

Principles and Practice of Clinical Research

A Global Journal in Clinical Research



PPCR

ISSN: 2378-1890

A Practical Guide to Perform a Systematic Literature Review and Meta-analysis

A.F.H. Ortiz^{*1}; E.C. Camacho²; J.C. Rojas¹; T.C. Camacho¹; S.Z. Guevara¹; N.T.R. Cuenca³; A.V. Perdomo¹; V.C. Herazo¹; R.G. Malo¹.

¹ Universidad El Bosque, Bogotá, Colombia. ² Universidad Javeriana, Bogotá, Colombia. ³ Fundación Universitaria de Ciencias de la Salud (FUCS), Bogotá, Colombia.

*Corresponding author: Andrés Felipe Herrera Ortiz, Universidad El Bosque, Bogotá, Colombia. E-mail: afherreraor@gmail.com

Received September 1, 2021; accepted October 27, 2021; published December 31, 2021.

Abstract:

Nowadays, systematic literature reviews/meta-analyses of clinical trials are considered the best evidence in clinical research; thus, if performed appropriately, they can save resources by avoiding the development of unnecessary clinical trials. Nevertheless, to carry out a systematic literature review /meta-analysis, researchers must deeply understand its methodology. This narrative review aims to act as a learning tool for new researchers to perform systematic literature reviews/meta-analyses for categorical variables, addressing this matter in a step-by-step fashion using STATA 17. Readers must understand that the scope of this topic is much broader. Therefore, concepts such as meta-analysis for numerical outcomes or meta-analysis for survival analysis are not addressed. The authors hope that the information presented in this manuscript serves as pillars to enhance research worldwide, especially in places where formal research training is uncommon. If readers want to extend their knowledge regarding this subject, we recommend reading the "The Cochrane Handbook for Systematic Reviews of Interventions and the Center for Reviews and Dissemination" or the book "Finding What Works in Health Care: Standards for systematic reviews."

Keywords: Systematic literature review, Meta-analysis, Research, Methodology, STATA, Forest plot.

DOI: <https://doi.org/10.21801/ppcrj.2021.74.6>

INTRODUCTION

A systematic literature review (SLR) is a qualitative synthesis of information regarding a specific research question; it is carried out based on a prespecified protocol to objectively select the articles that will be part of our study (Manterola et al., 2013). In some cases, SLR may end up in a meta-analysis (MA) if sufficient evidence would allow us to use statistical methods to perform a quantitative synthesis of the information. MA represents the pooled effect of all the studies included in the SLR to provide an overall result; therefore, it is an objective evaluation of the information based on numerical data (Haidich, 2010).

SLR/MA can be used for different purposes, such as keeping physicians updated regarding a specific topic, evaluating the methodological quality of published studies, detecting knowledge gaps when nothing has been published, or answering knowledge gaps when there are discrepancies in the literature (Impellizzeri, 2012).

Nowadays, SLR/MA of clinical trials are considered the best evidence in clinical research; thus, if performed appropriately, they can save resources by avoiding the development of unnecessary clinical trials (Lau et al., 1992; Wallace, 2010). Nevertheless, to carry out a SLR/MA, researchers must deeply understand its methodology, which represents a challenge, especially in developing countries where research mentoring is

not so strong. That is why this narrative review aims to act as an easy and accessible learning tool for new researchers to perform SLR/MA for categorical variables, addressing this matter in a step-by-step fashion using STATA 17. Concepts such as MA for numerical outcomes or MA for survival analysis will not be discussed at this opportunity.

STEPS TO CARRY OUT A SYSTEMATIC LITERATURE/ META-ANALYSIS

The process to perform a SLR/MA involve several steps (Drucker et al., 2016; Khan et al., 2003):

1. Define the research question.
2. Write and publish the protocol of the SLR.
3. Carry out the literature search.
4. Remove the duplicated articles.
5. Perform the screening for eligibility.
6. Assess the methodological quality of the selected papers.
7. Carry out the data extraction and qualitative synthesis
8. Perform the quantitative synthesis (Only if MA is going to be carried out).
9. Assess publication bias (Only if MA is going to be carried out).
10. Perform sensitivity analysis (Only if MA is going to be carried out).
11. Publish the SLR/MA.
12. Update the SLR/MA.

We are going to discuss each of these steps separately throughout the manuscript.

1. DEFINE THE RESEARCH QUESTION

The research question should focus on topics that generate certain discrepancies in literature or clinical practice. Additionally, the research question should be designed so that the outcomes might be interesting for the scientific community regardless if they are positive or negative. Researchers must build the research question based on the PICOT acronym: P (Population), I (Intervention), C (Comparison), O (Outcome), T (Time) (Munn et al., 2018; Sargeant et al., 2020).

The investigators must classify their research questions according to those described in Table 1 to facilitate the building of the SLR/MA search strategy. Based on the question categorization, the investigator will determine the types of studies that might be

suitable to include in the SLR/MA to answer the research question.

Table 1. Types of research questions and the studies to be included to answer it

Types of research question

Incidence question: The authors want to determine the incidence of a disease

Prevalence question: The researchers want to determine the prevalence of a disease

Therapy question: The investigators want to test a specific drug or intervention

Screening question: The authors want to evaluate a screening test for a specific disease

Diagnostic accuracy question: The researchers want to test the diagnostic accuracy of a specific test compared to a gold standard

Prognostic question: The authors want to evaluate the prognosis of a disease

Harm question: The researchers want to evaluate if an intervention has adverse effects on the participants

Etiology question: The investigators wish to assess the etiology of a specific condition

Studies to be included to answer the research question

Include cohorts studies and case-control studies.

