



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

Lorenzo Bertizzolo, Patrick M. Bossuyt, Ignacio Atal, Philippe Ravaud, Agnès Dechartres

### **► To cite this version:**

Lorenzo Bertizzolo, Patrick M. Bossuyt, Ignacio Atal, Philippe Ravaud, Agnès Dechartres. External factors may influence Cochrane reviewers when classifying the risk of bias of original reports. *Journal of Clinical Epidemiology*, 2020, 123, pp.1-8. <10.1016/j.jclinepi.2020.03.008>. <hal-03490764>

**HAL Id: hal-03490764**

**<https://hal.science/hal-03490764v1>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

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

Lorenzo Bertizzolo<sup>1, 2</sup>, Patrick M. Bossuyt<sup>2</sup>, Ignacio Atal<sup>1, 3, 4</sup>, Philippe Ravaud<sup>1, 3, 5, 6, 7</sup>, Agnes Dechartres<sup>8</sup>

1. Université de Paris, CRESS, INSERM, INRA, F-75004 Paris, France
2. Department of Clinical Epidemiology and Biostatistics, Academic University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands
3. Cochrane France, Paris, France
4. Centre de Recherches Interdisciplinaires (CRI) – Université Paris Descartes
5. Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu AP-HP (Assistance Publique - Hôpitaux de Paris), Paris, France
6. Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France
7. Department of Epidemiology, Columbia University, Mailman School of Public Health, New York, USA
8. Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpitaux Universitaires Pitié Salpêtrière - Charles Foix, Département Biostatistique Santé Publique et Information Médicale, F75013, Paris, France

Corresponding author: Lorenzo Bertizzolo, M.D., MPH

INSERM, U1153 - Centre d'Epidémiologie Clinique –  
Hôpital Hôtel-Dieu,  
1, place du parvis Notre Dame, 75004 Paris, France

Tel: +33 (0)1 42 34 78 25

E-mail: lorenzo.bertizzolo@gmail.com

Keywords: Risk of bias; systematic reviews; research on research; reporting; matched case-control.

Declarations 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.

Word count: 2881

## **Abstract (199 words)**

**Objective:** To explore contextual factors associated with high or low risk of bias judgement 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 RCTs included in two Cochrane reviews for which there was disagreement on the risk of bias judgement 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 1.39-98.76). Earlier trials in a review were more often assigned a low risk (OR 0.37 [0.15-0.96]). 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 0.95-0.998) for authors).

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

### **What is new?**

- 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.
- 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.
- Authors of systematic reviews should be aware that their risk-of-bias judgement 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 judgement, 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” judgement can be expected in case of incomplete or unclear information, many authors may feel confident enough to consider the information adequate to express a judgement of low or high risk of bias. Erroneous decisions apart, the RoB tool

clearly leaves ample space for personal interpretation and judgement 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 judgements when evaluating the same incomplete or unclear information. These characteristics could be related to the reviewers, the review, or the study itself, when compared to 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 judgement 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 Cochrane 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, incomplete outcome data), we compared the risk of bias judgements 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 analysed the support for the corresponding judgement in the two reviews, to evaluate whether it was the same or not. In case of a difference in the support for the judgement, 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 attributed this disagreement to a difference in interpretation related to incomplete or unclear information in the study report. In such situations, a judgement “unclear” would be expected.

### ***2.4. Definition of cases and controls***

A case was defined as a judgement 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 judgement 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:

#### Data regarding 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 analysed in the review (*only pharmacological/only non-pharmacological/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 to 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, 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 to the other studies included in the review. We followed the same process explained above for the year of publication of the RCTs included in the review, calculating the rank percentile for the RCT of interest.
- Risk of bias judgement of the RCT in comparison to the other judgements for the same item in the review. We evaluated whether the majority of the judgements for the item within the review equals “unclear risk of bias”.

## ***2.6. Statistical Analysis***

We evaluated characteristics potentially associated with judgement of high versus unclear risk of bias and low versus 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 analysed derive from the data we extracted and are listed in the previous paragraph. The results are expressed as odds ratios (OR) with 95% confidence intervals (CI) for each factor, considering separately disagreements of an assessment of high versus unclear risk of bias and low versus unclear risk of bias. Quantitative variables were analysed as such, without dichotomization. We conducted a single joint analysis across different items, then subgroup analyses for each item separately.

