

Meta-analysis: A Brief Introduction

Mike W.-L. Cheung, PhD

Assistant Professor

Department of Psychology

National University of Singapore

10 January 2008

(Last update: 9 April 2009)

1.1 A brief history

1. Karl Pearson (1904) was probably the first one using meta-analytic techniques to combine the correlations between inoculation for typhoid fever and mortality for five independent samples.
2. The term “meta-analysis” was coined by [Gene V. Glass](#) in educational research.
3. It is known as validity generalization in Industrial and Organizational Psychology (Hunter & Schmidt, 2004).
4. It is also related to systematic review in medical research.

1.2 Types of data analyses

1. Primary data analysis: Data are collected and analyzed for specific research questions.
2. Secondary data analysis: Existing data sets are re-analyzed to address research questions that are different from the original ones.
3. Meta-analysis: “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass, 1976, p.3).

2.1 Problems in empirical research

An artificial example on the correlations between income and self-esteem

Study	Sample size (n_i)	Correlation (r_i)	p values
1	26	.13	.53
2	42	.35	.03
3	20	-.10	.68
4	40	.30	.07

- 1.If you were the researchers conducting Studies 2 and 4, you might argue that there was a positive relationship.
- 2.If you were the researchers conducting Studies 1 and 3, you might conclude that there was no linear relationship.
- 3.The results look contradicting.

2.2 Problems of individual study

1. Measurement artifacts: unreliability, restriction of range
2. Small sample sizes
3. Low statistical power

2.3 Issues in narrative review

1. Subjectivity: How to judge whether the above studies are consistent?
2. Confirmation bias: Researchers tend to focus on studies consistent with their own expectations.
3. No quantifiable summary on the findings

2.4 Objectives of meta-analysis

1. Draw general conclusions in a particular topic.
2. Test the homogeneity (consistency) of the existing findings.
3. Estimate an average effect size.
4. Test potential moderators if the studies are heterogeneous.

2.5 Steps in meta-analysis

1. Formulate research questions.
2. Search relevant studies.
3. Classify and code the studies.
4. Convert the reported statistics into a common effect size.
5. Conduct a meta-analysis.
6. Interpret the results.
7. Conduct sensitivity analyses

2.6 Formulating research questions

1. When to conduct a meta-analysis?

1. There are sufficient number of studies addressing similar research questions.
2. The outcomes are quantifiable, e.g., correlation, odds ratio.
3. The outcomes are comparable across studies.

2. Is it too obvious to conduct a meta-analysis?

2.7 Searching relevant studies

1. Well-defined inclusion/exclusion criteria:

1. Keywords

2. Key authors

3. Time period

2. Computerized databases, e.g., PsycInfo, PubMed, Medline, Web of Science, Dissertation abstracts for unpublished papers

3. Manual searching on specific journals

2.8 Coding the studies

1. Study characteristics and potential moderators:

1. Samples: race, age (mean, *SD* and range), gender ratio, sample sizes
2. Methodology: experimental vs. observational studies, clinical trials
3. Measurement used: how to measure the outcome variables?

2. Effect sizes:

1. What effect sizes should we use?
2. Some effect sizes can be converted from one type into the others (see Rosenthal, 1994).

3.1 Effect sizes

1. Effect size is a scale-free measure on the strength of relationship.
2. Effect size is less sensitive to the sample size. However, its sampling variance does depend on the sample size.
3. Meta-analysis relies heavily on the availability of the effect sizes and their sampling variances. If the effect sizes are not available from some of the studies, these studies may have to be excluded from the meta-analysis.
4. Some common effect sizes:

<i>Effect sizes</i>	<i>Estimates (T_i)</i>	<i>Sampling variances (v_i)</i>
<p>Experimental studies: Standardized mean differences</p> <ol style="list-style-type: none"> Cohen' d (biased) Hedges' g (less biased; preferable in meta-analysis) 	$d_i = \frac{\bar{X}_{1i} - \bar{X}_{2i}}{s_i} \text{ where}$ $s_i = \sqrt{\frac{(n_{1i} - 1) s_{1i}^2 + (n_{2i} - 1) s_{2i}^2}{(n_{1i} - 1) + (n_{2i} - 1)}}$ $g_i = 1 - \frac{3}{(4m_i - 1)d_i} \text{ where}$ $m_i = n_{1i} + n_{2i} - 2$	<p>Sampling variance of g_i:</p> $v_i = \frac{n_{1i} + n_{2i}}{n_{1i} * n_{2i}} + \frac{g_i^2}{2(n_{1i} + n_{2i})}$

<i>Effect sizes</i>	<i>Estimates (T_i)</i>	<i>Sampling variances (v_i)</i>
Observational studies: Pearson correlation (r)	$r_i = \frac{cov(x, y)}{SD_x SD_y}$	$v_i = \frac{(1 - r_{ij}^2)^2}{n_i}$
Fisher's z score (z)	$z_i = 0.5 * \log \frac{(1 + r_i)}{(1 - r_i)}$	$v_i = \frac{1}{n_i - 3}$

Note. Hunter and Schmidt (2004) uses n_i as the weights in pooling Pearson r s. Their approach is very popular in Industrial and Organizational Psychology.

In clinical trials:

	Failure/Dead	Success/Alive
New treatment	a	b
Control	c	d

1. Odds of the new treatment: a/b

1. It expresses the likelihood of an occurrence (Failure) relative to the likelihood of a nonoccurrence (Success).

2. Odds range from 0 to ∞ (asymmetric):

3. 0 (surely success) \rightarrow **1.0 (50-50)** \rightarrow ∞ (surely failure)

2. We are usually interested in comparing two groups.

3. Odds ratio (OR) = odds of treatment / odds of control = $\frac{a/b}{c/d} = \frac{ad}{bc}$

1. It is interpreted as “the odds of new treatment is XX times as likely as the odds of control.”

2. When some of the cells are 0, 0.5 may be added into all the cells.

3. OR ranges from 0 to ∞ (asymmetric):

4. **0** (preferring new treatment) \rightarrow **1.0 (the same)** \rightarrow ∞ (preferring control)

5. $\ln(\text{OR})$ ranges from $-\infty$ to ∞ (symmetric):

6. $-\infty$ (preferring new treatment) \rightarrow **0.0 (the same)** \rightarrow ∞ (preferring control)

