

Bundesgesundheitsbl 2014 · 57:1432–1438
 DOI 10.1007/s00103-014-2065-6
 Online publiziert: 8. November 2014
 © Die Autor(en) 2014. Dieser Artikel ist auf
 Springerlink.com mit Open Access verfügbar.

B. Buchberger¹ · E. von Elm² · G. Gartlehner³ · H. Huppertz¹ · G. Antes⁴ · J. Wasem¹ · J.J. Meerpohl⁴

¹ Fakultät für Wirtschaftswissenschaften, Lehrstuhl für Medizinmanagement, Universität Duisburg-Essen, Essen, Deutschland

² Institut universitaire de médecine sociale et préventive (IUMSP) Lausanne, Cochrane Schweiz, Lausanne, Schweiz

³ Department für Evidenzbasierte Medizin und Klinische Epidemiologie, Donau-Universität Krems, Österreichische Cochrane Zweigstelle, Krems, Österreich

⁴ Deutsches Cochrane Zentrum, Universitätsklinikum Freiburg, Freiburg, Deutschland

Bewertung des Risikos für Bias in kontrollierten Studien

Elektronisches zusätzliches Material

Die Online-Version dieses Artikels (doi: 10.1007/s00103-014-2065-6) enthält zusätzliches Material, das für autorisierte Benutzer zugänglich ist.

Entscheidungen im Gesundheitswesen treffen Ärzte, Angehörige der Gesundheitsfachberufe, Patienten und Kostenträger nach Abwägung von Nutzen und Schaden alternativer Behandlungsstrategien. Gemäß international anerkannten Standards erreicht der praktizierende Arzt für seine Patienten die bestmögliche Behandlung unter Anwendung der Methoden der evidenzbasierten Medizin (EbM), die sowohl seine persönlichen Erfahrungen als auch die Patientenpräferenzen und die externe Evidenz aus Studien berücksichtigen [1]. Da die individuelle klinische Erfahrung von Zufall und Bias beeinflusst wird und sie somit fehleranfällig ist, fordert auch der deutsche Gesetzgeber ein ärztliches Handeln entsprechend den Methoden der EbM (Übersicht 1).

Validität und Bias

Die Qualität einer kontrollierten Studie wird von verschiedenen Dimensionen bestimmt, die von Design, Durchführung und Analyse bis zur klinischen Relevanz und Berichtsqualität reichen. Eine wesentliche Dimension ist die Validität von Studienergebnissen [7], die aufgrund von systematischen Fehlern (auch Verzerrung oder Bias) eingeschränkt sein kann. Das Risiko für Bias lässt sich je-

doch schon bei der Studienplanung und -durchführung verringern (s. unten). Zu unterscheiden sind interne und externe Validität (Übersicht 2): Die interne Validität (Ausmaß des Risikos für Bias) bedingt die externe Validität und beide die Validität weiterführender Analysen (wie z. B. im Rahmen der gesundheitsökonomischen Evaluation).

Methoden zur Verringerung von Bias und Auswirkungen von Bias in kontrollierten Studien

Zum Nachweis kausaler Zusammenhänge zwischen Intervention und Studienergebnis ist es notwendig, eine möglichst große Ähnlichkeit der Behandlungsgruppen hinsichtlich prognostisch relevanter Faktoren herzustellen, die lediglich durch eine zufällige Variabilität beeinträchtigt wird.

Selection-Bias. Um bei der Auswahl der Studienpopulation einen Selection-Bias zu vermeiden, der zu Gruppenunterschieden in wesentlichen Patientencharakteristika, wie beispielsweise dem Alter oder Erkrankungsstatus, führen kann, werden die Teilnehmer den Interventionsgruppen zufällig zugeteilt (randomisiert). Für eine erfolgreiche Randomisierung, d. h. Ähnlichkeit der zu vergleichenden Gruppen, sind drei Faktoren maßgeblich: eine ausreichend große Studienpopulation, die Wahl einer adäquaten Randomisierungsmethode (s. Anhang 1) und die Geheimhaltung des Randomisierungsschemas

(verdeckte Gruppenzuteilung, englisch: *Allocation concealment*). Die Kenntnis des Randomisierungsplans, beispielsweise einer Zufallszahlentabelle, birgt die Gefahr einer selektiven Zuordnung von Patienten zu Behandlungsgruppen auf der Basis prognostischer Faktoren: Studienteilnehmer, die für eine Behandlung vorgesehen waren, mögen als ungeeignet erscheinen und werden zurückgewiesen, wohingegen andere vorsätzlich einer geeignet erscheinenden Gruppe zugeteilt werden [7].

Performance-Bias. Ein Performance-Bias wird durch Unterschiede in der Behandlung der Patienten zusätzlich zur Intervention und in den Rahmenbedingungen hervorgerufen, oftmals nur in einer Behandlungsgruppe. Er kann durch Verblindung des Studienpersonals reduziert werden.