Include cross-sectional studies

Include clinical trials

Include clinical trials

Include clinical trials, cross-sectional studies

Include clinical trials, cohorts studies, case-control studies

Include clinical trial, cohorts studies, case-control studies

Include cohorts studies and case-control studies

2. WRITE AND PUBLISH THE PROTOCOL OF THE SYSTEMATIC LITERATURE REVIEW

The elaboration of a SLR/MA protocol is the most crucial section of the article, and it must be developed before starting the study. The protocol acts as a protective barrier for arbitrary decisions while conducting the SLR/MA. As a general rule, the protocol should be described so that the reader can replicate it. The most used guide for developing SLR/MA protocols is the PRISMA-P; it acts as a checklist that includes all the relevant information that must be addressed in the protocol (Shamseer et al., 2015).

Currently, many authors have supported the protocol's public availability to guarantee transparency and avoid duplicated reviews; therefore, the University of York created PROSPERO, a web page to register all SLR/MA protocols.

Sideri et al. published a SLR that included previous SLR in the orthodontic field. They found that SLR/MA posted in PROSPERO were associated with an increased methodological quality (Sideri et al., 2018). It is important to highlight that the PROSPERO webpage provides several protocol templates to develop SLR/MA, which authors can use to facilitate their work (Shamseer et al., 2015).

The step-by-step process to develop a SLR/MA protocol is beyond the scope of this review; therefore, we strongly suggest readers turn to the PRISMA-P checklist and the protocol templates available on the PROSPERO webpage.

3. CARRY OUT THE LITERATURE SEARCH

If the researcher is willing to develop a search strategy, it is recommended first to look in PROSPERO if other SLR/MA have been published that require an update; if there are, the researcher can use this information to build a new search.

SLR/MA should include all the available literature, ideally without language restriction, including published and unpublished literature. The search strategy must be built based on the following points (Bramer et al., 2017; Sayers, 2008):

1. Database: The search must be conducted in multiple databases using MESH terms, keywords, and synonyms linked by Boolean connectors and prespecified inclusion and exclusion criteria. Based on our experience, the most widely used databases are MEDLINE, EMBASE, SCOPUS, Cochrane Library, Web of Science, LILACS (Latin-American literature), African Index Medicus (African literature).
2. Unpublished literature: It can be searched in ProQuest and OpenGrey.
3. Snowballing technique: It is based on searching the references of the included papers to identify additional citations that can be helpful for your research.
4. Hand searching: The researcher must identify one to four high-impact journals to manually search all

the volumes and issues to identify proper papers to include in the SLR/MA.

At least two independent authors must carry out the search strategy; both researchers must keep track of all the records while performing the search (Herrera Ortiz et al., 2021; Pautasso, 2013). For example, how many papers were retrieved in each database, the inclusion, and exclusion criteria, the years the search strategy covered, etc. The results retrieved by both authors must be compared, and if discrepancies are present, these must be resolved by agreement or by a third author (Pautasso, 2013). As a general rule, the reader must be able to replicate the search strategy reported in your paper and obtain the same results (Pautasso, 2013).

4. DUPLICATES REMOVAL:

Once the search is accomplished, researchers can face many scientific papers coming from different databases. Some of these articles may be duplicated; therefore, duplication removal is mandatory. Citation removal can be easily accomplished using the Mendeley duplication removal tool or EndNote (Herrera Ortiz et al., 2021; Kwon et al., 2015). To use it, the researchers must load all the articles obtained by the search strategy on Mendeley or EndNote and click on the check for duplicates button.

5. THE SCREENING FOR ELIGIBILITY

When all the duplicated articles have been removed, the screening for eligibility can start. The first screening phase should be carried out by title and abstract; the eligible papers must be read in full to guarantee that the inclusion and exclusion criteria are fulfilled (Shamseer et al., 2015). The researchers must specify in the results section the reasons for excluding the studies.

Ideally, the screening of the papers should be done systematically and be completed by two independent authors; if discrepancies are presented, these can be resolved by a third author (Jahan et al., 2016).

The entire search and selection process must be recorded in a PRISMA flow diagram obtained on the PRISMA webpage.

6. ASSESS THE METHODOLOGICAL QUALITY OF THE SELECTED PAPERS

Before introducing how to assess the risk of bias in the included studies, it is mandatory first to explain the main biases associated with a SLR/MA. The types of biases in a SLR/MA can be classified as 1. Biases related to the studies included, and 2. Biases associated with the SLR/MA methods (Metabiases) (Ayorinde AA, et al., 2020; Drucker et al., 2016; Jahan et al., 2016). Table 2 provides a breakdown of some of these biases. Readers must understand that this table shows the most important biases, but the information regarding this topic is much broader.

Table 2. Biases associated with a SLR (Ayorinde AA, et al., 2020; Drucker et al., 2016; Jahan et al., 2016).

Biases related to the SLR/MA methods (Metabiases)

1. **Time-lag bias:** is caused by delays in the publication process; therefore, the article may not be available for inclusion when the SLR/MA is conducted.

2. **Location bias:** is caused by the different degrees of journal indexations. Some articles are published in regional journals with none or few indexations, making it harder for the researchers to be aware of them.

3. **Language bias:** is caused when researchers search only studies in a particular language (for example, English), excluding papers written in other languages.

