A medical writer's guide to meta-analysis

Adam Jacobs
Premier Research, Wokingham, UK

Correspondence to:
Adam Jacobs
Senior Principal Statistician
Premier Research
Fishponds Road
Wokingham
RG41 2GY
UK
adam.jacobs@premier-research.com

Abstract
Meta-analysis is a statistical technique for summarising the results of multiple studies in a quantitative manner. It should not be confused with a systematic review, though in practice the two are often found together. The main pitfalls with meta-analyses are being sure that the studies being combined are similar enough that it makes sense to combine them, and being sure that all relevant studies have been included.

Meta-analysis is a statistical technique for combining the results of more than one study. It should be immediately obvious how useful this is: it is very rare that a single study gives us a definitive answer in medicine. To get a good idea of whether an intervention works to treat or prevent disease, or whether a particular environmental factor is associated with an increased risk of disease, for example, it is frequently necessary to take account of many studies to get a better overall picture.

By combining studies in this way, not only can we reduce the risk of being fooled by a study with unusual results as a result of a statistical fluke or bad study design, we can also get more precise estimates of the magnitude of effects. It is entirely possible, for example, that several individual studies have looked at a particular intervention but been underpowered to detect its effects, and each of them alone failed to find a significant effect, but if you combine all the studies in a meta-analysis you could find that the overall result is that a statistically significant effect can be confirmed.

Meta-analysis should not be confused with systematic review, although the two often go together. A systematic review is an attempt to find and review the entirety of literature on a particular topic using a thorough literature search, often looking for unpublished as well as published studies. This guards against any cherry-picking (at least in theory) and ensures that decisions are made on the totality of evidence.

Often, a systematic review will include a meta-analysis. Once all the relevant studies have been identified, their results can be combined using a meta-analysis to give a numerical summary. However, it is possible to do a systematic review without a meta-analysis: typically, results will be presented in narrative form with no attempt made to produce a precise numerical summary of the results. This might be done, for example, if all the studies identified had such different methods, interventions, or study populations that trying to combine them into a single estimate does not make sense.

Equally, it is possible to do a meta-

analysis without a systematic review. Sometimes studies may be chosen in a non-systematic way and yet still combined in a meta-analysis. Obviously when interpreting the results of such an analysis it is important to ask questions about what other studies might exist and why they were not included, but there may sometimes be legitimate reasons for meta-analysis of data that have not been chosen through the methods of systematic review.

One of the most important decisions for the meta-analyst is when it makes sense to combine data and when it doesn't.

I say that "in theory" a systematic review guards against cherry-picking, but in practice a systematic review is not an absolute guarantee. An important process in a systematic review is setting the inclusion criteria for the studies that will be included. There are no hard and fast rules about what inclusion criteria should be, and some judgement is always required. For example, do you require a minimum sample size for each study, and if so, what size? Will you include just trials against placebo or also trials against active comparators? Will you only include randomised trials or will you also include observational research? Should there be a minimum study duration? Will you include studies on all patients with cancer, all patients with advanced cancer, or only on those with confirmed metastatic disease? The possibilities are endless, and there are no right answers: the best choice will depend very much on individual circumstances.

And here is the problem. If you know the literature in a particular area well – as many systematic reviewers do – you will know what the important studies are. You will therefore know, when you decide on your inclusion criteria, that a particular choice of inclusion criteria will exclude specific studies that you already know about. If you have an agenda, then you can still cherry pick your data subtly by choosing inclusion criteria to exclude the studies that you don't like. So just because a systematic review has been conducted thoroughly and scrupulously in accordance with its inclusion criteria, there is still no guarantee that all relevant trials have been included. It's always worth reading the inclusion criteria carefully and making your own mind up about how reasonable they are.

One of the most important decisions for the meta-analyst is when it makes sense to combine data and when it doesn't. By combining a wide range of studies you can get apparently more statistically precise estimates, as you have more data. However, that statistical precision may be illusory. If you are investigating the efficacy of a particular treatment in different study populations, for example, an overall estimate may conceal the fact that the treatment works really well in some patients and is harmful in others. So when looking at a meta-analysis it is always worth looking at the detail of the individual studies and asking if they are investigating the same thing. If they are not, then an overall estimate may be meaningless.

