

HOW TO PERFORM A META-ANALYSIS: A PRACTICAL STEP-BY-STEP GUIDE USING R SOFTWARE AND RSTUDIO

COMO REALIZAR UMA METANÁLISE: UM GUIA PRÁTICO PASSO A PASSO UTILIZANDO O SOFTWARE R E O RSTUDIO

DIEGO ARIEL DE LIMA¹ , CAMILO PARTEZANI HELITO² , LANA LACERDA DE LIMA¹ , RENATA CLAZZER³ ,
ROMEU KRAUSE GONÇALVES⁴ , OLAVO PIRES DE CAMARGO² 

1. Universidade Federal Rural do Semi-Árido, Faculty of Medicine, Mossoró, RN, Brazil.

2. Universidade de São Paulo, Faculty of Medicine, Hospital das Clínicas, Institute of Orthopedics and Traumatology, São Paulo, SP, Brazil.

3. Universidade Estadual do Rio Grande do Norte, Faculty of Medicine, Mossoró, RN, Brazil.

4. Instituto de Traumatologia e Ortopedia Romeu Krause, Knee studies, Recife, PE, Brazil.

ABSTRACT

Meta-analysis is an adequate statistical technique to combine results from different studies, and its use has been growing in the medical field. Thus, not only knowing how to interpret meta-analysis, but also knowing how to perform one, is fundamental today. Therefore, the objective of this article is to present the basic concepts and serve as a guide for conducting a meta-analysis using R and RStudio software. For this, the reader has access to the basic commands in the R and RStudio software, necessary for conducting a meta-analysis. The advantage of R is that it is a free software. For a better understanding of the commands, two examples were presented in a practical way, in addition to revising some basic concepts of this statistical technique. It is assumed that the data necessary for the meta-analysis has already been collected, that is, the description of methodologies for systematic review is not a discussed subject. Finally, it is worth remembering that there are many other techniques used in meta-analyses that were not addressed in this work. However, with the two examples used, the article already enables the reader to proceed with good and robust meta-analyses. **Level of Evidence V, Expert Opinion.**

Keywords: Meta-Analysis. Guideline. Software.

RESUMO

Metanálise é uma técnica estatística adequada para combinar resultados provenientes de diferentes estudos, seu uso vem crescendo e ganhando cada vez mais importância no meio médico. Assim, não apenas saber interpretar metanálise, como também saber realizar uma, mesmo que simples, é fundamental na atualidade. Portanto, o objetivo principal deste artigo é apresentar os conceitos básicos que a norteiam e servir de guia para a condução de uma metanálise utilizando os softwares R e RStudio. Para isso, através do presente artigo o leitor tem acesso aos comandos básicos existentes nos softwares R e RStudio, necessários para a condução de uma metanálise. A grande vantagem do R é o fato de ser um software livre. Para um melhor entendimento dos comandos, dois exemplos foram apresentados de forma prática, além de revisados alguns conceitos básicos dessa técnica estatística. É suposto que os dados necessários para a metanálise já foram coletados, ou seja, descrição de metodologias para revisão sistemática não é assunto discutido. Por fim, vale lembrar que existem muitas outras técnicas utilizadas em metanálises que não foram abordadas neste trabalho. Todavia, com os dois exemplos utilizados, o artigo já habilita o leitor a proceder boas e robustas metanálises. **Nível de Evidência V, Opinião do Especialista.**

Descritores: Metanálise. Guia. Software.

Citation: Ariel de Lima D, Helito CP, Lima LL, Clazzer R, Gonçalves RK, Camargo OP. How to perform a meta-analysis: a practical step-by-step guide using R software and RStudio. Acta Ortop Bras. [online]. 2022;30(3): Page 1 of 9. Available from URL: <http://www.scielo.br/aob>.

INTRODUCTION

Scientific research has been growing in all areas of knowledge, and in medicine it is no different. The same theme may be researched in several medical centers around the world. With the expansion of evidence-based medicine, the more studies on the same topic, the better the medical practices related to it.¹ However, the existence of many studies on the same subject may limit the access of medical professionals to all of them, either due

to the time or fees. Studies that aggregate the results of two or more studies on the same issue, in addition to facilitating and gathering evidence, would reduce the individual errors (biases) of each study, producing a powerful synthesis on a specific topic. The tool to achieve this is meta-analysis.²

Meta-analysis uses statistical methods to summarize the results of independent studies. By combining information from all relevant studies on the same topic, a meta-analysis can estimate

All authors declare no potential conflict of interest related to this article.

The study was conducted at Faculty of Medicine, Universidade Federal Rural do Semi-Árido (UFERSA).

Correspondence: Diego Ariel de Lima. Av. João da Escóssia, 1300, Mossoró, RN, Brazil, 59607330. arieldelima.diego@gmail.com

Article received on 02/15/2021, approved on 07/02/2021.



the effects of a given intervention more accurately than each study individually.³

In 1904, by arguing that studies on the preventive effect of inoculations against enteric fever were too small to allow a reliable conclusion (making the error size too great and the power of the studies too low), Karl Pearson, through correlations, combined the data from five studies, thus creating the first known meta-analysis.⁴ But it was only in the 1970s that the term meta-analysis was first used, becoming increasingly popular since then.⁵

Therefore, the main objective of this article is to present the basic concepts that guide a meta-analysis and to serve as a guide for conducting a meta-analysis using the R and RStudio software.

The data of a meta-analysis

For studies to be combined through a meta-analysis, it is necessary to define which results will be combined. We shall work with 2 examples:

As example 1, two surgical techniques seek to improve knee stability, technique A (experimental) and technique B (control). Let us say that there is a test to ascertain the stability of the knee (X test) and that if the test is positive, it means that the knee is unstable, similar to the pivot-shift test.⁶ Assuming that three authors decided to compare the two techniques (A and B), using the stability test X before and after surgery in both techniques (Table 1).

Table 1. Number of patients with positive X test before and after surgery of techniques A and B.

Author	PREOP – Number of patients subjected to technique A with positive X test	POSTOP – Number of patients subjected to technique A with positive X test	PREOP – Number of patients subjected to technique B with positive X test	POSTOP – Number of patients subjected to technique B with positive X test
1	18	8	21	18
2	30	10	60	31
3	42	12	45	20

In this example 1, we will work with discrete quantitative variables, which assume only values belonging to an enumerable set, which can assume only a countable finite or infinite number of values. Discrete variables are usually the result of counts. Examples: number of children, number of bacteria per milliliter of urine and number of cigarettes smoked per day.⁷

As example 2, we will work with continuous quantitative variables, which assume any value in a certain range of variation, for which fractional values make sense. They should usually be measured by means of some instrument. Examples: weight (scale), height (ruler), time (clock), blood pressure and age. Continuous variables are usually expressed in the form of an average of values followed by a measure of dispersion, typically the standard deviation.⁷

In example 2, we consider that there is a functional score w, such as the IKDC⁸ in which the higher the score, the better the result and which would serve to evaluate the post-operative clinical outcome of a given surgical technique. Assuming that three authors decided to compare two techniques, A (experimental) and B (control), using the functional score W in the postoperative period of the two techniques (Table 2).