Detection-Bias. Auch der Detection-Bias, der entsteht, wenn bekannt ist, welcher Behandlungsgruppe ein Patient zugeordnet ist und dies die Ergebnismessung beeinflusst, kann durch Verblindung verringert werden. Allerdings ist in vielen nicht medikamentösen Studien eine Verblindung nur schwer zu realisieren und wird entsprechend selten vorgenommen, obwohl zumindest eine verblindete Endpunkterhebung (auch Zielgrößen- oder Outcome-Erhebung) meistens möglich ist. Daraus folgt nicht zwingend ein gravierender Qualitätsmangel der Studie, jedoch können viele Endpunkte durch das

Übersicht 1: Evidenzbasierte Medizin (EbM) in der deutschen Gesetzgebung

SGB V

- § 35 Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen:
Abs. 1: Das Bundesministerium für Gesundheit regelt [...] das Nähere zur Nutzenbewertung. Darin sind insbesondere festzulegen: Grundsätze für die Bestimmung der zweckmäßigen Vergleichstherapie und des Zusatznutzens, [...] und die Voraussetzungen, unter denen Studien bestimmter Evidenzstufen zu verlangen sind; Grundlage sind die internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie.
- § 137 f Strukturierte Behandlungsprogramme bei chronischen Krankheiten:
(2) Der G-BA nach § 91 erlässt Richtlinien zu den Anforderungen an die Ausgestaltung von Behandlungsprogrammen nach Absatz 1. Zu regeln sind insbesondere Anforderungen an die
1. Behandlung nach dem aktuellen Stand der medizinischen Wissenschaft unter Berücksichtigung von evidenzbasierten Leitlinien oder nach der jeweils besten, verfügbaren Evidenz sowie unter Berücksichtigung des jeweiligen Versorgungsektors [...]
- § 139 a IQWiG:
(3) 3 Bewertungen evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten
- (4) Das Institut hat zu gewährleisten, dass die Bewertung des medizinischen Nutzens nach den international anerkannten Standards der evidenzbasierten Medizin [...] erfolgt.

Verfahrensordnung G-BA

- § 5 Abs. 2: Der G-BA ermittelt den allgemein anerkannten Stand der medizinischen Erkenntnisse auf der Grundlage der evidenzbasierten Medizin

G-BA: Gemeinsamer Bundesausschuss (oberstes Beschlussgremium der gemeinsamen Selbstverwaltung der Ärzte, Zahnärzte, Psychotherapeuten, Krankenhäuser und Krankenkassen in Deutschland, überprüft mit den Methoden der EbM den Leistungskatalog der gesetzlichen Krankenversicherung und entscheidet über die Erstattung neuer Untersuchungs- und Behandlungsmethoden), IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, SGB: Sozialgesetzbuch

Die wissenschaftliche Belegbarkeit der Wirksamkeit medizinischer Maßnahmen als Ergänzung zur ärztlichen Erfahrung [2] kann im Praxisalltag nicht nur den Patienten helfen, sondern auch bei der Erstellung von Gutachten nützlich sein und zum Schutz bei Regressforderungen beitragen. Allerdings muss der einzelne Arzt in der Lage sein, die Validität von Studienergebnissen einzuschätzen [3]. Notwendig hierfür ist auch eine nachvollziehbare Darstellung des Designs, der Durchführung, Analyse und Interpretation einer Studie in der entsprechenden Publikation, d. h. eine gute Berichtsqualität. Erst eine solche ermöglicht es, Ergebnisse von Studien adäquat zu beurteilen.

Ein Verständnis der Konzepte von Validität und Bias und dafür, wie Studien angelegt und durchgeführt werden sollen [4], ist für alle Ärzte essentiell, um ihre klinischen Entscheidungen auf der Basis von Literatur rechtfertigen zu können [5]. Die nun vorliegende deutsche Übersetzung des Risk-of-Bias (RoB)-Tool (Kap. 8.5 aus dem Cochrane Handbook for Systematic Reviews of Interventions [6]; s. Anhang) erleichtert den Zugang zu den Methoden der EbM für klinisch tätige Ärzte und Entwickler von Leitlinien, aber auch für einzelne Patienten, Patientenvertreter oder Selbsthilfegruppen im Sinne von *Patient advocacy* sowie für Entscheidungsträger im Gesundheitswesen oder Mitarbeiter von Verbänden. In der Lehre lassen sich mithilfe der Übersetzung wesentliche Inhalte und die Anwendung des Bewertungsinstruments einfacher und schneller vermitteln, und zwar sowohl im Medizinstudium als auch im Kontext der Akademisierung von Gesundheitsfachberufen, die aufgrund der komplexer werdenden Herausforderungen im Gesundheitssystem vorangetrieben wird.

