Traditional Weighted Random Effects Meta-Analysis Practice Must Change.

Meta-analysis of collections of randomized clinical trials is at the apex of most evidence pyramids. You can Google "evidence pyramid" to see this. The first link below is a peer reviewed paper that tells you that weighted random effects meta-analysis is not based on sound statistical principles and offers a classical survey sampling approach based on ratio estimation as a rigorous alternative. It is distribution-free, not relying on the truth of any statistical model. The second link provides free well-documented and user-friendly Excel templates for convenient analysis. That paper should be read by all consulting biostatisticians. This is a major public health issue as 800 such meta-analyses of collections of randomized clinical trials are published in peer review journals every year. These are major contributors to public health policy. It is **dangerous** to conduct business as usual, as the mainstream will generally use random effects methods with weights that are inversely proportional to purported estimated variance. The problem is that the mainstream presumes incorrectly (see "here is the problem" below) that the weights are constants at least to a strong degree of approximation. The fact is they are seriously random variables, and this leads to the potential for bias and incorrect standard error of the main effect size estimate. This is proven beyond any doubt in the reference linked below. In addition, a clinical and translational research paper, that is in-press, proves this in another way, at a level easily understood by non-statisticians. It should be noted that the reference linked below obtained review material from nine sources, three of whom were world-renowned meta-analysts. One of these would actually lose money if practice changed. All nine agreed that the weights are seriously random variables. While I cannot yet give you a preprint, the in-press paper that intends to change research practice, has been accepted for publication by the Journal of Clinical and Translational Research. Finally, in part 2 below, we go over the five assumptions inherent in the mainstream (weights inversely proportional in the estimated variance) and demonstrate that Assumptions A1 and A2 ensure that Assumption A4 is false, making their assumptions incompatible with each other!

There are three crucial factors that are of interest in this in-press paper.

1. To derive their statistical properties, the mainstream uses "off-label" statistical methods. Therefore, these lack a rigorous evidence basis.
2. When it comes to analyzing a set of clinical trials for relative risk, a review of a sample of 31 meta-analyses by our ratio estimation method and by the published mainstream method showed significantly shorter confidence intervals in the log scale by our ratio methods. This even accepts the warts of the mainstream methods. **Ratio estimation methods proved to be both more rigorous and more precise.**
3. The paper also provides two published examples where had our ratio estimates been available and used, incorrect public health conclusions, affecting hundreds of thousands of patients, could have been averted. Since these were both published over 10 years ago, it was not the fault of the research groups that these failures occurred. But the next potential disaster. where publication of a similar failure of the mainstream occurs, would be on the analyst's shoulders, assuming s/he knew about my publications and chose to

ignore them. From my analysis of prior papers, I would project about 25% of mainstream random-effects meta-analyses of collections of randomized clinical trials lead to incorrect qualitative or quantitative conclusions.

Here is the problem:

The global estimator of the mean effect size θ is $\hat{\theta}$, defined by

$$\hat{\theta}=\Sigma\ W_j\hat{\theta}_j \qquad (1)$$

With $\hat{\theta}_j$, the estimated effect size for study j and the weights W_j satisfying $\Sigma W_j=1$. The estimated variance of $\hat{\theta}$ is claimed to be

$$\mathrm{Var}(\hat{\theta})=\Sigma W_j^{\,2}\,\mathrm{Var}(\hat{\theta}_j) \qquad (2)$$

$$E(\hat{\theta})=\Sigma\ W_j\ E(\hat{\theta}_j) \quad = \quad \theta \qquad (3)$$

Formula (3) cannot be legitimately applied when the weights are seriously random variables. As noted in the article linked below, we can quantify the bias. There is no adequately powered diagnostic test to preclude correlation between weights and estimates for a study. Worse, the randomness of the weights makes formula (2) mathematically incorrect.