The basics of a meta-analysis

In a meta-analysis, the results of two or more independent studies are combined. The results of medical studies can be demonstrated in numerous ways. The two most common are the results expressed by measure of association and the results expressed by mean difference.

Table 2. Result of the postoperative w score of techniques A and B.

Author	Technique A – number of participants	Technique A – post op w score (Mean)	Technique A – w score (Standard deviation)	Technique B – number of participants	Technique B – post op w score (Mean)	Technique B – w score (Standard deviation)
1	18	96.30	1.80	30	90.30	3.73
2	30	86.90	9.30	60	84.30	9.80
3	42	79.20	18.80	45	76.70	17.20

Measures of association were developed with the objective of evaluating the relationship between a risk factor and its outcome. Among these measures we highlight the Relative Risk (RR) and the odds ratio (OR).⁷ RR and OR estimate the magnitude of the association between exposure to the risk factor and the outcome, indicating how many times the occurrence of the outcome in the exposed is greater than that among the unexposed.

For example, the result of a hypothetical study showed that smokers (exposed to the risk factor: cigarette) have a 5 times greater chance (RR), that is, 400% more, of progressing to lung cancer than non-smokers (unexposed).

When there is no difference between the exposed and unexposed, we say that the RR is equal to 1. When exposure to a factor increases the chances of an event occurring, as in the example above of smokers, the RR is greater than 1. When exposure to a factor decreases the chances of an event occurring, the RR is less than 1 (however, it is not negative, that is, it varies from 0 to < 1).⁹

Simply put, if we have a $RR > 1$, the RR expresses how many times the exposure can lead to the outcome. In the smokers' example above, the RR is equal to 5. When the RR is less than 1, the relative risk reduction (RRR), also known as efficacy, can also be calculated using the following formula: $RRR \text{ or Efficacy} = (1 - RR) \times 100$. If in a study the RR of 0.27 is found as a result, we can say that in this study the exposure to a factor decreased 73% the risk of an event occurring $(1 - 0.27) \times 100 = 73\%$.⁹

Another way to express the results of a survey is through the mean difference (MD). In some studies, the outcome is measured through score scales such as IKDC.⁸ These scales produce numerical scores for each patient, rather than dichotomous "yes/no" results. As we have seen above, this type of variable is called continuous, and it is common to calculate its mean in the two groups to be compared. In our example 2, to evaluate the best result technique (highest w score), A or B, it is necessary to compare the means of the w scores of the two groups throughout the study. One of the problems of this type of outcome measured by continuous variable is that, although it is possible to affirm that patients who used the A technique had a higher score in the w score, it is difficult to extract a clinical meaning from this difference. It is easier to understand a 25% increase in the return to sport using technique A than a difference of 6 points on a functional scale/score. When there is no difference between the averages of the groups, we say that the MD is equal to 0.

After obtaining the results of the studies chosen to compose the meta-analysis, the measures are aggregated based on the weighting of the results of all individual studies. This weighting is given by the sample size (number of patients) of each study, culminating in the measure of general association: the result of our meta-analysis.^{7,10} It is worth remembering that in a meta-analysis, only equal association measures should be compared: RR with RR or OR with OR. It is not possible to compare RR of one study with MD of another study.^{7,10}

Confidence interval and p-value

When performing a clinical study, it is unlikely that the actual magnitude is exactly that found in the study. This happens due

to the natural occurrence of random variations inherent to the researcher and/or the research situation. That is, the relative risk value found may be, and typically is, greater or lesser than the true value. For this reason, it is essential to measure the statistical accuracy of the data, which will allow the reader to perceive the confidence of the data presented.⁷

The confidence interval is a range of possible values for the actual magnitude of the effect. In clinical biomedical studies, the minimum accepted confidence interval is 95%, typically expressed as 95% CI. That is, a study with 95% CI means that if we take a random sample and build 100 confidence intervals, 95 would contain the real parameter.¹¹ In terms of accuracy, the narrower the confidence interval, the greater the accuracy of the results. Among the factors that can increase the accuracy of the confidence interval, the sample size is inserted, that is, the larger the sample, the greater the accuracy.¹²

The confidence intervals present information similar to those derived from the p-value (statistical significance). If the relative risk value 1 (equal effects of the intervention and control group) is present between the lower and upper limit of the confidence interval, then the p-value will be greater than or equal to 0.05 (statistically non-significant difference). However, if the relative risk value 1 is not within the confidence interval interpolated by the lower and upper limits, then the p-value will be less than 0.05 (statistically significant difference).

Fixed-effects models and random-effects models

In meta-analysis there are basically two types of models that can be adopted: the fixed effects model and the random effects model.² The fixed-effect model assumes that the effect of interest is the same in all studies and that the differences observed between them are due only to sampling errors, the so-called variability within the studies. In a simplified way, it is as if the methods with fixed effects considered that the variability between the studies occur only by chance and ignored the heterogeneity between them.³

Random effect models assume that the effect of interest is not the same in all studies. In this sense, they consider that the studies that are part of the meta-analysis form a random sample of a hypothetical population of studies. However, although the effects of the studies are not considered equal, they are connected through a probability distribution, usually supposed to be normal. For this reason, they create combined results with a greater confidence interval (but less precision), and thus are the most recommended models. Despite having this advantage, methods with random effects are criticized for attributing greater weight to smaller studies.³

There is no formal rule for choosing the model. Generally, when there is no important diversity or heterogeneity, studies with greater statistical power (greater population and greater intervention effect) have more "weight." In this case, the fixed-effects model is used, which assumes that all studies showed the same effect: for example, when the objective is to estimate a treatment effect for a specific population, not extrapolating this effect to other populations.¹³ When there is diversity and heterogeneity among the studies, it is more recommended to use the random effects model, which distributes weight in a more uniform way, valuing the contribution of small studies. For example, when the researcher combines several studies that have the same objective, but that were not conducted in the same way. In this case, it is possible to extrapolate the effects to other populations, which makes for a more comprehensive analysis.¹³

Heterogeneity

In a meta-analysis, usually preceded by a systematic review, however similar the selected studies may seem, they are not considered identical as to the effect of interest. For example,

in a meta-analysis of studies in which the efficacy of a new surgical procedure is being tested, there may be a difference in the selected groups: one group may be healthier in one study than in another, the age group of patients may vary from study to study, among other factors that may influence the effect of treatment.

