Identifying the ‘incredible’! Part 1

assessing the risk of bias in outcomes included in systematic reviews

Büttner, Fionn; Winters, Marinus; Delahunt, Eamonn; Elbers, Roy; Lura, Carolina B; Khan, Karim M; Weir, Adam; Ardern, Clare L

Published in:
British Journal of Sports Medicine

DOI (link to publication from Publisher):
10.1136/bjsports-2019-100806

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Identifying the "incredible"! Part 1: assessing the risk of bias in outcomes included in systematic reviews

Editorial

Fionn Büttner[1], Marinus Winters[2], Eamonn Delahunt[1, 3], Roy Elbers[4], Carolina Bryne Lura[2], Karim M Khan[5], Adam Weir[6, 7, 8], Clare L. Ardern[9, 10, 11]

1 School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
2 Research Unit for General Practice in Aalborg, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
3 Institute for Sport and Health, University College Dublin, Dublin, Ireland
4 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
5 Centre for Hip Health and Mobility, University of British Columbia, Vancouver, British Columbia, Canada
6 Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar
7 Department of Orthopaedics, Erasmus MC University Medical Center for Groin Injuries, Rotterdam, The Netherlands
8 Sport medicine and exercise clinic Haarlem (SBK), Haarlem, The Netherlands
9 Division of Physiotherapy, Linköping University, Linköping, Sweden
10 School of Allied Health, La Trobe University, Melbourne, Australia
11 Division of Physiotherapy, Karolinska Institute, Stockholm, Sweden

Correspondence to:
Fionn Cléirigh Büttner
School of Public Health, Physiotherapy and Sports Science
University College Dublin
Belfield
Dublin 4
Ireland
E: fionn.cleirigh-buttner@ucdconnect.ie

Word count: 1241

Keywords: review [MeSH], meta-analysis [MeSH], bias [MeSH]

INTRODUCTION

Systematic reviews fulfil a vital role in modern medicine.(1) However, the results of systematic reviews are only as valid as the studies they include.(2) Pooling flawed, or biased, studies can compromise the credibility of systematic review findings. Bias is a systematic deviation in the true results of a research study that can manifest due to limitations in study design, conduct, or analysis.(3)

The results of sport and exercise medicine (SEM) research, like results in other fields, are vulnerable to bias.(4) It is important that systematic review authors assess for bias in a way that enables a judgement about whether a review outcome is at risk of bias due to methodological limitations in included studies. This two-part education primer focuses on how systematic review authors can perform and interpret risk of bias assessments to avoid misleading systematic review conclusions. In this editorial, we introduce the concept of risk of bias, and the principles of assessing risk of bias.

BIAS: THE BASICS

Different biases have effects that vary in direction and magnitude.(3,5) It is challenging to precisely determine how bias over- or under-estimates an individual study's true findings. In fact, bias does not always result in distorted study findings and one can never be certain that bias is present when a study has methodological limitations. However, methodological limitations in study design, conduct, or analysis can be consistently associated with inflated research findings.(5) Due to this uncertainty, study outcomes are considered to be at *risk* of bias rather than 'biased'.

Studies with 'some concerns' or 'high' risk of bias in design, conduct, analysis, or reporting are at greater risk of inflated findings compared to studies at 'low' risk of bias, negatively affecting the probability that study findings accurately reflect reality.(5,6) Assessing the risk of bias of study outcomes that are included in a systematic review allows readers to interpret the credibility of review findings.

DON'T CONFUSE RISK OF BIAS WITH STUDY QUALITY

Risk of bias is a clearly-defined term and refers to the perceived risk that the results of a research study deviate from the truth.(3) Unfortunately, risk of bias is often conflated with study quality, despite being distinct constructs (Table 1).

Study quality is a vague and multi-dimensional term that loosely indicates how closely a research study is conducted to the highest possible methodological standards.(3) Quality refers to several areas of study methodology, with each area having different implications for how one should interpret a study's methodological rigor (Table 1).(7) A risk of bias assessment should not be replaced by an assessment of study quality.(8) When critically appraising a research study, assessors should prioritise how closely a study's findings approximate the truth (i.e., risk of bias) over how well the study was conducted (i.e., quality) (Table 1).

USE DOMAIN-BASED RISK OF BIAS ASSESMSENT TOOLS INSTEAD OF QUALITY SCALES AND CHECKLISTS

A plethora of assessment tools are available to critically appraise a research study.(9) However, not all of these tools are appropriate to assess risk of bias. This can confuse researchers about which tool is the most suitable tool to use. Broadly, three types of tools exist to assist researchers and readers in critically appraising a study: (1) quality scales, (2) quality checklists, and (3) domain-based risk of bias tools.(3) We explain why domain-based risk of bias tools are preferred over quality scales and checklists.

