Journal of Clinical Epidemiology 152 (2022) 47–55

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Adherence of systematic reviews to Cochrane RoB2 guidance was frequently poor: a meta epidemiological study

Silvia Minozzi[a], Marien Gonzalez-Lorenzo[b,*], Michela Cinquini[b], Daniela Berardinelli[c], Celeste Cagnazzo[d,e], Stefano Ciardullo[f,g], Paola De Nardi[h], Mariarosaria Gammone[i], Paolo Iovino[j,k], Alex Lando[l], Marco Rissone[m], Giovanni Simeone[n], Marta Stracuzzi[o], Giovanna Venezia[p], Lorenzo Moja[q], Giorgio Costantino[r], The University of Milan Post Graduate Course on Systematic Review Working Group

[a]*Department of Epidemiology, Lazio Regional Health Service, Rome, Italy*
[b]*Laboratorio di Metodologia delle revisioni sistematiche e produzione di Linee Guida, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy*
[c]*Department of Clinical and Biological Sciences, San Luigi Hospital, University of Turin, Turin, Italy*
[d]*Department of Sciences of Public Health and Pediatrics, University of Turin, Turin, Italy*
[e]*Division of Paediatric Onco-Haematology, Stem Cell Transplantation and Cellular Therapy, Città della Salute e della Scienza Hospital, Turin, Italy*
[f]*Department of Medicine and Rehabilitation, Policlinico di Monza, Monza, Italy*
[g]*Department of Medicine and Surgery, University of Milano Bicocca, Milan, Italy*
[h]*Gastrointestinal Surgery, IRCCS San Raffaele Scientific Institute, Milan, Italy*
[i]*S.I.T.R.A, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy*
[j]*Department of Biomedicine and Prevention University of Rome Tor Vergata, Rome, Italy*
[k]*School of Nursing, Midwifery and Paramedicine Faculty of Health Science, Australian Catholic University, Melbourne, Australia*
[l]*Laboratory of Rehabilitation Technologies, IRCCS San Camillo Hospital, Venice, Italy*
[m]*Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy*
[n]*Local Health Unit of Brindisi, Pediatric Department, Brindisi, Italy*
[o]*Pediatric Infectious Disease Unit, Department of Pediatrics, Luigi Sacco Hospital, University of Milan, Milan, Italy*
[p]*Gastroenterology Unit, S.Croce e Carle Hospital, Cuneo, Italy*
[q]*Department of Biomedical Sciences for Health, University of Milan, Milan, Italy*
[r]*Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, University of Milan, Milan, Italy*

Accepted 6 September 2022; Published online 23 September 2022

Abstract

Objectives: To assess whether the use of the revised Cochrane risk of bias tool for randomized trials (RoB2) in systematic reviews (SRs) adheres to RoB2 guidance.

Methods: We searched MEDLINE, Embase, Cochrane Library from 2019 to May 2021 to identify SRs using RoB2. We analyzed methods and results sections to see whether risk of bias was assessed at outcome measure level and applied to primary outcomes of the SR as per RoB2 guidance. The relation between SR characteristics and adequacy of RoB2 use was examined by logistic regression analysis.

Results: Two hundred-eight SRs were included. We could assess adherence in 137 SRs as 12 declared using RoB2 but actually used RoB1 and 59 did not report the number of primary outcomes. The tool usage was adherent in 69.3% SRs. Considering SRs with multiple primary outcomes, adherence dropped to 28.8%. We found a positive association between RoB2 guidance adherence and the methodological quality of the reviews assessed by AMSTAR2 (p-for-trend 0.007). Multivariable regression analysis suggested journal impact factor [first quartile vs. other quartiles] was associated with RoB2 adherence (OR 0.34; 95% CI: 0.16-0.72).

* Corresponding author. Laboratorio di Metodologia delle revisioni sistematiche e produzione di Linee Guida, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy. Tel.: +390239014635; fax: +390233200231.

E-mail address: mariengonzalezlorenzo@gmail.com (M. Gonzalez-Lorenzo).

https://doi.org/10.1016/j.jclinepi.2022.09.003

Conclusions: Many SRs did not adhere to RoB2 guidance as they applied the tool at the study level rather than at the outcome measure level. Lack of adherence was more likely among low and very low quality reviews.