Übersicht 2: Wesentliche Aspekte von Validität und Bias in kontrollierten Studien (Quelle: mod. nach Jüni et al. [7])

Interne Validität: Ausmaß, in dem systematische Fehler verringert werden konnten

Bias: Systematischer Fehler, Verzerrung

- **Selection-Bias:** Verzerrte Zuteilung zu Studiengruppen
- **Performance-Bias:** Verzerrung durch ungleiche Behandlungen abgesehen von Interventionen
- **Detection-Bias:** Verzerrte Erfassung von Ergebnissen
- **Attrition-Bias:** Verzerrung durch Gruppenunterschiede in der Anzahl und Begründung von Studienabbrüchen
- **Reporting-Bias:** Verzerrung durch selektives Berichten von positiven Ergebnissen

Externe Validität: Ausmaß, in dem eine Verallgemeinerung der Studienergebnisse möglich ist

- **Patienten:** Alter, Geschlecht, Schweregrad und Risikofaktoren, Komorbidität
- **Behandlungsplan:** Dosierung, Häufigkeit und Art der Verabreichung, Art des Präparats innerhalb von Arzneimittelklassen, Begleitbehandlungen
- **Setting:** Versorgungsstufe (primär, sekundär, tertiär), Erfahrung und Spezialisierung des Leistungserbringers
- **Zielgrößen:** Typ oder Definition und Dauer der Nachbeobachtung

Wissen um die Behandlung eines Studienteilnehmers beeinflusst werden. Folglich entsteht durch die fehlende Verblindung ein Risiko für Bias [8].

Verblindung und verdeckte Gruppenzuteilung sind unterschiedliche Aspekte, die oft miteinander verwechselt werden. Die verdeckte Gruppenzuteilung muss bereits bei der Aufnahme eines Teilnehmers in eine Studie gewährleistet sein, um einen Selection-Bias zu vermeiden. Die Verblindung muss hingegen nach der

Aufnahme beginnen, um Performance- und Detection-Bias zu verringern.

Attrition-Bias. Ein Attrition-Bias entsteht durch den Studienabbruch einzelner Teilnehmer, weil dieser in der Regel nicht zufällig erfolgt und weil er bei hohen Ausfallzahlen sowie Gruppenunterschieden in der Häufigkeit von Studienabbrüchen den Effekt einer Randomisierung aufhebt. Eine Verletzung des Studienprotokolls oder ein Loss-to Follow-up können unterschiedlichste Ursachen haben, wie beispielsweise mangelnde Wirksamkeit oder schwer zu tolerierende Nebenwirkungen, aber auch Desinteresse oder Unzuverlässigkeit der Patienten. Um die potentielle Über- oder Unterschätzung eines Interventionseffektes aufgrund von Studienabbrüchen zu verringern, sollte eine Intention-to-treat-Analyse durchgeführt werden. Dazu werden zur Auswertung alle Patienten den Behandlungsgruppen zugerechnet, denen sie ursprünglich randomisiert zugeordnet waren. Zur Kompensation fehlender Messwerte von ausgeschiedenen Teilnehmern stehen verschiedene Methoden zur Verfügung [9–11].

Auswirkung von Bias

Eine Zusammenfassung von Daten aus sieben metaepidemiologischen Studien (▣ **Abb. 1**) ergab, dass eine inadäquate oder fehlende Randomisierung zu einer Überschätzung der Behandlungseffekte um durchschnittlich 11 % führt [12]. Durch eine fehlende oder inadäquate verdeckte Gruppenzuteilung werden Behandlungseffekte um 7 %, und im Fall einer fehlenden oder inadäquaten doppelten Verblindung um 13 % überschätzt [12].

In einem gewissen Ausmaß hängt die Bedeutung der Verblindung für das Bias-Risiko allerdings von der Art der Untersuchungsparameter und vom Studiendesign ab. Im Fall des Endpunktes „Mortalität“ beispielsweise ist eine fehlende Verblindung mit einem deutlich geringeren Risiko für einen Bias verbunden als bei einem Endpunkt, der einen Interpretationsspielraum lässt. Ähnlich verhält es sich mit Verblindungen von medikamentösen Interventionen, die von unverwechselba-

Bundesgesundheitsbl 2014 · 57:1432–1438 DOI 10.1007/s00103-014-2065-6
© Springer-Verlag Berlin Heidelberg 2014

B. Buchberger · E. von Elm · G. Gartlehner · H. Huppertz · G. Antes · J. Wasem · J.J. Meerpohl

Bewertung des Risikos für Bias in kontrollierten Studien

Zusammenfassung

Hintergrund. Praktisch tätige Ärzte treffen täglich eine Vielzahl von medizinischen Entscheidungen. Diese basieren überwiegend auf dem in der Ausbildung Erlernten und persönlicher Erfahrung, sollen heutzutage jedoch auch Patientenpräferenzen und die wissenschaftliche Evidenz für den Nutzen einer Intervention berücksichtigen. Mit dem Ziel der bestmöglichen Versorgung von Patienten bilden diese drei Aspekte gleichgewichtet die Grundlage für das Konzept der evidenzbasierten Medizin (EbM). Ohne Verständnis der methodischen Hintergründe ist die Gefahr von Fehleinschätzungen der Evidenz jedoch hoch und kann Behandlungsfehler zur Konsequenz haben.

