REVISTA DE ENFERMAGEM REFERÊNCIA
homepage: https://rr.esenfc.pt/rr/
ISSNe: 2182.2883

THEORETICAL ARTICLE/ESSAY

How to measure and explore heterogeneity in a meta-analysis: Key methodological strategies

Como medir e explorar a heterogeneidade de uma meta-análise: Estratégias metodológicas fundamentais

Cómo medir y explorar la heterogeneidad en un metaanálisis: Estrategias metodológicas clave

Eduardo Santos [1, 2]
https://orcid.org/0000-0003-0557-2377

Daniela Cardoso [1, 2]
https://orcid.org/0000-0002-1425-885X

João Apóstolo [1, 2]
https://orcid.org/0000-0002-3050-4264

[1] Health Sciences Research Unit: Nursing (UICISA: E), Nursing School of Coimbra (ESEnfC), Coimbra, Portugal

[2] Portugal Centre for Evidence-Based Practice: a JBI Centre of Excellence, Coimbra, Portugal

Corresponding author
Eduardo Santos
E-mail: ejf.santos87@gmail.com

Received: 20.05.21
Accepted: 21.12.21

Abstract

Background: Systematic reviews including several studies will have some diversity, even if they address a similar topic. Studies have different designs, participants, interventions/exposures, and expected outcomes. This diversity is called heterogeneity. For its relevance, we will put forward some methodological strategies to measure and explore it.

Objective: To demonstrate how to measure and explore heterogeneity in a meta-analysis.

Main topics under analysis: We present the concept of heterogeneity in meta-analysis, as well as its types, measurement models, and conditions for its application. We also put forward several methodological options for exploring heterogeneity, using practical examples to operationalize them.

Conclusion: Exploring heterogeneity in a meta-analysis is an essential step in the development of a systematic review to increase the consistency of its results and, consequently, the strength of its recommendations.

Keywords: meta-analysis; review literature as topic; effectiveness; epidemiologic methods; evidence-based practice

Resumo

Enquadramento: As revisões sistemáticas ao incluírem diversos estudos, ainda que abordem uma questão semelhante, irão reunir inevitavelmente algum grau de diversidade. Os estudos têm diferentes desenhos, assim como participantes, intervenções/exposições ou mesmo outcomes/ resultados esperados. Essa diversidade é denominada de heterogeneidade e pela sua pertinência consideramos essencial apresentar algumas estratégias metodológicas para a medir e explorar.

Objetivo: Demonstrar como se mede e explora a heterogeneidade de uma meta-análise.

Principais tópicos em análise: É apresentado o conceito de heterogeneidade em meta-análise, bem como as suas tipologias, modelos de medida e condições para a sua aplicação. Adicionalmente apontamos diversas opções metodológicas que permitem explorar a heterogeneidade, recorrendo a exemplos práticos para as operacionalizar.

Conclusão: A exploração da heterogeneidade de uma meta-análise assume-se como um passo imprescindível na realização de uma revisão sistemática que melhora a consistência dos seus resultados e, consequentemente, a força das suas recomendações.

Palavras-chave: metanálise; literatura de revisão como assunto; efetividade; métodos epidemiológicos; prática clínica baseada em evidências

Resumen

Marco contextual: Las revisiones sistemáticas, al incluir varios estudios, aunque aborden una cuestión similar, reunirán inevitablemente cierto grado de diversidad. Los estudios tienen diferentes diseños, así como participantes, intervenciones/exposiciones o incluso resultados esperados. Esta diversidad se denomina heterogeneidad y, debido a su relevancia, consideramos imprescindible presentar algunas estrategias metodológicas para medirla y explorarla.

Objetivo: Demostrar cómo medir y explorar la heterogeneidad en un metaanálisis.

Principales temas en análisis: Se presenta el concepto de heterogeneidad en el metaanálisis, así como sus tipologías, modelos de medición y condiciones para su aplicación. Además, señalamos varias opciones metodológicas que permiten explorar la heterogeneidad, utilizando ejemplos prácticos para operacionalizarlas.

