International Journal of Clinical Pharmacy (2019) 41:785–792
https://doi.org/10.1007/s11096-019-00818-2

RESEARCH ARTICLE

Usability and sensitivity of the risk of bias assessment tool for randomized controlled trials of pharmacist interventions

Fernanda S. Tonin[1] · Livia A. Lopes[1] · Inajara Rotta[2] · Aline F. Bonetti[1] · Roberto Pontarolo[3] · Cassyano J. Correr[3] · Fernando Fernandez-Llimos[4]

Received: 3 July 2018 / Accepted: 27 March 2019 / Published online: 9 April 2019

Abstract

Background The Cochrane collaboration risk of bias assessment (RoB) tool is used in several fields to evaluate the methodological quality of studies. Its strengths and challenges are discussed. *Objective* To assess the sensitivity of the RoB tool in studies of pharmacist interventions. *Setting* DEPICT database was used to pool randomized controlled trials (RCTs) of complex interventions. *Method* A Guide for RoB Judgment in Pharmacy Services was created to help in the interpretation and judgment of bias criteria. The evaluation of bias (low, unclear, high risk) was performed by RCT. Sensitivity analyses were performed to assess the influence of different interpretations of eight elements of judgment in the RoB tool. Paired analysis and estimations of the effect size (95% confidence interval) of the criteria modifications compared to the original analyses were calculated. *Main outcome measure* Changes in the interpretations of judgment in the RoB tool. *Results* Overall, 8.3, 45.4, and 46.3% of the studies were determined to have low, unclear, and high risk of bias, respectively. High risk of bias was caused by attrition and detection domains. The number of studies classified with high risk of bias significantly increased for five of the eight interpretations, while unclear risk of bias increased for three interpretations (with a negligible effect size in all of them). Lack of blinding, loss of participants, and the use of subjective and self-reported outcomes were the main elements resulting in high risk of bias. *Conclusion* The RoB tool is useful for evaluating RCTs of pharmacist interventions if adapted criteria for judgment are used. Ignoring these adjustments produces a floor-effect with studies classified with high risk of bias.

Keywords Cochrane collaboration · Methodology · Outcome assessment · Pharmacists · Risk of bias · RoB tool

Impacts on practice

- The criteria for judgment of the Cochrane collaboration risk of bias assessment (RoB) tool should be adapted to each subject area to avoid a floor-effect and risk of biased results.
- A Guide for RoB Judgment in Pharmacy Services will help in the interpretation and judgment of bias criteria in randomized controlled trials of clinical pharmacy interventions.
- High risk of bias in trials of pharmacist interventions is especially caused by lack of blinding, use of self-reported outcomes, losses of participants, and use of subjective outcomes.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11096-019-00818-2) contains supplementary material, which is available to authorized users.

✉ Fernando Fernandez-Llimos
f-llimos@ff.ul.pt

[1] Pharmaceutical Sciences Postgraduate Programme, Federal University of Paraná, Curitiba, Brazil

[2] Hospital de Clínicas, Federal University of Paraná, Curitiba, Brazil

[3] Department of Pharmacy, Federal University of Paraná, Curitiba, Brazil

[4] Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, University of Lisbon, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal

Introduction

The assessment of research quality involves evaluation of the internal and external validity of a given study [1, 2]. Quality assessment has also been used to define a threshold for the selection of primary studies, such as randomized controlled trials (RCTs) for systematic reviews, to weigh and explore study results, and to guide the interpretation of findings for clinical practice [3–6]. However, while the impact of RCT bias on evidence synthesis has been largely recognized, the approaches to methodological quality assessment have been inconsistent and controversial [7–10]. A wide variety of tools have been developed to evaluate the quality of RCTs in different health areas, but many of them have not been developed using scientifically rigorous methods nor have they been validated [10–13]. Subsequently, a RCT may be rated very differently by different tools, which can impact treatment recommendations [14, 15].

