

Practical Meta-Analysis

David B. Wilson
American Evaluation Association
Orlando, Florida, October 3, 1999

1

The Great Debate

- 1952: Hans J. Eysenck concluded that there were no favorable effects of psychotherapy, starting a raging debate
- 20 years of evaluation research and hundreds of studies failed to resolve the debate
- 1978: To prove Eysenck wrong, Gene V. Glass statistically aggregate the findings of 375 psychotherapy outcome studies
- Glass (and colleague Smith) concluded that psychotherapy did indeed work
- Glass called his method "meta-analysis"

2

The Emergence of Meta-Analysis

- Ideas behind meta-analysis predate Glass' work by several decades
 - R. A. Fisher (1944)
 - "When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance" (p. 99).
 - Source of the idea of cumulating probability values
 - W. G. Cochran (1953)
 - Discusses a method of averaging means across independent studies
 - Laid-out much of the statistical foundation that modern meta-analysis is built upon (e.g., inverse variance weighting and homogeneity testing)

3

The Logic of Meta-Analysis

- Traditional methods of review focus on statistical significance testing
- Significance testing is not well suited to this task
 - highly dependent on sample size
 - null finding does not carry to same "weight" as a significant finding
- Meta-analysis changes the focus to the **direction** and **magnitude** of the effects across studies
 - Isn't this what we are interested in anyway?
 - Direction and magnitude represented by the effect size

4

When Can You Do Meta-Analysis?

- Meta-analysis is applicable to collections of research that
 - are empirical, rather than theoretical
 - produce quantitative results, rather than qualitative findings
 - examine the same constructs and relationships
 - have findings that can be configured in a comparable statistical form (e.g., as effect sizes, correlation coefficients, odds-ratios, etc.)
 - are “comparable” given the question at hand

5

Forms of Research Findings Suitable to Meta-Analysis

- Central Tendency Research
 - prevalence rates
- Pre-Post Contrasts
 - growth rates
- Group Contrasts
 - experimentally created groups
 - comparison of outcomes between treatment and comparison groups
 - naturally occurring groups
 - comparison of spatial abilities between boys and girls
- Association Between Variables
 - measurement research
 - validity generalization
 - individual differences research
 - correlation between personality constructs

6

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

7

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.

8

Which Studies to Include?

- It is critical to have an explicit inclusion and exclusion criteria (see handout)
 - the broader the research domain, the more detailed they tend to become
 - developed iteratively as you interact with the literature
- To include or exclude low quality studies
 - the findings of all studies are potentially in error (methodological quality is a continuum, not a dichotomy)
 - being too restrictive may restrict ability to generalize
 - being too inclusive may weaken the confidence that can be placed in the findings
 - must strike a balance that is appropriate to your research question

9

Searching Far and Wide

- The “we only included published studies because they have been peer-reviewed” argument
- Significant findings are more likely to be published than nonsignificant findings
- Critical to try to identify and retrieve all studies that meet your eligibility criteria
- Potential sources for identification of documents
 - computerized bibliographic databases
 - authors working in the research domain
 - conference programs
 - dissertations
 - review articles
 - hand searching relevant journal
 - government reports, bibliographies, clearinghouses

10

Strengths of Meta-Analysis

- Imposes a discipline on the process of summing up research findings
- Represents findings in a more differentiated and sophisticated manner than conventional reviews
- Capable of finding relationships across studies that are obscured in other approaches
- Protects against over-interpreting differences across studies
- Can handle a large numbers of studies (this would overwhelm traditional approaches to review)

11

Weaknesses of Meta-Analysis

- Requires a good deal of effort
- Mechanical aspects don't lend themselves to capturing more qualitative distinctions between studies
- “Apples and oranges”; comparability of studies is often in the “eye of the beholder”
- Most meta-analyses include “blemished” studies
- Selection bias poses continual threat
 - negative and null finding studies that you were unable to find
 - outcomes for which there were negative or null findings that were not reported
- Analysis of between study differences is fundamentally correlational

12

The Effect Size

- The effect size (ES) makes meta-analysis possible.
- The ES encodes the selected research findings on a numeric scale.
- There are many different types of ES measures, each suited to different research situations.
- Each ES type may also have multiple methods of computation.