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**

Figure 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 non-pharmacological 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 judgements; 31% (n=124) were between high and unclear judgements.

The most common disagreements in both the disagreements between low and unclear and high and unclear judgements 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 Figure 1*).

Most disagreements derived from two judgements expressed by the same Cochrane group. This occurred in two thirds of disagreements (n=261, 65%). In 376 of the 804 (47%) single judgements analysed, the most common judgement for that item in the review was “unclear” risk of bias. For only 15 judgements (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 judgements: low versus unclear and high versus unclear disagreements. Among the low versus 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 0.95-0.998] for one additional Cochrane review conducted by the review group or by the author, respectively).

Among the low risk of bias judgements, 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 judgements, there were significantly more reviews in which one of the authors had also co-authored a trial report compared to 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 judgements there were significantly fewer reviews that had assigned the majority of RCTs in their review at an unclear risk of bias (OR 0.12 [0.07-0.19]). Similarly, among the high risk of bias judgements in the high vs unclear comparison, there

were significantly fewer reviews that had evaluated the majority of 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 associated with different risk-of-bias judgments. The confidence intervals for these associations are close to the no-significance level, but the variables are quantitative. Therefore, the interpretation of the odds ratios is for one additional review conducted either by the review group or author so the effect may not be negligible.

### ***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 judgements 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 versus 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 analyse how these difficulties translate into real practice. More recently, Babic et al.(16) and Propadalo et al.(15) analysed 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 judgement, but they focused on an incorrect use of the RoB tool and did not analyse which factors may be associated with a different attitude toward the same information.

#### **4.4. Interpretation of results and implications**



We observed that judgements 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 judgement 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 judgements of unclear risk, instead of low risk.

The ORs for high versus unclear disagreements did not reach statistical significance as compared with those for low versus 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 judgements 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 judgements may help researchers in conducting more reproducible and meaningful assessments of risk of bias.

## **ACKNOWLEDGEMENTS:**

We 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. We also thank Laura Smales for English revision of the manuscript.

## **CONTRIBUTIONS:**

Lorenzo Bertizzolo was involved in the study conception, selection of trials, data extraction, data analysis, interpretation of results and drafting the manuscript.

Patrick Bossuyt was involved in the study conception, data analysis, interpretation of results and drafting the manuscript.

Ignacio Atal was involved in the study conception, data extraction, and drafting the manuscript.

Philippe Ravaud was involved in the study conception and drafting the manuscript.

Agnès Dechartres was involved in the study conception, selection of trials, data extraction, data analysis, interpretation of results and drafting the manuscript.

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.

**COMPETING INTERESTS:**

The authors declare that they have no competing interests in relation to this study.

**DATA SHARING:**

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

## REFERENCES

1. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126(5):376-80.
2. Hopewell S, Boutron I, Altman DG, Ravaud P. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *BMJ Open.* 2013;3(8):e003342.
3. Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D, Cochrane Bias Methods G. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. *Cochrane Database Syst Rev.* 2013(4):ED000058.
4. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343(oct18 2):d5928.
5. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J.* 1986;134(6):587-94.
6. Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Sterne JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. *Syst Rev.* 2016;5:80.
7. Wilkins AJ. Risk of bias in assessing Risk of Bias. *Ophthalmic Physiol Opt.* 2017;37(1):107-9.
8. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol.* 2013;66(9):973-81.

9. Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes J, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One*. 2014;9(5):e96920.
10. Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One*. 2011;6(2):e17242.
11. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339(oct19 1):b4012.
12. da Costa BR, Beckett B, Diaz A, Resta NM, Johnston BC, Egger M, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a prospective study. *Syst Rev*. 2017;6(1):44.
13. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract*. 2012;18(1):12-8.
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.
15. Propadalo I, Tranfic M, Vuka I, Barcot O, Pericic TP, Puljak L. In Cochrane reviews risk of bias assessments for allocation concealment were frequently not in line with Cochrane's Handbook guidance. *J Clin Epidemiol*. 2018.
16. Babic A, Tokalic R, Amilcar Silva Cunha J, Novak I, Suto J, Vidak M, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Med Res Methodol*. 2019;19(1):76.

17. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: a research on research study using cross-sectional design. *BMJ Open*. 2019;9(4):e028382.
18. Review Manager (RevMan) [Computer program]. Version 5.3 ed. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
19. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.
20. Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. *Syst Rev*. 2014;3:37.