Keywords: Systematic reviews; Randomized controlled trials; Risk of bias; Adherence; Meta-epidemiologic methods; Methodological quality

1. Introduction

Systematic Reviews (SRs) are essential for the development of guidelines and for healthcare decision-making given that they provide an overview of the available evidence on a specific topic [1]. Assessing risk of bias of included primary studies is essential during the conduct of a SR since it allows the identification of potential flaws that could limit the validity of the results of the review [1−3].

During the last decade, the most used tool to assess the risk of bias of randomized controlled trials (RCTs) in both Cochrane and non-Cochrane SRs was the Cochrane Risk of Bias tool (RoB1) [4]. This instrument covers seven domains (random sequence generation and allocation concealment for selection bias, blinding of participants and providers for performance bias, blinding of outcome assessor for detection bias, incomplete outcome data for attrition bias, selective outcome reporting for reporting bias, and other sources of bias) [4]. Several methodological studies have been published evaluating the appropriate use of this tool [5−13].

A revised version of this risk of bias tool (RoB2) was released in 2019 [14], with the purpose of overcoming some limitations afflicting the previous version, such as the inconsistent use of the tool (authors added or removed domains), an excessive use of ‘‘unclear judgment’’ and the lack of an overall judgment domain. The new tool incorporates advances in the theoretical understanding of bias, such as the need to assess separately selective reporting of specific measures of a given outcome and selective non-reporting of outcomes, and a more careful evaluation of risk of bias for studies that are not blinded. The new RoB2 supports the evaluation of potential biases arising from inadequate randomization process, deviations from intended interventions, missing data, measurement of the outcomes, selection of the reported results and overall bias. For each domain RoB2 guides the reviewer in formulating a judgement on the risk of bias, which can be expressed as ‘‘low’’, ‘‘high’’ or ‘‘some concern’’. The most relevant difference between RoB2 and the original version of the tool (RoB1) is the exclusion of any judgment at the overall study level. In fact, RoB2 requires that the risk of bias is always assessed at the individual outcome measure level. The rationale being that different outcome measures within the same study often have different risk of bias. RoB2 guidance recommends that all the outcomes included in the Summary of Findings table should be fully evaluated at the outcome measure level [15].

The objective of this descriptive study is to evaluate whether SRs published after the launch of RoB2 that used this tool complied with RoB2 guidance.

2. Methods

We performed this SR according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [16] and guidelines for reporting meta-epidemiological research [17].

2.1. Search strategy and selection criteria

We searched MEDLINE (via PubMed), Embase and Cochrane Library from 1st January 2019 to 31st May 2021. We chose this relatively recent time frame because the final version of RoB2 was published in 2019 [14]. A combination of free text and Mesh terms were used. No language restrictions were applied. The full search strategy for each database is presented in the Supplementary Material.

We included SRs, irrespective of their qualitative or quantitative nature, who aimed to evaluate the effect of any intervention and that included at least one randomized controlled trial. As a consensus definition of SR does not exist [18], we defined a review as systematic if the review objective, the inclusion-exclusion criteria and the search strategy were clearly stated in the methods and if the number of studies finally included was reported in the results section. We restricted our focus and selection to those SRs who declared in their methods to assess risk of bias by using the RoB2 tool. Indeed, our sample includes only SRs that were published after the RoB2 and its guidance became available in the public domain.

After combining the search results and removing duplicates, two pairs of reviewers (MR, FL, SV, EZ) independently screened titles and abstracts of all records to identify potentially relevant SRs. We used EndNote X9 software [19], to manage citations and Rayyan software [20] for the title/abstract screening process.

We retrieved the full-text versions of all potentially relevant SRs for definitive assessment of eligibility. Two reviewers (PDN, AL) independently assessed full texts of potentially relevant SRs. In case of disagreements a third reviewer (MGL) was involved.

What is new?