1

Examples of Different Types of Effect Sizes: The Major Leagues

- **Standardized Mean Difference**
 - group contrast research
 - treatment groups
 - naturally occurring groups
 - inherently continuous construct
- **Odds-Ratio**
 - group contrast research
 - treatment groups
 - naturally occurring groups
 - inherently dichotomous construct
- **Correlation Coefficient**
 - association between variables research

2

Examples of Different Types of Effect Sizes: Two from the Minor Leagues

- **Proportion**
 - central tendency research
 - HIV/AIDS prevalence rates
 - Proportion of homeless persons found to be alcohol abusers
- **Standardized Gain Score**
 - gain or change between two measurement points on the same variable
 - reading speed before and after a reading improvement class

3

What Makes Something an Effect Size for Meta-Analytic Purposes

- The type of ES must be comparable across the collection of studies of interest.
- This is generally accomplished through standardization.
- Must be able to calculate a standard error for that type of ES
 - the standard error is needed to calculate the ES weights, called inverse variance weights (more on this latter)
 - all meta-analytic analyses are weighted

4

The Standardized Mean Difference

$$\overline{ES} = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_{pooled}} \quad s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

- Represents a standardized group contrast on an *inherently continuous* measure.
- Uses the pooled standard deviation (some situations use control group standard deviation).
- Commonly called “d” or occasionally “g”.

5

The Correlation Coefficient

$$\overline{ES} = r$$

- Represents the strength of association between two *inherently continuous* measures.
- Generally reported directly as “r” (the Pearson product moment coefficient).

6

The Odds-Ratio

- The Odds-Ratio is based on a 2 by 2 contingency table, such as the one below.

Frequencies		
	Success	Failure
Treatment Group	a	b
Control Group	c	d

$$\overline{ES} = \frac{ad}{bc}$$

- The Odds-Ratio is the odds of success in the treatment group relative to the odds of success in the control group.

7

Methods of Calculating the Standardized Mean Difference

- The standardized mean difference probably has more methods of calculation than any other effect size type.

8

The different formulas represent degrees of approximation to the ES value that would be obtained based on the means and standard deviations

- Great
 - direct calculation based on means and standard deviations
 - algebraically equivalent formulas (t-test)
 - exact probability value for a t-test
 - approximations based on continuous data (correlation coefficient)
- Good
 - estimates of the mean difference (adjusted means, regression B weight, gain score means)
 - estimates of the pooled standard deviation (gain score standard deviation, one-way ANOVA with 3 or more groups, ANCOVA)
- Poor
 - approximations based on dichotomous data

9

Methods of Calculating the Standardized Mean Difference

Direction Calculation Method

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

10

Methods of Calculating the Standardized Mean Difference

Algebraically Equivalent Formulas:

$$ES = t \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad \text{independent t-test}$$

$$ES = \sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}} \quad \text{two-group one-way ANOVA}$$

exact *p-values* from a *t*-test or *F*-ratio can be converted into *t*-value and the above formula applied

11

Methods of Calculating the Standardized Mean Difference

A study may report a grouped frequency distribution from which you can calculate means and standard deviations and apply to direct calculation method.

12

Methods of Calculating the Standardized Mean Difference

Close Approximation Based on Continuous Data -- Point-Biserial Correlation. For example, the correlation between treatment/no treatment and outcome measured on a continuous scale.

$$ES = \frac{2r}{\sqrt{1-r^2}}$$

Methods of Calculating the Standardized Mean Difference

Estimates of the Numerator of ES -- The Mean Difference

- difference between gain scores
- difference between covariance adjusted means
- unstandardized regression coefficient for group membership

Methods of Calculating the Standardized Mean Difference

Estimates of the Denominator of ES -- Pooled Standard Deviation

$$s_{pooled} = se\sqrt{n-1} \quad \text{standard error of the mean}$$

Methods of Calculating the Standardized Mean Difference

Estimates of the Denominator of ES -- Pooled Standard Deviation

$$s_{pooled} = \sqrt{\frac{MS_{between}}{F}} \quad \text{one-way ANOVA } >2 \text{ groups}$$

$$MS_{between} = \frac{\sum \bar{X}_j^2 n_j - \frac{(\sum \bar{X}_j n_j)^2}{\sum n_j}}{k-1}$$