In 2008, the Cochrane Collaboration designed the risk of bias (RoB) assessment tool, which assesses the risk of bias for RCTs and was updated in 2011 [16]. This tool has been widely used in the literature [7] and is based on seven bias domains: random sequence generation and allocation concealment (both within the domain of selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias. For each domain, the tool makes users assign a judgment of "high," "low", or "unclear" risk of bias and document the basis for their judgments [16]. However, controversial opinions among researchers on the strengths and challenges of this tool persist [1, 17]. Mostly because, despite providing a standardized approach for quality assessment based on both theoretical and empirical grounds, the RoB tool presents low consistency among researchers (modest inter-observer agreement) [11, 18]. Additionally, the domains of attrition and reporting bias are usually noted as the most difficult to judge [18–20]. On the other hand, the subjective interpretation of this tool, along with its versatility, is an important benefit that allows its use in different fields [17, 20]. In an attempt to overcome some of the issues of the RoB tool, a revised version of risk of bias in randomized trials (RoB 2.0) is under development in a pilot study [21].

In contrast to what happens with drug therapy RCTs, there is only limited data on the specific instruments used to assess the methodological quality of trials of pharmacy services. Clinical pharmacy services are now recognized as an integral resource of the healthcare system [22], but RCTs may be more difficult to judge since they include a number of different pharmacist interventions [23]. Additionally, no literature on the validity, reliability, and responsiveness of the RoB tool exists in this field.

Aim of the study

We aimed to evaluate the usability and sensitivity of the Cochrane Collaboration RoB assessment tool for RCTs of pharmacist interventions by adjusting the criteria for judgment.

Ethics approval

Ethical approval for conducting this study was not necessary.

Methods

Study selection

To gather all the published RCTs assessing the impact of pharmacists' interventions, the Descriptive Elements of Pharmacist Interventions Characterization Tool (DEPICT) Project database of RCTs (second update) was used. Criteria for the inclusion of trials in the database are described in detail on the project website (depict.org) and in previous publications [24–26]. The final pool comprised 432 RCTs published in 517 articles.

Criteria for judgment and tool adjustments

According to the recommendations of the Cochrane Collaboration [16], the criteria for risk of bias judgment for each domain of the RoB tool should take into account the characteristics of the interventions under evaluation. To adapt the RoB criteria for judgment, we discussed with DEPICT project researchers all items of the RoB tool, considering the complexity and specificity of pharmacist interventions, before reaching a consensus to be applied on the judgements. Finally, a "Guide for RoB Judgment in Pharmacy Services" was created following the Cochrane Collaboration recommendations [16] and published literature [17, 27].

To refine this guide, a sample containing 10% of the RCTs from the DEPICT database was randomly selected using the Randomizer software (randomizer.org). The RoB tool was then applied for each trial of this sample by two independent researchers following the instructions of the "Guide for RoB Judgment in Pharmacy Services". Discrepancies between reviewers were discussed and the criteria of judgment of the "Guide for RoB Judgment in Pharmacy Services" were adjusted accordingly when necessary. When there was no consensus between the two researchers, a third was consulted. The final version of the "Guide for RoB Judgment in

Pharmacy Services" with a full description of the criteria for the assessment of the risk of bias is available in the online appendix.

Definition of the outcomes and tool application

The risk of bias in RCTs must be assessed for one specific outcome [16]. Since considering different outcomes of a given study may generate different risk of bias judgments, we established a sequential procedure to identify the primary outcome before applying the RoB tool:

1. Outcome considered by the authors as the primary outcome;
2. Outcome considered by the authors in the definition of the sample calculation;
3. Outcome mentioned by the authors in the objectives of the research;

Primary outcomes were later classified as objective or subjective based on their method of measurement [28]. Outcomes were considered objective if a) they were the result of a biochemical or physical measurement, or b) resulted from the strict application of a definition, which allowed no variability of interpretation. Subjective outcomes were those where the judgment depended on the researchers' or individuals' interpretation. Subjective outcomes were classified as validated (i.e., measured by an instrument with validated psychometric characteristics) and invalidated outcomes.

After defining outcomes, the RoB tool was applied to each RCT considering the final version of "Guide for RoB Judgment in Pharmacy Services". The seven domains (random sequence generation and allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias) were judged as "high," "low", or "unclear" risk of bias. All studies were considered as low risk of bias for the performance bias domain due to the difficulties of blinding participants/personnel for non-pharmacological interventions. The results obtained in this first assessment were labeled as "original analysis" of the risk of bias.