4. **Duplication bias:** Some articles appear duplicated on the web. If those papers are included twice in the systematic literature review/meta-analysis, it may lead to bias.

5. **Citation bias:** Studies with positive results are more likely to be cited. Therefore, using the snowballing technique will favor the location of studies favorable to that topic, which may bias the conclusions of the SLR/MA.

6. **Publication bias:** It is when only studies with positive results are published; therefore, the effects of the SLR/MA may be overestimated if unpublished literature (usually with negative results) is not included.

7. **Selective outcome reporting:** The author describes just the outcomes that favor the research question. For example, suppose a drug reveals mild and severe adverse reactions, but the author just reports the mild reactions and wholly ignores the severe ones to favor the intervention.

8. **Independent data extraction bias:** When only one author extracted the articles' data, which can generate subjectivity and lead to bias; therefore, the data extraction process must be carried out by at least two authors.

9. **Investigator experience bias:** During the literature search phase, an experienced researcher is prone to find more studies than an amateur one.

Biases related to the studies included (Clinical trials)

Selection bias: When the patients allocated in one group have specific characteristics that influence the outcome. It is solved by implementing randomization.

Performance bias: When health personnel provides better care to patients allocated to one group instead of the other (either the experimental or the control group). It can be solved by implementing blinding.

Attrition bias: When high dropouts are present during the study, and these dropouts are not included in the statistical analysis. This issue can be solved by performing an intention to treat analysis.

Detection bias: When the physician knows the patients' allocation; therefore, the data interpretation is not impartial. For example, when comparing treatment "A" against "placebo," if the researcher knows which patients are allocated to treatment "A" instead of "placebo," unconsciously, he will interpret the data in favor of treatment "A." This issue can be solved by implementing blinding.

It is essential to assess the risk of bias of the papers included in the SLR/MA. Several authors have described multiple methodological scales to evaluate the risk of bias of scientific papers (Arem Lim et al., 2003; Cashin, et al 2020; Ciapponi, 2015; Herrera Ortiz et al., 2021; Natalie A de Morton, 2009; Sterne et al., 2019). Nevertheless, none of these are universally accepted and must be assessed based on the reader's expertise (Pussegoda et al., 2017). Some of the most widely used scales are shown in Table 3 (Jørgensen et al., 2016; Ma et al., 2020):

Table 3. Some of the most widely used scales to evaluate the quality of included studies:

Scale Name	Characteristics	Advantages	Disadvantages
1. The methodological index for non-randomized studies (MINORS): (Arem Lim et al., 2003).	<p>MINORS scale has 12 items (the first eight are used in non-comparative studies, while all 12 items are used in comparative studies).</p> <p>Each item can be scored based on its quality:</p> <ul style="list-style-type: none"> • 0 (if not reported, has high risk of bias) • 1 (if reported but inadequate, has an unclear risk of bias) • 2 (reported and adequate, has a low risk of bias). 	It is used to evaluate the methodological quality of non-randomized clinical trials	MINORS scale does not generate a summary score; therefore, the result must be interpreted based on the researcher's experience
2. PEDro scale: (Cashin et al., 2020; A de Morton, 2009).	<p>The PEDro scale has 10 items; each can be scored as 1 (present) and 0 (absent).</p> <p>The score of each item can be added to generate a summary score. Results between:</p> <p>0-3 are considered poor 4-5 are considered fair 6-8 are considered good 9-10 are considered excellent</p>	It is a scale used to evaluate the methodological quality of randomized clinical trials.	Their results can not be used as a measurement of the validity of the study conclusions
3. QUADAS-2: (Ciapponi, 2015; Herrera Ortiz et al., 2021).	<p>The QUADAS-2 is divided into two domains: 1. The probability of bias (has 4 items) 2. The applicability concerns (has 3 items).</p> <p>The results in each item can be positive, negative, or unknown.</p> <p>As a general rule, if all the answers in a domain are positive, a low probability of bias is suggested; on the other hand, if a domain has at least one negative response, it may suggest bias within the study</p>	It is a scale developed to test the methodological quality of diagnostic studies	QUADAS-2 does not provide a summary score; therefore, the results must be interpreted based on the researcher's experience
4. Cochrane risk of bias tool: (Sterne et al., 2019).	The scale is constructed based on five domains (selection, performance, attrition, reporting, and others) with different items inside each domain.	It is a scale used to test the methodological quality of randomized controlled clinical trials	Cochrane risk of bias tool does not generate a summary score. Therefore, the degree of bias is judged based on the

7. HOW TO PERFORM THE DATA EXTRACTION AND QUALITATIVE SYNTHESIS

To perform the qualitative synthesis, at least two researchers need to extract the information of interest from all the articles included and save the data in a table (Herrera Ortiz et al., 2021; Soilemezi et al., 2018). Typically, a data extraction table contains the author, the year of publication, the type of study, the number of patients, the mean age of the participants, and other relevant information regarding the research question. The information that will be extracted and included in the table must be prespecified before running the SLR/MA (Xiao et al., 2019).

8. HOW TO PERFORM THE QUANTITATIVE SYNTHESIS

The quantitative synthesis must be performed if a MA is considered feasible. We assume that a MA is possible if there are at least 2 studies, the papers included provide enough information to calculate the effect measurement prespecified in the protocol, and all the studies measured the exact outcome you are interested in (we don't want to mix apples and oranges) (Borenstein et al., 2009; Haidich, 2010).