Key findings

- We assessed whether the evaluation of risk of bias was performed at outcome measure level and applied to the primary outcomes of the SR as per RoB2 guidance l in 208 SRs. Seventy-one (34%) were not evaluable (use of RoB1 despite declaring use of RoB2, number of primary outcomes not reported).
- Out of the 137 (66%) SRs evaluable, the use of the tool was judged adherent to the guidance in 95 (69.3%). When considering only those with more than one primary outcome, the proportion dropped to 28.8% (17/59).
- We found a positive association between the correct use of the tool and the overall methodological quality of the reviews assessed by AMSTAR 2 and with publication on journal in the first quartile impact factor, while no association was found with area of medicine and geographical region according to the first author's affiliation.

What this adds to what is known?

- Many systematic reviews published in the last 2 years incorrectly applied the revised Cochrane risk of bias tool (RoB2) as they assessed the risk of bias at the study level, rather than at the outcome measure level.

What is the implication, what should change now?

- Interventions are needed to improve the quality of published SRs and to promote the correct application of RoB2 tool.

2.2. Data extraction and adherence to RoB2 guidance

For each SR we sought general information (title, publication date, scientific journal) where the SR was published and its 2020 Impact Factor (IF) on journal citation report 2021 [21], first author's affiliation as a proxy for country of origin (categorized as Europe, Asia, Africa, North America, South America, Australia), area of medicine covered (categorized as Alternative medicine/acupuncture, Anesthesiology/Emergency medicine, Dentistry, Medical treatments/chronic disease, Medical treatments/infectious disease, Psychiatry/Psychology, Rehabilitation/physiotherapy/speech therapy, Surgery), intervention and comparison, number of RCTs included and assessment of certainty of evidence with the GRADE approach [2]. We then evaluated what authors reported in the methods section of their review related to risk of bias assessment. In particular, we tried to ascertain in what way they used the RoB2 and if its use adhered to RoB2 guidance. Adherence was satisfied if authors applied RoB2 to at least stated primary outcomes. Pragmatically, in each review, the number of RoB2 assessments should have been equal to or greater than the reported number of primary outcomes. Adherence was considered inadequate if RoB2 was applied at the study level (e.g., the number of Rob2 assessment was one per study and the number of stated primary outcomes was greater than one). We did not assess adherence for those SRs in which we could not identify the exact number of primary outcomes. We purposely did not make our operationalized definition of adherence too stringent. Two authors (SC, MS) independently evaluated adherence; discrepancies were resolved by consensus, in consultation with a senior epidemiologist (SM).

We assessed the methodological quality of each SR using the *A MeaSurement Tool to Assess systematic Reviews* (AMSTAR) 2 checklist [22]. Each review was judged as: (i) ''high'' quality (i.e., none or one non-critical weakness); (ii) ''moderate'' (i.e., more than one non-critical weakness); (iii) ''low'' (i.e., one critical flaw with or without non-critical weaknesses), or (iv) ''critically low'' (i.e., more than one critical flaw with or without non-critical weaknesses). We considered the following six items in order to assess critical weaknesses: ''*Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?", "Did the review authors use a comprehensive literature search strategy?", "Did the review authors provide a list of excluded studies and justify the exclusions?", "Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?", "If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?", "Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?*'' Then we followed the algorithm suggested in the original guidance [22] to derive the overall rating. For each review all questions were addressed leading to a detailed judgment of quality. The results of the AMSTAR 2 evaluation for each review are reported in the online Supplementary Material.

Extraction and all evaluations were done in duplicate by three pair of independent reviewers (MS, MRG, GS, MR, AL, PDN). Disagreements were resolved by consensus or by involvement of a third reviewer (SM), who was asked to settle controversial assessments. All database entries were checked by a third senior author (GC, SM).

2.3. Analysis

The characteristics of the reviews were summarized with descriptive statistics of means and standard deviations for

continuous variables, and frequencies and percentages for categorical variables. A Chi-squared test for trend was performed to evaluate association between methodological quality of the reviews, assessed with AMSTAR 2 and RoB2 use adherent to guidance. The relation between the reviews' characteristics and adherence to RoB2 guidance was examined by univariate logistic regression model analysis. The dependent variable in this logistic regression analyses was dichotomized in RoB2 use adherent to guidance vs. and not adherent (''not adherent'' as the reference). As independent variables we considered:

a. country of origin, based the first author's affiliation (Europe as the reference);
b. area of medicine addressed by the review (Alternative Medicine, acupuncture as the reference);
c. quartiles of IF. In order to obtain comparable results, IFs were categorized into quartiles specific for each SR discipline, according to the classification of the Journal of citation report (first quartile as the reference) [2];
d. methodological quality of the reviews according to AMSTAR 2 (high quality as the reference).

