

Meta-Analysis Workshop

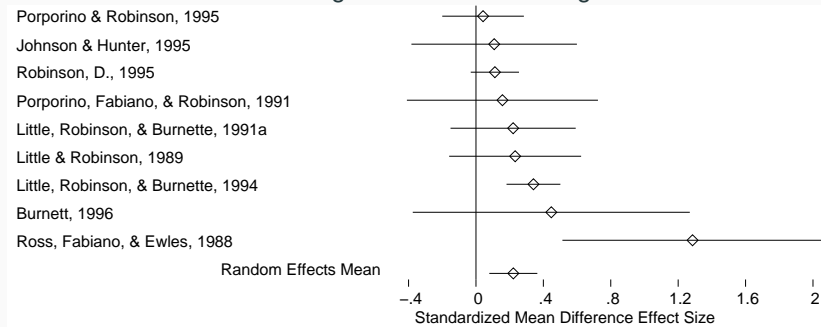
David B. Wilson, PhD

May 24–27, 2021

George Mason University
Department of Criminology, Law & Society
dwilsonb@gmu.edu

The End-Game

Forest-Plot of Standardized Mean Differences and 95% Confidence Intervals for the Effects of Cognitive-Behavioral Programs on Recidivism



Overview

- Historical background
- Logic of Meta-analysis
- Effect sizes
 - Common types
 - Computing standardized mean difference effect sizes
 - Computing odds ratio effect sizes
- Basic meta-analysis methods
- Random-effects versus fixed-effect model
- Moderator analysis
 - Analog to the ANOVA
 - Meta-analytic regression
- Forest plots
- Publication bias
- Cutting edge methods
 - Network meta-analysis
 - Robust standard errors for statistically dependent effect sizes
 - Regression coefficient and fully multivariate models

Historical Background

A Great Debate

- Eysenck 1952: Psychotherapy doesn't work
- Dizzying array of mixed results followed
- Glass (with Smith) averaged results from 375 studies
- Glass coined the term **meta-analysis**

Deep Roots

- Pearson (1904): averaged correlations between inoculation for typhoid fever and mortality
- Fisher (1944): independent studies individually may not be significant, yet the aggregate may be improbable
- W. G. Cochran (1953): developed methods of averaging means across studies (basis of modern methods)
- A. Wicker (1967) averaged correlations between attitudes and behavior
- Concurrent with Smith and Glass (1977) were
 - Hunter and Schmidt (1977) *Validity generalization*
 - Rosenthal and Rubin (1978) *Interpersonal expectancy effects*

Logic of Meta-analysis

- Narrative review methods:
 - Focuses on statistical significance
 - Lacks transparency and replicability
- Weakness of focusing on statistical significance:
 - A significant effect is a strong conclusion
 - A non-significant effect is a weak conclusion
 - How do you balance a collection of significant and non-significant effects?

- Meta-analysis:
 - Focuses on **direction** and **magnitude** of effect
 - Approaches task as a research endeavor
 - Examines pattern of evidence across studies
 - average effect
 - consistency of effects
 - relationship between study features and effects

Research Suitable to Meta-analysis

Forms of Research Findings Suitable to Meta-analysis

- Central tendency
- Prevalence rates
- Pre-post contrasts
- Growth rates
- Group contrasts (e.g., boys versus girls)
- Experimentally created groups (factorial designs, RCTs, etc.)

Forms of Research Findings Suitable to Meta-analysis

- Association between variables
 - Measurement (psychometric) research
 - Individual differences research
 - Correlation between personality constructs
 - Regression models
 - Fully multivariate (SEM, factor analysis)

Concept of Effect Size and Essential Features

Effect Size: The Key to Meta-analysis

- The effect size makes meta-analysis possible
- It is the “dependent variable”
- It standardizes findings across studies such that they can be directly compared

Effect Size: The Key to Meta-analysis

- Any standardized index can be an “effect size” (e.g., standardized mean difference, correlation coefficient, odds ratio) as long as it meets the following criteria:
 - comparable across studies (generally requires standardization)
 - represents the magnitude and direction of the relationship of interest
 - independent of sample size

Meta-analysis as a Systematic Review

Elements of a Systematic Review

- Explicit inclusion/exclusion criteria
- Exhaustive search for all eligible studies
- Coding and data extraction
- Assessing study quality
- Summarizing the evidence (ideally using meta-analysis)

The Replication Continuum

You must be able to argue that the collection of studies you are meta-analyzing examine the same relationship. This may be at a broad level of abstraction, such as the relationship between criminal justice interventions and recidivism or between school-based prevention programs and problem behavior. Alternatively it may be at a narrow level of abstraction and represent pure replications.

The closer to pure replications your collection of studies, the easier it is to argue comparability.

Effect Sizes

The Effect Size: The Heart and Soul of Meta-analysis

- Meta-analysis shifts the focus from statistical significance to the *direction* and *magnitude* of the effect
- Key to this is the effect size
- It is the dependent variable of meta-analysis
- Encodes research findings on a numerical scale
- Different types of effect sizes for different research situations
- Each type may have multiple methods of computation

- Main types of effect sizes
- Logic of the standardized mean difference
- Methods of computing the standardized mean difference
- Logic of the odds ratio and risk ratio
- Methods of computing the odds ratio
- Adjustments, such as for baseline differences
- Issues related to the variance estimate

Most Common Effect Size Indexes

- Standardized mean difference (d or g)
 - Group contrast (e.g., treatment versus control)
 - Inherently continuous outcome construct
- Odds ratio and Risk ratio (OR and RR)
 - Group contrast (e.g., treatment versus control)
 - Inherently dichotomous (binary) outcome construct
- Correlation coefficient (r)
 - Two inherently continuous constructs

Less Common Effect Size Indexes

- Raw (unstandardized) mean difference
- Proportion or prevalence (usually converted to a logit)
- Standardized gain score
- Standardized regression coefficient