Conclusión: Explorar la heterogeneidad de un metaanálisis es un paso esencial en la realización de una revisión sistemática que mejora la consistencia de sus resultados y, en consecuencia, la fuerza de sus recomendaciones.

Palabras clave: metaanálisis; literatura de revisión como asunto; efectividad; métodos epidemiológicos; práctica clínica basada en la evidencia

How to cite this article: Santos, E., Cardoso, D., & Apóstolo, J. (2022). How to measure and explore heterogeneity in a meta-analysis: Key methodological strategies. *Revista de Enfermagem Referência*, *6*(1), e21077. https://doi.org/10.12707/RV21077

Introduction

Systematic reviews seek to identify, appraise, and summarize all relevant studies that meet predefined criteria to answer one or more predefined questions (Aromataris & Pearson, 2014). Although there are numerous types of systematic reviews with methodological specificities, the most common type is the systematic review of effectiveness, which aims to assess the effectiveness of one or more interventions or therapies (Santos & Cunha, 2013; Tufanaru et al., 2017). These systematic reviews of effectiveness lack an aggregative synthesis technique, so they usually use meta-analysis to summarize statistical results with the positivist goal of reducing the subjectivity of traditional data obtained by narrative synthesis and synthesizing data from different studies into a single measure (Apóstolo, 2017; Santos & Cunha, 2013; Santos et al., 2016).

It should be noted that the terms *systematic review* and *meta-analysis* are often misused. In the literature, we found meta-analyses that were not based on a consistent synthesis process, that is, they were not the subject of a systematic review (Santos & Cunha, 2013). The meta-analysis is one of the final steps of what should be a rigorous process that should only be performed if the necessary conditions are met, that is, in the presence of several predetermined assumptions, such as the homogeneity of the included studies (Santos & Cunha, 2013; Tufanaru et al., 2017; Tufanaru et al., 2015). We can, therefore, state that there are two synthesis options in a systematic review of effectiveness: meta-analysis and narrative synthesis (Tufanaru et al., 2015).

There are several proposed methods for the design, development, publication, and dissemination of systematic reviews, of which we highlight those proposed by the JBI Collaborating Centers (Tufanaru et al., 2017) and the Cochrane Collaboration (Higgins et al., 2019). However, even though highly specific and comprehensive, reviewers usually find some of their steps to be *easy* and *quick* (Berman & Parker, 2002), with some inaccuracies and limitations that may render all the recommendations invalid (Imrey, 2020). Interestingly, this happens several times when we refer to the heterogeneity of the results of a meta-analysis. It is common for reviewers to only perform the meta-analysis rather than analyze or consider the heterogeneity, ignoring its implications and not exploring it properly (Higgins & Thompson, 2002; Santos & Cunha, 2013).

This article presents the fundamental concepts of heterogeneity and inconsistency in a meta-analysis, as well as the different options for measuring and exploring it. We will specifically address the types of heterogeneity, the applicable statistical tests, and their implications for the choice of the analysis model, as well as the exploration of heterogeneity through sensitivity, subgroup, and meta-regression analyses. Thus, this article aims to demonstrate how to measure and explore heterogeneity in a meta-analysis.

Development

Systematic reviews with meta-analysis can provide convincing and reliable evidence for healthcare. Their value is enhanced when the results of the included studies show clinically significant effects of similar magnitude (Higgins et al., 2019). When this condition is met, the studies are homogeneous. To meet this condition, systematic reviews with meta-analyses usually calculate and interpret the statistical tests of heterogeneity (Higgins et al., 2003). At their genesis, these tests seek to confirm whether there are genuine differences underlying the results of the studies (heterogeneity) or whether the variation in results is compatible with chance alone (homogeneity; Higgins et al., 2003; Santos & Cunha, 2013).