Methods of Calculating the Standardized Mean Difference

Estimates of the Denominator of ES -- Pooled Standard Deviation

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

standard deviation of gain scores, where r is the correlation between pretest and posttest scores

Methods of Calculating the Standardized Mean Difference

Estimates of the Denominator of ES -- Pooled Standard Deviation

$$s_{pooled} = \sqrt{\frac{MS_{error} \cdot df_{error} - 1}{1 - r^2} \cdot \frac{df_{error} - 2}{df_{error} - 2}}$$

ANCOVA, where r is the correlation between the covariate and the DV

Methods of Calculating the Standardized Mean Difference

Estimates of the Denominator of ES -- Pooled Standard Deviation

$$s_{pooled} = \sqrt{\frac{SS_B + SS_{AB} + SS_W}{df_B + df_{AB} + df_W}}$$

A two-way factorial ANOVA where B is the irrelevant factor and AB is the interaction between the irrelevant factor and group membership (factor A).

Methods of Calculating the Standardized Mean Difference

Approximations Based on Dichotomous Data

$$ES = probit(p_{group_1}) - probit(p_{group_2})$$

the difference between the probits transformation of the proportion successful in each group

converts proportion into a z-value

Methods of Calculating the Standardized Mean Difference

Approximations Based on Dichotomous Data

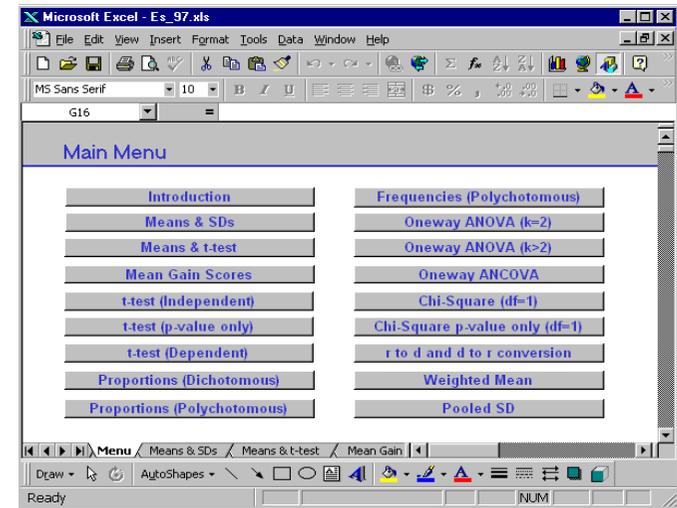
$$ES = 2\sqrt{\frac{\chi^2}{N - \chi^2}}$$

chi-square must be based on a 2 by 2 contingency table (i.e., have only 1 df)

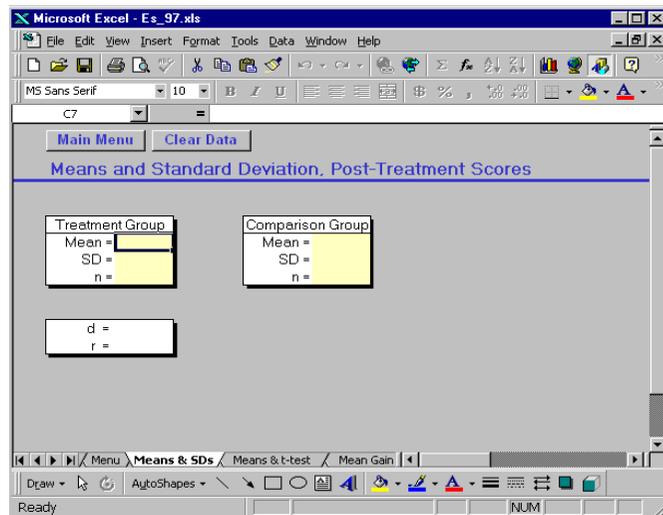
$$ES = \frac{2r}{\sqrt{1 - r^2}}$$

phi coefficient

21



22



23

Formulas for the Correlation Coefficient

- Results typically reported directly as a correlation.
- Any data for which you can calculate a standardized mean difference effect size, you can also calculate a correlation type effect size.
- See Appendix B for formulas.

