

Navigating heterogeneity in meta-analysis: methods for identification and management

Muhammad Ilmawan^{1*}

¹Medical Research Center, Deka Institute, Malang, Indonesia.

*Corresponding author: Muhammad Ilmawan (Email: milmawan@gmail.com)

Citation: Ilmawan M. Navigating heterogeneity in meta-analysis: methods for identification and management. *Deka in Medicine*. 2024; 1(2): e269

Received: July 26, 2024

Revised: August 11, 2024

Accepted: August 12, 2024

Published: August 15, 2024



Copyright: © 2024 by the authors.

This is an open access article distributed under the terms and conditions of the CC BY-SA 4.0

ABSTRACT

Although meta-analysis is a powerful way to synthesize research findings from multiple studies, the problem of heterogeneity usually arises due to variation in study outcomes. Differences between studies regarding heterogeneity in results can arise from populations, interventions, outcome measures, and methodologies both within and between the studies. This article aims to provide an overview of the methods for identifying and dealing with heterogeneity in meta-analyses to ensure accurate and reliable conclusions. The article aims to describe the application of several statistical methods for detecting heterogeneity, namely the Q statistic and the I² statistic. The Q statistic is used to test whether observed variability in effect sizes exceeds chance expectations, while the I² statistic quantifies the proportion of variability due to heterogeneity. Other methods include the DerSimonian-Laird between-studies variance in random-effects models and the T and T² methods, which use both observed and expected information about effect size dispersion. Methods for dealing with heterogeneity are discussed, including choices between using fixed- versus random-effects models, and techniques for assessing and dealing with outliers using methods such as the Hedges technique. Additionally, the article explores methods to investigate sources of heterogeneity through subgroup analysis and meta-regression. Recognizing limitations such as residual heterogeneity, publication bias, and study quality is also important in making meta-analytical findings more robust. In conclusion, these methods enable researchers to more effectively address heterogeneity issues in meta-analyses, thereby providing more reliable and valid conclusions that contribute to evidence-based practice.

KEYWORDS: Meta-analysis; heterogeneity; identification; management; statistical analysis.

INTRODUCTION

Meta-analysis has emerged as the foundation of evidence synthesis and has the role to aid in bringing together research results and interpreting them across different sectors. By combining results from various studies, meta-analysis aims to provide more precise estimates of effects and reveal patterns that might not be detected in individual studies.¹ The landscape of meta-analysis has evolved dramatically since its inception, reflecting significant advancements in research synthesis methods and the increasing demand for evidence-based practices.² Meta-analysis offers several benefits: it enhances statistical precision by pooling data from multiple sources, increases the power to detect effects that might not be apparent in smaller studies, and can help resolve controversies from conflicting findings in individual studies. Additionally, meta-analysis enables researchers to evaluate treatment effects associated with different variables by conducting subgroup analyses based on the data, and findings from multiple studies can help generate new hypotheses and research questions.³ However, meta-analysis is not without its limitations. Although it provides valuable

insights, researchers must be aware of its flaws, including heterogeneity, study quality, publication bias, overgeneralization, and dependence on available data. Heterogeneity is the most formidable problem among these.⁴

Heterogeneity refers to the variability in study outcomes, which can complicate the process of drawing unified conclusions from disparate studies.⁵ Gene Glass first introduced the principle of heterogeneity in meta-analysis in 1976 when he coined the term "meta-analysis."⁶ However, researchers like Julian Higgins and Simon Thompson further developed specific statistical methods and measures to quantify heterogeneity in subsequent years, particularly in the early 2000s. They provided frameworks for assessing and interpreting heterogeneity in meta-analyses, which have become foundational in the field.⁷ Several sources such as differences in study populations, variations in intervention implementations, discrepancies in outcome measurements, and diverse methodological approaches may impact to the heterogeneity in meta-analysis.⁸ For instance, a meta-analysis on the effectiveness of a particular drug might include studies with different dosages, patient demographics, and outcome assessment tools. Such variability can obscure true effect sizes and lead to misleading conclusions if not properly addressed. To ensure the validity and reliability of meta-analytic findings, it is crucial for recognizing and managing heterogeneity.⁷ The presence of significant heterogeneity raises questions about the comparability of studies and the generalizability of the results. Thus, it is crucial to implement systematic methods for identifying, quantifying, and managing heterogeneity to maintain the integrity of the meta-analysis.⁹ This article aims to discuss the various methods available for identifying and managing heterogeneity in meta-analyses, such as the use of statistical techniques like the Q-test and I^2 statistic, which help detect and quantify heterogeneity. Furthermore, advanced approaches like subgroup analysis, meta-regression, and the use of random-effects models are discussed, which can be employed to address and account for heterogeneity. By navigating the complexities of heterogeneity, researchers can enhance the robustness and credibility of their meta-analytic conclusions, thereby making a more meaningful contribution to evidence-based practice.

