Started 5 July 2021
Written for the germination database ...

Lizzie's guidelines on how to start a meta-analysis

Overview: Just like an experiment, when you do a meta-analysis you need to do it in a way that others could follow your methods and get (roughly) the same results. This is the basics of how Lizzie does it, but you can check for more info from places like PRISMA.

Without further ado, here are your basic steps.

1 Search for relevant papers

1. **Come up with reasonable search terms.** We need search terms to use in ISI (and potentially Google Scholar) that will capture most of the relevant papers, but will not return so many results that the list of papers to review is impossible. As a rule of thumb, I generally am hoping for search terms that return <1000 papers (but <2000 can work for bigger projects like this) and that at least 20% of papers look promising after an abstract review. If >80% of papers look relevant you may have too narrow search terms.

   Once you're happy with your search terms, I suggest running them by someone else (like Lizzie) along with how many total papers that is, and do a subsample to estimate how many look promising.

   For the germination project, check out the main `README` for the aim of the project, and the data/early folder has some search terms to start with.

2. **Record your search terms, the date you ran the search and what ISI edition you used.** Different universities pay for different access to ISI, you need to know what collections you searched and what years were included (universities also pay for a certain set of years).

3. **Download the search results.** You can generally download as a tabular form (Excel, tab delimited etc.) and a TXT (I usually download each). If you download the abstract of each article then you can do the next step without internet. No harm in also downloading a form that works with BIBTEX and your own reference software (try Mendeley or Zotero and/or ask in the lab what folks use).

4. **Review each abstract** and exclude any papers you can tell will not work for this meta-analysis. Create a key and record the reason for excluding each paper.

A few **don'ts**...

1. Do not exclude papers based on language; try to use a translation service to see if the paper is relevant.

2. If you cannot find a PDF of a paper, don't let that stop you! You can try https://unpaywall.org/ to see if a version may be lurking on the web. If that does not work, ask a librarian (tell them it's on the behalf of Lizzie). They should be able to help. If they can't, forward the email and exact article request to Lizzie.

3. Don't forget to include why you excluded a paper.

2 Develop a structure for the data you will scrape

This is seriously the most critical part of any meta-analysis according to Lizzie. You want to develop a database (aka an Excel or CSV file with metadata and column headers) that will allow you to enter all the data you'll eventually need to run analyses, but not to the point that you enter *tons* of data you will never need. Most people come up with databases without enough columns and spend the rest of the lives (okay, not really, it just feels that way), going back to the papers to enter new data they wish they had collected on the first pass.

So, review what the OSPREE database looked like, review a lot of papers to see what info you may want, ask Lizzie and spend a while finalizing this (Lizzie has literally spent days debating how to do this in working groups with lots of experts so this is an important and tricky skill). You'll likely miss something, that's okay! But spending more time here is better than entering all the data to find later that we wish we had a different format.

Okay, so steps to do this part:

1. Review other examples of similar databases then develop your own first attempt at your database's structure

2. Once you think you're happy, start entering papers (see below section)

3. As you enter papers, refine your meta-data (it needs to work so that others can understand it and enter data themselves)

4. As you enter papers, adjust your database structure

5. Check in with Lizzie for advice as you go.

6. When you think you're happy enter at least another 5-10 papers and make sure you're good.

Here's a list of common columns (they should have an underscore instead of white space).

1. ID – authoryear ID

2. first author

3. publication year

4. lat

5. lon

6. elevation

7. genus

8. species

9. varetc – variety, subspecies etc.

10. material

11. date of study

12. treatment

13. response value

14. response units

15. response sample size

16. response error value

17. response error type – SD, variance, SE etc. (this is a good example of where we should standardize what folks enter)

18. data scraped or downloaded

19. figtable of data

A few **do's**...

1. Be a splitter, not a lumper. Have a separate column for numbers and units (e.g., 10 m would be entered in one column as '10' and another units column as 'm')

2. When in doubt include an extra column.

3. Put the most common columns together at the start

4. Avoid special characters (they read into R as gobbly gook)! Write out Greek letters (e.g., micromol). Avoid white space in column names and wherever possible (for exmaple, R reads 'gnu' as different than 'gnu ').

5. Use lowercase letters as much as possible.

6. Encourage standard formatting whenever possible—if you see common methods or such, then develop a list of options for what people can enter. Much of the later work is standardizing everyone's formatting so getting people to enter stuff consistently to start helps a lot.

3   Scrape data from papers!

Eventually once you and at least someone else are happy with your database and have scraped a few papers, it's time to scrape the rest! For that, we will get help from the whole lab. Here's some basic steps (we need to work on this section):

1. Review each paper, highlighting info you may need to enter and noting any data publication info and/or figures with data we want.

2. When data is published, go get it! And see if you can extract the data we need.

3. If you need to scrape data from papers, we can. See `howtoscrapedata` for now (and let's expand this with your and Lizzie's help).

4. Work up some materials on how to do this, so others can learn from you ...