When this difference between groups happens, that is, when the variability between the studies is not just random, we say that the studies are heterogeneous. In the presence of heterogeneity, other meta-analysis techniques (such as subgroups and meta-regression) can be considered to explain the variability between groups. However, these types of analysis require a large number of studies. When it is not possible to count on so many studies, the random effects model is recommended, as seen in the topics above.¹⁴

Thus, it is clear that in choosing between the fixed effects model and the random effects model, the evaluation of heterogeneity plays an important role in this choice. The most used ways to verify the existence of heterogeneity in meta-analyses are by Cochran's Q test and Higgins and Thompson's I^2 statistic.³

Cochran's Q test

Cochran's Q test presents as null hypothesis the assertion that the studies that make up the meta-analysis are homogeneous, that is, the higher the Q value, the more heterogeneity. Thus, a problem is that the value of Q varies between 0 and infinity. A deficiency of this test is having a low power when the number of studies that make up the meta-analysis is small. On the other hand, when the number of studies is very large, it leads to false heterogeneities. In this test, a p-value is also calculated, which indicates whether or not heterogeneity is significantly different from zero.¹⁰

The I^2 Statistic

The I^2 statistic, proposed by Higgins and Thompson, is obtained from the Q statistic of the Cochran test and the number of studies. The I^2 statistic can vary from negative values to 100%. When the value is negative it is equal to 0. The p-value of I^2 is equivalent to the p-value of Q^2 .

Higgins et al. suggest a scale in which an I^2 value closer to 0% indicates non-heterogeneity among studies, while those closer to 25% indicates low heterogeneity, those closer to 50% indicates moderate heterogeneity and those closer to 75% indicates high heterogeneity among studies.²

Forest plot

The forest plot is a graphical and friendly way to demonstrate the results of a meta-analysis. It has two axes: the X and the Y (Figure 1). The Y-axis (vertical line), or central trend axis, is a line that indicates that at that point there is no difference between the interventions under study, that is, Relative Risk equal to 1 or Mean Difference equal to 0.

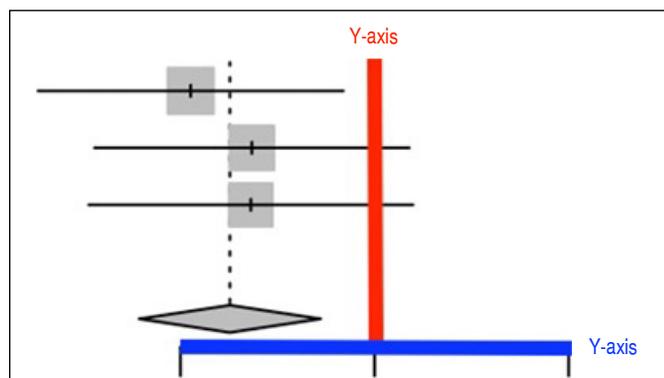


Figure 1. Axes of the forest plot.

The X-axis (horizontal line) is where the numerical dispersion of the meta-analysis results occurs. The X axis is cut in half by the Y axis and, as stated above, at this point (RR = 1 or MD = 0) there is no difference between interventions. What is to the right of this point favors an intervention and what is to the left favors another intervention. The further away from the Y-axis, the greater the effect/strength of this intervention (Figure 2).

Each individual study that makes up the meta-analysis is represented by three structures: a solid geometric shape (typically a square), a horizontal line, and a small vertical line in the center of the square (Figure 3).

The vertical line corresponds to the individual result of each study. If it is to the left of the Y axis, the result indicates a tendency of an intervention; if it is to the right of Y, it indicates a tendency for the other intervention; if it is in the center of Y, it indicates no difference between the two interventions under study (Figure 3).

The geometric shape (square) has its area as an estimate of the size of the individual effect of the study. That is, the larger the square, the greater the relative weight of the study in the meta-analysis (Figure 3).

The horizontal line corresponds to the individual confidence interval of each study. If the entire line is to the left of the Y axis, the result indicates that there is a statistically significant trend of an intervention ($p < 0.05$); if the entire line is to the right of Y, it indicates that there is a statistically significant trend for the other intervention ($p < 0.05$); if the line crosses or even "touches" the Y axis, it indicates that there is no statistically significant difference between the two interventions under study ($p > 0.05$) (Figure 3).

The diamond (rhombus), which appears below the studies, synthesizes the combined effect of all the studies that make up the meta-analysis. That is, the Diamond is the meta-analysis "in itself." The center of the Diamond corresponds to the result of the meta-analysis, and its location (to the left or right of the Y-axis) defines which intervention has the most "advantage." The Diamond width corresponds to the confidence interval of the meta-analysis. If any part of the Diamond of the meta-analysis crosses or even "touches" the Y-axis, it indicates that there is no statistically significant difference between the two interventions under study ($p > 0.05$) (Figure 3).

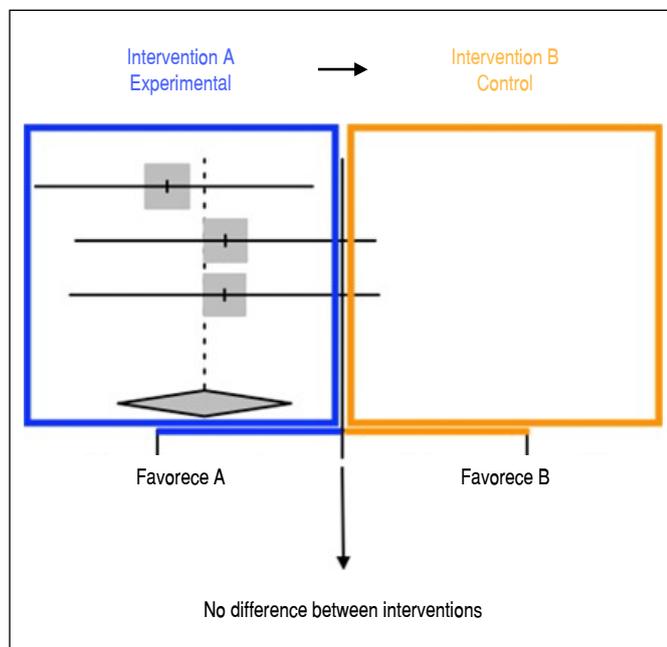


Figure 2. Forest plot intervention trends.

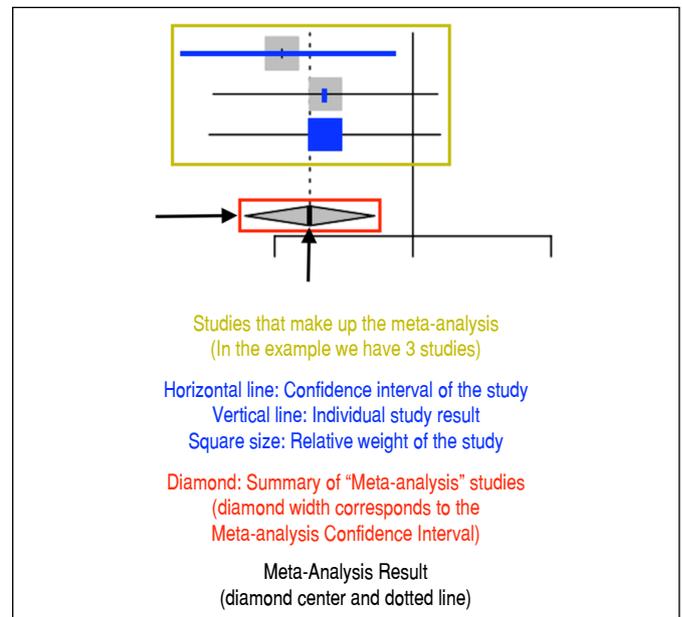


Figure 3. Forest plot "anatomy."

