





Methodologic quality and risk-of-bias in systematic reviews of healthcare interventions: a review of methods

Qualidade metodológica e risco de viés em revisões sistemáticas de intervenções em saúde: uma revisão de métodos

Cristine Miron Stefani¹ 
Liliana Vicente Melo de Lucas Rezende² 
Lia Rosana Honnef³ 
Luciana Butini Oliveira⁴ 

Nathan da Cruz Lopes⁵ 
Júlia Meller Dias de Oliveira⁶ 
Carla Massignan⁷ 
Graziela de Luca Canto⁸ 

¹Corresponding author. Universidade de Brasília (Brasília). Distrito Federal, Brazil. cmstefani@gmail.com

^{2,7}Universidade de Brasília (Brasília). Distrito Federal, Brazil. rezendeliliana@gmail.com, carmassignan@yahoo.com.br

^{3,5,6,8}Universidade Federal de Santa Catarina (Florianópolis). Santa Catarina, Brazil. lia.honnef@hotmail.com, nathan.lopes@grad.ufsc.br, julia_meller5@hotmail.com, delucacanto@gmail.com

⁴Instituto e Centro de pesquisas São Leopoldo Mandic (Campinas). São Paulo, Brazil. lubutini@uol.com.br

ABSTRACT | OBJECTIVE: To compare the characteristics of systematic reviews of healthcare interventions that assessed or did not methodologic quality/risk-of-bias of included studies. Additionally, to analyze methodologic features of those assessing the methodologic quality/risk-of-bias. **METHODS:** PubMed database was searched. From 25,571 systematic reviews retrieved, a random sample of 1,025 was screened. Frequencies were used to describe outcomes. Unadjusted and adjusted logistic regressions were performed to test the associations with methodologic quality/risk-of-bias results assessment. In a second analysis, systematic reviews that assessed methodologic quality/risk of bias were dichotomized according to the design of included studies (randomized clinical trials-only versus non-randomized studies of intervention or a combination of both). **RESULTS:** 303 systematic reviews were fully analyzed. Methodologic quality/risk of bias was assessed by 278 (92%). Methodologic quality/risk-of-bias assessment was associated with a higher number of databases searched (>4, P= 0.008), the presence of meta-analysis (P= 0.005), and the design of included studies (randomized clinical trials-only, P= 0.042). The chance of using a suitable tool and a tool designed for risk-of-bias assessment rather than methodologic quality was higher for randomized clinical trials-only systematic reviews (P< 0.05). The most used tool was Cochrane's RoB Tool without a clear studies' overall risk classification system. **CONCLUSION:** methodologic quality/risk-of-bias assessment was associated with included studies' design (randomized clinical trials-only), a meta-analysis of data, and the number of databases searched (>4). The most used tool was Cochrane's RoB Tool, with no clearly defined rating system. Methodologic quality/risk-of-bias assessment methods description, results, and impacts on meta-analysis, the certainty of evidence, and systematic reviews' results are still to be consistently addressed.

KEYWORDS: Systematic Review. Healthcare. Methodologic Quality. Risk of Bias. Critical Appraisal.

RESUMO | OBJETIVO: Comparar as características de revisões sistemáticas de intervenções em saúde que avaliaram ou não a qualidade metodológica/risco de viés dos estudos incluídos. Além disso, analisar características metodológicas daquelas que avaliam a qualidade metodológica/risco de viés. **MÉTODOS:** A base de dados PubMed foi pesquisada. De 25.571 revisões sistemáticas recuperadas, uma amostra aleatória de 1.025 foi selecionada. Frequências foram usadas para descrever os resultados. Regressões logísticas não ajustadas e ajustadas foram realizadas para testar as associações com a avaliação da qualidade metodológica/risco de viés. Em uma segunda análise, as revisões sistemáticas que avaliaram a qualidade metodológica/risco de viés foram dicotomizadas de acordo com o desenho dos estudos incluídos (somente ensaios clínicos randomizados versus estudos de intervenção não randomizados ou uma combinação de ambos). **RESULTADOS:** 303 revisões sistemáticas foram analisadas na íntegra. A qualidade metodológica/risco de viés foi avaliada por 278 (92%). A avaliação da qualidade metodológica/risco de viés foi associada ao maior número de bases de dados pesquisadas (> 4, P = 0,008), à presença de meta-análise (P = 0,005) e ao desenho dos estudos incluídos (somente ensaios clínicos randomizados, P = 0,042). A chance de usar uma ferramenta adequada e uma ferramenta projetada para avaliação de risco de viés em vez de qualidade metodológica foi maior para revisões sistemáticas que incluíram apenas ensaios clínicos randomizados (P <0,05). A ferramenta mais usada foi a RoB Tool da Cochrane, sem um sistema claro de classificação de risco geral dos estudos. **CONCLUSÃO:** a avaliação da qualidade metodológica/risco de viés foi associada ao desenho dos estudos incluídos (somente ensaios clínicos randomizados), meta-análise dos dados e número de bancos de dados pesquisados (>4). A ferramenta mais utilizada foi a RoB Tool da Cochrane, sem um sistema de classificação claramente definido. A avaliação da qualidade/risco de viés metodológico nos métodos e resultados das revisões sistemáticas, e os impactos na meta-análise e na certeza da evidência, ainda precisam ser abordados de forma mais consistente.

PALAVRAS-CHAVE: Revisão Sistemática. Cuidados de saúde. Qualidade Metodológica. Risco de viés. Avaliação crítica.