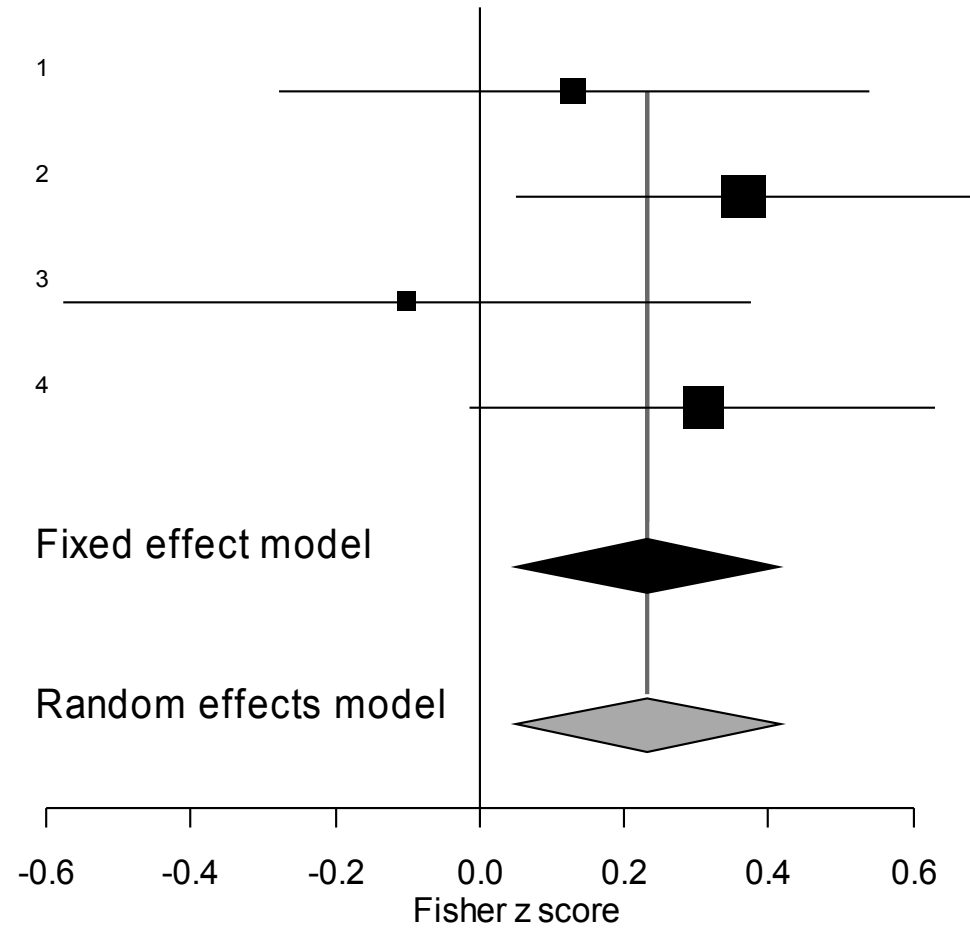
<i>Effect sizes</i>	<i>Estimates (T_i)</i>	<i>Sampling variances (v_i)</i>
Clinical trials: Odds ratio (OR)	$\ln(\text{OR})$	Sampling variance of $\ln(\text{OR}): \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$

1. Natural logarithm is used to “normalize” the sampling distribution of OR.
2. This is similar to the Fisher's z transformation on correlation coefficients.

3.2 Forest plots

1. They are usually used to visualize the results in a meta-analysis.
2. An observation from the next forest plot:
 1. The confidence interval (CI) of the pooled effect size is smaller than the CIs of individual studies.
3. Example 1: The artificial example on the correlations between income and self-esteem

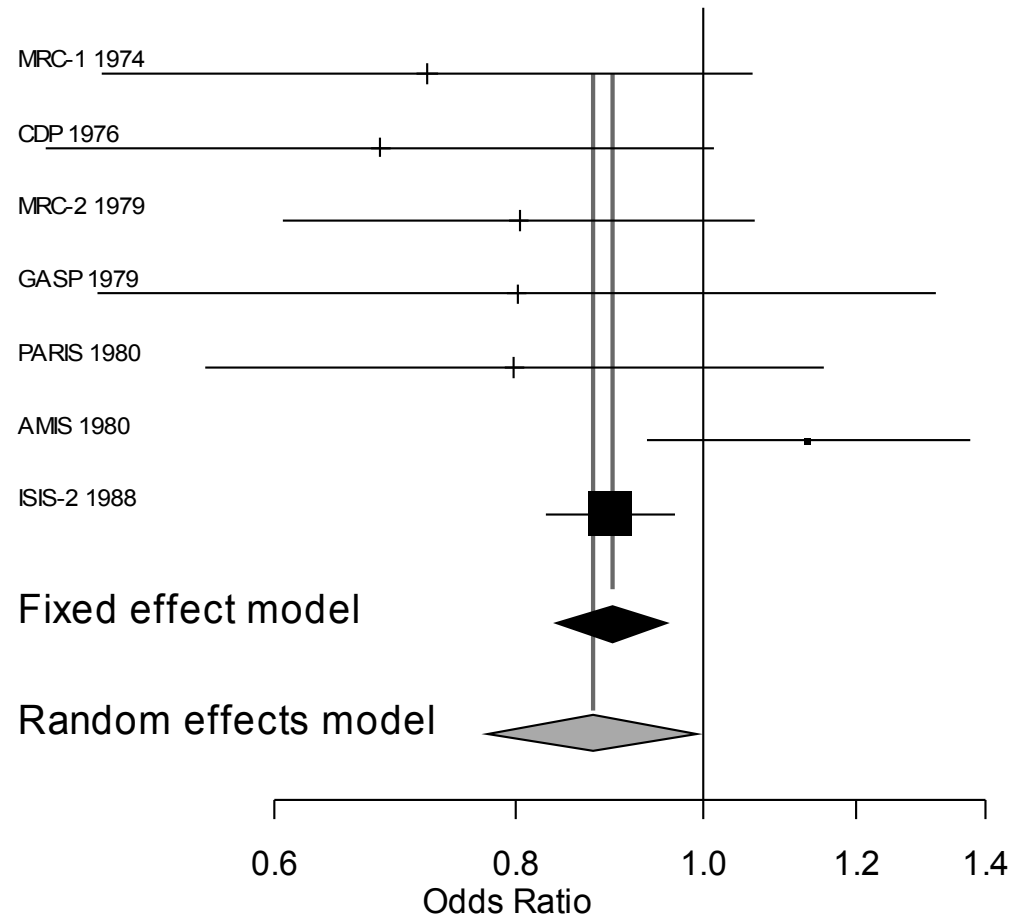
Study	Sample size (n_i)	Correlation (r_i)	p values
1	26	.13	.53
2	42	.35	.03
3	20	-.10	.68
4	40	.30	.07



An example from Fleiss (1993): Seven randomised controlled trials of the effectiveness of aspirin versus placebo in preventing death after heart attack.