Using a cut-off *P*-value of 0.05, we included all the variables in a multivariable model to explore potential predictors of the use of RoB2 adhering to formal guidance. Results are reported as odds ratios (ORs) with 95% CIs. Analyses were performed using SAS (Statistical Analysis System, SAS Institute Inc., Cary, NC, USA, Version 8.20) software. All tests were two-sided and $P < 0.05$ was used to determine statistical significance.

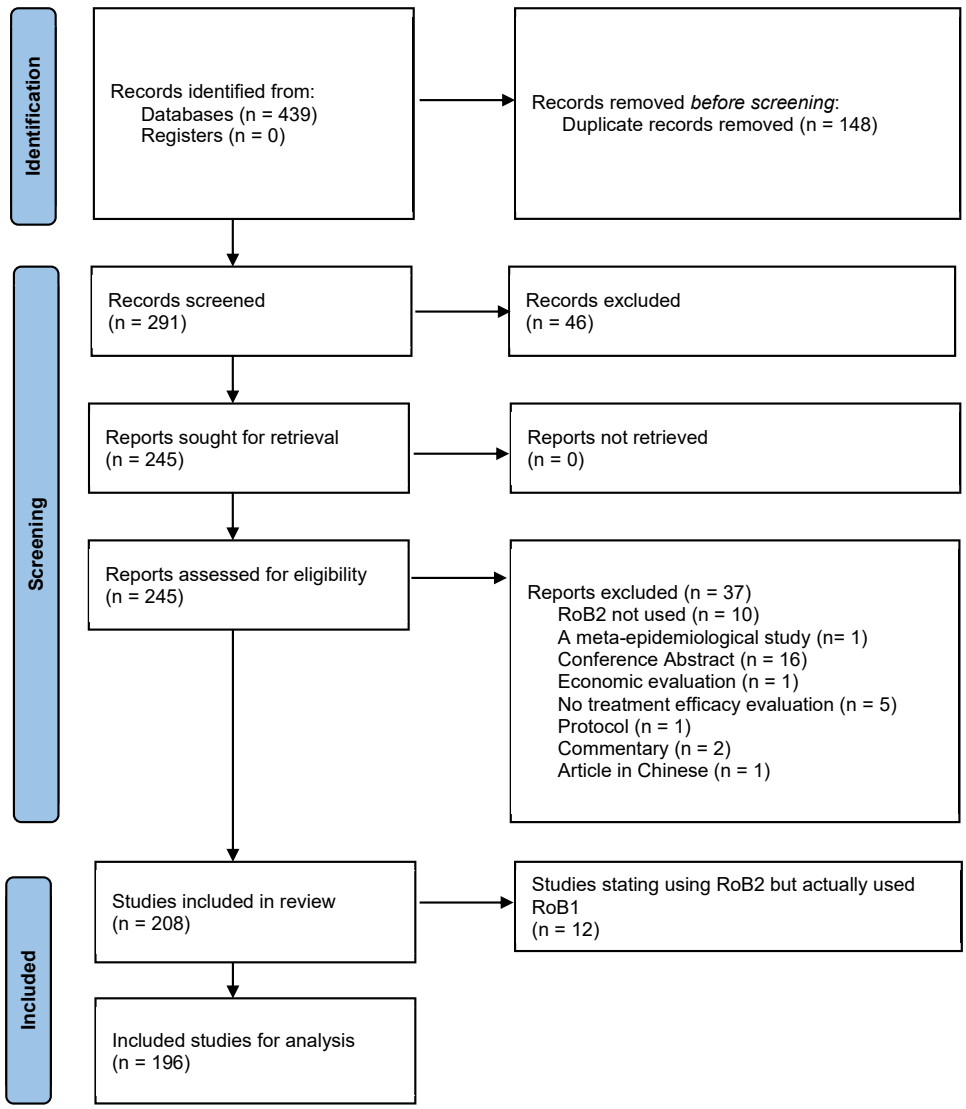

Fig. 1. Study flow from literature search.

