier der **inception bias (lead-time bias; Vorlaufzeitbias)** genannt. Dieser Bias tritt auf, wenn insbesondere bei Patienten in der Interventionsgruppe der Beobachtungszeitpunkt (gleichzusetzen mit Studienbeginn) nicht mit dem Behandlungsbeginn zusammenfällt. Dies kann der Fall sein, wenn Patienten in eine Studie eingeschlossen werden, die die zu untersuchende Intervention oder medizinische Maßnahme schon über einen längeren Zeitraum einnehmen beziehungsweise erhalten. In solchen Fällen liegt der Behandlungs- vor dem Studienbeginn und Ereignisse, die vor Studienbeginn auftraten werden dadurch nicht erfasst. Es ist offensichtlich, dass ein systematischer Ausschluss der initialen Beobachtungszeit zu einer Über- oder Unterschätzung des Nutzens beziehungsweise Schadens einer Intervention führen kann.

Eine weitere Sonderform des selection bias stellt der **immortal-time bias (time-dependent bias)** dar. Dieser Bias tritt auf, wenn bei der Endpunktauswertung Ereignisse, die in einer bestimmten Zeit auftreten (Wartezeit), nicht erfasst werden. Das kann insbesondere der Fall sein, wenn Patienten eine Intervention nur dann erhalten, wenn sie nach einem bestimmten vordefinierten Zeitpunkt (Studienbeginn, zum Beispiel Krankenhausentlassung), eine gewisse Zeit überleben. Die Kontrollen dieser Studie erhalten die Intervention hingegen unmittelbar nach Studienbeginn (Krankenhausentlassung). In einer solchen Studie würden diejenigen Patienten der Interventionsgruppe bei der Endpunktauswertung nicht berücksichtigt, die in der Zeit zwischen Krankenhausentlassung und Interventionsbeginn (Wartezeit) verstorben sind, was zu einer niedrigeren Mortalitätsrate in der Interventions- im Vergleich zur Kontrollgruppe führen würde. Ein anderes Beispiel sind Transplantationsstudien mit dem Endpunkt Überlebenszeit. Wird die Überlebenszeit von Patienten mit Transplantat mit der Überlebenszeit von Patienten verglichen, die kein Transplantat erhalten (Kontrollen), so muss berücksichtigt werden, dass transplantierte Patienten bereits zuvor lange genug überlebt haben müssen, um überhaupt ein Transplantat zu erhalten. Um zu dieser Gruppe zu zählen, dürfen sie also während der Wartezeit nicht verstorben sein.²²

Im Gegensatz zu randomisierten Studien, in denen der Beobachtungsbeginn in der Regel den Zeitpunkt der Randomisierung darstellt (auch wenn die Intervention erst später erfolgt [siehe Beispiel immortal-time bias] oder schon seit längerer Zeit eingenommen wurde [siehe Beispiel inception bias]), kann bei nicht-randomisierten Studien diese Biasform nur ausgeschlossen werden, wenn alle Ereignisse ab dem Zeitpunkt erfasst werden, zu dem die Patienten in die jeweilige Studie aufgenommen wurden beziehungsweise die entsprechende Intervention eingenommen haben.

Bias durch die Erfassung der Intervention

Information bias (Informationsbias): Dieser Bias entsteht durch eine fehlerhafte oder ungenaue Erhebung individueller Faktoren, wie zum Beispiel Risikofaktoren. Insbesondere kann sich beim information bias ein bestimmter Endpunkt auf eine verzerrte Expositionserfassung auswirken. Zum Beispiel erinnern sich Mütter von Kindern mit Fehlbildungen besser an potentielle Risikofaktoren während der Schwangerschaft als Mütter, die ein gesundes Kind zur Welt bringen.²³ Dies ist insbesondere bei Fall-Kontrollstudien der Fall. Erkrankte denken gründlicher über mögliche Expositionen nach als Gesunde (Nicht-Betroffene). Auch damit kann ein Zusammenhang zwischen Exposition und Erkrankung vorgetäuscht werden. Dieser systematische Fehler wird auch als **recall bias (Erinnerungsbias)** bezeichnet.

Auch der **interview bias** fällt in die Kategorie des information bias. Ein Interviewer begegnet häufig den befragten Fällen mit mehr Empathie als den Kontrollen. Dadurch bekommt er detailliertere Informationen von den Fällen. Dieser systematische Fehler kann vor allem bei der Erhebung unerwünschter Ereignissen auftreten.

Bias nach Behandlungszuteilung

Die häufigsten Biasformen nach Behandlungszuteilungen stellen **performance bias, detection bias, attrition bias** und **reporting bias** dar, und sind somit identisch zu den Biasformen der randomisierten Studien nach Behandlungszuteilung (*siehe Abschnitt 3.1: Wichtige Biasformen in randomisierten kontrollierten Studien*).

3.3 Auswirkung von Bias auf die Ergebnisse von klinischen Studien

Eine Zusammenfassung von Daten aus sieben meta-epidemiologischen Studien zeigte eine Überschätzung des Behandlungseffekts bei inadäquater oder fehlender Randomisierung um im Durchschnitt 11% (95%-KI 4% bis 18%).²⁴ Durch eine fehlende oder inadäquate verdeckte Gruppenzuteilung wurden Behandlungseffekte um 7% (95%-KI 1% bis 13%), und im Fall einer fehlenden oder inadäquaten doppelten Verblindung um 13% (95%-KI 4% bis 21%) überschätzt.²⁴

Das Ausmaß des Effekts einer fehlenden oder inadäquaten Verblindung hängt jedoch wesentlich von den Untersuchungsparametern und der Intervention ab. Im Fall des Endpunktes ‚Mortalität‘ ist eine fehlende Verblindung mit einem deutlich geringeren RoB verbunden als bei einem subjektiven Endpunkt, der Interpretationsspielraum lässt. Bei chirurgischen Interventionen ist eine Verblindung von Chirurg und Patient seltener realisierbar oder oft auch aus ethischen Gründen nicht zu rechtfertigen.^{4 25} Bei bestimmten Endpunkten, wie zum Beispiel der intraoperativen Blutungsmenge, kann auch unabhängiges Studienpersonal zur Endpunkterhebung eingesetzt werden.²⁶ Das Ergebnis einer meta-epidemiologischen Studie war, dass eine inadäquate oder fehlende Verblindung zu einer Überschätzung der Studienergebnisse von 25% (95%-KI 7% bis 39%) bei subjektiv erhobenen Endpunkten führt.²⁷

Untersuchungen zum Ausmaß des attrition bias sind bis dato wenig aussagekräftig und generell schwierig. Fehlen nämlich in Publikationen die Daten von Probanden und Informationen darüber, warum Teilnehmer eine Studie abgebrochen haben, oder sind die Angaben nicht nach Behandlungsgruppen differenziert berichtet, ist es nicht sicher möglich, das Ausmaß dieses Bias abzuschätzen.

Ob die industrielle Finanzierung von Studien als eigenständiges Kriterium berücksichtigt werden sollte, ist nicht abschließend geklärt. Autoren eines Cochrane-Reviews konnten zeigen, dass eine finanzielle Unterstützung durch Hersteller zu deutlich positiveren Ergebnissen hinsichtlich Wirksamkeit und Nebenwirkungen einer Intervention führt.²⁸ Da jedoch davon auszugehen ist, dass die Überschätzung der Effekte im Wesentlichen durch zum Beispiel selektives Berichten von Endpunkten oder andere der oben genannten Biasformen zustande kommt, ist „Sponsoring“ aktuell kein eigenständiger Bestandteil der Biasbewertung nach den Cochrane Methoden.^{29 30}

4. Bewertungsinstrumente

Zur standardisierten Bewertung des RoB steht eine Vielzahl von Instrumenten zur Verfügung, die sich im Wesentlichen drei Kategorien zuordnen lassen: Skalen, Checklisten und Komponentensysteme.³¹

Skalen

Anhand einer Skala werden verschiedene interne und externe Validitätsaspekte einer Studie mit Hilfe eines „Scores“ (Punktezahl) numerisch bewertet. Für die Gesamtbewertung der Studie werden die einzelnen Punkte addiert. Bewertungsverfahren auf Skalenbasis werden häufig wegen mangelnder Evidenz im Hinblick auf die Gewichtung der einzelnen Bewertungsaspekte durch die empirische Forschung nicht gestützt.^{32 33} Zudem bewerten Skalen eher die Berichtsqualität als das Verzerrungspotential, eine nicht im Detail dargestellte Methodik muss allerdings nicht unbedingt bedeuten, dass die Studie mangelhaft durchgeführt wurde (*siehe Abschnitt 2.4: Biasrisiko versus Berichtsqualität*). Einige Skalen, wie die bekannte Jadad Skala (Oxford Skala) für randomisierte kontrollierte Studien³⁴, berücksichtigen zudem nicht den wichtigen RoB Aspekt der verdeckten Gruppenzuteilung (allocation concealment).³³ Es wurden außerdem substantielle Unterschiede in den Ergebnissen von Metaanalysen, die auf die Verwendung unterschiedlicher Skalen bei randomisierten kontrollierten Studien zurückzuführen waren, berichtet.³³

Zur Bewertung von nicht-randomisierten Studien (insbesondere von Fall-Kontrollstudien und Kohortenstudien) wird häufig die **Newcastle Ottawa Skala (NOS)** herangezogen.³⁵ Die Kritikpunkte dieser Skala gleichen denen der Skalen für randomisierte kontrollierte Studien (siehe oben).