Submitted 08/31/2021, Accepted 04/12/2022, Published 05/18/2022

J. Évid-Based Healthc., Salvador, 2022;4:e4067

<http://dx.doi.org/10.17267/2675-021Xevidence.2022.e4067>

ISSN: 2675-021X

Assigned editor: Luis Claudio Correia

How to cite this article: Stefani CM, Rezende LVML, Honnef LR, Oliveira LB, Lopes NC, Oliveira JMD, et al. Methodologic quality and risk-of-bias in systematic reviews of healthcare interventions: a review of methods. J. Évid-Based Healthc. 2022;4:e4067. <http://dx.doi.org/10.17267/2675-021Xevidence.2022.e4067>



Introduction

In the 1990s, David Sakett popularized the concept of Evidence-Based Medicine¹ (EBM, nowadays expanded to Evidence-Based Practice). During the search for the best evidence to support health professionals in their clinical decision-making process and handling the vast amount of information available, Systematic Reviews (SR) arouse as the soundest source of information.

Over the years, the scientific method for planning, conducting, and reporting SR has improved considerably. Largely due to the introduction of reporting guidelines, such as the PRISMA Statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses²), checklists for critical appraisal, as the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews³), and systems for the analysis of the certainty/strength of the evidence, as the GRADE approach (Grading of Recommendations, Assessment, Development and Evaluations⁴).

All of them highlight essential items that should be addressed in SR and are unanimous in pointing out the importance of analyzing the risk-of-bias of included studies to ensure consistent results. Therefore, the assessment of bias in included studies and a careful analysis of its impacts on SR results are central to drawing conclusions in the most impartial and objective manner⁵ and should be carefully addressed by SR authors.

Poor design and/or execution of original studies of healthcare interventions often result in poor internal validity. Thus, assessing the internal validity of studies included in an SR should highlight the risk of under or overestimation of true intervention effect.⁶ Features of studies' design associated with effect overestimation include inadequate random sequence generation and concealment, unexplained or unexplored expressive loss to follow-up, and unblinded outcome assessment.^{7,8} All of these must be properly explored during the critical appraisal of included studies.

Both the first version of the PRISMA Statement² and the revised version⁸ emphasize the importance of risk-of-bias within and across studies assessment in SR. The new version, however, suggests some analytic strategies to examine the influence of included studies' risk of bias on SR results: (i) restricting the primary analysis to studies at low risk of bias (sensitivity analysis); (ii) stratifying studies according to risk-of-bias through subgroup analysis or meta-regression; or (iii) adjusting the result from each study to remove the bias.⁸

A methodologic study evaluated the adherence of SR to the PRISMA checklist and, besides most performed some risk-of-bias assessment within studies (72%), poor compliance to risk-of-bias across studies assessment (38%), and presentation of results of risk-of-bias across studies (30%) was observed.⁹ Another study evaluated 1,114 SR of oral health interventions, of which 61.4% assessed the risk of bias, more consistently in Cochrane reviews than in non-Cochrane reviews (100% versus 56.3% $P < 0.001$) and those published after the PRISMA Statement release.¹⁰

Cochrane's Handbook defines bias "as a systematic error, or deviation from the truth, in results" of included studies. The risk-of-bias analysis comprises the use of tools to evaluate the internal validity of included studies. On the other hand, methodologic quality is assessed by scales, not rarely mixing different concepts (risk-of-bias, imprecision, relevance, ethics, etc.), making their final scores hard to interpret.⁶

Eleven years after the first version of the PRISMA Statement² release and twelve years after the GRADE Approach⁴ introduction, and since both consider mandatory the methodologic quality and/or risk-of-bias (MQ/RoB) assessment and analysis in SR of healthcare interventions, uncertainty remains about the adherence of SR authors to this requirement. Therefore, this study aimed to compare the characteristics of systematic reviews of healthcare interventions assessing or not MQ/RoB. Additionally, to analyze methodologic features of SR that assessed the MQ/RoB of included, including:

- used tools and their suitability to the design of included studies;
- MQ/RoB assessment methodology adopted;
- presentation and interpretation of MQ/RoB assessment.

Methods

The search strategy (Box 1) was developed for PubMed database with the aid of an experienced librarian. The search was performed in September 1st 2020.

Box 1. Search strategy used in PubMed database

("Systematic Review"[Title/Abstract] OR "meta-analysis"[Title/Abstract] OR "metanalysis"[Title/Abstract] OR "meta analysis"[Title/Abstract]) AND ((y_1[Filter]) AND (fft[Filter]) AND (english[Filter])) AND (systematic [sb])

Filters applied: Full text, in the last 1-year, Systematic Reviews, English language. Sort by: Most Recent

The search returned 25,571 records. A random sample of 1,025 (4%) was retrieved for analysis. A random number generator (<https://www.openepi.com/Random/Random.htm>) was used to randomly define the 1,025 records to retrieve. The research team selected the randomly defined references by ticking the boxes on the PubMed results list and exported them to a reference manager (EndNote X9, version 3.3 - Clarivate, Thomson Reuters Corporation, Fairfax, VA, USA).

The inclusion process was done in a two-step approach. In the first step, five reviewers read titles and abstracts of retrieved studies, applying specific eligibility criteria. In the second step, four reviewers read the studies' full text, applying additional exclusion criteria, for final inclusion. At both steps, individual information was cross-checked by a second reviewer. Conflicts were solved by a third experienced reviewer.