The study involved no human participants and required no ethical approval.

3. Results

3.1. Characteristics of the studies

The bibliographic search retrieved 291 records after duplicates removal. After screening the titles and abstracts, we reviewed 245 full texts for further assessment, of which, 37 were excluded. The detailed reasons for exclusion are shown in the Supplementary Material (Characteristics of excluded studies table). Finally, 208 SRs were included, eleven (5.3%) of which were Cochrane reviews. The flow diagram of the selection process is shown in Figure 1 (References of included studies are reported in the Supplementary Material).

The most frequent investigated areas of medicine were medical treatments for chronic disease (n = 46, 22.1%), dentistry (n = 42, 20.2%) and psychiatry/psychology (n = 34, 16.3%), whereas 30 reviews (14.4%) addressed rehabilitation interventions. The reviews included a mean of 16.91 (SD = 32.43, range 2-424) RCTs. Mean number of outcomes investigated was 4.63 (SD = 3.26, range 1-23). Fifty-nine reviews (28.4%) did not specify the number of the primary outcomes. Among the remaining, the mean number of primary outcomes was 2.11 (SD 1.93, range 1-16). Based on the first authors' affiliations, the reviews mostly involved European (n = 89, 42.8%) and Asian countries (n = 60, 28.9%). Ninety-eight reviews (47.1%) were published in the first IF quartile (Q1), 64 (30.8%) in the second quartile (Q2), 38 (18.3%) in the third quartile (Q3) and 7 (3.4%) in the fourth quartile (Q4). Only one review (0.5%) was published in journals without IF. Ninety-two reviews (44.2%) assessed the certainty of evidence using the GRADE approach [2]. According to AMSTAR 2 checklist, a critically low, low, moderate and high-quality judgment was attributed to 122 (58.6%), 52 (25%), 20 (9.6%) and 14 (6.7%), respectively (Supplementary Material; Characteristic of included studies).

3.2. Adherence to RoB2 guidance

In 12 (5.8%) out of the 208 included studies, the authors declared using RoB2 in the methods section but actually used RoB1 domains. A total of 59/196 (30.1%) did not report the number of primary outcomes, thereby preventing us from evaluating the adherence to RoB2 guidance. The final number of reviews included to assess adherence was 137 (66%). The use of the tool was judged to adhere to RoB2 guidance in 95/137 reviews (69.3%) (Table 1). Of these, 78 SRs had only one primary outcome (Table 1). When considering those SRs with more than one primary outcome, the proportion of SRs with an appropriate use of the tool dropped to 28.8% (17/59). Seven (41.2%) of these 17 reviews, were Cochrane reviews.

Table 1. Proportion of reviews that applied the RoB2 tool adhering to guidance according to the number of primary outcomes evaluated and the number of RoB assessments performed

Criteria	*N.* of reviews
According to the number of primary outcomes	
1 primary outcome	78
Adherent, *n* (%)	78 (100)
Not adherent, *n* (%)	0 (0)
1 Rob, *n* (%)	78 (100)
>1 Rob, *n* (%)	0 (0)
>1 primary outcome	59
Adherent, *n* (%)	17 (28.8)
Not adherent, *n* (%)	42 (71.2)
1 Rob, *n* (%)	42 (71.2)
>1 Rob, *n* (%)	17 (28.8)
Number of primary outcomes not stated	59
Adherent, *n* (%)	NA
Not adherent, *n* (%)	NA
1 Rob, *n* (%)	59 (100)
>1 Rob, *n* (%)	0 (0)
According to the number of RoB assessments	
1 RoB assessment	179
Adherent, *n* (%)	78 (43.6)
Not adherent, *n* (%)	42 (23.5)
Can't tell, *n* (%)	59 (32.9)
1 Primary outcome, *n* (%)	78 (43.6)
>1 Primary outcome	42 (23.5)
Number of primary outcomes not stated, *n* (%)	59 (32.9)
>1 RoB assessments	17
Adherent, *n* (%)	17 (100)
Not adherent, *n* (%)	0 (0)
Can't tell, *n* (%)	0 (0)
1 Primary outcome, *n* (%)	0 (0)
>1 Primary outcome	17 (100)
Number of primary outcomes not stated, *n* (%)	0 (0)

Abbreviations: NA, not assessable; RoB, Risk of Bias.