HOW TO IDENTIFY HETEROGENEITY

The first step in addressing heterogeneity in meta-analysis is its identification. Several statistical tests and metrics are available for this purpose (Figure 1), each providing insights into the presence and extent of variability among study results. Accurately identifying heterogeneity is crucial for selecting appropriate methods to manage it, thereby ensuring the reliability of the meta-analytic conclusions.

Q statistic

The Q statistic, also known as Cochran's Q, is a key statistical test used in meta-analysis to evaluate the heterogeneity among the results of various studies.¹⁰ Introduced by William G. Cochran in 1954, this statistic was developed to measure the extent of variability across different studies included in a meta-analysis.¹¹ The primary purpose of the Q statistic is to determine whether the observed differences in effect sizes are greater than what would be expected by random chance alone.¹⁰ To calculate the Q statistic, one sums the squared deviations of each individual study's effect size from the overall pooled effect size, with each deviation weighted by the inverse of the study's variance. The resulting statistic follows a chi-squared distribution, where the degrees of freedom are equal to $k-1$, with k representing the number of studies included in the meta-analysis.¹² This method allows researchers to quantify the degree of heterogeneity and assess whether the variability among study results is substantial enough to require further investigation or adjustment.¹⁰

The interpretation of the Q statistic involves examining both the Q value and its associated p-value. Both metrics are influenced by the number of studies included in the meta-analysis. A significant Q statistic, typically characterized by a p-value less than 0.10 due to the conservative nature of the test, indicates the presence of heterogeneity among the study results. This suggests that the variability in effect sizes across the studies is greater than what would be expected by random chance alone.¹³ Alternatively, the Q statistic can be directly assessed: a low Q value signifies that the studies are relatively homogeneous, meaning their results are consistent and can be effectively combined. Conversely, a high Q value indicates substantial heterogeneity, suggesting that the observed differences among the study results exceed what would be expected due to sampling variability alone. This detailed interpretation helps researchers determine whether the variability in study outcomes is significant and if it warrants further investigation or adjustment in the meta-analysis.¹⁴

However, the Q test has its limitations. It has low power to detect heterogeneity when the number of studies is small and can be overly sensitive when the number of studies is large. This means that in large meta-analyses, even trivial amounts of heterogeneity may be flagged as significant, while substantial heterogeneity might go unnoticed in smaller analyses.¹⁵ Additionally, the Q statistic does not provide information on the extent of heterogeneity; it simply indicates whether heterogeneity is present. To address this limitation, researchers often use the I^2 index, which quantifies the percentage of total variation across studies attributable to heterogeneity rather than chance. The I^2 index complements the Q statistic by offering a clearer picture of the extent of variability due to heterogeneity.⁹

