External factors may influence Cochrane reviewers when classifying the risk of bias of original reports

Lorenzo Bertizzolo\textsuperscript{a,b,c,*}, Patrick M. Bossuyt\textsuperscript{c}, Ignacio Atal\textsuperscript{a,b,d,e}, Philippe Ravaud\textsuperscript{a,b,d,f,g,h}, Agnès Dechartres\textsuperscript{i}

\textsuperscript{a}INSERM, U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), Methods of Therapeutic Evaluation of Chronic Diseases Team (METHODS), Paris F-75004, France

\textsuperscript{b}Paris Descartes University, Sorbonne Paris Cité, France

\textsuperscript{c}Department of Clinical Epidemiology and Biostatistics, Academic University Medical Centers, University of Amsterdam, Amsterdam, the Netherlands

\textsuperscript{d}Cochrane France, Paris, France

\textsuperscript{e}Centre de Recherches Interdisciplinaires (CRI), Université Paris Descartes, University of Paris, Paris, France

\textsuperscript{f}Centre d’Épidémiologie Clinique, Hôpital Hôtel Dieu AP-HP (Assistance Publique - Hôpitaux de Paris), Paris, France

\textsuperscript{g}Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

\textsuperscript{h}Department of Epidemiology, Columbia University, Mailman School of Public Health, New York, USA

\textsuperscript{i}Sorbonne Université, Institut National de la Santé et de la Recherche Médicale (INSERM), Institut Pierre Louis d’Épidémiologie et de Santé Publique, AP-HP, Sorbonne Université, Hôpital Pitié Salpêtrière, Département de Santé Publique, Centre de Pharmacopédiatrie de l’AP-HP (Cephepi), CIC-1422, F75013, Paris, France

Accepted 12 March 2020; Published online 19 March 2020

Abstract

Objective: The objective of the study was to explore contextual factors associated with high or low risk-of-bias judgment in case of incomplete or unclear information in study reports.

Study Design and Setting: Research-on-research study, using matched case–control design, with a sample of 304 randomized controlled trials (RCTs) included in two Cochrane reviews for which there was disagreement on the risk-of-bias judgment related to incomplete or unclear information in the study report. A case was defined as an RCT judged at high or low risk of bias; a control was the same RCT judged at unclear risk. We used a conditional logistic regression model for analysis.

Results: Review authors being also authors of the RCT were more likely to assess an item at low risk of bias than unclear (OR: 11.71; 95% CI: 4.58–30.22). Earlier trials in a review were more often assigned a low risk (OR: 0.37; [0.15–0.89]). Review groups and authors that had completed a lower number of reviews slightly more often assigned a low risk, whereas others reported “unclear” (OR: 0.97, [95% CI: 0.95–0.99]) for groups and 0.97 (95% CI: 0.95–0.99) for authors.

Conclusions: Risk-of-bias assessment of RCTs in case of incomplete or unclear information may be affected by contextual factors.

Keywords: Risk of bias; Systematic reviews; Research on research; Reporting; Matched case-control

Conflict of interest: None.

Transparency declaration: The guarantor (Lorenzo Bertizzolo) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Ethics approval: Not applicable. This is a research on research study.

Lorenzo Bertizzolo is the guarantor. He had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding: This project received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 676207. The funding source had no involvement in any part of the study.

Data sharing: Raw data and analyses are available on request from the authors.

* Corresponding author. INSERM, U1153 - Centre d’Épidémiologie Clinique, Hôtel Hôtel-Dieu, 1, place du parvis Notre-Dame, 75004 Paris, France. Tel.: +33 (0)1 42 34 78 25; fax: +33 (0) 142 160 332.

E-mail address: lorenzo.bertizzolo@gmail.com (L. Bertizzolo).
What is new?

Key findings
- If review authors are authors of the RCT, they are more likely to consider the item at low risk of bias than unclear.
- Within a review, authors may differently judge a study that is more recent and larger than the other included studies.