Meta-analysis in R

R is a free programmable statistical software, with a focus on data analysis. It consists of a platform on which the so-called "packages" (similar to applications) can be installed to perform certain functions. There are thousands of packages with different functions implemented, not to mention the user collaborations that the software receives. This guide will use the *metapackage* ("application"), which is sufficient for a good and simple meta-analysis.

Installing the R

The first step is to access the page www.r-project.org and in the left menu, under download, choose the alternative "CRAN." Now choose any of the CRAN *mirrors*, preferably one from Brazil (ex: <http://cran.fiocruz.br/>). This will redirect to one of the software's download pages. In "Download and Install R," choose the desired platform (Linux, Mac, Windows), download the installer (Latest release) and run it.

R is not software with a user-friendly interface. Some basic operations can be laborious. Thus, our second step is to install another software: RStudio. RStudio provides a good interface for importing and viewing files, installing packages, and exporting charts. In a simplistic analogous way, it is as if the R software is a kind of "Command Prompt" and RStudio is a kind of "Windows system." To download R Studio, go to the following page: <http://www.rstudio.com/products/rstudio/download/> and under "Installers for ALL Platforms" choose the most appropriate platform (Windows, Mac or Linux) and run the installation. RStudio is not required to be installed, but as stated above it greatly optimizes time during a meta-analysis. There are *free* and paid versions, and the *free* version is enough for the basics we are proposing.

As stated above, the package we will use in our meta-analysis is the "meta." To install *meta* (Figure 4), open RStudio (remember to install R before), (A) click Packages; (B) click Install; (C) The box for installation will open and then type the name *meta*. Click install and after installing, make sure that the meta package is enabled, that is, with the "check" in the box next to its name. Installing the package is only necessary once, but whenever you restart RStudio, you must enable the package by checking this option in the box (Figure 5).

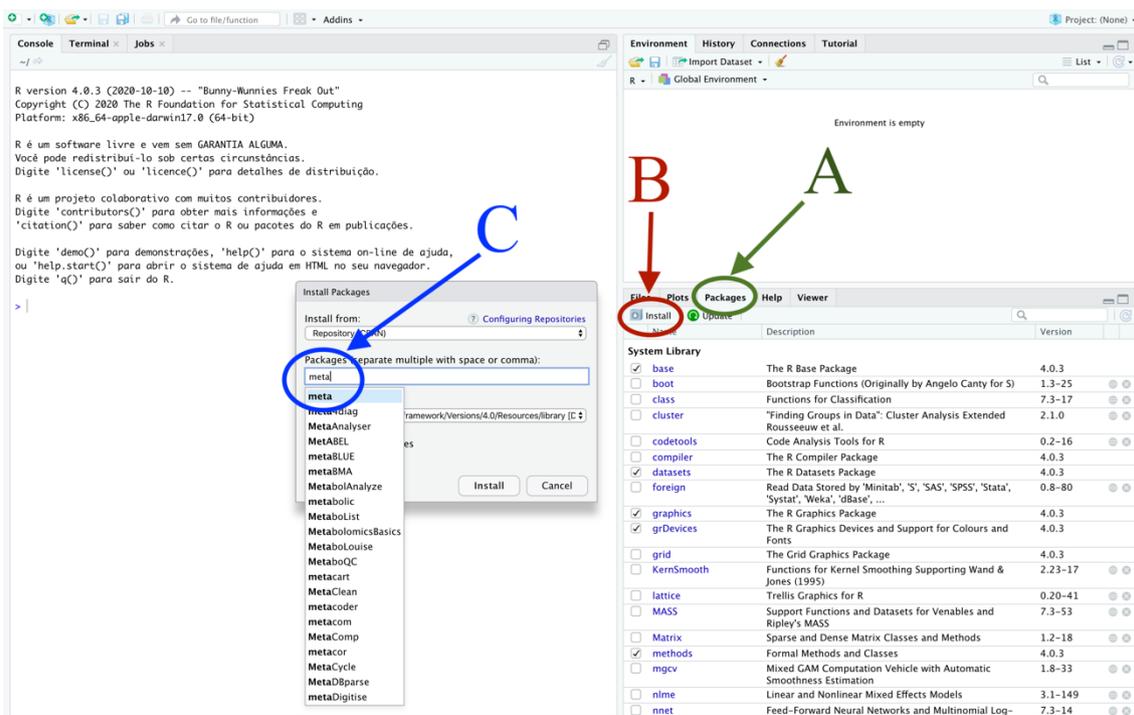


Figure 4. Installing the *meta* package.

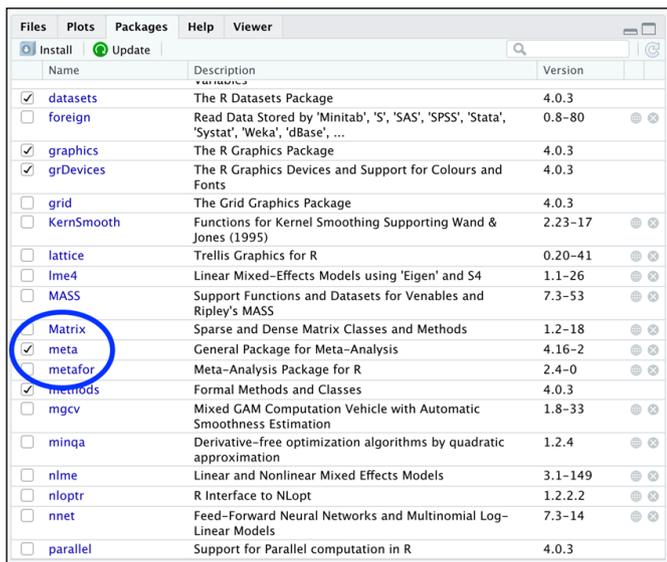


Figure 5. Enabling *meta*.

Building a database of example 1

The simplest way to create a database for analysis in R is to create a table in Microsoft Excel, Numbers (macOS), or another spreadsheet editor.

In example 1, knee stability is assessed with the Pre- and Post-operative X-test of 2 surgical techniques, A and B.

Thus, the database of example 1 will consist of a table with five columns, necessarily in this sequence (Figure 4):

Column 1: name of the studies: in this case, 3 studies;

Column 2: number of events in the experimental/treatment group (evtto – Number of patients subjected to technique A with a

positive X test POSTOP): in this case, 8, 10 and 12 patients, respectively in the 3 studies;

Column 3: total sample of the experimental/treatment group (ntto – Number of patients subjected to technique A with a positive X test PREOP): in this case, 18, 30 and 42 patients, respectively in the 3 studies;

Column 4: number of events in the control group (evcont – Number of patients subjected to technique B with a positive X test POSTOP): in this case, 18, 31 and 20 patients, respectively in the 3 studies;

Column 5: total sample of the control group (ncont – Number of patients subjected to technique B with a positive X test PREOP): in this case, 21, 60 and 45 patients, respectively in the 3 studies.