Study	Year of publication	OR	ln(OR)	Var of ln(OR)
MRC-1	1974	0.742046	-0.298344	0.038562
CDP	1976	0.699291	-0.357688	0.040969
MRC-2	1979	0.827038	-0.189905	0.020119
GASP	1979	0.820853	-0.197411	0.063957
PARIS	1980	0.819326	-0.199274	0.034693
AMIS	1980	1.118333	0.111839	0.009515
ISIS-2	1988	0.914173	-0.089736	0.001451

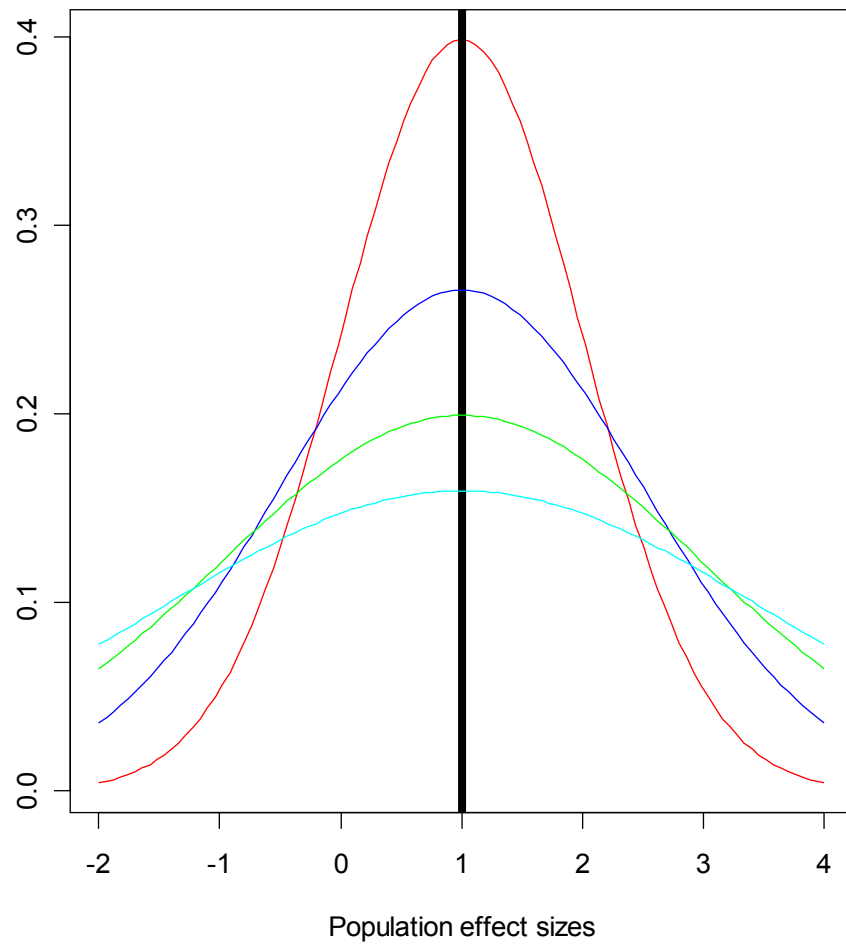
Note. OR=1 means the effectiveness of aspirin and placebo is the same.



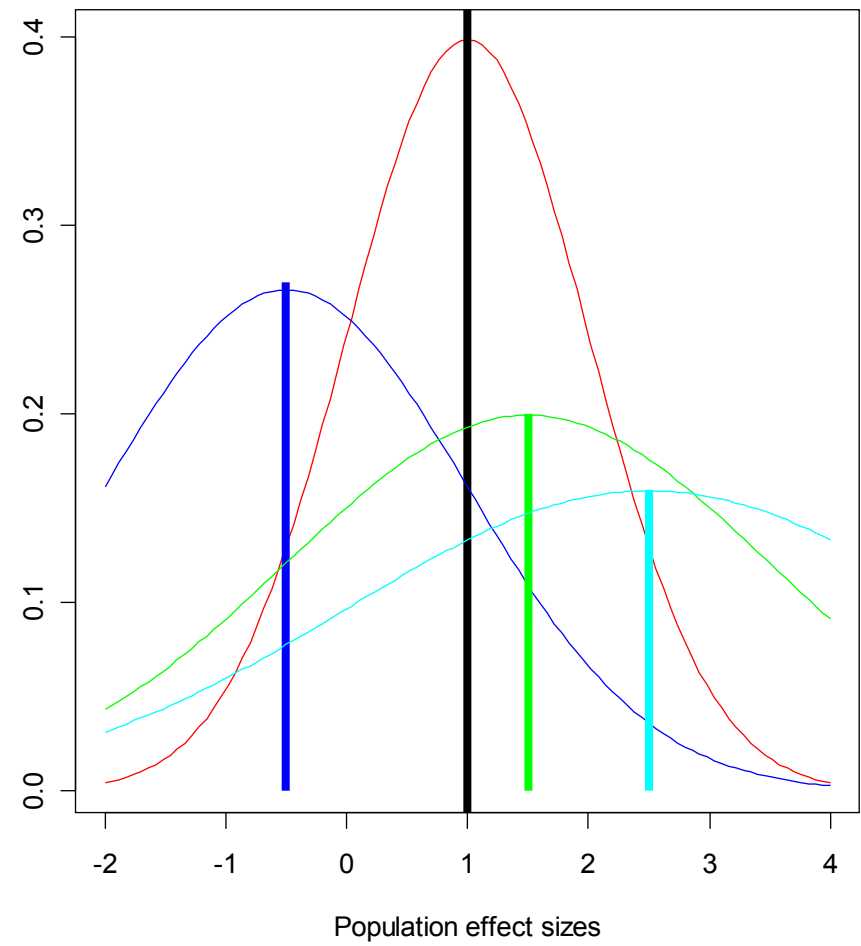
4.1 Statistical models for meta-analysis

	No moderator	With moderators
Fixed-effects models	Fixed-effects meta-analysis	Meta-regression (in medical research)
Random-effects models	Random-effects meta-analysis	Mixed-effects meta-analysis/Meta-regression