24

Formulas for the Odds Ratio

- Results typically reported in one of three forms:
 - frequency of successes in each group
 - proportion of successes in each group
 - 2 by 2 contingency table
- Appendix B provides formulas for each situation.

25

Data to Code Along with the ES

- The Effect Size
 - may want to code the data from which the ES is calculated
 - confidence in ES calculation
 - method of calculation
 - any additional data needed for calculation of the inverse variance weight
- Sample Size
- ES specific attrition
- Construct measured
- Point in time when variable measured
- Reliability of measure
- Type of statistical test used

26

Overview of Meta-Analytic Data Analysis

- Transformations, Adjustments and Outliers
- The Inverse Variance Weight
- The Mean Effect Size and Associated Statistics
- Homogeneity Analysis
- Fixed Effects Analysis of Heterogeneous Distributions
 - Fixed Effects Analog to the one-way ANOVA
 - Fixed Effects Regression Analysis
- Random Effects Analysis of Heterogeneous Distributions
 - Mean Random Effects ES and Associated Statistics
 - Random Effects Analog to the one-way ANOVA
 - Random Effects Regression Analysis

1

Transformations

- Some effect size types are not analyzed in their “raw” form.
- Standardized Mean Difference Effect Size
 - Upward bias when sample sizes are small
 - Removed with the small sample size bias correction

$$ES'_{sm} = ES_{sm} \left[1 - \frac{3}{4N - 9} \right]$$

2

Transformations (continued)

- Correlation has a problematic standard error formula.
- Recall that the standard error is needed for the inverse variance weight.
- Solution: Fisher's Zr transformation.
- Finally results can be converted back into “r” with the inverse Zr transformation (see Chapter 3).

3

Transformations (continued)

- Analyses performed on the Fisher's Zr transformed correlations.

$$ES_{Zr} = .5 \ln \left[\frac{1+r}{1-r} \right]$$

- Finally results can be converted back into “r” with the inverse Zr transformation.

$$r = \frac{e^{2ES_{Zr}} - 1}{e^{2ES_{Zr}} + 1}$$

4

Transformations (continued)

- Odds-Ratio is asymmetric and has a complex standard error formula.
 - Negative relationships indicated by values between 0 and 1.
 - Positive relationships indicated by values between 1 and infinity.
- Solution: Natural log of the Odds-Ratio.
 - Negative relationship < 0 .
 - No relationship = 0.
 - Positive relationship > 0 .
- Finally results can be converted back into Odds-Ratios by the inverse natural log function.

5

Transformations (continued)

- Analyses performed on the natural log of the Odds- Ratio:

$$ES_{LOR} = \ln[OR]$$

- Finally results converted back via inverse natural log function:

$$OR = e^{ES_{LOR}}$$

6

Adjustments

- Hunter and Schmidt Artifact Adjustments
 - measurement unreliability (need reliability coefficient)
 - range restriction (need unrestricted standard deviation)
 - artificial dichotomization (correlation effect sizes only)
 - assumes a normal underlying distribution
- Outliers
 - extreme effect sizes may have disproportionate influence on analysis
 - either remove them from the analysis or adjust them to a less extreme value
 - indicate what you have done in any written report

7

Overview of Transformations, Adjustments, and Outliers

- Standard transformations
 - sample size bias correction for the standardized mean difference effect size
 - Fisher's Z to r transformation for correlation coefficients
 - Natural log transformation for odds-ratios
- Hunter and Schmidt Adjustments
 - perform if interested in what would have occurred under "ideal" research conditions
- Outliers
 - any extreme effect sizes have been appropriately handled

8

Independent Set of Effect Sizes

- Must be dealing with an independent set of effect sizes before proceeding with the analysis.
 - One ES per study OR
 - One ES per subsample within a study

9

The Inverse Variance Weight

- Studies generally vary in size.
- An ES based on 100 subjects is assumed to be a more “precise” estimate of the population ES than is an ES based on 10 subjects.
- Therefore, larger studies should carry more “weight” in our analyses than smaller studies.
- Simple approach: weight each ES by its sample size.
- Better approach: weight by the inverse variance.