Necessary Characteristics of Effect Sizes

- Numerical values produced must be comparable across studies
- Must be able to compute its standard error
- Must not be a direct function of sample size

The Standardized Mean Difference Effect Size

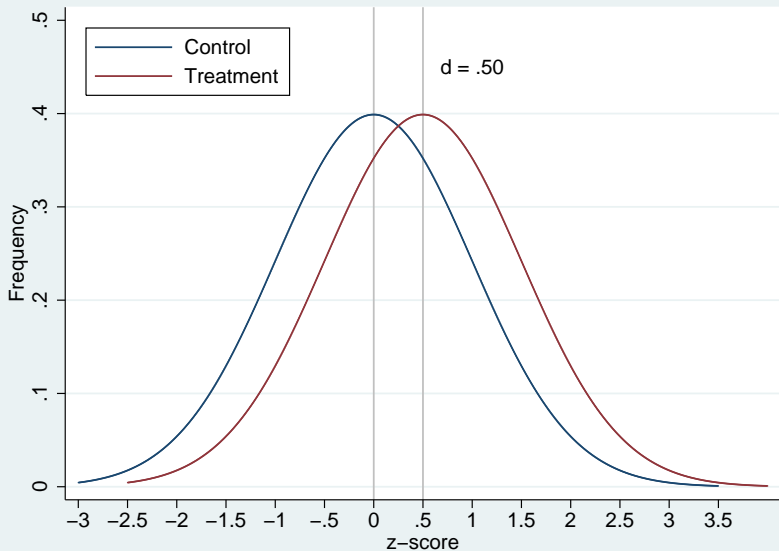
The Standardized Mean Difference

- Compares the means of two groups
- Standardizes this difference

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{\text{pooled}}}$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

Visual Example of d -type Effect Size



Small Sample Size Bias Adjustment

d is slightly upwardly biased when based on small samples. This can be fixed with Hedges' small sample size bias adjustment:

$$g = \left[1 - \frac{3}{4N - 9} \right] d = \left[1 - \frac{3}{4(n_1 + n_2 - 2) - 1} \right] d$$

The result is Hedges' g . The difference between Cohen's d and Hedges' g is this adjustment.

It is common practice to always use this adjustment for standardized mean differences.

Standard Error and Variance of g

The standard error (and variance) of g is mostly a function of sample size:

$$se_g = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2)}}$$

$$v_g = se_g^2$$

- Lots of methods of computing d
- Goal is to reproduce what you would get with the means, standard deviations, and sample sizes
- Some methods are straightforward, others not

Computing d from a t -test

Formula for d :

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}$$

Formula for t :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{n_1+n_2}{n_1 n_2}}}$$

Therefore:

$$d = t \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Computing d from a p -value from a t -test

- An exact p -value from a t -test can be converted into a t -value, assuming you have the degrees-of-freedom
- Then proceed as with prior slide

Computing d from dichotomous outcomes

- Some studies may report a dichotomous outcome
- d (and g) can be estimated from these data
- Several methods
 1. Logit
 2. Cox Logit
 3. Probit
 4. Arcsine (no longer recommended)
- 1–3 have a similar logic and produce similar results unless the outcome has a low or high base rate (i.e., near 0 or 1).

Logit Method of Computing d

- Logistic distribution similar to the normal distribution
- Logged odds ratios conform to the logistic distribution
- Standard deviation of the logistic distribution is

$$s_{logistic} = \frac{\pi}{\sqrt{3}} = 1.814$$

- Thus, we can rescale a logged OR into d

$$d = \frac{\ln(OR)}{1.814}$$

- Monte Carlo simulations suggest that this performs fairly well
- However, there are some advantages to the Cox logit method (divide by 1.65)

$$d = \frac{\ln(OR)}{1.65}$$

- The probit method uses the standard normal distribution rather than the logistic distribution
- Converts each proportion (i.e., successes in each group) into the corresponding z-score with p area under the left portion of the curve.
- Effect size is the difference between these to probits (z-values)

$$d = \text{probit}(p_1) - \text{probit}(p_2)$$

$$d = z_1 - z_2$$

More Complicated Estimates of d

- Can think of d as having two parts
 - Numerator: mean difference
 - Denominator: pooled standard deviation
- Trick is to get reasonable estimates of each
- You often have one but not the other

Estimates of numerator of d

- Covariate adjusted means
- Difference in gain score means
- Unstandardized regression coefficient for treatment dummy
- Standardized regression coefficient for treatment dummy converted to an unstandardized estimate

Estimates of denominator of d

- Gain score standard deviation

$$s_{\text{pooled}} = \frac{s_{\text{gain}}}{\sqrt{2(1-r)}}$$

$$r = \frac{(s_1^2 t^2 + s_2^2 t^2) - (\bar{X}_1 - \bar{X}_2) n}{2s_1 s_2 t^2} \quad (1)$$

- Overall standard deviation for the outcome
- Standard error

$$s = se\sqrt{n-1}$$

- One-way ANOVA

$$s_{\text{pooled}} = \sqrt{MS_w}$$

$$s_{\text{pooled}} = \sqrt{\frac{MS_B}{F}}$$

<https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>

Odds Ratio and Risk Ratio Effect Sizes

Logic of the Odds Ratio and Risk Ratio

- Odds ratio is the ratio of odds
- An odds is the probability of failure over the probability of success
- Risk ratio is the ratio of risks (probabilities)
- A risk is simply the probability of failure
- An odds ratio or risk ratio of 1 indicates no difference between the groups
- Odds ratios are more difficult to interpret than risk ratios
- Risk ratios are sensitive to base event rate; odds ratios are not
- Odds ratios are symmetrical with respect to the outcome (this is why they can be used in case-control studies)