The first line defines the name of the five variables (study, evtto, ntto, evcont and ncont). The name is indifferent; however, special characters (such as diacritics or cedillas) should not be used and, if possible, everything should be lowercase (Figure 6).

When saving the database, it must be saved in the format “CSV” (variables separated by a comma). For example 1 we will name the file “testex.csv” (Figure 7).

We then have the database of example 1 ready to be imported by RStudio. Now we will open RStudio and in the menu we will go to File, Import dataset, From Text (base)... Select the testex.csv file. Make sure the parameters are the same as in Figure 8 and click the import button. The *Name* field is equivalent to the name of the variable that will be assigned within the R with the database data, in this case, “testex.” Leave the Heading option checked as Yes so that the first row of the worksheet matches the name of the database columns.

Now the R imported the database within the variable “testex.” Type testex in the RStudio console and hit “enter/return” to see the assigned value inside this variable (Figure 9). Now we have our example 1 database imported into RStudio, ready for analysis.

	A	B	C	D	E
1	estudo	evtto	ntto	evcont	ncont
2	autor 1	8	18	18	21
3	autor 2	10	30	31	60
4	autor 3	12	42	20	45

Figure 6. Example 1 database worksheet (x-test). Note that in relation to Table 1, columns B and C are inverted, as well as D and E. This is due to the meta package requiring the study event to come first (in this case the number of patients with a positive x test POSTOP) and then the total sample (number of patients subjected to the surgical technique – positive PREOP).

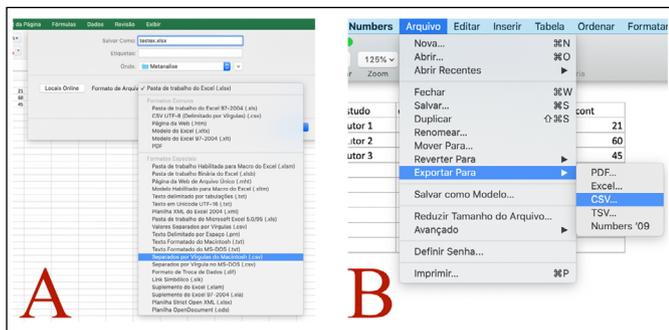


Figure 7. Exporting spreadsheet to CSV. A: Excel; B:Numbers.

Meta-analyzing Example 1 – Test X

Once the database is imported, we will proceed with the meta-analysis itself. We will use the *meta* package to run these analyses (remember to enable it, with the “check” in the box next to the name).

To perform the meta-analysis of example 1, which uses discrete quantitative variables and categorical outcome (instability improves or not with the procedure) we will use the “*metabin*” command.

We will create a variable for the metabin command of our example 1 meta-analysis, the *testex*. We will call it “*metanalisetestex*.” Thus, the command line will be:

```
metanalisetestex = metabin (evtto, ntto, evcont, ncont, estudo, data = testex)
```

Type the line above and hit “enter/return.” Remember that the names *testex* (database created from example 1) and *metanalisetestex* (variable created for the metabin command) are chosen by the author of the review, and can be any name; however, they are easy to remember and do not contain special characters.

Apparently, nothing happened, but RStudio saved the meta-analysis result within the *metanalisetestex* variable. By typing *metanalisetestex* into the console and enter/return, the software will show us the results (Figure 10).

As such, we have the results of the meta-analysis. Didactically, we can divide the results into four parts (Figure 11).

In the first part (Figure 11), we have each of the individual studies, with their relative risk (RR), confidence interval (95%CI) and weight

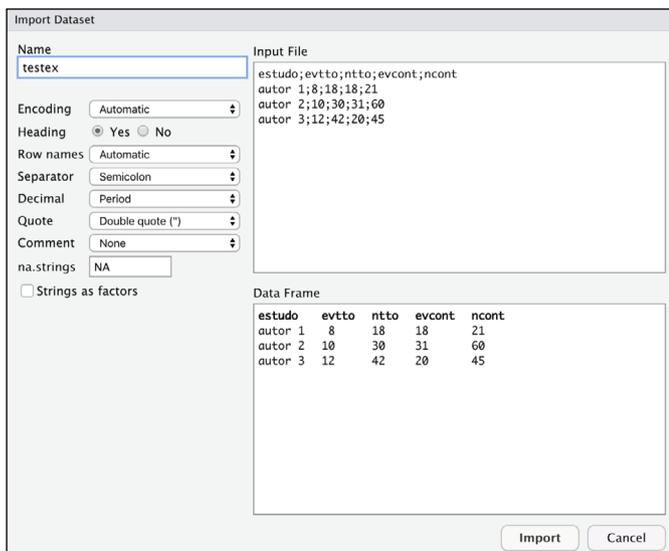


Figure 8. Importing the test CSV file into RStudio.

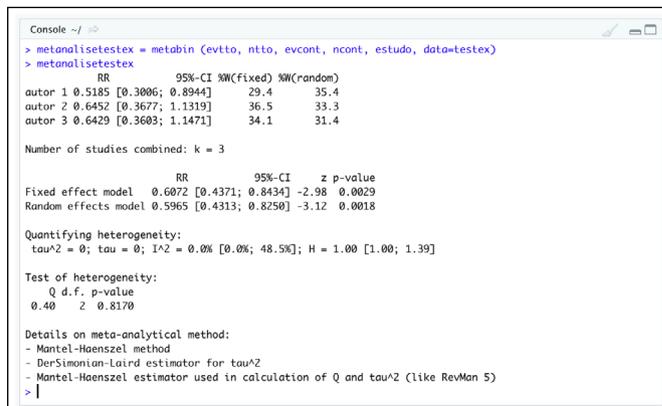


Figure 10. Results of the example 1 meta-analysis (x-test for stability of 2 surgical techniques).

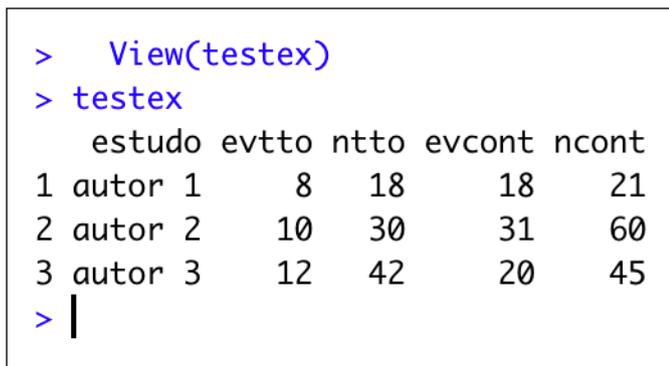


Figure 9. Database of example 1 (*testex*) in RStudio.