Fixed-effects model



Random-effects model



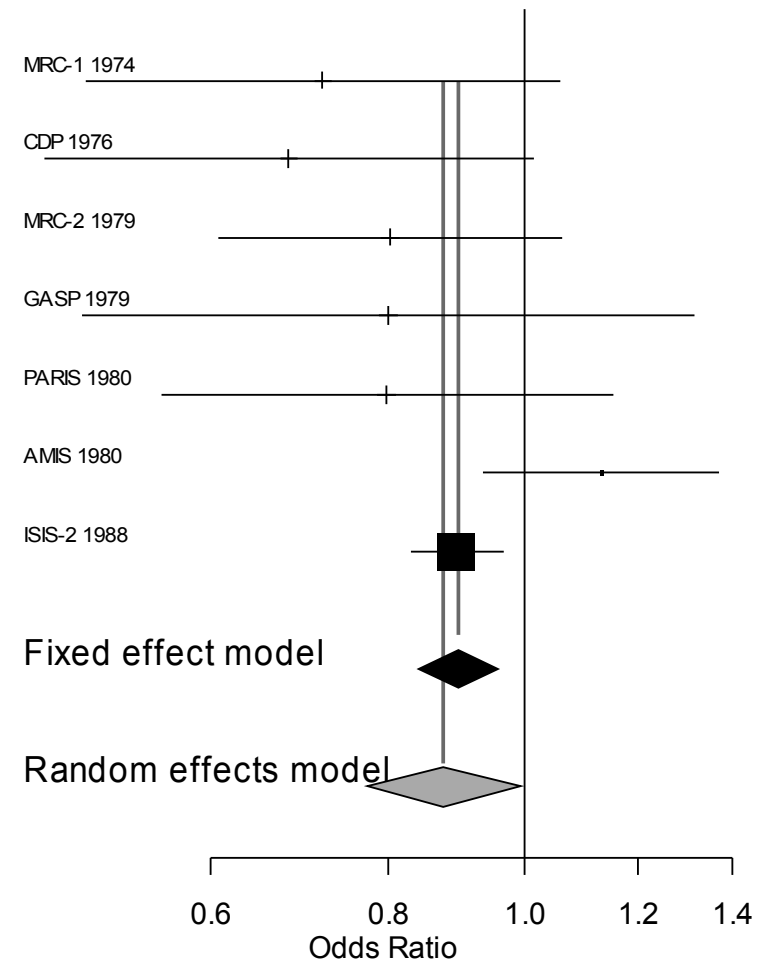
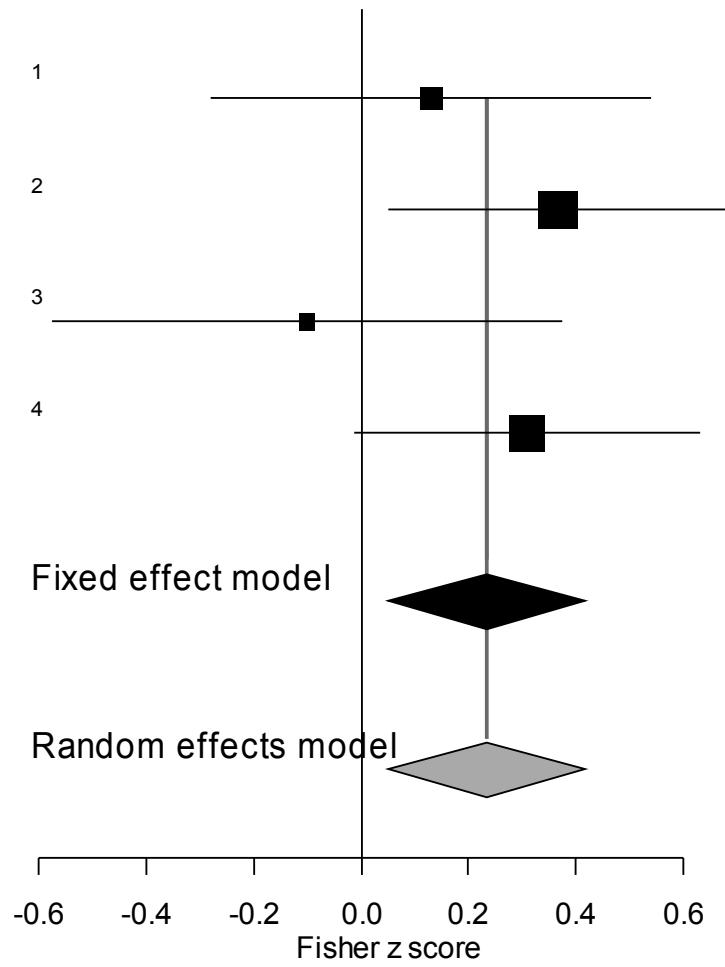
1. Fixed-effects models:

1. The population effect sizes are assumed equal for all studies.
2. The differences on the observed effect sizes are due to sampling error.
3. *We cannot generalize* the findings beyond the studies included in the meta-analysis.
4. Most published meta-analytic studies use fixed-effects models.

2. Random-effects models:

1. The population effect sizes may be different across studies.
2. The differences on the observed effect sizes are due to a combination of true difference (variance component) and sampling error.
3. *We can generalize* the findings beyond the studies included in the meta-analysis.

4. It is usually preferred methodologically.
3. Comparison of the standard errors (*SEs*) between fixed- and random-effects models:
 1. The *SEs* based on the fixed-effects models are correct only if the effect sizes are homogeneous.
 2. The *SEs* based on the fixed-effects models are under-estimated if the effect sizes are heterogeneous.
 3. The *SEs* based on the random-effects models are more appropriate under the random-effects models.
 4. The *SEs* based on the random-effects models are usually larger than those based on the fixed-effects models.



4.2 Fixed-effects meta-analysis

1. Model:

1. $T_i = \delta + e_i$,

2. where T_i is the observed effect size, δ is the population effect size, and $v_i = \text{var}(e_i)$ is the conditional sampling variance of T_i .

3. There is only one δ . That is, all effect sizes are assumed equal across studies.

4. v_i is assumed known in each study. This assumption is reasonable when the sample sizes are reasonably large (say > 30).

2. Pooled effect size:

$$1. \bar{T}_F = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i},$$

2. where $w_i = 1/v_i$ is the weight and k is the number of studies.

3. Sampling variance of \bar{T}_F :

$$1. s_F^2 = \frac{1}{\sum_{i=1}^k w_i}$$

2. We may test whether the pooled effect size is statistically significant by using $Z = \bar{T}_F / s_F$ which has a standard normal distribution.

Example: Correlation coefficient

Study	Sample size (n_i)	Correlation (r_i)	z_i	Var(z_i)	Weight: $w_i = 1/\text{var}(z_i)$
1	26	0.13	0.1307	0.0435	23
2	42	0.37	0.3654	0.0256	39
3	20	-.10	-0.1003	0.0588	17
4	40	0.31	0.3095	0.0270	37

1. Pooled effect size and its sampling variance:

$$1. \bar{T}_F = \frac{\sum w_i z_i}{\sum w_i} = 0.23 \quad \text{and} \quad s_F^2 = \frac{1}{\sum_{i=1}^k w_i} = 0.009$$

2. $Z = \bar{T}_F / s_F = 0.23 / 0.009 = 26.7$. Thus, the pooled effect size is statistically significant at .001.

4.3 Testing the homogeneity of the effect sizes

1. Fixed-effects model is appropriate only if the effect sizes are homogeneous, i.e., $H_0: \delta = \delta_1 = \delta_2 = \dots = \delta_k$.

2. If the effect sizes are heterogeneous, the calculated *SEs* are underestimated (smaller than the “true” *SEs*).

3. Under the assumption of homogeneity, the test statistic Q_T has a chi-square distribution with $df = \text{no. of studies} - 1$.

$$1. Q_T = \sum w_i (T_i - \bar{T}_F)^2$$

$$2. \text{In our example, } Q_T = \sum w_i (T_i - \bar{T}_F)^2 = 3.03, df = 3, p = .387$$

3. Thus, we cannot reject the null hypothesis that all effect sizes are equal.

4. It is found that the power of this homogeneity test is not very high.

5.If the number of studies is small, this test may not be powerful enough to reject the null hypothesis even it is wrong.

