

Model behavior: fixed ideas, random thoughts - rethinking model selection in meta-analysis

Geun Joo Choi and Hyun Kang

Department of Anesthesiology and Pain Medicine, College of Medicine, Chung-Ang University, Seoul, Korea

Received July 21, 2025; Revised September 8, 2025; Accepted September 9, 2025

Handling Editor: Francis S. Nahm

Correspondence: Hyun Kang

Department of Anesthesiology and Pain Medicine, College of Medicine, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea

Tel: +82-2-6299-2571, 2579, or 2586, Fax: +82-2-6299-2585, E-mail: roman00@naver.com

Meta-analysis is a powerful tool for synthesizing quantitative evidence from multiple studies, but its utility is shaped by a series of crucial methodological choices—none more prominent than the selection of the appropriate statistical model. A recent article by Vatvani et al. [1], published in the Korean Journal of Pain, examined the efficacy and safety of low-dose naltrexone for patients with fibromyalgia. This paper prompted a constructive discussion regarding model selection in meta-analysis. The authors employed a random-effects model, citing anticipated heterogeneity due to clinical and methodological differences among the included studies. In response, a letter to the editor raised concerns about the lack of empirical justification for this choice, advocating for the use of statistical thresholds—such as I^2 and Cochran's Q test—as criteria for model selection [2]. Hariyanto et al. [3] replied with a detailed defense of their approach, drawing upon both theoretical and practical considerations.

This discussion underscores the complex and evolving nature of statistical model selection in meta-analytic research. As editors, we welcome such scholarly engagement, which provides an opportunity to revisit foundational principles while recognizing the contextual factors that inform analytical decisions.

At the heart of this debate is a fundamental question:

What assumptions should guide the pooling of effect sizes across studies? The fixed-effect model assumes that all studies estimate the same true effect, and any observed variation is due to sampling error alone. In contrast, the random-effects model assumes that true effect sizes may differ across studies due to clinical, methodological, or contextual variability. The fixed-effect model, therefore, estimates the common effect in a specific set of studies, whereas the random-effects model estimates the average effect across a broader distribution of related studies [4].

The letter to the editor proposed a rule-based decision framework: apply a fixed-effect model when $I^2 \leq 50\%$ and $P_{\text{chi}^2} \geq 0.05$, and a random-effects model when $I^2 > 50\%$ and $P_{\text{chi}^2} < 0.05$. This approach is commonly taught in evidence-based medicine curricula and often used in practice. However, as Hariyanto et al. [3] correctly point out—and as recent methodological reviews have emphasized—relying solely on statistical thresholds may be too simplistic and, in some cases, misleading [4,5].

There are several reasons why relying exclusively on I^2 and Cochran's Q test can be problematic. First, the Cochran's Q test has limited statistical power, especially when the number of studies is small, increasing the risk of failing to detect true heterogeneity. Second, I^2 quantifies the proportion of the true effect variance to the ob-



served variance attributable to between-study heterogeneity, but it is a relative—rather than absolute—measure. It does not indicate the magnitude of heterogeneity on the original outcome scale. Moreover, I^2 often has wide and imprecise confidence intervals when few studies are available, reducing its reliability. These limitations suggest that heterogeneity statistics should inform—but not dictate—model selection.

That said, as the letter's authors rightly observe, applying empirical or even arbitrary thresholds—such as those based on I^2 —can help standardize interpretation and improve transparency [2]. While these rules may not fully capture the complexity of between-study variation, they can still serve as useful rules of thumb for clinicians, researchers, and readers aiming for consistent interpretation.

That said, as the letter's authors point out, considering both objectivity and practicality, setting empirical cutoffs for heterogeneity—such as using I^2 thresholds—can help standardize interpretation and improve transparency [2]. Although somewhat arbitrary, these thresholds may serve as useful heuristics for clinicians, researchers, and readers who seek more consistent guidance when interpreting meta-analytic findings.

The random-effects model is generally more conservative and robust in the presence of study-level variability. Notably, when the between-study variance (τ^2) is small, results from random-effects and fixed-effect models tend to converge, reducing the practical risk of misestimation. In the study by Vatvani et al. [1], the I^2 for the primary outcome was 33%, typically interpreted as low-to-moderate heterogeneity. Interestingly, the pooled estimates under both models were nearly identical [3]. This suggests that, even without applying empirical or arbitrary thresholds, the use of a random-effects model in this instance produced results that were not meaningfully different. As such, the model choice was unlikely to introduce bias or alter the clinical interpretation.