Logic of the Odds Ratio and Risk Ratio

Odds ratios and risk ratios contrast two groups on a dichotomous outcome

	Outcome	
	Success	Failure
Treatment	a	b
Control	c	d

where a , b , c , and d , are cell frequencies

Computing the Odds Ratio and Risk Ratio

Odds ratio is computed as:

$$OR = \frac{ad}{bc} = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$

Risk ratio is computed as

$$RR = \frac{a / (a + b)}{c / (c + d)} = \frac{p_1}{p_2}$$

Computing the Standard Error and Variance of the Odds ratio and Risk Ratio

Standard error of the logged odds ratio:

$$se_{\ln(OR)} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Standard error of the logged risk ratio:

$$se_{\ln(RR)} = \sqrt{\frac{1-p_1}{n_1 p_1} + \frac{1-p_2}{n_2 p_2}}$$

Variance:

$$v = se^2$$

Computing the Odds Ratio from Other Data

- Only so many ways you can report binary data
- Makes computing odds ratios and risk ratios relatively easy
- Generally have
 - Full two-by-two table
 - Percent of events (successes/failures) in each group
 - Proportion of events (successes/failures) in each group
 - Actual odds ratio or risk ratio
 - Convert d -type effect size
 - Chi-square

Computing the Odds ratio from a χ^2

Odds ratio can be computed from a χ^2 if you have the overall event rate for the sample (total number of successes or failures) and the sample size in each group

The frequency for cell a of the two by two can be computed as follows

$$a = N \left[p_x p_y \sqrt{\frac{\chi^2 p_x p_y (1 - p_x)(1 - p_y)}{N}} \right]$$

where p_x is the proportion of the sample in one of the two conditions, p_y is the proportion of the overall sample with a positive (or negative) outcome, and N is the total sample size.

Prevalence Rate as Effect Sizes

Prevalence Rate as Effect Size

- Prevalence can be expressed as a proportion
- Logit of a proportion makes for a good effect size

$$\text{logit} = \ln \left[\frac{p}{1-p} \right]$$

- The standard error of the logit is

$$SE_{\text{logit}} = \sqrt{\frac{1}{np} + \frac{1}{n(1-p)}}$$

- Final results can be back-converted into a proportion

$$p = \frac{e^{\text{logit}}}{e^{\text{logit}} + 1}$$

Issues Related to Variance Estimation

Issues Related to Variance Estimation

- Method of computing effect size may affect variance estimate
- For example
 - d computed from dichotomous data
 - numerator of d is adjusted for baseline or other covariates
- Study data may involve clusters (there are corrections formulas for this situation)
- Possible solution: rescale standard error from correct model provided in study (if available)

$$se_d = \frac{d}{z}$$

- where z is either a z or t test associate with the treatment effect from a complex statistical model.

Coding Protocol

Development of the Coding Protocol

- Like a survey but for studies
- A good coding protocol has transparency and replicability
- Goals:
 - Descriptive: characteristics of the studies, particularly ways in which the studies differ from each other
 - Extraction of findings: computation of effect sizes and related information

- Meta-analytic data is almost always hierarchical
 - Effect sizes nested within
 - Measures nested within
 - Samples or subsamples nested within
 - Studies
- Create separate coding forms and data tables for each level
- Idea is that of a relational database
- Must have identifying “keys” to connect the rows across tables
- Saves coding time, data cleanup time, and data manipulation time

- Paper forms
- Excel (please don't)
- Computer database (great choice!)
- Specialized software (e.g., Comprehensive Meta-analysis, RevMan)

- Reliability of coding is essential
- Best practice is to double-code everything and resolve any differences

Basic Meta-Analysis

- Goal:
 - Describe the distribution, including its mean
 - Establish a confidence interval around the mean
 - Test that the mean differs from zero
 - Test whether studies tell a consistent story (are homogeneous)
 - Explore the relationship between study features and effect size

Determining the Mean Effect Size

- Problem: some effect sizes are more accurate than others
- What we need is an index of precision
- Standard error is a direct measure of precision
- Hedges and Olkin's solution:
 - Weight by the inverse variance
 - Provides a statistical basis for:
 - Standard error of the mean effect size
 - Confidence intervals
 - Homogeneity testing

Some Preliminary Transformations

- Small sample size bias correction for the standardized mean differences:

$$ES_g = \left(1 - \frac{3}{4N-9}\right) ES_d$$

- Fisher's Z_r transform of correlations (ES_r):

$$ES_{Z_r} = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right)$$

- Log transform of odds ratios ES_{OR} :

$$ES_{\ln(OR)} = \log (ES_{OR})$$

- Log transform of risk ratios ES_{RR} :

$$ES_{\ln(RR)} = \log (ES_{RR})$$

Standard Errors

- Standardized mean difference g :

$$se_g = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{ES_g^2}{2(n_1+n_2)}}$$

- Correlation Z_r :

$$se_{Z_r} = \frac{1}{\sqrt{n-3}}$$

- Logged odds ratio $\ln(OR)$:

$$se_{\ln(OR)} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- Logged risk ratio $\ln(RR)$:

$$se_{\ln(RR)} = \sqrt{\frac{1-p_1}{n_1 p_1} + \frac{1-p_2}{n_2 p_2}}$$

Inverse Variance Weights

- Inverse variance weight w :

$$w = \frac{1}{se^2} = \frac{1}{v}$$

- At this point, we have for each study:
 - An effect size
 - An inverse variance weight
- Problem: statistical models assume independence
- Only include one effect size per study (or independent sample)
- Multiple analyses for different subsets of independent effects
 - Different outcome constructs
 - Different time periods

Inverse Variance Weighted Mean

Meta-analytic mean effect size is:

$$\overline{ES} = \frac{\sum w_i ES_i}{\sum w_i}$$

where ES_i is the effect size for each study (i) and w_i is the inverse variance weight

