Lab 12: Meta Analysis

PHS2000B, Spring 2023

Sudipta Saha

April 23, 2023

Overview

Thanks to Dena Jevadi, Matt Lee and generations of PHS TFs

- **Motivation for evidence synthesis**
- **What can evidence synthesis look like?**
- **How do we do evidence synthesis?**
  1. Inclusion Criteria
  2. Systematic search
  3. Extract qualitative and quantitative information
  4. Pool effect sizes (Meta-analysis)
  5. Assess bias in the literature
  6. Summarize the state of evidence

Motivation for evidence synthesis

Motivation

Evidence synthesis describes the practice of **qualitatively** (systematic/ narrative review) or **quantitatively** (meta-analysis) summarizing information across multiple studies.

- No **single** scientific study fills the gaps in the literature
- Studies may lack **internal** and/or **external** validity
- The **literature itself** may lack validity (i.e. publication bias)

Evidence synthesis addresses some of these concerns by providing a framework to assess the (near) totality of relevant literature for a given topic, **summarize results** across studies, **identify biases in the evidence base**, and **describe knowledge gaps** for researchers and implications for policy.

How are reviews used?

- Setting research priorities
- Defining research gaps
- Defining practice challenges
- Shaping the field
- Resource allocation
- Guideline Development

What can evidence synthesis look like?

What do you need to do?

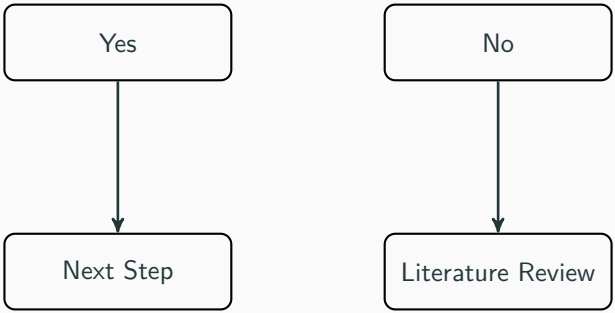
Do you want to systematically gather all evidence on a particular research topic?

Yes

No

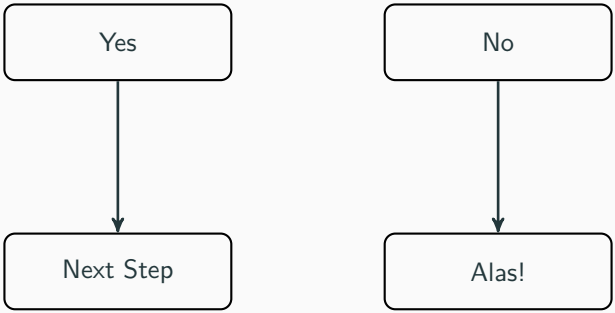
What do you need to do?

Do you want to systematically gather all evidence on a particular research topic?



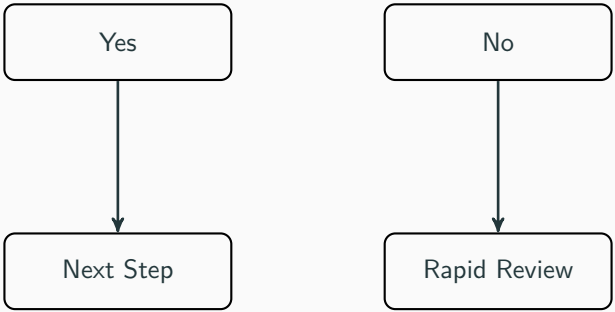
What do you need to do?

Do you have 3 or more people working on the review?



What do you need to do?

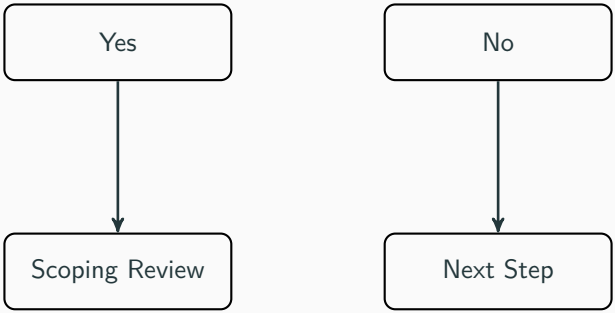
Do you have 12-18 months to complete the review?