This finding supports a more flexible, context-sensitive approach to model selection. As we have previously argued, heterogeneity is an intrinsic feature of meta-analysis and should be seen not merely as a statistical complication, but as a reflection of real-world variation in clinical practice, patient populations, and methodological design [4]. Therefore, decisions regarding model choice should also take into account the presence of clinical heterogeneity, variability in study designs, and the intended scope of generalizability [4]. When studies are drawn from diverse settings involving different interventions and populations—as in the case of low-dose

naltrexone—a random-effects model often better reflects the likely distribution of effects.

To improve the characterization and communication of heterogeneity, we also recommend supplementing I^2 with additional metrics such as τ (the standard deviation of true effects), τ^2 (its variance), and prediction intervals. Unlike I^2 , τ and the prediction interval are expressed in the same units as the outcome of interest, making them more intuitive for clinical interpretation. Prediction intervals, in particular, offer insight into the expected range of effects in future studies and may be more informative than a single summary estimate for decision-making [4].

Moreover, we advocate for the use of additional tools—such as τ , τ^2 , and prediction intervals—to better characterize and communicate heterogeneity. Unlike I^2 , which expresses heterogeneity as a percentage, τ and the prediction interval are expressed in the same units as the outcome of interest, allowing for more intuitive interpretation. Prediction intervals in particular offer insights into the range of likely effects in future studies, which can be more informative for clinicians and policymakers than a single summary estimate [4,5].

Another important point highlighted by the letter writers was the use of the term “random-effect model” instead of the technically correct “random-effects model.” We concur that precise terminology is essential in scholarly writing, as even seemingly minor errors can reflect a lack of clarity or attention to important distinctions in statistical concepts. In this case, adopting the correct term not only aligns with established conventions in meta-analysis but also reinforces the conceptual distinction that the model accounts for variability in multiple effect sizes. We appreciate the authors for bringing this to attention, as such corrections enhance both the clarity and credibility of scientific communication.

In conclusion, this dialogue highlights the importance of thoughtful, transparent, and context-aware decision-making in meta-analytic research. There is no one-size-fits-all rule for model selection. Both empirical measures and conceptual reasoning should guide analytical choices. We encourage authors to prespecify their analytical strategy in a registered protocol, justify their model selection using both statistical and clinical arguments, and perform sensitivity analyses to evaluate the robustness of their results. Editors and reviewers, in turn, should assess these decisions within the broader context of the research question and the characteristics of the included studies.

We thank both the letter authors and the original investigators for contributing to this important methodological conversation. Such discussion not only enhance the

quality of individual publications but also fosters a more thoughtful and rigorous scientific community.

Hyun Kang, <https://orcid.org/0000-0003-2844-5880>

DATA AVAILABILITY

Data sharing is not applicable to this article as no datasets were generated or analyzed for this paper.

CONFLICT OF INTEREST

Hyun Kang is a section editor of the Korean Journal of Pain. However, he was not involved in the selection of peer reviewers, the evaluation, or the decision-making process for this article.

FUNDING

No funding to declare.

AUTHOR CONTRIBUTIONS

Geun Joo Choi: Writing/manuscript preparation; Hyun Kang: Writing/manuscript preparation.

ORCID

Geun Joo Choi, <https://orcid.org/0000-0002-4653-4193>

REFERENCES

1. Vatvani AD, Patel P, Hariyanto TI, Yanto TA. Efficacy and safety of low-dose naltrexone for the management of fibromyalgia: a systematic review and meta-analysis of randomized controlled trials with trial sequential analysis. *Korean J Pain* 2024; 37: 367-78.
2. Doan HN, Kwak SG, Chang MC. Comments on "Efficacy and safety of low-dose naltrexone for the management of fibromyalgia: a systematic review and meta-analysis of randomized controlled trials with trial sequential analysis". *Korean J Pain* 2025; 38: 493-4.
3. Hariyanto TI, Vatvani AD, Patel P, Yanto TA. Authors' response to comments on "Efficacy and safety of low-dose naltrexone for the management of fibromyalgia: a systematic review and meta-analysis of randomized controlled trials with trial sequential analysis". *Korean J Pain* 2025; 38: 495-8.
4. Choi GJ, Kang H. Heterogeneity in meta-analyses: an unavoidable challenge worth exploring. *Korean J Anesthesiol* 2025; 78: 301-14.
5. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010; 1: 97-111.