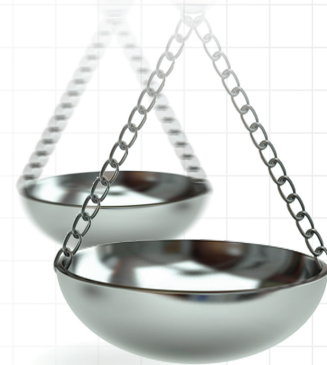




Assessing risk of bias judgments for blinding of outcome assessors in Cochrane reviews

Journal of **Comparative Effectiveness Research**

Ognjen Barcot^{*1}, Svjetlana Dosenovic², Matija Boric¹, Tina Poklepovic Pericic³, Marija Cavar⁴, Antonia Jelacic Kadic⁵ & Livia Puljak⁶

¹Department of Abdominal Surgery, University Hospital Split, Split, Croatia

²Department of Anesthesiology & Intensive Care, University Hospital Split, Croatia

³Department of Research in Biomedicine & Health, University of Split School of Medicine, Split, Croatia

⁴Department of Radiology, University Hospital Split, Split, Croatia

⁵Department of Pediatrics, University Hospital Split, Split, Croatia

⁶Center for Evidence-Based Medicine & Health Care, Catholic University of Croatia, Zagreb, Croatia

*Author for correspondence: ognjen.barcot@gmail.com

Aim: Adequate judging of risk of bias (RoB) for blinding of outcome assessors (detection bias) is important for supporting highest level of evidence. **Materials & methods:** Judgments and supporting comments for detection bias were retrieved from RoB tables reported in Cochrane reviews. We categorized comments, and then compared judgment and supporting comment with instructions from the Cochrane Handbook. **Results:** We analyzed 8656 judgments for detection bias from 7626 trials included in 575 reviews. Overall, 1909 judgments (22%) were not in line with the Cochrane Handbook. In 9% of trials, the authors split the detection bias domain according to outcomes. Here, prevalence of inadequate judgments was 19%. **Conclusion:** Interventions to improve RoB assessments in systematic reviews should be explored.

First draft submitted: 1 December 2019; Accepted for publication: 23 March 2020; Published online: 27 May 2020

Keywords: blinding • Cochrane • detection bias • risk of bias • systematic reviews • outcome assessors

Cochrane's risk of bias (RoB) tool is used for assessment of randomized controlled trials (RCTs) [1]. The tool comprises seven domains that are supposed to detect flaws in RCT methods. Two domains of Cochrane RoB tool assess whether key individuals were blinded participants and personnel (performance bias) and outcome assessors (detection bias) [1].

Detection bias, also called observer bias or ascertainment bias, could be particularly important if outcome assessors have strong preconceptions about an intervention, and when they need to assess subjective outcomes, including qualitative scoring, or recognizing patterns in images. Likewise, detection bias should not be very important for evaluating an objective outcome, such as death. A systematic review of Hrobjartsson *et al.* has shown that on average nonblinded outcome assessors of subjective binary outcomes in RCTs exaggerated odds ratios by 36% [2].

Even though blinding of outcome assessment is one of the key methodological components of RCTs, Kahan *et al.* have recently shown that blinding of outcome assessors was infrequently used in a cohort of analyzed trials, and when used it was often poorly reported [3].

Our research group has previously shown that Cochrane authors frequently do not assess RoB adequately in Cochrane reviews, including domains regarding randomization [4], allocation concealment [5], blinding of participants and personnel [6], selective reporting [7], attrition bias [8] and other bias [9]. In those studies, we compared judgments made by Cochrane authors with instructions from the Cochrane Handbook for Systematic Reviews of Interventions (Cochrane Handbook), which provides methodological guidance for conducting Cochrane reviews [1].

The aim of this study was to analyze whether judgments about the RoB associated with blinding of outcome assessors in Cochrane reviews of RCTs were adequate, in other words, in line with recommendations from the Cochrane Handbook.

Materials & methods

Study design

This was a primary methodological study, in which we analyzed methodology of Cochrane reviews published in the Cochrane Database of Systematic Reviews.

Inclusion & exclusion criteria

Cochrane Database of Systematic Reviews was searched for reviews of RCTs (or both RCTs and nonrandomized studies; but we analyzed RoB assessments only for RCTs) of interventions published from July 2015 to June 2016. Advanced search option was used to limit results to content type and publication date. We excluded diagnostic Cochrane reviews, overviews of systematic reviews, empty or withdrawn reviews and other Cochrane reviews containing no RCTs about interventions.