Standard error of the mean effect size is:

$$se_{\overline{ES}} = \sqrt{\frac{1}{\sum w_i}}$$

Some Basic Inferential Statistics

Confidence intervals can be constructed in the usual manner:

$$\overline{ES}_{lower} = \overline{ES} - se_{\overline{ES}}1.96$$

$$\overline{ES}_{upper} = \overline{ES} + se_{\overline{ES}}1.96$$

And a z-test can be performed as:

$$z = \frac{\overline{ES}}{se_{\overline{ES}}}$$

Example: Cognitive-Behavioral Programs for Adult Offenders

Author	Sample Size	<i>g</i> Effect Size	<i>w</i>
Burnett, 1996	60	0.45	5.684
Johnson & Hunter, 1995	98	0.11	15.934
Little & Robinson, 1989	180	0.23	25.049
Little et al 1991	152	0.22	27.852
Little et al 1994	1,381	0.34	150.485
Porporino & Robinson, 1995	757	0.04	65.529
Porporino et al 1991	63	0.16	11.953
Robinson, D., 1995	2,125	0.11	187.177
Ross et al 1988	45	1.28	6.441

Note: These studies are a subset of studies included in Wilson et al. (2005) and represent two specific treatment programs (Moral Reconciliation and Reasoning and Rehabilitation) and studies that were randomized or used high quality quasi-experimental designs.

Computations by Hand

	g	w	$g \times w$
	0.45	5.684	2.55780
	0.11	15.934	1.75274
	0.23	25.049	5.76127
	0.22	27.852	6.12744
	0.34	150.485	51.16490
	0.04	65.529	2.62116
	0.16	11.953	1.91248
	0.11	187.177	20.58947
	1.28	6.441	8.24448
Sum	2.94	496.104	100.7317

$$\bar{g} = \frac{\sum gw}{\sum w} = \frac{100.7317}{496.104} = 0.203$$

$$se = \sqrt{\frac{1}{\sum w}} = \sqrt{\frac{1}{496.104}} = 0.045$$

$$z = \frac{\bar{g}}{se} = \frac{0.203}{0.045} = 4.511$$

$$\bar{g}_{lower95} = 0.203 - 1.96(0.045) = 0.115$$

$$\bar{g}_{upper95} = 0.203 + 1.96(0.045) = 0.291$$

Example: Cognitive-Behavioral Programs for Adult Offenders

Stata output from “masum.ado”

```
. masum g, w(invw) model(FE)
```

No. of obs =	9	Homogeneity Analysis	
Minimum obs =	0.040	Q =	14.1885
Maximum obs =	1.280	df =	8
Minimum weight =	5.6840	p =	0.07698
Maximum weight =	187.1770	I ² =	43.62%

	Mean	-95%CI	+95%CI	se	z	p
Model FE	0.2030	0.1151	0.2910	0.0449	4.5225	0.0000

Fixed (Common) Effect model

Version 2021.04.18 of masum written by David B. Wilson

Homogeneity Testing

- Homogeneity analysis tests whether the assumption that all of the effect sizes are estimating the same population mean is a reasonable assumption.
- If homogeneity is rejected, the distribution of effect sizes is assumed to be heterogeneous.
 - Single mean ES not a good descriptor of the distribution
 - There are real between study differences, that is, studies estimate different population mean effect sizes.
 - Three options:
 - model between study differences
 - fit a random effects model
 - do both

Homogeneity Q Statistic

- Q is simply a weighted sums-of-squares:

$$Q = \sum w_i (ES_i - \overline{ES})^2$$

- There are easier computational formulas:

$$Q = \sum w_i ES_i^2 - \frac{(\sum w_i ES_i)^2}{\sum w_i}$$

- It is distributed as a chi-square with $k - 1$ degrees of freedom, where k is the number of effect sizes

Higgins' I^2 Statistic

- Q is statistically under powered when the number of studies is small, unless the sample sizes of those studies is very large
- Jonathan Higgins proposed an alternative
- It is based on Q but doesn't focus on statistical significance

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100$$

- This reflects the percentage of the total variance in effects due to heterogeneity rather than sampling error

Rough guide to interpreting I^2 :

I^2	Excess variability
0% to 40%	might not be important
30% to 60%	may represent moderate heterogeneity
50% to 90%	may represent substantial heterogeneity
75% to 100%	may represent considerable heterogeneity

Computations by Hand

g	w	$g \times w$	$g^2 \times w$
0.45	5.684	2.55780	1.1510
0.11	15.934	1.75274	0.1928
0.23	25.049	5.76127	1.3251
0.22	27.852	6.12744	1.3480
0.34	150.485	51.16490	17.3960
0.04	65.529	2.62116	0.1048
0.16	11.953	1.91248	0.3060
0.11	187.177	20.58947	2.2648
1.28	6.441	8.24448	10.5529
Σ	2.94	496.104	100.7317
			34.6416

$$Q = \sum w_i ES_i^2 - \frac{(\sum w_i ES_i)^2}{\sum w_i}$$

$$Q = 34.6416 - \frac{(100.7317)^2}{496.104}$$

$$Q = 34.6416 - 20.4531$$

$$Q = 14.19$$

$$df = 9 - 1 = 8, \quad p < .05$$

$$f^2 = 100 \times \frac{Q - df}{Q}$$

$$f^2 = 100 \times \frac{14.19 - 8}{14.19}$$

$$f^2 = 43.6\%$$

Example: Cognitive-Behavioral Programs for Adult Offenders

Stata output from “masum.ado”

```
. masum g, w(invw) model(FE)
```

No. of obs =	9	Homogeneity Analysis	
Minimum obs =	0.040	Q =	14.1885
Maximum obs =	1.280	df =	8
Minimum weight =	5.6840	p =	0.07698
Maximum weight =	187.1770	I ² =	43.62%

	Mean	-95%CI	+95%CI	se	z	p
Model FE	0.2030	0.1151	0.2910	0.0449	4.5225	0.0000

Fixed (Common) Effect model

Version 2021.04.18 of masum written by David B. Wilson

Random Effects versus Fixed Effect Models

Random versus Fixed Effect Models

- Fixed effect model assume:
 - there is one true population effect that all studies are estimating
 - all of the variability between effect sizes is due to sampling error
- Random effects model assume:
 - there are multiple (i.e., a distribution) of population effects that the studies are estimating
 - variability between effect sizes is due to sampling error + variability in the population of effects

Fixed versus Random: Which to Use?

