Meta-Analysis

Collin Nolte

April 27, 2021

Goals

- Understand basic approach and goals of meta-analysis
- Read and interpret common graphics
- Understand benefits and limitations of meta-analysis
- Evaluate quality of meta-analysis and draw conclusions

Review Articles

Three types: qualitative/narrative, systematic, and quantiative (meta-analysis)

Generally serve a number of goals:

- Summarize literature of topic up to some point
- Synthesize and condense material from multiple studies
- Explore and perhaps exlain conflicting results

As with any scientific endeavor, care needs to be taken to ensure that the review is effective and free from bias

Systematic Reviews

"A systematic review is a high-level overview of primary research on a particular research question that tries to identify, select, synthesize and appraise all high quality research evidence relevant to that question in order to answer it."

– Definition from Cochrane Collaboration

http://www.cochrane.org/about-us/evidence-based-health-care

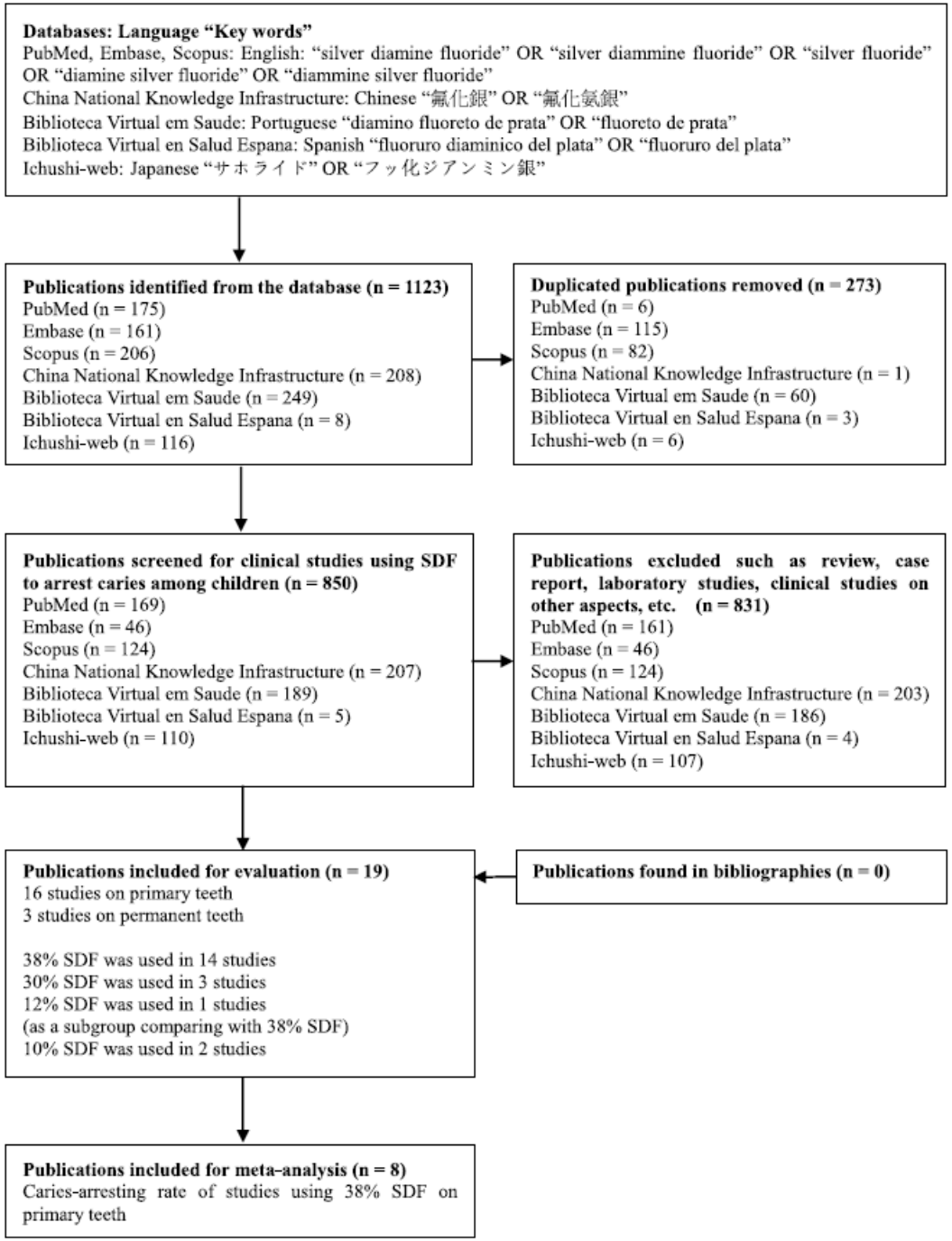
Systematic Reviews

To remove researcher bias and limit focus to high quality studies, important to develop a detailed research protocal *a priori*

This usually involves a comprehensive search to identify relevant studeis, with specific criteria for determining which studies are relevant

Should include rigorous methods for the appraisal, collection, and synthesis of data

Garbage in, garbage out


Databases: Language "Key words"
PubMed, Embase, Scopus: English: "silver diamine fluoride" OR "silver diammine fluoride" OR "silver fluoride" OR "diamine silver fluoride" OR "diammine silver fluoride"
China National Knowledge Infrastructure: Chinese "氟化銀" OR "氟化氨銀"
Biblioteca Virtual em Saude: Portuguese "diamino fluoreto de prata" OR "fluoreto de prata"
Biblioteca Virtual en Salud Espana: Spanish "fluoruro diaminico del plata" OR "fluoruro del plata"
Ichushi-web: Japanese "サホライド" OR "フッ化ジアンミン銀"
Publications identified from the database (n = 1123)
PubMed (n = 175)
Embase (n = 161)
Scopus (n = 206)
China National Knowledge Infrastructure (n = 208)
Biblioteca Virtual em Saude (n = 249)
Biblioteca Virtual en Salud Espana (n = 8)
Ichushi-web (n = 116)
Duplicated publications removed (n = 273)
PubMed (n = 6)
Embase (n = 115)
Scopus (n = 82)
China National Knowledge Infrastructure (n = 1)
Biblioteca Virtual em Saude (n = 60)
Biblioteca Virtual en Salud Espana (n = 3)