If the researcher is interested in a categorical outcome, the effect must be measured using: Relative risk (RR), odds ratio (OR), Risk difference, Diagnostic OR, or number needed to treat (De Sousa et al., 2009; Papageorgiou et al., 2015). If the investigator wants to assess a numerical outcome, the effect must be measured using: Mean difference, Cohen's d, or Hedge's g, among others (Andrade, 2020). Finally, if the researchers want to evaluate a time-to-event outcome, they must measure the effect using hazard ratio (Escrig-Sos, 2005).

This paper will explain how to conduct the quantitative synthesis for a MA in STATA 17 for categorical outcomes. We have chosen to present categorical outcomes because it will be easier for new researchers to understand its biostatistical concepts and perform them in STATA 17. The explanation for numerical outcomes is beyond the scope of this review and would require us to write another paper.

For this explanation, suppose we perform a MA of 16 studies to determine if drug "X" is better than drug "Y" to stop cigarette consumption.

The first of all is to know how to structure the database (Table 4):

Table 4. Database structure to perform a MA for categorical outcomes.

Studies	year	a	b	c	d
Rossen et al	1990	1	103	1	105
Mort et al	1972	5	96	4	86
Duncan et al	1990	15	277	5	207
Stain et al	1982	11	493	4	103
Portz et al	1986	8	106	5	63
Slam et al	1995	42	2157	5	924
Moreno et al	1994	20	134	7	102
Futt et al	1979	34	997	8	1099
Willman et al	1982	21	85	11	94
Manson et al	1985	12	73	11	67
Janzz et al	1987	20	116	12	94
Mora et al	1990	43	534	17	515
Herrero et al	1990	34	203	20	214
Madarin et al	1995	53	415	35	454
Futt et al	1983	43	718	35	624
jean et al	1984	77	435	50	491

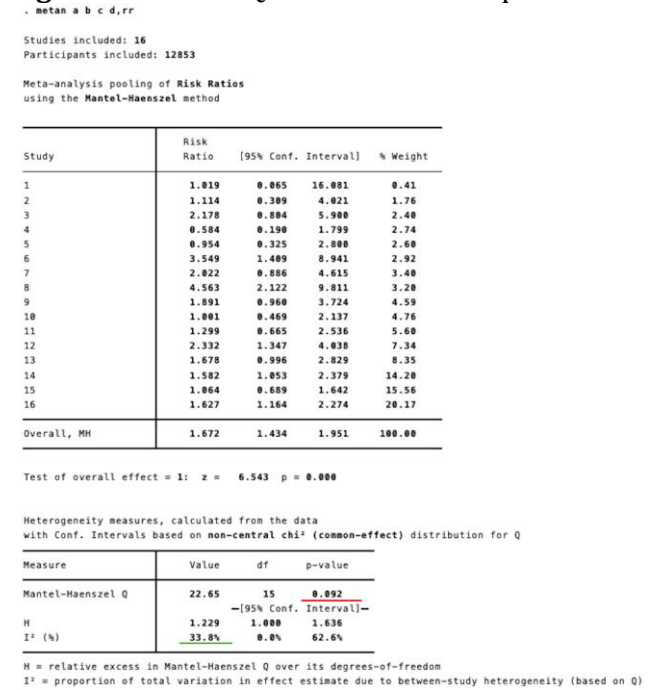
The database shows the studies, the year of publication, a (represent the number of patients that received treatment X and stopped smoking), b (represent the number of patients that received treatment X and continue smoking), c (represent the number of patients that received treatment Y and stopped smoking), d (represent the number of patients that received treatment Y and continue smoking).

Once the database has been created, it must be imported to STATA 17. Then researchers need to install the following complements: “ssc install metan”, and “ssc install metaninf”. Now the investigators must proceed to determine the heterogeneity of the data, which represents the variation between studies in a MA regarding sample size, sample characteristics, outcomes, study design, etc (Melsen et al., 2014).

The importance of determining heterogeneity in a MA rises in the fact that depending on the degree of heterogeneity; the researchers will choose different statistical methods to analyze the data (either fixed or random effect model). According to several authors, the heterogeneity of the studies can be assessed through the I2 value and the Cochran’s Q test (Borenstein et al., 2010; Melsen et al., 2014). If Cochran’s Q test is < 0.05, we consider that the MA presents high heterogeneity; if it is > 0.05, we consider the data homogeneous (Zeng et al., 2021). Moreover, we can measure the degree of heterogeneity with the I2 value; if it is < 50%, we assume a low heterogeneity, and a fixed-effect model must be used. If I2 square is >50%, we consider a moderate-high heterogeneity, and a random effect model must be employed (Melsen et al., 2014). The disadvantage of using a fixed-effect model is that it tends to give too much weight to the studies with a huge sample size, leading to a lack of robustness if one of the big studies is removed. In contrast, the random effect model tends to redistribute the weights of the studies almost proportionally.

The I2 value and the Cochran’s Q test can be calculated in STATA 17 with the following command: metan “RECEIVED DRUG X AND STOP SMOKING” “RECEIVED DRUG X AND CONTINUE SMOKING” “RECEIVED DRUG Y AND STOP SMOKING” “RECEIVED DRUG Y AND CONTINUE SMOKING”, rr (Figure 1). The words in capital letters must be replaced by the variables inputted in STATA 17 (a b c d); the effect measurement can be replaced if desired by OR, RR, etc. In this case, we decided to use RR. The software will automatically calculate the effect size selected and its 95% confidence interval when this command is applied.

