

1: Meta-analysis

Emma James

1 Introduction

The aim of this week's practical is to walk through some of the key steps involved in conducting a meta-analysis. As a basis for this, we will use a recently published meta-analysis that examined the relationship between procedural learning and reading ability.

The Procedural Deficit Hypothesis proposes that reading and language disorders stem from a difficulty in the procedural learning system - i.e., the system that underpins the learning of probabilistic knowledge (Ullman et al., 2020). Procedural learning has often been measured using Serial Reaction Time (SRT) tasks: participants respond to a repeating sequence of stimuli, and their improvements in speed over several blocks are measured as an index of learning. However, evidence that SRT performance is associated with reading ability is mixed, with some studies showing a strong association and others finding nothing at all. Many studies in this area suffer from small sample sizes, making it hard to draw sound conclusions.

A meta-analysis provides a valuable way of summarising this evidence. This meta-analysis was conducted by a former PhD student at York, Dr Catia Oliveira, together with Professor Lisa Henderson and Dr Emma Hayiou-Thomas. We will use part of their dataset to showcase different steps in conducting a meta-analysis.

Learning objectives

By the end of this practical, you will be able to:

1. Identify and extract effect sizes for pooling.
2. Calculate pooled effect sizes via fixed and random effects meta-analyses.
3. Use sample size to weight studies in a meta-analysis.
4. Produce a forest plot to display effect sizes.
5. Produce a funnel plot to inspect for publication bias.

2 Effect sizes

2.1 Extract effect sizes

In this week's practical section on the VLE, you should find a file named **ambr1_srt_gaps.csv**. Download this file and save it in your data folder.

Tip

If you don't have this folder on your current machine, repeat the set-up steps outlined in **0_SetUp.pdf** from Week 1.

Open this file in Excel to start with. This is a stripped back version of the published meta-analysis. You can see that we have the following columns:

- **Study:** the authors and year of publication
- **N:** the sample size
- **Age:** the average age of the sample
- **Task:** the reading task administered
- **Group:** whether the participants had developmental dyslexia (DD) or not (TD)
- **Cor:** the correlation coefficient, r

You will see that a few cells are blank. We need to extract this information from the relevant papers before we begin! Inspect the following papers to fill in the missing sample sizes and correlation coefficients:

- [Clark & Lum \(2017\)](#). (*Hint: Inspect Table 3. You want to know the correlation between the procedural learning (SRTT) and nonword (phonemic decoding efficiency) tasks.*)
- [Schmalz et al. \(2021\)](#). (*Hint: Inspect Table 2. The correlations in the meta-analysis reflect each of the word reading and pseudoword reading measures with the SRTT difference*)

Make sure you check carefully that all cells of the table are filled before proceeding. Then save your new version of the file as **ambr1_srt_complete.csv** in the same data folder.

2.2 Setting up in RStudio

Now we have extracted information from our identified studies, let's load the data into R. In your **ARM_Practicals** folder, double click the **ARM_Practicals.Rproj** file to open it in RStudio.

Create a new R script (using the button at the top-left of your window), and save it as **ARM2_meta-analysis.R** in your scripts folder.

Let's start by loading the necessary packages for today's exercises. You will need to first install the **metafor** package for meta-analysis by running `install.packages("metafor")`. You should only need to run this once on any device, so you don't need to add this to your script. We will also be using the **tidyverse** package. This is already installed on university machines, but you might need to install it in the same way if you are working on a personal device.

Once you've installed the packages, add the following lines to the top of your script and run them to load the necessary packages for today.

```
# Load packages
library(tidyverse) # for data cleaning
library(metafor)   # for meta-analysis
```

The next step is to load the data into R. If you have opened your R project and have saved the data in the data folder, you should be able to load it in as follows.

```
# Read in extracted effect sizes and study info
full_dat <- read.csv("./data/ambr1_srt_complete.csv")
```

The dataset includes studies of people with dyslexia as well as those without reading difficulties. Let's start by focusing just on those individuals without reading difficulties. We can do this using the `filter()` function from the *tidyverse* package that we loaded above. We tell R to filter (or subset) the dataset to keep the rows where the `Group` column is listed as "TD", and store the result as `td_dat`.

```
td_dat <- filter(full_dat, Group == "TD")
```

2.3 Compute sampling variance

We already have a correlation coefficient for each study, which is our effect size for this meta-analysis. It quantifies the direction and strength of relationship between SRT task performance and reading ability. However, some of the studies has much larger sample sizes than others. Use the `min()` and `max()` functions to identify the smallest and largest sample sizes (N) in the dataset. Remember, you can call e.g., `help(min)` to remind you how these functions work.

