

KEY CONCEPTS IN CLINICAL EPIDEMIOLOGY

Quality versus Risk-of-Bias assessment in clinical research

Luis Furuya-Kanamori^a, Chang Xu^b, Syed Shahzad Hasan^c, Suhail A. Doi^{b,*}

^aResearch School of Population Health, College of Health and Medicine, Australian National University, Canberra, Australia

^bDepartment of Population Medicine, College of Medicine, QU Health, Qatar University, Doha, Qatar

^cDepartment of Pharmacy, University of Huddersfield, Huddersfield, UK

Accepted 23 September 2020

Abstract

Assessment of internal validity safeguards implemented by researchers has been used to examine the potential reliability of evidence generated within a study. These safeguards protect against systematic error, and such an assessment has traditionally been called a *quality assessment*. When the results of a quality assessment are translated through some empirical construct to the potential risk of bias, this has been termed a *risk of bias assessment*. The latter has gained popularity and is commonly used interchangeably with the term *quality assessment*. This key concept paper clarifies the differences between these assessments and how they may be used and interpreted when assessing clinical evidence for internal validity. © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Bias assessment; Risk of bias; Quality assessment; Internal validity; systematic error

1. Introduction

Bias negatively impacts clinical epidemiologic research, and to prevent this requires the placement of methodological safeguards within research studies. Such safeguards may be within the design, conduct, and analysis of a study. Before a body of research can get its results embedded into evidence-based practice, there should be a *quality assessment* for the presence of such safeguards against bias in the various included studies on the topic to ensure the credibility of research results, and this is formally undertaken using a quality assessment. More recently, this assessment has also been called the *risk of bias assessment*. The distinction between them is that:

a) quality assessment is the assessment of the inclusion of methodological safeguards within a study, and

b) risk of bias assessment concerns the implication of the inclusion of such safeguards for study results.

This distinction is not just semantic and reflects differences related to how the tool is intended to be used. For example, the *quality assessment* also deals with the number or type of safeguards present and does not include signaling questions and criteria for judgments, both of which are required for *risk of bias* assessment as the latter requires a subjective degree of risk as the outcome of the assessment. The Cochrane collaboration had previously suggested a similar distinction suggesting that quality assessment should refer to the extent to which a study was designed, conducted, analyzed, interpreted, and reported to avoid systematic errors, while risk of bias assessment should refer to what flaws in the design, conduct, and analysis affect the study results [1]. It is important to note that the terms “quality” and “risk of bias” have been used interchangeably in the epidemiologic literature to describe the methodological conditions associated with the validity of study results and at times, purely reporting items have crept into quality assessments that have no bearing on systematic error [2]. Some researchers have taken this to suggest that quality assessment should refer to reporting such items [3] irrespective of whether these are safeguards against bias or not, and this is incorrect. We should henceforth

Funding: This work was made possible by Program Grant #NPRP10-0129-170274 from the Qatar National Research Fund (a member of Qatar Foundation) to Suhail Doi. The findings herein reflect the work and are solely the responsibility of the authors.

Conflicts of interest: There are no conflicts of interest.

* Correspondence to. Department of Population Medicine, College of Medicine Qatar University Drive, P.O. Box 2713, Doha, Qatar. Tel.: +974 66001271; fax: +974 44037854.

E-mail address: s.doi@gmx.net (S.A. Doi).

<https://doi.org/10.1016/j.jclinepi.2020.09.044>

0895-4356/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

use the term *quality assessment* when we refer to the measurement of the extent that methodological safeguards against bias have been implemented and *risk of bias assessment* when we refer to bias judgments based on such *quality assessment*. Judgments are usually made in terms of low/high risk of bias or high/low quality of a study, which are terms that have also been used interchangeably. Reporting checklists (and there are many) must be distinguished from a quality assessment tool, checklist or scale.

All evidence syntheses require that such assessment be formally done of the included studies so that users of such research are cognizant of how much the results can be trusted. When a meta-analysis is included, the quantitative results should also be interpreted in the light of either the *quality assessment* or *risk of bias assessment* of included studies. This can be taken a step further, and these assessments (particularly the *quality assessment*) can be used for bias adjustment of meta-analysis results [4].