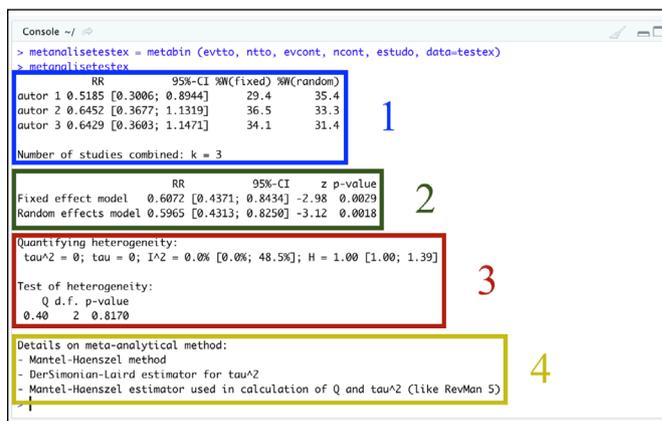


Figure 11. Four parts of the result of the example 1 meta-analysis (test x). 1: studies that make up the meta-analysis; 2: summary measure (the “result” itself) of the meta-analysis; 3: measures of heterogeneity of the meta-analysis; 4: tests used in the meta-analysis.

(%W) in the analyses by both the fixed effects model and the random effects model. In our example, three studies were combined ($k = 3$). In the second part (Figure 11), we have the summary measure of the meta-analysis, that is, the “result itself.” This part shows the relative risk (RR), the confidence interval (95% -CI) and the z-value (statistical test of the significance of the global effect, that is, a mathematical measure equivalent to the location and width of the diamond in the forest plot) for the fixed and random effects model, with their respective p-values (remembering that this p-value is what describes whether or not the study was statistically significant, with $p < 0.05$).

In the third part (Figure 11), we have measures of heterogeneity of the meta-analysis. The tau-squared (τ^2) and tau reflect the variability between studies in the meta-analysis of random effects, that is, the closer to zero the lower the variability between studies (this estimate is always calculated when the random effects model is used and its value does not have much interpretation applied). The I^2 statistic (I^2), followed by its standard deviation, as already mentioned, is an excellent indicator of heterogeneity. Similar to the I^2 statistic, the h (H) statistic and its standard deviation measure the heterogeneity of the studies, and when H is close to 1 we have evidence of homogeneity between the studies. Finally, in the third part, the value of the Q test (already mentioned above) is presented with its p-value (not to be confused with the p-value of the second part) and the degrees of freedom (d.f.), which is the number of studies minus 1 ($k-1$), which helps in the calculation of the I^2 statistic. Finally, in the fourth part (Figure 11), it is detailed which tests were used in the meta-analysis in question.

To create the forest plot of the meta-analysis, the *forest* command is used. By typing *forest* (meta-analysis name), RStudio will create a forest plot of the meta-analysis. In this case type in the console: *forest* (metanalisetestex)

If you want to omit the result/diamond of the fixed model from the forest plot (Figure 12), set the *comb.fixed* argument to false by typing the following command line in the console:

forest (metanalisetestex, *comb.fixed* = FALSE)

metanalisetestex in example 1. RStudio provides numerous ways to edit the forest plot. It is only necessary that, inside the parentheses, after the name of the variable that we attribute to our meta-analysis, a “comma” (,) is placed and the argument corresponding to what we want to edit in the forest plot. We emphasize that numerous edits can be made to the same forest plot, just follow the sequence “comma” (,) and the argument. For example, if we want the forest plot of example 1 (testex) to omit the diamond of the fixed-effect model result and the diamond of the random-effect model to be blue in color, my command will be:

forest(metanalisetestex, *comb.fixed* = FALSE, *col.diamond* = “blue”)
In Table 3, there are some useful commands to edit the forest plot (commands are in English):

Table 3. Commands for editing the forest plot in RStudio. Follow the sequence: *forest* (meta-analysis name, command 1, command 2, command 3, ..., command n).

Command	Function
test.overall = TRUE	Displays the p-value (which determines the statistical significance of the study) and the Z-value (“diamond width calculation”) in the fixed and random models
comb.fixed = FALSE	Omit in the chart the result/diamond of the fixed model
comb.random = FALSE	Omit the result/diamond from the random model in the graph
col.diamond = “blue”	Changes the color of the diamond (defaults to gray). Place the desired color in English between the “quotation marks.” In the example it is blue.
lab.e = “Medication A”	Rename the experimental groups of the studies (the default is Experimental). Place the desired name in quotation marks. In the example it is Medication A.
lab.c = “Medication B”	Change the name given to the control groups of the studies (the default is Control). Place the desired name in quotation marks. In the example it is Medication B.
xlab = “Favors A – Favors B”	Places a text below the X-axis (horizontal). Place the desired name in quotation marks. In the example it is Favors A – Favors B

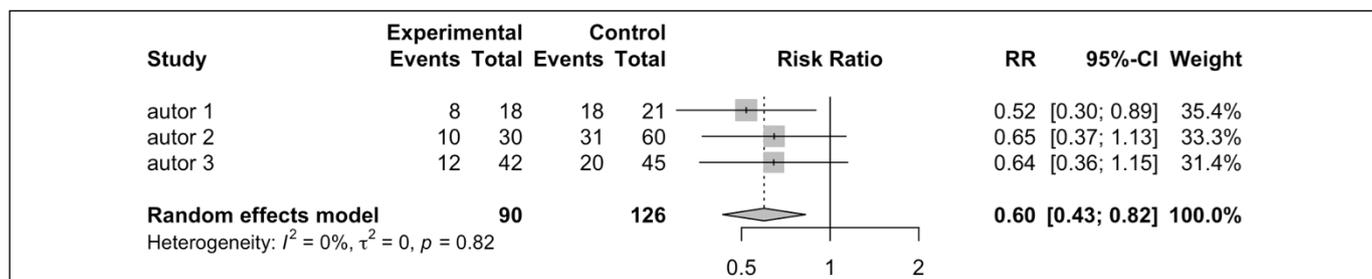


Figure 12. Forest plot of the meta-analysis of example 1 in the random effect model (x-test for stability of 2 surgical techniques, A and B).

The conclusion of the meta-analysis data from example 1 is that the risk is lower than the occurrence of persistence of instability (positive x test) in the Experimental group (technique A), RR = 0.5965 (“rounded” to 0.60 in the forest plot) (Figure 13). We can say that the use of technique A reduced the incidence of instability measured by the x test in the postoperative period by close to 40% (1-RR), compared to technique B [Relative Risk (RR) of 0.5965; confidence interval at the 95% level (95% CI) between 0.4313 and 0.8250; and p-value of 0.0018 (in the random effect model). The I^2 statistic indicates non-heterogeneity between studies ($I^2 = 0.0\%$, with a heterogeneity test p-value of 0.8170).