10

What is the Inverse Variance Weight?

- The standard error (SE) is a direct index of ES precision.
- SE is used to create confidence intervals.
- The smaller the SE, the more precise the ES.
- Hedges’ showed that the optimal weights for meta-analysis are:

$$w = \frac{1}{SE^2}$$

11

Inverse Variance Weight for the Three Major League Effect Sizes

- Standardized Mean Difference:

$$se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{\overline{ES}_{sm}}{2(n_1 + n_2)}} \quad w = \frac{1}{se^2}$$

- Zr transformed Correlation Coefficient:

$$se = \sqrt{\frac{1}{n-3}} \quad w = n - 3$$

12

Inverse Variance Weight for the Three Major League Effect Sizes

- Logged Odds-Ratio:

$$se = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad w = \frac{1}{se^2}$$

Where a, b, c, and d are the cell frequencies of a 2 by 2 contingency table.

Ready to Analyze

- We have an independent set of effect sizes (ES) that have been transformed and/or adjusted, if needed.
- For each effect size we have an inverse variance weight (w).

The Weighted Mean Effect Size

Study	ES	w
1	-0.33	11.91
2	0.32	28.57
3	0.39	58.82
4	0.31	29.41
5	0.17	13.89
6	0.64	8.55
7	-0.33	9.80
8	0.15	10.75
9	-0.02	83.33
10	0.00	14.93

- Start with the effect size (ES) and inverse variance weight (w) for 10 studies.

$$\overline{ES} = \frac{\sum (w \times ES)}{\sum w}$$

The Weighted Mean Effect Size

Study	ES	w	w*ES
1	-0.33	11.91	-3.93
2	0.32	28.57	9.14
3	0.39	58.82	22.94
4	0.31	29.41	9.12
5	0.17	13.89	2.36
6	0.64	8.55	5.47
7	-0.33	9.80	-3.25
8	0.15	10.75	1.61
9	-0.02	83.33	-1.67
10	0.00	14.93	0.00

- Start with the effect size (ES) and inverse variance weight (w) for 10 studies.
- Next, multiply w by ES.

The Weighted Mean Effect Size

Study	ES	w	w*ES
1	-0.33	11.91	-3.93
2	0.32	28.57	9.14
3	0.39	58.82	22.94
4	0.31	29.41	9.12
5	0.17	13.89	2.36
6	0.64	8.55	5.47
7	-0.33	9.80	-3.24
8	0.15	10.75	1.61
9	-0.02	83.33	-1.67
10	0.00	14.93	0.00
		269.96	41.82

- Start with the effect size (ES) and inverse variance weight (w) for 10 studies.
- Next, multiply w by ES.
- Repeat for all effect sizes.

The Weighted Mean Effect Size

Study	ES	w	w*ES
1	-0.33	11.91	-3.93
2	0.32	28.57	9.14
3	0.39	58.82	22.94
4	0.31	29.41	9.12
5	0.17	13.89	2.36
6	0.64	8.55	5.47
7	-0.33	9.80	-3.24
8	0.15	10.75	1.61
9	-0.02	83.33	-1.67
10	0.00	14.93	0.00
		269.96	41.82

- Start with the effect size (ES) and inverse variance weight (w) for 10 studies.
- Next, multiply w by ES.
- Repeat for all effect sizes.
- Sum the columns, w and ES.
- Divide the sum of (w*ES) by the sum of (w).

$$\overline{ES} = \frac{\sum (w \times ES)}{\sum w} = \frac{41.82}{269.96} = 0.15$$

The Standard Error of the Mean ES

Study	ES	w	w*ES
1	-0.33	11.91	-3.93
2	0.32	28.57	9.14
3	0.39	58.82	22.94
4	0.31	29.41	9.12
5	0.17	13.89	2.36
6	0.64	8.55	5.47
7	-0.33	9.80	-3.24
8	0.15	10.75	1.61
9	-0.02	83.33	-1.67
10	0.00	14.93	0.00
		269.96	41.82

- The standard error of the mean is the square root of 1 divided by the sum of the weights.

