

How to Distinguish Best Evidence from Bias: A Basic Guide to Understanding a Systematic Review

Cristina Pires Camargo, MD*
Rolf Gemperli, PhD†
Rod J. Rohrich, MD‡

Summary: A systematic review is a powerful tool to synthesize and show concise and robust evidence for clinical practice. Thus, the inclusion of biased, low-quality studies should be avoided, for otherwise, the resulting systematic review will not reflect the best medical evidence. Because the methodology of systematic review is relatively new, this review aims to present basic rules to guide the reader's interpretation of a systematic review. (*Plast Reconstr Surg Glob Open* 2020;8:e2625; doi: 10.1097/GOX.0000000000002625; Published online 24 February 2020.)

INTRODUCTION

The vast amount of published data in the medical databases makes it difficult to keep up with advances in clinical practice.¹ Therefore, a systematic review is the best way to find the best medical evidence on a particular subject with implications for our daily questions. However, it must be recognized that some systematic reviews are not actually systematic, and their conclusion consequently does not reflect the best medical evidence. The evidence derived from a good quality systematic review has the potential to save time and money and ultimately improve our patients' safety. The clinical practice based on the recommendations of SR is called evidence-based clinical practice. Some of the aims of evidence-based clinical practice include updating clinical practice, providing policymakers with high-quality evidence on the harm and benefits of interventions, and building an accurate scenario for researchers.²⁻⁴ Considering the impact of an SR on clinical practice and given that SRs are at the peak of the evidence pyramid, one should be very cautious in conducting an SR if there are not enough data to develop such a study design.

Another issue to consider is the fact that a systematic review is a secondary study; the unit of analysis is the primary study. Therefore, the author selects studies that show the best possible available evidence: randomized clinical trials, quasi-randomized clinical trials, and cohort studies (Table 1).^{2,3} Randomized clinical trials are the gold standard for scientific evidence because an RCT is the best

design for the comparison of interventions to treat a certain disease.

However, in many instances, an RCT is not available due to ethical issues. In this situation, the best evidence will be provided by observational studies and quasi-randomized studies. The reader can then reflect that in some medical conditions the literature cannot provide the best evidence, for instance, where the search strategy retrieved only case-series studies. When we face this situation, the authors can summarize the data but a statistical analysis [meta-analysis (MA)] is not possible.^{3,5}

A BRIEF HISTORY OF SYSTEMATIC REVIEWS

Archibald Cochrane, a Great British physician, seeded the idea of searching for the best evidence to offer the best treatment for the patients.⁶ Cochrane's obsession with effective treatment was the core of modern evidence-based medicine, of which the initial idea was born in one of his publications: *Effectiveness and Efficiency: Random Reflections on Health Services* of 1972. This book inspired the medical community to think critically about medical evidence and effectiveness. In 1976, under Cochrane's influence a young doctor, Ian Chamber, published the first systematic review (MA). These study results modified UK governmental healthcare.⁶

In the 1970s, some academic institutions started adopting best-evidence medicine. One of these pioneering institutions was McMaster University, which decided to create a department and invited Dr. David Sackett (1934–2015) to lead it.⁷ Since that time, SRs have come to be considered as providing the best evidence for medical purposes. Respectable organizations such as Cochrane's and Campbell's Collaborations have been working to assure the quality of SRs.^{8,9}

In the 1980s and 1990s, the term "systematic review" started to appear in the medical database. Some institutions, such as Cochrane's and Campbell's Collaborations, have defined guidelines to produce good-quality systematic reviews. More

From the *Laboratory of Microsurgery and Plastic Surgery (LIM-04), School of Medicine, Universidade de São Paulo, Brazil; †Plastic Surgery Division, School of Medicine, Universidade de São Paulo, Brazil; and ‡Dallas Plastic Surgery Institute; Dallas, Tex.

Received for publication July 18, 2019; accepted December 9, 2019.

Copyright © 2020 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](#), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/GOX.0000000000002625

Disclosure: The authors have no financial interest to declare in relation to the content of this article.

