Journal of Clinical Epidemiology 113 (2019) 104−113

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Risk of bias assessments for blinding of participants and personnel in Cochrane reviews were frequently inadequate

Ognjen Barcot[a], Matija Boric[a], Svjetlana Dosenovic[b], Tina Poklepovic Pericic[c], Marija Cavar[d], Livia Puljak[e,*]

[a]*Department of Abdominal Surgery, University Hospital Split, Split, Croatia*
[b]*Department of Anesthesiology and Intensive Care, Anesthesiology Resident, University Hospital Split, Split, Croatia*
[c]*Cochrane Croatia, University of Split School of Medicine, Split, Croatia*
[d]*Department of Radiology, University Hospital Split, Split, Croatia*
[e]*Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000 Zagreb, Croatia*

Accepted 18 May 2019; Published online 24 May 2019

Abstract

Objectives: The objective of this study was to analyze adequacy of judgments about risk of bias (RoB) for blinding of participants and personnel (performance bias) in Cochrane systematic reviews of randomized controlled trials (RCTs).

Study Design and Setting: We extracted judgments and supporting comments for performance bias from Cochrane reviews' RoB tables using automated data scraping. We parsed all intervention descriptions, judgments about risk of performance bias, and comments supporting judgments into simple categories. We assessed adequacy of RoB judgments against recommendations from the Cochrane Handbook.

Results: We analyzed judgments for performance bias of 10,429 RCTs included in 718 Cochrane reviews. Overall, 1,828 out of 6,918 judgments (26%) for performance bias were not in line with the Cochrane Handbook and were therefore considered inadequate. In reviews where Cochrane authors have split the performance bias domain into two subdomains, based on blinded individuals, we found lower prevalence of inadequate risk of bias judgments, with 9% of judgments for blinding of participants, and 5.8% judgments for the blinding of personnel subdomain being judged inadequately.

Conclusion: In Cochrane reviews, risk of bias assessments for blinding of participants and personnel were frequently not in line with Cochrane Handbook recommendations. Interventions to improve these assessments should be taken into consideration.

Keywords: Risk of bias; Cochrane; Systematic reviews; Blinding; Participants; Personnel; Performance bias

1. Introduction

Systematic reviews (SRs) are considered the highest level of evidence in medicine. One of the crucial aspects of SR methodology is the assessment of bias in included trials [1]. Cochrane is using a tool for assessing risk of bias (RoB) in randomized controlled trials (RCTs), which aims to detect flaws in their design, conduct, analysis, and completeness of reporting. Cochrane's RoB tool has seven domains, and two of those domains consider blinding—the first one is called "blinding of participants and personnel"—assessing performance bias, and the second one is "blinding of outcome assessors"—assessing detection bias [2].

Blinding is one of the key methodological considerations in RCTs because it is necessary for controlling bias. Blinding in clinical trials involves concealment of research design elements, such as group assignment, treatment agent, and research hypothesis, from participants, personnel, and outcome assessors. However, it has already been described that making judgments about the adequacy of blinding in RCTs may be difficult because of poor reporting [3].

Conflicts of interest: Tina Poklepovic Pericic and Livia Puljak are volunteer members of Cochrane Croatia. Other authors have no competing interests to declare.

Conflict of interest: The authors report no conflict of interest.

Ethics: This study involved an analysis of published literature; it did not involve any humans or animals and therefore approval of the research ethics committee was not sought.

* Corresponding author: Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000 Zagreb, Croatia, Tel.: + 385 (0) 1 370 66 00; fax: + 385 (0) 1 370 66 01.

E-mail addresses: livia.puljak@gmail.com; livia.puljak@unicath.hr (L. Puljak).

https://doi.org/10.1016/j.jclinepi.2019.05.012

What is new?

Key findings

- More than a quarter of risk of bias (RoB) assessments for blinding of participants and personnel in Cochrane reviews were not in line with Cochrane Handbook.
- The highest prevalence of inadequate judgments was found in reviews judged with low risk of bias.
- Higher prevalence of adequate assessments for this domain of Cochrane RoB tool was found in reviews that have split this domain into two subdomains, one for participants and one for personnel.

What this adds to what was known?

- Our previous studies have shown that Cochrane reviews have a high prevalence of inadequate and inconsistent assessments in other RoB domains.
- This study adds to this body of knowledge by analyzing RoB domain separately for blinding of participants and personnel.

What is the implication and what should change now?