**Table 1: Characteristics of the included reviews, N=376**

<b>Characteristics</b>	<i>n (%) except otherwise indicated</i>
<b>Most common Cochrane review groups</b>	
Pregnancy	45 (12)
Airways	22 (6)
Tobacco	21 (6)
Pain	20 (5)
Anaesthesia	15 (4)
Respiratory Infections	13 (3)
Mental Disorders	13 (3)
Gynaecology	13 (3)
Neonatal	13 (3)
Stroke	13 (3)
Drug Addiction	12 (3)
Cystic Fibrosis	12 (3)
<b>Number of reviews compiled by the authoring review group during 2011-2014*</b>	22 (16-37)
<b>Year of publication of the reviews*</b>	2012 (2012-2013)
<b>Intervention evaluated</b>	
Pharmacological	145 (39%)
Non-pharmacological	204 (54%)
Both	27 (7%)
<b>Mention of conflict of interest by reviewer</b>	147 (39%)
<b>Number of RCTs included in the reviews*</b>	16 (8-31)
<b>Year of publication of RCTs* included</b>	2003 (1998-2007)

\* median (1st and 3rd quartile)



**Table 2: Comparison of review characteristics between the different judgments (high versus unclear and low versus unclear risk of bias) for the included trials for all items of the Cochrane Risk of Bias tool. (n (%) except otherwise indicated)**

<b>All items</b>	<b>Low N=278</b>	<b>Unclear N=278</b>	<b>OR (95% CI)</b>
<b>Year of publication of the review*</b>	2012 (2012-2013)	2012 (2012-2013)	0.99 (0.8-1.22)
<b>Most common review Cochrane groups</b>	Tobacco: 28 (10%) Airways: 24 (9%) Pregnancy: 22 (8%) Pain: 17 (6%) Stroke: 13 (5%)	Pregnancy: 28 (10%) Pain: 26 (9%) Tobacco: 25 (9%) Airways: 22 (8%) Stroke: 18 (6%)	-
<b>Number of reviews compiled by review group*</b>	22 (15-29)	23 (18-42)	<b>0.97</b> <b>(0.95-0.99)</b>
<b>Number of reviews compiled by first author*</b>	3 (1-5)	3 (1-7)	<b>0.97</b> <b>(0.95-0.998)</b>
<b>Shared author between the review and the RCT</b>	9 (3%)	1 (0.4%)	<b>11.71</b> <b>(1.39-98.76)</b>
<b>Mention of conflict of interest</b>	118 (42%)	129 (46%)	0.77 (0.50-1.18)
<b>Number of RCTs included in the review*</b>	23 (10-45)	29 (13-55)	<b>0.99</b> <b>(0.99-0.998)</b>
<b>Percentile distribution of sample size of RCTs^</b>	0.58 (0.3)	0.61 (0.29)	0.37 (0.13-1.02)
<b>Percentile distribution of the year of RCTs^</b>	0.56 (0.29)	0.59 (0.28)	<b>0.37</b> <b>(0.15-0.96)</b>
<b>Majority of judgements are at “unclear” risk</b>	73 (26%)	192 (69%)	<b>0.12</b> <b>(0.07-0.19)</b>
<b>All items</b>	<b>High</b>	<b>Unclear</b>	<b>OR</b>

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

<sup>^</sup> mean (standard deviation)

\* median (1st and 3rd quartile)

625 disagreements analysed for reasons of disagreement in our previous publication

- 223 disagreements excluded:
- 16 input mistake
  - 25 missed information in the publication
  - 39 cannot determine reason of disagreement
  - 79 access to additional information
  - 64 reasons of different interpretation not part of our inclusion criteria

402 disagreements kept for analysis in the present study

278 disagreements between low and unclear assessments

124 disagreements between high and unclear assessments

62 (22%)  
Random  
Sequence  
Generation

74 (27%)  
Allocation  
Concealment

16 (6%)  
Blinding  
Participants  
and Personnel

37 (13%)  
Blinding of  
Outcome  
Assessment

89 (32%)  
Incomplete  
Outcome Data

9 (7%)  
Random  
Sequence  
Generation

29 (23%)  
Allocation  
Concealment

22 (18%)  
Blinding  
Participants  
and Personnel

23 (19%)  
Blinding of  
Outcome  
Assessment

41 (33%)  
Incomplete  
Outcome Data