Ichushi-web (n = 6)
Publications screened for clinical studies using SDF to arrest caries among children (n = 850)
PubMed (n = 169)
Embase (n = 46)
Scopus (n = 124)
China National Knowledge Infrastructure (n = 207)
Biblioteca Virtual em Saude (n = 189)
Biblioteca Virtual en Salud Espana (n = 5)
Ichushi-web (n = 110)
Publications excluded such as review, case report, laboratory studies, clinical studies on other aspects, etc. (n = 831)
PubMed (n = 161)
Embase (n = 46)
Scopus (n = 124)
China National Knowledge Infrastructure (n = 203)
Biblioteca Virtual em Saude (n = 186)
Biblioteca Virtual en Salud Espana (n = 4)
Ichushi-web (n = 107)
Publications included for evaluation (n = 19)
16 studies on primary teeth
3 studies on permanent teeth
38% SDF was used in 14 studies
30% SDF was used in 3 studies
12% SDF was used in 1 studies
(as a subgroup comparing with 38% SDF)
10% SDF was used in 2 studies
Publications found in bibliographies (n = 0)
Publications included for meta-analysis (n = 8)
Caries-arresting rate of studies using 38% SDF on primary teeth

Meta-analysis

Once this systematic review involves a *statistical aggregation* of results, we are in meta-analysis land

The idea here is to perform a "study on studies" that combines results in an *objectively verifiable manner*. In one sense, this effectively serves as a summary of current literature

By taking a high level approach, meta-analysis can possibly answer questions that individual studies themselves could not

Meta-analysis

For example, any individual study may be limited by sample size or population. Aggregating multiple studies is one way to *increase statistical power*

Different studies may also include different populations (sex, age, etc). Aggregation may shed light on how outcome changes in different subgroups, or if there are characteristics that have higher risk factors

In some cases gives insight into dose-response relationship

These are especially helpful when literature suggests conflicting results, or when planning future studies (clinical trials)

Examples

Here, combination of studies sheds light on potential risk factors associated with prostate cancer:

- Measures of sexual activity and STDs (Dennis & Dawson, 2002)
- Vasectomies (Dennis, Dawson & Resnick, 2002)
- Occupational exposure in tire and rubber manufacturing operations (Stewart et al, 2002)

Seemingly unrelated but can be combined to paint bigger picture

What are we interested in?