Overall, there are three types of heterogeneity (Higgins et al., 2019; Santos & Cunha, 2013): statistical heterogeneity (differences in outcomes) is the variability in the results of studies arising from clinical or methodological diversity, the wrong choice of treatment effect measures, or chance alone; methodological heterogeneity (differences in the designs of the included studies) consists of the variability in randomization, allocation concealment, blinding, and losses to follow-up/exclusions; clinical heterogeneity (differences in the characteristics of the studies, namely in participants, interventions/exposures, and outcomes) is the actual difference between studies due to their characteristics: participants (inclusion and exclusion criteria, diagnosis), interventions (type, dose, duration), and clinical outcomes (type, scale, cut-point, length of follow-up).

Heterogeneity can be intuitively identified by checking the graphs of meta-analyses for the clustering or dispersion of the individual effects of the studies (are the results similar? do the confidence intervals overlap?; Santos & Cunha, 2013). However, this procedure does not replace the application of the tests recommended for this purpose (Higgins et al., 2019).

Statistical tests for heterogeneity derive from applying the chi-square test (X^2; Marôco, 2021). The most commonly used tests are Cochran's Q-test and the I^2 test.

The Q-test assumes that the findings of the primary studies are equal (null hypothesis) and checks whether the data found refute this hypothesis. If the null hypothesis is confirmed, the studies are considered homogeneous ($p > 0.05$; Lau et al., 1998). However, the application of this test alone is not without *problems* and/or *disadvantages* because it is based on the X^2 distribution, imposing a low power of the test for meta-analyses (Higgins et al., 2003; Lau et al., 1998). Thus, the Q-test for meta-analyses involving a small number of studies may fail to detect heterogeneity. However, when the meta-analysis involves a large number of studies, the power of the test will be high and may reveal heterogeneity between studies that is statistically significant but clinically irrelevant (Higgins & Thompson, 2002; Higgins et al., 2019; Higgins et al., 2003; Lau et al., 1998). Based on these limitations, the I^2 test obtained using the Q-test was proposed (Higgins

& Thompson, 2002).
Although the I^2 test can range from negative values to 100%, for interpretative purposes, negative values are put equal to 0. An I^2 around 0% suggests no heterogeneity between studies (homogeneity), where <25%, 25-75%, and >75% reflect low, moderate, and high heterogeneity, respectively (Higgins & Thompson, 2002; Higgins et al., 2019; Higgins et al., 2003). The I^2 test does not have the limitations of the *Q*-test, so it is recommended as a test of heterogeneity for assessing the consistency of evidence in systematic reviews with meta-analyses (Higgins et al., 2019; Higgins et al., 2003).
As heterogeneity has implications for the results of systematic reviews and their recommendations (Guyatt et al., 2008), it should be explored and minimized or addressed if possible (Higgins et al., 2019; Santos & Cunha, 2013; Tufanaru et al., 2017). Several options are described for this purpose: Excluding studies with ambiguous inclusion criteria (Higgins et al., 2019; Santos & Cunha, 2013); Excluding studies with low methodological quality (Higgins et al., 2019; Santos & Cunha, 2013); Confirming the input of the meta-analysis data and re-analyze if there is any uncertainty about the results (Higgins et al., 2019; Santos & Cunha, 2013); Repeating the meta-analysis using different statistical models (fixed-effects or random-effects models; Higgins et al., 2019; Tufanaru et al., 2017; Tufanaru et al., 2015); Repeating the meta-analysis using different measures of effect (in case of statistical heterogeneity; Higgins et al., 2019; Santos & Cunha, 2013; Tufanaru et al., 2017; Tufanaru et al., 2015); and Performing subgroup analysis or meta-regressions that predict and confirm the sources of heterogeneity. For example, the characteristics of the participants (different age groups, gender, among others), the interventions (type, dose, frequency, intensity, route of administration, among others), and the outcomes (type, different scales, cut-off points, length of follow-up, among others). These examples represent the most *easily* controlled sources of clinical heterogeneity. Additionally, subgroups can be performed based on specific aspects of a design, such as performing subgroup meta-analyses in clinical trials, taking into account aspects such as randomization and blinding (this is an example of methodological heterogeneity; Higgins et al., 2019; Richardson et al., 2019; Santos & Cunha, 2013; Tufanaru et al., 2017; Tufanaru et al., 2015).