4.4 Random-effects meta-analysis

1. Model:

1. $T_i = \delta + u_i + e_i$,

2. where T_i is the observed effect size, δ is the “super” population effect size, $v_i = \text{var}(e_i)$ is the conditional sampling variance of T_i , and u_i is the study specific effect.

3. The meanings of δ are different under the fixed- and random-effects models.

4. $\tau^2 = \text{var}(u_i)$ indicates the variability on the effect sizes which is larger than the sampling error.

2. Estimating the variance component:

1. The first step is to estimate the variance component $\hat{\tau}^2$ (e.g.,

DerSimonian & Laird, 1986; Viechtbauer, 2005).

2. As the value on $\hat{\tau}^2$ cannot be interpreted directly, Higgins and Thompson (2002) proposed several indices to quantify the degree of heterogeneity. One of them is the I^2 index.

1. It can be interpreted as the proportion of total variation due to heterogeneity between studies.

2. $I^2 = 0$: the studies are homogeneous.

3. $I^2 = 0.3$: 30% of the total variation is due to heterogeneity between studies.

3. Once the estimated variance component $\hat{\tau}^2$ is available, we may compute a new weight $\tilde{w}_i = 1 / (v_i + \hat{\tau}^2)$.

4. Then we may apply what we have learned in fixed-effects model.

5. *Notes:*

1. When $\hat{\tau}^2 = 0$, random-effects model becomes fixed-effects model.

2. When $\hat{\tau}^2$ is small, results based on fixed- and random-effects models are similar.

3. When $\hat{\tau}^2$ is large, results based on fixed- and random-effects models are different.

3. Pooled effect size:

$$1. \bar{T}_R = \frac{\sum_{i=1}^k \tilde{w}_i T_i}{\sum_{i=1}^k \tilde{w}_i}, \text{ where } \tilde{w}_i = 1 / (v_i + \hat{\tau}^2) .$$

4. Sampling variance of \bar{T}_R :

$$1. \quad s_R^2 = \frac{1}{\sum_{i=1}^k \tilde{w}_i}$$

2. We may test whether the pooled effect size is statistically significant by using $Z = \bar{T}_R / s_R$ which has a standard normal distribution.

Our correlation coefficient example:

		95%-CI	z	p.value
Fixed effects model	0.2328	[0.0508; 0.4148]	2.5074	0.0122
Random effects model	0.2323	[0.0493; 0.4153]	2.4880	0.0128

Quantifying heterogeneity:

$\tau^2 = 4e-04$; $H = 1.01$ [1; 2.57]; **$I^2 = 1\%$** [0%; 84.8%]

Test of heterogeneity:

Q	d.f.	p.value
3.03	3	0.387

Our odds ratio example:

	OR	95%-CI	z	p.value
Fixed effects model	0.8969	[0.8405; 0.9570]	-3.2876	0.001
Random effects model	0.8763	[0.7743; 0.9917]	-2.0918	0.0365

Quantifying heterogeneity:

$\tau^2 = 0.0096$; $H = 1.29$ [1; 1.99]; $I^2 = 39.7\%$ [0%; 74.6%]

Test of heterogeneity:

Q	d.f.	p.value
9.95	6	0.1269

4.5 Mixed-effects model

1. Sometimes, we may want to model the effect sizes with the study characteristics.

2. Study characteristics may be in the form of:

1. Categorical, e.g., experimental vs. observational studies

2. Continuous, e.g., year of publication, duration of intervention

3. Model:

1. $T_i = \delta + \beta X + u_i + e_i$,

2. where T_i is the observed effect size, δ is the “super” population effect size, $v_i = \text{var}(e_i)$ is the conditional sampling variance of T_i , u_i is the study specific effect, and X is the covariate.

An example in Hox (2002): http://www.ats.ucla.edu/stat/HLM/examples/ma_hox/chapter8.htm

Study	d_i	v_i	$w_i=1/v_i$	rii	$ntot$	$weeks$
1	-0.264	0.086	11.63	0.9	47	3
2	-0.23	0.106	9.43	0.75	38	1
3	0.166	0.055	18.18	0.75	74	2
4	0.173	0.084	11.90	0.9	48	4
5	0.225	0.071	14.08	0.75	57	3
6	0.291	0.078	12.82	0.75	53	6
7	0.309	0.051	19.61	0.9	80	7
8	0.435	0.093	10.75	0.9	51	9
9	0.476	0.149	6.71	0.75	32	3
10	0.617	0.095	10.53	0.75	46	6
11	0.651	0.11	9.09	0.75	56	6
12	0.718	0.054	18.52	0.9	79	7
13	0.74	0.081	12.35	0.75	55	9
14	0.745	0.084	11.90	0.9	51	5

15	0.758	0.087	11.49	0.9	59	6
16	0.922	0.103	9.71	0.9	56	5
17	0.938	0.113	8.85	0.75	45	5
18	0.962	0.083	12.05	0.9	54	7
19	1.522	0.1	10.00	0.9	66	9
20	1.844	0.141	7.09	0.75	45	9

