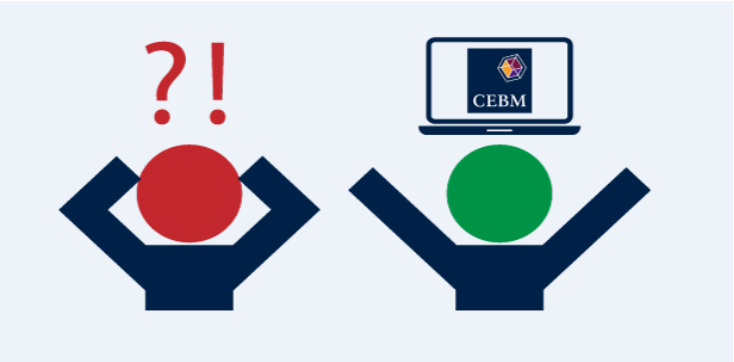
Tip for data extraction for meta-analysis – C1



What can I do if I have a missing mean, standard deviation or sample size?

Kathy Taylor

For meta-analysis of continuous outcomes in intervention studies, for each intervention group, you need to extract the mean and standard deviation (SD) of the outcome measure, and the number of participants in each group. Sometimes, you find that the wrong summary statistic is reported. When I say 'wrong', I mean that it's not what you want. It could be that:

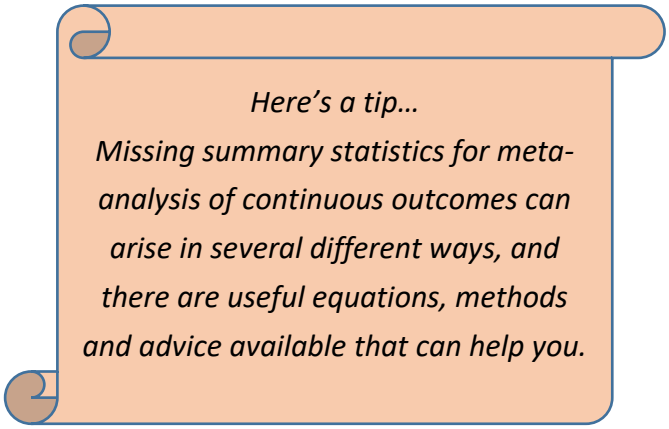
(1) the summary statistic you want is reported, but it's for the wrong time point
(2) the summary statistic you want is reported, but it's for the wrong group
(3) only a similar summary statistic is reported, not the statistical measure that you want, for example, it's the wrong measure of dispersion - an interquartile range instead of a SD
(4) neither the summary statistic you want nor a similar statistic is reported

Chapter 6 of the Cochrane Handbook and the review by Wiebe et al (and update by Weir et al) provide equations, methods and advice on how to deal appropriately with missing summary statistics when extracting data for meta-analysis for continuous outcomes. These resources show that multiple methods are in use across the systematic review community in handling missing means and SDs, with some readily implementable and others being more sophisticated and requiring statistical expertise. The former methods include algebraic recalculation (some exact and some approximate) and single imputation ('filling in') at the individual study level. The latter methods include multiple imputation at the meta-analysis level.

In a series of posts, I'll go through each of the 4 cases listed above and highlight some examples of equations that are readily implementable, provide some worked examples and derivations of equations (where they're not already derived in the source publications). I won't cover all the equations and methods, and refer you to the resources given above if you want to find out more about them and also the issues to consider when implementing them. In particular, it's important to be aware of the assumptions underlying these methods.

Note that I use the term "summary statistic" to refer to the variables needed as inputs to meta-analysis (for continuous outcomes, this is the means, SDs and numbers in each group), to differentiate between aggregated, summary data (which I cover in my blog) and individual patient data (which I won't cover). An individual patient data review is a particular type of systematic review, which is described in Chapter 26 of the Cochrane Handbook. Also, I'll refer to the pooled result of a meta-analysis as an "effect estimate". The Cochrane Handbook refers to this as an estimate of the "effect measure" or the "summary statistic" as meta-analysis summarises the results of studies.

Finally, here's a couple of reminders. As actual data are preferable to estimates it's worth trying to contact study authors, as stated before in post G2 (point number 4), to enquire about the data that you want. You might be lucky. Also, if you need to make any estimates (when using methods other than exact algebraic recalculations), don't forget to flag these studies, as I stated before in post G1 (point number 9), to remove as part of your sensitivity analysis.



In my next post, I'll focus on the 1st way of how a summary statistic that you want may be missing: ***the summary statistic you want is reported, but it's for the wrong time point.***

Dr Kathy Taylor teaches data extraction in Meta-analysis. This is a short course that is also available as part of our MSc in Evidence-Based Health Care, MSc in EBHC Medical Statistics, and MSc in EBHC Systematic Reviews.

Follow updates on this blog, related news, and to find out about other examples of statistics being made more broadly accessible on Twitter @dataextips