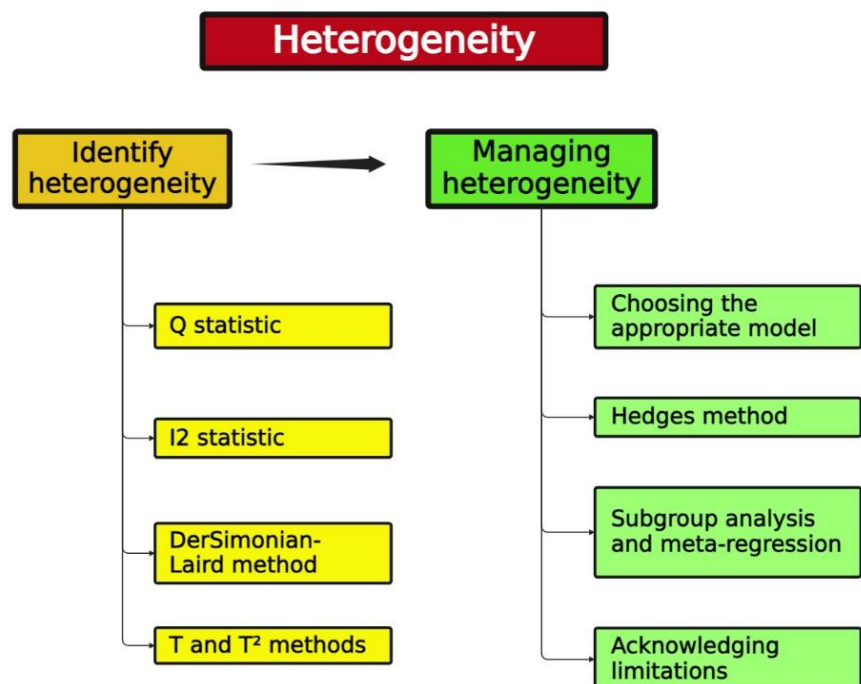


Figure 1. A summary of how to identify and manage heterogeneity in meta-analyses.

I^2 statistic

The I^2 statistic in meta-analysis is used to quantify the degree of heterogeneity among the results of different studies.¹⁶ Introduced by Higgins and Thompson in 2002, this statistic was developed to provide a more interpretable measure of heterogeneity,

complementing Cochran's Q statistic.⁷ The I^2 statistic offers a more intuitive understanding of heterogeneity by representing the proportion of total variability in effect sizes attributable to heterogeneity rather than random chance.⁹ Unlike the Q statistic, which is influenced by the number of studies included in the meta-analysis, the I^2 statistic remains unaffected by sample size, making it a more robust measure.¹² The I^2 statistic has several advantages, including its clear interpretability, comparability across different studies, guidance for modeling, complementarity with the Q statistic, and its ability to provide insights into the sources of heterogeneity.¹⁴

The interpretation of the I^2 statistic in meta-analysis is based on its value, which reflects the degree of heterogeneity among the study results. An I^2 value of 0% indicates no observed heterogeneity, while higher values signify increasing levels of heterogeneity. For example, an I^2 value exceeding 50% is generally regarded as indicating substantial heterogeneity.¹⁷ While the interpretation of I^2 values can be somewhat subjective, there are commonly accepted guidelines: an I^2 value between 0-25% suggests low heterogeneity, 25-50% indicates moderate heterogeneity, 50-75% represents substantial heterogeneity, and 75-100% signifies considerable heterogeneity.¹⁸ These guidelines help researchers assess the extent of variability among study results and determine the need for further investigation or adjustment.

The I^2 statistic has several limitations. First, it can be sensitive to the characteristics of the included studies, such as sample size and effect size. Variations in these factors can influence the I^2 value, potentially leading to misleading interpretations of heterogeneity.¹⁶ Second, while I^2 quantifies the extent of heterogeneity, it does not provide information about the direction or nature of this heterogeneity. Consequently, researchers cannot determine whether the observed differences in study results are due to systematic biases or true variations in effects.⁹ Third, the I^2 statistic can be affected by the number of studies included in the meta-analysis; in smaller meta-analyses, even minor differences in study results can lead to a high I^2 value that may not accurately reflect the true level of heterogeneity.¹⁶ Fourth, researchers may misinterpret I^2 values, especially if they do not consider the context of the studies involved. For example, a high I^2 value might be viewed as problematic without understanding the underlying reasons for heterogeneity, such as differences in populations or interventions.¹⁷ Finally, the I^2 statistic may not be suitable for all types of meta-analyses, particularly those involving studies with very different designs or outcomes. In such cases, the I^2 value might not provide meaningful insights into heterogeneity, limiting its usefulness for guiding analytical decisions.¹⁶