Resources for Rapid Review: Cochrane Guidance, WHO Guide.

What do you need to do?

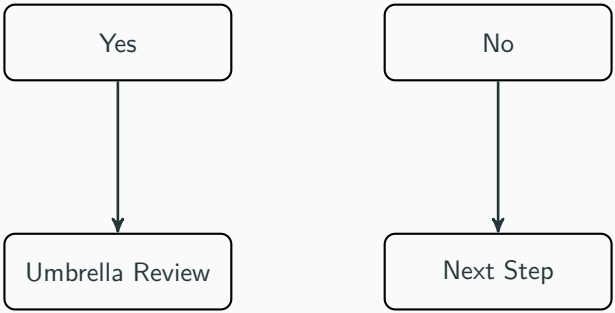
Do you have a broad topic or multiple research questions?



Resources for Scoping Review: Cochrane Guidance

What do you need to do?

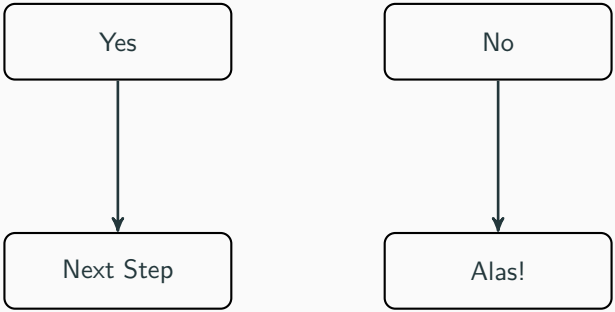
Do you want to review other published systematic reviews on your topic?



Resources for Umbrella Review: Cochrane Guidance

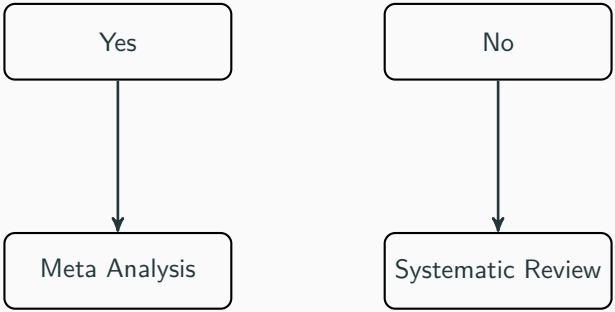
What do you need to do?

Do you have a well-formulated question?



What do you need to do?

Will you use statistical methods to summarize results?



Resources for Systematic Review: Cochrane Guidance

Resources for Meta Analysis: Cochrane Guidance

How do we do evidence synthesis?

A roadmap for evidence synthesis

1. Develop inclusion criteria
2. Systematic search of studies
3. Extract qualitative and quantitative information
4. Quantitatively pool effect sizes[1]
5. Assess risk of bias in the literature
6. Holistically assess the state of evidence

[1] Typically done in **meta-analyses**, otherwise called a **systematic review**.

1. Develop Inclusion Criteria

Goal of inclusion criteria is to **mitigate bias** in the review process, specific to the research question of interest. Higgins & Thomas (2021)[2] recommend answering the **"PICO"** criteria for:

- **Each synthesis**, planned at the **protocol stage** (i.e. defining the question that each specific synthesis aims to answer, and "should relate clearly and directly to the questions or hypotheses that are posed when the review is formulated")
- **Included studies**, planned at the **review stage** (i.e. what was actually investigated in the studies)

[2] See chapters 2 and 3

1. Develop Inclusion Criteria

The PICO criteria include:

Population: What kinds of participants must studies include?

Intervention: What kinds of interventions or exposures must studies assess?