Sensitivity and descriptive analyses

After performing the original analysis for the assessment of risk of bias in RCTs of pharmacist interventions, sensitivity analyses were performed to assess the influence of different interpretations of the RoB tool's criteria for judgmentin the results. Sensitivity analyses consisted in the modification of the criteria for judgement of eight elements potentially producing bias in the domains of "performance", "attrition", and "detection". Table 1 presents the detailed description of the modified criteria for judgement. Then, eight new RoB assessments using the "Guide for RoB Judgment in Pharmacy Services", but now considering each of the modified judgement criteria, were performed separately.

Descriptive analyses of the overall risk of bias results for the original and the sensitivity analyses (i.e. with the modified judgement criteria), along with the proportion of studies with domains of "high", "low", and "unclear" risks of bias, were performed. Normality of the distributions was evaluated by the Kolmogorov–Smirnov test. The paired analysis by McNemar–Bowker to test if a square table is symmetric, was used to evaluate the changes of the risk of bias during the sensitivity analyses. The results were considered statistically significant at $p < 0.05$ [29]. Cliff's delta

Table 1 Modifications in the interpretation of the guide for RoB judgment in pharmacy services criteria to perform sensitivity analyses

Domain	Modified judgment criteria	Support for judgment
Performance bias		
P01	No participant blinding in self-report outcomes	High risk of bias
P02	Contamination of the blinding in the intervention/control groups (e.g., single professional performed the evaluation in both groups)	High risk of bias
Attrition bias		
A01	Reduction of the cut-off for loss of participants/data from 20 to 10%*	High risk of bias
A02	Differences of 10% or more of losses between intervention and control groups	High risk of bias
A03	Difference of 20% or more of losses between intervention and control groups	High risk of bias
A04	Lack of reasons for loss of participants in the intervention/control groups, individually	Unclear risk of bias
A05	Lack of description of the statistical method by intention-to-treat to treat the lost data	Unclear risk of bias
Detection bias		
D01	Lack of validated instrument for subjective outcomes	High risk of bias

*For the original analyses of the overall risk of bias the categories were: low risk of bias ≤ 10% of losses of participants/data; unclear risk of bias between 10 and 20%; high risk of bias ≥ 20%

calculation, a non-parametric effect size measure that quantifies the amount of difference between two groups of ordinal observations beyond *p*-values interpretation [30], was used to estimate the effect size differences (with 95% confidence intervals—CI) between the risk of bias original analysis *versus* the results obtained in the sensitivity analyses with the eight modified judgement criteria. The magnitude of the changes between original *versus* sensitivity analyses was assessed using the thresholds $|d| < 0.147$ "negligible", $|d| < 0.33$ "small", $|d| < 0.474$ "medium", otherwise "large". The analyses were conducted in IBM SPSS Statistics v. 24.0 (Armonk, NY: IBM Corp.) and software R v.3.4.1/R studio 1.0.153.

Results

The initial overall risk of bias of the 432 RCTs showed that 8.3% of studies (n = 36) were of low risk of bias, while 196 (45.4%) presented unclear risk of bias, and 200 (46.3%) were high risk (see Fig. 1). The judgments for each domain are presented in Table 2. The domains with the highest risk of bias were "attrition bias" (n = 124) followed by "detection bias" (n = 60), whereas the domains most judged as unclear risk were from the selection bias domain—"allocation concealment" (n = 303) and "random sequence generation" (n = 205).

Despite initial analyses that considered all studies with low risk of bias for the "performance bias" domain, few studies described the blinding of participants (n = 25, 5.8%), the blinding of healthcare professionals (n = 1, 0.2%), or both (n = 3, 0.7%). In the "attrition bias" domain, the median percentages of losses in intervention and control groups were 11.6% (interquartile range [IQR] 3.2–23.1) and 11.0% (IQR 2.5–21.8), respectively. Almost one-third of studies (28.7%) were classified as high risk of bias for the "attrition bias" domain since losses of participants/data were greater than 20%. Approximately one-fifth of RCTs (21.3%) did not present sufficient information to calculate losses. The reasons for participant losses were described in 272 studies. Description of the statistical treatment of data was classified as intention-to-treat analysis in 120 studies. Of these, only 44 (36.7%) reported in detail how the analyses were performed. Study protocols were published by only 49 RCTs