Part 2: Incompatibility of Mainstream Assumptions

<u>Here is a review of the mainstream model assumptions:</u>

A1: The true primary effect sizes for each study are drawn independently from a single large "urn" of primary effect sizes. **This assumption tells us we are targeting the unweighted mean of all studies in the urn, θ.**

A2: The true primary effect sizes in the urn, θ_j, follow a normal distribution whose unweighted mean is the target parameter of interest, θ and whose variance is σ^2.

A3: The individual study provides an unbiased estimate of its study-specific true primary effect size, $\hat{\theta}_j$ and has an approximate normal distribution about its true primary effect size, θ_j.

A4: **Up to a strong approximation, the weights are "constants" rather than seriously random variables. In other words, if you repeat the total experiment under the same assumptions A1-A3 and same urn, this assumption presumes you obtain identical weights up to a strong approximation**. **This assumption is mandatory to use the above formulas for the mean and variance in the mainstream methods but will be shown to be false under assumptions A1-A3.** More on this below.

A5: There is no association between study weights, W_j, and true study effect sizes, θ_j. For example, if big studies tend to have higher (lower) effect sizes than smaller studies, the method will tend to overestimate (underestimate) respectively the overall effect size. This could lead to unacceptable bias.

How Mainstream Weighted Random Effects Methods Work

The "variance" for the estimates of effect size for each study consists of two components, the reasons why its individual study estimate of effect size differs from the true mean of the effect sizes in the urn: (a) Within study variance, which is estimated under assumption A3 and (b) Between study variance, which is the variance of the true effect sizes in the urn, per assumptions A1 and A2. The first component (a) depends on the accuracy of the within study estimate and varies from study to study. The second component (b) is the same for all studies. The overall estimate is the weighted average of the individual study estimates with weights inversely

proportional to the study's estimated variance, which is the sum of the estimated within study variance and the estimated between study variance. If all five assumptions were true, these weights would minimize the standard error (square root of the overall variance) of the estimate of the overall effect size, over all choices of weights that sum to one (a requirement for unbiasedness). Note that all other things being equal, larger between study variance pushes the weights closer to equal weights and smaller between study weights pushes the weights closer to fixed effects.

Why Assumption A4 is false.

What this assumption requires is that if we repeat the experiment under Assumptions A1-A3, the resulting weights will be the same up to a strong approximation.

Imagine a Meta-Analysis where we independently generate the data twice under Assumptions A1-A3. Clearly, the true study effect sizes for these two repetitions are sure to differ. It follows that the diversity (sample variance) of these true effect sizes will differ. All things being equal, the one with the greater sample variance in true effect sizes will have weights closer to equality than the other, thanks to a larger between study variance.

If someone "leaked" the sample variances of the true θ_j, in each replication, S_1^2 and S_2^2, we would certainly prefer to use them as the optimal (minimum variance unbiased estimators) of σ^2 in the two respective meta-analyses. These estimators would be superior to the meta-analysis derived estimators. The ratio of these two estimates is less volatile than the ratio of the two meta-analysis estimators. Are the two estimators of between study variance virtually the same? The ratio has an F-distribution with degrees of freedom M-1 and M-1 where M is the number of studies being combined. The impact on the between study variance is huge. For example, with

eight studies being combined, the larger sample variance would be at least 50% larger than the smaller with probability 61%. If Assumption A4 is true, these two replications should yield virtually the same sample variances.

Links:

Published Paper:

Meta-Analysis 2020: A Dire Alert and a Fix (juniperpublishers.com)

Link to Excel Templates for Ratio estimate Analysis. It does not offer forest plots but Comprehensive meta-analysis software + edit picture in Word can get you there.

This is user friendly, quality-assured software designed for clinicians. There is a user's guide, an excel template for relative risk and an excel template for differences in means or proportions. It can accommodate up to 50 studies.

https://biostat.ufl.edu/research/faculty-developed-software/

If you have any questions or comments, please contact me at shusterj@ufl.edu

Please indicate whether private or public, as I intend to share the public ones with the bloggers.

I will blog the link to the new article when published.

Best wishes,

Jon