- Interventions for ensuring adequate and consistent use of systematic review methodology in Cochrane reviews should be tested with the aim to improve RoB assessments.

For making judgments about RoB in Cochrane systematic reviews (Cochrane reviews), the Cochrane Handbook for Systematic Reviews of Interventions (Cochrane Handbook) provides guidance about various possible scenarios that should be judged either as low, unclear, or high RoB [2]. However, our previous analyses of other Cochrane RoB domains have shown that Cochrane authors do not necessarily make adequate judgments that are in line with the Cochrane Handbook [4–8]. Therefore, the aim of our study was to analyze the adequacy of judgments about the risk of performance bias in Cochrane reviews of RCTs by comparing judgments based on supporting comments with instructions from the Cochrane Handbook.

2. Methods

2.1. Study design

This was a primary methodological study that analyzed methods of published Cochrane reviews.

2.2. Inclusion and exclusion criteria

We retrieved Cochrane reviews of RCTs from The Cochrane Library via advanced search limited to interventions published from July 2015 to June 2016 ($N = 955$). We used the advanced search option because it enables the use of search limits such as content type, publication date, and so on. We excluded all Cochrane reviews ($N = 237$) that did not include RCTs about interventions (diagnostic Cochrane reviews, overviews of systematic reviews), as well as empty reviews, which only contained non-RCTs and reviews that were withdrawn during the analyzed period. If a Cochrane review included both randomized and non-randomized trials, we analyzed the RoB table for included RCTs only.

2.3. Screening for study eligibility

One author (O.B.) assessed all titles and abstracts to establish the eligibility of Cochrane reviews for inclusion. The second author (M.B.) verified the assessments of the first author.

2.4. Data extraction

Data extraction was automated in a stepwise manner in Microsoft Excel 2,010 (Microsoft, Redmond, WA, USA) using macrocommands written in Visual Basic for Applications (VBA, Microsoft, Redmond, WA, USA) by the first author (O.B.). Data scraping was done by automated copying all content from The Cochrane Library webpage for every eligible Cochrane review to a separate spreadsheet in MS Excel. More details about data extraction are available in Supplementary File 1.

2.5. Development and testing of the parser tool

By referring to the Cochrane Handbook (Section 8.11.2 and Box 8.11.a) [2], we developed four questions that needed to be answered for correct assessment of performance bias: (1) Who was blinded? (2) Was blinding achieved? (3) Outcome category? (4) Outcome influenced by blinding? Using these questions and the instructions from the Cochrane Handbook, and considering previous studies [3,9–15], we first made a new assessment of the risk of performance bias and then compared it with the assessments made by Cochrane authors in published Cochrane reviews by using the computer algorithm developed specifically for this study. Detailed methods about our assessment of performance bias and computer algorithm are available in Supplementary File 1.

Owing to a large number of analyzed trials, a user- and data-friendly interface was developed in MS Excel VBA by the first author (O.B.) to facilitate the transformation of complex textual data to variables (parsing). Using multiple checkboxes and limited option buttons, data from RoB

Table 1. Hypotheses, outcome measures, the statistical test used, result, and power

Variable/outcome	Null hypothesis	Method of analysis	Result	Conclusion
Main analysis: Number and adequacy of judgment for risk of bias associated with performance bias	Judgments assigned by Cochrane review authors are concordant to our calculation according to the Cochrane Handbook	Wilcoxon test (paired samples)	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Length of comment is the same for all (high, low, and unclear) groups of assigned judgments.	Kruskal–Wallis test	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: Length of comment impacts assigned judgment.
	Length of comment is the same for all (high, low, and unclear) groups of calculated judgments.	Kruskal–Wallis test	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: Length of comment impacts calculated judgment.
Secondary analysis: Length of comment and judgments	Length of comment is the same in the group when blinding was achieved as in group when blinding was not achieved	Kruskal–Wallis test	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: Length of comment impacts blinding achievement.
	Length of comment is the same in subgroup changing from assigned unclear RoB judgment to calculated high RoB judgment as in the group with unchanged high RoB judgment	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: Shortening the comment does not lower the risk judgment from high to unclear.
	Length of comment is the same in subgroup changing from assigned unclear RoB judgment to calculated high RoB judgment as in the group with unchanged unclear RoB judgment	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: Shortening the comment does not lower the risk judgment from high to unclear.
Secondary analysis: Splitting the outcomes	Proportion of cases with unlikeliness of lack of blinding influencing the outcome is the same in the subgroup with split outcomes as in the main sample	Chi-squared test	$P < 0.05$	Null hypothesis rejected: significant difference between groups. Alternative hypothesis accepted: cases tend to group in split outcome sample.
Secondary analysis: Splitting the domain	Distribution of assigned judgments is the same in the subgroup (personnel) as in the main sample	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Distribution of calculated judgments is the same in the subgroup (personnel) as in the main sample	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Distribution of assigned judgments is the same in the subgroup (participants) as in the main sample	Mann–Whitney	$P = 0.0612$	No significant difference between the groups found.
	Distribution of calculated judgments is the same in the subgroup (participants) as in the main sample	Mann–Whitney	$P = 0.1412$	No significant difference between the groups found.
	Distribution of assigned judgments is the same in the merged subgroup (participants and personnel) as in the main sample	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups.