Happily, this question of how comparable different studies are can be investigated statistically. A good meta-analyst will look for a measure of heterogeneity among the studies. It is expected that not all studies will give exactly the same result just because of normal random variation, but do the studies vary more than would be expected by chance? That's a simple question to ask, though not so simple to answer. Although it is possible to calculate a simple statistical test and calculate a P value, where a significant P value shows significant heterogeneity, the results of such a test are not straightforward to interpret, as there is a high risk of both false positive and false negative conclusions.

Higgins *et al.*[1] have proposed an alternative approach to quantifying heterogeneity, by calculating a measure known as the I^2 statistic, where 0 means that the studies are all identical and higher values (with a maximum of 100%) show increasing heterogeneity.

If you observe substantial heterogeneity, then it is reasonable to question the relevance of an overall estimate.

If you are looking at meta-analysis results you will come across things called "fixed effects estimates" and "random effects estimates". These are alternative statistical approaches for combining multiple studies, and are based on different assumptions.

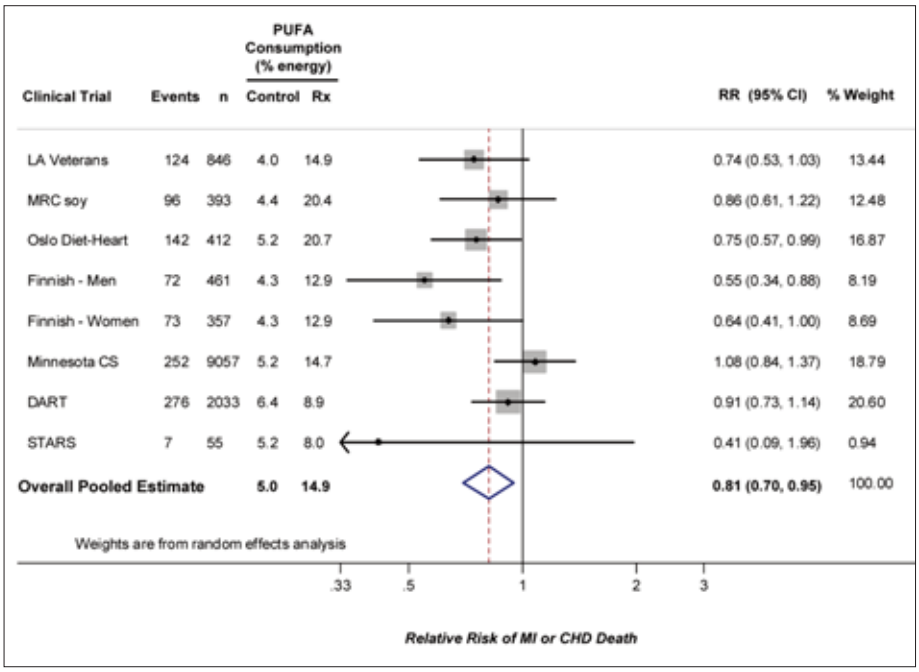
The fixed effects method makes the assumption that there is no important heterogeneity, and that all studies are essentially measuring the same thing. In other words, it assumes that any differences in estimates of treatment effects from one study to the next are due purely to statistical random variability. If in fact you observe that heterogeneity is low, then the fixed effect measure gives you a good summary of the results.

The random effects method assumes that heterogeneity is present, and the differences among studies are due partly to statistical random variability, but also due to differences in the "true" treatment effect that each study is measuring, as it is not assumed that all studies are measuring the same thing. In practical terms, the main difference between the two methods is that random effects estimation gives more weight to small studies that give different results to the average effect.

Interpreting the results of random effects meta-analyses is, as mentioned above, difficult. Although it gives you an estimate of the average effect, that treatment effect may depend on specific characteristics of the studies. If you want to apply the results to a real life situation, there is no guarantee that you will be applying it in an average situation. Your situation may match some studies far better than others.

If you observe substantial heterogeneity, then it is reasonable to question the relevance of an overall estimate.