Statistical models of meta-analysis and their impact on heterogeneity

There are two types of statistical models used in a meta-analysis: the fixed-effects model and the random-effects model (DerSimonian & Laird, 1986; Moayyedi, 2004; Santos & Cunha, 2013; Tufanaru et al., 2015). Fixed-effects models assume that the effect of interest is the same across all studies included in the meta-analysis and that the differences observed between them are only due to sampling error. In a simplified way, these models assume that any variability between studies is only due to chance and ignore their heterogeneity (Moayyedi, 2004; Santos & Cunha, 2013). On the other hand, contrary to fixed-effects models, random-effects models assume that the effect of interest is not the same in all studies, assuming some level of heterogeneity a priori. Even so, although the effects vary across studies, they usually follow a normal probability distribution (Moayyedi, 2004; Santos & Cunha, 2013). Due to these particularities, fixed-effects models generate individual study weights for the meta-analysis that are more sensitive to sample size and inversely proportional to the measure of variability estimated in the study, which is nullified by the assumption of no heterogeneity. On the other hand, random-effects models generate individual study weights that are *more sensitive* to smaller sample sizes, which ultimately increases the confidence intervals for the effect size in a meta-analysis (less *precision*; DerSimonian & Laird, 1986; Moayyedi, 2004; Santos & Cunha, 2013).
Choosing these models is essential for the heterogeneity analysis and the results of the meta-analysis because random-effects models should be used in the case of significant heterogeneity (DerSimonian & Laird, 1986; Santos & Cunha, 2013; Tufanaru et al., 2015). In contrast, fixed-effects models should only be used when there is no heterogeneity, that is, in the case of true homogeneity (Santos & Cunha, 2013; Tufanaru et al., 2015). Given the assumption that there is some degree of heterogeneity across studies, random-effects models should be used especially when there is a clear intention to generalize the results. Based on these decisions, the JBI proposed a decision flowchart (Tufanaru et al., 2015).

Methodological options to explore heterogeneity

As mentioned before, random-effects models should be preferred in the presence of significant heterogeneity (Tufanaru et al., 2015). Still, other options can be used to minimize the effects of heterogeneity on the meta-analytic value and the conclusions we can draw. The most commonly used methodological options are sensitivity analysis, subgroup meta-analysis, and meta-regression (Higgins et al., 2019; Santos & Cunha, 2013; Tufanaru et al., 2017).
Sensitivity analyses seek to explore the impact of different decisions on the outcomes/meta-analysis. These types of analyses may explore the impact of using different meta-analysis models or of excluding or including studies in meta-analyses based on sample size, methodological quality, or variance. If results remain consistent across different analyses, they can be considered robust. In opposition, different results across sensitivity analyses indicate that the results should be interpreted with caution and/or targeted for corrective action (Higgins et al., 2019; Tufanaru et al., 2017).
On the other hand, subgroup meta-analysis allows drawing valid conclusions from meta-analyses of heterogeneous studies when they are subgrouped, and there is no heterogeneity in the subgroups (i.e., the results of the individual studies within each subgroup are similar; Richardson et al., 2019). Thus, the interpretation of the subgroup analysis can lead to informative conclusions about the results of the meta-analysis that would not otherwise be possible. In addition, the analyses, in some

cases, require presentation by groups due to the clinical sense. For example, it may be necessary to estimate the effectiveness of the treatment for specific subgroups of patients (characteristics: gender, age, among others) or even of the intervention (type, dose, frequency, intensity, route of administration, among others; Richardson et al., 2019; Santos & Cunha, 2013).

There may be five scenarios in subgroup meta-analyses: 1) There is a statistically significant, quantitative subgroup effect; 2) There is a statistically significant, qualitative subgroup effect, with substantial unexplained heterogeneity; 3) There is no subgroup effect; 4) There is no subgroup effect, but there is unexplained heterogeneity; and 5) There is a statistically significant subgroup effect, with an uneven covariate distribution (Richardson et al., 2019). Finally, if studies are divided into subgroups, meta-regression can help to explore the sources of heterogeneity. Meta-regression is an extension of subgroup analyses that allows investigating the effect of continuous or categorical characteristics and the effects of multiple factors. Meta-regressions are similar in essence to simple regressions, in which an outcome/dependent variable is predicted according to the values of one or more explanatory variables. In meta-regression, the outcome variable is the effect estimate (for example, a mean difference, a risk difference). Explanatory variables are characteristics of studies that can influence the size of the intervention effect and are often referred to as 'potential effect modifiers' or covariates (Higgins et al., 2019).