(*Continued*)

Table 1. Continued

Variable/outcome	Null hypothesis	Method of analysis	Result	Conclusion
	Distribution of calculated judgments is the same in the merged subgroup (participants and personnel) as in the main sample	Mann–Whitney	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Proportion of unclear judgments is the same in the subgroup (participants) as in the main sample	Chi-squared test	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Proportion of unclear judgments is the same in the subgroup (personnel) as in the main sample	Chi-squared test	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Proportion of unclear judgments is the same in the subgroup (participants) as in the subgroup (personnel)	Chi-squared test	$P = 0.1023$	Statistic test demonstrates no significant difference between groups. Sample too small.
	Judgments in participant subgroup assigned by CSR authors are concordant to our calculation according to Cochrane handbook	Wilcoxon test (paired sample)	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Judgments in personnel subgroup assigned by CSR authors are concordant to our calculation according to Cochrane handbook	Wilcoxon test (paired sample)	$P = 0.2565$	No significant difference between the groups found.
	Judgments merged (participant and personnel) subgroup assigned by CSR authors are concordant to our calculation according to Cochrane handbook	Wilcoxon test (paired sample)	$P < 0.05$	Null hypothesis rejected: significant difference between groups.
	Proportion of concordant judgments is the same in the subgroup (participants) as in the subgroup (personnel)	Chi-squared test	$P < 0.05$	Statistic test demonstrates possible difference between the groups. Sample too small.
	Proportion of concordant judgments is the same in the merged subgroup (participants and personnel) as in the main sample	Chi-squared test	$P < 0.05$	Null hypothesis rejected: significant difference between groups.

Abbreviations: RoB, risk of bias.

tables (judgments, comments, citations) were parsed into ordinal or nominal variables.

The first author (O.B.) analyzed 500 random trials to pilot-test and adjust the tool; afterward, another author (S.D.) analyzed a new set of 500 trials against judgments from the Cochrane Handbook. Both analyses were verified by the third author (L.P.). During this exercise, we established an important set of rules for each of the four questions used for assessing performance bias.

2.6. Primary outcomes

We analyzed if number and adequacy of judgments for performance bias assigned by Cochrane authors were in line with recommendations from the Cochrane Handbook. The Handbook was used as a gold standard in our assessment; if the judgments of Cochrane authors did not fully adhere to the guidance from the Cochrane Handbook, we considered them inadequate.

2.7. Secondary outcomes

We analyzed end result of usage of undefined "tuple-blind" statements (uncertainty in supporting comments), impact of length of supporting comment (LOC) for assigned judgment vs. calculated judgment, influence of splitting judgment for different outcomes (splitting the outcomes), and influence of dividing blinding subjects to subdomains (splitting the domain).

2.8. Statistics

Descriptive data were presented as frequencies and percentages. Type I error $\alpha = 0.05$ and type II error $\beta = 0.2$