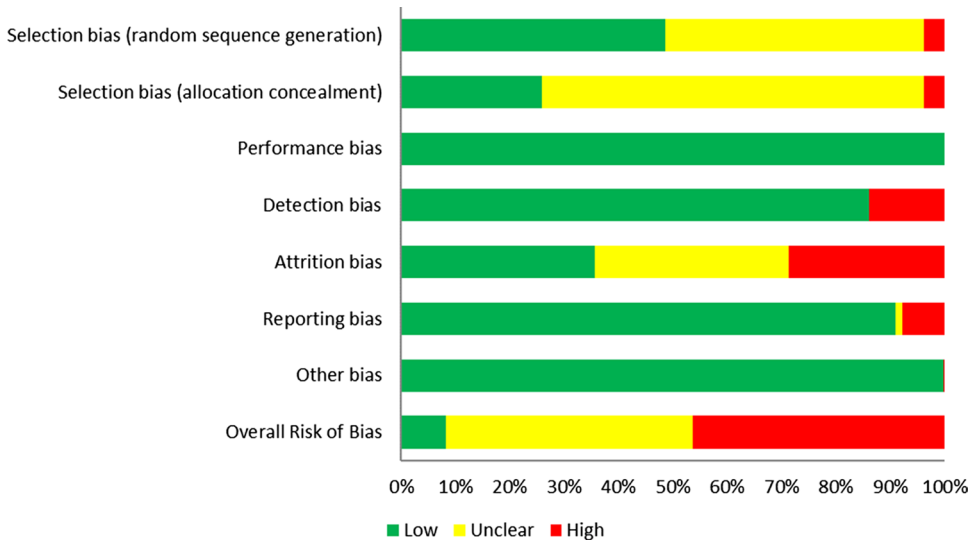

Fig. 1 Risk of bias graph: review authors' judgments of each risk of bias item. Values are presented as percentages across all included studies. The first seven lines represent the domains of the tool. Judgment of "high," "low", or "unclear" risk of bias was assigned for each RCT in each domain. The last line shows the overall risk of bias of included studies

Table 2 Original analyses of the risk of bias assessment across all included studies

	Low risk of bias n (%)	Unclear risk of bias n (%)	High risk of bias n (%)
Overall risk of bias	36 (8.3)	196 (45.4)	200 (46.3)
Selection bias	210 (48.6)	205 (47.5)	17 (3.9)
Allocation bias	112 (25.9)	303 (70.1)	17 (3.9)
Performance bias	432 (100.0)	0 (0.0)	0 (0.0)
Detection bias	372 (86.1)	0 (0.0)	60 (13.9)
Attrition bias	154 (35.6)	154 (35.6)	124 (28.7)
Reporting bias	393 (91.0)	5 (1.2)	34 (7.9)
Other bias	431 (99.8)	0 (0.0)	1 (0.2)

(11.3%); of these, 20 did not show agreement between the outcomes proposed in the protocol and article results. For "other bias", only one study was evaluated with high risk, because the trial was interrupted before the follow-up time without justification.

The sensitivity analyses revealed changes in the overall risk of bias for seven of the eight new elements of judgment criteria of the RoB tool when compared to the original analyses (see Table 3). For two of these new judgments (elements A04 and A05), the difference originated from an increase in studies classified with unclear risk of bias, while for five new judgments (elements P01, P02, A01, A02, and D01) there was an increase in the percentage of studies with overall high risk of bias.

The main differences in the "performance bias" domain were based on the existence of 404 (93.5%) not blinded studies with 124 presenting self-reported outcomes (P01). In 95 studies (22.0%), the pharmacist acted in both the intervention and control groups (P02). Differences in the "attrition bias" domain were primarily due to 204 studies (47.2%) with losses of participants or data of 10% or higher instead of 119 studies (27.5%) originally considered with losses of 20% or higher (A01). In 17 studies (3.9%), there were differences in losses between groups of 10% or higher (A02). Finally, in the "detection bias" domain, 195 studies (45.1%) reported subjective outcomes, of which 111 used a validated instrument (D01). Despite the statistically significant increase of risk classificartion, the effect size of these changes was negligible in all the sensitivity analyses except for the decrease of the cut-off to consider attitrion (A01) as high risk of bias (Table 3).

Discussion

With the evaluation of more than 400 RCTs, our study demonstrated that the RoB is a useful and sensitive tool for the assessment of RCTs of pharmacists' interventions. Testing the RoB in specific areas and intervention types is necessary because of the previously reported low reliability among users, which may impact the methodological quality assessment of primary studies included in systematic reviews [3, 4, 18].