These assessments require tools that include a list of safeguard items likely to influence internal validity (*quality assessment*) or estimated intervention effects (*risk of bias assessment*) that can be looked for in the published research. By and large, these safeguard items are the same, and the only reason why they may be labeled by either term is the intent of the particular assessment. Of note, a study that implemented all safeguards listed in a tool may not necessarily be unbiased, and one that applied none is not necessarily biased, and this applies to both assessment types.

The final point is that the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) group [5] uses the term “quality of evidence” to reflect the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation. Therefore, in addition to the *risk of bias assessment* of individual studies making up the body of evidence, four other issues are rated under the “quality of evidence” umbrella: inconsistency, indirectness, imprecision, and publication bias. The latter is

not actually a *quality assessment* in terms of what is meant by this term for individual study assessment of internal validity and uses “quality” in a more generic sense making this a distinct concept from what we mean when we undertake a *quality assessment* of individual studies. Of note, GRADE incorporates a *risk of bias assessment* within this generic quality construct.

2. Application

The use of either type of tool, despite similar content, differs slightly. The best use of a *quality assessment* tool is to enumerate the safeguards implemented in a study as a count. We are avoiding the use of the word score here because it has received bad publicity [6] although not for the right reasons [7]. Once the safeguards are counted, stratification of studies for subgroup analysis by such counts is not recommended [7]. There remains no quantitative utility for risk of bias judgments (apart from their descriptive use) when we take away stratification, and all bias adjustment methods [4] require a *quality assessment* as input. Such input requires conversion of the counts of safeguards from a *quality assessment* into a relative ranking of studies that can then be examined analytically [4]. Such ranking is achieved by dividing all study counts by the highest count in the group of studies; thus, study ranks vary from 1 downwards to zero. Again, a study of rank 1 is not necessarily devoid of bias, and *quality assessment* can only help with the relative assessment of studies. Bias quantification is not possible from a *quality assessment*. An example of a study ranking is given in Table 1.

It should be noted that the (seemingly “arbitrary”) application of rules to *quality assessment* tools turns them into the *risk of bias* assessment tools. The underlying safeguards are the same, but the *risk of bias* assessment tools needs an empirical construct for making the hierarchal risk of bias judgments (high/low risk which is equivalent to low/high quality). These constructs can vary, such as missing important items to the number of missing items

Table 1. An example of quality and risk of bias assessments

Study	Randomized	±1 ^a	Double blind	±1 ^a	Withdrawals and dropouts	Safeguard counts	Judgment (rule: >3 = high quality/low risk of bias)	Quality rank ^b
Al-Sunaidi 2007	1	1	0	0	1	3	Low	0.6
Cicinelli 1998	1	1	1	1	1	5	High	1
Giorda 2000	1	1	0	0	1	3	Low	0.6
Lau 1999	1	1	1	1	1	5	High	1
Vercellini 1994	1	1	0	0	1	3	Low	0.6

^a ±1 refers to an adequate description of the safeguard in the previous column.

^b Quality rank was calculated through the safeguard counts of each study divided the maximum count across studies.

Key points

- *Quality assessment* and *risk of bias assessment* have been used interchangeably in the epidemiologic literature.
- All assessment tools consist of a set of methodological safeguard items to be checked against the study for their presence in research (as reported).
- A *risk of bias assessment* tool has an additional empirical construct regarding how nonimplementation of safeguard(s) might possibly have impacted bias of study results (high/low-risk or low/high-quality studies).
- Such judgments must not be used to stratify studies for subgroup analysis as this induces a selection bias.
- At a minimum, *quality assessment* should be related to study results qualitatively or quantitatively.
- Bias-adjustment is one way of including such assessments into study results quantitatively.

or domains with missing items depending on the tool authors' judgments regarding possible influence on estimates of effect. One disadvantage of this process is that assessments are tool dependent and consistency (as opposed to agreement) is not an expectation as we move from tool to tool. An example of a risk of bias judgment is also given in Table 1. Of note, researchers can use a *risk of bias assessment* tool for quality assessment but not vice versa unless the empirical construct to be used for such judgments is first defined.