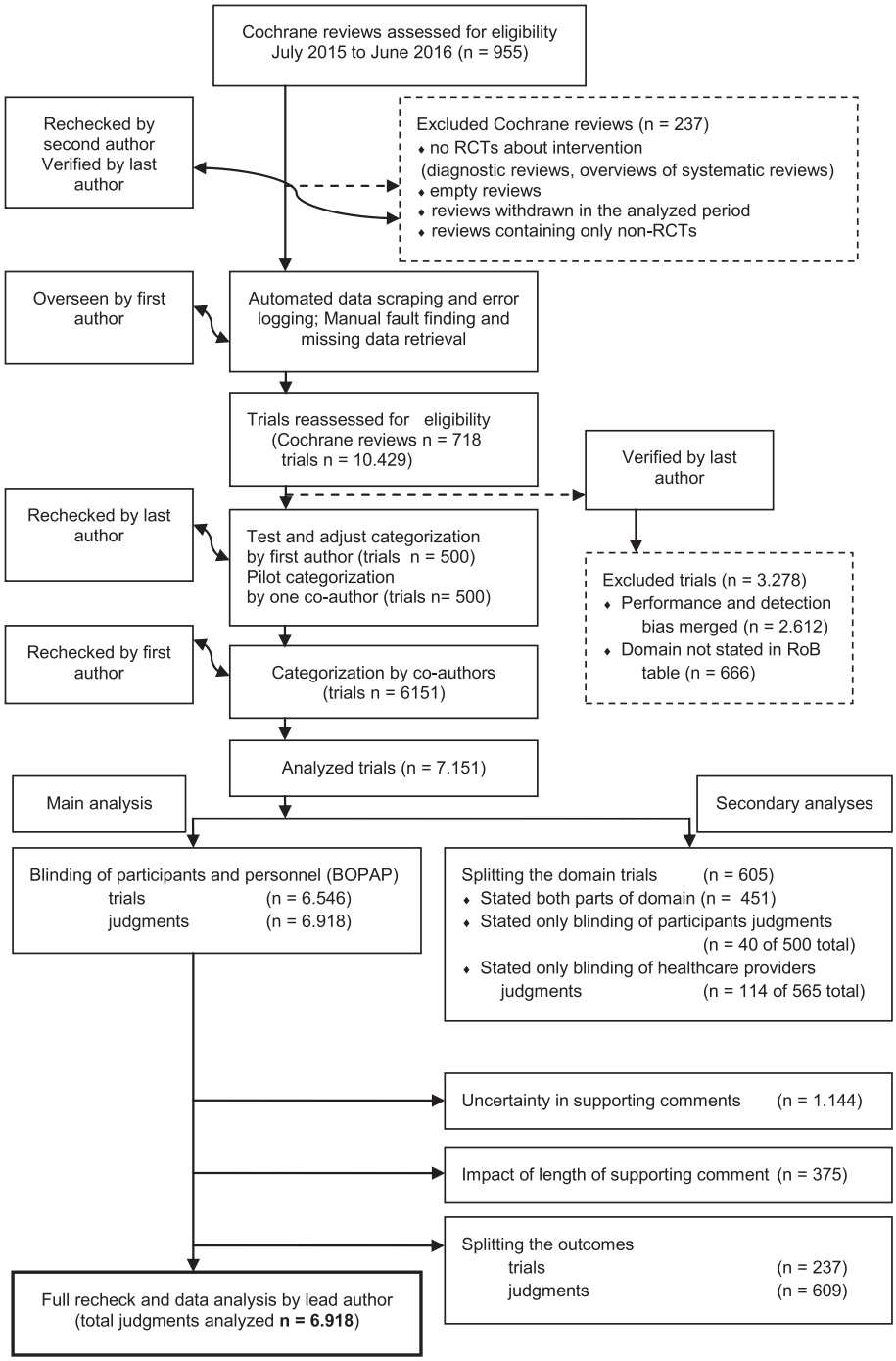

Fig. 1. Flow diagram of the progress through the phases of the study.

were used for all statistical tests. Normality of data sets was tested using the Kolmogorov-Smirnov test. For comparison of independent groups of nonparametric data, Mann-Whitney test was used; for more than two groups of data, Kruskal-Wallis test was used. For dependent groups of nonparametric data, Wilcoxon test (paired samples) was used. Statistical analyses were performed using MedCalc for Windows, version 12.5.0.0 (MedCalc Software, Ostend, Belgium). Detailed descriptions of the tests used are given in Table 1.

3. Results

We analyzed RoB tables of 10,429 RCTs included in 718 Cochrane reviews (Figure 1). Performance bias domain, as proposed by the Cochrane Handbook, containing judgment for blinding of participants and personnel was addressed in 6,546 RCT analyses. This was followed by 237 RCTs in which assessments have been made for more than one outcome (splitting the outcomes) thus contributing to the total of 6,918 judgments.