SR of healthcare interventions with or without meta-analysis was included. Exclusion criteria at step 1 comprised comments, opinions, letters, protocols, narrative reviews, scoping reviews, brief/rapid SR, overviews of reviews (umbrella reviews), network meta-analysis, association (including association and risk), proportion (including prevalence and incidence), descriptive (including phenomenon, case reports, case series, and surveys), prognostic, diagnostic, qualitative and mixed-methods, methodological, cost-effectiveness, in vitro and in vivo (in animal models), and non-healthcare SR (classified according to Munn et al.¹¹). Cochrane Collaboration SRs were also excluded since they were considered benchmarks.

In step 2, articles were excluded whether not fulfilling the key characteristics of SR according to Cochrane's Handbook¹² and Krnic Martinic et al.¹³ (Box 2).

Box 2. Exclusion criteria considered at step 2: key characteristics of systematic reviews^{12,13}

- 1) SR undertaken by a single reviewer;
- 2) SR not presenting a focused question (based on PICOS or another acronym);
- 3) SR not determining pre-specified eligibility criteria;
- 4) SR not using a pre-specified search strategy;
- 5) SR not applying the search strategy on at least 2 databases;
- 6) SR not presenting a detailed description of the studies' selection process (studies included and excluded in each step);
- 7) Studies to which full text was not obtained.

SR not assessing MQ/RoB were compared to those which, by any means (qualitative/narrative or quantitative through validated tools), assessed the MQ/RoB of included studies.

Data collected comprised author; country; health area; journal impact factor in 2020; funding; adherence to PRISMA Statement; public a priori protocol registration; the number of databases searched; search in grey literature; Methodologic Quality and/or Risk-of-bias (MQ/RoB) tool used and suitability to included studies' design; the number of reviewers involved in MQ/RoB assessment; strategies for conflicts solving; the moment of MQ/RoB assessment – before or after studies' inclusion; use of cut-off point or other systems to studies' quality/bias classification; and interpretation/use of risk-of-bias results (in discussion, meta-analysis, and/or GRADE approach). The same process described for the inclusion process was used for data extraction, and the same five reviewers worked on it.

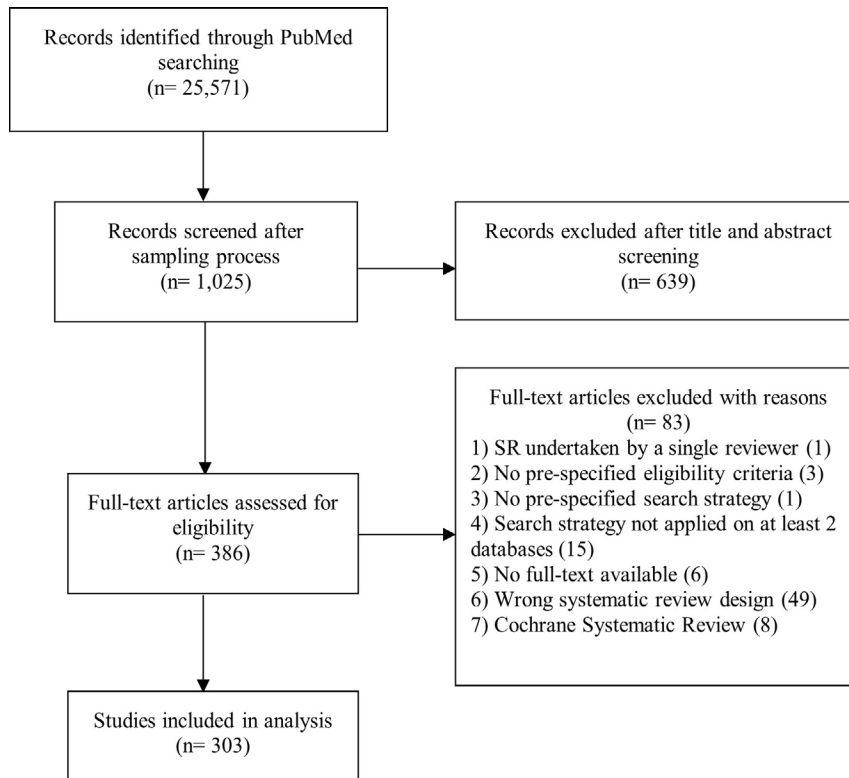
Frequencies were used to describe characteristics of included SR. Unadjusted and adjusted logistic regressions were performed to test the associations with MQ/RoB assessment. The dependent variable was MQ/RoB assessment dichotomized as yes (zero) and no (1). In a second analysis, SRs that assessed MQ/RoB were dichotomized according to the design of included studies - randomized clinical trials (RCT) only as yes (zero) and non-randomized studies of intervention (NRSI) or a combination of both (RCT+NRSI) as no (1). All variables presenting a p-value <0.20 in the unadjusted model entered the final model. Odds ratios (OR) and 95% confidence intervals (CI) were calculated at a 0.05 level of significance. The Statistical Package for Social Sciences program (SPSS for Windows, version 21.0, SPSS Inc. Chicago, IL, USA) was used for analyses.

This study protocol was registered in Open Science Framework platform (www.osf.io) under DOI 10.17605/OSF.IO/CH5T9.¹⁴ All supplementary files are also available in Open Science Framework (<https://osf.io/ch5t9/files/>).