A study reported in the BMJ [8] illustrates these points. The authors used the Jadad quality scale to assess the quality of five randomized controlled trials of paracervical anesthesia for pain control and considered six safeguards through the tool. They then checked the studies for the presence or absence of safeguards denoted 1 and zero respectively and counted how many were present

(Table 1). Finally, they also created their own rule for *risk of bias assessment* (>3 safeguards = low risk or high quality). We added the column on quality rank for illustration. The latter can be used for quantitative bias adjustment [4], and this is illustrated in Fig. 1. Of note, instead of the Jadad quality scale, researchers in such a study may now use the RoB 2 *risk of bias* tool from Cochrane, which has explicit albeit equally “arbitrary” rules for arriving at the judgment that obviates the creation of one by the authors as was done with this example.

CRedit authorship contribution statement

Luis Furuya-Kanamori: Writing - review & editing. **Chang Xu:** Writing - review & editing. **Syed Shahzad Hassan:** Writing - review & editing. **Suhail A. Doi:** Conceptualization, Methodology, Writing - original draft, Funding acquisition.

References

- [1] Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi M, Minozzi S. Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. *J Clin Epidemiol* 2018;99:24–32.
- [2] Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Seida JK, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.
- [3] Kamper SJ. Risk of bias and study quality assessment: Linking evidence to practice. *J Orthop Sports Phys Ther* 2020;50:277–9.
- [4] Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *J Clin Epidemiol* 2019;117:36–45.
- [5] Schünemann H, Brozek J, Guyatt G, Oxman AI: GRADE handbook for grading quality of evidence and strength of recommendations, 2013. The GRADE Working Group; 2013.
- [6] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
- [7] Stone J, Gurunathan U, Glass K, Munn Z, Tugwell P, Doi SAR. Stratification by quality induced selection bias in a meta-analysis of clinical trials. *J Clin Epidemiol* 2019;107:51–9.
- [8] Cooper NA, Khan KS, Clark TJ. Local anaesthesia for pain control during outpatient hysteroscopy: systematic review and meta-analysis. *BMJ* 2010;340:c1130.

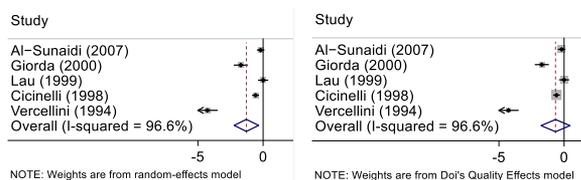


Fig. 1. Effects of bias adjustment via quality ranks on the results of the example meta-analysis of the weighted mean difference in pain scores after local anaesthesia for hysteroscopy. The left panel depicts a random-effects analysis and the right panel a bias-adjusted analysis.

Further Reading

- [9] Stone J, Gurunathan U, Glass K, Munn Z, Tugwell P, Doi SAR. Stratification by quality induced selection bias in a meta-analysis of clinical trials. *J Clin Epidemiol* 2019;107:51–9. [This paper explains why the use of stratification by a quality/risk of bias judgment is not good practice in clinical epidemiology].
- [10] Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *J Clin Epidemiol* 2019;117:36–45. [This paper explains what is meant by bias

adjustment and what the role of quality assessment could be in this process].

- [11] Whiting P, Savović J. The rationale for rating risk of bias should be fully reported: response. *J Clin Epidemiol* 2016;76:239. [This letter highlights the subjectivity of judgments inherent in use of a tool for risk of bias assessment and outlines the thinking of risk of bias developers regarding the difference between risk of bias and quality assessment].

- [12] Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6. [This paper reviews the old thinking around safeguard counts and their hazards in application to practice and may be useful for understanding the historical development in this area. Many of these principles have now been superseded by progress in this area (see the first and second recommended readings above)].