Table 1. Definitions of RCTs, Quasi-Randomized Clinical Trials, and Observational Studies

Study Design	Definition
Randomized clinical trial	To avoid bias, this study use randomization, concealed allocation and blinding
Quasi-randomized clinical trial	This study, by ethical reason, did not randomize the patients. But follow all other items as an RCT
Observational (cohort)	This study design entails some degree of selection bias

recently, in the 1990s, a group of scientists created an SR guideline: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) that comprises 27 items related to SR and MA. Even though SRs have strict rules, some reviews have not followed PRISMA guidelines. Therefore, the outcome of this type of review does not show the best evidence because these reviews collect data from biased studies. Due to this fact, this review aims to present some essential topics to bear in mind when analyzing an SR, and ultimately to clarify the correct interpretation of data.^{9,10}

HOW TO ANALYZE A SYSTEMATIC REVIEW

Unlike a narrative review, an SR follows rigid rules to find the best scientific evidence. These rules will assure the best evidence in primary studies. In a practical way the authors need to think in accordance with the “garbage-in, garbage-out” principle.^{3,11}

The guideline to assist authors is the PRISMA statement. This statement provides the minimum requirements for reporting SR and MA.

First, a rationale and background must be stated. An SR usually focuses on a clear research question which should be clinically relevant. The research question can be summarized using the PICOT/PECOT structure (Table 2).⁴

The introduction, background, and rationale will indicate existing knowledge, gaps in the literature, and finally the aim of the review.

Eligibility Criteria

Before the literature search, the authors must define the eligibility criteria for the inclusion of primary studies, which include the study design, the demographic characteristics and interventions to be compared, and the outcomes of interest.^{4,5,12} A good outcome should be clinically relevant for decision-making, and this outcome needs to be time-framed (ie, short-term or long-term outcome). As in other studies, the outcome is divided into primary and secondary outcome. Frequently, in an SR study, the primary outcome is an effective outcome and a co-primary outcome is a safety outcome (major adverse events). The secondary outcomes can also be minor effective outcomes (surrogate outcomes) and other safety issues (minor adverse events).^{4,5}

Table 2. PICOT/PECOT Acronym

P: population of interest	P: population of interest
I: intervention	E: exposition
C: comparator or control	C: comparator or control
O: outcome	O: outcome
T: time of follow-up	T: time of follow-up

Database

In an SR, the ideal scenario is to search all available databases. The most commonly searched databases are the following:

- Medline/ PUBMED¹³ (from 1946)
- EMBASE¹⁴ (from 1974)
- COCHRANE LIBRARY¹⁵ (the Cochrane Central Register of Controlled Trials (CENTRAL))
- LILACS¹⁶ (Latin American and Caribbean Health Science Information database, from 1982)

Some private databases are also checked, but the author has to consider the potential bias of these databases because a private institution has financial interests that can interfere in data acquisition. The choice of the databases will depend on the topic; for instance, in a psychological review it is crucial to consult the American Psychological Association. The most frequent private databases are Web of Science, SCOPUS, EBSCO, and APA.

Other reference sources are obtained through the cross-referencing process, that is, searching all available abstracts presented in the most important medical congresses related to the research question. In the congress abstract, you can find negative trials that will help to find the best evidence and avoid publication bias. Trials with negative results have less chance of being published.

The authors also need to search in register platforms to check ongoing studies, which can clarify research directions and some potential adverse events. The most commonly used protocol register platforms are: The metaRegister of Controlled Trials,¹⁷ The US National Institutes of Health Ongoing Trials Register,¹⁸ The Australian New Zealand Clinical Trials Registry,¹⁹ The World Health Organization International Clinical Trials Registry platform,²⁰ and The EU Clinical Trials Register.²¹

The ongoing and awaiting trials will guide the authors toward future perspectives. Moreover, these trials will be checked when the SR is eventually updated, as the complete protocols that were not published can reveal publication bias.

Register in a Database

As in other clinical studies, the SR protocol register is recommended. There is a dedicated platform for SR named PROSPERO—the International Prospective Register of Systematic Reviews.³ The goal of registering a protocol is to monitor fraud and misconduct (a change of outcomes or assessment tools in the middle of or after data acquisition).

Language

Although English is the most widespread language in science, language restrictions are not recommended,

because some conditions are more prevalent in certain non-English-speaking regions.

Period

Searching without a limited period is desirable, except in the case of new health conditions. However, if the focus of the research changes with time, for example, antibiotics and bacterial resistance, a sensitivity analysis (SA) (see below) is useful for studying the influence of a certain “old-fashioned” drug, surgery, or intervention.