Fragestellung. Im vorliegenden Beitrag werden das Konzept der systematischen Verzerrungen (Bias) und seine Bedeutung erläutert. Hierzu werden Ursachen, Auswirkungen und Methoden zur Minimierung von Bias dargestellt. Diese Inhalte sollen ein tieferes Verständnis vermitteln, um in der Praxis eine

bessere Einschätzung von Studien und ihren Ergebnissen sowie die Umsetzung ihrer Empfehlungen zu erlauben.

Schlussfolgerungen. Das Risk-of-Bias (RoB)-Tool der Cochrane Collaboration ist ein Instrument zur Bewertung des Verzerrungspotentials in kontrollierten Studien. Zu seinen Stärken zählen eine einfache Anwendung, kurze Bearbeitungszeit, hohe Transparenz der Beurteilung und eine leicht verständliche graphische Darstellung der Ergebnisse. Mit dieser Publikation wird die deutsche Fassung des RoB-Tool veröffentlicht. Sie soll die Anwendung des Instruments auch außerhalb von Expertenkreisen erleichtern und über die Einschätzung der Validität von Studienergebnissen zur Entscheidungsfindung bei medizinischen Fragestellungen beitragen.

Schlüsselwörter

Bias · Qualitätsbewertung · Instrument · Berichtsqualität · Cochrane Collaboration

Assessment of risk of bias in controlled studies

Abstract

Background. Practicing physicians are faced with many medical decisions daily. These are mainly influenced by personal experience but should also consider patient preferences and the scientific evidence reflected by a constantly increasing number of medical publications and guidelines. With the objective of optimal medical treatment, the concept of evidence-based medicine is founded on these three aspects. It should be considered that there is a high risk of misinterpreting evidence, leading to medical errors and adverse effects without knowledge of the methodological background.

Objectives. This article explains the concept of systematic error (bias) and its importance. Causes and effects as well as methods to minimize bias are discussed. This information should impart a deeper understanding, lead-

ing to a better assessment of studies and implementation of its recommendations in daily medical practice.

Conclusion. Developed by the Cochrane Collaboration, the risk of bias (RoB) tool is an assessment instrument for the potential of bias in controlled trials. Good handling, short processing time, high transparency of judgments and a graphical presentation of findings that is easily comprehensible are among its strengths. Attached to this article the German translation of the RoB tool is published. This should facilitate the applicability for non-experts and moreover, support evidence-based medical decision-making.

Keywords

Bias · Quality assessment · Tool · Reporting quality · Cochrane collaboration

ren Nebenwirkungen begleitet werden. Bei chirurgischen Interventionen ist eine doppelte Verblindung selten realisierbar, allerdings lohnt es sich, hier alle Möglichkeiten gewissenhaft zu überprüfen [7, 13]: Wenn z. B. die intraoperative Blutungsmenge untersucht werden soll, kann auch

ein Operateur für diesen Endpunkt verblindet werden [14].

Ob die Endpunkterhebung subjektiven oder objektiven Einflüssen unterliegt, hat ebenfalls Auswirkungen auf das Ausmaß des Bias-Risikos. Ergebnis einer metaepidemiologischen Studie war, dass eine inadäquate oder fehlende Ver-

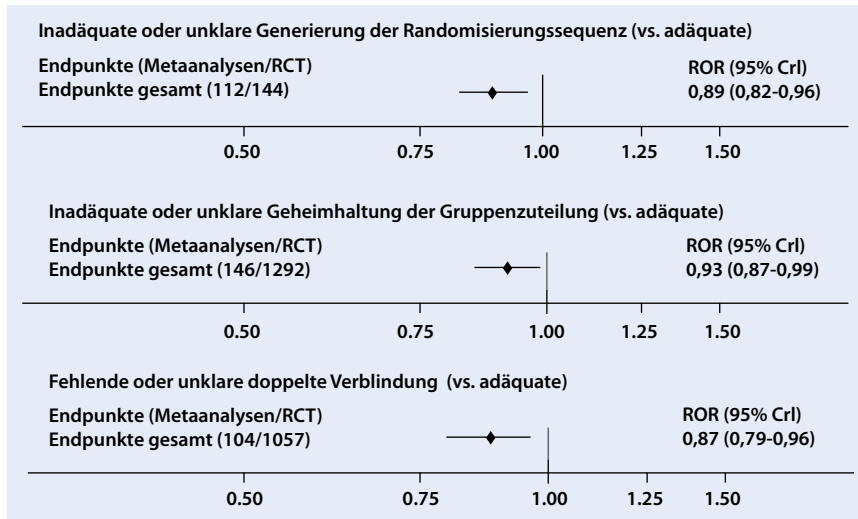


Abb. 1 ▲ Potenzielle Überschätzung von Ergebnissen aufgrund fehlender Kontrolle von Bias. *CrI* Credibility interval (Glaubwürdigkeitsintervall), *RCT* Randomised Controlled Trial, *ROR* Ratio of Odds Ratios. (Quelle: mod. nach Savović et al. [12])

deckung der Gruppenzuteilung zu einer Überschätzung subjektiv erhobener Endpunkte (beispielsweise durch Patienten berichtete oder durch Ärzte beurteilte Endpunkte wie vaskuläre Ereignisse, Pyelonephritis oder Atemnotsyndrom) von 30% führt, im Fall fehlender Verblindung zu einer Überschätzung von 25% [15].