Screening for study eligibility

Titles and abstracts of Cochrane reviews were assessed by first author (O Barcot) who established inclusion eligibility. These assessments were verified by the second author (S Dosenovic).

Data extraction

The first author (O Barcot) wrote series of macroinstructions in Visual Basic for Applications (VBA, Microsoft, WA, USA) to automate data scraping from The Cochrane Library webpage to Microsoft Excel 2010 (Microsoft) workbook. Automatic extraction of RoB tables for every eligible Cochrane review was done offline with new set of coded instructions, as described earlier [6]. Errors during data extraction were logged and checked manually.

Development & testing of parser tool

The first author (O Barcot) developed a special user interface (MS Excel VBA User Form) to facilitate parsing. In this case, natural language text (comments, citations) was transformed to ordinal or nominal variables, as described earlier [6]. The second author (S Dosenovic) analyzed 500 random trials in order to pilot test and adjust the tool. This analysis was verified by the first author (O Barcot). Prior to usage of the tool, other authors were instructed to follow specific rules established in pilot testing.

Assessment of adequacy of Cochrane authors' detection bias assessment

In the developed user interface, we made new assessment of detection bias for RCTs in which Cochrane authors provided full detection bias assessment, in other words, both judgment (RoB is low, high or unclear) and an accompanying comment. We followed instructions for rating detection bias from the Cochrane Handbook (Section 8.12.2) [1] and defined that two main questions need to be correctly answered to adequately assess the detection bias. The question #1 was: who is assessing the outcome? – because different outcome assessors can be used, and it has to be assessed whether they were blinded or not. The question #2 was: is there a possibility of RoB in the outcome assessment? – because not all outcomes are equally prone to detection bias. For example, for an outcome such as death, lack of blinding of an outcome assessor may not influence the outcome. To appraise whether Cochrane authors mentioned type of outcome that was assessed within analyzed detection bias domain, we categorized every outcome into one of six predefined categories: objectively measured/subject-independent outcomes, clinician-rated/reported/related outcomes, patient/self-reported/rated outcomes, subjective outcomes, all outcomes, not specified, based on consideration how subjective or objective the outcome was [10]. The latter two categories do not relate to specific type of outcome in the same way as the first four do. 'All outcomes' mean that the authors judged RoB for all outcomes together, in other words, they did not specify certain subgroups of outcomes in the name of the RoB domain. We categorized that outcomes were 'not specified' when authors did not mention any particular type outcome with the name of a domain in the RoB table. Not specifying outcome that a domain is assessing is the default setting of the RevMan [11], software that is used for writing/analyses of Cochrane reviews. Lastly, we compared our new assessments with the assessments made by the Cochrane authors.

Primary outcomes

Judgments for detection bias assigned by Cochrane authors were analyzed by number, type and adequacy. Definition standard in our assessment was the Cochrane Handbook. We considered that judgment from Cochrane authors was inadequate if it did not completely adhere to the Cochrane Handbook guidance.

Secondary outcomes

We analyzed prevalence of splitting of the detection bias domain (whether Cochrane authors split a detection bias domain into two or more sub-domains), types of outcomes for which splitting was used, and adequacy of detection bias judgments in different outcome categories.

Statistics

We presented all descriptive data as frequencies and percentages. For all statistical tests we used type I error $\alpha = 0.05$, and type II error $\beta = 0.2$. Statistical analyses were performed using MedCalc for Windows, version 12.5.0.0 (MedCalc Software, Ostend, Belgium). All datasets were tested for normality by the Kolmogorov–Smirnov test. The Mann-Whitney test was utilized for comparison of independent samples of nonparametric data, and the Wilcoxon test was used for paired samples. Difference in proportions was tested with Chi-squared test. Hypotheses, outcome measures, statistical tests used and its results are logged in [Supplementary File 1](#).

Results

Included reviews, trials, judgments & outcomes

Out of 955 retrieved Cochrane reviews, 227 were not eligible. In the remaining 728 reviews there were 10,523 trials. Additional 2897 trials were excluded: all of 2480 trials from 141 reviews in which performance and detection bias domains were merged; all of 379 trials from 11 reviews and additional 28 trials without detection bias domain stated in the RoB table; one review with four trials without RoB table and additional four trials without RoB table as well; two trials were duplicate entries due to computational error ([Figure 1](#)).