Checklisten

Checklisten, wie zum Beispiel die des **Scottish Intercollegiate Guidelines Networks (SIGN)**³⁶, sind in der Regel umfangreich und bewerten meist neben der internen auch die externe Validität einer Studie (zum Beispiel die Angemessenheit der Ein- und Ausschlusskriterien oder ob eine Fallzahlplanung *a priori* durchgeführt wurde). Da die externe Validität keine Aussage darüber zulässt, inwieweit man dem gemessenen Effekt vertrauen kann, muss sie von der RoB Bewertung unterschieden werden (*siehe Abschnitt 2.1: Biasrisiko versus externe Validität einer Studie*).³⁷

Komponentensysteme

In Komponentensystemen, wie zum Beispiel im System der RoB Bewertung der Cochrane Collaboration für randomisierte kontrollierte Studien³⁸ oder auch für nicht-randomisierte kontrollierte Studien (ACROBAT-NRSI; **A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions**³⁹), wird das Verzerrungspotential für verschiedene Komponenten (Domänen) separat dargestellt. Als weiteres Komponentensystem für nicht-randomisierte kontrollierte Studien sei an dieser Stelle noch RoBANS (**Risk of Bias Assessment tool for Non-randomized Studies**) genannt, welches als alternatives Instrument zum ACROBAT-NRSI der Cochrane Collaboration betrachtet wird.⁴⁰ Einen wesentlichen Vorteil der Komponentensysteme stellt die Möglichkeit dar, dass das Verzerrungspotential sowohl auf Studien- als auch auf Endpunktebene getrennt bewertet werden kann (*siehe Abschnitt 5.1: Randomisierte kontrollierte Studien und Abschnitt 5.2: Nicht-randomisierte Studien*).³

✓ Allen Instrumenten ist gemein, dass sie keine exakte Messung, sondern eine Einschätzung des Verzerrungsrisikos in klinischen Studien darstellen und dadurch immer eine subjektive Komponente beinhalten.

5. Bewertung des Biasrisikos nach den Cochrane Methoden

5.1 Randomisierte kontrollierte Studien

Randomisierte kontrollierte Studien 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.

Die RoB Bewertung von randomisierten kontrollierten Studien nach den Methoden der Cochrane Collaboration basiert auf einem Komponentensystem, dessen Kriterien auf theoretischen und empirischen Forschungsergebnissen basieren.⁴¹ Das Cochrane RoB Instrument setzt sich dabei aus sieben Domänen zusammen (*siehe Abschnitt 5.1.1: Domänen der Bewertung*).³⁸ Für jede Studie wird dabei jede einzelne Domäne entweder mit einem ‚geringen RoB‘ (low RoB), ‚hohen RoB‘ (high RoB) oder ‚unklaren RoB‘ (unclear RoB) bewertet. Die Einstufung ‚unklares RoB‘ erfolgt, wenn in einem Studienbericht relevante Einzelheiten zur Einschätzung des RoB nicht ausreichend dargestellt werden.⁴ Transparent werden die Bewertungen dadurch, dass jede Einstufung durch ein Zitat aus dem Originaltext oder einen Kommentar zu belegen ist.

Im Gegensatz zu den Bewertungsinstrumenten, die auf Skalen oder Checklisten basieren, liegen für das Cochrane RoB Instrument empirische Studien vor, die die Reliabilität, d.h. die interindividuelle Übereinstimmung der einzelnen RoB Bewertungen, evaluierten.^{42 43} Diese Untersuchungen zeigten jedoch, sowohl auf Studien- als auch auf Metaanalysen-Ebene, eine nur mäßig gute Interrater-Reliabilität, ausgedrückt mittels Cohens Kappa (κ). Interrater-Reliabilität für einzelne Domänen der Cochrane RoB Bewertung (auf Studienebene): Median $\kappa=0,19$; Bandbreite $\kappa=-0,04$ (andere Ursachen für Bias) bis $\kappa=0,62$ (Generierung der Randomisierungssequenz); Interrater-Reliabilität für die gesamte RoB Bewertung zwischen Cochrane Review-Autoren und ‚externen verblindeten‘ Autoren (auf Ebene von Metaanalysen): $\kappa=0,02$; 95%-KI $-0,06$ bis $+0,06$.⁴² Trotz dieser interindividuellen Diskrepanzen ermöglicht das Cochrane Komponentensystem – als eines der wenigen – eine Transparenz im Hinblick auf die Einstufung des RoB durch ein Zitat aus der zu bewertenden Studie oder einen Kommentar.

5.1.1. Domänen der Bewertung

1) Generierung der Randomisierungssequenz

Die Randomisierungssequenz wird idealerweise vor der Zuteilung der Studienteilnehmer in die unterschiedlichen Gruppen erstellt. Eine zufällige Zuteilung der Teilnehmer zur Interventions- und Kontrollgruppe verhindert systematische Unterschiede zwischen den Gruppen.

Kriterien für ein ‚geringes RoB‘:

- ✓ Tabelle mit Zufallszahlen
- ✓ Computergenerierte Zufallszahlen
- ✓ Stratifizierte oder Block-Randomisierung
- ✓ Einfach erstellte Sequenzen: Münzwurf, Mischen von Karten, würfeln, auslosen
- ✓ Minimierung^{44 45}

Kriterien für ein ‚hohes RoB‘:

- ✗ Quasi-Randomisierung: zum Beispiel nach Geburtsdatum, Wochentag, Aktennummer
- ✗ Nicht-zufällige Sequenz: zum Beispiel durch Arzt-Zuteilung, Patienten-Präferenz, Testresultat oder Verfügbarkeit der Intervention

Kriterien für ein ‚unklares RoB‘:

- ? Unzureichende Informationen für eine Beurteilung ‚geringes/hohes RoB‘

⇒ Selection bias, verursacht durch inadäquate Generierung einer Randomisierungssequenz!

2) Geheimhaltung und Unvorhersehbarkeit der Gruppenzuteilung

Durch die Geheimhaltung der Gruppenzuteilung (Behandlungsfolge) wird sichergestellt, dass die Randomisierungssequenz angemessen implementiert wird. Dadurch wird einer selektiven Auswahl der Studienteilnehmer in die einzelnen Behandlungsgruppen entgegengewirkt.

Kriterien für ein *geringes RoB*:

- ✓ Externe Zuteilung (Telefon-, Fax- oder durch Apotheken kontrollierte Randomisierung)
- ✓ Fortlaufend nummerierte, versiegelte, blickdichte Umschläge
- ✓ Fortlaufend nummerierte, identische Arzneimittelverpackungen

Kriterien für ein *hohes RoB*:

- ✗ Umschläge oder Verpackungen ohne geeignete Sicherheitsmaßnahmen
- ✗ Verwendung eines offenen Randomisierungsplans
- ✗ Alternierende oder rotierende Gruppenzuteilung
- ✗ Vorhersehbare Reihenfolge (wie zum Beispiel Geburtsdatum, Aktennummer)

Kriterien für ein *unklares RoB*:

- ? Unzureichende Informationen für eine Beurteilung *geringes/hohes RoB*; dies ist dann der Fall wenn die Methode der Geheimhaltung nicht oder nur ungenau beschrieben wurde, wie zum Beispiel bei der Verwendung von Umschlägen bei denen unklar bleibt, ob diese sequentiell nummeriert, blickdicht und versiegelt waren

⇒ **Selection bias, verursacht durch inadäquate Geheimhaltung der Gruppenzuteilung!**

3) Verblindung von Studienpersonal/-teilnehmern während der Behandlung

Verfügen Studienpersonal /-teilnehmer über Informationen zur Gruppenzugehörigkeit, kann dieses Wissen die Ergebnisse der Studie beeinflussen. Ist der behandelnde Arzt einer Studie nicht verblindet, besteht die Gefahr, dass Patienten der Interventionsgruppe eine intensivere Behandlung/Betreuung erfahren als Patienten der Kontrollgruppe. Als Folge kann die Wirksamkeit überschätzt werden. Auch Studienteilnehmer, die wissen, welcher Gruppe sie zugehören, können die Ergebnisse durch ein subjektiv gesteigertes positiveres/negativeres Empfinden verzerren.