Untersuchungen zum Einfluss des Attrition-Bias sind bis dato wenig aussagekräftig und generell schwierig. Fehlen nämlich in Publikationen die Informationen darüber, warum Teilnehmer eine Studie abbrechen, oder sind die Angaben nicht nach Behandlungsgruppen differenziert, ist es weder möglich, zwischen Berichtsqualität und methodischer Qualität zu unterscheiden, noch kann die Auswirkung der Studienabbrüche auf den Endpunkt eingeschätzt werden. Darüber hinaus ist zu berücksichtigen, dass der Studienabbruch eines Teilnehmers in manchen Fällen zwingend notwendig ist, er sich in anderen Fällen aber durchaus vermeiden lässt [7].

Ein großes und viel diskutiertes Bias-Risiko birgt auch die industrielle Finanzierung von Studien. Die Autoren eines Cochrane-Review kamen zu dem Schluss, dass eine finanzielle Unterstützung durch Hersteller zu deutlich positiveren Ergebnissen hinsichtlich von Wirksamkeit und Nebenwirkungen einer Intervention führt [16]. Die Diskussion um eine Aufnahme

dieses Bias in das RoB-Tool ist noch nicht abgeschlossen [17, 18].

Bewertungsinstrumente

Zur standardisierten Bewertung des Bias-Risikos steht eine Vielzahl von Instrumenten zur Verfügung, die sich im Wesentlichen drei Kategorien zuordnen lassen: Skalen, Checklisten und Komponentensysteme [19].

Skalen. In einer Skala wird jedes Item anhand eines Scores numerisch bewertet; für eine Gesamtbewertung der Studie können die einzelnen Bewertungen zu einem Gesamt-Score addiert werden. Die einfache und dennoch quantitative Bewertung einer Studie scheint für die Anwendung einer Skala zu sprechen, allerdings wird das Verfahren durch die empirische Forschung wenig gestützt [20, 21]: Weder eine implizite Gleichgewichtung aller Items noch eine explizit unterschiedliche Gewichtung einzelner Items konnte bisher gerechtfertigt werden. Darüber hinaus dienen Skalen oft der Bewertung von Aspekten, die eher auf die Berichtsqualität, ethische Probleme oder statistische Verfahren bezogen sind als auf das Risiko für Bias [21]. Zum Teil fehlen auch für die Bewertung wichtige Domänen, wie die verdeckte Gruppenzuteilung im Fall des bekannten Jadad-Score [22]. Ferner berichten Jüni et al. 1999 über substanzielle Un-

terschiede in den Ergebnissen von Metaanalysen, die auf die Verwendung unterschiedlicher Skalen zurückzuführen waren [21]. Zwar deuten Untersuchungen darauf hin, dass der Einsatz von Skalensystemen in systematischen Reviews zurückgeht [23, 24], aber noch ist deren Anwendung so verbreitet, dass auch in der aktuellen Version des Cochrane Handbook for Systematic Reviews of Interventions ausdrücklich davon abgeraten wird [6].

Checklisten. Checklisten bestehen aus mindestens zweizu bewertenden Items. Sie können sowohl qualitative Bewertungen von einzelnen Items als auch qualitative Bewertungen der ganzen Studie enthalten. Meistens beinhalten sie weit mehr als eine Abfrage von Aspekten zur Bewertung des Bias-Risikos, wie z. B. das Auflisten von Ein- und Ausschlusskriterien oder die Beurteilung der Angemessenheit von Fallzahlplanung oder Ergebnispräsentation [24].

Komponentensysteme. In Komponentensystemen kann ein Bias-Risiko für die verschiedenen Dimensionen separat dargestellt werden. So wird der unterschiedlichen Bedeutung einzelner Aspekte für bestimmte Studiendesigns, Interventionen oder Untersuchungsparameter Rechnung getragen (s. oben, unten und Anhang) [7].

Allen Instrumenten ist gemeinsam, dass sie keine exakte Messung, sondern eine Einschätzung und Bewertung des Bias-Risikos darstellen, die eine subjektive Komponente beinhalten.