Finally, we included in our study RoB tables from 575 Cochrane reviews (listed in [Supplementary File 2](#)), which included a total of 7626 trials ([Figure 1](#)). In those 7626 trials there were 8656 domains (judgments) for detection bias, because in some Cochrane reviews this domain was split (had multiple assessments for various types of outcomes). In 720 out of 7626 (9.4%) trials, RoB domain for detection bias was split into multiple subdomains based on different outcomes (ranging from 2 to 8). In those 720 trials, there were 1750 judgments for specific outcomes.

In the whole sample of 8656 detection bias judgments, for the majority (7110/8656; 82%) Cochrane authors specified in the RoB table that the judgment referred either to ‘all outcomes’ (6072/8656; 70%) or did not specify to which outcomes the domain was referring to (1038/8656; 12%). For the remaining 18% of judgments, Cochrane authors specified to which outcomes the domain (or subdomain) was referring to ([Table 1](#)).

Distribution of six categories of outcomes in the whole sample (8656 judgments) and a subsample of trials with split domain ($N = 720$ trials, $N = 1750$ judgments) was significantly different ($p < 0.05$; [Supplementary File 1](#)), as in this subsample only 4% judgments were specified for ‘all outcomes’ and 14% were not specified ([Table 2](#) and [Supplementary File 3](#)). In 231 trials with split domain (accounting for 540 judgments), risk of detection bias judgment was identical within all of their split outcomes (in a single trial, all of the RoB judgments were of the same level: all high, all low or all unclear).

Adequacy of Cochrane authors’ judgments for risk of detection bias

In the main analysis, among 8656 judgments for detection bias, there were 1679 (19%) that Cochrane authors judged with high risk, 3374 (39%) with low risk and 3603 (42%) with unclear RoB ([Table 3](#)).

Out of 8656 detection bias judgments, 6747 (78%) were judged adequately by Cochrane authors. Our assessment of adequacy for those judgments indicated that the highest prevalence of inadequate judgments was found for trials judged with low risk of detection bias (1287 of 3374; 38%), followed by those judged with high risk (239 of 1679; 14%) and those judged with unclear risk (383 of 3603; 11%; [Table 3](#)). For transparency purposes and ease of visualization we provided a table ([Supplementary File 4](#)) that presents a few detailed examples for each type of disagreement in decisions (between Cochrane authors and our team) for different types of outcome categories.

Adequacy of judgments for ‘all outcomes’ and judgments where no outcomes were specified was 77 and 78%, respectively ([Table 2](#) – right column). Among the remaining 18% of judgments, higher adequacy versus 78% adequacy in whole sample (all $p < 0.05$; [Supplementary File 1](#)) was observed among outcomes described as clinician-rated (84%), patient-rated (87%) and subjective outcomes (93%). These three groups of outcomes are considered subjective and we noticed significantly higher accuracy of judgments in merged group of outcomes compared with whole sample (86.9 vs 77.9%; $p < 0.05$; [Supplementary Files 1 & 5](#)) as well as in comparison of