Kriterien für ein *geringes RoB*:

- ✓ Gesicherte Verblindung und geringe Wahrscheinlichkeit für durchbrochene Verblindung
- ✓ Keine, unvollständige oder durchbrochene Verblindung; Einfluss der fehlenden Verblindung auf Endpunkt unwahrscheinlich (zum Beispiel bei objektiven Endpunkten)

Kriterien für ein *hohes RoB*:

- ✗ Keine, unvollständige oder durchbrochene Verblindung; Einfluss der fehlenden Verblindung auf Endpunkt wahrscheinlich (zum Beispiel bei subjektiven Endpunkten)
- ✗ Versuch der Verblindung von Studienteilnehmern/-personal, aber hohe Wahrscheinlichkeit für durchbrochene Verblindung; Einfluss der fehlenden Verblindung auf Endpunkt wahrscheinlich (zum Beispiel bei subjektiven Endpunkten)

Kriterien für ein *unklares RoB*:

- ? Unzureichende Informationen für eine Beurteilung *geringes/hohes RoB*

⇒ **Performance bias, verursacht durch Kenntnis der Gruppenzugehörigkeit!**

4) Verblindung bei der Endpunkterhebung/-bewertung

Sind sowohl Studienpersonal als auch -teilnehmer bei der Endpunktbeurteilung verblindet, wird einer unterschiedlichen Endpunktinterpretation, beeinflusst durch die Kenntnis der Interventions-beziehungsweise Kontrollbehandlung, entgegengewirkt. Vor allem grenzwertige Befunde oder minimale Veränderungen bei subjektiven Endpunkten können somit nicht überinterpretiert werden. Auch wenn – wie häufig bei nicht-medikamentösen Interventionen – eine Verblindung in der Interventionsphase nicht möglich ist, kann und sollte die Endpunkterhebung durchaus verblindet erfolgen. Das Ausmaß des Verzerrungsrisikos ist dabei vor allem abhängig vom Studienendpunkt: für objektive Endpunkte, wie zum Beispiel Mortalität, ist in der Regel eine fehlende Verblindung wesentlich unproblematischer als für subjektive Endpunkte, wie zum Beispiel Schmerz oder Lebensqualität.

Kriterien für ein ‚geringes RoB‘:

- ✓ Gesicherte Verblindung bei der Endpunkterhebung und geringe Wahrscheinlichkeit für Entblindung
- ✓ Keine Verblindung bei der Endpunkterhebung; Einfluss der fehlenden Verblindung auf Endpunkt unwahrscheinlich (zum Beispiel bei objektiven Endpunkten)

Kriterien für ein ‚hohes RoB‘:

- ✗ Keine Verblindung bei der Endpunkterhebung und Einfluss der fehlenden Verblindung auf Endpunkt wahrscheinlich (zum Beispiel bei subjektiven Endpunkten)
- ✗ Versuch der Verblindung bei Endpunkterhebung, aber hohe Wahrscheinlichkeit für durchbrochene Verblindung; Einfluss der fehlenden Verblindung auf Endpunkt wahrscheinlich (zum Beispiel bei subjektiven Endpunkten)

Kriterien für ein ‚unklares RoB‘:

- ? Unzureichende Informationen für eine Beurteilung ‚geringes/hohes RoB‘

⇒ **Detection bias, verursacht durch Kenntnis der Gruppenzugehörigkeit bei Endpunkterhebung!**

5) Fehlende Daten bei der Endpunkterhebung

Fehlende Daten bei der Endpunkterhebung, insbesondere wenn diese nicht „zufällig“ fehlen, können zu einer Verzerrung des Therapieeffektes führen (*siehe Abschnitt 3.1: Wichtige Biasformen in randomisierten kontrollierten Studien/attrition bias*).

Kriterien für ein ‚geringes RoB‘:

- ✓ Keine fehlenden Daten bei der Endpunkterhebung
- ✓ Die Ursachen für fehlende Daten stehen wahrscheinlich nicht im Zusammenhang mit der Intervention/Exposition oder den Ergebnissen; so ist zum Beispiel das Zensieren von Überlebensdaten in der Regel unproblematisch
- ✓ Die Anzahl fehlender Daten ist zwischen den Behandlungsgruppen gleich verteilt, außerdem sind die Ursachen für die fehlenden Daten zwischen den Gruppen vergleichbar
- ✓ Der Einfluss der fehlenden Daten auf das Ereignisrisiko, die Odds-Ratio oder die Risikodifferenz (bei dichotomen Endpunkten) beziehungsweise auf die Effektgröße (bei kontinuierlichen Endpunkten) ist statistisch und/oder klinisch nicht relevant
- ✓ Fehlende Daten wurden mit Hilfe geeigneter statistischer Methoden imputiert (ITT-Analyse)

Kriterien für ein ‚hohes RoB‘:

- ✗ Die Ursache für die fehlenden Daten steht wahrscheinlich in Zusammenhang mit den Interventionen oder Ergebnissen; entweder durch ein Ungleichgewicht in der Anzahl oder der Gründe der fehlenden Daten zwischen den Gruppen
- ✗ Der Einfluss der fehlenden Daten auf das Ereignisrisiko, die Odds-Ratio oder die Risikodifferenz (bei dichotomen Endpunkten) beziehungsweise auf die Effektgröße (bei kontinuierlichen Endpunkten) ist statistisch und/oder klinisch relevant
- ✗ Durchführung einer As-Treated-Analyse (AT-Analyse) bei einem wesentlichen (einflussreichen) Anteil von fehlenden Daten
- ✗ Fehlende Daten wurden mit Hilfe von ungeeigneten statistischen Methoden imputiert

Kriterien für ein ‚unklares RoB‘:

- ? Unzureichende Informationen für eine Beurteilung ‚geringes/hohes RoB‘

⇒ **Attrition bias, verursacht durch die Anzahl, Ursache oder den Umgang mit fehlenden Daten!**

6) Selektives Berichten von Endpunkten

Wenn Autoren oder Sponsoren von Studien ausgewählte positive Endpunkte berichten, wird dies ‚selektives Berichten‘ genannt. Selektives Berichten kann zu einer Überschätzung des Nutzens einer Intervention im Rahmen der Studie und sekundär des endpunkt-bezogenen Behandlungseffekts im Kontext einer Metaanalyse führen.¹⁷ Das Cochrane Handbuch schlägt vor, dass eine definitive Aussage für das Nicht-Vorliegen von selektivem Berichten (ergebnisunabhängige Berichterstattung) Zugang zum Protokoll erfordert, welches vor Studienbeginn entwickelt worden sein sollte.³⁸

Kriterien für ein ‚geringes RoB‘:

- ✓ Das Studienprotokoll ist verfügbar und alle vorab festgelegten Endpunkte, die für die Leitlinie oder systematische Übersichtsarbeit von Bedeutung sind, wurden in der Publikation berichtet
- ✓ Das Studienprotokoll ist nicht verfügbar, aber es ist eindeutig, dass die Publikation alle zu erwartenden Endpunkte enthält (*eine überzeugende Publikation dieser Art kommt selten vor*)

Kriterien für ein ‚hohes RoB‘:

- ✗ Nicht alle der vorab festgelegten primären Endpunkte wurden berichtet
- ✗ Einer/mehrere primäre Endpunkte wurden mit nicht vorab festgelegten Analysemethoden oder mit Teilmengen von Daten, wie Subgruppen oder -skalen, angegeben
- ✗ Einer/mehrere (primäre) Endpunkte wurden nicht vorab festgelegt
- ✗ Einer/mehrere Endpunkte, die für das Review von Bedeutung sind, wurden nicht vollständig berichtet und können daher nicht verwendet werden
- ✗ Ergebnisse eines für die Fragestellung ausschlaggebenden Endpunkts wurden nicht berichtet

Kriterien für ein ‚unklares RoB‘:

- ? Unzureichende Informationen für eine Beurteilung ‚geringes/hohes RoB‘ (es ist wahrscheinlich, dass ein Großteil der Publikationen in diese Kategorie fällt)

⇒ **Reporting bias, verursacht durch selektives Berichten von Endpunkten!**

7) Andere Ursachen für Bias

Neben den oben genannten Verzerrungspotentialen kann Bias auch andere Ursachen haben.

Kriterien für ein *geringes RoB*:

- ✓ Die Studie scheint weitgehend frei von anderen Ursachen für Bias

Kriterien für ein *hohes RoB*:

- ✗ Studientyp-spezifische Probleme (zum Beispiel Carry-Over-Effekt in Cross-Over-Studien, nicht-repräsentative Rekrutierung in Cluster-randomisierten Studien)
- ✗ Unterschiede in den Baseline-Charakteristika der eingeschlossenen Patienten (nicht durch Randomisierung erklärbar wie zum Beispiel Alter, Krankheitsstadium oder Schweregrad)
- ✗ Offensichtliches betrügerisches Verhalten der Autoren

Kriterien für ein *unklares RoB*:

- ? Unzureichende Informationen für andere Ursachen für Bias
- ? Rationale/Evidenz nicht ausreichend um ein potentiell erkanntes Problem als Bias zu deklarieren

⇒ **Bias, verursacht durch Probleme, die noch nicht an anderer Stelle erfasst wurden!**

5.1.2. Dokumentation und Berichten des Verzerrungspotenzials

Die Cochrane Collaboration empfiehlt die Bewertungen der einzelnen Domänen für jede Studie in einer ‚**RoB Tabelle**‘ darzustellen. Transparent werden die Bewertungen dadurch, dass die Entscheidung für jede Domäne mit einem Zitat aus dem Artikel begründet wird. Werden wichtige Aspekte in der Publikation nicht berichtet, wird dies ausdrücklich vermerkt. In Tabelle 2 ist eine beispielhafte RoB Bewertung für eine randomisierte kontrollierte Studie in einer von der Cochrane Collaboration empfohlenen ‚**RoB Tabelle**‘ dargestellt. Die RoB Bewertung kann für verschiedene in der Studie berichtete Endpunkte unterschiedlich ausfallen. Es wird deshalb empfohlen, dass die Bewertung des Verzerrungspotenzials für alle in der Studie berichteten Endpunkte für die Aspekte ‚Verblindung bei der Endpunkterhebung/-bewertung‘ und ‚fehlende Daten bei der Endpunkterhebung‘ separat durchgeführt wird.