DerSimonian-Laird method

The DerSimonian-Laird method is a foundational technique in meta-analysis that provides a framework for synthesizing study results while accounting for variability among studies.¹⁹ Introduced by Nan Laird and William DerSimonian in 1986, this method is employed in random-effects meta-analyses to address heterogeneity.²⁰ Unlike fixed-effects models, which assume that all studies estimate the same underlying effect, random-effects models acknowledge the existence of variability both within and between studies. The DerSimonian-Laird method estimates the between-study variance (τ^2) using the Q statistic and adjusts the weights of the studies to account for this additional variability. This adjustment yields more conservative and reliable estimates of the overall effect size, as well as more accurate confidence intervals and p-values, by incorporating the variability due to heterogeneity.²¹ Interpreting the DerSimonian-Laird method involves understanding its role in synthesizing varying effect sizes, estimating between-study variance, calculating weighted averages,

providing confidence intervals, assessing statistical significance, and recognizing its limitations.²²

The DerSimonian-Laird method offers several advantages. First, it is straightforward to implement, requiring only basic data summaries, such as effect sizes and variances, from each study. This simplicity makes it accessible to researchers without extensive statistical training, enabling them to conduct meta-analyses efficiently.²³ Second, unlike some more complex methods that involve iterative calculations, the DerSimonian-Laird method is non-iterative. This characteristic reduces computational burden and time, making it a practical choice for many researchers.²² Third, the method is particularly effective in estimating overall treatment effects, especially when the sample sizes of the studies are large, providing reliable estimates that can inform clinical and policy decisions.²³ Finally, the DerSimonian-Laird method explicitly accounts for heterogeneity among studies by incorporating both within-study and between-study variances. This capability to model variability allows researchers to draw more accurate conclusions about the effects being studied.²⁴

The DerSimonian-Laird method has several limitations. First, it can be inefficient when estimating between-study variance, particularly when the number of studies is small or when there is significant variability in study sizes. This inefficiency can lead to unreliable estimates of heterogeneity and may increase the risk of false positives in the conclusions drawn from the meta-analysis.²⁴ Second, a theoretical drawback of the DerSimonian-Laird method is that it often produces confidence intervals that are slightly too narrow. This is because the method does not fully account for the uncertainty associated with estimating heterogeneity, which can lead to misleading conclusions about the significance of the overall effect.²² Third, the method's results can be sensitive to the characteristics of the included studies, such as sample sizes and effect sizes. Variability in these factors can affect the accuracy of the pooled effect size and the estimated heterogeneity, making the method less robust in certain contexts.⁹ Fourth, although the DerSimonian-Laird method is designed for random-effects models, it assumes that true effects are normally distributed around the average effect. If this assumption does not hold, the results may be biased or inaccurate, compromising the validity of the meta-analysis.²⁴ Finally, in cases where studies are highly heterogeneous, the DerSimonian-Laird method may not adequately address the underlying differences among studies. Alternative methods that better account for extreme heterogeneity may be more appropriate, as they can provide more reliable estimates of the overall effect size and its associated uncertainty.²⁵

T and T² methods

The T and T² methods in meta-analysis are statistical approaches used to estimate treatment effects and the variability among studies.²⁶ These methods are discussed within the context of random-effects models and build on the foundational work of William DerSimonian and Nan Laird.²³ The T and T² methods provide alternative ways to quantify heterogeneity in meta-analyses. Specifically, the T method estimates the standard deviation (T) of the distribution of true effect sizes, reflecting the spread or dispersion of effect sizes around the overall mean.⁹ A larger T value indicates greater heterogeneity. In contrast, the T² method calculates the variance (T²) of the effect sizes, offering an estimate of how much true effect sizes differ from each other. As T² is the square of T, larger T² values also signify greater heterogeneity.²⁷ Both methods are valuable for providing a detailed understanding of effect size distribution, complementing the insights provided by the Q and I² statistics. They are especially

useful under the assumption of normally distributed effect sizes, allowing for more nuanced interpretations of variability among studies.¹²

The T and T² methods offer several advantages in meta-analysis. First, the T method provides a systematic approach to estimating the overall treatment effect across multiple studies. By pooling data, it enables researchers to obtain a more reliable estimate of the effect size than any single study, leading to more conclusive findings.²⁷ Second, the T² method specifically quantifies the between-study variance, offering insights into the degree of heterogeneity among the studies included in the meta-analysis. This information is crucial for understanding how much of the variability in effect sizes is due to true differences in study populations or interventions, rather than random error.⁹ Third, the T and T² methods support the use of random-effects models, which are particularly useful when studies are heterogeneous. This flexibility allows researchers to select the most appropriate model based on the characteristics of the studies, thereby enhancing the validity of the conclusions drawn from the meta-analysis.²⁸ Finally, by integrating results from multiple studies, the T method increases the statistical power to detect significant effects. This is particularly beneficial in fields where individual studies may have small sample sizes, as the combined data can reveal effects that might otherwise go unnoticed.⁹