Like most everything we've covered, there are a number of outcomes that may be of interest

This includes things like mean differences, differences in proportions, odds ratios, or correlation. We'll call them *effect sizes* to refer to them in general

Where we need to be super-duper cautious, though, is in combining results associated with significance probabilities.

Not all p-values are created equally

Big picture

Of course, we can't just average effects sizes across studies and call it a day

Each study will have some associated amount of variability (more later), as well as different sample sizes

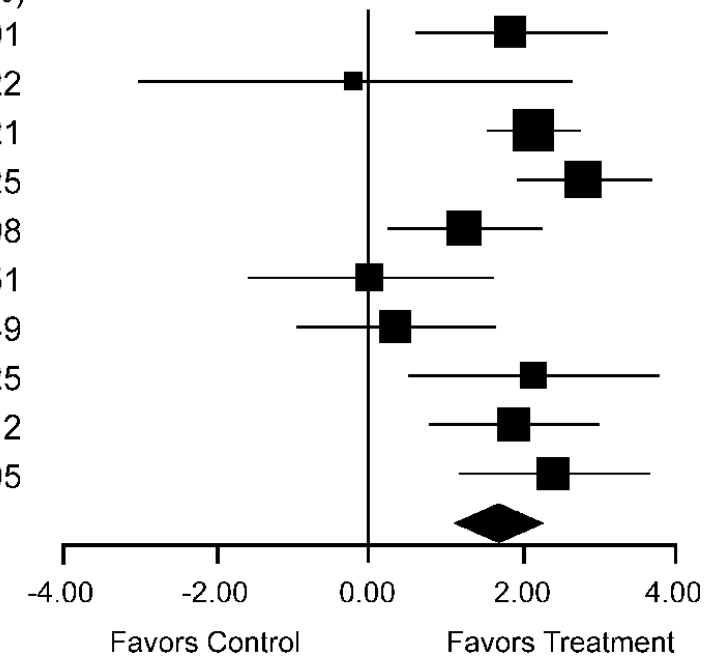
We need some way to bring all of this information together to get an informed estimate of the true *standard error*

Forest Plot

Study name	Difference in means	95% CI	P-value	Sample sizes Treatment	Sample sizes Control	Weight (%)
Study 1	1.847	[0.588, 3.106]	0.004	15	21	9.91
Study 2	-0.200	[-3.041, 2.641]	0.890	10	22	3.22
Study 3	2.159	[1.549, 2.769]	<0.001	15	19	16.21
Study 4	2.806	[1.913, 3.700]	<0.001	20	16	13.25
Study 5	1.240	[0.217, 2.263]	0.017	11	12	11.98
Study 6	0.000	[-1.606, 1.606]	1.000	14	18	7.51
Study 7	0.339	[-0.973, 1.651]	0.613	18	14	9.49
Study 8	2.139	[0.487, 3.791]	0.011	25	12	7.25
Study 9	1.900	[0.784, 3.016]	0.001	35	26	11.12
Study 10	2.400	[1.159, 3.641]	<0.001	40	21	10.05
Combined Estimate	**1.687**	**[1.130, 2.243]**	**<0.001**			

Statistics for each study

Difference in means and 95% CI



Heterogeneity: Chi^2=19.82, df=9, P=0.019; I^2=54.58%
Test for overall effect: P<0.0001

Forest Plot from Meta-Analysis of Hypothetical Data: Alveolar Ridge Preservation (mm)

Dawon, Philstrom & Blanchette (2016). The outcome is postextraction alveolar ridge preservation in mm. The effect size of interest is the difference in outcome between Tx and Control

Combining Effect Estimates

We saw in the forest plot that different studies have differing amounts of variability

What's of critical importance here is determining *where exactly that variability comes from*

In other words, the statistical methods we use to aggregate information will depend on whether or not we believe the effects to be homogenous among studies

Assessing Homogeneity

It's helpful to take our assumptions to the logical conclusions to determine their relevance. If we assume that studies are homogenous, we are claiming:

- These studies are measuring the exact same effect
- The variability in these measurements are all due to random error
- The samples in the study come from the same population
- and more!

