

## JAMA Guide to Statistics and Methods

# Random-Effects Meta-analysis

## Summarizing Evidence With Caveats

Stylianios Serghiou, MD; Steven N. Goodman, MD, PhD

**Questions involving medical therapies** are often studied more than once. For example, numerous clinical trials have been conducted comparing opioids with placebos or nonopioid analgesics in the treatment of chronic pain. In the December 18, 2018, issue of *JAMA*, Busse et al<sup>1</sup> evaluated the evidence on opioid efficacy from 96 randomized clinical trials and, as part of that work, used random-effects meta-analysis to synthesize results from 42 randomized clinical trials on the difference in pain reduction among patients taking opioids vs placebo using a 10-cm visual analog scale (Figure 2 in Busse et al).<sup>1</sup> Meta-analysis is the process of quantitatively combining study results into a single summary estimate and is a foundational tool for evidence-based medicine. Random-effects meta-analysis is the most common approach.

### Why Is Random-Effects Meta-analysis Used?

Each study evaluating the effect of a treatment provides its own answer in terms of an observed or estimated effect size. Opioids reduced pain by 0.54 cm more than placebo on a visual analog scale in 1 study<sup>2</sup>; this was the observed effect size and represents the best estimate from that study of the true opioid effect. The true effect is the underlying benefit of opioid treatment if it could be measured perfectly, and is a single value that cannot directly be known.

If a particular study was replicated with new patients in the same setting multiple times, the observed treatment effects would vary by chance even though the true effects would be the same in each. The belief that the true effect was the same in each study is called the fixed-effect assumption, whereby the fixed effect is the common, unknown true effect underlying each replication. A meta-analysis making the fixed-effect assumption is called a fixed-effect meta-analysis. The corresponding fixed-effect estimate of the treatment effect is a weighted average of the individual study estimates and is always more precise (ie, it has a narrower confidence interval [CI] than that of any individual study, making the estimate appear closer to the true value than any individual study).

However, medical studies addressing the same question are typically not exact replications and they can use different types of medication or interventions for different amounts of time, at different intensities, within different populations, and have differently measured outcomes.<sup>3</sup> Differences in study characteristics reduce the confidence that each study is actually estimating the same true effect. The alternative assumption is that the true effects being estimated are different from each other or heterogeneous. In statistical jargon, this is called the random-effects assumption. The plural in effects implies there is more than 1 true effect and random implies that the reasons the true effects differ are unknown.

A random-effects assumption is less restrictive than a fixed-effect assumption and reflects the variation or heterogeneity in the true effects estimated by each trial. This usually results in a more realistic estimate of the uncertainty in the overall treatment effect with

larger CIs than would be obtained if a fixed effect was assumed. A random-effects model can also be used to provide differing, study-specific estimates of the treatment effect in each trial, something that cannot be done under the fixed-effect assumption.

### Description of Random-Effects Meta-analysis

In a random-effects meta-analysis, the statistical model estimates multiple parameters. First, the model estimates a separate treatment effect for each trial, representing the estimate of the true effect for the trial. The assumption that the true effects can vary from trial to trial is the foundation for a random-effects meta-analysis. Second, the model estimates an overall treatment effect, representing an average of the true effects over the group of studies included. Third, the model estimates the variability or degree of heterogeneity in the true treatment effects across trials. Compared with a fixed-effect estimate, the random-effects estimate for the overall effect is more influenced by smaller studies and has a wider CI, reflecting not just the chance variation that is reflected in a fixed-effect estimate, but also the variation among the true effects.<sup>4</sup> In the report by Busse et al,<sup>1</sup> the random-effects average opioid benefit was -0.69 cm (95% CI, -0.82 to -0.56 cm).

Whether the variability observed in the estimates of treatment effect is consistent with chance variation alone is reflected in statistical measures of heterogeneity, often expressed as an  $I^2$ , the percentage of total variation in the random-effects estimate due to heterogeneity in the true underlying treatment effects. An  $I^2$  value greater than 50% to 75% is considered large.<sup>5</sup> Busse et al<sup>1</sup> report an  $I^2$  of 70.4%, reflecting the marked variation among studies, which is also demonstrated by nonoverlapping CIs around some individual treatment estimates.

A more natural heterogeneity measure is the standard deviation of the true effects, often denoted as  $\tau$ . A  $\tau$  of 0.35 cm can be derived from the data in Figure 2 in the article by Busse et al.<sup>1</sup> Given the overall random-effects estimate of -0.69 cm, this means that the true effects in individual studies could vary over the range of -0.69 cm  $\pm$  2  $\tau$  or -1.39 cm to 0.01 cm, namely a true benefit in some studies roughly twice as large as the average and no benefit in some others. This reflects the display provided in the study by Busse et al<sup>1</sup> in which 10 of 42 studies estimated a benefit larger than 1 cm, which was the minimum clinically importance difference. Quantifying the variability in treatment effects among studies helps readers decide whether combining these results makes sense. Like the proverbial person said to be at normal average temperature with 1 foot in ice and the other in boiling water, the estimated average effect can be nonsensical if the true individual study effects are too variable.

### Why Did the Authors Use Random-Effects Meta-analysis?

Meta-analyses incorporate some uncertainties that mathematical summaries cannot reflect. A sensible approach is to use the

statistical method least likely to overstate certainty almost regardless of perceptions or philosophy about true effects being fixed or random, which is why random-effects models are a frequent choice in meta-analyses.