Despite their advantages, the T and T² methods also have several limitations. First, these methods can be sensitive to the characteristics of the included studies, such as sample sizes and effect sizes. Variability in these factors can impact the accuracy of the pooled effect size and the estimated heterogeneity, potentially leading to misleading conclusions.²⁷ Second, both methods generally assume that true effects are normally distributed. If this assumption does not hold, the results may be biased or inaccurate, particularly when the studies included in the meta-analysis have very different designs or populations.²⁹ Third, the T² method, which estimates between-study variance, can be inefficient when the number of studies is small. This inefficiency may result in unreliable estimates of heterogeneity, making it challenging to draw meaningful conclusions from the analysis.¹⁹ Fourth, the T method aggregates results from different studies, which can lead to oversimplification. When studies vary significantly in methodology, population, or intervention, aggregating these results may obscure important differences and nuances, potentially leading to erroneous interpretations.²⁹ Finally, in cases where studies exhibit high heterogeneity, the T and T² methods may not adequately address the underlying differences among studies. Alternative methods that better account for extreme heterogeneity might be more suitable, as they can provide more reliable estimates of the overall effect size and its associated uncertainty.⁹

HOW TO MANAGE HETEROGENEITY

Effective management of heterogeneity is crucial for ensuring the robustness and validity of meta-analytic findings. This section discusses various methods for managing heterogeneity, including selecting the appropriate model, assessing and handling outliers, exploring sources of heterogeneity, and acknowledging the limitations (Figure 1).

Choosing the appropriate model: random-effects vs. fixed-effects models

One of the first steps in managing heterogeneity in meta-analysis is choosing the appropriate model: the fixed-effects model or the random-effects model.²⁸ The fixed-effects model assumes that all studies estimate the same underlying effect size.²¹ Introduced by Cochran in 1954, this model established foundational principles for

combining results from various studies under the assumption of a single true effect size.¹¹ It is particularly suitable when there is minimal to no heterogeneity among study results. The fixed-effects model gives greater weight to larger studies, providing a precise estimate of the common effect size.²¹ However, its effectiveness diminishes in the presence of significant heterogeneity, as it does not account for variability between studies. Using a fixed-effects model under such conditions can lead to biased estimates and misleading conclusions, emphasizing the importance of careful model selection.²⁸

The random-effects model, in contrast, assumes that effect sizes vary between studies rather than being identical.²⁸ Introduced by DerSimonian and Laird in 1986, this model accounts for both within-study and between-study variability, making it more suitable for meta-analyses with significant heterogeneity.²⁰ By incorporating this variability, the random-effects model provides a more conservative estimate of the overall effect size, as evidenced by wider confidence intervals that reflect the additional uncertainty.³⁰ The DerSimonian-Laird method is commonly used to estimate the between-study variance (τ^2) in random-effects models, adjusting the weights of the studies accordingly. This approach ensures that variability among studies is appropriately addressed, leading to more reliable and accurate meta-analytic conclusions.²⁰

Assessing and handling outliers: hedges method

Outliers can significantly impact the results of a meta-analysis, especially in the presence of heterogeneity. Identifying and managing outliers is crucial for ensuring the validity and accuracy of the findings.³¹ Outliers can be detected through visual inspection of forest plots or by using statistical tests. Forest plots, which display the effect sizes and confidence intervals of individual studies, are particularly useful for visually identifying studies that deviate markedly from the overall effect. This visual inspection aids in pinpointing studies that may disproportionately influence the meta-analysis results, enabling researchers to address these anomalies and enhance the robustness of their conclusions.³²

The Hedges method is a statistical approach for detecting and managing outliers in meta-analyses.³³ Introduced by Larry V. Hedges in the early 1980s, this method calculates the standardized residuals for each study and identifies those with residuals exceeding a specified threshold, such as 1.96 for a 95% confidence level.³⁴ Studies flagged as outliers can then be further examined to determine whether they should be excluded or down-weighted in the analysis. By removing or adjusting outliers, researchers can reduce the influence of anomalous data points, thereby enhancing the accuracy and reliability of the meta-analytic results.³⁵ It is also essential to document and justify any decisions to exclude or adjust studies to ensure transparency and uphold the integrity of the research process.³⁶