Gray Literature

“Gray literature” refers to unpublished articles. The sources for gray literature searches are theses and dissertations, medical reports, government reports, conference papers, and ongoing research.²² There is a database dedicated to gray literature.^{23,24} Depending on the medical issue, this database is incomplete, but it is worth checking because some sponsored papers with negative results could not be accepted for publication.

After including and excluding all the records and studies a diagram flow is mandatory (Fig. 1).

Intervention

As in other study designs, SRs must also search for primary and secondary endpoints. In general, the primary outcome must be an outcome for effectiveness and for safety.⁴

Effectiveness is more useful than efficacy. Efficacy reflects the effect of a certain intervention/diagnosis in a trial, but effectiveness indicates whether these interventions work in real life. For example, a new antibiotic has efficacy in treating MRSA infection (efficacy); however, this drug is too expensive or causes many adverse events, compromising its effectiveness.

Unit of Analysis

The unit of analysis is the primary study, preferably RCTs. The best evidence is obtained from RCTs, but, as previously discussed, quasi-randomized and observational

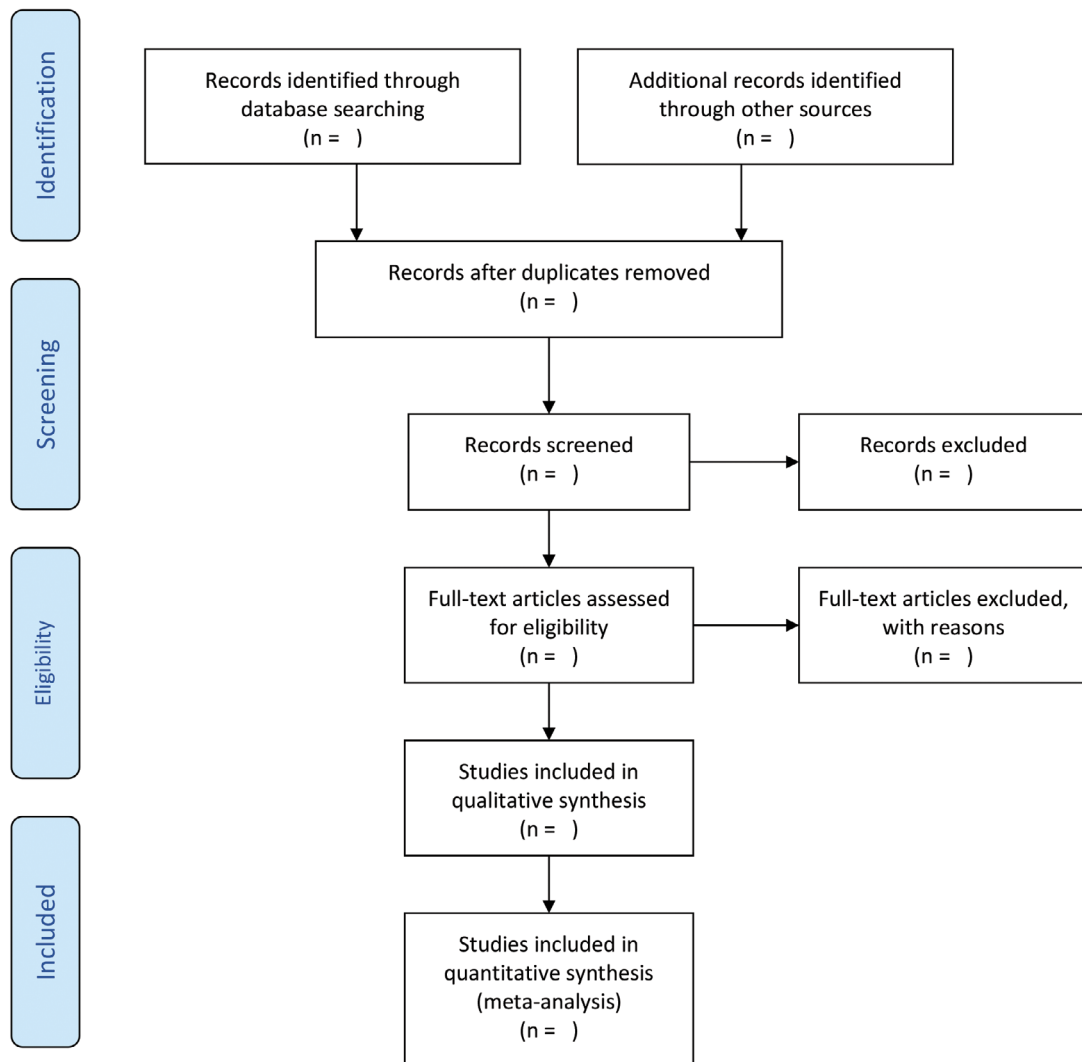


Fig.1. PRISMA flow diagram.

studies can be selected if RCTs are not available. However, the researcher should not pool and analyze the results from RCTs, quasi-randomized, and observational studies together in SR. Randomization can produce homogeneous groups regarding known and unknown variables. An observational study creates heterogeneous groups. Therefore, a cause-effect analysis is not adequate in this last study design.