To capture this uncertainty around the correlation estimates, we should compute their variance ahead of running the meta-analysis. Our meta-analysis package, *metafor*, provides a handy function for doing this.

```
td_dat_var <- escalc(measure = "COR",    # specify effect size measure
                      ri = Cor,        # specify correlation variable
                      ni = N,          # specify sample size variable
                      data = td_dat)  # dataset
```

If you look to your Environment window (top right), you should now see an object called `td_dat_var`. If you click the grid to inspect it, you will see two extra columns at the end:

- **yi**: your effect size, which in this case is identical to the one in the `Cor` column as we haven't asked it to do any kind of transformation
- **vi**: your sampling variances, which have now been computed based on the correlation coefficient and sample size

Now we're ready to run the analysis!

3 Computing pooled effect sizes

In the lecture, you learned about fixed and random effects meta-analyses. Let's have a go at running both.

3.1 Fixed effects meta-analysis

In a fixed effects analysis, we assume that all studies are measuring the same underlying true effect size, and calculate the average effect size across the included studies. The `metafor` package terms this an “equal effects” model, as a better descriptive name. Use the `rma()` function to run this as follows:

```
# Fixed effects model
fixed_mod <- rma(yi = yi, vi = vi,
                  method = "FE", data = td_dat_var)
```

We've specified the variables with our effect size estimates and sampling variance (already conveniently named the same by `escalc()` function above). We've then specified the “FE” method for fixed effects, and the dataset to use.

We can use the `summary()` function on the model object to inspect the results.

```
summary(fixed_mod)
```

Fixed-Effects Model (k = 73)

logLik	deviance	AIC	BIC	AICc
23.9041	82.2057	-45.8082	-43.5177	-45.7518

I^2 (total heterogeneity / total variability): 12.41%
 H^2 (total variability / sampling variability): 1.14

Test for Heterogeneity:
 $Q(df = 72) = 82.2057$, p-val = 0.1926

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0444	0.0178	2.5003	0.0124	0.0096	0.0793 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The output tells us that we ran a Fixed-Effects model with 73 effect sizes. The pooled estimate for the correlation coefficient is ~0.04, showing a very small yet statistically significant correlation.

3.2 Random effects meta-analysis

The random effects model assumes that the samples in the meta-analysis are a random sample from a larger population of studies that could plausibly have been/will be conducted. We can run this by simply changing the method in the meta-analysis function. The package provides us with several different options that have different ways of estimating the heterogeneity between studies (feel free to read more about these!), but for now we will stick with the default option.

```
# Random effects model
random_mod <- rma(yi = yi, vi = vi,
                     method = "REML", data = td_dat_var)
summary(random_mod)
```

Random-Effects Model (k = 73; tau^2 estimator: REML)

logLik	deviance	AIC	BIC	AICc
25.3819	-50.7639	-46.7639	-42.2105	-46.5900

tau^2 (estimated amount of total heterogeneity): 0.0068 (SE = 0.0049)

tau (square root of estimated tau^2 value): 0.0827

I^2 (total heterogeneity / total variability): 22.78%

H^2 (total variability / sampling variability): 1.30

Test for Heterogeneity:

Q(df = 72) = 82.2057, p-val = 0.1926

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
0.0361	0.0208	1.7361	0.0825	-0.0047	0.0769 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

We can see from the results that the coefficient estimate has shrunk slightly and is no longer significant. These are both valid models but have different inferences: the fixed effects model tells us that there is a very small but significant correlation across the available datasets, but when we consider heterogeneity between effect sizes in the random effects model, we see that this effect size is likely even smaller in the wider population.