Basic forest plot editing

As we have seen, to create a forest plot in RStudio just type in the *forest* command line and between the parentheses put the name of the variable that we assign to our meta-analysis, in the case

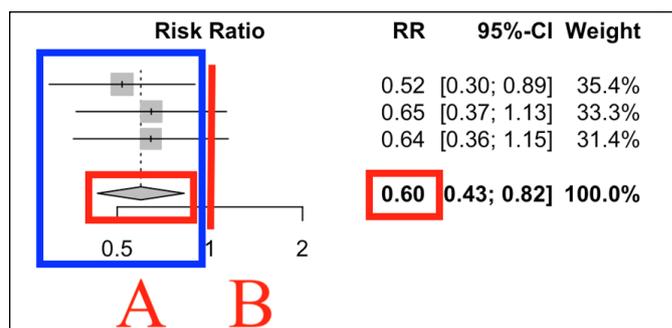


Figure 13. Forest plot of the meta-analysis of example 1 in the random effect model demonstrating advantage for technique A, with RR of 0.6 (x-test for stability of 2 surgical techniques, A and B).

Building a database of example 1

In example 2, three authors compared two surgical techniques, A (experimental) and B (control), using the functional w score in the postoperative period in both techniques, in which the higher the score, the better the result.

Thus, the database of example 1 will consist of a table with five columns, necessarily in this sequence (Figure 4):

Column 1: name of the studies: in this case, 3 studies;

Column 2: total sample of the experimental/treatment group (ne – Number of patients subjected to technique A): in this case, 18, 30 and 42 patients, respectively in the 3 studies;

Column 3: continuous quantitative variable of the event in the experimental/treatment group (me – Mean of the w score in the POSTOP period of patients subjected to technique A): in this case, 96.30; 86.90 and 79.20, respectively in the 3 studies;

Column 4: standard deviation of the continuous quantitative variable of the event in the experimental/treatment group (SDE – Standard deviation of the w score in the POSTOP period of patients subjected to technique A): in this case, ± 1.80 ; ± 9.30 and ± 18.80 , respectively in the 3 studies;

Column 5: total sample of the control group (nc – Number of patients subjected to technique B): in this case, 30, 60 and 45 patients, respectively in the 3 studies;

Column 6: continuous quantitative variable of the event in the control group (mc – Mean w score in the POSTOP of patients subjected to technique B): in this case, 90.30; 84.30 and 76.70, respectively in the 3 studies;

Column 7: standard deviation of the continuous quantitative variable of the event in the control group (sdc – Standard deviation of the w score in the POSTOP of patients subjected to technique B): in this case, ± 3.73 ; ± 9.80 and ± 17.20 , respectively in the 3 studies;

The first line defines the name of the seven variables (study, ne, me, sde, nc, mc and sdc). The name is indifferent; however, special characters (such as diacritics or cedillas) should not be used and, if possible, everything should be lowercase (Figure 14).

When saving the database, it must be saved in the “CSV” format (as seen above). For example 2 we will name the file “scorew.csv.” Now we will open RStudio and in the menu we will go to File, Import dataset, From Text (base)... Select the scorew.csv file. Make sure the parameters are the same as in Figure 8 and click the import button. The *Name* field is equivalent to the name of the variable that will be assigned within the R with the database data, in this case, “scorew.” Leave the Heading option checked as Yes so that the first row of the worksheet matches the name of the database columns.

Type scorew in the RStudio console and hit “enter/return” to see the assigned value inside this variable (Figure 16). Now we have our example 2 database imported into RStudio, ready for analysis.

	A	B	C	D	E	F	G
1	estudo	ne	me	sde	nc	mc	sdc
2	autor 1	18	96,3	1,8	30	90,3	3,73
3	autor 2	30	86,9	9,3	60	84,3	9,8
4	autor 3	42	79,2	18,8	45	76,7	17,2

Figure 14. Example 2 database worksheet (w score). The first line presents the names of the seven variables. study: names of the studies involved; ne: number of patients subjected to technique A; me: mean score W in the POST-OP of patients subjected to technique A; sde: standard deviation of score W in the POST-OP of patients subjected to technique A; nc: number of patients subjected to technique B; mc: mean score W in the POST-OP of patients subjected to technique B; sdc: standard deviation of score W in the POST-OP of patients subjected to technique B. Remember to remove the \pm sign of standard deviations.

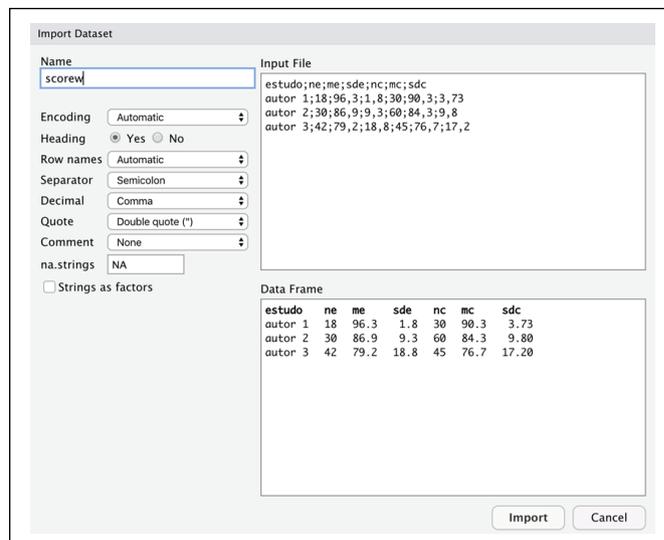


Figure 15. Importing the CSV scorew file into RStudio.

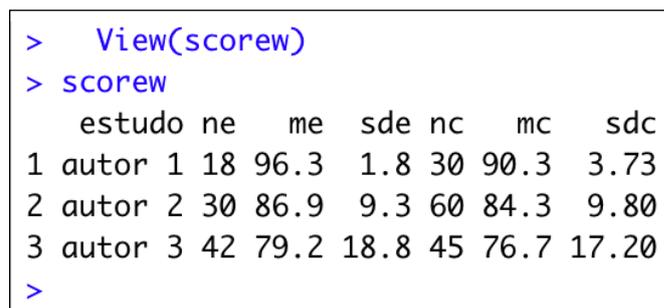


Figure 16. Database of example 2 (scorew) in RStudio.

Meta-analyzing Example 2 – w score

To perform the meta-analysis of example 2, we will use the “*metacont*” command of the *meta* package (remember to enable it, with the “check” in the box next to the name).

We will create a variable for the metacont command of our meta-analysis of example 2, the scorew. We will call it “metanalisescorew.” Thus, the command line will be:

```
metanalisescorew = metacont (ne, me, sde, nc, mc, sdc, study, data = scorew)
```

Type the line above and hit enter/return and RStudio will save the result of the meta-analysis inside the metanalisescorew variable. By typing metanalisescorew into the console and enter/return, the software will show us the results (Figure 17).