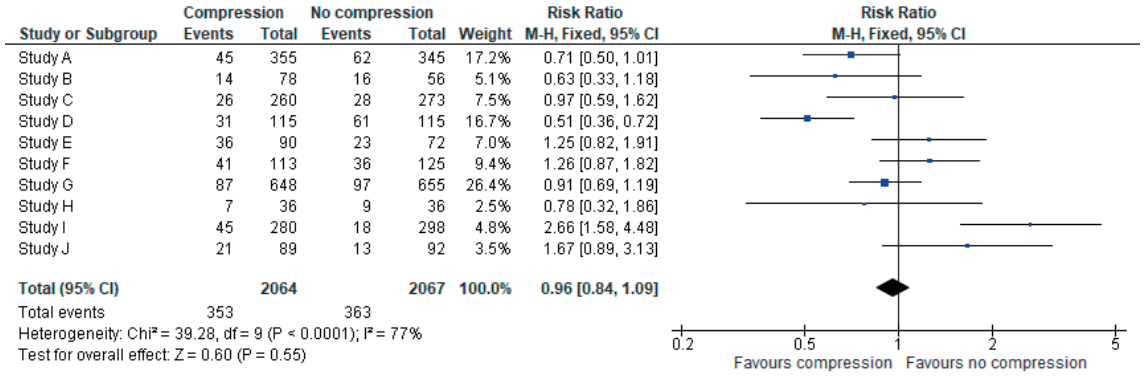
Practical example

To illustrate the concepts above more objectively, we will use a fictitious example of a systematic review with meta-analysis that sought to assess the effectiveness of compression therapy and standard care (no compression) in the treatment of venous leg ulcers. Ten clinical trials were included in the review process, and all methodological assumptions of a systematic review were met. It should be noted that we used the same intervention and comparator across studies and the outcome of interest was ulcer healing. Under these conditions, the RevMan 5.1.7 software was used to perform the meta-analyses for dichotomous variables using the classical Mantel-Haenszel statistical method.

Given that we were unaware if there was homogeneity between the studies included in the meta-analysis, we recommend assuming that the effect between the included studies is the same until proven otherwise, that is, we used fixed-effects models for this purpose (Higgins et al., 2019; Santos & Cunha, 2013; Tufanaru et al., 2017). Figure 1 illustrates the dichotomous fixed-effects meta-analysis for dummy data that revealed high heterogeneity ($I^2 = 77\%$).

Figure 1

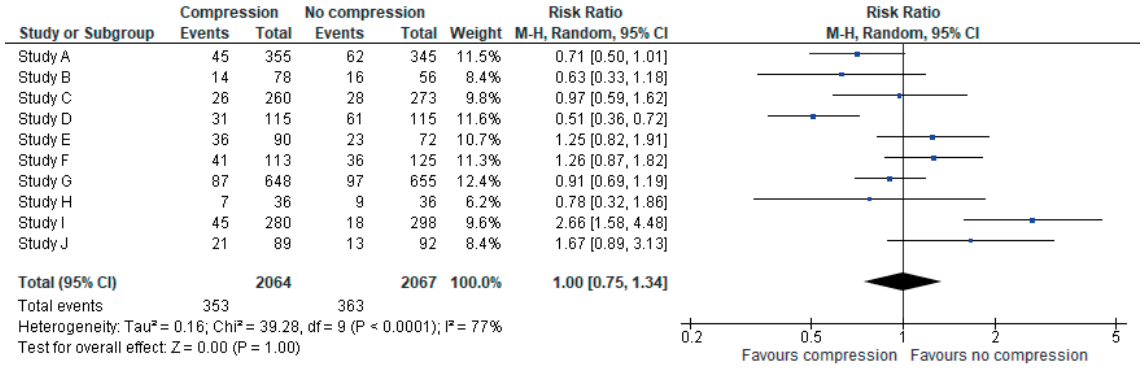
Dichotomous fixed-effects meta-analysis for dummy data with high heterogeneity



In this case of high heterogeneity, we recommend exploring it. If it cannot be assumed that the studies are homogeneous, random-effects models should be used, as previously mentioned (Figure 2).