In these situations, our statistical model corresponds to a *fixed effects model*

Assessing Heterogeneity

On the other hand, there are a number of reasons to suspect heterogeneity may be present:

- Different endpoints/length of study
- Varying definitions of disease or outcome
- Treatment regimens
- Study protocols and designs (which patients excluded, incentive, etc.)

Assuming heterogeneity, the observed variance in effects includes variability due to the study rather than variability in the outcome alone

In these situations, our statistical model corresponds to a *random effects model*

Assessing Heterogeneity

There are a number of statistical procedures that can be used to determine degrees of heterogeneity:

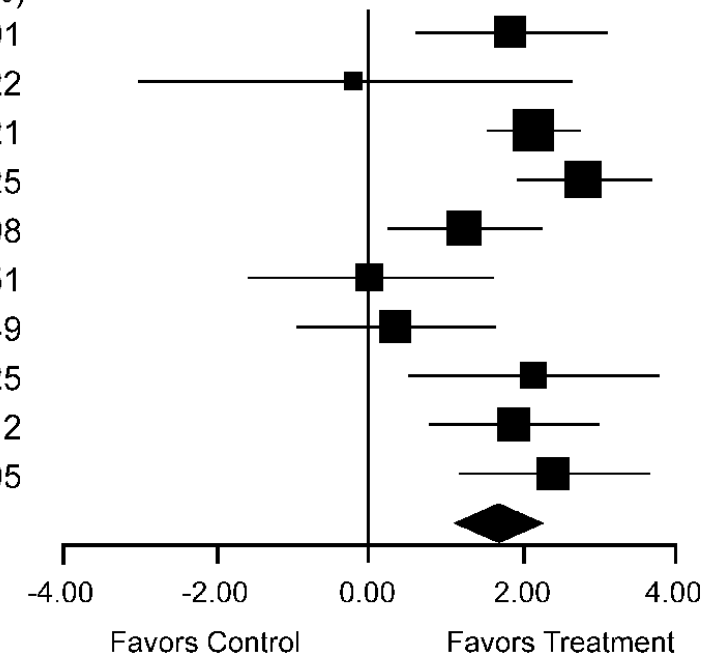
- *Q statistic* – a χ^2 test for heterogeneity
- I^2 – proprotion of observed variability among effect sizes that reflects real differences in effect size
- τ^2 (tau squared) – an estimate of the variance of the true effect sizes

Forest Plot

Study name	Difference in means	95% CI	P-value	Sample sizes Treatment	Sample sizes Control	Weight (%)
Study 1	1.847	[0.588, 3.106]	0.004	15	21	9.91
Study 2	-0.200	[-3.041, 2.641]	0.890	10	22	3.22
Study 3	2.159	[1.549, 2.769]	<0.001	15	19	16.21
Study 4	2.806	[1.913, 3.700]	<0.001	20	16	13.25
Study 5	1.240	[0.217, 2.263]	0.017	11	12	11.98
Study 6	0.000	[-1.606, 1.606]	1.000	14	18	7.51
Study 7	0.339	[-0.973, 1.651]	0.613	18	14	9.49
Study 8	2.139	[0.487, 3.791]	0.011	25	12	7.25
Study 9	1.900	[0.784, 3.016]	0.001	35	26	11.12
Study 10	2.400	[1.159, 3.641]	<0.001	40	21	10.05
Combined Estimate	**1.687**	**[1.130, 2.243]**	**<0.001**			