Exploring sources of heterogeneity: subgroup analysis and meta-regression

Understanding the sources of heterogeneity provides valuable insights into the variability among study results and helps refine meta-analytic conclusions.³⁷ Subgroup analysis is a technique used to explore these sources by dividing studies into meaningful subgroups based on specific characteristics, such as population demographics, intervention types, or study quality.³⁸ By conducting separate meta-analyses for each subgroup, researchers can identify whether certain characteristics are associated with different effect sizes. This method can reveal patterns that might not be apparent in the overall analysis, allowing researchers to tailor their conclusions to specific contexts and populations.³⁷ Analyzing these subgroup differences helps

researchers gain a deeper understanding of the underlying causes of heterogeneity, leading to more nuanced and applicable insights.⁹

Meta-regression builds on the methodology of subgroup analysis by enabling the simultaneous inclusion of multiple covariates, thus providing a more detailed understanding of the factors that contribute to heterogeneity in meta-analytic studies.³⁹ First introduced by Gene V. Glass in the 1970s, this technique represents a significant advancement in meta-analytic methods.⁴⁰ Meta-regression models investigate the relationship between various study characteristics—such as sample size, duration of follow-up, methodological quality, and other relevant covariates—and the effect sizes reported in the studies.⁴¹ This approach elucidates which specific factors are associated with variability in effect sizes across different studies. The method is particularly effective in uncovering complex interactions between study characteristics and effect sizes that might not be detected through simpler subgroup analyses.²⁷ For instance, meta-regression can reveal how variations in study design or participant demographics may influence observed effect sizes, thereby providing a more comprehensive understanding of heterogeneity. However, the reliability of meta-regression is contingent upon the availability of a sufficient number of studies to ensure robust estimates. Therefore, while meta-regression offers valuable insights, its results must be interpreted with caution to avoid overfitting and ensure that the conclusions drawn are both accurate and applicable.⁴²

Acknowledging limitations

No meta-analysis is without limitations, and recognizing these limitations is crucial for maintaining research integrity.³⁷ Common limitations include heterogeneity, publication bias, study quality, and generalizability.³ Despite efforts to manage heterogeneity, some degree of variability may persist, and it is essential to address the potential impact of this residual heterogeneity on the findings.⁹ Publication bias—where studies with positive results are more frequently published than those with negative or null results—can skew the meta-analysis.⁴³ Techniques such as funnel plots and Egger's test should be employed to assess and mitigate publication bias.⁴⁴ The quality of the included studies also influences the reliability of the meta-analytic conclusions; conducting sensitivity analyses and excluding low-quality studies can help address this issue.⁴⁵ Additionally, the findings of a meta-analysis may not be generalizable to all populations and settings. Researchers must consider the characteristics of the included studies and discuss the applicability of the results.⁴⁶ By acknowledging these limitations, researchers can provide a more balanced interpretation of their findings and offer informed recommendations for future research.

CONCLUSION

In conclusion, heterogeneity in meta-analysis can significantly impact results and must be carefully managed. Key methods for detecting heterogeneity include the Q statistic, I² statistic, DerSimonian-Laird method, and T and T² methods. To address heterogeneity, researchers can choose between fixed-effects and random-effects models, handle outliers using the Hedges method, and explore sources of variability through subgroup analysis and meta-regression. Acknowledging and addressing the limitations of these methods is also crucial for ensuring accurate and reliable meta-analytic conclusions.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Data used in our study were presented in the main text.

COMPETING INTERESTS

Not applicable.