- A random-effects model becomes a fixed-effect model when distributions is homogeneous
- Assumptions of fixed effects model rarely plausible
 - Consequence: standard error that is too small; confidence intervals that are too narrow
- Bottom-line: Use the random-effects model unless you can make a strong *a priori* case for a fixed effect model

Computing a Random Effects Model

- Fixed effects model: weights are a function of sampling error
- Random effects model: weights are a function of sampling error + study level variability
- We call this study level variability τ^2
- Thus, we need a new set of weights

$$w = \frac{1}{se^2 + \tau^2}$$

Computing a Random Effects Model

- First, compute τ^2 (random effects variance component):

$$\tau^2 = \frac{Q - df_Q}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

- Second, re-compute the inverse variance weights:

$$w_i = \frac{1}{se_i^2 + \tau^2}$$

- Third, re-compute meta-analytic results using new weight

Random Effects Computations by Hand

$$\tau^2 = \frac{Q - df_Q}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

$$\tau^2 = \frac{14.19 - 8}{496.104 - \frac{63848.76}{496.104}}$$

$$\tau^2 = 0.0168$$

Random Effects Computations by Hand

Step 1. Compute v if you don't already have such a column.

g	w	v
0.45	5.684	.175932
0.11	15.934	.062759
0.23	25.049	.039922
0.22	27.852	.035904
0.34	150.485	.006645
0.04	65.529	.015260
0.16	11.953	.083661
0.11	187.177	.005343
1.28	6.441	.155255

Random Effects Computations by Hand

Step 2. Add τ^2 to each v .

g	w	v	τ^2	$v + \tau^2$
0.45	5.684	.175932	.0168	.192732
0.11	15.934	.062759	.0168	.079559
0.23	25.049	.039922	.0168	.056722
0.22	27.852	.035904	.0168	.052704
0.34	150.485	.006645	.0168	.023445
0.04	65.529	.015260	.0168	.032060
0.16	11.953	.083661	.0168	.100461
0.11	187.177	.005343	.0168	.022143
1.28	6.441	.155255	.0168	.172055

Random Effects Computations by Hand

Step 3. Compute new weight ($1/(v + \tau u^2)$). In this example, these are DerSimonian and Laird weights (hence the “DL”)

g	w	v	τu^2	$v + \tau u^2$	w_{DL}
0.45	5.684	.175932	.0168	.192732	5.18854
0.11	15.934	.062759	.0168	.079559	12.56931
0.23	25.049	.039922	.0168	.056722	17.62992
0.22	27.852	.035904	.0168	.052704	18.97387
0.34	150.485	.006645	.0168	.023445	42.65269
0.04	65.529	.015260	.0168	.032060	31.19111
0.16	11.953	.083661	.0168	.100461	9.95411
0.11	187.177	.005343	.0168	.022143	45.16195
1.28	6.441	.155255	.0168	.172055	5.81208

Random Effects Computations by Hand

Step 4 and 5. Multiply the new weights by the effect size and sum columns.

	g	w	v	τ^2	$v + \tau^2$	w_{DL}	$g \times w_{DL}$
	0.45	5.684	.175932	.0168	.192732	5.18854	2.3348
	0.11	15.934	.062759	.0168	.079559	12.56931	1.3826
	0.23	25.049	.039922	.0168	.056722	17.62992	4.0549
	0.22	27.852	.035904	.0168	.052704	18.97387	4.1743
	0.34	150.485	.006645	.0168	.023445	42.65269	14.5019
	0.04	65.529	.015260	.0168	.032060	31.19111	1.2476
	0.16	11.953	.083661	.0168	.100461	9.95411	1.5927
	0.11	187.177	.005343	.0168	.022143	45.16195	4.9678
	1.28	6.441	.155255	.0168	.172055	5.81208	7.4395
Sum	2.94					189.1336	41.6961

Step 6. Re-compute mean effect size and standard error as before.

Comment:

Because the variance is larger, the weight is smaller. Also, across studies, the weights will be more similar. With a large enough τ^2 the weights will become virtually equal.

Example: Cognitive-Behavioral Programs for Adult Offenders

Stata output from “masum.ado”

```
. masum g, w(invw) model(DL)
```

```
No. of obs =          9
Minimum obs =         0.040
Maximum obs =         1.280
Minimum weight =       5.1874
Maximum weight =      45.0727

Homogeneity Analysis
Q =      14.1885
df =           8
p =      0.07698
I^2 =      43.62%
tau^2 =     0.01684
se (tau^2) =  0.02158
tau =     0.12978
```

```
-----+-----
          |      Mean   -95%CI   +95%CI      se      z      p
-----+-----
Model DL  |  0.2205   0.0779   0.3631   0.0728   3.0303   0.0024
-----+-----
```

```
Random effects: Dersimonian-Laird
```

```
Version 2021.04.18 of masum written by David B. Wilson
```

Numerous Estimators for τ^2

There are numerous estimators for the random effects variance component.