Figure 2

Dichotomous random-effects meta-analysis for dummy data with high heterogeneity

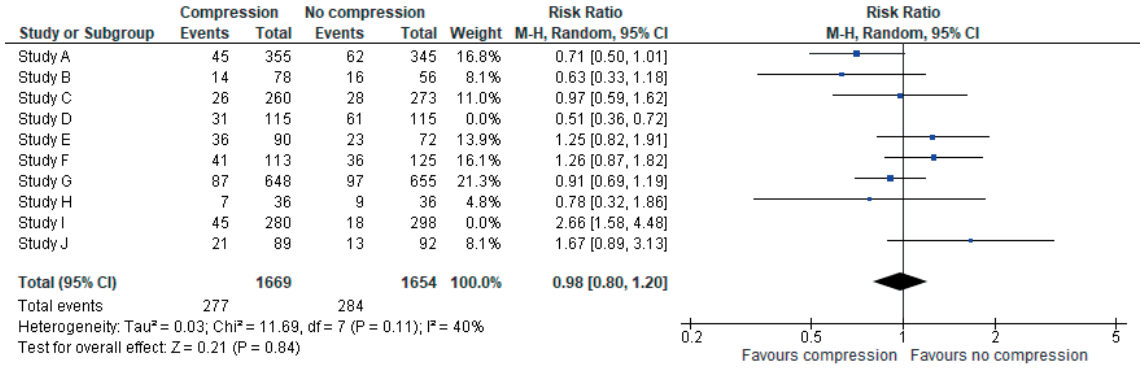


Although the levels of heterogeneity remain the same, the individual weights of each study have changed, and the total value of the meta-analysis has a greater confidence interval.

Then, some methodological options can be used to minimize the effects of heterogeneity, such as sensitivity analyses or subgroup meta-analyses. If choosing sensitivity analyses, it is clear that two studies (D and I) are responsible for a marked increase in heterogeneity (I^2 of 40 to 77%). If there is a justified criterion (e.g., low methodological quality), their removal will improve heterogeneity to a moderate level (Figure 3).

Figure 3

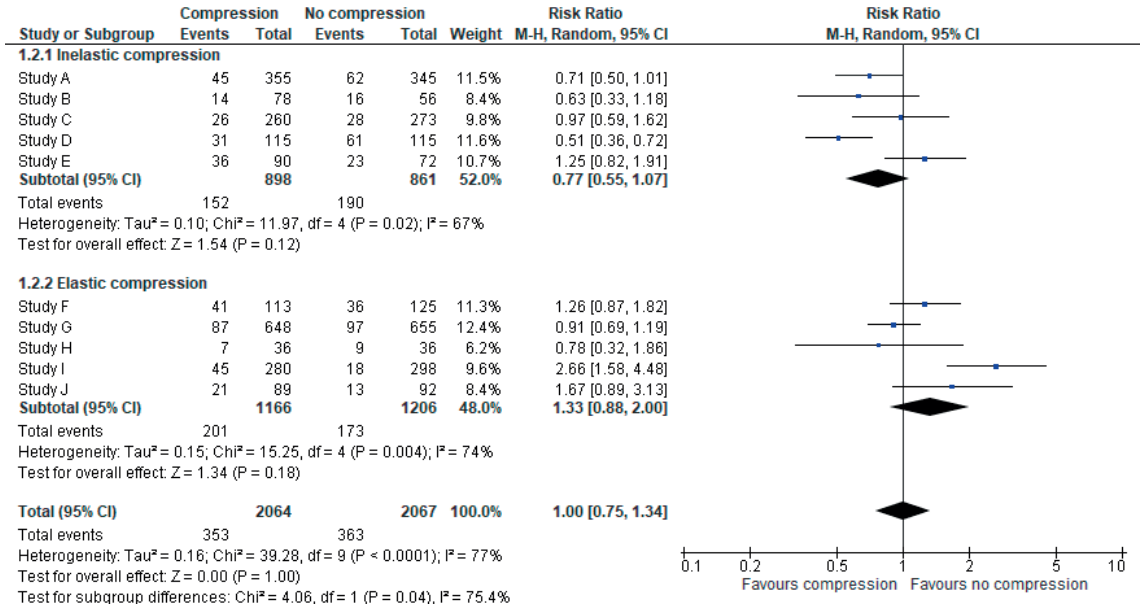
Dichotomous random-effects meta-analysis for dummy data with moderate heterogeneity after sensitivity analysis (removal of studies D and I)