Thus we have the results of the meta-analysis of example 2, the w score. As we saw with example 1, we can divide the results of example 2 into four parts: 1. Studies that make up the meta-analysis; 2. Summary measure (the “result” itself) of the meta-analysis; 3. Measures of heterogeneity of the meta-analysis; and 4. Tests used in the meta-analysis. However, in example 2, because continuous quantitative variables are used, the result is not expressed as relative risk (as in example 1) but as mean difference (MD). That is, author 1 demonstrated a mean of 6 more “points” in the w score when using the A technique in relation to B; author 2 demonstrated a mean of 2.6 more “points” in the w score when using the A technique in relation to B; and author 3 demonstrated a mean of 2.5 more “points” in the w score when using the A technique in relation to B. By typing *forest* (meta-analysis name), RStudio will create a forest plot of the meta-analysis. In this case type in the console: `forest (metanalisescorew)`

```

> metanalisescorew = metacount (ne, me, sde, nc, mc, sdc, estudo, data=scorew)
> metanalisescorew
      MD      95%-CI %W(fixed) %W(random)
autor 1 6.0000 [ 4.4274; 7.5726]      84.3      65.8
autor 2 2.6000 [-1.5502; 6.7502]      12.1      25.0
autor 3 2.5000 [-5.0882; 10.0882]      3.6      9.3

Number of studies combined: k = 3

      MD      95%-CI      z      p-value
Fixed effect model  5.4619 [4.0182; 6.9056] 7.42 < 0.0001
Random effects model 4.8266 [2.3891; 7.2640] 3.88  0.0001

Quantifying heterogeneity:
tau^2 = 1.7081 [0.0000; >100.0000]; tau = 1.3069 [0.0000; >10.0000]
I^2 = 30.1% [0.0%; 92.7%]; H = 1.20 [1.00; 3.71]

Test of heterogeneity:
      Q      d.f.      p-value
2.86      2      0.2391

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
- Jackson method for confidence interval of tau^2 and tau
>

```

Figure 17. Results of the meta-analysis of example 2 (functionality w score of two surgical techniques).

If you want to omit the fixed model result from the graph, set the *comb.fixed* argument to false by typing the following command line in the console:
 forest (metanalisescorew, comb.fixed = FALSE)

The conclusion of the meta-analysis data from example 2 is that the experimental group (subjected to technique A) presented on average 4.8266 more “points” in the w score (MD, random effect model) in relation to the control group (subjected to technique B), MD = 4.8266 (“rounded” to 4.83 in the forest plot) (Figure 18). It is worth highlighting that in this example, what is to the right of the Y-axis is advantageous for technique A. We can say that the use of technique A has a better clinical result, measured by the w score in the postoperative period, compared to technique B [Mean Difference (MD) of 4.8266; confidence interval at the 95% level (95% CI) between 2.3891 and 7.2640; and p-value of 0.0001 (in the random effect model)]. The I^2 statistic indicates non-heterogeneity between studies ($I^2 = 0.0\%$, with a heterogeneity test p-value of 0.8170).

CONCLUSIONS

Through this article, the reader has access to the basic commands existing in the R and RStudio software, necessary for conducting a meta-analysis. The great advantage of R is the fact that it is a free software. For a better understanding of the commands, two examples were presented in a practical way, in addition to reviewing some basic concepts of this statistical technique. It is assumed that the data necessary for meta-analysis have already been collected, that is, description of methodologies for systematic review is not the discussed subject. Finally, it is worth remembering that there are many other techniques used in meta-analysis that were not addressed in this work. However, with the two examples used, the article already enables the reader to perform good and robust meta-analyses.

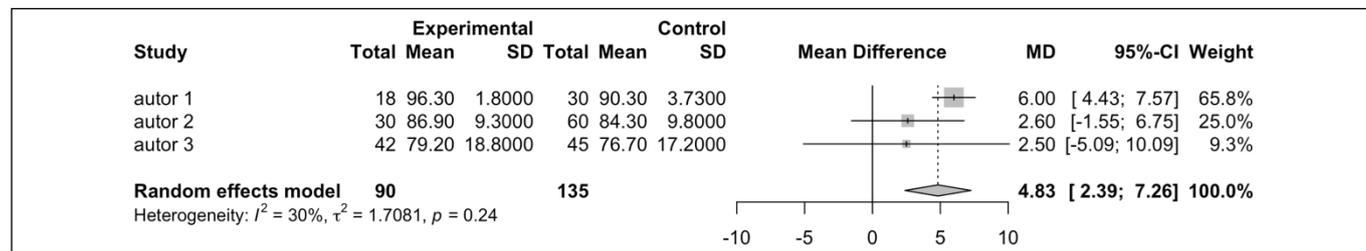


Figure 18. Forest plot of the meta-analysis of example 2 in the random effect model (functionality w score of two surgical techniques).

AUTHORS' CONTRIBUTIONS: Each author contributed individually and significantly to the development of this article. DAL: intellectual concept of the article, writing, critical review of the intellectual content and final approval of the version of the manuscript to be published; CPH: substantial contribution to the design of the work and final approval of the version of the manuscript to be published; LLL: writing, critical review of the intellectual content and final approval of the version of the manuscript to be published; RC: writing, critical review of the intellectual content and final approval of the version of the manuscript to be published; RKG: intellectual concept of the article and final approval of the version of the manuscript to be published; OPC: final approval of the version of the manuscript to be published.

REFERÊNCIAS

- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-2.
- Rodrigues CL, Ziegelmann PK. Metanálise: um guia prático. *Rev HCPA & Fac Med Univ Fed Rio Gd do Sul*. 2010;30(4):436-47.
- Santos EJF, Cunha M. Interpretação crítica dos resultados estatísticos de uma meta-análise: estratégias metodológicas. *Millenium*. 2013;(44):85-98.
- Simpson RJS, Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J*. 1904;2(2288):1243-6.
- Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res*. 1976;5(10):3-8.
- Vaudreuil NJ, Rothrauff BB, de Sa D, Musahl V. The pivot shift: current experimental methodology and clinical utility for anterior cruciate ligament rupture and associated injury. *Curr Rev Musculoskelet Med*. 2019;12(1):41-9.
- Fletcher RH, Fletcher SW, Wagner EH. *Epidemiologia clínica: elementos essenciais*. 3rd ed. Porto Alegre: Artmed; 1996.
- Hefti E, Müller W, Jakob RP, Stäubli HU. Evaluation of knee ligament injuries with the IKDC form. *Knee Surg Sports Traumatol Arthrosc*. 1993;1(3-4):226-34.
- Coutinho ESF, Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. *Braz J Psychiatry*. 2005;27(2):146-51.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to meta-analysis*. Hoboken: Wiley; 2011.
- Lanska DJ. *Epidemiology and biostatistics: an introduction to clinical research*. JAMA. 2010;303(18):1869.
- Jekel JF, Katz DL, Elmore JG. *Epidemiology, biostatistics, and preventive medicine*. Philadelphia: Saunders; 2001.
- Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet*. 1998;351(9096):123-7.
- Roever L. Compreendendo os estudos de metanálise na pesquisa clínica. *Rev Soc Bras Clin Med*. 2016;14(4):245-9.