Out of the 196 reviews, 179 (91.3%) reported one RoB2 table. Only 78 of these (43.6%) adhered to RoB2 guidance appropriately, as they had only one primary outcome (Table 1).

A higher proportion of reviews adhering to RoB2 guidance was found in high quality review (71.4%) compared to moderate (68%), low (49%) or critically low (42%) (P-value for trend 0.007) (Figure 2). No differences in adherence were identified when the analyses were stratified by country and journal impact factor quartile, whereas a higher frequency of scarce adherence was found within reviews concerning the fields of alternative medicine and emergency medicine/anesthesiology (Table 2).

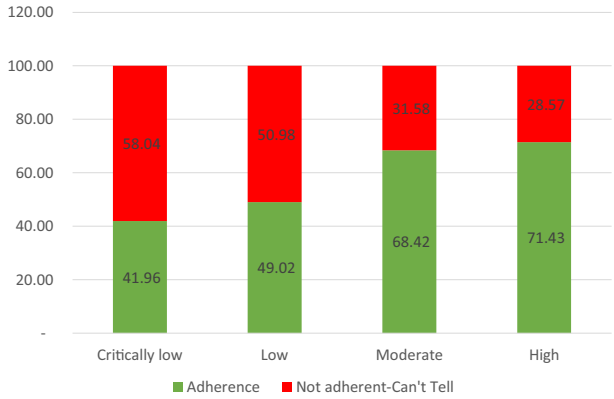

Fig. 2. Adherence to RoB2 tool guidance according to methodological quality (AMSTAR 2).

3.3. Logistic regression

The univariate models included all 196 SRs that applied Rob2. For the analysis, we combined reviews with use of RoB2 adherent to guidance and reviews for which adherence could not be evaluated (''can't tell group'') because no specification was provided about the number of primary outcomes in the reviews and only one RoB2 evaluation was provided.

Taken AMSTAR 2 high methodological quality as the reference, RoB2 evaluation was less likely to adhere to methodological guidance if the SR quality was judged as critically low (OR 0.29, 95% CI 0.09–0.98) (Table 3). The only quartile of IF associated with a significant higher probability of appropriate RoB2 use adherent to guidance was the first (OR 0.35, 95% CI 0.18-0.68). No relation was found in RoB2 use adherence to guidance for geographical region or area of medicine. Results from the multivariable regression analyses reinforced a potential relation between adherence to RoB2 guidance and the first quartile of IF (OR 0.34, 95% CI 0.16-0.72).

4. Discussion

The aim of this systematic review was to evaluate if the use of the RoB2 tool adheres to use guidance across

Table 2. Adherence to RoB2 tool guidance according to reviews characteristics of 196 reviews included in the analyses

Reviews characteristics	All # (column%)		Adherent # (row%)		Not adherent + Can't tell # (row%)	
Amstar 2						
Critically low	112	57.14%	47	41.96%	65	58.04%
Low	51	26.02%	25	49.02%	26	50.98%
Moderate	19	9.69%	13	68.42%	6	31.58%
High	14	7.14%	10	71.43%	4	28.57%
Impact factor						
Q1	91	46.67%	55	60.44%	36	39.56%
Q2	59	30.26%	20	33.90%	39	66.10%
Q3	38	19.49%	16	42.11%	22	57.89%
Q4	7	3.59%	3	42.86%	4	57.14%
Area of medicine						
Alternative Medicine, Acupuncture	10	5.10%	3	30.00%	7	70.00%
Dental Care	40	20.41%	19	47.50%	21	52.50%
Medical Area (chronic disease)	45	22.96%	22	48.89%	23	51.11%
Medical Area (infectious)	20	10.20%	11	55.00%	9	45.00%
Rehabilitation, Physiotherapy, Speech Therapy	25	12.76%	11	44.00%	14	56.00%
Surgery	15	7.65%	8	53.33%	7	46.67%
Psychiatry/Psychology	32	16.33%	18	56.25%	14	43.75%
Anaesthesiology, Emergency	9	4.59%	3	33.33%	6	66.67%
Country (First author's affiliation)						
Europe	81	41.33%	37	45.68%	44	54.32%
North America	21	10.71%	11	52.38%	10	47.62%
South America	26	13.27%	15	57.69%	11	42.31%
Africa	5	2.55%	3	60.00%	2	40.00%
Oceania	12	6.12%	6	50.00%	6	50.00%
Middle East	1	0.51%	0	0.00%	1	100.00%
South East Asia	17	8.67%	8	47.06%	9	52.94%
Far East	33	16.84%	15	45.45%	18	54.55%