Another methodological option is subgroup meta-analyses. In this case, although the intervention is compression therapy, there are different materials to administer it (elastic and non-elastic/inelastic bandages/compression). As previously mentioned, this aspect may translate into an important source of clinical heterogeneity, so it must be analyzed (Figure 4).

Figure 4

Dichotomous random-effects meta-analysis for dummy data with qualitative subgroup effect and high unexplained heterogeneity

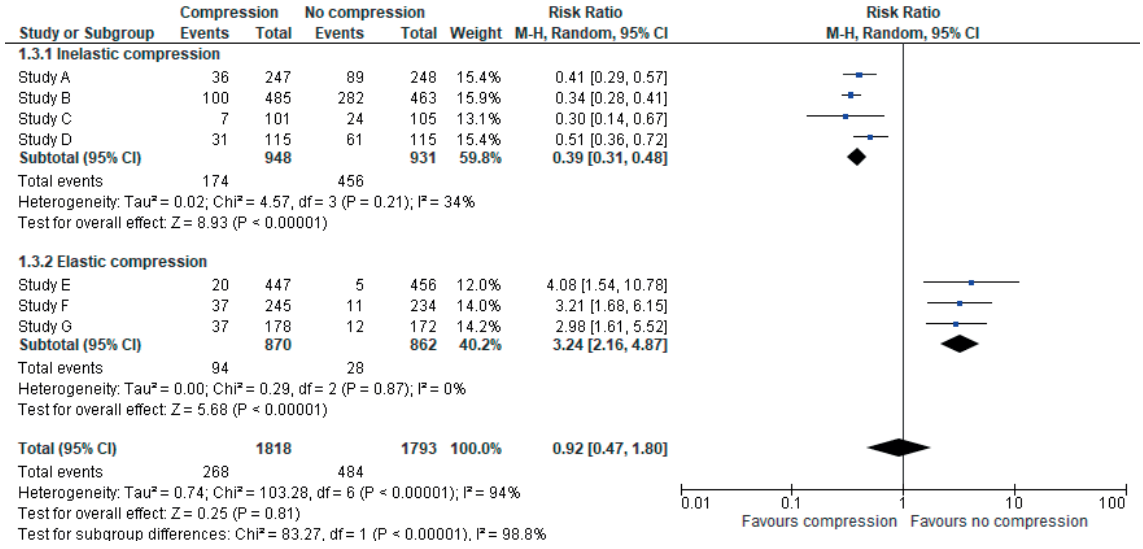


The subgroup meta-analysis revealed a statistically significant subgroup effect ($p = 0.04$), even though heterogeneity remains high within each subgroup (I^2 of 67 to 77%). In this case, the subgroup meta-analyses do not seem to justify the levels of heterogeneity. However, researchers may decide to keep them for clinical purposes.

Finally, to illustrate the quantitative subgroup effect and mild heterogeneity, Figure 5 shows another example of subgroup meta-analysis (which does not follow the previous one), in which there is a significant subgroup effect ($p < 0.00001$) that justified the levels of heterogeneity within each subgroup (I^2 from 0 to 34%).