Zusätzlich können graphische Darstellungen mittels einer ‚**RoB Zusammenfassung**‘ (Abb. 1) und einem ‚**RoB Graphen**‘ (Abb. 2) einen Überblick über das Verzerrungspotential der eingeschlossenen Studien verschaffen. Innerhalb einer ‚**RoB Zusammenfassung**‘ wird dabei die Bewertung für jede einzelne Domäne und Studie tabellarisch dargestellt. Der ‚**RoB Graph**‘ gibt hingegen einen Überblick über das gesamte Verzerrungspotential über alle Studien hinweg für ausgewählte Endpunkte. Sowohl die ‚**RoB Zusammenfassung**‘ als auch der ‚**RoB Graph**‘ können mit Hilfe der frei zugänglichen Software Review Manager (RevMan; <http://ims.cochrane.org/revman>) erstellt werden. Die mit einer Verkehrsampel assoziierte Darstellung (grün-gelb-rot) sorgt dabei für eine globale Verständlichkeit. Weiterführende Angaben zur Darstellung des Verzerrungspotenzials sind auch dem Cochrane Handbook for Systematic Reviews of Interventions in den Abschnitten 8.6 bis 8.17 zu entnehmen.³⁸

Da die tabellarische und graphische RoB Darstellung nicht die damit (möglicherweise) verbundene Ergebnisverzerrung abbildet, ist es wichtig, die RoB Bewertung auch narrativ in der Systematischen Übersichtsarbeit beziehungsweise Leitlinie zu beschreiben und zu diskutieren.

Tabelle 2. Beispiel für eine Cochrane RoB Tabelle für eine einzelne randomisierte fiktive Studie mit verschiedenen Endpunkten

Bias Domäne	Bewertung	Begründung der Bewertung
Generierung der Randomisierungssequenz	Geringes RoB	Zitat: "patients were randomly allocated." <i>Kommentar:</i> Höchstwahrscheinlich erfolgt. In früheren Berichten derselben Untersucher wurde die Randomisierungsmethode bereits klar beschrieben.
Geheimhaltung und Unvorhersehbarkeit der Gruppenzuteilung	Hohes RoB	Zitat: "... using a table of random numbers." <i>Kommentar:</i> Sagt nichts zur verdeckten Gruppenzuteilung aus, die Tabelle war höchstwahrscheinlich vom Zuteiler einsehbar => höchstwahrscheinlich nicht erfolgt.
Verblindung von Studienpersonal/-teilnehmern während der Behandlung	Geringes RoB	Zitat: "double blind, double dummy"; "High and low dose tablets or capsules were indistinguishable in all aspects of their outward appearance. For each drug an identically matched placebo was available (the success of blinding was evaluated by examining the drugs before distribution)." <i>Kommentar:</i> Höchstwahrscheinlich erfolgt.
Verblindung bei der Endpunkterhebung (subjektiver [patientenberichteter] Endpunkt)	Geringes RoB	<i>Kommentar:</i> Höchstwahrscheinlich wurde die oben beschriebene Verblindung auch bei der Endpunkterhebung aufrechterhalten.
Verblindung bei der Endpunkterhebung (objektiver Endpunkt [Mortalität])	Geringes RoB	Informationen zu diesem Endpunkt aus der Patientenakte. Review Autoren sind überzeugt, dass (aufgrund des objektiven Endpunkts) kein Bias vorliegt.
Fehlende Daten bei der Endpunkterhebung (kurzfristiger Endpunkt, 2-6 Wochen)	Hohes RoB	nach 4 Wochen: Interventionsgruppe: 17 Studienabbrecher von 110 Patienten (9 wegen fehlender Wirksamkeit); Kontrollgruppe: 7 Studienabbrecher von 113 Patienten (2 wegen fehlender Wirksamkeit). <i>Kommentar:</i> Ungleiche Verteilung der Studienabbrecher zwischen den Gruppen.
Fehlende Daten bei der Endpunkterhebung	Hohes RoB	12 Wochen: Interventionsgruppe: 31 von 110; Kontrollgruppe: 18 von 113; <i>Kommentar:</i> Die Abbruchgründe und die Anzahl der Studienabbrecher unterscheiden sich zwischen den Gruppen.
Selektives Berichten von Endpunkten	Hohes RoB	Zitat: "Three rating scales for cognition listed in Methods, but only one (with statistically significant results) is reported." <i>Kommentar:</i> Selektives Berichten liegt höchstwahrscheinlich vor.

Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien
Version 1.0

	Generierung der Randomisierungssequenz (selection bias)	Verdeckte Gruppenzuteilung (selection bias)	Verblindung von Studienpersonal/-teilnehmern (performance bias)	Verblindung der Endpunkterhebung/-bewertung (detection bias)	Fehlende Daten bei der Endpunkterhebung/-bewertung (attrition bias)	Selektives Berichten von Endpunkten (reporting bias)	Andere Ursachen für Bias
Schmucker et al. 2012	+	?	?	+	?	?	?
Braun et al. 2003	?	?	?	?	?	?	?
Meerpohl et al. 2014	+	+	-	+	+	+	+
Mahlke et al. 2009	+	+	+	+	+	+	+

Abb. 1. Cochrane RoB Zusammenfassung für fiktive Studien; ⊕geringes RoB; ⊖hohes RoB; ?unklares RoB.

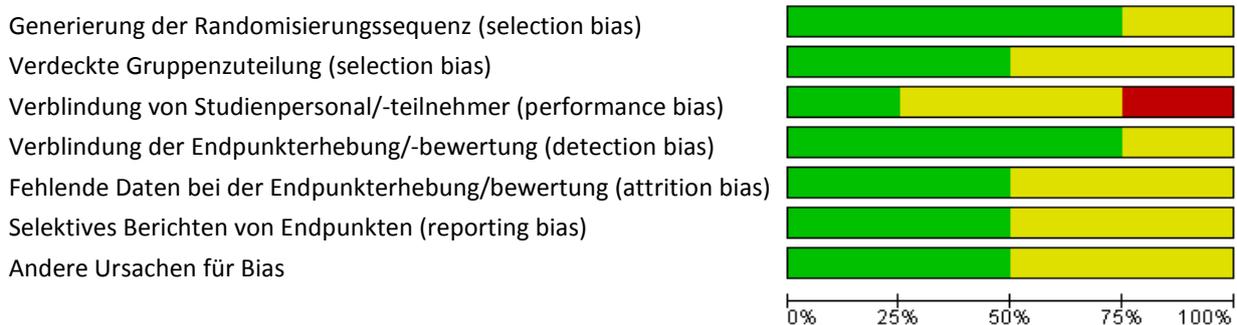


Abb. 2. Cochrane RoB Graph für fiktive Studien (über alle Studien hinweg); geringes RoB; hohes RoB; unklares RoB.

5.2 Nicht-randomisierte Studien

5.2.1 Vergleichende nicht-randomisierte Studien

Bei der Einteilung nicht-randomisierter Studien unterscheidet man grundsätzlich zwischen *vergleichenden* und *nicht-vergleichenden* Studien. Zu den vergleichenden nicht-randomisierten Studien zählen beispielsweise quasi-randomisierte kontrollierte Studien, vergleichende Kohortenstudien, historisch kontrollierte Studien und Fall-Kontrollstudien.

Eine **quasi-randomisierte Studie** setzt dabei im Gegensatz zu einer randomisierten Studie keinen für die Studie erzeugten Zufallsmechanismus ein, der die Gruppenzuteilung bestimmt. Darüber hinaus erfolgt keine Geheimhaltung und Unvorhersehbarkeit der Gruppenzuteilung (*siehe Abschnitt 3.1: Wichtige Biasformen in randomisierten kontrollierten Studien*). Die Quasi-Randomisierung erfolgt zum Beispiel dadurch, dass Studienpatienten abwechselnd den Therapiegruppen zugeteilt werden oder die Zuteilung nach geradem versus ungeradem Geburtsdatum vorgenommen wird.

Mit Hilfe von **Kohortenstudien** lässt sich der Effekt einer Intervention oder Exposition im Zeitverlauf untersuchen. Kohortenstudien sind in diesem Sinne meist prospektiv angelegt. Allerdings gibt es auch retrospektive Kohortenstudien, in denen die Exposition aus der Vergangenheit erfasst wird. Grundsätzlich sind prospektive Studiendesigns retrospektiven Designs vorzuziehen.