- DerSimonian-Laird method-of-moments most widely used
- Doesn't always perform well when the number of studies is small and heterogeneity is high
- Restricted maximum likelihood (REML)
 - Widely recommended alternative
 - Requires iteration; not suitable for hand calculation
 - Is now widely available in most software programs
- Roughly 7 other estimators with various pros and cons

Comparison of Random Effects and Fixed Effect Models

- Biggest difference you will notice is the significance levels and confidence intervals
- As heterogeneity increases
 - p -value for the mean effect size will increase
 - confidence interval will get wider
- A mean that was significant under a fixed effect model might not be significant under random effects model
- Random effects model is more conservative
- Mean effect size may differ between models if sample size and effect size are related

Summary of Basic Meta-analysis

- Apply any transformations on effect sizes
 - Small sample size bias correction (d)
 - Fisher's Z_r transformation (r)
 - Natural log (OR and RR).
- Compute inverse variance weight
- Compute mean, standard error, and confidence intervals
- Compute homogeneity statistics (Q and I^2)
- Assume random effects unless you can make a strong *a priori* case for a fixed effect model
- Next up, moderator analysis

Moderator Analysis

Moderator Analysis

- Modeling between study variability
 - Categorical models (analogous to a one-way ANOVA)
 - Regression models
- Fixed and random effects versions of each (latter often called “mixed” models)
- Number of studies needed

Analog to the ANOVA

Analog to the ANOVA

- Useful for a single categorical independent variable
- Produce a separate mean effect size for each category
- Recall that Q is a sum-of-squares
- The total sum-of-squares (Q) can be partitioned
 - Variability between groups (Q_{Between})
 - Residual variability within groups (Q_{Within})

Analog to the ANOVA

- Q_{Between} analogous to an F -test between means
- Q_{Within} assesses whether residual distribution homogeneous

Analog to the ANOVA Equations

Recall that the overall (or total) Q is computed as

$$Q_{Total} = \sum w_i (ES_i - \overline{ES})^2$$

$$Q_{Total} = \sum w_i ES_i^2 - \frac{(\sum w_i ES_i)^2}{\sum w_i}$$

Analog to the ANOVA Equations

To compute the Q_{Within} , we need to compute the Q within each group and then sum across groups

$$Q_{\text{Within}} = \sum w_{ij} (ES_{ij} - \overline{ES}_j)^2$$

Analog to the ANOVA Equations

To compute the Q_{Between} , simply subtract Q_{Within} from Q_{Total}

$$Q_{\text{Between}} = Q_{\text{Total}} - Q_{\text{Within}}$$

Recall that Q is distributed as a χ^2 . The degrees-of-freedom are computed just as they are in a one-way ANOVA:

$$df_{\text{Within}} = k - j$$

$$df_{\text{Between}} = j - 1$$

$$df_{\text{Total}} = k - 1$$

where k is the number of effect sizes and j is the number of categories or groups.

Analog to the ANOVA Interpretation

- Q_{Between} reflects the differences in the means across categories (analogous to the F from a one-way ANOVA)
- Q_{Within} indicates whether the distributions within categories are homogeneous (overall)
 - For the random effects model, Q_{Within} should be based on a fixed effect model or a significance test of τ^2
- Test is often statistically under-powered (particularly in the random effects case)
- The random effects version estimates τ^2 based on the within groups heterogeneity

Using my Stata “ado” file `maanova`, you can run an analog to the ANOVA.

For a simple fixed effect model

```
maanova g iv, w(invw) model(FE)
```

where `g` is the effect size, `iv` is the categorical or grouping variable and `invw` is the inverse variance weight.

Analog to the ANOVA in Stata

For a mixed (random) effects model, add the `model` option. You can specify either a method-of-moments estimator of τ^2 or the maximum likelihood estimator as `model(mm)` and `model(ml)` respectively.

```
maanova g iv, w(invw) model(FE)
maanova g iv, w(invw) model(DL)
maanova g iv, w(invw) model(REML)
maanova g iv, w(invw) model(REML) print(EXP)
```

If you are analyzing logged odds ratios or risk ratios, add the `print(exp)` option to get the mean effects reported as odds ratios or risk ratios, as shown in the last example above.

Random Effects Analog to the ANOVA

Experimental versus Quasi-experimental Studies in the Domestic Violence (Page 1 of 2)

```
. maanova d random, w(w) model(DL)
```

```
No. of obs =          14
Minimum effect size =  -0.4671
Maximum effect size =   1.5286
Minimum weight =       2.8878
Maximum weight =       5.6175
Common tau^2 =         0.1626
SE tau^2 =            0.1000
Common tau =          0.4033
I^2 =                 76.20%
```

```
-----+-----
Category |      Mean  -95%CI  +95%CI   se    z    p    k
-----+-----
Total    |  0.2784   0.0251   0.5316  0.1292  2.1543  0.0312  14
-----+-----
0.0000  |  0.2912  -0.0586   0.6409  0.1785  1.6315  0.1028   7
1.0000  |  0.2643  -0.1029   0.6314  0.1873  1.4107  0.1583   7
-----+-----
```

Random Effects Analog to the ANOVA

Experimental versus Quasi-experimental Studies in the Domestic Violence (Page 2 of 2)

```
-----+-----
Category      | Q-within  p(Q-within)  df
-----+-----
      0.0000  |    42.2240    0.0000    6
      1.0000  |     8.1873    0.2247    6
-----+-----
```

Q-within based on fixed effect weights

Analog-to-the-ANOVA Table

```
-----+-----
Source        |          Q          p(Q)      df
-----+-----
Between (Model) |    0.0108    0.9172      1
Within (Error) |    50.4113    0.0000     12
-----+-----
```

Random effects: Dersimonian-Laird

Q-within based on fixed effect weights

Meta-analytic Regression

Meta-analytic Regression

- Conceptually identical to multiple regression
 - Effect size is the dependent variable
 - Study moderator variables are the independent variables
- Can handle multiple variables simultaneously
- Don't use standard OLS regression procedures (even if weighted)
- Must use specialized software

Using my Stata “ado” file `mareg`, you can run a meta-analytic multiple regression model following the examples below

For a simple fixed effect model

```
mareg g iv1 iv2, w(invw) model(FE)
```

where `g` is the effect size, `iv1` and `iv2` are the independent variables, and `invw` is the inverse variance weight (you can alternatively specify `var(variance)` or `se(standard error)`).