1. Potential moderators:

1. *rii*: reliability of the outcome measures

2. *ntot*: total number of participants

3. *weeks*: duration of the experiments

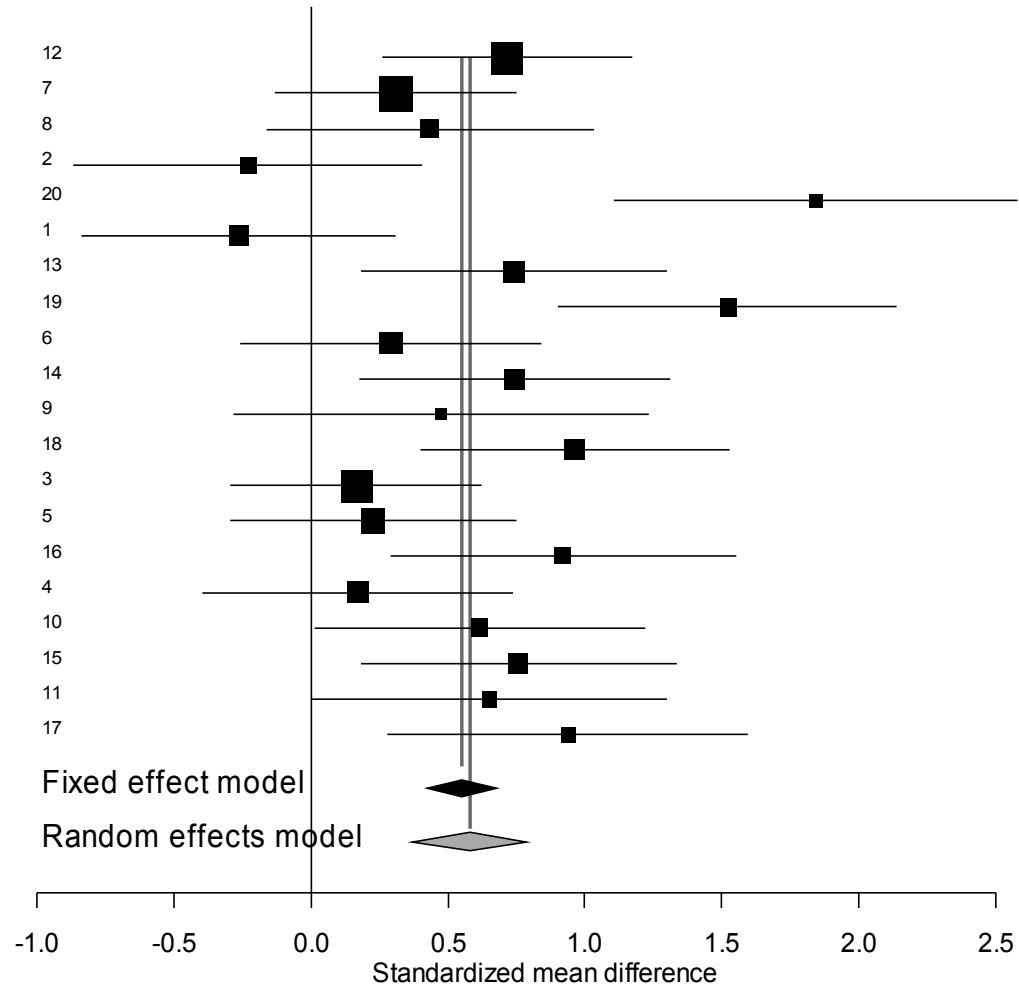
2. Research questions:

1. Are the effect sizes homogeneous?

2. If heterogeneous, are the moderators significant in predicting the effect sizes?

3. Which moderators are significant in predicting the effect size?

4. The model is $T_i = \delta + \beta_1 * rii + \beta_2 * ntot + \beta_3 * weeks + u_i + e_i$



3. Results based on the fixed- and random-effects model:

1. The effect sizes are heterogeneous.

2. $\hat{\tau}^2 = 0.1366$ and $I^2 = 0.617$

3. We may want to model the effect size by including the covariates.

		95%-CI	z	p.value
Fixed effects model	0.5502	[0.4228; 0.6776]	8.4650	< 0.0001
Random effects model	0.5795	[0.3715; 0.7875]	5.4609	< 0.0001

Quantifying heterogeneity:

$\tau^2 = 0.1366$; $H = 1.62$ [1.27; 2.06]; $I^2 = 61.7\%$ [37.8%; 76.4%]

Test of heterogeneity:

Q	d.f.	p.value
49.59	19	0.0002

4. Results based on the mixed-effects model:

1. The variance component $\hat{\tau}^2$ drops from 0.1366 to **0.0493** after including the covariates.

2. The omnibus test on all covariates is significant. It suggests that some of the covariates may be useful in predicting the effect size.

3. *Weeks* is significant in predicting the effect size.

Estimate of (Residual) Heterogeneity: 0.0493

Omnibus Test of all Moderators:

QME = 16.3944

df = 3

p-value = 9e-04

Individual Moderator Tests:

	estimate	SE	zval	pval	CI_L	CI_U
intrcpt	0.3843	0.9234	0.4161	0.6773	-1.4256	2.1942
rii	-0.5510	1.2010	-0.4588	0.6464	-2.9049	1.8029
ntot	-0.0036	0.0070	-0.5076	0.6118	-0.0174	0.0102
weeks	0.1506	0.0376	4.0107	0.0001	0.0770	0.2242

5.1 Extensions of meta-analysis

1. The techniques discussed so far focus on one single effect size.
2. Studies and research questions may involve more than one effect size.
3. I am going to illustrate two extensions of meta-analysis:
 1. Multivariate meta-analysis
 2. Meta-analytic structural equation modeling

5.2 Multivariate meta-analysis

1. An example: Hedges and Olkin (1985) reported four studies that examined the effectiveness of open education.
2. Effectiveness is defined by two measures: attitudes toward school and reading achievement.
3. We have to analyze these two outcomes together (Becker, 1992; 1995; Hedges & Olkin, 1985)

<u>Study</u>	<u>Effect sizes</u>		<u>Variances and covariances among effect sizes</u>		
	$d_{attitude}$	$d_{achievement}$	$\text{Var}(d_{attitude})$	$\text{Cov}(d_{attitude} , d_{achievement})$	$\text{Var}(d_{achievement})$
1	0.458	0.100	0.0513	0.0319	0.0501
2	0.363	0.241	0.0354	0.0222	0.0351
3	0.162	-0.121	0.0546	0.0344	0.0545
4	0.294	0.037	0.0286	0.0179	0.0286

4.Results:

1.The effect sizes are homogeneous across these studies with

$$\chi^2(6)=2.25, p=.90 .$$

2.The weighted mean effect for attitude is significant while the weighted mean effect for achievement is non-significant.

3.Thus, open education is only useful in improving the attitude toward school.

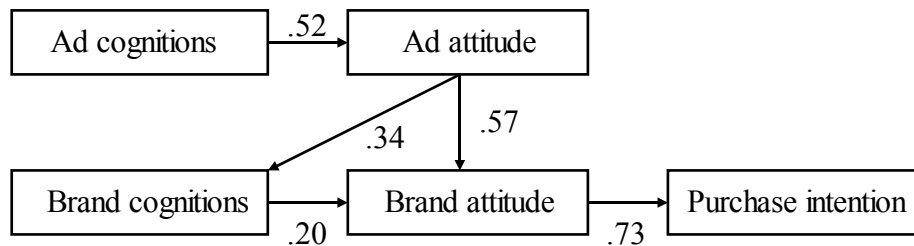
	beta	SE	z	p values
Attitude	0.32096945	0.09949242	3.2260694	0.001255029
Achievement	0.07779711	0.09912832	0.7848122	0.432563706

5.3 Meta-analytic structural equation modeling (MASEM)

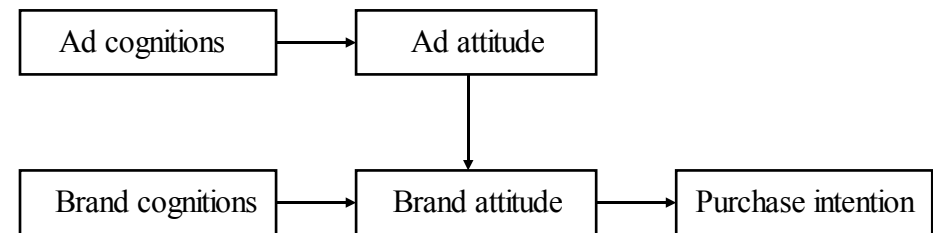
1. Structural equation models (SEM) are often used to fit hypothetical models in behavioral sciences.
2. Researchers may propose different models that make the comparison difficult.
3. MASEM combines both meta-analytic techniques and SEM together.
4. An example: Brown and Stayman (1992):
 1. Antecedents and consequences of attitudes toward the advertisement (Ad)
 2. No. of variables: 5
 3. No. of studies: 47
 4. Pooled sample size across studies: +4,600