Table 2. Adequacy of judgments for performance bias

Assigned judgment[a]	Calculated judgment[b]			Total *N* (%)
	High risk	Low risk	Unclear risk	
High risk	2,386	5	183	2,574 (37.2%)
Low risk	96	1,094	1,143	2,333 (33.7%)
Unclear risk	375	26	1,610	2,011 (29.1%)
Total	2,857 (41.3%)	1,125 (16.3%)	2,936 (42.4%)	6,918 (100.0%)
Inadequate judgments	471 (16.5%)	31 (2.8%)	1,326 (45.2%)	1,828 (26.4%)

[a] Judgment provided by the Cochrane review authors.
[b] Judgment in line with the Cochrane Handbook.

In secondary analyses, we observed an additional 605 RCTs in which authors split the domain to either blinding of participants ($n = 500$) or blinding of personnel ($n = 565$). Most of these ($n = 451$) included both parts of the split domain.

3.1. Number and adequacy of judgment for risk of bias associated with performance bias

A total of 6,918 RoB judgments from Cochrane reviews were considered in our main analysis. The largest proportion of judgments, 2,574 (37.2%), were assigned high risk, followed by the 2,333 (33.7%) with low risk, and 2,011 (29.1%) assigned with unclear risk of performance bias (Table 2).

Our calculation of performance bias showed a significant difference ($P < 0.0001$, Table 1) in judgments between Cochrane authors and calculations of our algorithm. We found that Cochrane authors made inadequate judgments of performance bias in 1,828 (26.4%) trials. Our analysis yielded "worse" RoB judgment in 1,614 (88.3%) of those trials (i.e., our algorithm changed judgment from low to unclear, or unclear to high), and "better" RoB judgments in 214 (11.7%) trials (i.e., our algorithm changed judgment from unclear to low, or high to unclear), as shown in Table 3. We found that 7.3% of high-risk judgments made by Cochrane authors were calculated as unclear or low, whereas 19.9% of the assigned unclear risk judgments were calculated as either high or low risk. Half of the judgments (53.1%) assigned low risk of performance bias were calculated to be of unclear or high risk.

Table 3. Difference in judgment provided by the Cochrane review authors and judgment in line with the Cochrane Handbook (total 6,918 judgments)

Direction of change of RoB judgment	*N* (%)	Median LOC[b] (characters)	95% CI
To higher RoB	1,614 (23.3%)		
Low to high	95 (4.1%[a])	129	[111, 175]
Unclear to high	375 (18.6%[a])	104	[76, 118]
Low to unclear	1,144 (49.0%[a])	56	[51, 63]
To lower RoB	214 (3.1%)		
High to unclear	183 (7.1%[a])	16	[15, 22]
Unclear to low	26 (1.3%[a])	170	[104, 289]
High to low	5 (0.2%[a])	169[c]	[60, 278]
Unchanged	5,090 (73.6%)		
High	2,386 (92.7%[a])	71	[67, 73]
Unclear	1,610 (80.1%[a])	39	[33, 43]
Low	1,094 (46.9%[a])	129	[129, 135]

Abbreviations: LOC, length of comment; RoB, risk of bias.
[a] Percentage of assigned judgment subgroup.
[b] Length of comment.
[c] Normal distribution, arithmetic mean.

Cochrane authors frequently renamed RoB domain for performance bias to clarify the outcomes that the domain was evaluated for; only in 22.1% of analyzed performance bias domains, Cochrane authors did not specify which outcome domain was targeted. In 73.1% of the analyzed judgments, Cochrane authors specified that the performance bias domain refers to "all outcomes" (Table 4).

3.2. Uncertainty in supporting comments

The largest subgroup (Table 2) of judgments not in line with Cochrane Handbook was the one where Cochrane authors assigned low RoB judgment, while our algorithm calculated unclear risk of bias ($N = 1,143$, 62.6%). This was due to massive uncertainty (98.3% of cases) on whether both, participants and personnel, were blinded. In 84.6% of the cases, this uncertainty came from two statements: double-blinding and placebo.

3.3. Impact of length of supporting comment

We demonstrated that LOC supporting the judgment has an impact on assigned judgment, achievement of blinding and calculated judgment (Table 5). Shortest comments with a median of 44 characters were found in RoB tables with RCTs with unclear risk of performance bias assigned. The longer the comment (LOC over 90), the higher the chance for successful blinding was described, along with low RoB judgment for performance bias assigned by Cochrane authors or calculated according to the Cochrane Handbook. Comments with LOC between 62 and 82 usually supported high RoB judgment or described unsuccessful blinding.