For example, some studies may have used different doses. You may find that the high dose studies give greater treatment effects



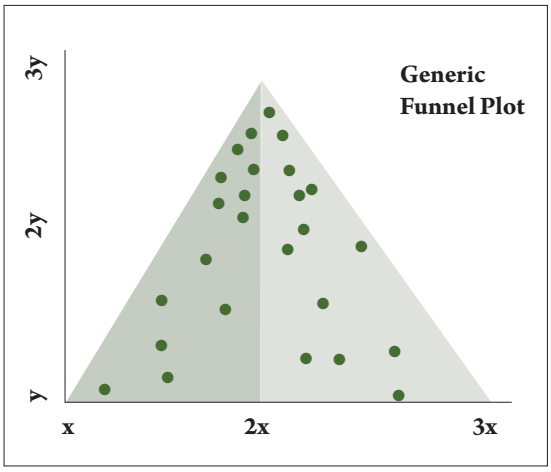
Source: Mozaffarian et al.[4]

Figure 1: Forest plot of the effects of replacing saturated fat with polyunsaturated fat on coronary heart disease. *Abbreviations: CHD: coronary heart disease; MI: myocardial infarction; PUFA: polyunsaturated fatty acid; RR: relative risk*

than the low dose studies. The relevant estimate is therefore not an average, but the treatment effect for the dose level that you are interested in. That's a fairly obvious example, but there can be many other more subtle factors that can affect treatment effects, such as the inclusion criteria for the study, treatment duration, concomitant medications, healthcare setting, etc.

One way to deal with the problem of heterogeneity is to determine the major cause of heterogeneity and to present separate estimates for different groups. For example, Annane *et al.*[2] did a systematic review and meta-analysis to investigate the effects of corticosteroids on overall mortality at 28 days in patients with severe sepsis and septic shock. Their overall meta-analyses did not find a significant effect on mortality (relative risk 0.92, 95% confidence interval 0.75 to 1.14, P = 0.46), but it also found significant heterogeneity (I^2 = 58%, P = 0.003). When they divided their studies into those that had used long courses of low dose corticosteroids or short courses of high dose corticosteroids, they found that there was indeed a significant reduction in mortality in the studies that had used long courses of low doses (relative risk: 0.80, 0.67 to 0.95, P = 0.01), but not in the studies with short courses of high doses. Ignoring the heterogeneity would have meant missing the important difference between the difference dosing regimens.

That said, use of corticosteroids in sepsis is complex and controversial, and Annane *et al*'s analysis is unlikely to be the last word. Although a meta-analysis can give more reliable results than a single study, even a meta-analysis is often not sufficient to settle a medical question once and for all. There is probably considerably more heterogeneity that needs to be unpicked in this case, including genetic features of the patient and the nature of the infecting organism.[3]



Source: Wikipedia

Figure 2: Hypothetical symmetric funnel plot

One very common way in which the results of results of meta-analyses are presented is with a graph known as a forest plot. The example in Figure 1 is typical.

This shows the results of a meta-analysis on the effects on coronary heart disease (CHD) of increasing polyunsaturated fat in place of saturated fat.[4] There is a lot of information in that one graph. We can see details of each study, including the name of the study, the number of patients, and the number of CHD events. We also see how extensive the dietary changes were in each study as figures for % polyunsaturated fatty acid consumption in the control and intervention groups. We then see the results presented both graphically and in text. The central blob of each line shows the estimated relative risk from each study, and the extent of the horizontal line shows the 95% confidence interval. The size of the central blob shows how much weight the study provides (mainly a function of the number of patients in each study), the bigger the blob, the more that study contributes to the overall analysis. We then get the same information in text form to the right of the graph.

At the bottom, we see the overall estimate. Again, we see the relative risk and its confidence interval, presented both graphically and in text form. That's the important number to take away from meta-analyses, though as stated previously, it may be hard to interpret in the presence of significant heterogeneity among studies. The forest plot gives us another means of assessing heterogeneity by simply eyeballing the spread of the estimates from the individual studies.

Lastly, no discussion of meta-analyses would be complete without a few words about publication bias. Meta-analyses will never give a true summary of all the research that has been done if some studies are excluded. We know that not all studies are published. The claim by the All Trials campaign that only 50% of studies are published is of course nonsense and the real