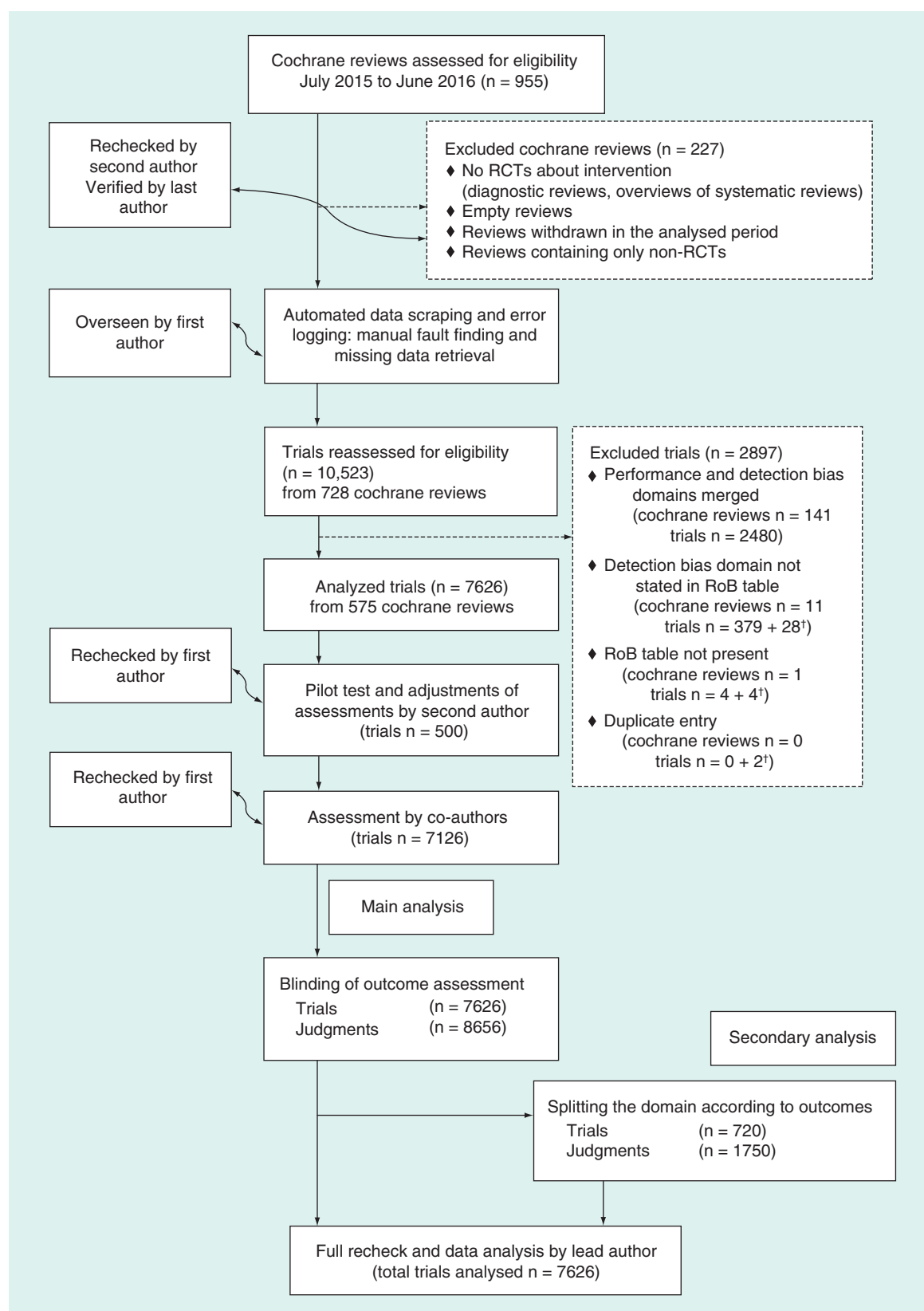


Figure 1. Flow diagram of the progress through the phases of the study.

†Only trials excluded, not whole Cochrane reviews.

RCT: Randomized controlled trials; RoB: Risk of bias.

Table 1. Distribution of judgments by Cochrane authors, according to categories of outcomes and judgments.

Outcome category	Reassessment of judgments in this study							
Judgment by Cochrane authors	High risk		Low risk		Unclear risk		Total	
All outcomes	1116	(18%)	1255	(21%)	3701	(61%)	6072	(70%)
High risk	908	(81%)	5	(0%)	180	(5%)	1093	(18%)
Low risk	41	(4%)	1208	(96%)	977	(26%)	2226	(37%)
Unclear risk	167	(15%)	42	(3%)	2544	(69%)	2753	(45%)
Clinician-related/-rated/-reported	144	(22%)	185	(28%)	340	(51%)	669	(8%)
High risk	130	(90%)	1	(1%)	2	(1%)	133	(20%)
Low risk	1	(1%)	182	(98%)	85	(25%)	268	(40%)
Unclear risk	13	(9%)	2	(1%)	253	(74%)	268	(40%)
Not specified	239	(23%)	247	(24%)	552	(53%)	1038	(12%)
High risk	221	(92%)	2	(1%)	35	(6%)	258	(25%)
Low risk	2	(1%)	240	(97%)	172	(31%)	414	(40%)
Unclear risk	16	(7%)	5	(2%)	345	(63%)	366	(35%)
Objectively measured/subject independent	0	(0%)	408	(100%)	0	(0%)	408	(5%)
High risk	0	–	9	(2%)	0	–	9	(2%)
Low risk	0	–	292	(72%)	0	–	292	(72%)
Unclear risk	0	–	107	(26%)	0	–	107	(26%)
Patient/self-reported/-rated	154	(72%)	55	(26%)	5	(2%)	214	(2%)
High risk	131	(85%)	5	(9%)	0	(0%)	136	(64%)
Low risk	1	(1%)	50	(91%)	0	(0%)	51	(24%)
Unclear risk	22	(14%)	0	(0%)	5	(100%)	27	(13%)
Subjective	59	(23%)	116	(45%)	80	(31%)	255	(3%)
High risk	50	(85%)	0	(0%)	0	(0%)	50	(20%)
Low risk	1	(2%)	115	(99%)	7	(9%)	123	(48%)
Unclear risk	8	(14%)	1	(1%)	73	(91%)	82	(32%)
Total	1712	(20%)	2266	(26%)	4678	(54%)	8656	(100%)
Bold font is used to emphasize subgroups.								