Results

Of 1,025 randomly selected studies, 386 (38%) were initially classified as SR of healthcare interventions at step 1 (title and abstracts reading). In step 2 (full-text reading), 49 SR were excluded due to misclassification as SR of healthcare interventions, 20 were lacking key characteristics of SR; eight were Cochrane's reviews, and the full text was not available for six SR. The remaining 303 (79%) were SR of healthcare interventions, with methodological quality/risk-of-bias (MQ/RoB) assessment of included studies (n= 278, 92%) or not (n= 25, 8%). A flowchart of the inclusion process can be found in Figure 1, and a list of excluded studies with reasons is available in the S1 Table.

Figure 1. Flowchart of literature search and selection criteria



Descriptive characteristics and statistic comparison between SR of intervention with or without MQ/RoB are available at Table 1.

Table 1. Characteristics and statistic comparison of the included systematic reviews according to MQ/RoB assessment (n= 303) (to be continued)

	SR with MQ/RoB assessment (n= 278)	SR with no MQ/RoB assessment (n= 25)	Unadjusted logistic regression*	Adjusted logistic regression
Country			P= 0.057	
High income	167 (60%)	20 (80%)	OR 2.659	NS
Low/Medium income	111 (40%)	5 (20%)	CI 0.969-7.292	
Health Area				
Medicine	219 (79%)	20 (80%)	P> 0.2	-
Psychology	22 (8%)	3 (12%)		
Dentistry	16 (6%)	1 (4%)		
Other	21 (7%)	1 (4%)		
Impact Factor				
≤ Median (2.366)	138 (50%)	16 (64%)	P> 0.2	-
> Median (2.366)	140 (50%)	9 (36%)		
Funding for SR				
Yes	109 (39%)	9 (36%)	P> 0.2	-
No	169 (61%)	16 (64%)		

Table 1. Characteristics and statistic comparison of the included systematic reviews according to MQ/RoB assessment (n= 303) (conclusion)

	SR with MQ/RoB assessment (n= 278)	SR with no MQ/RoB assessment (n= 25)	Unadjusted logistic regression*	Adjusted logistic regression
Databases searched			P= 0.031	P= 0.008
≤ Median (4)	182 (65%)	22 (88%)	OR 3.868	OR 6.046
> Median (4)	96 (35%)	3 (12%)	CI 1.129-13.252	CI 1.598-22.879
Grey literature search				
Yes	111 (40%)	8 (32%)	P> 0.2	-
No	167 (60%)	17 (68%)		
Protocol Registration			P= 0.110	
Yes	123 (44%)	6 (24%)	OR 2.169	NS
No/Not informed	155 (56%)	19 (76%)	CI 0.840-5.599	
PRISMA adherence				
Yes	215 (77%)	17 (68%)	P> 0.2	-
No/Not informed	63 (23%)	8 (32%)		
Included studies' design				
			P= 0.014	P= 0.042
RCT only	152 (55%)	7 (28%)	OR 0.322	OR 0.356
RCT + NRSI / NRSI only	126 (45%)	18 (72%)	CI 0.130-0.796	CI 0.131-0.964
Meta-Analysis			P= 0.000	P=0.005
Yes	214 (77%)	7 (28%)	OR 8.598	OR 5.300
No	64 (23%)	18 (72%)	CI 3.438-21.502	CI 1.672-16.804
GRADE (or other)			P= 0.070	
Yes	59 (21%)	1 (4%)	OR 6.466	NS
No	219 (79%)	24 (96%)	CI 0.857-48.786	

MQ/RoB (Methodologic Quality/Risk-of-Bias); RCT (Randomized Clinical Trial); NRSI (Non-Randomized Studies of Intervention); OR (Odds Ratio); CI (Confidence Interval); NS (Non-Significative).

*Variables with p value <0.20 were included in adjusted model.

The main area of interest was medicine (n= 239, 79%), and the country with the highest number of publications was China (n= 71, 23%). Most SR was published in journals with available Impact Factor (n= 233, 77%) and most authors received no funding for developing the SR or provided no information on funding (n= 185, 61%).

PRISMA Statement adherence was reported by 232 SR (77%), and more than half (n= 184, 61%) did not mention the availability of a published a priori review protocol. Of 119 SR that mentioned a protocol, 97% registered it at PROSPERO (International Prospective Register of Systematic Reviews, <https://www.crd.york.ac.uk/prospERO/>).

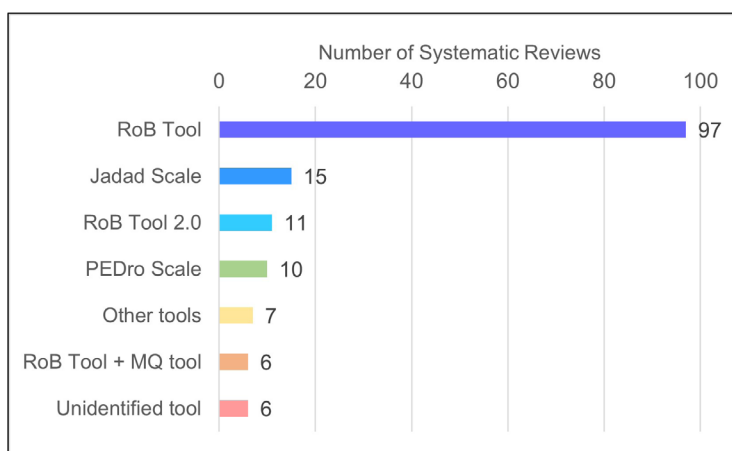
The mean number of databases searched was 4.1 (range 2 to 12) and most included SR did not search grey literature (n= 184, 61%). About half of SR (n= 160, 53%) included only randomized clinical trials (RCT). Most SR performed a meta-analysis (n= 221, 77%), and did not use the GRADE approach (or other analysis of evidence certainty) (n= 243, 80%).