Moreover, case series and case reports should not be included in an SR, because they are usually highly biased due to the lack of randomization, allocation, blinding, and a comparator group.

Study Selection (SA)

The study selection process demands 2 independent authors. Any disagreement between the authors on the studies to be included in the SR should be resolved by a third author or by the 2 authors during a consensus meeting. Some useful tools are available to select papers: Endnote, Medley, and Rayyan, for example.

If the authors use review manager software [Review Manager (RevMan) (Computer program). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014], they will fill in the study information in a table of study characteristics, and a risk of bias will be generated for each individual study. The authors will also have to address the reason for every study exclusion.

In the clinical trial register, we can find trials in the following phases: recruitment started and completed. Consequently, the trial protocol in recruitment and the started phase will be classified in the SR as ongoing studies. The completed protocols that have not been published yet will be classified in the SR as “awaiting classification studies.”

Data Extraction

According to the pre-specified protocol, the authors will extract data from the included articles. It is desirable to create a document that allows the collection of data (Table 3).

In this phase, data extraction, some information may be missing. The PRISMA and Cochrane Handbook advises contacting the study authors requesting further information. All correspondence must be documented.

Assessment of Risk of Bias

Bias means a systematic error that causes a non-truthful outcome. It can either underestimate or overestimate the intervention effect. Scales and checklists are used to assess the risk of bias of the primary studies. Risk of bias assessment relates to the evaluation of the internal validity of the primary study. A well-known scale, the Jadad Scale (a seven-domain scale), assigns grades for each study domain (Table 3). The Jadad Scale comprises 5 positive points and 2 negative points, and the final grade will reflect the chance of the risk of bias. The Jadad Scale seems somewhat simplistic.^{5,25}

Currently, the Cochrane Collaboration Handbook recommends the use of a seven-domain-based evaluation of bias. This evaluation tool is called RoB (Risk of Bias; Table 4).²⁶

Table 3. An Example Data Extraction Form

Source Eligibility Methods	Study ID/Review ID Inclusion and exclusion criteria Study design Study duration Randomization (sequence generation) Blinding
Participants	Setting (outpatients or inpatients) No. participants and number of participants per group Demographic data (age, gender, race, ethnicity) Diagnostic criteria Comorbidities
Interventions	Specify the interventions
Outcomes	Describe each outcome (type, number of patients) Unit of measurement No. participants per group
Result	Missing data Summary of the data
Other information	Funding Bias Correspondence information* Study violation

*indicates information obtained by contacting the author via e-mail. ID, identification

Risk of Bias of Observational Studies

In some areas of healthcare (surgery, physiotherapy), randomized clinical trials are not feasible for ethical reasons. In these cases, the best evidence is provided by observational studies (cohort, case-control). However, observational studies carry a higher risk of selection bias. Consequently, a different tool than RoB must be applied to assess risk of bias in observational studies.

Several scores and scales have been proposed, such as the Newcastle-Ottawa Scale²⁷ and the Downs-Black Scale.²⁸ Although these scales are intended for the evaluation of non-RCT studies, the Newcastle-Ottawa and Downs-Black Scales' assessments regarding external and internal validity depend on the SR author's judgment. To overcome this obstacle, the Cochrane collaboration created a new score (Risk Of Bias In Non-Randomized Studies of Interventions, or ROBINS-I) to evaluate this bias. This scale considered the effectiveness and safety of an intervention from non-randomized studies.²⁹

Different types of study designs demand different scores and scales, as shown in Table 5.