Bold font is used to emphasize subgroups.

Table 2. Distribution and adequacy of judgments according to outcomes.

Types of outcomes	Judgments for domains that were split for various outcomes											
	Overall			Different			All the same			Whole sample		
	N	(%N)	Adeq	N	(%N)	Adeq	N	(%N)	Adeq	N	(%N)	Adeq
All	77	(4%)	68%	48	(4%)	52%	29	(5%)	93%	6072	(70%)	77%
Clinician RRR	657	(38%)	84%	427	(35%)	82%	230	(43%)	89%	669	(8%)	84%
Not specified	247	(14%)	79%	166	(14%)	71%	81	(15%)	95%	1038	(12%)	78%
Objective†	407	(23%)	71%	317	(26%)	64%	90	(17%)	97%	408	(5%)	72%
Patient RRR	173	(10%)	84%	117	(10%)	83%	56	(10%)	88%	214	(2%)	87%
Subjective	189	(11%)	92%	135	(11%)	89%	54	(10%)	98%	255	(3%)	93%
Total	1750		81%	1210		75%	540		92%	8656		78%

† Objectively measured/subject independent.

Bold font is used to emphasize totals and subtotals.

Adeq: Adequacy; RRR: -rated, -related, -reported.

this merged subgroup to subsample with split domain (85.7 vs 80.6%; $p < 0.05$). Finally, judgments of objective outcomes had lower adequacy (72%) compared with the whole sample (78%; $p < 0.05$; [Supplementary File 1](#)).

Subsample of trials with split domain where all judgments in a trial were the same (i.e., all judgments were low RoB; [Table 2](#)) showed higher adequacy of judgments (92 vs 78% in whole sample $p < 0.05$). On the other hand, judgments in a subsample of trials that split domain in various outcomes and judged the risk of detection bias differently were as adequate as the whole sample (76 vs 78%; $p = 0.0553$).

Table 3. Adequacy of judgments for detection bias.

Judgment by Cochrane authors	Reassessment of judgments in this study			Total N (%)	Inadequate judgments N (%)
	High risk	Low risk	Unclear risk		
High risk	1440	22	217	1679 (19.4%)	239 (14.2%)
Low risk	46	2087	1241	3374 (39.0%)	1287 (38.1%)
Unclear risk	226	157	3220	3603 (41.6%)	383 (10.6%)
Total	1712 (19.8%)	2266 (26.2%)	4678 (54.0%)	8656 (100.0%)	1909 (22.1%)

Bold font is used to emphasize totals.

Table 4. Impact of length of comment to assigned judgment, calculated judgment and achievement of blinding.

Observation of impact	Impact	N (%)	Median LOC (characters)	95% CI
RoB judgment assigned by Cochrane authors	High	1712 (19.8%)	78	[75–85]
	Unclear	4678 (54.0%)	45	[44–45]
	Low	2266 (26.2%)	102	[97–108]
RoB judgment calculated in this study	High	1679 (19.4%)	70	[64–74]
	Unclear	3603 (39.0%)	43	[40–44]
	Low	3374 (39.0%)	94	[91–97]
Blinding achieved	No	1710 (19.8%)	79	[76–85]
	Unclear	5436 (62.8%)	48	[46–49]
	Yes	1510 (17.4%)	106	[101–113]

LOC: Length of comment; RoB: Risk of bias.