What this adds to what was known?
- This is the first study exploring review, author, and study characteristics that may affect risk-of-bias assessment of a randomized controlled trial in case of incomplete or unclear information in the trial report.

What is the implication and what should change now?
- Authors of systematic reviews should be aware that their risk-of-bias judgment can be affected by factors that are not specific to the study itself.

1. Introduction

Evaluation of risk of bias in individual studies is an essential part of the systematic review process [1–3], which allows researchers to determine how strongly to make conclusions [1–3]. Cochrane has developed a tool, the Risk of Bias (RoB) tool, for assessing risk of bias in randomized controlled trials (RCTs) [4]. The tool has become the reference method for evaluating risk of bias [5]. This tool examines trial characteristics that are organized in seven items: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. Each trial is evaluated at “high,” “low,” or “unclear” risk of bias on each of these seven items.

Previous studies showed a lack of reproducibility of assessments with the RoB tool [6–14]. Some recent studies also suggested that the support for judgment, and therefore the assessments themselves, frequently do not agree with recommendations from the Cochrane Handbook [15,16]. In a previous project [17], we evaluated the reasons for disagreements in risk-of-bias assessment of trials included in several Cochrane reviews. Our results confirmed suboptimal agreement in risk-of-bias assessment, ranging from 81% agreement for random sequence generation to 57% for incomplete outcome data. More than two-thirds of disagreements could be related to differences in interpretation of the same information and these disagreements were mostly due to incomplete or unclear information in the published trial report [17].

Although an “unclear” judgment can be expected in case of incomplete or unclear information, many authors may feel confident enough to consider the information adequate to express a judgment of low or high risk of bias. Erroneous decisions apart, the RoB tool clearly leaves ample space for personal interpretation and judgment of evidence from study reports.

We wondered whether contextual factors, in particular identifiable characteristics of the review team or the review itself, might be associated with differences in risk-of-bias judgments when evaluating the same incomplete or unclear information. These characteristics could be related to the reviewers, the review, or the study itself, when compared with the other studies included in the review. For example, authors may consider studies that are larger or newer differently than smaller or older studies, or different review groups may have different approaches for handling incompleteness of information.

We aimed to explore which factors, if any, are associated with high or low risk-of-bias judgment in case of incomplete or unclear information in reports of primary studies for which an “unclear” risk-of-bias judgment would have been expected. We evaluated these by using a sample of Cochrane reviews.

2. Methods

This was a research-on-research study on risk-of-bias assessment that used a large collection of published Cochrane systematic reviews.

2.1. Study design

The study followed a matched case–control design with a sample of RCTs for which the risk of bias was assessed differently in two separate systematic reviews.

2.2. Data sources

Data were collected from the set of all Cochrane reviews published or updated between March 2011 and September 2014. The same set was used for a previous study [17]. The process of selection, evaluation of agreement, and in-depth analysis of disagreements was described in detail elsewhere [17].

Briefly, for each review, we obtained all data entered by review authors in RevMan, the software used for managing Cochrane reviews [11], as xml files [18]. We identified RCTs assessed for risk of bias in two separate systematic reviews. When more than two different reviews included the same study, we randomly selected one of the possible pairs. For each of the five key risk-of-bias items (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome...
assessment, and incomplete outcome data), we compared the risk-of-bias judgments between the two reviews to evaluate whether they agreed (e.g., both reviews assessing allocation concealment concluded “low risk”) or not. For each disagreement, we analyzed the support for the corresponding judgment in the two reviews, to evaluate whether it was the same or not. In case of a difference in the support for the judgment, we examined the study report to evaluate whether the reason for the disagreement could be related to having access to different information (e.g., the study protocol, additional published reports, or contact with authors) or to differences in interpretation of the same information.

We classified each case of a difference in interpretation according to the likely reason: “confusion with another item of the tool,” “incomplete or unclear information in the study report,” or other. This assessment was conducted independently by two researchers; disagreements were resolved by consensus. More details can be found in the report of the previous study [17].