Table 4. Distribution of outcomes by category and calculated judgment

			Calculated risk of bias judgment		
			High	Low	Unclear
Outcome category	***N***	**%**	***N***	***N***	***N***
All outcomes	5,054	73.1%	1,998	809	2,247
Not specified	1,527	22.1%	644	222	661
Objective	96	1.4%	48	42	6
Subjective	78	1.1%	59	4	15
Complications/adverse events	53	0.8%	43	4	6
Nonmortality outcomes	38	0.5%	24	14	0
Mortality	32	0.5%	5	27	0
Patient-reported/private phenomena	17	0.2%	15	1	1
Behavioral	13	0.2%	13	0	0
Assessor/clinician related	10	0.1%	8	2	0
Total	6,918	100.00%	2,857	1,125	2,936

We observed an increase in number of higher risk judgments and unclearness of blinding achievements. However, respective medians of LOCs of these groups moved apart (Table 3). Furthermore, out of 1,614 judgments where we calculated higher judgment of RoB, only 375 changed from unclear to high with a median LOC being 104 (95% CI 76 to 118). These judgments had significantly longer comments than the ones relating to unclear ($P < 0.05$) or high RoB ($P < 0.05$).

3.4. Splitting the performance bias domain based on different outcomes

Among all analyzed RCTs, we found 237 RCTs with more than one RoB judgment for performance bias domain, where multiple judgments were made for different outcomes. In those 237 RCTs, we found a total of 609 judgments for different outcomes. Distribution of judgments in this subgroup of trials was significantly different to main sample ($P < 0.001$)—see Table 6, with 61.7% (vs. 39.3%) judged as high, 20.9% (vs. 15.8%) as low, and 17.4% (vs. 44.8%) as unclear risk of performance bias.

In 139 RCTs, all judgments for risk of performance bias regarding different outcomes were the same, whereas in 98 RCTs, we found at least one judgment different between the subdomains for performance bias. Our assessment, based on the Cochrane Handbook, resulted in 77 RCTs (with 175 overlapping judgments) that did not have unanimous judgments. Out of 127 domains judged with low risk of performance bias both by Cochrane authors and by us, 69 (54.3%) were due to the comment that it was unlikely that blinding of participants and personnel may influence the outcome.

3.5. Splitting the domain based on blinded individuals—subdomains for participants and personnel

In 605 analyzed RCTs, Cochrane authors did not use the ''blinding of participants and personnel domain'' as one domain for performance bias; instead, they split the performance bias domain into the subdomains based on those two

Table 5. Impact of LOC to assigned judgment, calculated judgment, and achievement of blinding

Type of analysis	RoB judgment	*N* (%)	Median LOC (characters)	95% CI
RoB judgment assigned by Cochrane authors	High	2,574 (37.2%)	66	[62, 69]
	Unclear	2,011 (29.0%)	44	[44, 48]
	Low	2,333 (33.8%)	95	[90, 100]
RoB judgment calculated in this study	High	2,856 (41.3%)	73	[72, 75]
	Unclear	2,937 (42.4%)	43	[40, 44]
	Low	1,125 (16.3%)	129	[129, 136]
Blinding achieved	No	3,118 (45.1%)	78	[75, 82]
	Unclear	3,008 (43.5%)	44	[42, 45]
	Yes	792 (11.4%)	142	[133, 149]

Abbreviations: LOC, length of comment; RoB, risk of bias.

Table 6. Difference in the risk of bias judgment distribution when outcomes are split

Judgment	*N* (%)			*N* (%)		
Outcome	Assigned RoB judgment by Cochrane review authors			Calculated judgment in line with Cochrane handbook		
	Single	Split	Combined	Single	Split	Combined
High risk	2,253 (35.8%)	319 (52.5%)	2,573 (37.2%)	2,479 (39.3%)	375 (61.7%)	2,856 (41.3%)
Low risk	2,147 (34.1%)	185 (30.4%)	2,333 (33.7%)	998 (15.8%)	127 (20.9%)	1,125 (16.3%)
Unclear risk	1,902 (30.2%)	104 (17.1%)	2,007 (29.0%)	2,825 (44.8%)	106 (17.4%)	2,932 (42.4%)
Total	6,302 (100%)	608 (100%)	6,913 (100%)	6,302 (100%)	608 (100%)	6,913 (100%)

groups of individuals—one subdomain for blinding of participants and/or one subdomain for blinding of personnel. However, not all the trials had used both of those subdomains: 40 RCTs had only a subdomain for blinding the participants and 114 RCTs had only a subdomain for blinding the personnel. In 451 RCTs, both of those subdomains were used: 460 judgments for blinding of participants (higher number is due to nine trials with further division of the subdomain based on outcomes) and 451 judgments for blinding of personnel subdomain. We found a significantly lower proportion of unclear judgments in both subdomain for personnel (22.7%, $P < 0.01$) and subdomain for participants (18.4%, $P < 0.001$), in contrast to 29.1% in the main sample (Table 7).