Length of comment (LOC) supporting the judgment has an impact on assigned judgment, achievement of blinding and calculated judgment. Shortest comments with a median of 45 characters were found in RoB tables of trials with unclear risk of detection bias assigned (Table 4). The longer the comment (LOC over 91) – the higher the chance for successful blinding was described, along with low RoB judgment for detection bias assigned by Cochrane authors or calculated according to the Cochrane Handbook.

Discussion

The main finding of this study is that 22% of RoB judgments for detection bias in analyzed Cochrane reviews were inadequate, because those judgments were not supported by accompanying comments. We found more adequate judgments in trials where Cochrane authors divided the RoB domain for detection bias into different categories of outcomes, such as subjective or objective outcomes.

When we previously analyzed the domain for blinding of participants and personnel, in other words, performance bias, we concluded that there were four aspects that need to be considered for making a judgment: who was blinded; was blinding achieved; outcome category and whether outcome may be influenced by blinding [6].

The same principles can be applied for making judgments about detection bias. The first question is who was/were outcome assessors(s) and whether this was clearly specified in a trial report. A trial usually has multiple outcomes and therefore there may be multiple outcome assessors. The second question is whether outcome assessors were blinded. The third and fourth questions are which types of outcomes are assessed, and whether results for those outcomes can be influenced by a lack of blinding of an assessor. All this information is often not reported in a trial, or is poorly reported, as it has been shown by Kahan *et al.* [3].

However, trialists are not the only ones with poor descriptions of outcomes that were judged. Descriptive terms used by Cochrane authors to describe outcomes that were assessed are often not sufficiently specific. An example is descriptor ‘objective’ outcome. In English dictionary [12], the term objective (as an adjective) means ‘not influenced,’ and thus suggesting a strict logical conclusion of low RoB. If an outcome is erroneously categorized as objective, this automatically leads to faulty judgment. One such example is a length of stay [13] – a numerical value that may be considered objective, but it is actually highly dependent on subjective decisions of an attending clinician. Moustgaard *et al.* [10] specifically discussed this issue of objectiveness versus clinical relevance. The same study stated three definitions of a term ‘subjective’ and, according to these, ‘objective’ terms were defined as opposites.

We have shown that Cochrane authors have rarely described outcomes as either subjective or objective. It is also not sufficient to simply indicate that an outcome was assessed by a clinician. Some Cochrane authors described an outcome as clinician-rated, with specific examples of such outcome, which clarifies the nature of an outcome and enables easier judgment whether an outcome assessment is prone to subjectivity. A statement that an outcome was patient-reported or self-reported, implies that assessment was subjective. Clear descriptions of outcomes that were assessed can be built into newer versions of RoB assessment tools in form of programmed rules.

Although the central part of this study (and the observed domain) was to emphasize the differentiation of outcomes as a more specific factor we must also mention how length of supporting comment (LOC) opacifies the justification of decision. Even in our previous study [6], we demonstrated that LOC impacts adequate (transparent) description of successful blinding and the same applies for the detection bias domain as well.

Should we split detection bias domain per outcomes, which outcomes & why?

We noticed significantly higher accuracy of judgments for outcomes that were described as subjective, compared with other categories of outcomes, both in trials with split domains and in the entire sample of trials. We also found an inversion in proportions of adequate judgments depending on whether a domain for detection bias was split into subdomains or not. In trials where the domain was not split, proportions of adequate judgments for specified outcomes versus those that indicated they judged all outcomes or did not specify types of outcomes were 18 versus 82%, respectively. Among trials that had split domain for various outcomes, the proportion of adequacy was inverse – 82% when subcategories of outcomes were specified versus 18% for all outcomes or when outcomes were not specified. Furthermore, in a subgroup of trials with split domains, there was higher prevalence of inadequate judgments for ‘objective outcomes’ than for other groups of outcomes. We believe this is due to erroneous categorization of outcomes due to lack of specific instructions in the Cochrane Handbook about which outcomes should be considered objective.