A complete list of included SRs with descriptive characteristics can be found in the S2 Table.

The adjusted logistic regression showed that the chance of not performing the MQ/RoB assessment was higher for SR searching 4 or fewer databases (OR 6.046, CI 1.598-22.879, P= 0.008). Also higher for SR without meta-analysis (OR 5.300, CI 1.672-16.804, P =0.005). And lower for SR including RCT only (OR 0.356, CI 0.131-0.964, P= 0.042).

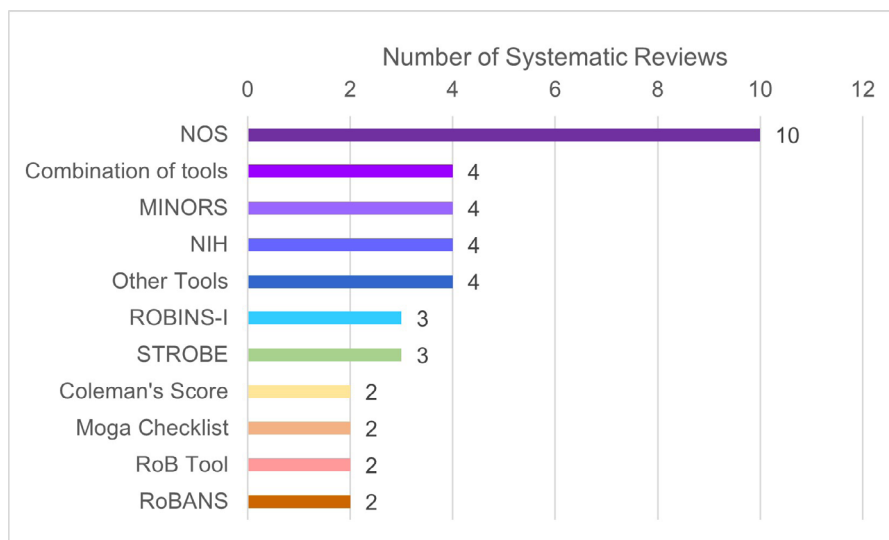
From 278 SR assessing the MQ/RoB of included studies, 152 (55%) included only RCT, 86 (31%) both RCT and NRSI, and 40 (14%) only NRSI. Most used MQ/RoB tool in SR including only RCT was Cochrane’s RoB Tool (2011)¹⁵ (n= 101, 66%). SR including only NRSI used many MQ/RoB different tools, yet the most used was The Newcastle-Ottawa Scale¹⁶ (NOS) (n= 10, 25%). Most SRs including both RCT and NRSI applied different tools for RCT and NRSI, and the most used combination of tools was the Cochrane’s RoB Tool (2011)¹⁵ for RCT and the Newcastle-Ottawa¹⁶ (NOS) for NRSI (n= 13, 15%). Most used tools and combinations for MQ/RoB assessment according to included studies’ designs can be found in Figure 2 (RCT-only), Figure 3 (NRSI-only), and Figure 4 (RCT+NRSI), and a complete list in S4 Table. Main findings of comparison between SR of intervention that assessed MQ/RoB including RCT-only or RCT + NRSI/NRSI-only are available in Table 2.

Figure 2. Methodologic Quality/Risk-of-Bias tools used in Systematic Reviews including Randomized Clinical Trials only



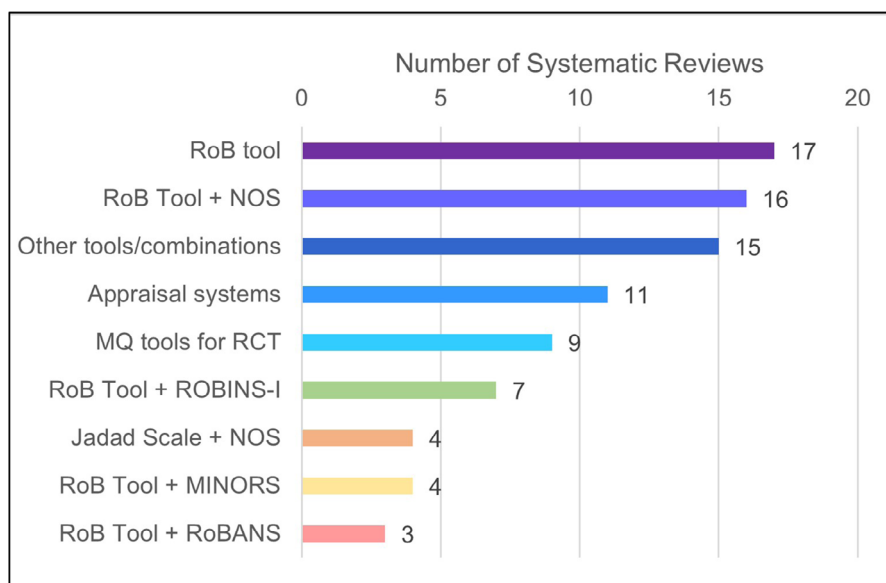
Note: RoB Tool and RoB Tool 2.0 refer to Cochrane’s tools. MQ tool refers to any methodologic quality tool.

