

Collecting data for meta-analysis – Part I

[Print](#)

I am collecting data for meta-analysis again. The first database is complete and preliminary analyses have been done, revealing very promising findings. Starting up another meta-analytic database brings back the problems I frequently faced over the last few months. But let's start from the beginning - meta-analysis itself.

Briefly, [meta-analysis](#) is a "statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" ([Gene Glass, 1976](#)). You start from a specific question, for example, "Is trait A related to B (e.g. body length and body weight)?" or "Is there a difference between group C and group D?" – e.g. control and treatment groups for a specific type of treatment. Then you collect all available data. You put them in a database. You analyse the data using some existing software and you are ready to publish new exciting results. Sounds simple and easy, but in practice it can be a lengthy and sometimes frustrating process.

Conducting meta-analytical research involves many steps. And each of these steps itself can be quite complicated and time consuming. For detailed description on "how it should ideally look like" you can refer to multiple books published on the topic. I find "[The handbook of research synthesis and meta-analysis](#)" by Cooper H., Hedges L.V. and Valentine J.C (2nd edition 2009) very informative and well written. But, the real question is – how does meta-analysis look in practice?

From what I done so far, I can say that knowledge of the research field and question really helps. But if you are not an expert to start off, you can still learn on the way. In both cases, you will end up reading a lot. Talking to people that know about given type of research helps heaps too.

Good tools are important for this job. You will need some reference management and flexible database software, especially for working with large collections of data. Although, if you expect to be able to find only a couple of papers on your topic, you can get by with entering everything in an Excel workbook. Simple Excel file is a good way to start developing a big database because it makes you think which fields and tables you will need and what the relations between them should look like. Later on, you can just import created Excel tables into proper database software and make them look beautiful and easier to use for everybody (that's the main advantage of a proper database).

Deciding which information from papers should be entered into a database and in what form takes time and practice. It is a good idea to do some pilot testing and revising selection and coding protocols couple times to get them right (otherwise you can end up with GIGO – "garbage in, garbage out"). Though, this does not mean that later, half way into your data collection, you will realise that you missed some important bits and you have to change the database structure again. Including censoring fields at different levels of the hierarchical database structure from the very beginning will be helpful by allowing you to easily exclude some records from the final dataset without having to delete them completely. Backing-up all files regularly is always essential.

I now see that I will not be able to fit everything I wanted to write in one blog. So, in my next two

blogs there will be something about: how much literature searching is enough, publication bias, effect sizes and reporting quality, and why high impact journals can be worse when it comes to collecting data.

Details

Written by Malgorzata (Losia) Lagisz

Category: General

Created: 16 May 2010

Last Updated: 27 August 2015

Hits: 1569098