Part 5: Meta-analysis

Following on from the meta-analysis lab…

Your trusty assistant calls you in a panic to tell you that he accidently messed up when he sent you the data that we looked at in the meta-analysis labs! He confesses that he accidently sent you his office mate's data on meta-analysis of European drinking habits instead! The real data is in a file called "systematic review data for assignment.csv" that he has now sent to you. Your assistant is worried because he knows you've done lots of work on this analysis already, but you reassure him and say that it shouldn't be too much extra work to adapt the code you already have to look at the proper data set you now have.

Read in the new proper data and have a good look it.

How many studies/papers did your assistant find?

What is the maximum number of species considered in any one study?

How many studies included that maximum number of species?

How many total results/records/species has your assistant recorded?

How many different authors?

Which author has the most studies included in the dataset?

Which environment had the most studies?

Which method was used in most studies?

Within this dataset, which author had studied the fewest species?

Were the same number of reps used in each study?

Were the same number of reps used for each species within each study?

You've recently read a paper on standardising data for metanalysis. You didn't do this when you looked at the old dataset because you were thinking that the data was already quite standardised, since it is all relative growth rates, rather than absolute growth amounts. But as you've thought about it more, you think that perhaps it makes sense to standardise all the relative growth rates by the mean base relative growth rate for each species. This should avoid the possible issue that species will exhibit larger effects due to conspecific soil, just because they grow more generally. You know this shouldn't change the results much anyway, but decide to do it for the sake of avoiding possible reviewer criticism.

First make a forest plot of your data as it comes in the data sheet. Next standardise the data by dividing all the columns from 'meanbase' to 'ub' (ie six columns) by the meanbase growth rate (ie make everything relative to the mean base relative growth rate for that species). Then make the forest plot again on the standardised data. What is the effect of the standardisation?

From now on, use the standardised effect size for everything!

For how many records/species was a significant difference found?

For what proportion of records/species was a significant difference found?

What is the mean unweighted (but standardised) effect size across the studies?

What is the mean standardised effect size across the studies, when we weight by the number of reps used for that species in the study?

Make some forest plots using colouring and ordering to get a visual impression of whether effect size seems to vary with environment, plant types, method, author and/or study. Just from the forest plot, which environment looks like it has the smallest effect size overall (ie closest to zero)?

Estimate the effect sizes for each environment by calculating the mean standardised effect for each environment, weighted by the number of replicates. What is the estimated effect size for species in arid environments?

Now estimate the effect sizes for each environment by calculating the mean standardised effect for each environment, weighted by the inverse standard error. Does the relative ranking of the effect sizes for the four environments change?

Do the results of these estimates match what you saw in the forest plot about which environment looks like it has the smallest effect size overall?

Now imagine the very likely situation where none of the non-significant papers had actually been published ie create a new data set that only includes the significant studies. Make a forest plot of this reduced data set and check that it makes sense.

Now what is the mean unweighted (but standardised) effect size across the studies?

Why is this estimate bigger than the previous estimate based on the non-significant studies as well?

Estimate the effect sizes for each environment by calculating the mean standardised effect for each environment, weighted by the number of replicates. What is the estimated effect size for species in arid environments?

For which environment has estimated effect size changed the most by excluding the non-significant results?

Now try looking at the effects of environment, plant type, and method using the metareg function, using the data set with the non-significant results excluded. Consider only additive effects. You might want to use Forest plots too. Write down your conclusions from this analysis. (There is a list of possible conclusions to select from in the online quiz, so probably best to look at the options as soon as you can.)

Now try looking at the effects of environment, plant type, method and author using the metareg function, using the data set with the non-significant results excluded. Consider only additive effects, except also include an interaction between environment and plant type. You might want to use Forest plots too. Write down your conclusions from this analysis. (There is a list of possible conclusions to select from in the online quiz, so probably best to look at the options as soon as you can.)