$$se_{\overline{ES}} = \sqrt{\frac{1}{\sum w}} = \sqrt{\frac{1}{269.96}} = 0.061$$

Mean, Standard Error, Z-test and Confidence Intervals

Mean ES

$$\overline{ES} = \frac{\sum (w \times ES)}{\sum w} = \frac{41.82}{269.96} = 0.15$$

SE of the Mean ES

$$se_{\overline{ES}} = \sqrt{\frac{1}{\sum w}} = \sqrt{\frac{1}{269.96}} = 0.061$$

Z-test for the Mean ES

$$Z = \frac{\overline{ES}}{se_{\overline{ES}}} = \frac{0.15}{0.061} = 2.46$$

95% Confidence Interval

$$\begin{aligned} Lower &= \overline{ES} - 1.96(se_{\overline{ES}}) = 0.15 - 1.96(.061) = 0.03 \\ Upper &= \overline{ES} + 1.96(se_{\overline{ES}}) = 0.15 + 1.96(.061) = 0.27 \end{aligned}$$

Homogeneity Analysis

- 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.
 - Two options:
 - model between study differences
 - fit a random effects model

Q - The Homogeneity Statistic

Study	ES	w	w*ES	w*ES^2
1	-0.33	11.91	-3.93	1.30
2	0.32	28.57	9.14	2.93
3	0.39	58.82	22.94	8.95
4	0.31	29.41	9.12	2.83
5	0.17	13.89	2.36	0.40
6	0.64	8.55	5.47	3.50
7	-0.33	9.80	-3.24	1.07
8	0.15	10.75	1.61	0.24
9	-0.02	83.33	-1.67	0.03
10	0.00	14.93	0.00	0.00
		269.96	41.82	21.24

- Calculate a new variable that is the ES squared multiplied by the weight.
- Sum new variable.

Calculating Q

We now have 3 sums:

$$\sum w = 269.96$$

$$\sum (w \times ES) = 41.82$$

$$\sum (w \times ES^2) = 21.24$$

Q can be calculated using these 3 sums:

$$Q = \sum (w \times ES^2) - \frac{[\sum (w \times ES)]^2}{\sum w} = 21.24 - \frac{41.82^2}{269.96} = 21.24 - 6.48 = 14.76$$

Interpreting Q

- Q is distributed as a Chi-Square
- df = number of ESs - 1
- Running example has 10 ESs, therefore, df = 9
- Critical Value for a Chi-Square with df = 9 and p = .05 is:
 - 16.92
- Since our Calculated Q (14.76) is less than 16.92, we **fail to reject** the null hypothesis of homogeneity.
- Thus, the variability across effect sizes does not exceed what would be expected based on sampling error.

Heterogeneous Distributions: What Now?

- Analyze excess between study (ES) variability
 - categorical variables with the analog to the one-way ANOVA
 - continuous variables and/or multiple variables with weighted multiple regression
- Assume variability is random and fit a random effects model.

Analyzing Heterogeneous Distributions: The Analog to the ANOVA

Study	Grp	ES	w	w*ES	w*ES^2
1	1	-0.33	11.91	-3.93	1.30
2	1	0.32	28.57	9.14	2.93
3	1	0.39	58.82	22.94	8.95
4	1	0.31	29.41	9.12	2.83
5	1	0.17	13.89	2.36	0.40
6	1	0.64	8.55	5.47	3.50
			151.15	45.10	19.90
7	2	-0.33	9.80	-3.24	1.07
8	2	0.15	10.75	1.61	0.24
9	2	-0.02	83.33	-1.67	0.03
10	2	0.00	14.93	0.00	0.00
			118.82	-3.29	1.34

- Calculate the 3 sums for each subgroup of effect sizes.

A grouping variable (e.g., random vs. nonrandom)

Analyzing Heterogeneous Distributions: The Analog to the ANOVA

Calculate a separate Q for each group:

$$Q_{GROUP_1} = 19.90 - \frac{45.10^2}{151.15} = 6.44$$

$$Q_{GROUP_2} = 1.34 - \frac{-3.29^2}{118.82} = 1.25$$

Analyzing Heterogeneous Distributions: The Analog to the ANOVA

The sum of the individual group Qs = Q within:

$$Q_w = Q_{GROUP_1} + Q_{GROUP_2} = 6.44 + 1.25 = 7.69$$

$$df = k - j = 10 - 2 = 8$$

Where k is the number of effect sizes and j is the number of groups.