FUNDING SOURCE

Not applicable.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTION

Conceptualization: MI; Data Curation: MI; Formal Analysis: MI; Investigation: MI; Project Administration: MI; Resources: MI; Methodology: MI; Software: MI; Visualization: MI; Supervision: MI; Validation: MI; Writing – Original Draft Preparation: MI; Writing – Review & Editing: MI. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

REFERENCES

1. Wang XM, Zhang XR, Li ZH, et al. A brief introduction of meta-analyses in clinical practice and research. *J Gene Med* 2021;23(5):e3312.doi: 10.1002/jgm.3312. PMID: 33450104
2. Gurevitch J, Koricheva J, Nakagawa S, et al. Meta-analysis and the science of research synthesis. *Nature* 2018;555(7695):175-182.doi: 10.1038/nature25753. PMID: 29517004
3. Greco T, Zangrillo A, Biondi-Zoccai G, et al. Meta-analysis: pitfalls and hints. *Heart Lung Vessel* 2013;5(4):219-225.doi. PMID: 24364016
4. Yadav S. Challenges and concerns in the utilization of meta-analysis software: Navigating the landscape of scientific synthesis. *Cureus* 2024;16(1):e53322.doi: 10.7759/cureus.53322. PMID: 38435898
5. Ranganathan P, Aggarwal R. Study designs: Part 8 - Meta-analysis (I). *Perspect Clin Res* 2020;11(4):178-181.doi: 10.4103/picr.PICR_283_20. PMID: 33489837
6. Glass GV. Primary, secondary, and meta-analysis of research. *Educational researcher* 1976;5(10):3-8.doi: 10.3102/0013189X0050100. PMID:
7. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21(11):1539-1558.doi: 10.1002/sim.1186. PMID: 12111919
8. Tong G, Guo G. Meta-analysis in sociological research: Power and heterogeneity. *Sociol Methods Res* 2022;51(2):566-604.doi: 10.1177/0049124119882479. PMID: 35754525
9. Linden AH, Honekopp J. Heterogeneity of research results: A new perspective from which to assess and promote progress in psychological science. *Perspect Psychol Sci* 2021;16(2):358-376.doi: 10.1177/1745691620964193. PMID: 33400613
10. Kulinskaya E, Hoaglin DC. On the Q statistic with constant weights in meta-analysis of binary outcomes. *BMC Med Res Methodol* 2023;23(1):146.doi: 10.1186/s12874-023-01939-z. PMID: 37344771
11. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954;10(1):101-129.doi: 10.2307/3001666. PMID:
12. van Aert RCM, van Assen M, Viechtbauer W. Statistical properties of methods based on the Q-statistic for constructing a confidence interval for the between-study variance in meta-analysis. *Res Synth Methods* 2019;10(2):225-239.doi: 10.1002/jrsm.1336. PMID: 30589219
13. Pereira TV, Patsopoulos NA, Salanti G, et al. Critical interpretation of Cochran's Q test depends on power and prior assumptions about heterogeneity. *Res Synth Methods* 2010;1(2):149-161.doi: 10.1002/jrsm.13. PMID: 26061380
14. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, et al. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006;11(2):193-206.doi: 10.1037/1082-989X.11.2.193. PMID: 16784338
15. von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med Res Methodol* 2015;15(1):35.doi: 10.1186/s12874-015-0024-z. PMID: 25880989