Table 3. Results of the univariate logistic regression analysis

Independent variable	reference category	Effect	Odds ratio (95% CI) $n = 196$
AMSTAR 2 evaluation			
	High	Moderate	0.87 (0.19–3.92)
		Low	0.39 (0.11–1.39)
		Critically low	0.29 (0.09–0.98)
IF quartile			
	1	2	0.35 (0.18–0.68)
		3	0.51 (0.24–1.11)
		4	0.67 (0.16–2.86)
Geographical region			
	Europe	Africa	1.78 (0.28–11.25)
		Far east	0.99 (0.44–2.23)
		North America	1.31 (0.50–3.42)
		Oceania	1.19 (0.35–4.00)
		South America	1.67 (0.66–4.19)
		Middle east	0.00 (0.00–∞)
		South east Asia	1.07 (0.39–2.91)
Area of medicine			
	Alternative Medicine, Acupuncture	Dental Care	0.87 (0.39–1.93)
		Medical Area (chronic disease)	2.31 (0.67–7.94)
		Medical Area (infectious)	0.57 (0.30–1.07)
		Rehabilitation, Physiotherapy, Speech Therapy	3.68 (1.94–6.95)
		Surgery	0.96 (0.19–4.79)
		Psychiatry/Psychology	0.96 (0.19–4.79)
		Anaesthesiology, Emergency	0.96 (0.19–4.79)

recently published SRs. We found that most of the SRs included did not properly use the RoB2 tool. Out of 208 reviews included, the assessment of guidance adherence was only possible for 137 (66%) studies since the remaining did not specify which of the included outcomes were considered primary ($n = 59$) or the authors declared using the RoB2 tool but actually used RoB1 domains ($n = 12$). Out of the evaluable reviews, the percentage with an appropriate use of Rob2 was 69.3%; however, it should be noticed that most of these reviews had a single primary outcome (82%); this prevented us from discriminating whether the authors performed a single RoB assessment because they were applying it to the primary outcome, or whether it was applied at the study level (which would conceptually be inappropriate). Furthermore, all the reviews for which guidance adherence was not evaluable included more than one outcome in total and all of them performed a single RoB assessment, thereby raising suspicion that the tool was applied at the study level (i.e., inappropriately) in most of these studies. When we restricted the analysis to reviews with at least two primary outcomes where the approach taken by SR authors is less ambiguous, the proportion of reviews with an appropriate use of the tool decreased to 29%.

These findings may have several explanations. It is possible that authors adopting the RoB2 are still under the influence of the old quality assessment paradigm, which privileged for at least 2 decades scales and checklist operating at study level [23]. Another possible explanation is that the reviews authors simply adopted the new tool without careful reading of the tool's guidance and without understanding the relevant conceptual changes that occurred from the first to the second version of the instrument. Previous studies have shown that the RoB2 is generally difficult to apply and time consuming, even by experienced reviewers [24,25]. The association between scarce adherence to RoB2 guidance and low SR methodological quality assessed by AMSTAR suggests that poor operationalizations of SR methods is likely to afflict the whole study, with risk of bias assessment not being an exception. Similar results were found in a recent study assessing the association between the quality of SRs evaluated by AMSTAR 2 and the correct use of the tool to assess risk of bias of non-randomized studies (ROBINS-I) [26,27]. Finally, our results are in line with research showing some inertia by the peer review system to align to new guidance [28,29].