Comparison: What kinds of comparator or placebo must studies use?

Outcome: How must studies define and measure the outcome?

As part of this process, inclusion criteria should also **consider study design** and the types of questions they can answer. For causal questions:

- Randomized studies (most preferred)
- Longitudinal observational studies that measure exposure before outcome and control for confounding
- Cross-sectional (least preferred)

2. Systematic search for studies

After specifying the PICO parameters, important to devise a **search strategy** that covers all potential sources of evidence:

- Need to craft specific **search terms** that are both sensitive and specific, especially for popular questions
- At least **2 reviewers** should independently screen studies against inclusion criteria (start with titles & abstracts, then skim full-text as needed)
- Choice of database will depend on research question, although **Google Scholar** is an option that also captures the gray literature.
  - The (free) script `sortgs.py` is a Python program[3] that can help identify highly-cited publications for a given search. Best to use as a check to see if you've missed any obvious papers, since sorting by no. citations is obviously subject to publication bias
  - A more formal (but paid and not Harvard sponsored) option is DistillerSR

[3] https://github.com/WittmannF/sort-google-scholar

2. Systematic search for studies

Also want to note that you should:

- Talk
- To
- A
- Librarian

3. Extracting Information

There are a number of **software options** for organizing extracted information, including DistillerSR, Covidence, Endnote, etc. For a full comparison of the different options for biomedical research, see Van der Mierden et al. (2019)

For meta-analysis, quantitative results need to be extracted and may be transformed to meet assumptions about the population effects[4]:

- Risk ratios (log transform)
- Odds ratios (log transform)
- Pearson's correlations (Fisher's z)
- Standardized mean differences, i.e. Cohen's d

Higgins & Thomas (2021) Chapter 6 breaks down how to extract and standardize by type of outcome (dichotomous, continuous, ordinal, count, rate, time-to-event).

[4] i.e., effects are normal and independent of their standard errors

4. Pooling the effect sizes (Meta-analysis)

In a **meta-analysis**, we need a way to pool across different studies while accounting for differences in precision (i.e. up-weighting larger studies and down-weighting smaller ones). Let:

k : Number of studies in the meta-analysis

$\hat{\theta}_i$: Point estimate of study i

$\hat{\sigma}_i$: Standard error for $\hat{\theta}_i$ of study i

μ : True effect size in all studies

$\hat{\mu}$: Meta-analytic (pooled) estimate

4. Pooling the effect sizes (Meta-analysis)

There are two primary approaches to meta-analysis:

- **Fixed-effects**: Studies measure the **same** effect μ, with some study-specific random error ε_i. All the effect estimates from the various studies are assumed to come from the same common underlying distribution
- **Random-effects**: Studies measure **different** effects, which are comprised of a shared effect μ and a study-specific *random effect* γ_i (as well as random error ε_i). Assumes a distribution of study specific effects (each study a realization from underlying study distribution)

4. Pooling the effect sizes (Fixed-effects meta-analysis)

Fixed-effects: Studies measure the **same** effect μ, with some study-specific random error ε_i, i.e.:

$$\hat{\theta}_i = \mu + \varepsilon_i$$

Where the goal is to obtain an estimate of μ, $\hat{\mu}$:

$$\hat{\mu} = \frac{\sum_i^k w_i \hat{\theta}_i}{\sum_i^k w_i}$$

Where $w_i = 1/\hat{\sigma}_i^2$, i.e. we weight each study by the **inverse of the study-specific estimated effect variance**.