Fall-Kontrollstudien stellen unter den vergleichenden nicht-randomisierten Studien eine Besonderheit dar, da die Studienteilnehmer aufgrund ihrer Endpunkte ausgewählt wurden. Diese Studien bieten daher die Möglichkeit, um zum Beispiel die Ursache einer seltenen Erkrankung zu untersuchen und kommen auch bei Screening- und Impffragestellungen zur Anwendung.

Die Bewertung systematischer Fehler muß spezifische Aspekte entsprechend dem jeweiligen Studiendesign berücksichtigen. Die Bewertung des Verzerrungspotentials von vergleichenden, nicht-randomisierten kontrollierten Studien nach den Methoden der Cochrane Collaboration basiert auf einem Komponentensystem (ACROBAT-NRSI), genau wie bei der Bewertung randomisierter Studien. Die Kriterien dieser Komponenten basieren ebenfalls auf theoretischen und empirischen Forschungsergebnissen und wurden von der ‚Cochrane bias methods group‘ und der ‚Cochrane methods group for non-randomised studies‘ entwickelt.³⁹ Das noch in der Pilotierungsphase befindliche Instrument ACROBAT-NRSI setzt sich aus sieben Domänen zusammen (*siehe Abschnitt 5.2.1.1: Domänen der Bewertung*). Da bei vergleichenden nicht-randomisierten Studien grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden kann (sondern eine hohe Wahrscheinlichkeit für systematische Unterschiede zwischen den Gruppen besteht), stellt vor allem die **Kontrolle von möglichen Störgrößen** einen wesentlichen Aspekt der

RoB Bewertung dar. Als Störgröße wird dabei ein Faktor betrachtet, der sowohl mit einer Intervention (oder Exposition) als auch mit dem beobachteten Endpunkt einer Studie assoziiert sein kann. Zu den bekannten und/oder unbekannt prognostisch relevanten Störfaktoren zählen beispielsweise Alter, Geschlecht, Ko-Morbiditäten, Ko-Interventionen, bestimmte Krankheitsindikatoren oder -marker, Adipositas oder auch der sozioökonomische Status. Die Identifizierung von möglichen Störgrößen erfordert Fachwissen und muss je nach Krankheitsspektrum und eingesetzter Intervention spezifisch diskutiert werden.

Die ersten beiden Domänen von ACROBAT-NRSI adressieren Bias, der aus Gruppenunterschieden als Folge der Nicht-Randomisierung (bedingt durch messbare und nicht-messbare Störfaktoren) resultiert (*siehe Abschnitt 3.2: Wichtige Biasformen in nicht-randomisierten Studien/Bias vor Behandlungszuteilung*). Die dritte Domäne adressiert das RoB, welches bei der Erfassung der Intervention auftreten kann (*siehe Abschnitt 3.2: Wichtige Biasformen in nicht-randomisierten Studien/Bias durch die Erfassung der Intervention*). Diese drei Domänen sind spezifisch für die Bewertung nicht-randomisierter Studien. In den restlichen vier Domänen wird das RoB (ab Behandlungsbeginn) adressiert, dass auch in randomisierten Studien auftreten kann.

Transparent unterstützt wird die Bewertung der einzelnen Domänen in ACROBAT-NRSI dadurch, dass unter jeder Domäne vordefinierte Fragen beantwortet und mit einem Studienzitat begründet werden sollen. Für diese Fragen stehen folgende Antwortmöglichkeiten zur Verfügung:

- **Ja (J)**
- **Wahrscheinlich Ja (J?)**
- **Wahrscheinlich Nein (N?)**
- **Nein (N)**
- **Unklar/keine Information (?)**

Die Antwortmöglichkeiten **Ja und Wahrscheinlich Ja** (genauso wie **Nein und Wahrscheinlich Nein**) bringen ähnliche Implikationen mit sich. Abschließend erfolgt die Bewertung des Verzerrungspotentials für jede einzelne Domäne und die Gesamtbewertung (*siehe Abschnitt 5.2.1.2: Dokumentation und Berichten des Verzerrungspotenzials*).

Da sich, wie oben bereits erwähnt, das Instrument ACROBAT-NRSI noch in der Pilotierungsphase befindet, wird im vorliegenden Manual auch die NOS zur Bewertung von nicht-randomisierten Studien dargestellt (*siehe Abschnitt 6: Bewertung des Biasrisikos nach der NOS*).³⁵

5.2.1.1 Domänen der Bewertung (nach Acrobat-NRSI)

1) RoB durch Störfaktoren

Unter dieser Domäne wird zwischen (1) Störfaktoren, die die zu erhaltende Intervention schon vor Behandlungsbeginn prognostizieren (Baseline-Störfaktoren wie Alter, Geschlecht, Komorbidität; selection bias) und (2) Störfaktoren, die nach Behandlungsbeginn auftreten und für den Endpunkt prognostisch sind (zeitabhängige Störfaktoren wie unerwünschte Wirkungen; attrition bias), unterschieden. Zeitabhängige Störfaktoren können in Cross-Over-Studien oder in Studien, in denen Patienten die Behandlungsgruppe wechseln beziehungsweise die Studie abbrechen, auftreten.

- a) Kann ein verzerrtes Ergebnis, verursacht durch jegliche Störfaktoren, ausgeschlossen werden?
J / J? / N? / N
Wenn **J** oder **J?**: es kann von einem geringen RoB ausgegangen werden => es müssen keine weiteren Fragen unter dieser Domäne berücksichtigt werden.

Nur für quasi-randomisierte und Kohortenstudien zutreffend (nicht für Fall-Kontrollstudien):

- b) Wurden die Patienten entsprechend ihrer ursprünglichen Behandlungsgruppe analysiert?
Anmerkung: Wenn eine adäquate ITT-Analyse verwendet wurde, ist attrition bias unproblematisch.
J / J? / N? / N / ?
Wenn **J** oder **J?**: weiter mit d)
Wenn **N?**, **N**, oder **?**: weiter mit c) => zeitabhängige Störfaktoren müssen berücksichtigt werden
- c) Ist es wahrscheinlich, dass die Ursachen für den Behandlungswechsel oder Studienabbruch nicht mit relevanten Störfaktoren in Verbindung stehen?
J / J? / N? / N
Wenn **J** oder **J?**: weiter mit d)
Wenn **N?** oder **N**: weiter mit g) => zeitabhängige Störfaktoren müssen berücksichtigt werden

Wenn keine zeitabhängigen (nur Baseline)-Störfaktoren vorliegen (d.h. kein Hinweis auf attrition bias):

- d) War die Analysemethodik angemessen, um für alle relevanten Störfaktoren zu adjustieren?
J / J? / N? / N / ? Nur wenn **J** oder **J?**: weiter mit e) und f)
- e) Wurden die adjustierten Störfaktoren valide und zuverlässig erfasst?
J / J? / N? / N / ?
- f) Wurde eine Adjustierung für Post-Interventions-Variablen vermieden?
J / J? / N? / N / ?
Nur wenn zeitabhängige Störfaktoren vorliegen (d.h. Hinweis auf attrition bias):
- g) War die Analysemethodik angemessen, um für alle relevanten Störfaktoren zu adjustieren?
J / J? / N? / N / ? Nur wenn **J** oder **J?**:
- h) Wurden die adjustierten Störvariablen valide und zuverlässig erfasst?
J / J? / N? / N / ?

⇒ **Selection bias (und/oder attrition bias), verursacht durch messbare Störfaktoren!**
gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

2) Bias durch Selektion der Teilnehmer in die Studiengruppen

- a) Erfolgte der Einschluss der Teilnehmer unabhängig von Intervention oder Endpunkt (z.B. konsekutiv)?
J / J? / N? / N
- b) Erfolgte der Behandlungsbeginn gleichzeitig mit der Nachbeobachtung?
J / J? / N? / N / ? Nur wenn N oder N? bei a) oder b):
- c) Erfolgte in der Studie eine Adjustierung im Hinblick auf selection bias?
J / J? / N? / N / ?
- d) Wurden die Kontrollen aus derselben Grundgesamtheit ausgewählt, aus der auch die Fälle stammten? (nur für Fall-Kontrollstudien zutreffend)
J / J? / N? / N / ?

⇒ **Sonderformen des selection bias, verursacht durch nicht-messbare Störfaktoren!**
gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

3) Bias durch die Erfassung der Intervention

- a) Wurde der Interventionsstatus (Dauer, Dosierung und Applikationsform) klar definiert?
J / J? / N? / N
- b) Wurde die Information über den Interventionsstatus zum Zeitpunkt der Intervention erfasst (oder retrospektiv aufgenommen)?
J / J? / N? / N / ?
- c) Erfolgte die Erfassung des Interventionsstatus unbeeinflusst vom Endpunkt?
J / J? / N? / N / ?