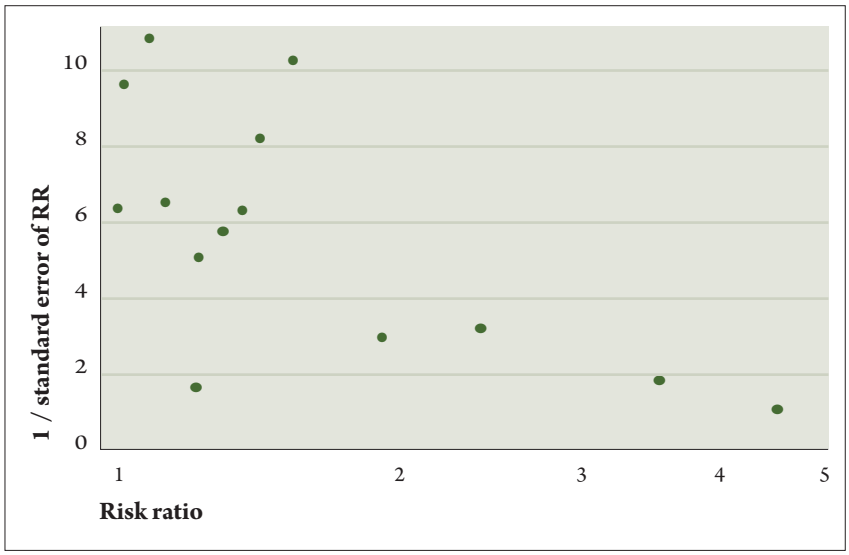


Figure 3: Funnel plot of studies investigating link between industry sponsorship and results favourable to the sponsor's product. *Abbreviations: RR: relative risk*

figure is probably much higher,[5] but nonetheless, the proportion of trials that are published is certainly less than 100%, and we know that studies reporting negative results are less likely to be published than positive studies.[6] If a meta-analysis includes only positive trials and ignores negative ones, then it will give an over-optimistic estimate of the true treatment effect.

A careful meta-analyst will therefore try to tell whether there is any evidence that publication bias has occurred. One way to do this is with a funnel plot, in which the treatment effect of individual studies is plotted on the x axis against the size of the study on the y axis. If all studies are published, the results would look roughly like an inverted funnel, with a greater spread of studies towards the bottom of the plot, where small sample sizes means that considerable variation in results is likely, and a smaller spread towards the top, where large sample sizes would keep results close to the "true" result (Figure 2).

If there is publication bias, it is likely that small negative studies will be unpublished, whereas small positive studies will be published. Large studies are more likely to be published whatever they show, as once you've gone to all the trouble of doing a large study you are more likely to be motivated to write it up. This can give rise to asymmetry in the funnel plot. Figure 3 shows one example of an asymmetric funnel plot.

I created this funnel plot from data provided in a Cochrane review of the effect of pharmaceutical industry sponsorship on publications.[7] The reviewers claimed that trials sponsored by pharmaceutical companies were more likely to be favourable to the sponsor's product than independent studies. Certainly the results of their meta-analysis showed that very strongly, but how much can we trust that result with such strong evidence of publication bias?

Meta-analysis is undoubtedly a useful technique that can provide important insights when summarising the medical literature. However, it is not a magic bullet, and must be interpreted with the same caution you would apply to any other results. Obviously if a meta-analysis is based on poor quality studies, the result will also be questionable. But in addition, it is also important to be aware of whether the studies are sufficiently similar that a meta-analysis makes sense, and crucially, whether all relevant studies have been included.

References

1. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
2. Annane D, Bellissant E, Bollaert E, Briegel J, Keh D, Kupfer Y. Corticosteroids for severe sepsis and septic shock: a systematic review and meta-analysis BMJ. 2004;329(7464):480.
3. Morgan, Paul (@drpaulmorgan). "@statsguyuk Very! The host-response mechanism is a mix of genetics (cytokine response) and the individual bug. Lots of interesting research!". [Tweet,10 June 2016, 8:12 AM]. Available from: https://twitter.com/drpaulmorgan/status/741166150147538944.
4. Mozaffarian D, Micha R, Wallace S. Effects on Coronary Heart Disease of Increasing Polyunsaturated Fat in Place of Saturated Fat: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. PLoS Med. 2010;7(3):e1000252.
5. Jacobs A. Zombie statistics on half of all clinical trials unpublished. [Blogpost; 29 August 2015]. Available from: http://www.statsguy. co.uk/zombie-statistics-on-half-of-all-clinical-trials-unpublished.
6. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, *et al.* Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8).
7. Jacobs A. Cochrane review on industry sponsorship. [Blogpost; 20 December 2012]. Available from: http://dianthus.co.uk/cochrane-review-on-industry-sponsorship.

Author information

Adam Jacobs was previously a medical writer, and was president of EMWA in 2004 to 2005. He now works as a medical statistician at Premier Research. He still teaches regular workshops for EMWA on statistical topics. You can find him on Twitter at @statsguyuk.