Statistics for each study

Difference in means and 95% CI



Heterogeneity: Chi^2=19.82, df=9, P=0.019; I^2=54.58%
Test for overall effect: P<0.0001

Forest Plot from Meta-Analysis of Hypothetical Data: Alveolar Ridge Preservation (mm)

Dawon, Philstrom & Blanchette (2016). The outcome is postextraction alveolar ridge preservation in mm. The effect size of interest is the difference in outcome between Tx and Control

Stratified Analysis

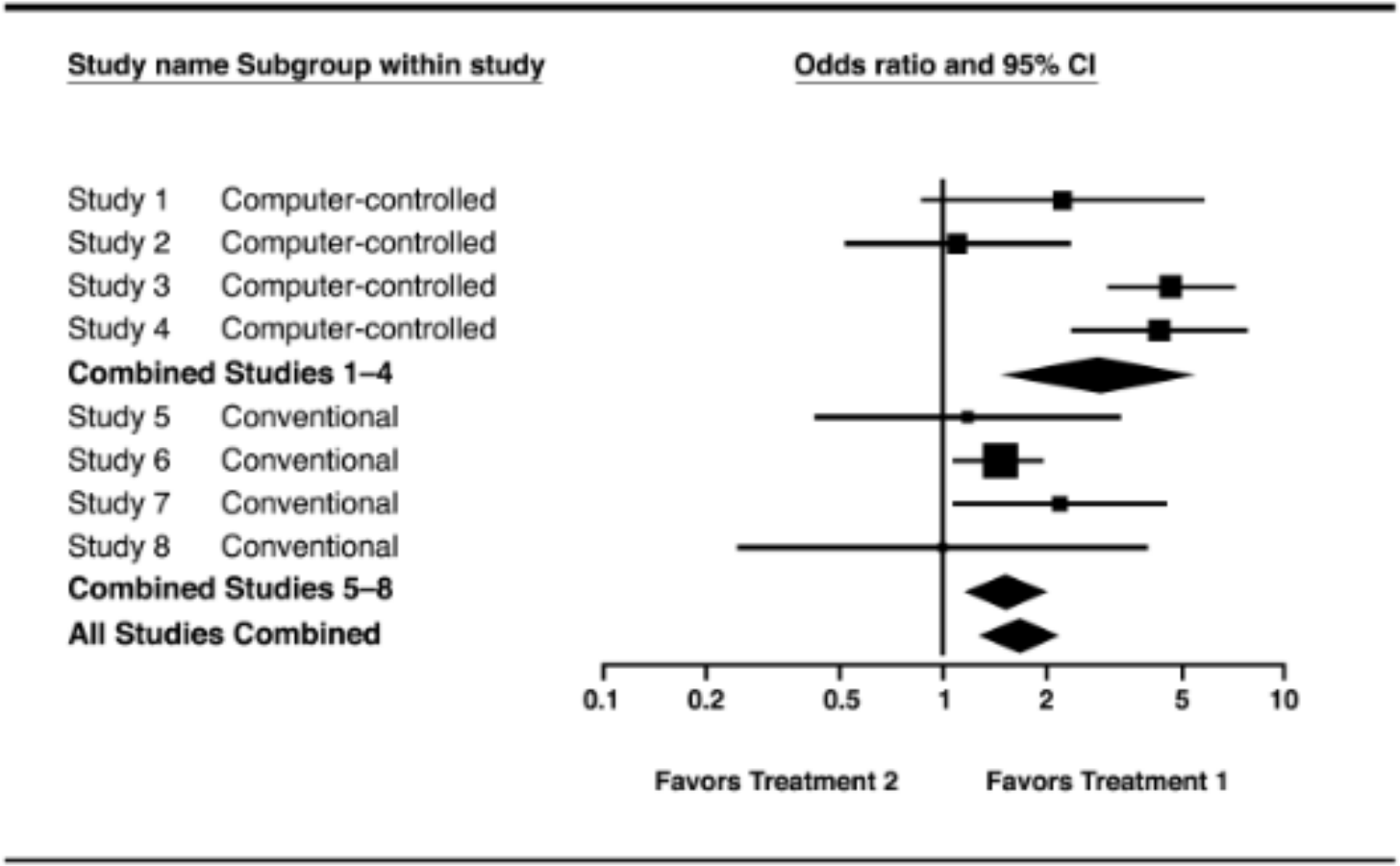
One way to help control for heterogeneity is to consider adding stratified levels to the study