Figure 1. Cochran’s Q test and I2 value output.



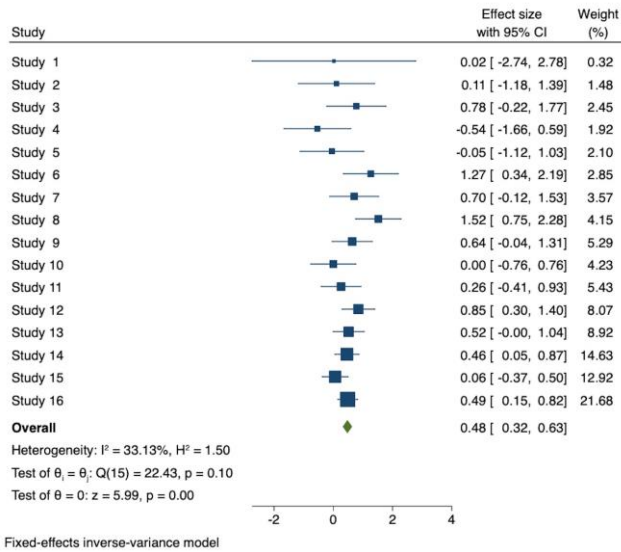
The Cochran’s Q test reported a value of 0.092 (red underline); therefore, we can assume homogeneity in our data. Moreover, the I2 value is 33.8% (green underline), for which a fixed-effect model must be used.

The next step would be to set up our data to operate as MA data; it can be achieved using the following command: meta set _ES _seES, fixed. In this case, we decided to use a fixed-effect model because the Cochran Q test and the I2 value revealed homogeneity in the data. Nevertheless, if the authors would like to use a random effect model, they just need to replace "fixed" with "random" in the command.

Now it is time to generate the forest plot of our MA to weigh all the values retrieved from the studies, to obtain a unique measurement. The forest plot can be generated with the following command: meta forestplot, fixed (Figure 2). This forest plot will represent the estimates as log RR instead of RR.

If a random effect model is meant to be used, the authors just need to replace the word "fixed" with "random."

Figure 2. Forest plot output.



Forest plot expressed in the form of log RR. The way forest plots are interpreted is by assuming that each square symbolizes the sample size of each study, the bigger the sample, the bigger the square. The horizontal lines that delimit each square are the confidences intervals. If the confidence interval of a study passes through zero, its results are not statistically significant. If the confidence interval is below zero, it means that the intervention has protective effects, and if it is above zero, it suggests that the intervention is a risk factor. The diamond in green represents the overall effect of all studies, showing an log RR value of 0.48 (0.32 - 0.63), suggesting that drug X has protective effects to stop smoking compared to drug Y. Another interpretation could be done transforming the log RR to RR exponentiating the values.

9. ASSESSING PUBLICATION BIAS

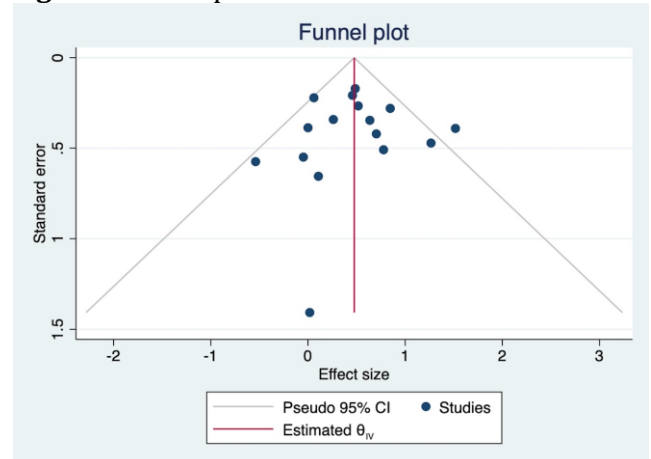
There are different ways to assess publication bias. The researcher must choose one method based on the number of studies included in the MA.

Funnel plot: It must be employed when > 10 studies are included in the quantitative synthesis (Figure 3) (Sterne et al., 2011). The effect size is shown on the x-axis, while the standard error is shown on the y axis (The smaller the standard error, the higher the sample size the study has). A funnel plot without publication bias must have a pyramidal appearance suggesting that the researchers include studies with small and big sample sizes and positive and negative outcomes (Sterne et al., 2011). The command to generate the funnel plot is: meta funnelplot.

Begg Test: It is a nonparametric test that only reaches enough statistical power to detect publication bias when the researchers include > 25 studies (Gjerdevik et

al., 2014). The command to calculate the Begg test is: meta bias, begg (Figure 4).

Figure 3. Funnel plot



This funnel plot shows studies with high sample sizes and positive and negative effects; studies with small sample sizes are missing. We would have to use another test to objectively quantify if the results are enough to be considered publication bias.

Figure 4. Begg test output

. meta bias, begg

```
Effect-size label: Effect size
Effect size: _ES
Std. err.: _seES
```

Begg's test for small-study effects

```
Kendall's score = 0.00
SE of score = 22.211
z = -0.05
Prob > |z| = 1.0000
```

The Begg test shows a non-statistically significant p-value (1.000). We can think that there is no publication bias; nevertheless, the Begg test does not have enough power to detect publication bias when the MA includes < 25 studies; therefore, the Egger test must be carried out.