4. Pooling the effect sizes (Fixed-effects meta-analysis)

Let's look at an example of conducting a fixed-effects meta-analysis in practice. The `metafor` package in `R` contains a vignette dataset from a meta-analysis by Banger-Drowns et al. (2004). In each study:

- The **treatment group** received some "writing-to-learn" training
- The **control group** received conventional instruction
- The **outcome**, some conventional measure of academic achievement, was converted to standardized mean differences (Cohen's d), where higher scores indicate beneficial treatment effects

4. Pooling the effect sizes (Fixed-effects meta-analysis)

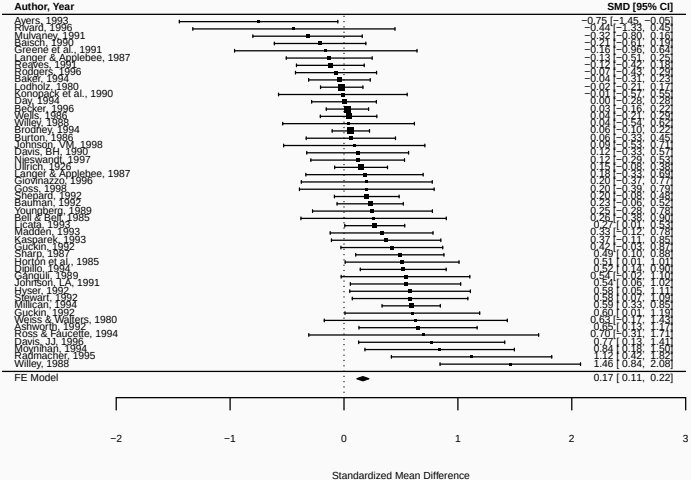
First, reading in the data:

```
library(tidyr)
library(metafor)
dat <- dat.bangertdrowns2004 %>%
    drop_na(yi) %>%
    arrange(yi)  # no missing outcomes, sort by effect size
# ?dat.bangertdrowns2004 # data dictionary
```

A sensible preliminary step is to look at a **forest plot** across the studies (we'll need to fit the model first to do this), which depicts each study's estimate and confidence interval:

```
#Fit model
fe <- rma.uni(yi=yi, # effect estimate for each study
              vi=vi, # variance for each study
              method = "FE", data = dat)
forest(fe, header = "Author, Year",
    slab = paste(dat$author, dat$year, sep = ", "),
    xlim = c(-3, 2), cex = 0.5)
```

4. Pooling the effect sizes (Fixed-effects meta-analysis)



4. Pooling the effect sizes (Fixed-effects meta-analysis)}

Finally, let's look at the **fixed effects model results** by printing our `fe` object

Question

How would you interpret this estimate, $\hat{\mu} = 0.166$?

4. Pooling the effect sizes (Fixed-effects meta-analysis)}

Finally, let's look at the **fixed effects model results** by printing our `fe` object

Question

How would you interpret this estimate, $\hat{\mu} = 0.166$?

The estimate of the true population standardized mean difference is 0.166 (assuming all studies are measuring the same thing)

4. Pooling the effect sizes (Fixed-effects meta-analysis)

A **key limitation** of fixed effects meta-analysis is that it essentially **ignores** the fact that each observation came from a (potentially very) different study, $\Rightarrow$ i.e assumes no **heterogeneity** in the population effects.

In our writing-to-learn example, studies differed in:

- Grade level (college, high, middle, elementary)
- Treatment length (1 week - 24 weeks)
- Subject (Nursing, Earth Science, Math, Social Studies, etc.)
- Whether treatment was given in class or at home
- Whether feedback was provided

It's more likely that these studies are measuring effects for different populations and/or treatments, so we need a way to incorporate this heterogeneity: **random-effects meta-analysis**

4. Pooling the effect sizes (Random-effects meta-analysis)

In **random-effects meta-analysis**, we assume that each study has it's **own** population effect, i.e.:

$$\hat{\theta}_i = \underbrace{\mu + \gamma_i}_{\text{study specific pop. effect}} + \varepsilon_i$$

Where μ is now the **mean effect across studies**. We also denote the standard deviation of the population effects as τ, which describes the amount of **heterogeneity**. We can estimate μ and τ using maximum likelihood[5] by also assuming:

$$\begin{aligned}
\gamma_i &\sim \mathcal{N}(0, \tau^2) \\
\varepsilon_i &\sim \mathcal{N}(0, \hat{\sigma}_i^2) \\
\gamma_i &\perp\!\!\!\perp \varepsilon_i \\
\hat{\theta}_i &\sim \mathcal{N}(\mu, \tau^2 + \hat{\sigma}_i^2)
\end{aligned}$$