Das Risk-of-Bias-Tool der Cochrane Collaboration

Das RoB-Tool der Cochrane Collaboration ist ein Komponentensystem, dessen einzelne Bewertungskriterien auf der Grundlage theoretischer und empirischer Forschungsergebnisse sowie von ausführlichen Diskussionen unter Methodikern im Bereich der klinischen Forschung ausgewählt wurden [25]. Es setzt sich aus sieben Domänen zusammen, die zum Teil bereits dargestellt wurden (s. oben): Generierung der Randomisierungssequenz, verdeckte Gruppenzuteilung, Verblindung von Teilnehmern und Studienper-

| | Generierung der Randomisierungssequenz (Selection-Bias) | Verdeckte Gruppenzuteilung (Selection-Bias) | Verblindung von Teilnehmern und Studienpersonal (Performance-Bias) | Verblindung der Endpunkterhebung (Detection-Bias) | Unvollständige Daten zu Endpunkten (Attrition-Bias) | Selektives Berichten zu Endpunkten (Reporting-Bias) | Andere Ursachen für Bias |
|-----------------|---|---|--|---|---|---|--------------------------|
| Buchberger 2012 | - | - | - | - | - | ? | - |
| Gartlehner 2013 | + | ? | ? | ? | + | + | - |
| Huppertz 2012 | ? | ? | - | ? | - | + | - |
| Meerpohl 2013 | + | + | + | + | + | + | ? |
| von Elm 2013 | + | + | + | + | + | + | + |
| Wasem 2013 | ? | ? | ? | ? | + | ? | - |

| | |
|--|-----------------|
| - | Hohes Risiko |
| ? | Unklares Risiko |
| + | Geringes Risiko |

Abb. 2 ◀ Risiko für Bias in fiktiven Studien

sonal, Verblindung der Endpunkterhebung, unvollständige Daten zu Endpunkten, selektives Berichten und andere Ursachen für Bias.

Selektives Berichten zu Endpunkten wird auch als Outcome-Reporting-Bias bezeichnet. Es resultiert daraus, dass vorzugsweise positive und signifikante Ergebnisse berichtet werden. Die Motive hierfür können Interessen von Editoren, Studienautoren oder Herstellern sein. Ein solcher Bias kann zu einer Überschätzung der Wirksamkeit von Interventionen führen.

Andere Ursachen für Bias können beispielsweise Abweichungen vom Studienprotokoll sein, die nicht die klinische Praxis widerspiegeln. Hierzu zählen z. B. eine

übertriebene Post-hoc-Dosiserhöhung, eine Intervention vor Randomisierung der Teilnehmer mit Auswirkungen auf die nachfolgende Intervention, der Einsatz eines unempfindlichen Messinstrumentes, das zur Unterschätzung sowohl von Nutzen als auch Schaden der Intervention führt, und letztendlich auch Betrag [6].

Für jede Domäne und Studie ist eine Beurteilung mit „geringes Risiko für Bias“, „hohes Risiko für Bias“ oder „unklares Risiko für Bias“ erforderlich. Darüber hinaus können die Aspekte „Verblindung von Teilnehmern und Studienpersonal“, „Verblindung der Endpunkterhebung“ und „unvollständige Daten zu Endpunkten“ separat für einzelne oder zusammenge-

fasste Endpunkte bewertet werden, wenn dies aufschlussreich ist oder die Ergebnisse für einen Endpunkt zu unterschiedlichen Zeitpunkten ermittelt wurden. Für die Domäne „andere Ursachen für Bias“ wird eine Gesamtbeurteilung empfohlen [6] (zum unterschiedlichen Verzerrungspotenzial subjektiv und objektiv erhobener Endpunkte s. oben). Werden in einer Publikation relevante Einzelheiten zur Einschätzung von Bias nicht ausreichend dargestellt, muss die Beurteilung des Bias-Risikos „unklar“ lauten [8].

Transparent werden die Bewertungen dadurch, dass jede Entscheidung durch ein Zitat aus dem Publikationstext oder durch einen Kommentar zu belegen ist. Es wird empfohlen, die Bewertungen der einzelnen Domänen je Studie in einer „Bias-Risiko-Tabelle“ darzustellen und auf detaillierte Angaben im Text zu verzichten (s. Anhang, Tab. 8.5.c). Zusätzlich verhilft die für das Tool vorgesehene graphische Darstellung zu einem Überblick über das gesamte Verzerrungspotential. Die Assoziation mit einer Verkehrsampel sorgt für weltweite Verständlichkeit (■ **Abb. 2**). Weiterführende Angaben sind dem Cochrane Handbook for Systematic Reviews of Interventions in den Abschn. 8.6 bis 8.17 zu entnehmen [6].

Berichtsqualität versus interne Validität

Zur klaren Abgrenzung zwischen Mängeln in der Berichtsqualität und tatsächlichen methodischen Schwächen von Studien startete eine Gruppe von Herausgebern wissenschaftlicher Zeitschriften, klinischen Forschern, Epidemiologen und Methodikern zu Beginn der 1990er-Jahre eine Initiative zur Verbesserung der Qualität von Berichten zu randomisierten kontrollierten Studien. Ergebnis war das CONSORT (Consolidated Standards of Reporting Trials)-Statement, eine Leitlinie für die Berichterstattung, die 1996 erstmals veröffentlicht und 2010 überarbeitet wurde [26]. Sie enthält eine Checkliste mit 25 Items zu wesentlichen Informationen, die in Publikationen über randomisierte Studien enthalten sein sollten, sowie das Modell eines Flussdiagramms, in dem Informationen über die

Anzahl von Patienten in vier Stadien einer klinischen Studie abgebildet sind. Bereits wenige Jahre nach Veröffentlichung des CONSORT-Statements nahm die Anzahl unklarer Berichte in drei hochrangigen Zeitschriften (*British Medical Journal*, *Journal of the American Medical Association*, *The Lancet*), die die obligatorische Verwendung der Checkliste vorsehen, signifikant ab [27].