Several tools to assess methodological quality of RCTs exist [4, 7, 31]. However, results obtained with different tools may not be equivalent. Armijo-Olivo et al. reported substantial differences in the results of the RoB compared with the Physiotherapy Evidence Database (PEDro) score [32], which may impact the evidence-based decision-making process. Studies have suggested that a consistent approach using the RoB may be preferable in the evaluation of RCTs, because summary quality scores such as PEDro may introduce bias by diluting the effect of items that are important for assessment of the risk of bias in a specific area [33, 34].

Despite the availability of consolidated guides for the conduction of RCTs in the literature, bias is closely linked

Table 3 Comparison between the overall risk of bias in the original analysis versus the sensitivity analyses

Overall risk of bias	Low risk of bias n (%)	Unclear risk of bias n (%)	High risk of bias n (%)	Cliff's delta* (95% CI)	*P* value**
Original analysis	36 (9.0)	196 (44.7)	200 (46.3)		
Sensitivity analyses—performance bias domain					
P01: No participant blinding in self-report outcomes	33 (7.6)	146 (33.6)	253 (58.6)	0.116 (0.047–0.184)	< 0.001
P02: Contamination of the blinding in the intervention/control groups	28 (6.5)	156 (36.1)	248 (57.4)	0.111 (0.043–0.179)	< 0.001
Sensitivity analyses—attrition bias domain					
A01: Reduction of the cut-off for loss of participants/data from 20% to 10%	24 (5.6)	87 (20.1)	321 (74.3)	0.272 (0.206–0.335)	< 0.001
A02: Differences of 10% or more of losses between intervention and control groups	35 (8.1)	190 (44.0)	207 (47.9)	0.017 (−0.053–0.085)	0.030
A03: Difference of 20% or more of losses between intervention and control groups	36 (9.0)	196 (44.7)	200 (46.3)	0 (−0.069–0.069)	–
A04: Lack of reasons for losses in intervention/control group	31 (7.2)	201 (46.5)	200 (46.3)	−0.006 (−0.075–0.063)	0.025
A05: Lack of description of the statistical method by intention-to-treat to treat the lost data	21 (4.9)	211 (48.8)	200 (46.3)	−0.019 (−0.087–0.050)	< 0.001
Sensitivity analyses—detection bias domain					
D01: Lack of validated instrument for subjective outcomes	35 (8.1)	150 (34.7)	247 (52.7)	0.101 (0.032–0.169)	< 0.001

*Magnitude of the changes's thresholds (original versus sensitivity analyses): $|d| < 0.147$ "negligible", $|d| < 0.33$ "small", $|d| < 0.474$ "medium", otherwise "large". Values with 95% confidence interval (CI)

**McNemar-Bowker test

to flaws in the design, analysis, and report of primary studies [35]. In our study, the majority of the RCTs were of unclear or high risk of bias, with notable methodological flaws in the selection bias domain (random sequence generation and allocation concealment). These components are frequently missing in clinical trials [36], with less than 40% of trials reporting the randomization process, and only 10% performing proper allocation [37, 38]. Poor randomization methods cause exaggerated treatment effects, are open to subversion by researchers or clinicians, and have a knock-on effect in systematic reviews [39, 40]. For example, one of the included studies reported as "open randomized controlled trial" described the randomization process as "recruited patients were randomised sequentially by day of recruitment into a control and two treatment groups"[article #14], which does not look like a probabilistic randomization. In our study, decreasing the cut-off from 20 to 10% for losses of participants/data ("attrition bias" domain) resulted in the greater increase of studies classified as high risk of bias among the eight sensitivity analyses. In contrast with clinical trials of pharmacological interventions with low attrition rates [41–43], clinical pharmacy interventions are more likely to lose participants [37, 38]. For example, one of the included studies [article #432] reported having 84% patients lost in intervention group during follow-up (starting with 706, ending with 116). Although this study provided a brief comment about the reasons for this attrition, the statistical analysis was 'per protocol' and not by 'intention to tract', as desirable. Describing the reasons for losses of participants and data in a RCT and its limitations should be required by editors and peer-reviews for publication [44, 45]. Our study also demonstrated that blinding is a concern in the performance domain of risk of bias for pharmacist intervention studies. With approximately 94% of non-blinded studies strictly using the original RoB criteria for judgment, this would result in the classification of almost all pharmacist intervention studies as high risk of bias. Blinding is especially important in the assessment of non-objective outcomes such as self-reported outcomes. In pharmacists' interventions, blinding participants or personnel was shown to be more difficult than blinding the outcome evaluator [37, 46]. An alternative to reduce the risk of bias in non-blinded studies using non-objective outcomes could be the use of validated instruments [5, 28]. Unfortunately, only 57% of pharmacist intervention RCTs uses validated instruments to assess non-objective outcomes.