[5] In practice, we'll use restricted ML (REML)

4. Pooling the effect sizes (Random-effects meta-analysis)

Can be helpful to think in multilevel terms:

$$\hat{\theta}_i \sim \mathcal{N}(\mu, \tau^2 + \hat{\sigma}_i^2)$$

where is the standard deviation of the population effects, which describes the amount of heterogeneity Notice here that the variability of the point estimates is then due to both genuine heterogeneity in the true effects τ^2 AND sampling error due to finite sample sizes.

- individual data points within each study are at level 1 and the studies are at level 2
- γ_i captures between study variation (level 2 residual), so τ^2 captures the variance at level 2 (γ_i) can be thought of as a random intercept)
- ε_i is the specific observations residual at level 1 and $\widehat{\sigma_i^2}$ is the within study, between individual variance

4. Pooling the effect sizes (Random-effects meta-analysis)

Under these assumptions, the ML estimators of μ and τ are given by:

$$\hat{\mu} = \frac{\sum_i^k (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-1} \widehat{\theta_i}}{\sum_i^k (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-1}} = \frac{\sum_i^k w_i \hat{\theta}_i}{\sum_i^k w_i}$$

Where $w_i = (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-1}$, and:

$$\hat{\tau}^2 = \max \left\{ 0, \frac{\sum_i^k (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-2} \left[(\hat{\theta}_i - \hat{\mu})^2 - \hat{\sigma}_i^2 \right]}{\sum_i^k (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-2}} \right\}$$

You don't need to worry about the estimation for this class. But notice that estimation of μ depends on estimation of τ^2 (and vice versa), so in practice software will estimate each iteratively until convergance $\Rightarrow$ it's important to adjust the standard error for $\hat{\mu}$ using the **Knapp-Hartung adjustment** that accounts for the uncertainty in estimating τ^2.

4. Pooling the effect sizes (Random-effects meta-analysis)

Like its fixed-effects counterpart, implementing random-effects meta-analysis in R is as easy as changing a couple of arguments. Going back to our writing-to-learn example:

```
library(metafor)
re <- metafor::rma.uni(yi, vi, method = "REML", data = dat, knha = TRUE)
```

To get a 95% confidence interval for $\widehat{\tau}$, we can use the `tau_CI()` function from the `MetaUtility` package.

```
library(MetaUtility)
MetaUtility::tau_CI(re)
```

4. Pooling the effect sizes (Random-effects meta-analysis)

```
## tau^2 (estimated amount of total heterogeneity): 0.0499 (SE = 0.0197)

## tau (square root of estimated tau^2 value):      0.2235

## I^2 (total heterogeneity / total variability):   58.37%

## H^2 (total variability / sampling variability):  2.40

## Test for Heterogeneity:

## Q(df = 47) = 107.1061, p-val < .0001

## Model Results:

## estimate      se    tval    pval   ci.lb   ci.ub

##   0.2219  0.0495  4.4810  <.0001  0.1223  0.3216  ***
## 95% CI for tau:

## [1] 0.1062656 0.2976404
```

4. Pooling the effect sizes (Random-effects meta-analysis)

Question

How would you interpret these estimates, $\hat{\mu} = 0.222$ and $\hat{\tau} = 0.224$?

The "writing to learn" intervention is associated on average across studies with increased academic achievement by SMD = 0.222 (95% CI: [0.12, 0.32]). The estimated SD of effects across studies is $\hat{\tau} = 0.224$ (95% CI: [0.11, 0.30]).

