Journal of Clinical Epidemiology 81 (2017) 72–76

Journal of Clinical Epidemiology

There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review

Vanessa M.B. Jordan*, Sarah F. Lensen, Cynthia M. Farquhar

Department Obstetrics and Gynaecology, Auckland University, Private Bag 92019, Auckland 1142, New Zealand

Accepted 30 August 2016; Published online 10 September 2016

Abstract

Objectives: To assess the consistency in risk of bias (RoB) judgments across Cochrane reviews for studies appearing in more than one Cochrane review in the field of subfertility.

Study Design and Setting: We retrieved any study that had been used more than once in systematic reviews present on the Cochrane Database of Systematic Reviews in the area of subfertility. We then retrieved the recorded RoB assessments for these studies and looked at the consistency of judgments made between different authoring teams on the same trials.

Results: From the 156 bias judgments that were completed by at least two separate groups of authors, 45% of these judgments differed. For the domains of random sequence generation and incomplete outcome data, there was reasonably high level of agreement (71% and 79%, respectively). However, for the domain of blinding, agreement was reached in only 35% of cases.

Conclusion: This assessment of how consistently the RoB is being applied in Cochrane reviews has shown that, especially in some domains, there are large discrepancies in how RoB is being evaluated. Further work needs to be undertaken to improve the application of this tool.

Keywords: Risk of bias; Internal validity; Reliability; Systematic reviews; Meta-analysis; Cochrane; Quality

1. Introduction

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was developed as a list of essential items for reporting of systematic reviews [1]. One of the items included on this checklist concerns the assessment of risk of bias (RoB) of each of the included studies within a systematic review. The purpose of assessing RoB is to report on the strength of the evidence that has been collated, both narratively and within the meta-analysis [2]. This can be done by restricting the primary analysis only to those studies with low RoB or by stratifying the analysis based on RoB, allowing readers to interpret this information [3].

The method for assessing RoB for randomized controlled trials (RCTs) recommended by Cochrane is the Cochrane Collaboration's RoB tool [4]. This tool was first published in 2008 and was updated in 2010. This RoB tool is a domain-based evaluation. It requires two authors independently to extract information across seven domains. These are random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. For each of these domains, authors are requested to make a judgment of low RoB, unclear RoB, or high RoB and then to support these judgments with text from the original trials or correspondence with primary trial authors [4].

Hartling et al. [5] undertook an assessment of the Cochrane collaboration RoB tool and reported a variation of interrater agreement of the seven domains that ranged from substantial agreement for the random sequence generation ($k = 0.74$) to slight agreement for the selective reporting domain ($k = 0.13$). The developers of the tool have acknowledged that there can be difficulties in assessing the level of RoB, particularly for the domains of incomplete outcome data and selective reporting of outcomes [6]. It was considered that the development of author guidance and training materials may increase the agreement levels between independent author judgments [3]. However, a

Conflict of interest: None of the above authors have any financial conflicts of interest with regard to this work. All authors are members are associated with the Cochrane Collaboration.

* Corresponding author. Tel.: +64 9 9239490; fax: +64 9 303 5969.
E-mail address: v.jordan@auckland.ac.nz (V.M.B. Jordan).

http://dx.doi.org/10.1016/j.jclinepi.2016.08.012
0895-4356/

What is new?

Key findings

- The Cochrane RoB tool shows a lack of consistency when utilized in systematic reviews.
- Judging RoB for sequence generation showed the highest level of consistency when using the Cochrane RoB tool.
- Judging RoB for blinding showed the lowest level of agreement using the Cochrane RoB tool.
- In order to gain consistency of RoB judgement the Cochrane RoB tool needs to be improved.

more recent study has also shown that after training on the use of the RoB tool, there was only fair agreement for most of the domains (k = 0.24–0.37) with the exception of sequence generation which again showed substantial agreement (k = 0.79) [7]. Therefore, we were interested to evaluate the use of the RoB tool by authoring teams that had not been trained specifically to use the RoB tool as part of a wider RoB evaluation project. The objective of the study was to assess the consistency in RoB judgments across Cochrane reviews for studies appearing in more than one Cochrane review in the field of subfertility.

2. Methods

2.1. Study selection

In 2013, the specialized register for the Cochrane Gynecology and Fertility group was searched to locate all the systematic reviews in the field of subfertility. For each of the identified reviews, a list of all their included studies was compiled in an Excel (Version 14, Microsoft Office Professional Plus 2010, USA) spreadsheet.

Once duplicate studies had been identified (the same study appearing in at least two reviews), judgments and supporting comments were extracted from the RoB tables for five of the required ROB domains: random sequence generation, allocation concealment, blinding of participants, personnel and outcome assessment, incomplete outcome data, and selective reporting. If judgments were coded with the previous Cochrane convention of yes, no, or unclear, these were recoded with low risk = yes, unclear = unclear, and high risk = no. Agreement was if the judgment was identical within the two reviews, and agreement with regard to the supporting information was if the same text from the original article had been extracted in both reviews.

We also noted the date of first publication to determine if the differing versions of RevMan (Version 5.3, Copenhagen, Norway 2014), the statistical software used to create Cochrane systematic reviews [8], had influenced the judgment differences found.

3. Results

Eighty-four Cochrane systematic reviews in the field of subfertility were identified and retrieved from the Cochrane Menstrual Disorders and Subfertility register. There were 981 included studies over the 84 reviews. We identified 46 duplicates in 34 systematic reviews. Not all systematic reviews assessed all five RoB domains for their included studies. The only domain with judgments within all 46 duplicates was allocation concealment. Overall, there were 156 bias judgments that were completed by at least two separate groups of authors and appeared in systematic reviews present on the Cochrane Library. In 45% of cases, these judgments differed. As the topic of these reviews was infertility, the primary outcome for all the reviews was identical, that of live birth.

3.1. Random sequence generation

There were 28 duplicates that had judgments within more than one systematic review in the random sequence generation domain. Of the judgments, 71% were in agreement, and of these, 85% were rated as low risk by both review teams (Table 1). The remaining eight (29%) assessments that were in disagreement were a combination of low risk vs. unclear risk. In this domain, the support information given for reviewers judgments differed for each of the eight contradictory assessments (Table 2).

3.2. Allocation concealment

All 46 duplicated studies had judgments on allocation concealment. Of these judgments 59% showed agreement across the different systematic reviews (Table 1). Fourteen (50%) duplicates rated the studies as low risk, 12 (44%) as unclear risk, and 1 (4%) as high risk. Of the remaining duplicates, 41% differed in their ROB judgments (Table 1). From these, 15 (79%) duplicates differed between choosing low risk and unclear, and 4 (21%) duplicates differed between unclear and high risk. There were no judgments where low risk and high risk were graded for the same study. The support information given for each of the

Table 1. Judgment recorded in RoB

Risk of bias domain	Judgments that agree/total number (%)	Judgments that disagree/total number (%)
Random sequence generation	20/28 (71)	8/28 (29)
Allocation concealment	27/46 (59)	19/46 (41)
Blinding	11/31 (35)	20/31 (65)
Incomplete outcome data	16/23 (70)	7/23 (30)
Selective reporting	12/28 (43)	16/28 (57)

Abbreviation: RoB, risk of bias.