4.1. Study limitations

Our review has several limitations. We developed a review protocol where the most relevant methods for the conduct of the review were established in advance. However, it was not registered in PROSPERO. Our definition of adherence to RoB2 guidance was simple, limited to one aspect: use at outcome level vs. study level, and indiscriminative when applied to SRs with a single primary outcome. Furthermore, our approach hinges on the specification of primary outcomes in the reviews. To improve our evaluation, we should have contacted SR authors who did not specify the primary outcomes to ask how many of them were considered primary, to investigate their deep comprehension of RoB2 guidance and barriers related to its use. However, this was not possible for resource limitation. We did not further explore the level of comprehension of each domain and signaling questions. Finally, we could have missed some SRs that applied RoB2 tool but did not specify it in the abstract; however, we don't think that our search strategy introduced a selection bias which could compromise the validity of our results.

4.2. Concluding remarks

Our results suggest that many SRs' authors did not apply RoB2 at the outcome measures level. Its use might be influenced by old approaches in which the risk of bias was assessed at the study level instead of at the outcome level. This change in paradigm requires multiple interventions to support a better implementation of the new tool in the upcoming generation of SRs.

Acknowledgments

This project was developed during a post graduate academic course focused on SR methodology, which took place in 2021 at the University of Milan. The project involved both teachers and students, equally contributing to the finalization of this manuscript.

Appendix A

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.09.003.

References

[1] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (update February 2021). 2021. Available at https://training.cochrane.org/handbook/current. Accessed January 12, 2022.

[2] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94.

[3] Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hróbjartsson A, Savović J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One 2016;11:e0159267.

[4] Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

[5] Jørgensen L, Paludan-Müller AS, Laursen DR, Savović 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.

[6] Barcot O, Boric M, Poklepovic Pericic T, Cavar M, Dosenovic S, Vuka I, 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 2019;19:170.

[7] 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 2019;106:10−7.

[8] 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 2019;113:104−13.

[9] 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.

[10] Babic A, Tokalic R, Amílcar Silva Cunha J, Novak I, Suto J, Vidak M, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. BMC Med Res Methodol 2019;19:76.

[11] Babic A, Pijuk A, Brázdilová 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.

[12] Puljak L, Ramic I, Arriola Naharro C, Brezova J, Lin YC, Surdila AA, et al. Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews. J Clin Epidemiol 2020;123:114−9.

[13] Barcot O, Dosenovic S, Boric M, Pericic TP, Cavar M, Jelicic Kadic A, et al. Assessing risk of bias judgments for blinding of outcome assessors in Cochrane reviews. J Comp Eff Res 2020;9(8):585−93.

[14] Sterne JAC, Savovič J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:I4898.

[15] Higgins JPT, Savović J, Page MJ. Sterne JAC on behalf of the RoB2 Development Group. Revised Cochrane risk-of-bias tool for randomized trials (RoB2). 2019. Available at https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2. Accessed January 4, 2022.

[16] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

[17] Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med 2017;22(4):139−42.

[18] Krnic Martinic M, Pieper D, Glatt A, Puljak L. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. BMC Med Res Methodol 2019;19:203.

[19] https://endnote.com/. Accessed December 16, 2021.

[20] https://www.rayyan.ai/. Accessed December 16, 2021.

[21] https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/. Accessed March 21, 2022.

[22] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that

include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.

[23] Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. BMJ 2005;330:1053.

[24] Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. The revised Cochrane risk of bias tool for randomized trials (RoB2) showed low interrater reliability and challenges in its application. J Clin Epidemiol 2020;126:37−44.

[25] Minozzi S, Dwan K, Borrelli F, Filippini G. Reliability of the revised Cochrane risk-of-bias tool for randomised trials (RoB2) improved with the use of implementation instruction. J Clin Epidemiol 2021; 141:99−105.

[26] Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.

[27] Igelström E, Campbell M, Craig P, Katikireddi SV. Cochrane's risk of bias tool for non-randomized studies (ROBINS-I) is frequently misapplied: a methodological systematic review. J Clin Epidemiol 2021;140:22−32.

[28] Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. BMJ 2014;348:g3804.

[29] Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, Zhang S, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. J Multidiscip Healthc 2013;6:169−88.