4. Pooling the effect sizes (Random-effects meta-analysis)

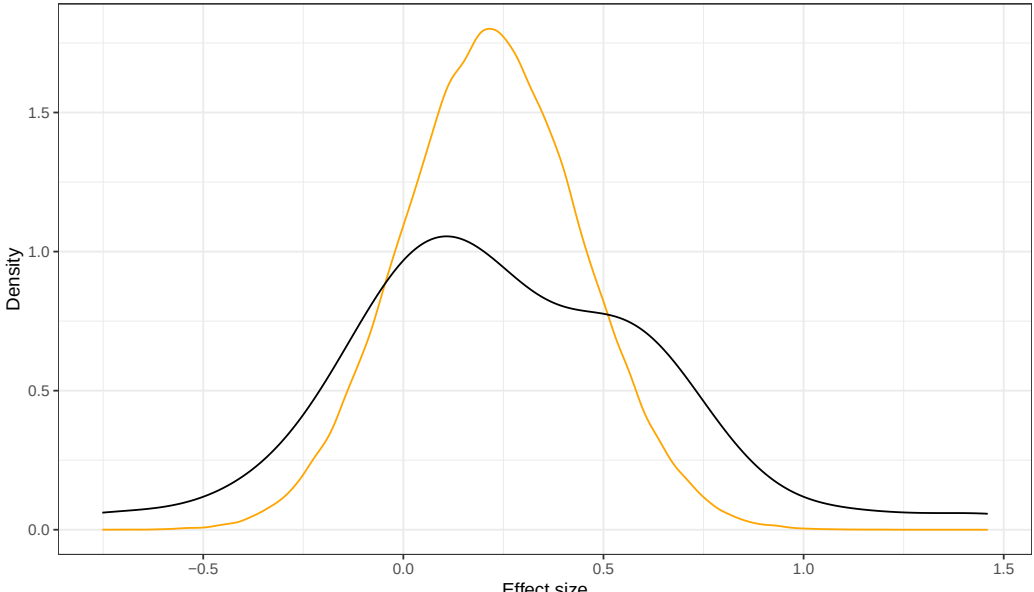
We can also compare the estimated distribution of the study-specific population effects, $\mu + \gamma_i$ (orange), to the observed distribution (black) in order to get a sense for how well our model predicts the data:

```
library(ggplot2)

muhat <- re$b
vhat  <- re$tau2
trues <- rnorm(n = 1000, mean = muhat, sd = sqrt(vhat))

ggplot() +
    geom_density( aes(x=trues), color = "orange" ) +
    geom_density( aes(x=dat$yi) ) +
    theme_bw() +
    xlab("Effect size") +
    ylab("")
```

4. Pooling the effect sizes (Random-effects meta-analysis)



4. Pooling the effect sizes (Random-effects meta-analysis)

Last, we should report some estimate of the **heterogeneity between studies**.

- At the very least, $\widehat{\tau}$ itself
- I^2: The proportion of the variation in the **observed effects** that is due to variation in the **true effects**
  - However, I^2 **doesn't** actually tell us **how much** the effects vary on an *absolute* scale, it tells us what proportion of the observed variation would remain if we could eliminate all sampling error.
- Borenstein et al. (2017) recommend reporting **95% prediction intervals**, or an interval that contains the population effect of a randomly drawn new study with 95% probability.

In `metafor`, this can be done using the `predict()` function (called `cr.lb` and `cr.ub` for "credibility interval"[6]):

[6] see https://rdrr.io/cran/metafor/man/predict.rma.html

4. Pooling the effect sizes (Random-effects meta-analysis)

```
predict(re)
```

```
##
##    pred     se  ci.lb  ci.ub   pi.lb  pi.ub
##  0.2219 0.0495 0.1223 0.3216 -0.2386 0.6824
```

4. Pooling the effect sizes (Random-effects meta-analysis)

- Another way to report heterogeneity is to provide a **percentage of meaningfully strong effects**, which is the predicted proportion of effects that are above some clinically relevant threshold.
- Can be implemented using `MetaUtility` package (see documentation for the `prop_stronger()`) function.