2.3. Selection of eligible studies

For the purpose of this study, we focused on trials for which we observed a disagreement between the two reviews in one or more of the five key risk-of-bias items (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, and incomplete outcome data) and attributed this disagreement to a difference in interpretation related to incomplete or unclear information in the study report. In such situations, a judgment “unclear” would be expected.

2.4. Definition of cases and controls

A case was defined as a judgment of “high risk of bias” or “low risk of bias” for an RCT in one of the reviews. A matched control was defined as a judgment of “unclear risk of bias” for the same RCT, for the same item, in the second, different review. Cases and controls were identified for each item.

2.5. Data extraction

We collected the following information regarding characteristics of reviews and studies included in the review:
- Year of publication of the review.
- Cochrane review group: There are 53 Cochrane review groups within Cochrane in charge of conducting Cochrane reviews. Each group is dedicated to a specific topic (e.g., breast cancer). We identified the Cochrane review group that authored the review and the number of Cochrane reviews conducted by this group during the period of interest, from March 2011 to September 2014. We used the number of reviews conducted by the review group as a measure of the volume/experience of the Cochrane review group.
- Authors: We collected the name of the first author and searched the number of Cochrane reviews conducted by this author up to the end of the period of interest (September 2014). The first reviews included in the Cochrane Database of Systematic Reviews date from 1995; therefore, we have identified all Cochrane reviews from 1995 to 2014. We also evaluated whether authors of the review were also authors of included RCTs (shared authorship RCT/review).
- Size of the review (number of RCTs included in the review).
- Type of interventions analyzed in the review (only pharmacological/only nonpharmacological/both).
- Mention of any type of conflict of interest by review authors.

Data regarding the RCT in comparison to the other RCTs included in the review:
- Sample size of the RCT in comparison with the other studies included in the review: From the xml file containing all the information reported in the review, we extracted the sample size of all studies included in the review. We calculated the distribution of the sample size for the RCTs included and then calculated the percentile of the sample size for the RCT of interest, expressed in a decimal number from 0 to 1.
- Year of publication of the RCT in comparison with the other studies included in the review: We followed the same process explained previously for the year of publication of the RCTs included in the review, calculating the rank percentile for the RCT of interest.
- Risk-of-bias judgment of the RCT in comparison with the other judgments for the same item in the review: We evaluated whether most of the judgments for the item within the review equals “unclear risk of bias.”

2.6. Statistical analysis

We evaluated characteristics potentially associated with judgment of high vs. unclear risk of bias and low vs. unclear risk of bias by using a conditional logistic regression model, accounting for the matched design. We evaluated each factor separately in a univariate analysis. The factors analyzed derive from the data we extracted and are listed in the previous paragraph. The results are expressed as odds ratios (ORs) with 95% confidence intervals (CIs) for each factor, considering separately disagreements of an assessment of high vs. unclear risk of bias and low vs. unclear risk of bias. Quantitative variables were analyzed as such, without dichotomization. We conducted a single joint analysis across different items and then subgroup analyses for each item separately.
The analyses were conducted with Stata 13.1 [19]. P < 0.05 was considered statistically significant. We did not adjust for multiple comparisons in this exploratory study.

3. Results

3.1. Selection process

Fig. 1 shows the details of the selection process. Our previous project [17] showed 625 disagreements, with the same trial judged at low or high risk of bias in one review and unclear in a second review. Among these disagreements, 402 were attributed to a different interpretation of the same information, which was considered related to incomplete or unclear information. These 402 disagreements related to 304 different RCTs, included in 376 different reviews.