4 Weighting studies

We saw in Section 2 that the variation in sample size was very large. We should consider this in our analysis: a very large study likely provides a better estimate of the true effect size than a very small study.

4.1 Under the hood: computing inverse variance

One of the most common ways to do this is via inverse-variance, which the `rma()` function does by default. If you inspect your `td_dat_var` data frame, you will see that the smaller samples have the highest variance (`vi`). You can click on the column header to order the data by this variable, so that you can see this for yourself.

To use this variance in the analysis, the model uses the *inverse (1/variance)*, so that smaller samples have the smallest weighting. The function does this under the hood, but you can see what this looks like by computing it for yourself:

```
td_dat_inv <- mutate(td_dat_var, inv_var = 1/vi)
```

Here we used the `mutate()` function (from the *tidyverse* package) to make a change to the existing data frame `td_dat_var`. We provide the name of the new variable, `inv_var`, and specify that this variable should be computed as 1 divided by the variance variable `vi`. We assigned the new version of the dataset to a new object, `td_dat_inv`.

Inspect this new object by clicking the grid icon in your Environment window. You should now see that your `inv_var` column has larger numbers for studies with larger sample sizes (`N`).

4.2 Changing weights in the meta-analysis

Because weighting by inverse variance is a very common and sensible approach, the `rma()` meta-analysis function does this by default. This means you've already weighted your samples in your above models!

To compare, try fitting two new models (fixed, random) and include the argument `weighted = FALSE`. Use a different object name to store the model output, so that you can inspect the differences in output when you call `summary()`.

Alternatively, you can try weighting by sample size alone, by setting `weights = N`.

5 Displaying results with forest plots

An easy way to visualise the results is to use a forest plot. The *metafor* package provides a handy `forest()` function for this: we simply give it the name of our meta-analysis model object. We can also use the `slab` argument to provide study labels, which we can take from the `Study` column of the dataset we used.

```
forest(fixed_mod, slab = td_dat_var$Study)
```

Try this now for the different models you ran, and see how the pooled estimate at the bottom changes in whether it crosses the 0 line.

6 Inspecting for publication bias

Finally, let's consider whether this analysis might be affected by publication bias. We can do this using a funnel plot:

```
funnel(fixed_mod)
```

Take a look at [this linked guide for interpreting the funnel plots](#). What do you conclude?

You can also test for asymmetry using Egger's regression test, which is also a built-in function in the *metafor* package. See if you can find the relevant function online and apply it here.

7 Summary

In this practical, we have extracted relevant results from published studies and used them in both fixed and random effects meta-analyses. We saw how these models can weight different studies in computing the pooled estimates, and plotted them using forest plots. Finally, we considered whether there might be evidence of publication bias in this literature. Before you leave, make sure you have commented your code so that future-you remembers what you have done.

We ran a basic version of these models here, but there is much more you can do to correct for small sample sizes and adjust for the fact that some estimates come from the same study (i.e., there were different correlation coefficients for different measures of reading in the same sample). If you are interested in these issues, you can [access the full paper here](#) or [download the full dataset and analysis scripts](#).

7.1 Challenge yourself

Often we are interested in how effect sizes vary across different conditions. In this case, the researchers were interested in whether the relationship between procedural learning (SRT) and reading ability might differ depending on the participant groups tested.

If you finish the practical in good time or want to push your skills further beyond the end of today's session, see if you can test whether participant *Group* is a moderator in your analysis. Use the original data file that you loaded in (*full_dat*), which also includes effect sizes for samples with developmental dyslexia. You can look at the help files or use Google to find out how to include this extra predictor.

7.2 Further learning

You can find the core reading for this week on the VLE. If you come to do meta-analysis in the future or want further details on the different steps that we have covered, then there is an excellent textbook available for free online: [Harrer et al. \(2021\). *Doing Meta-Analysis with R: A Hands-On Guide*.](#)

Feedback

These practical sessions have been heavily edited this year. Please take 20 seconds to rate the difficulty and length of these practical activities [via this survey](#), and let us know of any issues you encountered or aspects you didn't understand (positive feedback is also welcome!). You will need to be logged in with your university account to respond, but your submission will be anonymous. For more general questions about the practical, please post them on the VLE Discussion Board.