⇒ **Bias, verursacht durch die Erfassung der Intervention!**
gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

4) Bias durch Abweichungen in der Interventionsphase

- a) Lag eine Balance der relevanten Ko-Interventionen zwischen den Studiengruppen vor?
J / J? / N? / N
- b) Kann der Anteil der Wechsler zu anderen Interventionen als gering betrachtet werden?
J / J? / N? / N / ?
- c) Lagen höchstens geringfügige Probleme in der Implementierung der Interventionen vor?
J / J? / N? / N / ? Nur wenn N oder N? bei a) oder b) oder c):
- d) Wurde für Abweichungen in den Interventionen adjustiert?
J / J? / N? / N / ?

⇒ **u.a. performance bias, verursacht durch Abweichungen während der Behandlung!**
gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

5) **Bias durch fehlende Daten**

- a) Sind die Endpunktdaten hinreichend vollständig?
J / J? / N? / N / ?
- b) Sind die Daten zum Interventionsstatus hinreichend vollständig?
J / J? / N? / N / ?
- c) Sind die Daten zu anderen Analysevariablen hinreichend vollständig?
J / J? / N? / N / ? Nur wenn N oder N? bei a), b) oder c):
- d) Ist der Anteil der fehlenden Daten vergleichbar zwischen den Studiengruppen?
J / J? / N? / N / ?
- e) Wurden angemessene statistische Methoden eingesetzt um für fehlende Daten zu korrigieren?
J / J? / N? / N / ?

⇒ **Attrition bias, verursacht durch fehlende Daten!**

gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

6) **Bias bei der Endpunkterhebung**

- a) Handelte es sich um einen objektiven Endpunkt?
J / J? / N? / N
- b) Erfolgte die Endpunkterhebung verblindet in Bezug auf die Intervention?
J / J? / N? / N / ?
- c) Erfolgte die Endpunkterhebung in den Gruppen in vergleichbarer Weise?
J / J? / N? / N / ?
- d) Wurden systematische Fehler bei der Endpunkterhebung identifiziert, die in Zusammenhang mit der Intervention stehen könnten?
J / J? / N? / N / ?
Nur für Fall-Kontrollstudien zutreffend:
- e) Basierte die Definition von Fällen/Kontrollen auf objektiven Kriterien?
J / J? / N? / N / ?
- f) Erfolgte die Definition von Fällen/Kontrollen in Unkenntnis des Interventionsstatus (verblindet)?
J / J? / N? / N / ?

⇒ **Detection bias bei der Endpunkterhebung!**

gering (+ +) mäßig (+) erheblich (-) kritisch (- -) unklar (?)

7) **Bias durch selektives Berichten von Endpunkten**

- a) Kann selektives Berichten für multiple Endpunktmessungen ausgeschlossen werden?
J / J? / N? / N
- b) Kann selektives Berichten für multiple Analysen des Interventions-Endpunkt-Zusammenhangs ausgeschlossen werden?
J / J? / N? / N / ?
- c) Kann selektives Berichten für verschiedene Subgruppen ausgeschlossen werden?
J / J? / N? / N / ?
- d) Kann selektives Berichten für multiple Definitionen der Intervention ausgeschlossen werden?
(nur für Fall-Kontrollstudien zutreffend)
J / J? / N? / N / ?

⇒ **Reporting bias, verursacht durch selektives Berichten!**

gering (+ +)

mäßig (+)

erheblich (-)

kritisch (- -)

unklar (?)

5.2.1.2 Dokumentation und Berichten des Verzerrungspotenzials

Nach Beantwortung der vordefinierten Fragen für jede der sieben Domänen erfolgt die Bewertung des Verzerrungspotenzials jeder einzelnen Domäne (für jeden Studienendpunkt) nach folgender Klassifizierung (Tabelle 3):

- **geringes RoB (low RoB):** ++
Studie ist (für die jeweilige Bias Domäne) mit einer randomisierten Studie vergleichbar
- **mäßiges RoB (moderate RoB):** +
Studie stellt korrekt durchgeführt eine nicht-randomisierte Studie dar
- **erhebliches RoB (serious RoB):** -
Studie weist einige schwerwiegende Probleme auf
- **kritisches RoB (critical RoB):** --
Studie weist zu viele schwerwiegende Probleme auf
- **unklares RoB (unclear RoB):** ?
aufgrund mangelnder Information ist keine Einschätzung möglich

Werden dabei alle Fragen einer Domäne mit *Ja* oder *Wahrscheinlich Ja* beantwortet, dann liegt ein geringes RoB vor. Dies würde bedeuten, dass das Vertrauen in den Effektschätzer der nicht-randomisierten Studie für die entsprechende Domäne mit der Ergebniszuverlässigkeit einer randomisierten Studie vergleichbar ist. Diese Einstufung ist jedoch nur in sehr seltenen Fällen möglich. Wird eine der Fragen einer Domäne mit *Nein* oder *Wahrscheinlich Nein* beantwortet, liegt ein RoB vor, welches von den Autoren, je nach Ausmaß im Hinblick auf die klinische Fragestellung, eingeordnet werden muss (**mäßiges**, **erhebliches** oder **kritisches** RoB).

✓ Die Gesamtbewertung einer nicht-randomisierten Studie erfolgt endpunktabhängig. Die Einschätzung der Domäne mit dem höchsten RoB (d.h. der schlechteste Wert) bestimmt dabei die Gesamtbewertung (vergleiche dazu Tabelle 3 und 4).

Tabelle 3. Beispiel für eine Cochrane RoB Tabelle für eine einzelne nicht-randomisierte fiktive Studie mit verschiedenen Endpunkten (EP)

Bias Domäne	Beurteilung pro EP	Begründung der Beurteilung
1) Bias durch Störfaktoren	EP1: erhebliches Risiko	Zitat: "only counts available." <i>Kommentar:</i> Störfaktoren wurden für diesen EP nicht adjustiert.
	EP2: mäßiges Risiko	Zitat: "appropriately adjusted" <i>Kommentar:</i> In der Datenanalyse wurde für diesen EP eine Adjustierung der wichtigsten Störfaktoren vorgenommen.
2) Bias durch Selektion der Teilnehmer in die Studiengruppen	geringes Risiko (gilt für alle EP)	Zitat: "same issues thought to apply to all" <i>Kommentar:</i> selection bias kann ausgeschlossen werden, Publikation ist für diese Domäne (auf Studien-, aber nicht auf Endpunktebene) mit einer randomisierten Studie vergleichbar.
3) Bias durch die Erfassung der Intervention/Exposition	geringes Risiko (gilt für alle EP)	Zitat: "same issues thought to apply to all" <i>Kommentar:</i> Die Intervention/Exposition wurde durch die Patientenakte und durch bildgebende Verfahren erfasst.
4) Bias durch Abweichungen in der Interventionsphase	mäßiges Risiko (gilt für alle EP)	Zitat: "same issues thought to apply to all" <i>Kommentar:</i> Es lagen offensichtlich keine Abweichungen/Gruppenwechsler in der Interventionsphase vor.
5) Bias durch fehlende Daten bei der Endpunkterhebung/-bewertung	EP1: geringes Risiko	Zitat: "everyone followed up through records" <i>Kommentar:</i> Endpunktdaten sind soweit vollständig, bzw. es wurden statistische Methoden eingesetzt um für fehlende Daten zu korrigieren.
	EP2: unklares Risiko	Zitat: "due to attrition, participants missed" <i>Kommentar:</i> Es lagen Unterschiede in der Anzahl und/oder Ursache fehlender Daten zwischen den Behandlungsgruppen vor, Ausmaß aber nicht abschätzbar.
6) Bias bei der Endpunkterhebung	EP1: geringes Risiko	Zitat: "objective measure" <i>Kommentar:</i> Es wurde ein objektiver EP (Mortalität) betrachtet – auch eine fehlende Verblindung wäre hier eher unkritisch.
	EP2: kritisches Risiko	Zitat: "lack of blind outcome assessment" <i>Kommentar:</i> Es wurde ein subjektiver EP (Schmerz) betrachtet – durch die fehlende Verblindung ist von einem erheblichen RoB auszugehen.
7) Bias durch selektives Berichten von Endpunkten	EP1: mäßiges Risiko	Zitat: "unlikely to be manipulated" <i>Kommentar:</i> Dieser EP wurde vorab im Protokoll definiert und transparent berichtet.
	EP2: erhebliches Risiko	Zitat: "cut-point used without justification" <i>Kommentar:</i> Für die Bewertung des EP Schmerz wurde keine validierte Skala verwendet (ja/nein).

EP: Endpunkt.