Splitting detection bias judgment into multiple subdomains seems to have some specifics. First, this subgroup seemed to have higher accuracy of judgments, compared with the whole sample, but this is due to distribution of outcomes toward subjective ones. Subjective outcomes had higher accuracy of judgments but larger number of cases diminishes the effect of lower adequacy of misjudged objective outcomes. Second, the authors frequently assigned the same RoB judgment for all outcome types within a single trial, regardless of the outcome category. Third, although this was rare, in 4% of reviews with split outcomes, the authors also used subdomain ‘all outcomes’ together with one or more specific outcomes as well. If the authors decided to split domain into multiple subdomains based on specific outcomes, it is unclear why they would then, additionally, use subdomain for ‘all outcomes.’

While splitting of outcomes may be associated with errors due to erroneous categorization of certain outcomes, it provides more information to the readers than overall judgment for all outcomes, and therefore we can conclude that assessment of overall risk of detection bias for all outcomes should not be used. At least, the authors should define the category and clarify that all outcomes belong to the same category, and that this was the reason why they were all judged together.

Cochrane has recently published new version of its RoB tool [14]. At the time of the submission of this manuscript (December 2019), the new tool has not been implemented yet in Cochrane reviews. Even the authors of developing Cochrane protocols that were not published yet were not obliged to use the RoB 2.0 version. The RoB 2.0 has changed content, structure and type of applicable judgments. The current (old) RoB tool has seven domains, and each domain can be judged with three types of judgments – risk is considered to be high, low or unclear. The RoB 2.0 tool has five domains; each domain has from three to seven signaling questions; signaling questions are responded with five potential answers (yes, probably yes, no, probably no, no information). Based on the answers on signaling questions each domain is scored as low RoB, high RoB or having some concerns. Overall RoB assessment is also provided. Domain #4 is ‘Bias in measurement of the outcome’; this domain has three signaling questions referring to blinding of outcome assessors, asking whether outcome assessors were blinded to the intervention, whether assessment of an outcome could be influenced by lack of blinding of outcome assessor and whether it is likely that the assessment was actually influenced by the lack of blinding of outcome assessor. These three signaling questions represent major change from the current RoB tool. Based on the results of this study, we expect that these three signaling questions should facilitate detection bias judgments, and should help in making more transparent, adequately explained decisions.

Results of our study can help improve adequacy of RoB assessment in systematic reviews. Even though Cochrane has announced RoB 2.0 tool, the ‘old’ tool is still in use. Additionally, the ‘old’ tool is also used by the majority of

non-Cochrane reviews [15]. With the availability of our data, presented in this study, once the RoB 2.0 tool will be used in all Cochrane reviews, we will be able to compare in future whether adequacy of judgments of Cochrane authors for this particular domain has improved with the new tool. Therefore, our results can be useful to authors who will continue to use the current RoB tool, as well as from the research methodology perspective for measuring adequacy of methods in the new RoB 2.0 tool, and comparing it with the results for the old tool. New research methods continue to be developed, but it is important to make sure that they are better than previous methods.

Our study is important because it is yet another confirmation that more attention should be paid to Cochrane methods used by Cochrane authors. Erroneous judgments of RoB will lead to erroneous conclusions, which ultimately can send misleading messages to consumers, healthcare workers and decision makers. Editors and peer-reviewers could help ensure adequate use of systematic review methods.

Limitations & strengths

In this study, we have used software for data extraction, but it is possible that we have made errors in data interpretation. We tried to avoid this possibility by double-checking all decisions by two independent authors. We also need to state that in our study we did not try to detect possible bias in original trials, for example using the Berger-Exner test [16–18], because we did not have proper input for it to be executed. Even in our previous works on bias arising from inadequate random sequence generation [4] and allocation concealment [5] this was the case. Our dataset of Cochrane authors' comments and judgments, on which we performed analysis, did not include analysis of full texts of original trials. Instead, we only relied on the content of RoB tables from Cochrane reviews, and these tables only include information that the Cochrane authors have reported in them. This could have helped verifying if the judgment of Cochrane authors was adequate, when authors have failed to provide adequate explanation in their supporting comment.

Conclusion

We found that Cochrane reviews frequently had inadequate judgments for risk of detection bias. We expect that the new version of the Cochrane RoB tool 2.0, which uses three signaling questions to facilitate detection bias judgments, should help in making more transparent, adequately explained decisions. It would be worthwhile to explore interventions that would help ensure adherence to methodological guidance among systematic review authors. RoB judgments are incorporated into systematic review conclusions, and it is in the interest of the entire medical community to have trustworthy evidence.