When we combined judgments from these two subdomains, there was a 92.7% match in judgment categories (high, low, unclear) assigned by Cochrane authors and those calculated by our algorithm (Table 7). This was higher compared to the frequency of matching judgments for a single performance bias domain in the main data set, where accordance in judgments was 73.6%.

There was no statistically significant difference between assigned and calculated judgments in the subdomain for blinding of personnel ($P = 0.2565$, Table 1).

When RoB was judged as high, the overall match in the two groups of judgments was 96.0% (Table 7); when judged low, the match was 88.9%. When RoB was judged as unclear, similarly, high level of agreement was found only in the subdomain for blinding of personnel, with the 93.8% match; in the subdomain for blinding of participants, the match was 82.6%. This difference was significant ($P < 0.05$), but the sample was very small (see Table 1).

We found 451 cases of overlapping judgments for the same RCTs from same Cochrane reviews. Only 82 RCTs had their split domain differing in a total of 170 judgments. Overall, health care provider subgroup tended to expose the higher risk of bias with 28 being high, 47 unclear, and seven having low risk of bias judgments, whereas participant's blinding subgroup showed tendency of the overall lower RoB judgment, with 12 high-risk, 17 unclear, and 59 low-risk assignments.

Table 7. Risk of bias judgments distribution difference between split domain and main data set

Judgment assigned by Cochrane review authors	Judgment calculated according to the Cochrane handbook	Participants	Personnel	Sum of both split domains	Main data set
		N (%)	*N* (%)	*N* (%)	*N* (%)
High risk		237 (47.4%)	319 (56.5%)	556 (52.2%)	2,574 (37.2%)
	High risk	227 (95.8%)	307 (96.2%)	534 (96.0%)	2,386 (92.7%)
	Low risk	0 (0%)	0 (0%)	0 (0%)	5 (0.2%)
	Unclear risk	10 (4.2%)	12 (3.8%)	22 (4.0%)	183 (7.1%)
Low risk		171 (34.2%)	118 (20.9%)	289 (27.1%)	2,333 (33.7%)
	High risk	4 (2.3%)	3 (2.5%)	7 (2.4%)	95 (4.1%)
	Low risk	152 (88.9%)	105 (89.0%)	257 (88.9%)	1,094 (46.9%)
	Unclear risk	15 (8.8%)	10 (8.5%)	25 (8.7%)	1,144 (49.0%)
Unclear risk		92 (18.4%)[a]	128 (22.7%)[a]	220 (20.7%)[a]	2,011 (29.1%)[a]
	High risk	10 (10.9%)	6 (4.7%)	16 (7.3%)	375 (18.7%)
	Low risk	6 (6.5%)	2 (1.6%)	8 (3.6%)	26 (1.3%)
	Unclear risk	76 (82.6%)	120 (93.8%)	196 (89.1%)	1,610 (80.0%)
Matching judgments		455 (91.0%)	532 (94.2%)	987 (92.7%)	5,090 (73.6%)
Total		500 (100.0%)	565 (100.0%)	1,065 (100.0%)	6,918 (100.0%)

[a] $P < 0.05$.

4. Discussion

This study found that in a quarter of analyzed Cochrane reviews, the risk of bias assessments for blinding of participants and personnel were not in line with recommendations from the Cochrane Handbook, the key source of methodological guidance for Cochrane authors.

Our results can be used to improve future RoB assessments, and to guide further revisions of the RoB tool. We found a particularly high number of discrepancies when Cochrane authors judged the risk as low, mostly because Cochrane authors accepted statements such as ''double-blind'' and ''placebo'' as sufficient proof for this domain. Cochrane Handbook specifically indicates that the term double-blind is used inconsistently and makes it impossible to know who was blinded. However, many Cochrane authors have associated this expression with low RoB. The literature suggests seven groups of key individuals that may be blinded within a trial [9−12], and therefore, review authors should clearly specify whether a trial has provided information regarding specific groups of individuals.