Tabelle 4. Cochrane RoB Zusammenfassung (Profil) für eine einzelne nicht-randomisierte fiktive Studie mit verschiedenen Endpunkten

Bias Domäne	Endpunkt 1	Endpunkt 2
1) Bias durch Störfaktoren	-	+
2) Bias durch Selektion der Teilnehmer in die Studiengruppen	++	++
3) Bias durch die Erfassung der Intervention	++	++
4) Bias durch Abweichungen in der Interventionsphase	+	+
5) Bias durch fehlende Daten bei der Endpunkterhebung/-bewertung	++	?
6) Bias bei der Endpunkterhebung	++	--
7) Bias durch selektives Berichten von Endpunkten	+	-
Gesamtbewertung	-	--

geringes (++) mäßiges (+) erhebliches (-) kritisches (--) unklares (?) RoB

5.2.2 Nicht-vergleichende Studien

Unter nicht-vergleichende Studien fallen insbesondere einarmige Kohortenstudien ohne klar definierte Vergleichsgruppe und Fallserien (beziehungsweise Verlaufsbeobachtungen). Da die Bewertung des Nutzen- und Schadensverhältnisses einer Intervention eine Kontrollgruppe erfordert, lässt sich aus nicht-vergleichenden Studien in der Regel keine Aussage zur Wirksamkeit einer Intervention ableiten. Nicht-vergleichenden Studien sollten als erster Informationsgewinn, vor allem im Hinblick auf potentielle Schäden, zu einer Intervention betrachtet werden. Ausnahmen dabei bilden Interventionen bei Krankheitsbildern, die dramatische Effekte zeigen, wie zum Beispiel die Substitution von Insulin bei Patienten mit hyperglykämischer Krise bei Diabetes mellitus Typ 1. Nicht-vergleichende Querschnittstudien sind zum Beispiel für die Schätzung von Krankheitsprävalenzen geeignet, nicht jedoch für die Ableitung von Wirksamkeit. Für nicht-vergleichende Studien liegen in der Literatur (noch) keine eindeutigen Kriterien vor, nach denen das Verzerrungspotenzial auf Studienebene beurteilt werden soll. Aus methodischer Sicht kann jedoch festgehalten werden, dass auch bei nicht-vergleichenden Studien folgende Merkmale das Vertrauen in die Studienergebnisse erhöhen:

- (i) prospektive Planung mit Protokoll, in dem Einschlusskriterien und Interventionen sowie interessierende Endpunkte hinterlegt sind
- (ii) konsekutiver Patienteneinschluss
- (iii) transparentes, nicht-selektives Berichten in Bezug auf Patientencharakteristika, Intervention und Ergebnis.

6. Bewertung des Biasrisikos nach der Newcastle Ottawa Skala (NOS)

Das Verzerrungspotential einer Fall-Kontrollstudie (*siehe Abschnitt 6.1: Bewertung von Fall-Kontrollstudien*) oder einer Kohortenstudie (*siehe Abschnitt 6.2: Bewertung von Kohortenstudien*) ist insbesondere abhängig von der Strukturgleichheit der beiden zu vergleichenden Gruppen beziehungsweise Kohorten. Die Bewertung von nicht-randomisierten Studien nach der NOS vergibt daher ein Maximum von zwei Sternen für den Aspekt ‚Vergleichbarkeit‘ (siehe II., unten). Eine Studie erhält dagegen nur einen Stern für jeden gelisteten Unterpunkt der Aspekte ‚Selektion der Studienteilnehmer‘ und ‚Expositionserfassung‘ (bei Fall-Kontrollstudien) beziehungsweise ‚Endpunkterfassung‘ (bei Kohortenstudien). Insgesamt kann eine Fall-Kontrollstudie beziehungsweise eine Kohortenstudie neun Sterne erhalten.

6.1 Bewertung von Fall-Kontrollstudien

I. Selektion der Studienteilnehmer

1) Wurden die ‚Fälle‘ adäquat definiert?

- a) Ja (unabhängige Validierung, z.B. durch 2 Personen, Patientenakte(n), Dokumentation anhand bildgebender Verfahren) *
- b) Nein (z.B. Falldefinition anhand ‚record linkage‘ [z.B. anhand von ICD-Kodierung] oder Angaben des Patienten ohne vorhandene Patientenakte *oder* wenn keine Angaben vorhanden sind)

2) Sind die ‚Fälle‘ repräsentativ?

- a) Ja (z.B. konsekutive oder alle ‚Fälle‘ die in einem bestimmten Bezirk, Einzugsbereich oder einer vordefinierten Zeitspanne aufgetreten sind; randomisiertes Sample der vorliegenden ‚Fälle‘) *
- b) Nein (Potential für selection bias wahrscheinlich *oder* wenn keine Angaben vorhanden sind)

3) Sind die ‚Kontrollen‘ repräsentativ, erfolgte eine adäquate Auswahl der ‚Kontrollen‘?

- a) Ja (‚Kontrollen‘ stammen aus einer vergleichbaren Population wie die ‚Fälle‘ [jedoch ist der Endpunkt bei den ‚Kontrollen‘ nicht aufgetreten]) *
- b) Nein (Kontrollen aus dem Krankenhaus (Patienten) oder wenn keine Angaben vorhanden sind)

4) Wurden die Kontrollen adäquat definiert?

- a) Ja (der Endpunkt [z.B. Krebserkrankung], der bei den ‚Fällen‘ zum ersten mal eingetreten ist, darf bei den Kontrollen bei der Endpunkterhebung nicht vorhanden sein) *
- b) Nein (keine Angabe vorhanden, ob der Endpunkt bereits in der Kontrollgruppe aufgetreten ist)

II. Vergleichbarkeit

1) Ist die Vergleichbarkeit der ‚Fälle‘ und ‚Kontrollen‘ gegeben?

- a) Ja (die ‚Fälle‘ und ‚Kontrollen‘ wurden bereits bei der Auswahl ziemlich genau aufeinander abgestimmt [Matching] *oder* die Studie kontrolliert die wichtigsten Störfaktoren in der Datenanalyse (z.B. Alter, Geschlecht, Häufigkeit der Medikamenteneinnahme, Ko-Morbidität) * (an dieser Stelle können maximal 2 Sterne vergeben werden: ein Stern, wenn für den wichtigsten Störfaktor kontrolliert wurde und ein weiterer Stern, wenn für einen weiteren entscheidenden Störfaktor kontrolliert wurde [siehe Abschnitt 6.3: Tabellenvorlage])
- b) Nein (die Aussage: „no differences between groups or that differences were not statistically significant“ sind nicht ausreichend, um von einer Vergleichbarkeit auszugehen)

III. Expositionserfassung

1) Erfolgte eine valide Erfassung der Exposition?

- a) Ja (z.B. anhand der Patientenakte) *
- b) Ja (durch ein Interview z.B. der Kollegen oder Freunde, die im Hinblick auf den ‚Fall-Kontroll-Status‘ verblindet waren) *
- c) Nein (durch ein Interview z.B. der Kollegen oder Freunde, die im Hinblick auf den ‚Fall-Kontroll-Status‘ nicht verblindet waren)
- d) Nein (Angaben des Patienten ohne vorhandene Patientenakte *oder* wenn keine Angaben vorhanden sind)

2) Erfolgte die Erfassung der ‚Fälle‘ und ‚Kontrollen‘ identisch?

- a) Ja (z.B. durch identische standardisierte diagnostische Methoden) *
- b) Nein

3) Kann die ‚Non-Response-Rate‘ als valide betrachtet werden?

- a) Ja (für beide Gruppen liegt die Rate vor) *
- b) Nein (keine Ereignisraten angegeben)
- c) Nein (unterschiedliche Ereignisraten, in der Studie wird jedoch nicht weiter darauf eingegangen)

6.2 Bewertung von Kohortenstudien

I. Selektion der Studienteilnehmer

- 1) Ist die exponierte Kohorte repräsentativ für die zu untersuchende Intervention/Exposition?
 - a) Ja und Wahrscheinlich Ja (sollen z.B. unerwünschte Wirkungen der Östrogen-Exposition in der Post-Menopause untersucht werden, muss eine Frauen-Kohorte ausgewählt werden, die repräsentativ für die Einnahme dieser Östrogene ist. Frauen, die z.B. einer ethnischen Minderheit angehören, wären in diesem Fall nicht repräsentativ). *
 - b) Nein (selektiertes Sample wie z.B. freiwillig oder Krankenschwestern *oder* wenn keine Angaben vorhanden sind)

- 2) Ist die nicht-exponierte Kohorte repräsentativ, wurde sie adäquat ausgewählt?
 - a) Ja (die nicht-exponierte Kohorte stammt aus einer vergleichbaren Grundgesamtheit wie die exponierte Kohorte) *
 - b) Nein (die nicht-exponierte Kohorte stammt aus einer anderen Grundgesamtheit, z.B. aus dem Krankenhaus (Patienten) oder wenn keine Angaben vorhanden sind)

- 3) Erfolgte eine valide Erfassung der Exposition?
 - a) Ja (z.B. anhand der Patientenakte) *
 - b) Ja (durch ein strukturiertes Interview) *
 - d) Nein (narrative Angaben der Studienteilnehmer [ohne vorhandene Patientenakte] *oder* wenn keine Angaben vorhanden sind)

- 4) Ist es wahrscheinlich, dass der gemessene Endpunkt nicht zu Studienbeginn vorhanden war?
 - a) Ja (z.B. diagnostische Maßnahmen erfolgten) *
 - b) Nein (keine Angabe vorhanden, ob der Endpunkt bereits am Anfang der Studie vorhanden war)

II. Vergleichbarkeit

1) Ist die Vergleichbarkeit der exponierten und nicht-exponierten Kohorte gegeben?