3.2. Characteristics of included reviews

Table 1 shows the characteristics of the Cochrane systematic reviews included in our sample. The 376 reviews were compiled by 49 different Cochrane review groups, with the largest number (45; 12%) from the “pregnancy” group (Supplementary Table 1 gives a complete list of the different Cochrane review groups, with names shortened for simplicity in Tables 1 and 2). More than half of the reviews (n = 204, 54%) evaluated only nonpharmacological treatments. The number of RCTs included in reviews varied from 1 to 171 (median 16); only one-quarter of reviews had included more than 30 RCTs. Overall, 147 reviews (39%) reported a conflict of interest among authors of the review.

3.3. Characteristics of the disagreements

Across the 402 disagreements in risk-of-bias items, 69% (n = 278) were between low and unclear judgments; 31% (n = 124) were between high and unclear judgments. The most common disagreements in both the disagreements between low and unclear and high and unclear judgments concerned incomplete outcome data (n = 89, 32% for low vs. unclear and n = 41, 33% for high vs. unclear disagreement) and allocation concealment (n = 74, 27% and n = 29, 23%, respectively) (details in Fig. 1).

Most disagreements derived from two judgments expressed by the same Cochrane group. This occurred in two-thirds of disagreements (n = 261, 65%). In 376 of the 804 (47%) single judgments analyzed, the most common judgment for that item in the review was “unclear” risk of bias. For only 15 judgments (2%), one of the authors of the systematic review was also among the authors of the RCT being evaluated.

3.4. Analysis of factors associated with disagreements

Table 2 compares review and study characteristics for the different judgments: low vs. unclear and high vs. unclear disagreements. Among the low vs. unclear disagreements, review groups and authors that had completed a lower number of reviews assigned significantly more often a low risk of bias where others reported “unclear,” (OR: 0.97 [95% CI: 0.95–0.99] and OR: 0.97 [95% CI:
for one additional Cochrane review conducted by the review group or by the author, respectively).

Among the low risk-of-bias judgments, there were significantly a lower number of RCTs per review (OR: 0.99 [95% CI: 0.99–0.998]), but the OR close to 1 shows in fact a minor effect. Among the low risk-of-bias judgments, there were significatively more reviews in which one of the authors had also coauthored a trial report compared with controls with unclear assignments (OR: 11.71 [1.39–98.76]). There were also significantly fewer recent RCTs that were assigned a low risk of bias (OR: 0.37 [0.15–0.96]). A similar difference, although not significant, was observed for including larger trials (OR: 0.37 [0.13–1.02]).

Among the low risk-of-bias judgments, there were significantly fewer reviews that had assigned most RCTs in their review at an unclear risk of bias (OR: 0.12 [0.07–0.19]). Similarly, among the high risk-of-bias judgments in the high vs. unclear comparison, there were significantly fewer reviews that had evaluated most RCTs at an unclear risk of bias (OR: 0.11 [0.05–0.24]).

In the Supplementary Table 2, we present reviews and study characteristics for cases and controls, for each of the five key RoB items separately. The results for the separate items were affected by small sample sizes. The number of disagreements varied from 9 (random sequence generation) to 41 (incomplete outcome data) for high/unclear disagreements and 16 (blinding of participants and personnel) to 89 (incomplete outcome data) for low/unclear disagreements. We observed a significant difference for only the items random sequence generation, allocation concealment, and incomplete outcome data, with similar findings as in the overall analysis.

4. Discussion

4.1. Main findings

In this study, we evaluated factors or characteristics that are potentially associated with differences in risk-of-bias assessments of RCTs included in systematic reviews, in case of incomplete or unclear information given in reports. If review authors are also authors of the RCT, they are more likely to consider a low risk of bias than an unclear risk. If publication year of the RCT is more recent relative to the other RCTs included in the review, reviewers are more likely to judge an item “low” risk of bias than “unclear.” The number of Cochrane reviews conducted by the review group and by the first author was statistically