Summary points

- More than a fifth of risk of bias (RoB) assessments for blinding of outcome assessors were not in line with Cochrane Handbook.
- Highest prevalence of inadequate judgments was found for outcomes categorized as objective.
- Splitting the domain according to outcomes does not increase overall adequacy of judgments, but gives better insight of types of outcomes and actual risks.
- Our previous research has shown similar proportions of inadequate judging as in this trial, but mostly due to insufficient information supporting the judgment. This research demonstrated the importance of categorization of outcomes to achieve adequate judging in Cochrane systematic reviews.
- Observation of trials that have split this domain according to outcomes demonstrated that better focus on outcomes leads to more consistent judging and higher accuracy when subjective outcomes are judged.
- The same observation revealed that faulty categorization of an outcome that is considered to be objective leads to most of inadequate judging.
- Software solutions could be used to encourage authors to primarily define the outcomes and reinvestigate actual objectivity. Obligatory usage of dichotomous splitting of this domain could bring more information about risk of detection bias. This would ensure consistent methodological approach to assessment of risks of bias in Cochrane reviews.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/ce-2019-0181

Author contributions

L Puljak and O Barcot were responsible for the study design. O Barcot, S Dosenovic, M Boric, T Poklepovic Pericic, M Cavar and A Jelacic Kadic contributed for the acquisition, analysis or interpretation of data for the work. O Barcot and L Puljak have written the first draft of the manuscript. O Barcot, S Dosenovic, M Boric, T Poklepovic Pericic, M Cavar, A Jelacic Kadic and L Puljak were responsible for the critical revision of the manuscript. O Barcot, S Dosenovic, M Boric, T Poklepovic Pericic, M Cavar, A Jelacic Kadic and L Puljak have approved final version of the manuscript to be published. O Barcot, S Dosenovic, M Boric, T Poklepovic Pericic, M Cavar, A Jelacic Kadic and L Puljak were responsible for the agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved as stated in Author Disclosure Form FSG.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

- Higgins JPT, Green S (Eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. (2011). The Cochrane Collaboration, London, UK. <http://handbook-5-1.cochrane.org/>
- Hrobjartsson A, Thomsen AS, Emanuelsson F *et al*. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 344, e1119 (2012).
- Kahan BC, Rehal S, Cro S. Blinded outcome assessment was infrequently used and poorly reported in open trials. *PLoS ONE* 10(6), e0131926 (2015).
- Barcot O, Boric M, Poklepovic Pericic T *et al*. Risk of bias judgments for random sequence generation in Cochrane systematic reviews were frequently not in line with Cochrane Handbook. *BMC Med. Res. Methodol.* 19(1), 170 (2019).
- Propadalo I, Tranfic M, Vuka I, Barcot O, Poklepovic Pericic T, 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.* 106, 10–17 (2019).
- Barcot O, Boric M, Dosenovic S, Poklepovic Pericic T, Cavar M, Puljak L. Risk of bias assessments for blinding of participants and personnel in Cochrane reviews were frequently inadequate. *J. Clin. Epidemiol.* 113, 104–113 (2019).
- Saric F, Barcot O, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. *J. Clin. Epidemiol.* 112, 53–58 (2019).
- Babic A, Tokalic R, Amilcar Silva Cunha J *et al*. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Med. Res. Methodol.* 19(1), 76 (2019).
- Babic A, Pijuk A, Brazdilova L *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.* 19(1), 77 (2019).
- 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.* 67(12), 1327–1334 (2014).
- Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration (2014). <https://revman.cochrane.org/>
- Cambridge Dictionary (2019). <https://dictionary.cambridge.org/dictionary/english/objective>
- De Jong JD, Westert GP, Lagoe R, Groenewegen PP. Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work? *Health Serv. Res.* 41(2), 374–394 (2006).
- Sterne JaC, Savovic J, Page MJ *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366, l4898 (2019).
- Puljak L, Ramic I, Arriola Naharro C *et al*. Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews. *Journal of Clinical Epidemiology* 123, 114–119 (2020).
- Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. *Control. Clin. Trials* 20(4), 319–327 (1999).
- Mickenausch S, Fu B, Gudehithlu S, Berger VW. Accuracy of the Berger-Exner test for detecting third-order selection bias in randomised controlled trials: a simulation-based investigation. *BMC Med. Res. Methodol.* 14, 114 (2014).
- Berger V. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. John Wiley & Sons, Chichester, UK (2005).