Figure 3. Methodologic Quality/Risk-of-Bias tools used in Systematic Reviews including Non-Randomized Studies of Intervention only



Note: RoB Tool and RoB Tool 2.0 refer to Cochrane’s tools. MQ tool refers to any methodologic quality tool.

Figure 4. Methodologic Quality/Risk-of-Bias tools used in Systematic Reviews including both Randomized Clinical Trials and Non-Randomized Studies of Intervention



Note: RoB Tool refers to Cochrane's tools.

Table 2. Main findings of comparison between systematic reviews that assessed MQ/RoB according to the design of included studies (n=278)

	RCT only (n= 152)	RCT + NRSI / NRSI only (n= 126)	Unadjusted logistic regression*	Adjusted logistic regression
Suitable tool for studies' design			P= 0.000	P= 0.000
Yes	151 (99%)	98 (78%)	OR 0.023	OR 0.009
No	1 (1%)	28 (22%)	CI 0.003-	CI 0.001-
			0.173	0.077
Tool designed to assess				
MQ	30 (20%)	61 (49%)	P= 0.000	P= 0.000
RoB	116 (76%)	37 (29%)	OR 14.631	OR 29.255
2 tools (MQ/RoB)	4 (3%)	25 (20%)	CI 5.623-	CI 9.148-
Other	2 (2%)	3 (2%)	38.067	93.559

MQ/RoB (Methodologic Quality/Risk-of-Bias); RCT (Randomized Clinical Trial); NRSI (Non-Randomized Studies of Intervention); OR (Odds Ratio); CI (Confidence Interval); NI (not informed).

*Variables with P-value <0.20 were included in adjusted model.

Methodological quality tools were used by 91 (33%) SR (20% of SR including only RCT), while risk-of-bias tools were used by 151 (54%) SR (80% of SR including only RCT). Most SR (n= 249, 90%) used a proper MQ/RoB tool for the included studies' design.

The MQ/RoB assessment was performed by two independent reviewers in 152 (55%) SR, while 108 (39%) SR did not inform the number of assessors. Strategy to conflict solving was not informed by 163 (59%) SR.

MQ/RoB assessment was more often performed after the final inclusion of studies in SR (n= 239, 86%). The use of clear systems to classify the overall MQ/RoB of included studies or outcomes was described by only 79 (28%) SR.

Meta-analysis was carried out in 214 (77%) SR, of which only 53 (25%) evaluated the impact of MQ/RoB of included studies on meta-analysis results, mainly through sensitivity analysis (n= 36). GRADE approach or other methods for evaluation of certainty/strength/level of the body of evidence was used by only 59 (21%) SR. Most SRs using the GRADE approach (n= 46) described the impact of RoB on the certainty of evidence (n= 32, 70%).

Most SR presented results from MQ/RoB assessment in text or figures (n= 252, 91%). Seven SRs presented results as supplementary files, and 19 (7%) SRs did not present MQ/RoB results at all. Less than half of SR discussed the results of the MQ/RoB assessment (n= 138, 47%).

A list of qualitative characteristics of SR that assessed MQ/RoB can be found in the S3 Table.

Adjusted logistic regression of SR assessing the MQ/RoB of included studies indicated that the chance of choosing an unsuitable MQ/RoB tool was lower for RCT-only SR (OR 0.009, CI 0.001-0.077, P= 0.000); while the chance of using a tool designed for RoB assessment was higher (OR 29.26, CI 95% 9.15 - 93.56, P= 0.000).

Discussion

This review aimed to evaluate the MQ/RoB assessment in a sample of SR of healthcare interventions. SR assessing MQ/RoB were compared to those not assessing it. MQ/RoB assessment was associated with more than four databases searched, design of included studies (RCT-only), and meta-analysis performed. The decision of performing or not a meta-analysis goes through the MQ/RoB assessment, once a meta-analysis based on low quality/high risk-of-bias studies may impair the SR findings and conclusion⁶, thus, the finding of SR not assessing MQ/RoB being also prone to lacking meta-analysis is not surprising.

Increasing the number of searched databases, as well as searching the grey literature, (the so-called comprehensive search) is a manner of reducing the publication bias and meeting the criterion of

finding “all references” answering a certain focused question.¹⁷ This finding, together with the association to RCT design and meta-analysis performing, may suggest that SR assessing MQ/RoB achieved higher methodologic standards.

Considering only the SR that assessed MQ/RoB of included studies, it was noticed that most performed a meta-analysis (77%), a number slightly above the 68% reported by Pussegoda et al.⁹ This could be explained by the focus on SR of interventions and by the inclusion of the term “meta-analysis” in the search strategy.

About 80% of included SR declared adherence to PRISMA Statement² or other SR reporting checklists. Besides no strict analysis of presumed adherence was performed in this study, one criterion of the PRISMA checklist is the mention of a public a priori protocol. Yet less than half of included SR mentioned it. A study analyzing SR published between 1990 and 2014 identified that only 6% fulfilled this criterion.¹⁰ Considering the time elapsed since the PRISMA Statement introduction², SR authors were expected to give more importance to a protocol registration, especially considering the recent evidence that the existence of a registered protocol improves the quality of the SR reporting.¹⁸

Regarding the tools for MQ/RoB assessment, the most used was the Cochrane’s RoB Tool¹⁵ (n= 156, 56%). Considering the SR including only RCT, 70% used the Cochrane’s RoB Tool¹⁵, besides the version 2.018 has been available since 2016 (revised version²⁰ in 2019). Since the included studies were published between 2019/and 2020, it was expected that RoB Tool 2.0 would be more popular. RoB Tool 2.019, differently from its first version, included a classification system of overall RoB at the outcome level.²¹ The lack of a system for RoB Tool¹⁵ resulted in many studies classifying the overall risk of bias with unclear rating criteria (38%).