Table 1. Characteristics of the included reviews, N = 376

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n (%) Except otherwise indicated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Most common Cochrane review groups</strong></td>
<td></td>
</tr>
<tr>
<td>Pregnancy</td>
<td>45 (12)</td>
</tr>
<tr>
<td>Airways</td>
<td>22 (6)</td>
</tr>
<tr>
<td>Tobacco</td>
<td>21 (6)</td>
</tr>
<tr>
<td>Pain</td>
<td>20 (5)</td>
</tr>
<tr>
<td>Anesthesia</td>
<td>15 (4)</td>
</tr>
<tr>
<td>Respiratory infections</td>
<td>13 (3)</td>
</tr>
<tr>
<td>Mental disorders</td>
<td>13 (3)</td>
</tr>
<tr>
<td>Gynecology</td>
<td>13 (3)</td>
</tr>
<tr>
<td>Neonatal</td>
<td>13 (3)</td>
</tr>
<tr>
<td>Stroke</td>
<td>13 (3)</td>
</tr>
<tr>
<td>Drug addiction</td>
<td>12 (3)</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>12 (3)</td>
</tr>
<tr>
<td><strong>Number of reviews compiled by the authoring review group during 2011 –2014</strong></td>
<td>22 (16–37)</td>
</tr>
<tr>
<td><strong>Year of publication of the reviews</strong></td>
<td>2012 (2012–2013)</td>
</tr>
<tr>
<td><strong>Intervention evaluated</strong></td>
<td></td>
</tr>
<tr>
<td>Pharmacological</td>
<td>145 (39%)</td>
</tr>
<tr>
<td>Nonpharmacological</td>
<td>204 (54%)</td>
</tr>
<tr>
<td>Both</td>
<td>27 (7%)</td>
</tr>
<tr>
<td><strong>Mention of conflict of interest by reviewer</strong></td>
<td>147 (39%)</td>
</tr>
<tr>
<td><strong>Number of RCTs included in the reviews</strong></td>
<td>16 (8–31)</td>
</tr>
<tr>
<td><strong>Year of publication of RCTs</strong> included</td>
<td>2003 (1998–2007)</td>
</tr>
</tbody>
</table>

Abbreviation: RCTs, randomized controlled trials.

a Median (1st and 3rd quartiles).
associated with different risk-of-bias judgments. The CIs for these associations are close to the no-significance level, but the variables are quantitative. Therefore, the interpretation of the ORs is for one additional review conducted either by the review group or author so the effect may not be negligible.

Table 2. Comparison of review characteristics between the different judgments (high vs. unclear and low vs. unclear risk of bias) for the included trials for all items of the Cochrane Risk of Bias tool

<table>
<thead>
<tr>
<th>All items</th>
<th>Low N = 278</th>
<th>Unclear N = 278</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of publication of the review</td>
<td>2012 (2012–2013)</td>
<td>2012 (2012–2013)</td>
<td>0.99 (0.8–1.22)</td>
</tr>
<tr>
<td>Most common review Cochrane groups</td>
<td>Tobacco: 28 (10%)</td>
<td>Pregnancy: 28 (10%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Airways: 24 (9%)</td>
<td>Pain: 26 (9%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pregnancy: 22 (8%)</td>
<td>Tobacco: 25 (9%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pain: 17 (6%)</td>
<td>Airways: 22 (8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroke: 13 (5%)</td>
<td>Stroke: 18 (6%)</td>
<td></td>
</tr>
<tr>
<td>Number of reviews compiled by the review group</td>
<td>22 (15–29)</td>
<td>23 (18–42)</td>
<td>0.97 (0.95–0.99)</td>
</tr>
<tr>
<td>Number of reviews compiled by the first author</td>
<td>3 (1–5)</td>
<td>3 (1–7)</td>
<td>0.97 (0.95–0.99)</td>
</tr>
<tr>
<td>Shared author between the review and the RCT</td>
<td>9 (3%)</td>
<td>1 (0.4%)</td>
<td>11.71 (1.39–98.76)</td>
</tr>
<tr>
<td>Mention of conflict of interest</td>
<td>118 (42%)</td>
<td>129 (46%)</td>
<td>0.77 (0.50–1.18)</td>
</tr>
<tr>
<td>Number of RCTs included in the review</td>
<td>23 (10–45)</td>
<td>29 (13–55)</td>
<td>0.99 (0.99–0.99)</td>
</tr>
<tr>
<td>Percentile distribution of the sample size of RCTs</td>
<td>0.58 (0.3)</td>
<td>0.61 (0.29)</td>
<td>0.37 (0.13–1.02)</td>
</tr>
<tr>
<td>Percentile distribution of the year of RCTs</td>
<td>0.56 (0.29)</td>
<td>0.59 (0.28)</td>
<td>0.37 (0.15–0.96)</td>
</tr>
<tr>
<td>Majority of judgments are at “unclear” risk</td>
<td>73 (26%)</td>
<td>192 (69%)</td>
<td>0.12 (0.07–0.19)</td>
</tr>
</tbody>
</table>