In this example, reserachers interested in comparing efficacy of anesthetic in two treatments. The outcome was "anesthetic success"

However, in some of the studies, anesthetic delivery was computer controlled, while the other studies used conventional delivery

Stratified Forest Plots

Figure 2A



Forest Plot from Meta-Analysis of Hypothetical Data: Comparison of Two Anesthetics

Dawson, Pihlstrom & Blanchette (2016)

Exploratory analysis

From this last forest plot, an interesting thing pops up – possibly we could explore the difference in effect between computer-controlled and conventionally delivered anesthesia

This type of sub-analysis could be extended in any number of ways – effects of different populations, variants in protocol, or types of control

Studies with different doses of active drug can also be stratified, giving an assessment of the possible dose-response curve

Where these may not be definitive conclusions themselves, these types of exploratory analyses can create launch points for further investigation

Assessment of Quality – Bias

The most important type of bias in meta-analysis studies comes from *publication bias*

Consider that:

- Negative studies are less likely to be published
- When negative studies are published, they often include less information
- Researchers performing the meta-analysis could potentially limit themselves to findings that verify their own assumptions

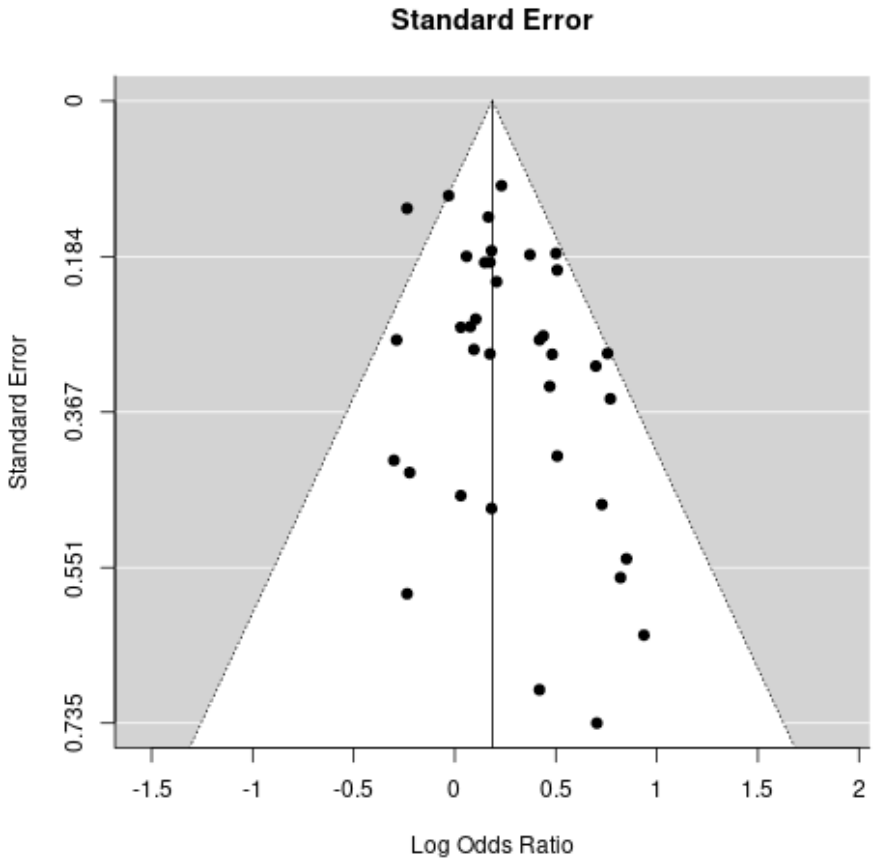
This last point in particular reinforces the importance of specifying a search strategy a priori, as we saw earlier in the lecture

