Meta-Analysis of Two Studies With Random Effects?

Ding-Geng Chen,

BA, MS, PhD University of North Carolina-Chapel Hill Chapel Hill, North Carolina

University of Pretoria, South Africa

Di Fang,

BA, PhD University of Arkansas, Arkansas

Jeffrey R. Wilson*,

PhD* Arizona State University, Arizona

*Corresponding author: Jeffrey R. Wilson, PhD Arizona State University, AZ.

The authors declare that they have no conflict of interest

A meta-analysis (MA) combines similar studies resulting in a larger number of subjects to improve the degree of belief in the significance declared. Its major purpose is to increase the number of observations and the associated statistical power, thereby increasing the precision for the estimates of the effect size as it relates to an association or an intervention. As commonly known, there are discrepancies between MAs and the large randomized clinical trials. The conclusions drawn are subject to bias because they are affected by the small size of clinical studies. However, large randomized clinical trials are the most reliable way of obtaining reproducible results; in other words, we expect the same results if we repeated the experiment. On the other hand, large trials do not guarantee that the protocol or the conclusions were appropriate. Although it is intuitive to believe an MA of similar trials is more likely to result in valid conclusions, studies show this is not always the case. By the same argument, adding studies with diverse protocols makes an MA less reliable. Because an MA is a summation, its reliability depends on the combined trials. Inclusion/exclusion criteria, conclusions, reliability of the results, and applicability for the conclusions affect the bias. Hence, we cannot declare that MA represents the final and accurate viewpoint on an area of research. Several statistical methods similar to what have been used to perform analyses on individual subject data have been modified to improve the reliability of MA.

There are 2 basic models used in MA: fixed effects and random effects. The fixed-effects model assumes that the effect sizes from all studies are the same and any variation is due to random error from each study (i.e., within-study variation) with no between-study variation. This implies that when an MA is conducted on these studies, a higher weighting is given to the larger studies and smaller weighting is assigned to smaller studies; because we have the same effect size in the larger studies, we have access to more reliable information. However, if the

observed effect size varies from 1 study to the next, then the fixed-effects model is not appropriate and the random-effects model should be used.

On the other hand, the random-effects model assumes the studies were taken from a population with varying effect sizes, which signifies the between-study variation. Therefore, the random-effect MA model incorporates both between-study variation and within-study variation. As explained by Borenstein et al [1] and Chen and Peace [2], the estimate from random-effects MA is a global estimate of the effect size from all studies that represents the average of a distribution of all values of effect sizes.

Incorporating the between-study variation, the total variation for the estimated global effect size under the random- effects MA model is typically larger than the estimate under the fixed-effects model. Whenever the between-study variation is statistically significantly greater than 0, the estimated variance, standard error, and confidence interval for the global effect size will be larger under random-effects as opposed to fixed-effects MAs.

The choice of MA models depends on the goals and purpose of the study. By definition, the fixed-effects model is a great choice if we believe all the studies are homogeneous without ignorable between-study variation and identical purpose of studies, and our goal is to compute the average effect size. On the other hand, if the goal of the analysis is to generalize to a set of populations, one needs the random-effects model. In such cases one has data from a series of studies performed by different researchers. It is expected that all the studies will have varying protocols that affect the results differently, and therefore we should not assume a common effect size. In addition, however, if the number of studies is very small, then there is no way to obtain estimates of the between-study variation with any degree of precision. In such a case, the fixed-effects model may be the natural choice. As such, we must resort to treating the included studies as the only studies of interest.

Some researchers suggest it is good practice to start with the fixed-effects model and then move to a random-effects model based on certain statistical significance. However, Borenstein et al [1] and Chen and Peace [2] reported that such practice should be discouraged. However, if the study effect sizes came from a distribution of effect sizes, then the random-effects model is the logical one to use. However, if the between-study dispersion turns out to be insignificant among studies, then the random-effects model reduces to the fixed-effects model. Hence, there is no real reason to avoid starting with the random-effects model. Statistically, a Q statistic used with the random-effects model allows one to quantify the between-study heterogeneity with an associated C^2 distribution to test the statistical significance of the between-study variation. However, if the number of studies is small and the within-studies variance is large, the Q statistic based on the c^2 test may have low power even if the between-study variance is substantial. In this case, using the Q statistic as a criterion for selecting the model is problematic, because the model selection is based on the conceptual perspective and realistic goals of the studies.

This discussion then leads to the MA reported in Pundir et al [3]. According to the description, the authors did an excellent job searching the most common databases, such as MEDLINE and EMBASE, with additional searches to conference proceedings and archived randomized controlled trials. Unfortunately, within 502 studies only 2 can be used for quantitatively MA. The authors then proceeded to fit an MA with a random-effects model and made conclusions that the excision group had a significantly greater reduction in symptoms of dysmenorrhea (MD, .99; 95% confidence interval, —.02 to 2.00; p = .05) along with other outcomes. This is obviously not correct because the 95% confidence interval is [—.002 to 2.00], which covered 0, and therefore this is not statistically significant at 5% level, more so because the authors reported a p = .05. We took the data and redid an MA using the random-effects MA model. In fact, we obtained a p = .0544 and suspected the authors rounded to .05.

Because there are only 2 studies, it would be difficult in statistical MA to justify using a random-effects meta-model to estimate the between-study heterogeneity. In addition, all the c^2 tests based on the Q statistic reported in the study were not statistically significant, which would lead one to use a fixed-effects MA to increase the statistical power. Therefore, we highly recommend the authors reconfigure the analyses using a fixed-effects MA model. In fact, we performed the fixed-effects MA for the outcome of reduction in symptoms of dysmenorrhea and found that the excision group had a significantly greater reduction in symptoms of dysmenorrhea where the MD is .994 (standard error, .504; p = .0485) with 95% confidence interval [.007-1.982].

References

1. Borenstein M, Hedges LV, Higgins JPT, et al. *Introduction to Meta-Analysis.* Wiley and Sons; 2009.
2. Chen DG, Peace KE. *Applied Meta-Analysis with R. Chapman and Hall/ CRC Biostatistics Series;* 2013.
3. Pundir J, Omanwa K, Kovoor E, et al. Laparoscopic excision versus ablation for endometriosis-associated pain-updated systematic review and meta-analysis. *J Minim Invasive Gynecol.* 2017;24:747-756.