The studies in the report by Busse et al<sup>1</sup> demonstrate substantial variability, being both qualitatively and quantitatively different. Tominaga et al<sup>6</sup> examined the effect of tapentadol extended-release tablets in Japanese patients with either chronic osteoarthritic knee pain or lower back pain, whereas Simpson and Włodarczyk<sup>7</sup> examined the effect of transdermal buprenorphine in Australian patients with diabetic peripheral neuropathic pain. These studies used different opioids to treat different sources of pain in culturally different populations that may assess pain differently; indeed, the variability in observed effects between the 2 studies suggests that the differences seen are probably beyond chance variation.

Taken together, these features provide substantial evidence that these studies are not examining the same effect, consistent with the random-effects assumption. Thus, a fixed effect is not plausible and a random-effects meta-analysis is the appropriate method.

### What Are Limitations of a Random-Effects Meta-analysis?

First, the random-effects model does not explain heterogeneity, it merely incorporates it. The standard recommendation is that researchers should attempt to reduce heterogeneity<sup>8,9</sup> by using subgroups of studies or a meta-regression; however, such methods represent exploratory data-dependent exercises and their results must be interpreted accordingly.

Second, there are many approaches to calculating the random-effects estimates. Although most produce similar estimates, the DerSimonian-Laird method is the most widely used and it produces CIs that are too narrow and *P* values that are too small when there are few studies (<10-15) and sizable heterogeneity; accordingly, this approach is not optimal in the setting of few studies and high heterogeneity and often may be contraindicated.<sup>10</sup>

Third, small studies more strongly influence estimates from random-effects than from fixed-effect models; in fact, the larger the heterogeneity, the larger their relative influence. If smaller studies are judged as more likely to be biased, this can be a substantial concern.

### What Are the Caveats to Consider When Assessing the Results of a Random-Effects Meta-analysis?

The overall summary effect of a random-effects meta-analysis is representative of the study-specific true effects without the estimate representing a true effect (ie, there may be no population of patients or interventions for which this summary value is true). This is why a random-effects meta-analysis should be interpreted with consideration of the qualitative and quantitative heterogeneity, particularly the range of effects calculated using  $\pm 2 \times \tau$ . If the range is too broad, or if *I*<sup>2</sup> exceeds 50% to 75%, the meta-analytic estimate might be too unrepresentative of the underlying effects, potentially obscuring important differences. Busse et al<sup>1</sup> assessed qualitative heterogeneity partly through their assessment of directness and judged it to be minimal, although that may not capture every dimension of importance.

Other caveats apply to all meta-analyses and include whether the analyses include all relevant studies, whether the studies are representative of the population of interest, whether study exclusions are justified, and whether study quality was adequately assessed. Sometimes heterogeneity reflects a mixture of diverse biases, with few of the studies properly estimating even their own true effects. Busse et al<sup>1</sup> addressed this by using a risk-of-bias assessment.

### How Should the Results of a Random-Effects Meta-analysis Be Interpreted in This Particular Study?

Three of 8 meta-analyses reported in Busse et al<sup>1</sup> (using 7 different outcome measures) have an *I*<sup>2</sup> of 50% or greater, and the meta-analysis we have been discussing that used the outcome of pain has an *I*<sup>2</sup> of 70.4%, reflecting conflict or a high degree of variability among studies. The meta-analysis by Busse et al<sup>1</sup> provided strong evidence against opioids increasing pain and suggested that opioids are generally likely to reduce chronic noncancer pain by a modest 0.69 cm more than placebo (less than the 1 cm minimum clinically important difference). However, in view of the amount of heterogeneity, it is possible that in some settings and patients, the benefit of opioids could be lesser or greater than this random-effects estimate. As such, physicians should consider this summary result with caution, and in conjunction with the effect from the subset of studies most relevant to the patients they need to treat.

#### ARTICLE INFORMATION

**Author Affiliations:** Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California (Serghiou, Goodman); Meta-research Innovation Center at Stanford, Stanford, California (Serghiou, Goodman); Department of Medicine, Stanford University School of Medicine, Stanford, California (Goodman).

**Corresponding Author:** Steven N. Goodman, MD, PhD, 150 Governor's Ln, Stanford, CA 94305 (steve.goodman@stanford.edu).

**Section Editors:** Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, JAMA.

**Published Online:** December 19, 2018. doi:10.1001/jama.2018.19684

**Conflict of Interest Disclosures:** None reported.

#### REFERENCES

- Busse JW, Wang L, Kamaleldin M, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. *JAMA*. 2018;320(23):2448-2460. doi:10.1001/jama.2018.18472
- Wen W, Sitar S, Lynch SY, et al. A multicenter, randomized, double-blind, placebo-controlled trial to assess the efficacy and safety of single-entity, once-daily hydrocodone tablets in patients with uncontrolled moderate to severe chronic low back pain. *Expert Opin Pharmacother*. 2015;16(11):1593-1606.
- Serghiou S, Patel CJ, Tan YY, et al. Field-wide meta-analyses of observational associations can map selective availability of risk factors and the impact of model specifications. *J Clin Epidemiol*. 2016;71:58-67.
- Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health*. 2014;17(2):53-57.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.
- Tominaga Y, Koga H, Uchida N, et al. Methodological issues in conducting pilot trials in chronic pain as randomized, double-blind, placebo-controlled studies. *Drug Res (Stuttg)*. 2016;66(7):363-370. doi:10.1055/s-0042-107669
- Simpson RW, Włodarczyk JH. Transdermal buprenorphine relieves neuropathic pain. *Diabetes Care*. 2016;39(9):1493-1500. doi:10.2337/dc16-0123
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis. *Stat Med*. 1999;18(20):2693-2708.
- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
- Cornell JE, Mulrow CD, Localio R, et al. Random-effects meta-analysis of inconsistent effects. *Ann Intern Med*. 2014;160(4):267-270.