All items | High N = 124 | Unclear N = 124 | OR (95% CI) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of publication of the review</td>
<td>2012 (2012–2013)</td>
<td>2012 (2011–2013)</td>
<td>1.15 (0.8–1.68)</td>
</tr>
<tr>
<td>Most common review Cochrane groups</td>
<td>Pregnancy: 19 (15%)</td>
<td>Pregnancy: 20 (16%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Pain: 10 (8%)</td>
<td>Tobacco: 10 (8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tobacco: 9 (7%)</td>
<td>Drugs Addiction: 7 (6%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gynaeco. Cancer: 7 (6%)</td>
<td>Anesthesia: 6 (5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Developmental: 6 (5%)</td>
<td>Pain: 6 (5%)</td>
<td></td>
</tr>
<tr>
<td>Number of reviews compiled by the review group</td>
<td>22 (16–42)</td>
<td>22 (16–39)</td>
<td>1.00 (0.98–1.02)</td>
</tr>
<tr>
<td>Number of reviews compiled by the first author</td>
<td>2 (1–5)</td>
<td>3 (1–6)</td>
<td>1.01 (0.98–1.03)</td>
</tr>
<tr>
<td>Shared author between the review and the RCT</td>
<td>3 (2%)</td>
<td>2 (2%)</td>
<td>1.60 (0.23–10.87)</td>
</tr>
<tr>
<td>Mention of conflict of interest</td>
<td>64 (52%)</td>
<td>54 (44%)</td>
<td>1.51 (0.86–2.68)</td>
</tr>
<tr>
<td>Number of RCTs included in the reviews</td>
<td>21 (12–45)</td>
<td>26 (10–47)</td>
<td>1.00 (0.99–1.01)</td>
</tr>
<tr>
<td>Percentile distribution of the sample size of RCTs</td>
<td>0.53 (0.3)</td>
<td>0.54 (0.3)</td>
<td>0.65 (0.14–3.01)</td>
</tr>
<tr>
<td>Percentile distribution of the year of RCTs</td>
<td>0.55 (0.28)</td>
<td>0.58 (0.29)</td>
<td>0.24 (0.05–1.20)</td>
</tr>
<tr>
<td>Majority of judgments are at “unclear” risk</td>
<td>30 (24%)</td>
<td>81 (65%)</td>
<td>0.11 (0.05–0.24)</td>
</tr>
</tbody>
</table>

Abbreviation: RCTs, randomized controlled trials.

n (%) except otherwise indicated.
a Mean (standard deviation).
b Median (1st and 3rd quartiles).
4.2. Strengths and limitations

The main strength of our study is that it relies on real-life data. We did not directly explore the intention or attitude of researchers in handling incomplete information but evaluated associations between features and judgments by the observation of high-quality data from systematic reviews. Our sample size of studies and our careful work in collecting and categorizing data are also strengths.