Weitere Orientierungshilfen für Autoren, Herausgeber und Gutachter wurden in der Folge für unterschiedlichste Studientypen publiziert. Sie sind auf den Internetseiten des EQUATOR-Netzwerks abrufbar (<http://www.equator-network.org/>). Diese Checklisten tragen zu einer transparenteren Berichterstattung bei, die für Entscheidungsträger von großem Nutzen sein kann [28]. Aufgrund ihres unkritischen Gebrauchs muss jedoch entschieden darauf hingewiesen werden, dass Checklisten keinen Ersatz für Bewertungsinstrumente darstellen [29].

Schlussfolgerung und Ausblick

Unter den Bewertungsinstrumenten für kontrollierte Studien fällt das RoB-Tool der Cochrane Collaboration durch eine hohe Transparenz, Erfassung der wesentlichen Bias-Arten, gute Handhabung mit lediglich sieben Items und die weltweit verständliche graphische Ergebnisdarstellung auf. Hartling et al. [30] kamen in ihrer Untersuchung der Reliabilität allerdings zu dem Ergebnis, dass Unterschiede in den Bewertungen des Bias-Risikos auf Unterschiede in der Interpretation des Instruments zurückzuführen waren. Ihrer Einschätzung nach besteht Bedarf für eine Überarbeitung der Erläuterungen zur Anwendung des Instruments, die detaillierter formuliert werden sollten. Dem trägt die Cochrane Collaboration durch fortlaufende Diskussionen über die Methodik und regelmäßige Aktualisierung Rechnung.

Die vorliegende deutsche Fassung des RoB-Tools basiert auf einer von den drei deutschsprachigen Cochrane-Zentren konsentierten Übersetzung. Zur Verbesserung der Sprachgenauigkeit und Verständlichkeit für Leser außerhalb von Expertenkreisen wurde diese von einem unabhängigen Anbieter ins Englische zu-

rückübersetzt. Die Zusammenführung der Erstübersetzung, konsentierten Fassung und Rückübersetzung zu einer endgültigen Version ist an die Methoden von Beaton et al. [31] angelehnt. Ziel der Übersetzung ist es, das Bewertungsinstrument der Cochrane Collaboration auf Deutsch verfügbar zu machen. Die mangelnde Berücksichtigung der Qualität von Evidenz hat bei Expertenempfehlungen in der Vergangenheit zu großen Fehlern mit schwerwiegenden Folgen geführt. Zum Beispiel wurde im Fall der Hormonersatztherapie auf der Basis von Beobachtungsstudien mit inkonsistenten Ergebnissen fälschlicherweise davon ausgegangen, dass sie das Risiko für kardiovaskuläre Ereignisse reduzieren würde. Eine Dekade lang wurden viele postmenopausale Frauen falsch behandelt, bis in zwei randomisierten Studien gezeigt wurde, dass die Hormonersatztherapie nicht zu einer entsprechenden Risikoreduktion führt, sondern das Risiko für kardiovaskuläre Ereignisse möglicherweise sogar erhöht [32–34]. Als Beitrag zu einer rationalen Entscheidungsfindung in der klinischen Praxis und im Gesundheitswesen [35] ist ein weit verbreitetes Verständnis des Konzepts und der Methoden der EbM wünschenswert.

Anhang

Deutsche Übersetzung des Risk-of-Bias (RoB)-Tool (Kap. 8.5 aus dem Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0).

Korrespondenzadresse

Dr. B. Buchberger

Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Medizinmanagement
Universität Duisburg-Essen
Schützenbahn 70, 45127 Essen
barbara.buchberger@medman.uni-due.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. B. Buchberger, E. von Elm, G. Gartlehner, H. Huppertz, G. Antes, J. Wasem und J.J. Meerpohl geben an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel unterliegt den Bedingungen der Creative Commons Attribution Noncommercial License. Dadurch sind die nichtkommerzielle Nutzung, Verteilung und Reproduktion erlaubt, sofern der/die Originalautor/en und die Quelle angegeben sind.