16. Thorlund K, Imberger G, Johnston BC, et al. Evolution of heterogeneity (I²) estimates and their 95% confidence intervals in large meta-analyses. *PLoS One* 2012;7(7):e39471. doi: 10.1371/journal.pone.0039471. PMID: 22848355
17. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560. doi: 10.1136/bmj.327.7414.557. PMID: 12958120
18. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007;335(7626):914-916. doi: 10.1136/bmj.39343.408449.80. PMID: 17974687
19. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016;7(1):55-79. doi: 10.1002/jrsm.1164. PMID: 26332144
20. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-188. doi: 10.1016/0197-2456(86)90046-2. PMID: 3802833
21. Zhai C, Guyatt G. Fixed-effect and random-effects models in meta-analysis. *Chin Med J (Engl)* 2024;137(1):1-4. doi: 10.1097/CM9.0000000000002814. PMID: 37612263
22. Jackson D, Law M, Barrett JK, et al. Extending DerSimonian and Laird's methodology to perform network meta-analyses with random inconsistency effects. *Stat Med* 2016;35(6):819-839. doi: 10.1002/sim.6752. PMID: 26423209
23. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials* 2015;45(Pt A):139-145. doi: 10.1016/j.cct.2015.09.002. PMID: 26343745
24. Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. *Res Synth Methods* 2018;9(3):382-392. doi: 10.1002/jrsm.1297. PMID: 29504289
25. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. *PLoS One* 2013;8(7):e69930. doi: 10.1371/journal.pone.0069930. PMID: 23922860
26. Welton NJ, Soares MO, Palmer S, et al. Accounting for heterogeneity in relative treatment effects for use in cost-effectiveness models and value-of-information analyses. *Med Decis Making* 2015;35(5):608-621. doi: 10.1177/0272989X15570113. PMID: 25712447
27. Scammacca N, Roberts G, Stuebing KK. Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Rev Educ Res* 2014;84(3):328-364. doi: 10.3102/0034654313500826. PMID: 25309002
28. Dettori JR, Norvell DC, Chapman JR. Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider. *Global Spine J* 2022;12(7):1624-1626. doi: 10.1177/21925682221110527. PMID: 35723546
29. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172(1):137-159. doi: 10.1111/j.1467-985X.2008.00552.x. PMID: 19381330
30. Partlett C, Riley RD. Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Stat Med* 2017;36(2):301-317. doi: 10.1002/sim.7140. PMID: 27714841
31. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods* 2010;1(2):112-125. doi: 10.1002/jrsm.11. PMID: 26061377
32. Kadlec D, Sainani KL, Nimphius S. With great power comes great responsibility: Common errors in meta-analyses and meta-regressions in strength & conditioning research. *Sports Med* 2023;53(2):313-325. doi: 10.1007/s40279-022-01766-0. PMID: 36208412
33. Lin L, Chu H, Hodges JS. Alternative measures of between-study heterogeneity in meta-analysis: Reducing the impact of outlying studies. *Biometrics* 2017;73(1):156-166. doi: 10.1111/biom.12543. PMID: 27167143
34. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Orlando, FL: Academic press; 1985.
35. Gress TW, Denvir J, Shapiro JI. Effect of removing outliers on statistical inference: implications to interpretation of experimental data in medical research. *Marshall J Med* 2018;4(2):9. doi: 10.18590/mjm.2018.vol4.iss2.9. PMID: 32923665
36. Prager EM, Chambers KE, Plotkin JL, et al. Improving transparency and scientific rigor in academic publishing. *Brain Behav* 2019;9(1):e01141. doi: 10.1002/brb3.1141. PMID: 30506879
37. Mikolajewicz N, Komarova SV. Meta-analytic methodology for basic research: A practical guide. *Front Physiol* 2019;10(1):203. doi: 10.3389/fphys.2019.00203. PMID: 30971933
38. Wang X, Piantadosi S, Le-Rademacher J, et al. Statistical considerations for subgroup analyses. *J Thorac Oncol* 2021;16(3):375-380. doi: 10.1016/j.jtho.2020.12.008. PMID: 33373692
39. Tipton E, Pustejovsky JE, Ahmadi H. A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Res Synth Methods* 2019;10(2):161-179. doi: 10.1002/jrsm.1338. PMID: 30589224
40. Glass GV. Integrating findings: the meta-analysis of research. *Rev Res Educ* 1977;5(1):351-379. doi: 10.3102/0091732x005001351. PMID:
41. Mathur MB, VanderWeele TJ. Meta-regression methods to characterize evidence strength using meaningful-effect percentages conditional on study characteristics. *Res Synth Methods* 2021;12(6):731-749. doi: 10.1002/jrsm.1504. PMID: 34196505
42. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-1573. doi: 10.1002/sim.1187. PMID: 12111920
43. Nair AS. Publication bias - Importance of studies with negative results! *Indian J Anaesth* 2019;63(6):505-507. doi: 10.4103/ija.IJA_142_19. PMID: 31263309
44. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics* 2018;74(3):785-794. doi: 10.1111/biom.12817. PMID: 29141096

-
45. De Cassai A, Boscolo A, Zarantonello F, et al. Enhancing study quality assessment: an in-depth review of risk of bias tools for meta-analysis-a comprehensive guide for anesthesiologists. *J Anesth Analg Crit Care* 2023;3(1):44.doi: 10.1186/s44158-023-00129-z. PMID: 37932825
 46. Russo MW. How to review a meta-analysis. *Gastroenterol Hepatol (N Y)* 2007;3(8):637-642.doi. PMID: 21960873