For NRSI, the most used tool was the NOS (Newcastle-Ottawa Scale¹⁶), yet it did not reach one-third of NRSI-only SR (n= 35/126, 28%). It is possible to notice that there is some uncertainty among authors on the most suitable tool for MQ/RoB assessment for NRSI. The chance of an RCT-only SR choosing an MQ/RoB tool suitable for this study design was much higher

than RCT plus NRSI/NRSI-only SR (108 times higher, $P= 0.000$). This confirms the perception registered in Cochrane's Handbook that "assessing the risk-of-bias in an NRSI has long been a challenge and has not always been performed or performed well".²²

It is true that NRSI comprises a gamma of different study designs. However, if the aim of the SR is to assess the effectiveness of a certain intervention, ROBINS-I (Risk Of Bias In Non-randomized Studies - of Interventions²³) tool should be used for all included NRSI, regardless of the study design.²⁴ Even though the NOS¹⁶ tool is still accepted by the Cochrane Collaboration²⁵, considering that a risk-of-bias tool is preferable, the ROBINS-I tool should be chosen.²⁴

Some tools suit more than one study design (Coleman's Methodology Score²⁶ and Downs & Black Scale²⁷, e.g.), however, they are MQ checklists, rather than RoB tools. The same is true for appraisal systems with tools for each study design, such as the JBI Critical Appraisal Tools (available at <https://jbi.global/critical-appraisal-tools>). The results of this review showed that SR including NRSI-only or both RCT+NRSI were prone to choose methodologic quality appraisal tools regardless of the study's design (48%), while RCT-only SR used mainly RoB tools (70%) (29 times higher chance, $P= 0.000$).

Ideally, the process for reaching RoB judgments should be clearly described in text⁶, mentioning the number of reviewers involved in MQ/RoB, if they worked independently or not and strategies for conflict solving.⁸ The Cochrane Collaboration, as well as the Joanna Briggs Institute, recommend that MQ/RoB assessments are performed independently by at least two reviewers, to reduce errors in assessments and the influence of individual preconceptions. And disagreement is solved through discussion till consensus, the participation of a third reviewer, or both.^{6,28} However, many included SR simply did not mention how many reviewers were involved in the activity (39%), or how the conflicts were solved (59%).

Most SR assessed MQ/ROB of included studies after the studies' inclusion process (86%), and only about 2% of SR described the assessment of MQ/RoB as a tool for a final decision on studies' inclusion. Assessing the MQ/RoB before studies' inclusion

assures that only moderate to good quality (or low risk-of-bias) studies are included, and besides this practice is recommended by some SR guidelines (as the JBI Manual for Evidence Synthesis²⁹, the Cochrane Collaboration recommends that all studies answering a certain focused question should be included. Including all eligible studies may result in precise but biased results, due to study characteristics. Instead, including only the studies at low risk of bias may produce an unbiased imprecise result. Therefore, strategies to cope with bias should be defined a priori and SR authors should judiciously interpret and discuss how studies' biases affected the SR results.⁶

Surprisingly, a few SR (9%), besides performing MQ/RoB assessment, simply did not report the results in text. Others provided a supplementary file with MQ/RoB results; however, free access was not often assured. Ideally, MQ/RoB results should be clearly exposed, preferably by tables or graphs, alongside the text, allowing readers full comprehension and appraisal of the MQ/RoB assessment process.⁶

MQ/RoB assessment results should be used for meta-analysis interpretation and certainty of evidence analysis and be discussed in the text.⁸ SR authors should not perform analyses and interpretations disregarding the results of the MQ/RoB assessment.⁶ Among those SRs that meta-analyzed the data, only 25% evaluated the impact of MQ/RoB of included studies on the results, mainly through sensitivity analysis. Despite sensitivity analysis being considered the main procedure to assess the bias impact in SR results, alternative ways to address it include metaregression, subgroup analysis, and careful discussion of findings in the light of MQ/RoB results.⁶ Nevertheless, less than half of the authors discussed the impact of MQ/RoB on the SR results.

All strategies to cope with bias issues, however, share the risk of being ignored in SR's conclusions, so the MQ/RoB assessment should be incorporated into some certainty of evidence measurement, as the GRADE approach e.g.⁶, and SR authors should make judgments not only for the risk-of-bias within studies but across studies as well.^{2,6} Notwithstanding, twelve years after the GRADE approach introduction⁴, only 17% of the included SR used it to rate the certainty of the evidence.

Conclusion

Most SR of healthcare interventions assessed MQ/RoB of included studies. MQ/RoB assessment was associated with included studies' design (RCT-only), a meta-analysis of data, and the number of databases searched (>4). The most used tool was Cochrane's RoB Tool, with no clearly defined rating system. SR including only RCT used more suitable tools and more RoB than MQ tools when compared to SR including other studies designs. MQ/RoB assessment methods description, results, and impacts on meta-analysis, the certainty of the evidence, and SR results are still to be consistently addressed.