For a mixed (random) effects model, add the `model` option. You can specify one of 8 estimators for τ^2 . The two most common are the restricted maximum likelihood estimator (`model(REML)`) and the DerSimonian and Laird method-of-moments estimator `model(DL)`, respectively.

```
mareg g iv1 iv2, w(invw) model(DL)  
mareg g iv1 iv2, w(invw) model(REML)
```

Meta-analytic Regression

```
. mareg g txvari txvar2 random, var(v) model(DL)
```

```
No. of obs =          38
Minimum effect size =  -0.6400
Maximum effect size =   0.9684
Minimum weight =       8.4864
Maximum weight =      20.9782
tau^2 =                0.0470
se (tau^2) =           0.0193
tau =                  0.2169
I^2 =                  92.51%
```

```
Random effects: Dersimonian-Laird
```

```
-----+-----
Variable |           b           se           z           P(z)  [-----95% CI-----]
-----+-----
txvari   |    0.37730  0.14363   2.62690  0.00862   0.09579   0.65881
txvar2   |    0.19846  0.08198   2.42072  0.01549   0.03777   0.35915
random   |    0.07296  0.08327   0.87619  0.38093  -0.09025   0.23617
_cons    |   -0.67358  0.23890  -2.81949  0.00481  -1.14182  -0.20534
-----+-----
```

```
Overall Model Statistics
```

```
-----+-----
Source   |           Q           p(Q)           df
-----+-----
Model    |    14.7194           0.0021           3
Error    |   453.7534           0.0000           34
-----+-----
```

```
Q-within based on fixed effect weights
```

Publication Selection Bias

Publication Selection Bias

- Statistically significant effects are more likely to be published than nonsignificant effects
- Has many forms
 - outcomes select for inclusion in a report
 - analyses selected for inclusion in a report
 - author's desire to write-up a study
 - reviewers' recommendations for publication
 - editor's decision to accept for publication
- Existence of this bias is well established
- Affects all form of reviewing the literature
- Essential that you search for and included unpublished studies that meet eligibility criteria
- Essential that you explore the risk of publication selection bias in your data

- Statistic approaches to assessing publication bias
 - Funnel plot: Scatterplot of the effect size against the inverse standard error of effect size
 - Trim-and-fill method (Tweedie and Duvall)
 - Fail-safe N (Not recommended!)

Funnel Plot

- As sample size increases, smaller effects become statistically significant
- Thus, smaller effects from smaller studies are more likely to be missing
- Statistical significance is directly a function of the standard error
- Funnel plot is a scatter plot if the inverse of the standard error (large values reflect larger samples) and effect size
- Expectation is that the plot should be symmetric around the mean effect size
- I.e., it should look like an upside-down funnel
- Asymmetry is suggestive of publication selection bias

There is an installable command called `metafunnel` that will produce funnel plots.

If you have not already done so, run the following command:

```
ssc install metafunnel
```

Using the `sexoffender.dta` data file, generate a funnel plot with the following command:

```
metafunnel lgores vlgores, var
```

Does this look asymmetrical?

Trim and Fill Method

- Method for assessing the impact of publication selection bias on results
- Based on the same idea as the funnel plot
- Nonparametric method for adjusting the mean effect size
- Rectifies the funnel plot asymmetry
 - Assumes the effects to the left of the mean are more likely to be missing
 - Initially trims effect sizes iteratively to make funnel plot symmetric
 - Replaces trimmed effect sizes with pairs on the left side of the mean

There is an installable command called `metatrim` that will produce funnel plots.

If you have not already done so, run the following command:

```
ssc install metatrim
```

Using the `sexoffender.dta` data file, run a trim and fill analysis with the following command:

```
metatrim lgores vlgores, var
```

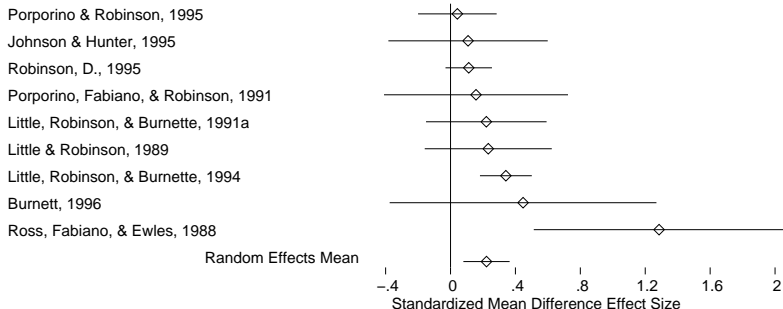
What affect did this procedure have on the estimated overall mean effect size?

Forest Plots

- Visual representation of results
- Row for each study that shows
 - study label
 - sample size; may include other information
 - effect size (dot, square, diamond)
 - confidence interval (horizontal line)
- Row for the overall mean results
 - effect size (dot, square, diamond)
 - confidence interval (horizontal line)

Example Forest-Plot

Forest-Plot of Standardized Mean Differences and 95% Confidence Intervals for the Effects of Cognitive-Behavioral Programs on Recidivism



Creating a Forest-Plot in Stata

The “ado” file `maforest` can create basic forest plots.

The basic syntax is:

```
maforest effect se
```

where *effect* is the effect size, *se* is the standard error of the effect size ($se = \sqrt{v} = 1/w$). If you have not already compute *se*, you must compute these before using the `maforest` command.

Creating a Forest-Plot in Stata

Syntax for `maforest`

`maforest effect se`

Options:

<code>xmin(num)</code>	Minimum value for the effect size axis
<code>xmax(num)</code>	Maximum value for the effect size axis
<code>xtics(numlist)</code>	List of values for the effect size axis
<code>sortby(varlist)</code>	Variable to sort rows
<code>col1(varname)</code>	Variable (such as author name) for column 1
<code>col1lbl("string")</code>	Column label for column 1
<code>col2(varname)</code>	Variable (such as sample size) for column 2
<code>col2lbl("string")</code>	Column label for column 2
<code>xlabel("string")</code>	Label for axis (default is "Effect Size")
<code>logxaxis</code>	Exponential to the axis (effect sizes are logged)
<code>font(large medium small tiny)</code>	Font size
<code>mean(fe mm ml)</code>	Add mean effect

Forest Plot Stata Examples

Use the `cbt.dta` data file.