Egger test: It has more statistical power than the Begg test and is generally used to detect publication bias when the MA includes between 10 - 25 studies (Sterne et al., 2011). If there are < 10 studies, the assessment of publication bias must be interpreted with caution because they can lead to errors.

If the regression intercept α of the egger test is highly different from 0, or if the P-value is < 0.05,

the test suggests publication bias (Irwig et al., 1997; Lifeng Lin et al., 2016). The command to calculate the Egger test is: meta bias, egger (Figure 5).

Figure 5. Egger test output

```
. meta bias, egger

Effect-size label: Effect size
Effect size: _ES
Std. err.: _seES

Regression-based Egger test for small-study effects
Fixed-effects model
Method: Inverse-variance

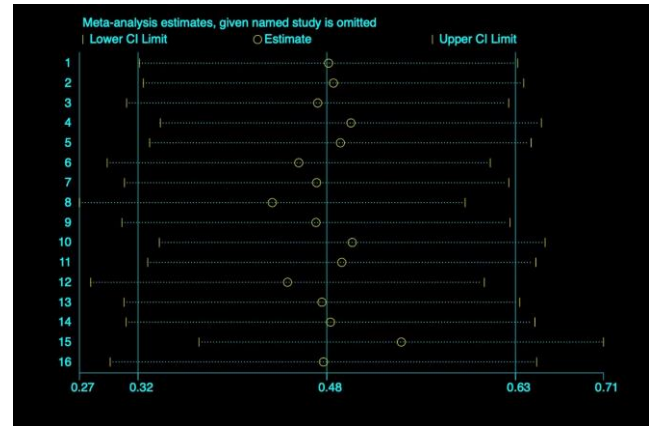
H0: beta1 = 0; no small-study effects
      beta1 =    0.02
SE of beta1 =    0.602
          z =    0.03
Prob > |z| =    0.9793
```

The Egger test shows a regression intercept α (displayed as beta1) of 0.02, which is pretty similar to 0; therefore, we can assume no publication bias in this MA. Also, if we check the P-value, we can see a non-statistically significant result ($P= 0.9793$), reinforcing the assumption that there is no publication bias.

10. HOW TO PERFORM A SENSITIVITY ANALYSIS

Sensitivity analysis is carried out to assess if the results of the MA are robust (Spinelì et al., 2021). It is performed by repeating the MA as many times as there are studies, always excluding one article to see if the overall result gives similar outcomes. If the overall effect keeps constant, we can assume that the results are robust (Spinelì et al., 2021). The command to generate a sensitivity analysis is: metaninf _ES _seES, fixed (Figure 6). If the authors would like to calculate a sensitivity analysis for a random effect model, they just need to replace the word "fixed" with "random" in the command.

Figure 6. Sensitivity analysis output.



The yellow circle in the sensitivity analysis plot represents the RR value, and the horizontal line indicates the confidence interval. The way to interpret these results is by assuming that we will obtain the results shown in the first point if we remove the first study. If we remove the second study, we will get the results obtained in the second point, and so on. As seen when the sensitivity analysis is performed, the results showed an RR value with a confidence interval below 1 in all the points. Therefore, we can say that the MA has robust data to conclude that drug X is better than drug Y to stop smoking.

11. PUBLISH THE SYSTEMATIC LITERATURE REVIEW/META-ANALYSIS

Andersen et al. published a SLR that included previous SLR/MA registered in PROSPERO. They showed that the median time from the development of the SLR/MA protocol until the publication of the final article was of 16 months (Andersen et al., 2021). However, if it was a Cochrane review, the median time widened to 24 months (Andersen et al., 2020). Therefore, there is an increased risk of a MA being outdated when it is finally published (Andersen et al., 2020).

12. UPDATE THE SYSTEMATIC LITERATURE REVIEW/META-ANALYSIS

Certain entities such as the Cochrane collaboration established the necessity to update SLR/MA every 2 years (Kaveh G et al., 2007). Nevertheless, some authors disagree with this opinion, arguing that the requirement of an update of a SLR/MA depends on the amount of published material per year; therefore, the necessity may vary according to topics and innovation (Garner et al., 2016).

CONCLUSIONS

SLR/MA are valuable tools to keep physicians updated, evaluate the methodological quality of the published evidence, and detect and solve

knowledge gaps. The present manuscript addresses the methodology of a SLR/MA in a step-by-step fashion, starting from the research question to the final publication and update of the MA. By this means, we don't tell that MA are infallible tools to solve a research question; there are cases in the literature in which large randomized clinical trials have refused the results of MA (LeLorier et al., 1997). The main limitation of this article is its design since it is a narrative review, and therefore, the information included has been subjectively selected by the authors. However, the information presented in this review is highly generalizable and replicable. The authors hope that the information presented in this manuscript serves as pillars to enhance research worldwide, especially in places where formal research training is uncommon. If readers want to extend their knowledge regarding this subject, we recommend reading the "The Cochrane Handbook for Systematic Reviews of Interventions and the Center for Reviews and Dissemination" or the book "Finding What Works in Health Care: Standards for systematic reviews."