Statistical Analysis

Data Synthesis

The authors will deal with qualitative and quantitative data. Qualitative data, if relevant, will be discussed in the text. Depending on some assumptions, quantitative data can be summarized in an MA. An MA is a statistical tool designed to compare numerical, categorical, and frequency proportion data extracted from similar studies. MAs increase the power of the study and decrease imprecision, raising questions that were not posed by the primary studies. Their major disadvantage is the potential to mislead the reader, amplifying the bias of the primary studies. This can happen if we mix “apples and oranges,” for example, different types of studies, case-series, or animal and human studies.^{4,5,30}

Table 4. The Description and Definition of Bias Domains for RCTs According to RoB (Risk of Bias)

Type of Bias	Domains	Justificative
Selection	Random sequence generation	Describe the randomization method
Selection	Allocation concealment	Check if the authors performed concealment allocation
Performance	Blinding participants and staff	Describe how the authors blind the participants and study staff
Detection	Blinding outcome assessment	Describe how the authors maintain blindness to avoid problems in the evaluation
Attrition	Incomplete outcome report	Describe the reason for drop-outs
Reporting	Selective reporting	Check if all the outcomes described in the methods show results
Other	Conflict of interest or any concern about the study	Describe any issue related to bias (study violation, conflict of interest)

The MA tests by pair-wise comparisons, intervention versus placebo, or two types of intervention. The MA will evaluate the difference between 2 treatments, also known as the effect or treatment.

When performing a data analysis, whether qualitative or quantitative, the reviewer must keep in mind the following 4 questions:

- The direction of the effect;
- the size of the effect;
- the effect consistency across studies; and
- the evidence for the strength of the effect.^{5,31}

An MA is able to answer the first 3 questions, while the last will depend on the reviewer’s judgment.

The reviewer must use his or her judgment to select homogeneous studies to be assessed with an MA. Different study populations and methodological tools can lead to different outcomes. These differences are classified as methodological heterogeneity, and depend exclusively on reviewer selection.

After the review analysis and study selection, MA will check for statistical heterogeneity. Both heterogeneities contribute to the final statistical heterogeneity. To compare different studies, a certainty evaluation is mandatory in terms of measurable heterogeneity. To deal with this measurement we can use the Cochrane *Q*, *I*.^{5,31,32}

Depending on the *I* value, 4 possible situations can happen:

- 0%–40%: heterogeneity might not be important;
 - 30%–60%: may represent moderate heterogeneity*;
 - 50%–90%: may represent substantial heterogeneity*;
- (* Cochrane Handbook classification, most of the authors only considered 40% of heterogeneity as threshold).

The statistical analysis will consider each study according to the effect direction, sample size, and heterogeneity. If the reader encounters some studies with directional discrepancies (beneficial effect or not), he or she must check whether the data extraction was correctly done.^{31,32}

When selecting studies with small sample sizes, such as <50 events or subjects, the confidence interval can be too

wide and the risk of an outcome by chance can become significant (see GRADE evaluation).³²

The value of each study depends on the sample size and precision, which will receive weight values. These weight values will be evaluated to compound the diamond.^{31,32} (Fig. 2)

Regarding the MA of observational studies, the results must consider selection bias. The characteristics of nonrandomization studies (cohort, case-control) can complicate the analysis in terms of robustness of the results to increase the evidence of certainty. Different populations (exposed group versus non-exposed group), create methodological and statistical heterogeneity. Recently, some researchers use a statistical technique propensity score to minimize population heterogeneity in retrospective observational studies. This statistical method matches the participants’ characteristics to create a pseudorandomization. The idea is to calculate scores using known covariates and according to these scores, match participants from different groups. The propensity score increases the strength of observational studies to prove the cause-effect. However, the Propensity score has some limitations. Some methods decrease the number of participants and can cause an error type II (less power). Several meta-analyses based on observational studies analyses studies in a match and unmatched way to compared the outcome directions and a possible intervention or exposure overestimation. More studies are needed to compare propensity score methods’ effectiveness.^{33,34}

Funnel Plot

The funnel plot will inform the reader if there is publication bias. To perform a funnel plot it is necessary to have at least 10 studies. In the case of no publication bias, we expect a homogeneous distribution of studies on both sides of the funnel plot.^{5,35}

A funnel plot has a cutoff line; this graphic plots the standard error (y-axis) and the logarithm of the odds ratio (x-axis). Studies with positive results are grouped on one side while those with negative results are on the other side.