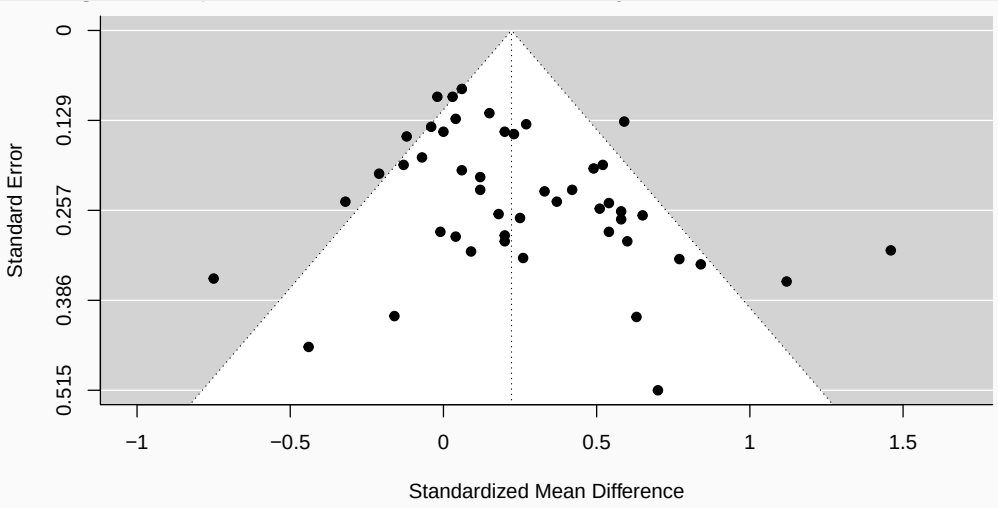
5. Assess bias in the literature

When synthesizing evidence, we need to think about biases from two sources:

- Biases **within** studies
  - Every meta-analysis should provide a qualitative **rating** for each study's risk of bias (e.g. ROBINS-I, GRADE, Cochrane RoB)
- Biases **across** studies
  - **Funnel plots** can provide evidence for/against publication bias (i.e. selective publication of positive or clinically significant results) by plotting each study's estimate vs. their standard error.
  - Ideally, symmetry in a funnal plot indicates lack of publication bias, whereas asymmetry might reflect some selective pressure to publish stronger estimates.

5. Assess bias in the literature

```
funnel(re)
```



5. Assess bias in the literature

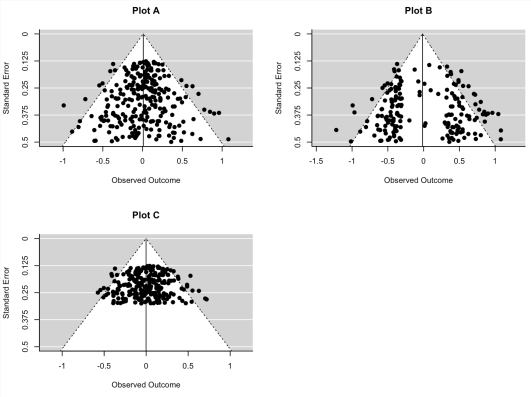
However, funnel plots have limitations and **should not be** the only source of evidence in determining whether there is publication bias

- May not detect publication bias due to statistical significance
- Assume no heterogeneity in the population effect estimates across studies

Let's consider Q1 of the Problem Set

Consider three different journals with different publication mechanisms:

- *PLOS ONE* accepts papers based on criteria not related to effect size or standard error.
- *Science Magazine* only accepts studies with large effect sizes, regardless of direction (positive or negative).
- *The Journal of High-Powered Research* only accepts studies with power above a specific threshold to detect an effect size of Cohen's d = 0.20. However, once a study surpasses that threshold, JHPR does not select based on results.



Reminders

1. **Fill out exam timing quiz on Canvas - Quizzes tab**
2. **Meta Analysis PS due on Friday Apr 28th**
3. **Extra credit assignment released tomorrow April 25th and due Friday, April 28th**