Funnel Plot

We can consider visual way to examine publication bias

We start by plotting effect size against standard error (or sample size; there are variations)

If there is no publication bias, the plot *should* be shaped like a funnel

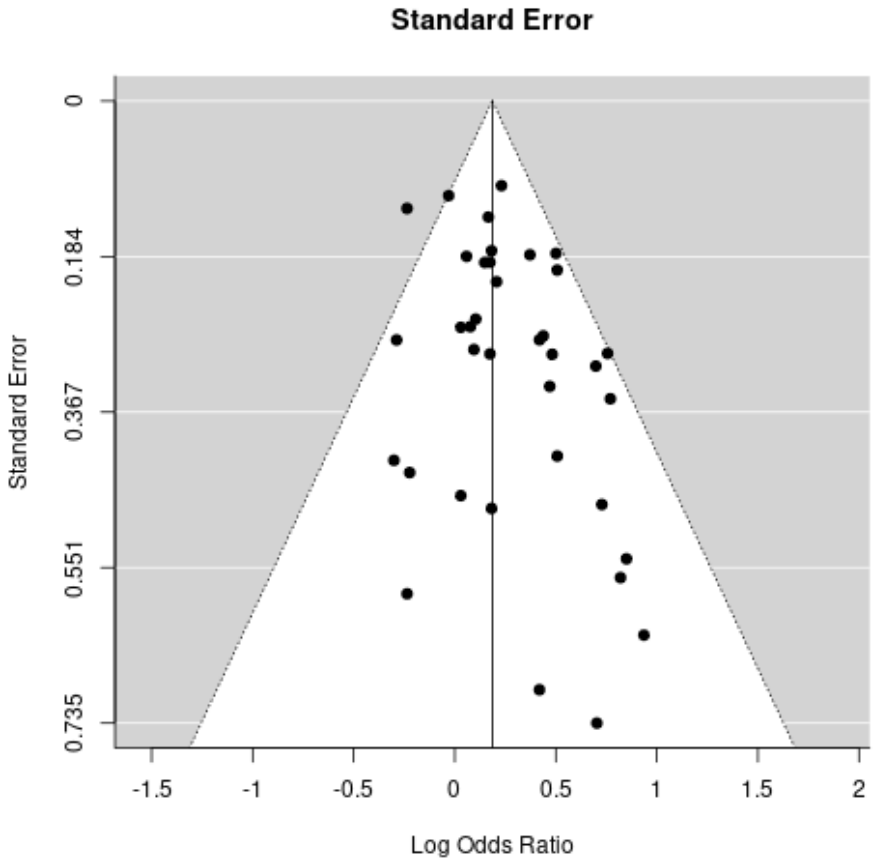


Why a funnel?