Studies with small sample sizes will plotted in the lower part (base) of the funnel plot, because they show a wider confidence interval. It is worth noting that several studies with small sample sizes are equally distributed in both sides, because studies with small sample sizes show less power and can show no effect (negative study) due to type II error.^{5,35}

Table 5. The Risk of Bias Scales by Study Design

Study	Risk of Bias Tool
Randomized clinical trial	RoB-Cochrane
Observational studies	ROBINS-I
Systematic review	ROBI
Diagnose test accuracy	QUADAS-2

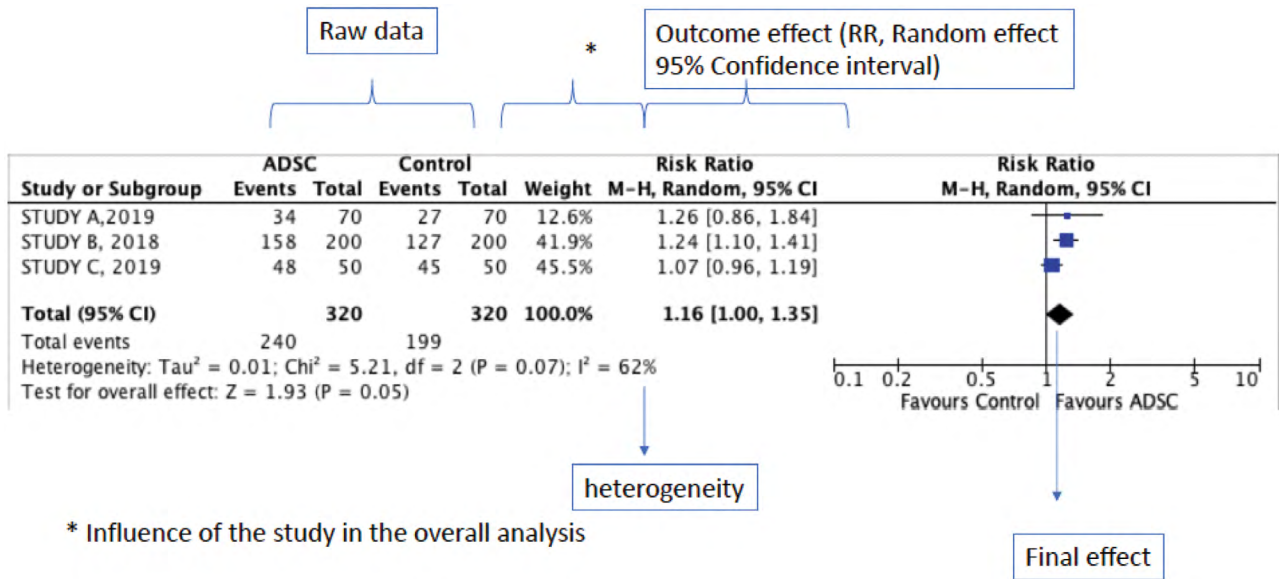


Fig. 2. Forest plot illustration.

The upper part of the funnel plot will be populated with larger-sample studies. If the funnel plot shows more positive studies, this suggests publication bias, whereby most studies with negative results are not published (Fig. 3).^{5,35}

Sensitivity Analysis (SA)

Put simply, SA is the same thing as MA, performed, however, under reviewer scrutiny. If the reviewer observes that some studies showed an arbitrary or unclear result, an SA is performed excluding these studies, that is, a second analysis is performed segregating these studies. The aim of this test is to show the best evidence without unclear data.^{31,36}

For example, studies with a high dropout rate are excluded. When we faced a high dropout rate, even the authors provided a reason for the dropout rate; the statistical analysis was compromised due to an available small sample.

If the SA showed the same outcome as the MA, it improves the certainty of the SR.^{5,34} SA is a crucial tool to determine whether a study could lead to a misinterpretation of the data.

Subgroup Analysis

Certain health conditions can be affected by subpopulation characteristics (gender, age, ethnicity, etc.). A subgroup analysis can point out these subgroups' behavior.³⁷

It is important to clarify the differences between SA and subgroup analysis. SA only removes some studies in the MA. Subgroup analysis is a statistical tool to study some relevant study subpopulation characteristics.^{4,5,37}

Meta-regression

Meta-regression is an extension of subgroup analysis. This tool will test the association of some clinically relevant factors and their influence on the outcome. Although the