- a) Ja (die exponierte und nicht-exponierte Kohorte wurden bereits bei der Auswahl ziemlich genau aufeinander abgestimmt [Matching] *oder* die Studie kontrolliert für die wichtigsten Störfaktoren in der Datenanalyse (z.B. Alter, Geschlecht, Häufigkeit der Medikamenteneinnahme, Ko-Morbidität, ethnische Herkunft) * (an dieser Stelle können maximal 2 Sterne vergeben werden: ein Stern, wenn für den wichtigsten Störfaktor kontrolliert wurde und ein weiterer Stern, wenn für einen weiteren entscheidenden Störfaktor kontrolliert wurde [siehe Abschnitt 6.3: Tabellenvorlage])
- b) Nein (die Aussage: „no differences between groups or that differences were not statistically significant“ sind nicht ausreichend, um von einer Vergleichbarkeit auszugehen)

III. Endpunkterfassung

1) Erfolgte eine valide Erfassung der Endpunkte?

- a) Ja (unabhängige oder verblindete Erhebung, z.B. durch 2 Personen, Patientenakte(n), Dokumentation anhand bildgebender Verfahren) *
- b) Ja (anhand ‚record linkage‘ [z.B. anhand von ICD-Kodierung] in der Patientenakte) *
- c) Nein (Angaben des Patienten [ohne dass eine Patientenakte verfügbar ist] *oder* wenn keine Angaben vorhanden sind)

2) Konnte in der Beobachtungszeit der Endpunkt überhaupt auftreten?

- a) Ja (*a priori* muss eine ausreichende Beobachtungszeit festgelegt werden, damit sicher gestellt ist, dass der Endpunkt auch in dieser Zeit auftreten kann, z.B. sollte bei der Bewertung der Verträglichkeit von Brustimplantaten ein Minimum von 5 Jahren angesetzt werden) *
- b) Nein

3) Wurden fehlende Daten adäquat berücksichtigt? (siehe Abschnitt 5.1.1: Domänen der RoB Bewertung/ Fehlende Daten bei der Endpunkterhebung)

- a) Ja (es liegen keine fehlenden Daten bei der Endpunkterhebung vor) *
- b) Ja (der Einfluss der fehlenden Daten auf den Effekt ist statistisch und/oder klinisch nicht relevant und/oder steht wahrscheinlich nicht in Zusammenhang mit der Exposition) *
- c) Nein (der Einfluss der fehlenden Daten auf den Effekt ist statistisch und/oder klinisch relevant und/oder steht wahrscheinlich in Zusammenhang mit der Exposition)
- d) Nein (keine Angaben dazu vorhanden)

6.3 Tabellenvorlage

Tabelle 5. NOS RoB Tabelle für nicht-randomisierte Studien

Fall-Kontrollstudien			
	Selektion	Vergleich- barkeit	Expositions- erfassung
Studie 1	****	**	***
Studie 2	**	*	*
Kohortenstudien			
	Selektion	Vergleich- barkeit	Endpunkt- erfassung
Studie 1	*	**	***
Studie 2	***	**	**

Insgesamt kann eine Fall-Kontroll- beziehungsweise eine Kohortenstudie neun Sterne erhalten.

7. Quellen

1. Deutsches Cochrane Zentrum, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften-Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin. "Manual Systematische Literaturrecherche für die Erstellung von Leitlinien". 1. Auflage 2013. Verfügbar: DCZ: <http://www.cochrane.de/de/manual-systematische-literaturrecherche>, AWMF: <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html>, ÄZQ: <http://www.aezq.de/aezq/publikationen/koooperation>, DOI: 10.6094/UNIFR/2013/2, <http://www.freidok.uni-freiburg.de/volltexte/9020/>, Zugriff: 07.03.2016.
2. Buchberger B, von Elm E, Gartlehner G, et al. Assessment of risk of bias in controlled studies. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014;57(12):1432-8.
3. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323(7303):42-6.
4. 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.
5. GCP - Good Clinical Practice. Secondary GCP - Good Clinical Practice. http://www.dgrw-online.de/files/leitlinien_gcp_korrektur_2002_deutsche_version.pdf, Zugriff: 07.03.2016.
6. ICH Guidelines - The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. <http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html>, Zugriff: 07.03.2016.
7. Deutsche Gesellschaft für Epidemiologie 2004. http://www.gesundheitsforschung-bmbf.de/media/Empfehlungen_GEP.pdf, Zugriff: 07.03.2016.
8. Blumle A, von Elm E, Antes G, et al. Measurement and assessment of study quality and reporting quality. Z Evid Fortbild Qual Gesundheitswes 2014;108(8-9):495-503.
9. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276(8):637-9.
10. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152(11):726-32.
11. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev 2012;11:MR000030.
12. Langer G, Meerpohl JJ, Perleth M, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. Z Evid Fortbild Qual Gesundheitswes 2012;106(5):357-68.
13. Chavalarias D, Ioannidis JP. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol 2010;63(11):1205-15.
14. Deutsches Netzwerk Evidenzbasierte Medizin e.V. www.ebm-netzwerk.de/was-ist-ebm/images/dnebm-glossar-2011.pdf, Zugriff: 07.03.2016.

15. Meerpohl JJ, Schell LK, Bassler D, et al. Evidence-informed recommendations to reduce dissemination bias in clinical research: conclusions from the OPEN (Overcome failure to Publish nEgative fINdings) project based on an international consensus meeting. *BMJ Open* 2015;5(5):e006666.
16. Schmucker C, Schell LK, Portalupi S, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS One* 2014;9(12):e114023.
17. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330(7494):753.
18. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Stat Med* 2001;20(24):3931-46.
19. Akl EA, Johnston BC, Alonso-Coello P, et al. Addressing dichotomous data for participants excluded from trial analysis: a guide for systematic reviewers. *PLoS One* 2013;8(2):e57132.
20. Akl EA, Shawwa K, Kahale LA, et al. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ Open* 2015;5(12):e008431.
21. Jones B, Jarvis P, Lewis JA, et al. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313(7048):36-9.
22. Wolkewitz M, Allignol A, Harbarth S, et al. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. *J Clin Epidemiol* 2012;65(11):1171-80.
23. Werler MM, Pober BR, Nelson K, et al. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989;129(2):415-21.
24. Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157(6):429-38.
25. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359(9307):696-700.
26. Otto C, Schiffer G, Tjardes T, et al. Blood loss and operative duration using monopolar electrosurgery versus ultrasound scissors for surgical preparation during thoracoscopic ventral spondylodesis: results of a randomized, blinded, controlled trial. *Eur Spine J* 2014;23(8):1783-90.
27. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336(7644):601-5.
28. Lundh A, Sismondo S, Lexchin J, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2012;12:MR000033.
29. Bero LA. Why the Cochrane risk of bias tool should include funding source as a standard item. *Cochrane Database Syst Rev* 2013;12:ED000075.
30. Sterne JA. Why the Cochrane risk of bias tool should not include funding source as a standard item. *Cochrane Database Syst Rev* 2013;12:ED000076.
31. Dreier M, Borutta B, Stahmeyer J, et al. Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Schriftenreihe Health Technology Assessment. Köln: Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), 2010.
32. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;3(12):i-iv, 1-98.

33. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054-60.
34. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17(1):1-12.
35. Newcastle Ottawa Quality Assessment Scale (NOS).
http://www.ohri.ca/programs/clinical_epidemiology/nosgen.pdf, Zugriff: 07.03.2016.
36. Scottish Intercollegiate Guidelines Network. <http://www.sign.ac.uk/methodology/checklists.html>, Zugriff: 06.04.2016.
37. Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008;8:22.
38. Higging J, Green S. Cochrane Handbook of Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. <http://handbook.cochrane.org/>, Zugriff:07.03.2016.
39. Sterne JA, Higging 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. <http://www.riskofbias.info>, Zugriff: 07.03.2016.
40. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. J Clin Epidemiol 2013;66(4):408-14.
41. Hrobjartsson A, Boutron I, Turner L, et al. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev 2013;4:ED000058.
42. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PLoS One 2014;9(5):e96920.
43. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2013;66(9):973-81.
44. Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. Control Clin Trials 1988;9(4):365-74.
45. Schulz KF, Grimes DA. Reihe Epidemiologie 6: Generierung von Randomisierungslisten in randomisierten Studien: Zufall, nicht Auswahl. Z ärztl Fortbild Qual Gesundheitswes 2007;101:419-26.

8. Weiterführende Informationen und Praxishilfen

Leitlinien zur Verbesserung der Berichterstattung verschiedener Studientypen einschließlich Systematischer Übersichtsarbeiten sind abrufbar unter: <http://www.equator-network.org/>.

equator network Enhancing the QUALity and Transparency Of health Research

Visit the EQUATOR Spanish Website

Home Library Toolkits Courses & events News Blog About us Contact

Essential resources for writing and publishing health research

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		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 291 reporting guidelines](#)

Organização Pan-Americana da Saúde
Organização Mundial da Saúde
ESCRITÓRIO REGIONAL PARA AS Américas

[Recursos em Português](#)

[Visit our new resource pages in Portuguese](#)

Abb. 3. Screenshot der Equator Webseite (<http://www.equator-network.org/>).