Our sensitivity analyses, with modifications of the RoB criteria for judgment, demonstrated the need for more explicit criteria to evaluate the methodological quality of trials of complex interventions such as pharmacists' interventions. Depending on the interpretation of the criteria for judgment, risk of bias was changed for seven of the eight scenarios. Recent studies have recommended further testing of psychometric properties (e.g., validity, reliability, and responsiveness) of the RoB in a broad range of health areas [32, 47]. The development of a guidance to assess the risk of bias has also been suggested, as well as standardized training for researchers to improve agreement rates [17, 18]. Our results reinforced the usability of the RoB tool for complex interventions but highlight the need for the development of specific criteria judgment's guidance. However, the use of these adjustments on the criteria for judgement may also be controversial. In inter-area comparisons, these adjustments may underestimate the potential risk of bias of clinical pharmacy interventions. On the other hand, classifying as high risk of bias all the studies because of the impossibility of blinding the allocation to the intervention provider (i.e., the pharmacist) would result in a massive RoB floor-effect. The use of RoB tool by clinical pharmacy RCT researchers could also have a pedagogical effect helping to reduce avoidable flaws (e.g., prevent high attrition rates, select objective outcome measures) and, thus increase the overall quality of clinical pharmacy studies. Version 2.0 of the RoB tool [21], which is under development, may be a promising alternative to reduce the subjectivity of some of the domains. Usability and sensitivity studies of the RoB may continue to improve RoB 2.0 during its development.

This study has some limitations. We aimed to evaluate only the usability and sensitivity of the RoB tool adapted to pharmacist interventions. Yet, other parameters of the validity, reliability, and responsiveness of this tool have been extensively evaluated in the literature. To properly apply the RoB tool it is crucial to identify the main outcome, which is not possible when authors do not declare it in the RCT. To minimize the effect of this limitation, we used objective outcome selection criteria. Criticism exists about the use of standardised regression coefficients like Cliff's delta because this statistics removes the unit of measurement from its value, which may induce bias in epidemiological studies [48]. However, Cliff's delta is probably the most robust effect size measure for ordinal conditions, like the RoB results [30].

Conclusion

The RoB tool was useful in the evaluation of the risk of bias of RCTs of pharmacists' interventions if adapted criteria for judgment are used. Ignoring these adjustments produces a floor-effect with the majority of studies classified as high risk of bias.

Funding This work was supported by Brazilian National Council of Technological and Scientific Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES). The funding sources had no role in the study design, data collection, data analyses, data interpretation, or writing of the report. The corresponding

author had full access to all of the data in the study and was responsible for making the final decision to submit the manuscript for publication.

Conflicts of interest All authors declare that they have no conflict of interest.