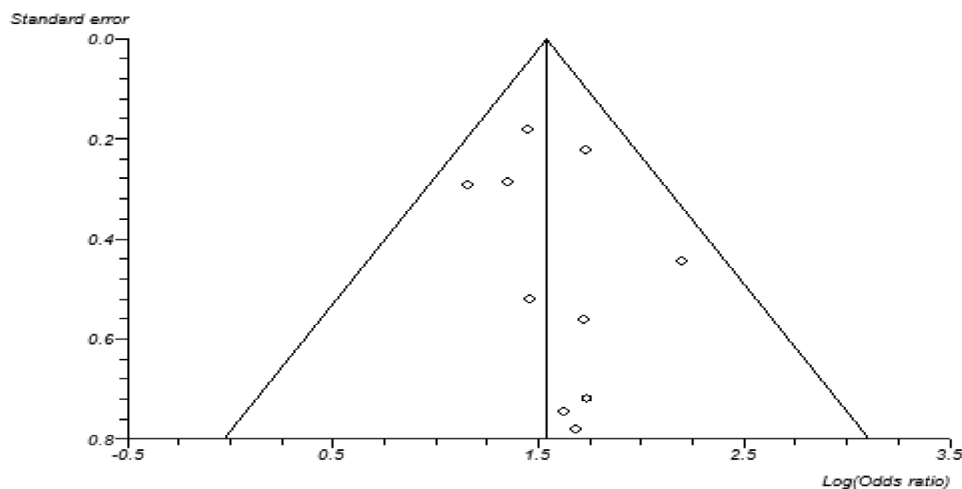


Fig. 3. The upper graph shows a homogeneous study distribution, suggesting no publication bias. The lower graph shows a heterogeneous study distribution (publication bias).

Table 6. GRADE Domains Description

GRADE Domains	
Limitation	Assess the Risk of Bias outcome (randomization, allocation, blinding, conflict of interest)
Indirectness	Assess how well the review outcome answers the clinical research question
Inconsistency	Heterogeneity regarding clinical aspects (population, intervention) and methodological aspects (<i>I</i> index)
Imprecision	Refers to sample size, quantity of number of events. As a rule of thumb in an SR (dichotomous variable), if the review shows <300 events the certainty is downgraded one level
Publication bias	Consider the size of included studies, funnel plot outcome

idea is interesting, meta-regression is performed only if there are 10 or more relevant studies assessed by MA.^{38,39}

Certainty of the Evidence

The certainty of evidence will assess the strength of the evidence (high, moderate, low, and very-low evidence).²⁷ The Grading Recommendations Assessment, Development, and Evaluation (GRADE) group approaches to rating certainty evidence are as follows. This tool uses 5 domains for potential certainty, and downgrading: study limitations, indirectness, inconsistency, imprecision, and publication bias (Table 6).⁴⁰⁻⁴²

Final Considerations and Conclusions

The idea of this article is to empower readers' critical thinking to decide the best evidence for their clinical practice. Although this article has simplified some steps to facilitate the basic comprehension of a systematic review, after reading this article, the reader will be able to identify whether an SR shows the best medical evidence and whether a particular SR is adequate by analyzing all the items described above. Also, it is interesting to note that surgical SR has to be analyzed cautiously because of multiple particularities that can impair the certainty of evidence: Observational studies and operator dependence, among other factors.