Four alternative models of Ad attitude

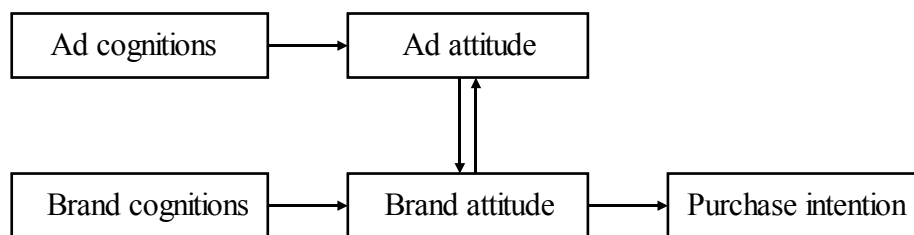
Dual mediation hypothesis



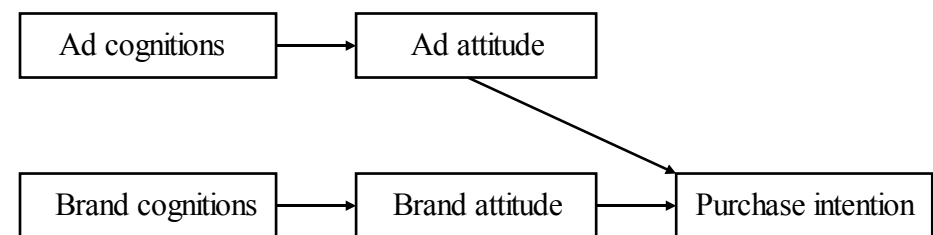
Affect transfer hypothesis



Reciprocal mediation hypothesis



Independent influences hypothesis



5. A two-step procedure on fitting MASEM (Cheung & Chan, 2005, 2009)

1. First stage:

1. Test the homogeneity of correlation matrices

2. Estimate a pooled correlation matrix

2. Second stage:

1. Fit proposed models on the pooled correlation matrix

6.1 Issues in meta-analysis

1. Although meta-analysis is very powerful, it is not without problems.

2. Combining apples and oranges

1. Studies are usually with various designs, samples and measures.

2. The combined effect size sometimes may not make sense.

3. Possible solutions:

1. Clearer definitions of inclusion criteria;

2. Designs and samples as potential moderators, e.g., experimental vs. observational studies, cultural groups

3. Publication bias

1. File drawer problems:

1. Non-significant findings are less likely to be submitted and accepted

for publication.

2. Published findings are more likely to be significant than unpublished findings.

2. Possible solutions:

1. Include unpublished findings whenever possible

2. Fail-safe-N

3. Funnel plot

4. Trim-and-fill method

6.2 Fail-safe-N

1. The no. of new studies (k_0) with null effect required to “neutralize” a significant effect.

2. $k_0 > -k + \frac{(\sum_{i=1}^k Z_i)^2}{1.96^2}$, where k is the no. of studies in a meta-analysis and

Z_i is the standard z score for the i th study.

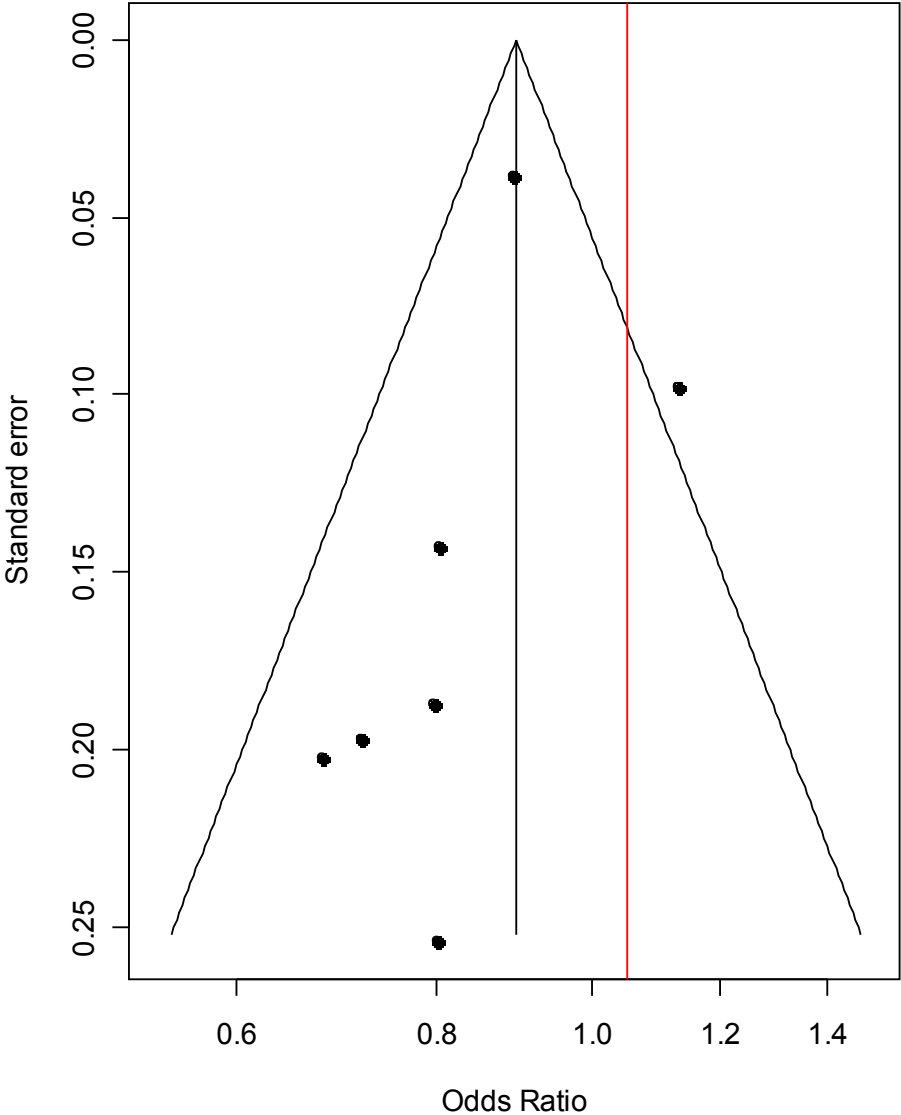
3. For examples, the fail-safe-Ns for the correlation and the odds ratio examples are 1 and 8, respectively.