Quality scales and quality checklists vary substantially in content, complexity, and rating criteria, and often include items that are not related to bias.(10) Quality scales assign numeric values to scale items and combine information about several methodological features in a study to produce a summary score.(9) For example, the PEDro scale includes items related to internal validity (e.g., random allocation) and reporting (e.g., clear description of participant eligibility criteria). A lack of a random allocation undermines the credibility of a study's findings (i.e., there is a 'high risk' of selection bias).

Unclear eligibility criteria challenge a study's reproducibility and make it difficult to judge to whom the study findings are applicable (i.e., external validity). In the presence of good reporting but poor methodological conduct, quality assessment may overestimate the credibility of study findings.(11,12)

Quality checklists contain items that relate to study quality without assigning numeric values or producing a summary score.(3,9) For example, the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies contains items relating to reporting, sample size, statistical power, precision, external validity, and internal validity (bias); requiring "yes", "no", or "other" responses to each item. Such quality checklist items do not solely address risk of bias and are not intended to be summed to produce one numeric score. However, review authors frequently modify quality checklists (by assigning arbitrary numeric values) to generate summary scores and summarise study quality. Summary scores do not inform the reader *which* biases might be present.(12)

Using quality scales and quality checklists is discouraged because different scales tend to generate conflicting conclusions when applied to the same studies.(11) Quality scales are also prone to misleading conclusions when relying on cut-off thresholds that arbitrarily categorize study quality as 'high', 'moderate', or 'low'.(13)

Domain-based risk of bias assessment tools are currently the commonly accepted and preferred method to judge the credibility of study findings.(3) Domain-based tools evaluate study limitations in specific domains that represent different biases (e.g., bias arising from the randomization process).(3,5) Domain-based tools overcome many shortcomings of quality scales, as they evaluate individual components that relate to study design, conduct, and analysis rather than a single summary score.(14) Several study design-specific, domain-based risk of bias assessment tools have been developed.(15–20) The Cochrane Risk of Bias tool 2 (RoB2) is a rigorously developed, domain-based risk of bias assessment tool that assesses the limitations of randomised controlled trials across five bias domains.(21) Each bias domain possesses strong empirical evidence that study limitations may distort study findings.(5)

RISK OF BIAS ASSESSMENT METHOD

Risk of bias assessments should be performed for each outcome of interest rather than as one general assessment for each study.(22) If a study includes multiple outcomes and time-points, separate risk of bias assessments should be undertaken for each included outcome (Table 2). Bias can impact review outcomes differently,(5,22) underscoring the need for separate risk of bias assessments when multiple outcomes are reported upon (Table 2). Cochrane recommends two approaches to risk of bias assessments.(3) Both approaches involve a domain-based risk of bias assessment of separate outcomes, assessing:

(1) Individual review outcomes, in each individual study, based on individual risk of bias domains.

(2) Individual review outcomes, across included studies (i.e., meta-analysis level), based on individual risk of bias domains.

In part two, we demonstrate both risk of bias assessment methods.

SUMMARY

In this editorial, we introduced risk of bias as the perceived risk that the results of a research study may under- or over-estimate the truth. Systematic review authors should perform a domain-based risk of bias assessment that reflects risk of bias instead of assessing study quality. If a research study reports upon multiple outcome measures, separate risk of bias assessments should be performed for each outcome measure.

In part 2 of this risk of bias education primer, we:

1. Evaluate the prevalence and methods of risk of bias assessments in systematic reviews published in *BJSM*.
2. Perform a risk of bias assessment on a sample of RCTs in a systematic review.
3. Illustrate the impact that different critical assessment tools have on risk of bias assessment findings, and ultimately, systematic review findings.
4. Provide recommendations for systematic review authors undertaking risk of bias assessments.