Figure 5

Dichotomous random-effects meta-analysis for dummy data with quantitative subgroup effect and mild heterogeneity



In this case, there was no high heterogeneity in the inelastic compression subgroup ($I^2 = 0\%$), but there was mild heterogeneity in the elastic compression subgroup ($I^2 = 34\%$).

Conclusion

This article presents the concept of heterogeneity in meta-analysis. It suggests several methodological options to explore it, namely sensitivity analysis, subgroup meta-analysis, and meta-regression, although the latter is not discussed in detail.

Exploring the heterogeneity of a meta-analysis is an essential step in a systematic review (meta-analysis) as it increases its homogeneity, improves its quality and the consistency of its results, and, consequently, strengthens its recommendations.

In conclusion, we can state that by exploring and addressing heterogeneity, researchers obtain information that is highly relevant to clinical practice, rather than restricting themselves almost *blindly* to the overall value of meta-analysis or the results of individual studies. The use of subgroup meta-analyses is particularly relevant because it improves specific practices for specific groups.

Author contributions

Conceptualization: Santos, E., Cardoso, D., Apóstolo, J.
Methodology: Santos, E., Cardoso, D., Apóstolo, J.
Supervision: Apóstolo, J.
Writing – original draft: Santos, E.
Writing – review and editing: Santos, E., Cardoso, D., Apóstolo, J.

References

Apóstolo, J. (2017). *Síntese da evidência no contexto da translação da ciência*. Escola Superior de Enfermagem de Coimbra.

Aromataris, E., & Pearson, A. (2014). The systematic review: An overview. *The American Journal of Nursing, 114*(3), 53-58. https://doi.org/10.1097/01.NAJ.0000444496.24228.2c

Berman, N. G., & Parker, R. A. (2002). Meta-analysis: Neither quick nor easy. *BMC Medical Research Methodology, 2*, 10-10. https://doi.org/10.1186/1471-2288-2-10

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schunemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ, 336*(7650), 924-926. https://doi.org/10.1136/bmj.39489.470347.AD

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions version 6.0.* www.training.cochrane.org/handbook

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557-560. http://doi.org/10.1136/bmj.327.7414.557

Imrey, P. B. (2020). Limitations of meta-analyses of studies with high heterogeneity. *JAMA Network Open, 3*(1), e1919325-e1919325. https://doi.org/10.1001/jamanetworkopen.2019.19325

Lau, J., Ioannidis, J. P., & Schmid, C. H. (1998). Summing up evidence: One answer is not always enough. *Lancet, 351*(9096), 123-127. https://doi.org/10.1016/s0140-6736(97)08468-7

Marôco, J. (2021). *Análise estatística com o SPSS Statistics* (8ª ed.). ReportNumber.

Moayyedi, P. (2004). Meta-analysis: Can we mix apples and oranges? *The American Journal of Gastroenterology, 99*(12), 2297-2301. https://doi.org/10.1111/j.1572-0241.2004.40948.x

Richardson, M., Garner, P., & Donegan, S. (2019). Interpretation of subgroup analyses in systematic reviews: A tutorial. *Clinical Epidemiology and Global Health, 7*(2), 192-198. https://doi.org/10.1016/j.cegh.2018.05.005

Santos, E. J., & Cunha, M. (2013). Interpretação crítica dos resultados estatísticos de uma meta-análise: Estratégias metodológicas. *Millenium, 44*, 85-98.

Santos, E. J., Ferreira, R. J., & Marques, A. A. (2016). How to perform and interpret a network meta-analysis for indirect and mixed comparisons: Key methodological strategies. *Revista de Enfermagem Referência, 4*(8), 133-140. https://dx.doi.org/10.12707/RIV15055

Tufanaru, C., Munn, Z., Aromataris, E., Campbell, J., & Hopp, L. (2017). Systematic reviews of effectiveness. In E. Aromataris & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer's Manual* (Chapter 3). The Joanna Briggs Institute.

Tufanaru, C., Munn, Z., Stephenson, M., & Aromataris, E. (2015). Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *International Journal of Evidence-Based Health, 13*(3), 196-207. https://doi.org/10.1097/xeb.0000000000000065