Acknowledgements

The authors would like to thank the librarian Karyn Munyk Lehmkuhl for her assistance in developing the search strategy.

Author's Contribution

Stefani CM participated in the conceptualization of the study, literature screening, data collection, writing, and critical review of the text. Rezende LVML participated in the conceptualization of the study, literature screening, data collection, and critical review of the text. Honnef LR participated in the literature screening, data collection, and organization of supplementary material. Oliveira LB participated in the study conceptualization, literature screening, data collection, critical text review, and literature screening. Lopes NC participated in data collection, and organization of reference lists. Oliveira JMD participated in the conceptualization of the study, literature screening, and critical review of the text. Massignan C participated in the conceptualization of the study, literature screening, statistical analysis, writing, and critical review of the text. Canto GL participated in the study conceptualization, writing, critical review of the text, and study coordination.

Competing Interests

No financial, legal, or political conflicts involving third parties (government, corporations and private foundations, etc.) have been declared for any aspect of the submitted work (including, but not limited to grants and funding, advisory board participation, study design, preparation of the manuscript, statistical analysis, etc.).

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-2. <https://doi.org/10.1136/bmj.312.7023.71>
2. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
3. 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. <https://doi.org/10.1136/bmj.j4008>
4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-6. <https://doi.org/10.1136/bmj.39489.470347.ad>
5. Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting a review [Internet]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020). London: Cochrane; 2020. Available from: www.training.cochrane.org/handbook
6. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020). London: Cochrane; 2020. Available from: www.training.cochrane.org/handbook
7. Page MJ, Higgins JP, Clayton G, Sterne JA, Hróbjartsson A, Savovic J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PloS one*. 2016;11(7):e0159267. <https://doi.org/10.1371/journal.pone.0159267>
8. 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. <https://doi.org/10.1136/bmj.n71>
9. Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev*. 2017;6(1):131. <https://doi.org/10.1186/s13643-017-0527-2>

10. 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(9):720-8.e1. <https://doi.org/10.1016/j.adaj.2016.03.017>
11. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol.* 2018;18(1):5. <https://doi.org/10.1186/s12874-017-0468-4>
12. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011] [Internet]. London: The Cochrane Collaboration; 2011. Available from: <http://handbook-5-1.cochrane.org/>
13. Martinic MK, 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(1):203. <https://doi.org/10.1186/s12874-019-0855-0>
14. Stefani CM, Rezende LVML, Honnef LR, Oliveira LB, Lopes NC, Oliveira JMD, et al. Methodologic quality and risk of bias in Systematic Reviews of Healthcare Interventions: a methodologic review. 2021. <https://doi.org/10.17605/OSF.IO/CH5T9>
15. Green S, Higgins JPT, Alderson P, et al. *Cochrane handbook: cochrane reviews: Ch 8: assessing risk of bias in included studies.* *Cochr Handbook Syst Rev Intervent.* 2011:3–10. <https://training.cochrane.org/handbook/archive/v5.1/>
16. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis, 2000 [Internet]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
17. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence-- publication bias. *J Clin Epidemiol.* 2011;64(12):1277-82. <https://doi.org/10.1016/j.jclinepi.2011.01.011>
18. Santos MBF, Agostini BA, Bassani R, Pereira GKR, Sarkis-Onofre R. Protocol registration improves reporting quality of systematic reviews in dentistry. *BMC Med Res Methodol.* 2020;20(1):57. <https://doi.org/10.1186/s12874-020-00939-7>
19. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials In: Chandler J, McKenzie J, Boutron I, Welch V, editors. [Internet] *Cochrane Methods.* Cochrane Database of Systematic Reviews. 2016. <https://doi.org/10.1002/14651858.CD201601>
20. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:4898. <https://doi.org/10.1136/bmj.l4898>
21. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial [Internet]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). London: Cochrane; 2022. Available from: www.training.cochrane.org/handbook
22. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions version 6.1* (updated September 2020). London: Cochrane; 2020. Available from: www.training.cochrane.org/handbook
23. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ* 2016;355:i4919. <https://doi.org/10.1136/bmj.i4919>
24. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane Handbook for Systematic Reviews of Interventions version 6.1* (updated September 2020). London: Cochrane; 2020. Available from: www.training.cochrane.org/handbook
25. Sterne J, Higgins J, Reeves B, Savović J, Turner L. Cochrane Scientific Committee. Recommendation statement/report. Review of the development of the risk of bias tool for nonrandomised studies for interventions – ROBINS-I [Internet]. Available from: https://methods.cochrane.org/sites/default/files/public/uploads/scientific_committee_statement_report_robins_i_fin.pdf
26. Coleman BD, Khan KM, Maffulli N, Cook JL, Wark JD. Studies of surgical outcome after patellar tendinopathy: clinical significance of methodological deficiencies and guidelines for future studies. Victorian Institute of Sport Tendon Study Group. *Scand J Med Sci Sports* 2000;10(1):2–11. <https://doi.org/10.1034/j.1600-0838.2000.010001002.x>
27. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health.* 1998;52(6):377-84. <https://doi.org/10.1136/jech.52.6.377>
28. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. *JBIM Manual for Evidence Synthesis.* JBI; 2020. <https://doi.org/10.46658/JBIMES-20-04>
29. Aromataris E, Munn Z, editors. *JBIM Manual for Evidence Synthesis.* JBI; 2020. <https://doi.org/10.46658/JBIMES-20-01>