4. If we can retreat 1 and 8 unpublished studies with null effect, the pooled effect sizes would become non-significant in these two examples.

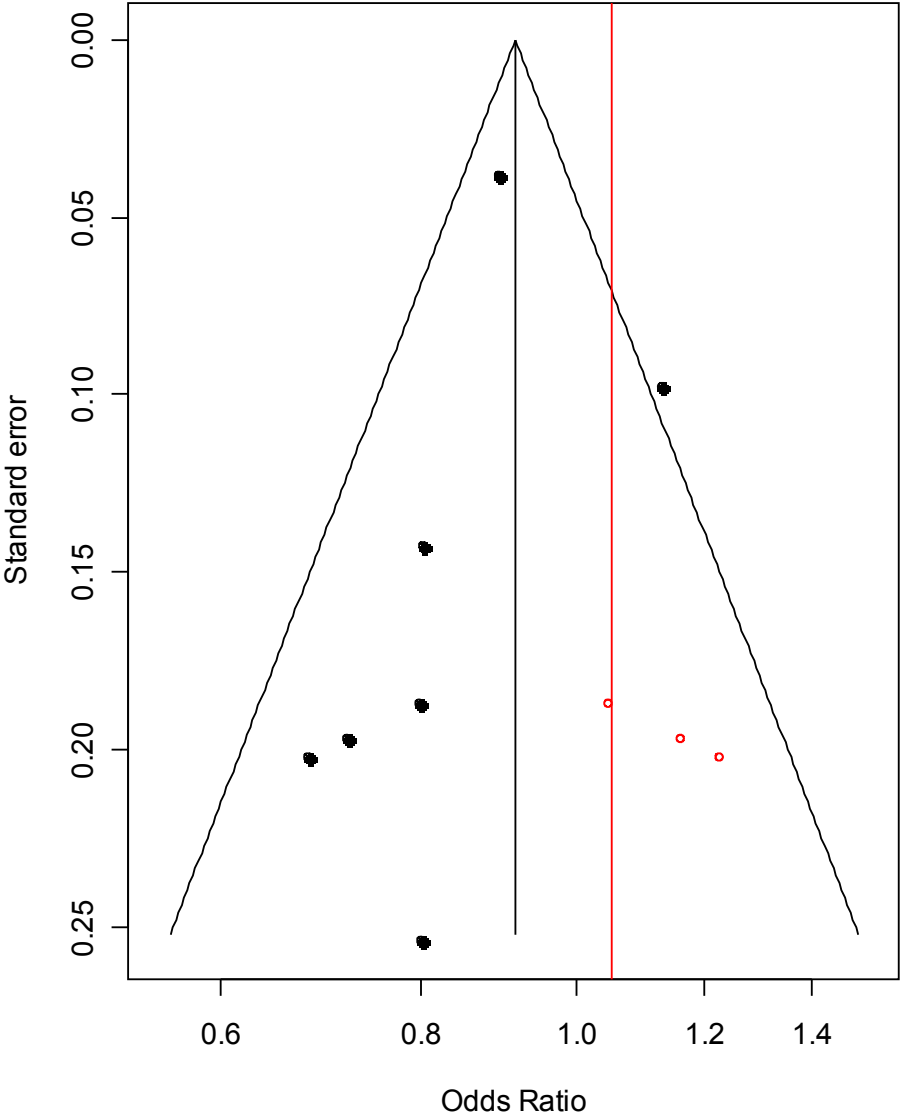
6.3 Funnel plot and trim-and-fill method

1. If there is no publication bias, the effect sizes are expected to be symmetrically distributed under a funnel plot.
2. In the presence of publication bias, one side of the effect sizes is missing.
3. Trim-and-fill method (Duval & Tweedie, 2000) may be used to address the sensitivity of publication bias.

Original effect sizes



With trimmed and filled effect sizes



4. Results:

1. Original data:

	OR	95%-CI	z	p.value
Fixed effects model	0.8969	[0.8405; 0.9570]	-3.2876	0.001
Random effects model	0.8763	[0.7743; 0.9917]	-2.0918	0.0365

2. Trimmed and filled data:

	OR	95%-CI	z	p.value
Fixed effects model	0.9140	[0.8587; 0.9727]	-2.8290	0.0047
Random effects model	0.9231	[0.8252; 1.0327]	-1.3978	0.1622

6.4 Software for meta-analysis

1. SEM approach:

1. Mplus: Fixed-, random-, and mixed-effects meta-analysis (Cheung, 2008) <http://www.statmodel.com/download/MCheung.pdf>

2. LISREL: Fixed-effects meta-analysis (Cheung, in press)

<http://courses.nus.edu.sg/course/psycwlm/internet/FEMA.pdf>

2. General statistical packages:

1. SPSS, SAS and Stata macros:

<http://mason.gmu.edu/~dwilsonb/ma.html>

2. SAS (PROC MIXED), e.g., Sheu and Suzuki (2001)

3. HLM (Variance known models), e.g., Chapter 8 of Hox (2002)

4. R/S-plus:

1. meta and rmeta libraries

2. [MiMa: http://www.wvbauer.com/downloads.html](http://www.wvbauer.com/downloads.html)

3. Stand-alone packages:

1. Comprehensive meta-analysis: <http://www.meta-analysis.com/>

2. <http://faculty.ucmerced.edu/wshadish/Meta-Analysis%20Links.htm>

References

- Becker, B.J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17*, 341-362.
- Becker, B.J. (1995). Corrections to “Using results from replicated studies to estimate linear models”. *Journal of Educational Statistics, 20*, 100-102.
- Brown, S.P., & Stayman, D.M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research, 19*, 34 –51.
- Cheung, M.W.L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods, 13*, 182-202.
- Cheung, M.W.L. (in press). Fixed-effects meta-analyses as multiple-group structural equation models. *Structural Equation Modeling*.
- Cheung, M.W.L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*, 40-64.
- Cheung, M.W.L., & Chan, W. (2009). A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling. *Structural Equation Modeling, 16*, 28-53.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188.
- Duval, S., & Tweedie, R. (2000). Trim and Fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.
- Fleiss, J.L. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research, 2*, 121–145.
- Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.

- Hox, J.J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, N.J.: Lawrence Erlbaum Publishers.
- Hunter, J.E., & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.)*. Thousand Oaks, Calif.: Sage Publications.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243-1246.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.231-244). New York: Russell Sage Foundation.
- Rosenthal, R., & DiMatteo, M.R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- Schmidt F.L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologists*, 47, 1173-1181.
- Sheu, C.F., & Suzuki, S. (2001). Meta-analysis using linear mixed models. *Behavioral Research Methods, Instruments, & Computers*, 33, 102-107.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30, 261-293.

Thank you!
Questions are welcome!