Furthermore, our results indicate that more accurate judgments could be made if the performance bias domain is by default split into two subdomains, for participants and personnel. In the version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) [16], which is being developed, signaling questions are used and there are separate signaling questions about blinding of study participants, and ''carers and people delivering the intervention'' [16]. Findings of our study support this revision.

We found that Cochrane authors have customized some of the analyzed domains by specifying outcomes that the domain refers to. Review Manager (RevMan), software used for production of Cochrane Reviews, could be revised to prompt authors to choose whether their judgment refers to all outcomes, subjective or objective outcomes so that in future there would be no cases where this is not specified. However, splitting the domain based on different outcomes should be abandoned.

Owing to our finding that shorter length of supporting comments is associated with more erroneous judgments, a minimum character limit could be introduced in RevMan and authors warned when the comment is too short; a cutoff value we recommend is 44 characters.

Besides interventions at the level of the Cochrane RoB tool, interventions aimed at editorial/peer-review level could be introduced to help Cochrane authors in assessing RoB adequately. This study is part of a series of analyses about adequacy of Cochrane authors' use of Cochrane RoB tool [4−8]; because of the massive amount of data and messages that analysis of each domain entails, analysis of each domain is presented separately.

4.1. Limitations and strengths

Although we used software extensively, we cannot exclude the possibility of human error in data interpretation. Therefore, every categorization made by one author was checked by the first author (O.B.). Furthermore, it is possible that Cochrane authors' RoB judgments were correct, but appeared inadequate due to insufficient details in supporting comments; this could be verified in included trials or by contacting authors. In this study, we were focused only on Cochrane reviews because the use of Cochrane RoB tool is obligatory in Cochrane reviews. Nevertheless, our results can be useful also to authors of non-Cochrane systematic reviews who aim to use Cochrane RoB tool.

5. Conclusions

In Cochrane reviews, risk of bias assessments for blinding of participants and personnel, that is, performance bias, were frequently not in line with recommendations. Interventions for improving these assessments and their consistency should be considered.

Acknowledgments

Authors' contributions: Study design was carried out by L.P. and O.B. Data collection and analysis were performed by O.B., M.B., S.D., T.P.P., and M.C. Writing the first draft of the manuscript was carried out by O.B., T.P.P., and L.P. Critical revision of the manuscript was carried out by O.B., M.B., S.D., T.P.P., M.C., and L.P. Approval of the final version of the manuscript was performed by O.B., M.B., S.D., T.P.P., M.C., and L.P.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.05.012.

References

[1] 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:d5928.

[2] Version 5.1.0 [updated March 2011]. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: The Cochrane Collaboration; 2011. Available at. http://handbook-5-1.cochrane.org/. Accessed November 22, 2018.

[3] Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials 2006;3:360−5.

[4] Babic A, Pijuk A, Brazdilova L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, et al. The judgement of biases included in the category ''other bias'' in Cochrane systematic reviews of interventions: a systematic survey. BMC Med Res Methodol 2019;19:77.

[5] Babic A, Tokalic R, Silva Cunha JA, 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;19:1−10.

[6] Barcot O, Boric M, Poklepovic Pericic T, Cavar M, Dosenovic S, Vuka I, et al. Judgments of risk of bias associated with random sequence generation in trials included in Cochrane systematic reviews are frequently erroneous. BioRxiv 2018366674.
[7] Propadalo I, Tranfic M, Vuka I, Barcot O, Poklepovic Pericic T, Puljak L. In Cochrane reviews risk of bias assessments for allocation concealment was frequently not in line with Cochrane's Handbook guidance. J Clin Epidemiol 2019;106:10−7.
[8] Saric F, Barcot O, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. J Clin Epidemiol 2019;112:53−8.
[9] Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. J Clin Epidemiol 2002;55:787−90.
[10] Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001; 285:2000−3.
[11] Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Ann Intern Med 2002;136:254−9.
[12] Meinert CL. Masked monitoring in clinical trials–blind stupidity? N Engl J Med 1998;338:1381−2.
[13] Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg 2010;53: 345−8.
[14] Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol 2004;57:543−50.
[15] Moustgaard H, Bello S, Miller FG, Hrobjartsson A. Subjective and objective outcomes in randomized clinical trials: definitions differed in methods publications and were often absent from trial reports. J Clin Epidemiol 2014;67:1327−34.
[16] Higgins JPT, Sterne JAC, Savovic J, Page MJ, Hrobjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V, editors. Cochrane Methods Cochrane Database of Systematic Reviews.