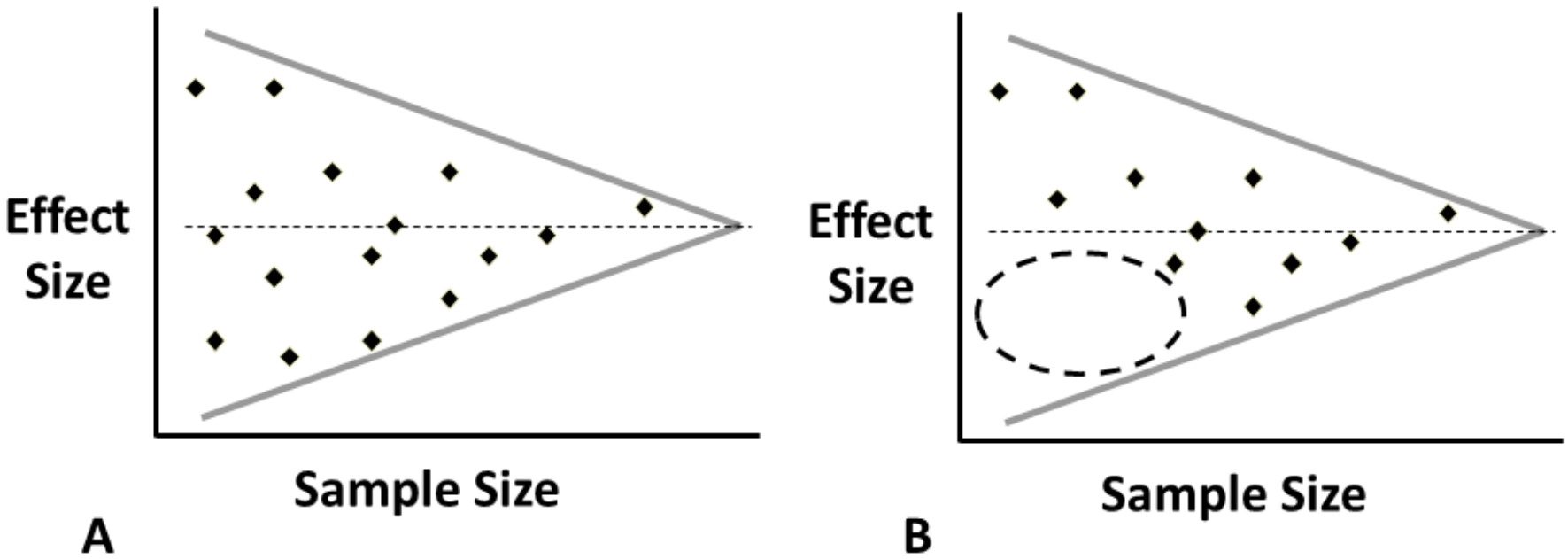
First, assuming a true effect size of δ, regardless of variance, studies should by symmetric to this line

Studies with larger S.E. should be more variable, with smaller S.E being less – this is indicated by tops and bottom of triangle

Asymmetry suggests the possibility of publication bias



Funnel Plot



Dawon, Philstrom & Blanchette (2016)

Quality Assessment

There is an entire field of literature for assessing quality of meta-analysis. General things to check for include:

- Clearly defined research questions
- Study participants described
- Does the conclusion make sense
- Methods for imputation and missing data
- Randomization, masking/blinding, study procedures done correctly
- Did authors use correct statistical methods?

Meta-analysis Quality Checklists

PRISMA – Preferred Reporting Items for Systematic reviews and Meta-Analyses

QUOROM – QUality Of Reporting Of Meta-Analyses

MOOSE – Meta-analysis Of Observational Studies in Epidemiology

Considering GWAS

Next week, we will be considering genome-wide association studies (GWAS)

This involves looking at *millions* of SNPs across thousands of individuals looking for assocations of phenotype and gene expression

Some of the issues with this we have already discussed, such as population stratification, multiple testing and Type I error, and some we have not, including imputation and handling missing data

In light of this, consider the potential difficulty in performing GWAS based meta-analysis

GWAS Meta-analysis issues

- Different platforms for recording expression (microarrays vs sequencing)
- Different SNP panels included in different studies
- Population stratification/different expression by sex
- No standard definition of phenotypes
- Handling of missing data

Final Notes

Like everything else, meta-anlaysis isn't without controversy

"Meta-analysis-schmeta-analysis . . . " (Shapiro, 1994)

"Statistical alchemy for the 21^{st} *century . . . "* (Feinstein, 1995)

Really, no problem so long as the limitations are kept in mind:

1. Meta-analysis cannot produce simple solutions to complex problems.
2. Meta-analysis approaches cannot make up for shortcomings of the studies or the data upon which the analysis is based
3. Meta-analysis is simply a tool for summarizing existing research within a prescribed set of bounds

Resources and references

Deb Dawson's GENE:6234 lecture notes (2020)

Needleman provides a guide to systematic reviews (Journal of Clinical Periodontology 2002, 29 (Suppl 3):6-9).

Dawson DV, Pihlstrom BL, Blanchette DR. Understanding and evaluating meta-analysis. J Am Dent Assoc 2016; 147(4):264-70

COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS - http://handbook.cochrane.org and http://www.cochranelibrary.com