Table 1 – Key terms relating to risk of bias and critical appraisal

Key terms	Explanation
Risk of bias	Bias is a systematic deviation from the truth in the results of a research study. Bias may occur due to limitations in study design, conduct, analysis, or reporting.(3) Bias is associated with under-estimated or over-estimated study findings. Multiple sources of bias exist, and different biases can vary in direction and magnitude. Assessing bias can never conclusively determine whether a study's findings under-estimate or over-estimate a true result because study findings can be unbiased despite methodological limitations. Therefore, *risk of bias*, rather than bias, is assessed to determine the likelihood that bias is present. Risk of bias is synonymous with the term internal validity.
Study quality	Study quality is the extent to which a study is conducted to the highest methodological standards possible. Study quality evaluates multiple constructs of study methodology including reporting completeness, ethical approval, statistical power, precision, and internal and external validity.(7,9)
Risk of bias & study quality	The terms 'risk of bias' and 'study quality' are often used interchangeably. However, both terms are distinct constructs. Discrepancies between study quality and risk of bias are highlighted when performing a risk of bias assessment. Blinding participants in RCTs can be challenging and often impossible in SEM research (e.g., randomising professional football players to receive a Nordic hamstring exercise programme, or not). A RCT that cannot blind participants might be considered high-quality because it may be the only way for trial investigators to conduct such a RCT. However, risk of bias targets the extent to which study findings should be believed, irrespective of researchers' (in)ability to prevent methodological shortcomings that may affect study findings. Because participants were not blinded, the trial outcome is at 'high' risk of bias – this fact is inescapable.(3)
Reporting quality	Reporting quality refers to the extent to which an original research article provides complete and transparent information about the design, conduct, analysis, and results of a study. Complete reporting facilitates a comprehensive assessment of a study's internal and external validity, and study design-specific reporting guidelines exist to guide systematic review reporting.(23) Good- or poor-quality reporting in a study does not imply that the study's outcomes are at 'low' or 'high' risk of bias, respectively.(24) For example, there is a difference between reporting whether a methodological procedure, such as randomization, was performed, and whether it was performed appropriately to sufficiently minimize risk of bias (e.g., by using simple randomization from a computer-generated random numbers table, with an equal allocation ratio).

Table 2 – Concepts in the assessment of risk of bias

Concept	Explanation
Domain-based risk of bias assessments	Domain-based risk of bias assessments identify specific study limitations that can introduce different biases (e.g., bias arising from the randomization process or bias due to deviation from intended interventions). Specific risk of bias domains provide insight into why a study outcome might be distorted, and by how much. For example, RCTs at 'high' risk of bias due to inadequate allocation concealment will be associated, on average, with over-estimated trial outcomes in favour of the experimental group compared to RCTs at 'low' risk of bias.(5)
Assessing review outcomes separately.	Study limitations that inform judgements of 'some concerns' or 'high' risk of bias can distort outcome measures differently.(5) For example, pain is more likely to be over-estimated when a patient is aware of their allocation to a specific intervention group (due to lack of patient blinding) than if they were not aware of their group allocation.(5,22) Conversely, a patient's awareness of their allocation to an intervention group is less likely to influence an outcome such as re-injury.(5,22) Systematic review authors should perform separate risk of bias assessments for each outcome rather than assessing all review outcomes at once with one, general risk of bias assessment. A domain-based risk of bias assessment for separate outcomes evaluates the judgements of each risk of bias domain for separate review outcome types.

References

1. Mulrow CD. The medical review article: state of the science. Ann Intern Med. 1987 Mar;106(3):485–8.

2. Weir A, Rabia S, Ardern C. Trusting systematic reviews and meta-analyses: all that glitters is not gold! Br J Sports Med. 2016 Sep;50(18):1100–1.

3. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Wiley; 2008. 672 p.

4. Bandholm T, Henriksen M, Thorborg K. Slow down to strengthen sport and exercise medicine research. Br J Sports Med. 2017 Oct 1;51(20):1453–1453.

5. Page MJ, Higgins JPT, Clayton G, et al. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. PLOS ONE. 2016 Jul 11;11(7):e0159267.

6. Turner L, Boutron I, Hróbjartsson A, et al. The evolution of assessing bias in Cochrane systematic reviews of interventions: celebrating methodological contributions of the Cochrane Collaboration. Syst Rev. 2013 Sep 23;2:79.

7. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol. 2004 Sep 16;4(1):22.

8. Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. J Clin Epidemiol Elmsford. 2012 Mar;65(3):236–8.

9. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. Control Clin Trials. 1995 Feb 1;16(1):62–73.

10. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. BMJ. 2001 Jul 7;323(7303):42–6.

11. Jüni P, Witschi A, Bloch R, et al. The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis. JAMA. 1999 Sep 15;282(11):1054–60.

12. da Costa BR, Hilfiker R, Egger M. PEDro's bias: summary quality scores should not be used in meta-analysis. J Clin Epidemiol. 2013 Jan;66(1):75–7.

13. Detsky AS, Naylor CD, O'Rourke K, et al. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol. 1992 Mar 1;45(3):255–65.

14. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostat Oxf Engl. 2001 Dec;2(4):463–71.

15. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

16. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. The BMJ [Internet]. 2016 Oct 12 [cited 2017 Jun 1];355. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5062054/

15. Whiting P, Savović J, Higgins JPT, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016 Jan 1;69(Supplement C):225–34.

18. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 2013 Feb 19;158(4):280.

19. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006 Mar 6;6(1):9.

20. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1–33.

21. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug 28;366:l4898.

22. 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 Mar 13;336(7644):601–5.

23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. The BMJ [Internet]. 2009 Jul 21 [cited 2017 Jul 13];339. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2714672/

24. Soares HP, Daniels S, Kumar A, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ. 2004 Jan 1;328(7430):22–4.