The variable you will need are:

<code>author</code>	Label with the author names
<code>totaln</code>	Total sample size
<code>g</code>	Hedges' g effect size
<code>se</code>	Standard error of the effect size

Forest Plot Stata Examples

```
maforest se  
maforest g se, col1(author) col2(totaln)  
maforest g se, col1(author) col2(totaln)  
    xtics(-3,-2,-1,0,1,2,3)  
    col1lbl("Author") col2lbl("Sample Size")
```

See exercise handout for more.

Features of a Quality Meta-analysis

What to look for in a quality meta-analysis

- Explicit inclusion and exclusion criteria
- Inclusion and exclusion criteria well justified
- Not restricted to published studies
- Search strategy well explicated
- Search includes multiple sources (e.g., databases, hand-searches, contact with authors, etc.)
- Used a detailed coding protocol
- Assessed coder reliability (e.g., double-coding)

What to look for in a quality meta-analysis

- What to look for in a quality meta-analysis
- Maintained statistical independent in the analysis of effect sizes
- Use proper meta-analytic methods (e.g., inverse variance weighting)
- Tested for heterogeneity in effect sizes
- Used a random effects model or was clearly justified in using a fixed effect model
- Used proper methods of testing moderator effects (e.g., analog-to-the-ANOVA, meta-regression)
- Assessed for publication bias
- If included methodologically “flawed” studies, performed sensitivity analysis on results

Cutting Edge Methods

- Network Meta-analysis
- Meta-analysis of dependent effect sizes
- Regression coefficients and fully multivariate meta-analysis

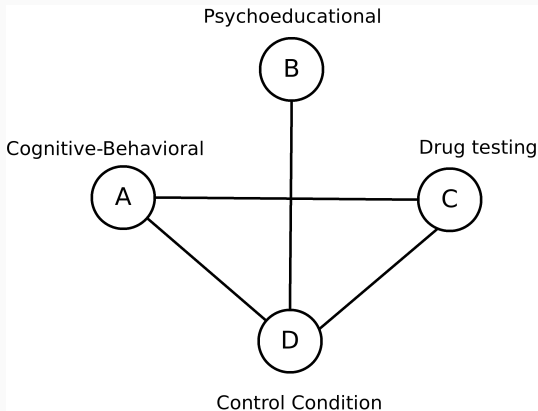
Network Meta-analysis

Network Meta-analysis

- Extension of regular meta-analysis to two or more treatment conditions
- Allows to the mixing of treatment versus treatment with treatment versus control studies
- Makes use of all available comparisons in the network to estimate the relative effectiveness of each treatment
- Well suited to comparing the effectiveness of multiple drugs for treating a common condition
- There are Stata commands for network meta-analysis

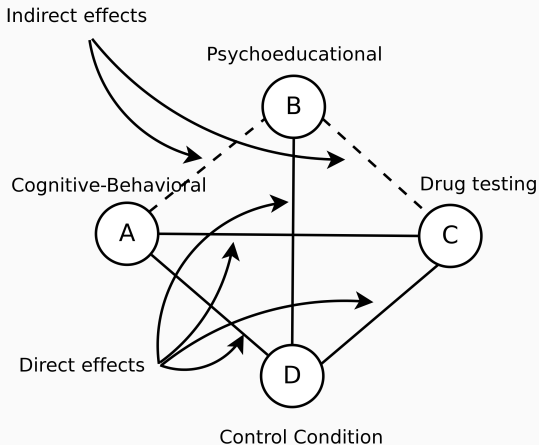
Network Meta-analysis

Example of a simple network of treatments

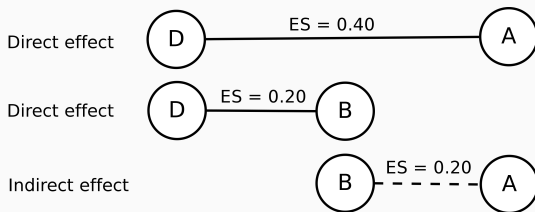


Network Meta-analysis

Example of direct and indirect effects



Example showing how indirect effects are estimated (conceptually)



Dependent Effect Sizes

Dependent Effect Sizes

- Hedges, Tipton, and Johnson (2010) developed robust variance estimation method
- Allows for dependent effect sizes
 - That is, multiple effect sizes per study or sample!
- Does not require knowledge of the covariance structure
- Has been implemented in Stata and R
- `ssc install robumeta`

Dependent Effect Sizes

- Method includes meta-regression models
- Works similar to `ma:reg` but requires a study ID
- Can test for moderator effects both between and within studies
 - Difference in effect sizes between boys and girls

Regression Coefficients and Multivariate Models

Regression Coefficients

- Numerous examples in the literature
- Debate among meta-analysis scholars about best methods
- Challenges
 - What effect size?
 - How comparable must the models be?
 - How should you compute the effect size variance?
- Possible effect sizes
 - Standardized regression coefficient (generally preferred)
 - Semi-partial correlation coefficient (similar to above)
 - Partial correlation coefficient (models must include the same variables)
 - Modified Cohen's d (assumes independent variable is dichotomous)
 - For logistic regression models, unstandardized regression coefficient

If you are interested in fully multivariate models and not just a bivariate relationship from a multivariate model, you can meta-analyze the underlying correlation matrix of interest.

Then use the matrix for your model (regression, factor analysis, structural equation modeling, etc.).

Complication: not all studies will have all correlations of interest.

Questions and Open Discussion