Our study has some limitations. Our analysis of disagreements originating from incomplete information is not fully objective even if the classification was conducted in duplicate by two authors and reconciled, following a precise methodology, to reduce the subjectivity. We did not correct for multiple testing, because this analysis was exploratory, so the results should be interpreted carefully. Moreover, because of the low number of disagreements, particularly for high vs. unclear disagreements (n = 124), our analysis may lack power. We did not perform a multivariate analysis on our sample because of the exploratory nature of this study and because the different variables may interact between each other in an unstudied way. We are aware that our sample of reviews is not up to date, and we cannot exclude that a sample of more recent reviews would have given different results regarding the agreement in the risk-of-bias assessment and the relative importance of different reasons of disagreement. We nevertheless believe that the general point of our study—contextual factors may influence judgment regarding risk of bias—is not affected by the age of the sample. We evaluated factors/characteristics that are not part of the risk-of-bias assessment and authors are not aware of them. Therefore, there is no reason why these factors would disappear in a more recent sample.

4.3. Comparison with other studies

To our knowledge, this is the first study to explore the potential influence of characteristics pertaining to review groups, studies, and reviewer experience on risk-of-bias assessment. The research of Jørgensen et al. [6] and Savović et al. [20] was based on surveys and user comments about the use of the tool. The authors reported on the difficulties reviewers experience in judging incomplete or unclear information, but they did not analyze how these difficulties translate into real practice. More recently, Babic et al. [16] and Propadalo et al. [15] analyzed the support for judgments provided by reviewers. Their results also show how the same information in the study report can lead to a different risk-of-bias judgment, but they focused on an incorrect use of the RoB tool and did not analyze which factors may be associated with a different attitude toward the same information.

4.4. Interpretation of results and implications

We observed that judgments of unclear risk of bias more frequently came from reviews that judged most of the RCTs included in the review at unclear risk. One possible explanation is that when the risk-of-bias assessment of a study is not completely straightforward (hence resulting in a disagreement between two different reviews), the reviewers may tend to go back to the judgment they used for most of the RCTs they assessed. This situation could represent a “halo” effect of the assignments for the other trials in the review. We may even suppose that the factors we identified, or possibly other factors, have a role in the risk-of-bias assessment regardless of the clarity or completeness of information.

An interesting point is that two-thirds of the disagreements originated within the same Cochrane review group. Nevertheless, we also showed that a higher number of reviews conducted by review groups and by authors, thereby suggesting more experience with risk-of-bias assessment, seems to have influenced reviewers in their judgment. This finding could suggest that the level of training may differ across author groups or it may be influenced by the experience of the reviewers. Of note, the influence of “experience” was toward more frequent judgments of unclear risk, instead of low risk.

The ORs for high vs. unclear disagreements did not reach statistical significance as compared with those for low vs. unclear disagreements. This finding might be explained in part by a lack of power of the study owing to the smaller sample size for the former comparison, but other interpretations are possible.

4.5. Conclusions

Our results indicate that identifiable review and author factors may play a role in risk-of-bias judgments of RCTs in systematic reviews. Specific factors that influenced reviewers in this analysis were the number of reviews they previously conducted and the number authored by the review group to which they belong. Reviewers also tended to judge a study at unclear risk of bias if they already did so for most of the studies included in the same review and tend to judge an item at “low” risk if they were involved in the RCT they are assessing. Awareness of contextual characteristics that can affect risk-of-bias judgments may help researchers in conducting more reproducible and meaningful assessments of risk of bias.

CRediT authorship contribution statement

Lorenzo Bertizzolo: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing - original draft, Writing - review & editing. Patrick M. Bossuyt: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing. Ignacio Atal: Data curation, Software, Writing - review & editing. Philippe
Ravaud: Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing - review & editing. Agnès Dechartres: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Acknowledgments

The authors thank David Tovey, former editor in chief of the Cochrane Library, for agreeing to share data from Cochrane reviews; Javier Mayoral Campos, system administrator; the Cochrane Central Executive for preparing files; and all Cochrane reviewers who collected data. The authors also thank Laura Smales for English revision of the manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.03.008.

References

[14] Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. J Clin Epidemiol 2017; 81:72–6.