Cristina Pires Camargo

Av Brigadeiro Luis Antonio 4161 01402-001

São Paulo, Brazil

E-mail- cristinacamargo@usp.br

REFERENCES

- Sackett D. How to read clinical journals: I. why to read them and how to start reading them critically. *Can Med Assoc J*. 1981;124:555-558.
- Guyatt G. Evidence-based medicine. *Ann Intern Med*. 1991;14(Suppl 2):A-16.
- Moher D, Liberati A, Tetzlaff J, et al. The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6:e1000097.
- Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration; 2011. www.cochrane-handbook.org.
- Zhang L, Gerson L, Maluf-Filho F. Systematic review and meta-analysis in GI endoscopy: why do we need them? How can we read them? Should we trust them? *Gastrointest Endosc*. 2018;88:139-150.
- Stavrou A, Challoumas D, Dimitrakakis G. Archibald Cochrane (1909-1988): the father of evidence-based medicine. *Interact Cardiovasc Thorac Surg*. 2014;18:121-124.
- Collier R. Dr. David Sackett, a giant among giants (1934-2015). *CMAJ*. 2015;187:640-641.
- Shamseer L, Moher D, Clarke M, et al; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647.
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
- Smith R, Rennie D. Evidence based medicine—an oral history. *BMJ*. 2014;348:g371.
- InformedHealth.org. What are systematic reviews and meta-analyses?. Cologne: Institute for Quality and Efficiency in Health Care (IQWiG);2016. <https://www.ncbi.nlm.nih.gov/books/NBK390295>. Accessed March 17, 2019.
- Greenhalgh T. How to read a paper: getting your bearings (deciding what the paper is about). *BMJ*. 1997;315:24-26.
- National Institutes of Health. PubMed [Database]. Bethesda, MD: National Library of Medicine. <http://pubmed.gov>. Accessed November 8, 2019.
- Embase. Excerpta Medica Database. OVID [Internet]. Amsterdam: Elsevier;1974. <https://www.embase.com>. Accessed November 8, 2019.
- Cochrane Central Register of Controlled Trials (CENTRAL) - Cochrane Library [Internet]. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2015. <http://www.cochranelibrary.com>. Accessed November 8, 2019.
- Latin American and Caribbean Center on Health Sciences Information [Internet]. São Paulo: Pan American Health Organization; 1984. Available from: <http://lilacs.bvsalud.org>. Accessed November 8, 2019.
- International Standard Randomised Controlled Trial Number Isrctn registry [Internet]. London: BioMed Central; 2000. Available from: <http://www.controlled-trials.com>. Accessed November 8, 2019.
- ClinicalTrials.gov [Internet]. Bethesda, Maryland: U.S. National Library of Medicine; 2000. <http://www.clinicaltrials.gov>. Accessed November 8, 2019.
- Australian and New Zealand Clinical Trials Registry [Internet]: Sydney, NSW: NHMRC Clinical Trials Centre, University of Sydney; 2005. <https://www.anzctr.org.au>. Accessed November 8, 2019.
- International Standards for Clinical Trial Registries - Version 3.0 [Internet]. Geneva: World Health Organization; 2018. <http://www.who.int/trialsearch>. Accessed November 8, 2019.
- EU Clinical Trials Register [Internet]. Amsterdam: European Medicines Agency; 2004. <https://www.clinicaltrialsregister.eu>. Accessed November 8, 2019.
- Paez A. Grey literature: an important resource in systematic reviews. *J Evid Based Med*. 2017;10:233-240.
- Grey Literature Report [Internet]. New York: The New York Academy of Medicine; 1999. <http://www.greylit.org>. Accessed November 8, 2019.
- Open Grey Repository [Internet]. Paris: Institut de l'Information Scientifique et Technique- Laboratoire CNRS; 2011. <http://www.opengrey.eu>. Accessed November 8, 2019.

25. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1–12.
26. Higgins JP, Altman DG, Gøtzsche PC, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
27. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed March 18, 2019
28. 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;52:377–384.
29. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
30. Akonbeng A K. Understanding systematic reviews and meta-analysis. *Archives of Disease in Childhood*. 2005;90:845–848.
31. Borenstein M, Hedges LV, Higgins JPT, et al. *Introduction to Meta-Analysis*. Chichester (UK): John Wiley & Sons; 2008.
32. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172:137–159.
33. Rosenbaum PR, Rubin DB. The central role of propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
34. Kuss O, von Salviati B, Börgermann J. Off-pump versus on-pump coronary artery bypass grafting: a systematic review and meta-analysis of propensity score analyses. *J Thorac Cardiovasc Surg*. 2010;140:829–835, 835.e1.
35. Sedgwick P. Meta-analyses: how to read a funnel plot. *BMJ*. 2013;346:f136. 2.
36. El-Kadiki A, Sutton AJ. Role of multivitamins and mineral supplements in preventing infections in elderly people: systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2005;330:871.
37. Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93–98.
38. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med*. 2004;23:1663–1682.
39. Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *Online J Curr Clin Trials*. 1994;Doc No 134.
40. Montgomery P, Movsisyan A, Grant SP, et al. Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the GRADE approach in global health. *BMJ Glob Health*. 2019;4(Suppl 1):e000848.
41. Ryan R, Hill S. How to GRADE the quality of the evidence. 2016. Cochrane Consumers and Communication Group; <http://cccr.org/cochrane.org/author-resources>. Accessed December 2016.
42. Higgins JPT, Altman DG, Sterne JAC, eds. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Churchill R, Chandler J, et al eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0*. Cochrane; 2017. Available from <http://www.training.cochrane.org/handbook>.