REFERENCES

- Andersen, M. Z., Fønnes, S., Andresen, K., & Rosenberg, J. (2021). Most published meta-analyses were made available within two years of protocol registration. *European Journal of Integrative Medicine*, 44(1), 101342. <https://doi.org/10.1016/j.eujim.2021.101342>
- Andersen, M. Z., Gülen, S., Fønnes, S., Andresen, K., & Rosenberg, J. (2020). Half of Cochrane reviews were published more than 2 years after the protocol. *Journal of Clinical Epidemiology*, 124(1), 85–93. <https://doi.org/10.1016/j.jclinepi.2020.05.011>
- Andrade, C. (2020). Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis: As Simple as It Gets. *The Journal of Clinical Psychiatry*, 81(5), 1–4. <https://doi.org/10.4088/JCP.20f13681>
- Arem Lim, K. S., Mile Ini, E. N., Amien Orestier, D. F., Abrice Wiatkowski, F. K., Ves Anis, Y. P., Acques Hipponi, J. C., Slim, K., Nini, E., Forestier, D., Kwiatkowski, F., Panis, Y., & Chipponi, J. (2003). Methodological index for non-randomized studies (minors): Development and validation of a new instrument. *ANZ J. Surg.* 73(1), 712–716. <https://doi.org/10.1046/j.1445-2197.2003.02748.x>
- Ayorinde AA, Williams I, Mannion R, Song F, Skrybant M, Lilford RJ, C. Y. (2020). Publication and related biases in health services research: a systematic review of empirical evidence. *BMC Med Res Methodol*, 20(1), 137–149. <https://doi.org/10.1186/s12874-020-01010-1>
- Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H. R. (2009). When Does it Make Sense to Perform a Meta-Analysis? In *Introduction to Meta-Analysis* (pp. 357–364). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470743386.ch40>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews*, 6(1), 1–12. <https://doi.org/10.1186/s13643-017-0644-y>
- Cashin, A. G., & McAuley, J. H. (2020). Clinimetrics: Physiotherapy Evidence Database (PEDro) Scale. *Journal of Physiotherapy*, 66(1), 59–59. <https://doi.org/10.1016/j.jphys.2019.08.005>
- Ciapponi, A. (2015). QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica QUADAS-2: an instrument for the evaluation of the quality of diagnostic precision studies. *Evid Act Pract Ambul*, 18(1), 22–30. <https://doi.org/10.51987/evidencia.v18i1.6341>
- De Sousa, M. R., & Ribeiro, A. L. P. (2009). Systematic review and meta-analysis of diagnostic and prognostic studies: A tutorial. *Arquivos Brasileiros de Cardiologia*, 92(3), 235–245. <https://doi.org/10.1590/S0066-782X2009000300013>
- Drucker, A. M., Fleming, P., & Chan, A. W. (2016). Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. *Journal of Investigative Dermatology*, 136(11), 109–114. <https://doi.org/10.1016/j.jid.2016.08.021>
- Escrig-Sos, J. (2005). Sobre cómo analizar la credibilidad de un ensayo clínico o metaanálisis cuyo resultado principal se ofrezca en odds ratio, riesgo relativo o hazard ratio. *Cirugia Espanola*, 78(6), 351–356. [https://doi.org/10.1016/S0009-739X\(05\)70953-2](https://doi.org/10.1016/S0009-739X(05)70953-2)
- Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Akl, E. A., Beyene, J., Chang, S., Churchill, R., Dearness, K., Guyatt, G., Lefebvre, C., Liles, B., Marshall, R., Mart\`inez Garc\`ia, L., Mavergames, C., Nasser, M., Qaseem, A., Sampson, M., Soares-Weiser, K., ... Schünemann, H. J. (2016). When and how to update systematic reviews: consensus and checklist. *BMJ*, 354(1), 1–10. <https://doi.org/10.1136/bmj.i3507>
- Gjerdevik, M., & Heuch, I. (2014). Improving the error rates of the Begg and Mazumdar test for publication bias in fixed effects meta-analysis. *BMC Medical Research Methodology*, 14(1), 1–16. <https://doi.org/10.1186/1471-2288-14-109>
- Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia*, 14(1), 29–37. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/>
- Herrera Ortiz, A. F., Fernández Beaujon, L. J., García Villamizar, S. Y., & López Fonseca, F. F. (2021). Magnetic resonance versus computed tomography for the detection of retroperitoneal lymph node metastasis due to testicular cancer : A systematic literature review. *European Journal of Radiology Open*, 8(1), 1–7. <https://doi.org/10.1016/j.ejro.2021.100372>
- Herrera Ortiz, A. F., Rincón Cuenca, N. T., & Fernández Beaujon, L. (2021). Brain Changes in Magnetic Resonance Imaging Caused by Child Abuse a Systematic Literature Review. *Revista Cuarzo*, 27(1), 27–33. <https://doi.org/10.26752/cuarzo.v27.n1.522>
- Impellizzeri, F. M., & Mario Bizzini. (2012). Systematic review and meta-analysis: a primer. *International Journal of Sports Physical Therapy*, 7(5), 493–503. <https://pubmed.ncbi.nlm.nih.gov/23091781/>
- Irwig, L., Macaskill, P., Berry, G., & Glasziou, P. (1997). Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. *BMJ (Clinical Research Ed.)*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Jahan, N., Naveed, S., Zeshan, M., & Tahir, M. A. (2016). How to Conduct a Systematic Review: A Narrative Literature Review. *Cureus*, 8(11), 864–870. <https://doi.org/10.7759/cureus.864>
- Jørgensen, L., Paludan-Müller, A. S., Laursen, D. R. T., Savović, J., Boutron, I., Sterne, J. A. C., Higgins, J. P. T., & Hróbjartsson, A. (2016). 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. *Systematic Reviews*, 5(1), 80–93. <https://doi.org/10.1186/s13643-016-0259-8>
- Kaveh G. Shojania, Margaret Sampson, Mohammed T. Ansari, et al. (2007). How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. *Annals of Internal Medicine*, 147(4), 224–233. <https://doi.org/10.7326/0003-4819-147-4-200708210-00179>

- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five Steps to Conducting a Systematic Review. *Journal of the Royal Society of Medicine*, 96(3), 118–121. <https://doi.org/10.1177/014107680309600304>
- Kwon, Y., Lemieux, M., McTavish, J., & Wathen, N. (2015). Identifying and removing duplicate records from systematic review searches. *Journal of the Medical Library Association*, 103(4), 184–188. <https://doi.org/10.3163/1536-5050.103.4.004>
- Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction. *New England Journal of Medicine*, 327(4), 248–254. <https://doi.org/10.1056/NEJM199207233270406>
- LeLorier, J., Grégoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *The New England Journal of Medicine*, 337(8), 536–542. <https://doi.org/10.1056/NEJM199708213370806>
- Lifeng Lin, & Chu, H. (2016). Quantifying Publication Bias in Meta-Analysis. *Physiology & Behavior*, 176(1), 139–148. <https://doi.org/10.1111/biom.12817>
- Ma, L.-L., Wang, Y.-Y., Yang, Z.-H., Huang, D., Weng, H., & Zeng, X.-T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Military Medical Research*, 7(1), 7–18. <https://doi.org/10.1186/s40779-020-00238-8>
- Manterola, C., Astudillo, P., Arias, E., & Claros, N. (2013). Revisión sistemática de la literatura. Qué se debe saber acerca de ellas. *Cirugía Española*, 91(3), 149–155. <https://doi.org/10.1016/j.ciresp.2011.07.009>
- Melsen, W. G., Bootsma, M. C. J., Rovers, M. M., & Bonten, M. J. M. (2014). The effects of clinical and statistical heterogeneity on the predictive values of results from meta-analyses. *Clinical Microbiology and Infection*, 20(2), 123–129. <https://doi.org/10.1111/1469-0691.12494>
- Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*, 18(5), 1–9. <https://doi.org/10.1186/s12874-017-0468-4>
- Natalie A de Morton. (2009). The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. *Australian Journal of Physiotherapy*, 55(2), 129–133. [https://doi.org/10.1016/S0004-9514\(09\)70043-1](https://doi.org/10.1016/S0004-9514(09)70043-1)
- Papageorgiou, S. N., Tsiranidou, E., Antonoglou, G. N., Deschner, J., & Jäger, A. (2015). Choice of effect measure for meta-analyses of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects. *Journal of Clinical Epidemiology*, 68(5), 534–541. <https://doi.org/10.1016/j.jclinepi.2015.01.004>
- Pautasso, M. (2013). Ten Simple Rules for Writing a Literature Review. *PLOS Computational Biology*, 9(7), 1–4. <https://doi.org/10.1371/journal.pcbi.1003149>
- Pussegoda, K., Turner, L., Garrity, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L. M., Hróbjartsson, A., Altman, D. G., & Moher, D. (2017). Identifying approaches for assessing methodological and reporting quality of systematic reviews: a descriptive study. *Systematic Reviews*, 6(1), 117–129. <https://doi.org/10.1186/s13643-017-0507-6>
- Sargeant, J. M. ., & O'Connor, A. M. (2020). Scoping Reviews, Systematic Reviews, and Meta-Analysis: Applications in Veterinary Medicine. *Frontiers in Veterinary Science*, 7(11), 1–14. <https://doi.org/10.3389/fvets.2020.00011>
- Sayers, A. (2008). Tips and tricks in performing a systemic review. *British Journal of General Practice*, 58(538), 425–425. <https://doi.org/10.3399/bjgp08X277168>
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Altman, D. G., Booth, A., Chan, A. W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Gøtzsche, P. C., Grimshaw, J. M., Groves, T., Helfand, M., ... Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ (Online)*, 349(1), 1–25. <https://doi.org/10.1136/bmj.g7647>
- Sideri, S., Papageorgiou, S. N., & Eliades, T. (2018). Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *Journal of Clinical Epidemiology*, 100(1), 103–110. <https://doi.org/10.1016/j.jclinepi.2018.01.003>
- Soilemezi, D., & Linceviciute, S. (2018). Synthesizing Qualitative Research: Reflections and Lessons Learnt by Two New Reviewers. *International Journal of Qualitative Methods*, 17(1), 1–14. <https://doi.org/10.1177/1609406918768014>
- Spineli, L. M., Kalyvas, C., & Papadimitropoulou, K. (2021). Quantifying the robustness of primary analysis results: A case study on missing outcome data in pairwise and network meta-analysis. *Research Synthesis Methods*, 12(4), 475–490. <https://doi.org/10.1002/jrsm.1478>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *The BMJ*, 366(1), 1–8. <https://doi.org/10.1136/bmj.l4898>
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Online)*, 343(1), 1–8. <https://doi.org/10.1136/bmj.d4002>
- Wallace, D. K. (2010). Evidence-based medicine and levels of evidence. *The American Orthoptic Journal*, 60(1), 2–5. <https://doi.org/10.3368/aoj.60.1.2>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Zeng, Q. C., Sun, Q., Zhang, M., Tang, Y., & Long, H. C. (2021). Relation between IL-8 level and obstructive sleep apnea syndrome. *Open Medicine (Poland)*, 16(1), 683–691. <https://doi.org/10.1515/med-2020-0240>