References

1. Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol. 2013;13:116.
2. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017;390:415–23.
3. 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:b4012.
4. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ: Rockville, 2008.
5. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ. 2001;323:42–6.
6. Sylvester RJ, Canfield SE, Lam TB, Marconi L, MacLennan S, Yuan Y, et al. Conflict of evidence: resolving discrepancies when findings from randomized controlled trials and meta-analyses disagree. Eur Urol. 2017;71:811–9.
7. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015;8:2–10.
8. 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.
9. Mansournia MA, Higgins JP, Sterne JA, Hernan MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiology. 2017;28:54–9.
10. Yamato TP, Maher C, Koes B, Moseley A. The PEDro scale had acceptably high convergent validity, construct validity and inter-rater reliability in evaluating methodological quality of pharmaceutical trials. J Clin Epidemiol. 2017;86:176–81.
11. Hempel S, Miles JN, Booth MJ, Wang Z, Morton SC, Shekelle PG. Risk of bias: a simulation study of power to detect study-level moderator effects in meta-analysis. Syst Rev. 2013;2:107.
12. Saltaji H, Ospina MB, Armijo-Olivo S, Agarwal S, Cummings GG, Amin M, et al. Evaluation of risk of bias assessment of trials in systematic reviews of oral health interventions, 1991–2014: a methodology study. J Am Dent Assoc. 2016;147:720–8.
13. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. BMJ. 2013;346:f1798.
14. 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;18:12–8.
15. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther. 2008;88:156–75.
16. Higgins JPT, Green S. In: Cochrane Handbook for systematic reviews of interventions version 5.1.0. http://handbook-5-1.cochrane.org/. Accessed 09 Jan 2019.
17. 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;6:44.
18. 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;66:973–81.
19. Manchikanti L, Hirsch JA, Cohen SP, Heavner JE, Falco FJ, Diwan S, et al. Assessment of methodologic quality of randomized trials of interventional techniques: development of an interventional pain management specific instrument. Pain Physician. 2014;17:E263–90.
20. 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.
21. Higgins J, Sterne J, Savović J, Page M, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. Cochrane Database Syst Rev. 2016;10:30–1.
22. Vo TH, Charpiat B, Catoire C, Juste M, Roubille R, Rose FX, et al. Tools for assessing potential significance of pharmacist interventions: a systematic review. Drug Saf. 2016;39:131–46.
23. Tsuyuki RT. Designing pharmacy practice research trials. Can J Hosp Pharm. 2014;67:226–9.
24. Correr CJ, Melchiors AC, de Souza TT, Rotta I, Salgado TM, Fernandez-Llimos F. A tool to characterize the components of pharmacist interventions in clinical pharmacy services: the DEPICT project. Ann Pharmacother. 2013;47:946–52.
25. Rotta I, Salgado TM, Felix DC, Souza TT, Correr CJ, Fernandez-Llimos F. Ensuring consistent reporting of clinical pharmacy services to enhance reproducibility in practice: an improved version of DEPICT. J Eval Clin Pract. 2015;21:584–90.
26. Rotta I, Souza TT, Salgado TM, Correr CJ, Fernandez-Llimos F. Characterization of published randomized controlled trials assessing clinical pharmacy services around the world. Res Soc Adm Pharm. 2017;13:201–8.
27. da Costa BR, Resta NM, Beckett B, Israel-Stahre N, Diaz A, Johnston BC, et al. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a study protocol. Syst Rev. 2014;13(3):144.
28. 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.
29. Hoffman JI. The incorrect use of Chi square analysis for paired data. Clin Exp Immunol. 1976;24:227–9.
30. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. Psychol Bull. 1993;114:494–509.
31. Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Comparison of tools for assessing the methodological quality of primary and secondary studies in health technology assessment reports in Germany. GMS Health Technol Assess. 2010;6:Doc07.
32. Armijo-Olivo S, da Costa BR, Cummings GG, Ha C, Fuentes J, Saltaji H, et al. PEDro or Cochrane to assess the quality of clinical trials? A meta-epidemiological study. PLoS ONE. 2015;10:e0132634.

33. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054–60.
34. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ. 2004;328:22–4.
35. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;2010(340):c332.
36. Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol. 2016;77:24–37.
37. Glynn LG, Murphy AW, Smith SM, Schroeder K, Fahey T. Self-monitoring and other non-pharmacological interventions to improve the management of hypertension in primary care: a systematic review. Br J Gen Pract. 2010;60:e476–88.
38. Bos JM, van den Bemt PM, de Smet PA, Kramers C. The effect of prescriber education on medication-related patient harm in the hospital: a systematic review. Br J Clin Pharmacol. 2017;83:953–61.
39. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: Are we getting better? BMJ. 2016;355:i5663.
40. Clark T, Davies H, Mansmann U. Five questions that need answering when considering the design of clinical trials. Trials. 2014;15:286.
41. Groenwold RH, Moons KG, Vandenbroucke JP. Randomized trials with missing outcome data: how to analyze and what to report. CMAJ. 2014;186:1153–7.
42. Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? Qual Life Res. 2016;25:2977–83.
43. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials. 2004;1:368–76.
44. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391.
45. Hewitt CE, Kumaravel B, Dumville JC, Torgerson DJ, Trial attrition study group. Assessing the impact of attrition in randomized controlled trials. J Clin Epidemiol. 2010;63:1264–70.
46. Jarvinen TL, Sihvonen R, Bhandari M, Sprague S, Malmivaara A, Paavola M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol. 2014;67:769–72.
47. 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:e96920.
48. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. Am J Epidemiol. 1986;123:203–8.