Literatur

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312(7023):71–72
2. Antes G, Bassler D, Forster J (2003) Evidenz-basierte Medizin. Praxis-Handbuch für Verständnis und Anwendung der EBM, 1. Aufl. Thieme, Stuttgart
3. Du Prel JB, Röhrig B, Blettner M (2009) Kritisches Lesen wissenschaftlicher Artikel. *Dtsch Arztebl Int* 106(7):100–105
4. Röhrig B, du Prel JB, Blettner M (2009) Studiendesign in der medizinischen Forschung. *Dtsch Arztebl Int* 106(11):184–189
5. Jones DJ (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. <http://www.bmj.com/rapid-response/2011/11/03/risk-bias-assessment-should-be-taught-all-junior-doctors>. Zugegriffen: 20. Nov. 2013
6. Higgins JPT, Green S (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org. Zugegriffen: 18. November 2013
7. Jüni P, Altman DG, Egger M (2001) Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 323(7303):42–46
8. Higgins JP, Altman DG, Gøtzsche P, Cochrane Bias Methods Group, Cochrane Statistical Methods Group et al (2011) The Cochrane Collaborations's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928. doi:10.1136/bmj.d5928
9. Unnebrink K, Windeler J (2001) Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statist Med* 20(24):3931–3946
10. Sterne JAC, White IR, Carlin JB et al (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338:b2393
11. Carpenter JR, Kenward MG (2014) Missing data in clinical trials – a practical guide. Birmingham: National Health Service Co-ordinating Centre for Research Methodology. <http://www.missingdata.org.uk>. Zugegriffen: 20. Mai 2014
12. Savović J, Jones HE, Altman DG et al (2012) Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 157:429–438
13. Schulz KF, Grimes DA (2002) Blinding in randomized trials: hiding who got what. *Lancet* 359:696–700
14. Otto C, Schiffer G, Tjardes T, Kunter H, Eysel P, Paffrath T (2014) Blood loss and operative duration using monopolar electrosurgery versus ultrasound scissors for surgical preparation during thoracoscopic ventral spondylosis: results of a randomized, blinded, controlled trial. *Eur Spine J* 23:1783–1790
15. Wood L, Egger M, Gluud LL et al (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336(7644):601–605

16. Lundh A, Sismondo S, Lexchin J, Busuioac OA, Bero L (2012) Industry sponsorship and research outcome. *Cochrane Database Syst Rev* (12):MR000033. doi:10.1002/14651858.MR000033.pub2
17. Bero LA (2013) Why the Cochrane risk of bias tool should include funding source as a standard item [editorial]. *Cochrane Database Syst Rev* (12):ED000075
18. Sterne JAC (2013) Why the Cochrane risk of bias tool should not include funding source as a standard item [editorial]. *Cochrane Database Syst Rev* (12):ED000076
19. Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U (2010) Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum, Schriftenreihe Health Technology Assessment, Bd 102. Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), Köln
20. Moher D, Cook DJ, Jadad AR et al (1999) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess* 3(12):i-iv, 1-98
21. Jüni P, Witschi A, Bloch R, Egger M (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 282(11):1054-1060
22. Jadad AR, Moore RA, Carroll D et al (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 17(1):1-12
23. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A (2005) On behalf of the Metaquality Study Group. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 330(7499):1053
24. Lundh A, Gøtzsche PC (2008) Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 8:22. doi:10.1186/1471-288-8-22
25. Hróbjartsson A, Boutron I, Turner L, Altman DG, Moher D (2013) Assessing risk of bias in randomized clinical trials included in Cochrane reviews: the why is easy, the how is a challenge. *Cochrane Database Syst Rev* 4:ED000058. doi:10.1002/14651858.ED000058
26. Ziegler A, König IR (2011) Leitlinien für Forschungsberichte: Eine Anwendung des CONSORT 2010 Statements. *Dtsch Med Wochenschr* 136:e2-8
27. Turner L, Shamseer L, Altman DG et al (2012) Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 11:MR000030. doi:10.1002/14651858.MR000030.pub2
28. Meerpohl JJ, Blümle A, Antes G, von Elm E (2009) Leitlinien für Forschungsberichte sind auch für Leser medizinischer Fachartikel hilfreich. *Dtsch Med Wochenschr* 134:2078-2083
29. Da Costa BR, Cevallos M, Altman DG, Rutjes AW, Egger M (2011) Uses and misuses of the STROBE statement: bibliographic study. *BMJ Open* 1:e000048. doi:10.1136/bmjopen-2010-000048
30. Hartling L, Hamm MP, Milne A et al (2013) Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 66(9):973-981
31. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25(24):3186-3191
32. Guyatt GA, Oxman AD, Vist GE et al (2008) GRADE: an emerging consensus of evidence and strength of recommendations. *BMJ* 336(7650):924-926
33. Hulley S, Grady D, Bush T et al (1998) Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 280:605-613
34. Rossouw JE, Anderson GL, Prentice RL et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 288:321-333
35. Raspe H (2005) Konzept und Methoden der Evidenz-basierten Medizin: Besonderheiten, Stärken, Grenzen, Schwächen und Kritik. Institut für Sozialmedizin. Universitätsklinikum Schleswig-Holstein. http://www.ebm-netzwerk.de/site/was-ist-ebm/images/konzept_ebm_raspe.pdf. Zugriffen: 19. April 2014