

Common mistakes in Meta-Analysis and How to Avoid Them

Fixed-effect vs. Random-effects

INTRODUCTION

There are two popular statistical models for meta-analysis, the fixed-effect model and the random-effects model.

Under the *fixed*-effect model we assume that there is one *true* effect size that underlies all the studies in the analysis, and that *all* differences in *observed* effects are due to sampling error. While we follow the practice of calling this a fixed-effect model, a more descriptive term would be a *common*-effect model. In either case, we use the singular (*effect*) since there is only one true effect.

By contrast, under the random-effects model we allow that the *true* effect size might differ from study to study. For example, the effect size might be higher (or lower) in studies where the participants are older, or more educated, or healthier than in other studies, or when a more intensive variant of an intervention is used. The term “Random” reflects the fact that the studies included in the analysis are assumed to be a random sample of all possible studies that meet the inclusion criteria for the review. And we use the plural (*effects*) since we are working with multiple true effects.

To understand the difference between the models, consider a hypothetical case where the sample size in each study was extremely large, so that the sampling error was trivial. For all intents and purposes, we’d be looking at the *true* effect size for each study. Under the fixed-effect model, these effects would all be essentially identical. Under the random-effects model, these effects would still vary.

HOW SHOULD WE SELECT A STATISTICAL MODEL?

The choice of a statistical model should depend on the sampling frame that was used to select studies for the analysis. If we are working with one population, then we should use the fixed-effect model. If we are working with a universe of populations, we should use the random-effects model. Consider the following two cases.

Case 1. A pharmaceutical company plans to enroll 1,000 patients. Because the staff can only work with 100 patients at time, it randomly divides the patients into 10 groups of 100 patients, and runs the identical study with each. We know all studies are based on the same population, since that’s how we selected them. The fixed-effect model matches the way the studies were sampled.

Case 2. We locate a series of published studies that were performed by different people in different locations at different times. While the studies all looked at similar interventions, it stands to reason that the true impact of this intervention will differ from study to study. If the studies were conducted in different hospitals it’s likely that the patient population (age, co-morbid diseases) varied from one hospital to the next, and that the intervention was therefore more effective in some hospitals than in others. It’s also possible that the intervention itself (the precise dosage, length of follow-up) differed from one hospital to the next and that this could have an impact on the effect size. While the difference

in effect-size from one hospital to the next could be small, it was probably not zero. And once there is *any* difference, the random-effects model is the model that fits the data.

The situation depicted in the first case, where all studies are sampled from the same population, is relatively rare. In the overwhelming majority of meta-analyses, the sampling frame is similar to that depicted in the second case.

THE MISTAKE TO AVOID

Some researchers start the analysis by selecting the fixed-effect model. They then test perform a statistical test for heterogeneity in effect sizes (the Q-test).

- If the test for heterogeneity *is not* statistically significant, they conclude that the fixed-effect model is consistent with the data, and use this model in the analysis.
- If the test for heterogeneity *is* statistically significant they conclude that the fixed-effect model is not consistent with the data, and use the random-effects model in the analysis.

This approach is fundamentally flawed for two reasons.

Reason 1

If we want to choose a model based on the sampling frame, then we should choose the model based on our understanding of how the studies were sampled, and not the results of a statistical test. If we are working with studies that assess the impact of an intervention in different populations then logic tells us that the random-effects model is the model that fits the data, and it's the model that we should choose.

To suggest that a non-significant p -value justifies the use a fixed-effect analysis is to suggest that the lack of significance *proves* that the null is correct (that the studies all share a common effect size). As we all learned in our first statistics class, the lack of significance does *not* prove that the null is true. And here, logic tells us that the null is probably false.

Reason 2.

The “flawed” approach uses the fixed-effect model as the starting point, and requires evidence (a significant test of heterogeneity) to shift to the random-effects model.

In fact, the *random-effects* model should be the logical starting point. The random-effects model says that the true effect size *may or may not* vary from study to study, and thus does not assume that either is the case. As part of the analysis we estimate the amount of variance in true effects across studies, and the estimate may or may not be zero.

By contrast, the fixed-effect model *requires* that the true effect size *does not* vary from study to study. Therefore, the fixed-effect model is more restrictive. It imposes a constraint that is neither necessary nor plausible.

WHY DOES IT MATTER WHICH MODEL WE USE?

If we should be using the random-effects model and (by mistake) employ the fixed-effect model, then it's likely that

- The estimate of the mean will be incorrect
- The standard error will be incorrect
- The test of significance for the mean will be incorrect
- The confidence interval about the mean effect will be too narrow

More fundamentally, the choice of a model defines the goals of the analysis.

The choice of a model determines the meaning of the summary effect

- Under the fixed-effect model there is only one true effect. The summary effect is an estimate of that value.
- Under the random-effects model there is a distribution of true effects. The summary effect is an estimate of that distribution's *mean*.

One of the most important goals of a meta-analysis is to determine how the effect size varies across studies.

- When we use the fixed-effect model we can estimate the common effect size but we cannot discuss how the effect size varies, since this model assumes that the true effect size is the same in all studies.
- By contrast, if we elect to work with the random-effects model, we can ask not only "What is the mean effect size" but also "How does the effect size vary across populations." In many cases, this question is key to understanding the effectiveness of the intervention.

IN SUM,

The selection of the correct statistical model is critically important.

- We should choose the model that fits the sampling frame.
- We should not choose a model based on the statistical test for heterogeneity.