The difference between the Q total and the Q within is the Q between:

$$Q_b = Q_T - Q_w = 14.76 - 7.69 = 7.07$$

$$df = j - 1 = 2 - 1 = 1$$

Where j is the number of groups.

Analyzing Heterogeneous Distributions: The Analog to the ANOVA

All we did was partition the overall Q into two pieces, a within groups Q and a between groups Q.

$Q_b = 7.69$	$df_b = 1$	$Q_{CV,.05}(1) = 3.84$	$p_b < .05$
$Q_w = 7.07$	$df_w = 8$	$Q_{CV,.05}(8) = 15.51$	$p_w > .05$
$Q_T = 14.76$	$df_T = 9$	$Q_{CV,.05}(9) = 16.92$	$p_T > .05$

The grouping variable accounts for significant variability in effect sizes.

Mean ES for each Group

The mean ES, standard error and confidence intervals can be calculated for each group:

$$ES_{GROUP_1} = \frac{\sum (w \times ES)}{\sum w} = \frac{45.10}{151.15} = 0.30$$

$$ES_{GROUP_2} = \frac{\sum (w \times ES)}{\sum w} = \frac{-3.29}{118.82} = -0.03$$

Analyzing Heterogeneous Distributions: Multiple Regression Analysis

- Analog to the ANOVA is restricted to a single categorical between studies variable.
- What if you are interested in a continuous variable or multiple between study variables?
- Weighted Multiple Regression Analysis
 - as always, it is weighted analysis
 - can use “canned” programs (e.g., SPSS, SAS)
 - parameter estimates are correct (R-squared, B weights, etc.)
 - F-tests, t-tests, and associated probabilities are **incorrect**
 - can use Wilson/Lipsey SPSS macros which give correct parameters and probability values

Meta-Analytic Multiple Regression Results From the Wilson/Lipsey SPSS Macro (data set with 39 ESs)

```

**** Meta-Analytic Generalized OLS Regression ****
----- Homogeneity Analysis -----
      Q      df      P
Model  104.9704   3.0000   .0000
Residual 424.6276  34.0000   .0000
    
```

Partition of total Q into variance explained by the regression "model" and the variance left over ("residual").

```

----- Regression Coefficients -----
      B      SE   -95% CI  +95% CI      Z      P      Beta
Constant  -.7782  .0925  -.9595  -.5970  -8.4170  .0000  .0000
RANDOM     .0786  .0215  .0364   .1207   3.6548  .0003  .1696
TXVAR1    .5065  .0753  .3590   .6541   6.7285  .0000  .2933
TXVAR2    .1641  .0231  .1188   .2094   7.1036  .0000  .3298
    
```

Interpretation is the same as will ordinal multiple regression analysis.
If residual Q is significant, fit a mixed effects model.

Review of Weighted Multiple Regression Analysis

- Analysis is weighted.
- Q for the model indicates if the regression model explains a significant portion of the variability across effect sizes.
- Q for the residual indicates if the remaining variability across effect sizes is homogeneous.
- If using a “canned” regression program, must correct the probability values (see manuscript for details).

33

Random Effects Models

- **Don't panic!**
- It sounds far worse than it is.
- Three reasons to use a random effects model
 - Total Q is significant and you assume that the excess variability across effect sizes derives from random differences across studies (sources you cannot identify or measure).
 - The Q within from an Analog to the ANOVA is significant.
 - The Q residual from a Weighted Multiple Regression analysis is significant.

34

The Logic of a Random Effects Model

- Fixed effects model assumes that all of the variability between effect sizes is due to sampling error.
- Random effects model assumes that the variability between effect sizes is due to sampling error **plus** variability in the population of effects (unique differences in the set of true population effect sizes).

35

The Basic Procedure of a Random Effects Model

- Fixed effects model weights each study by the inverse of the sampling variance.

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

- Random effects model weights each study by the inverse of the sampling variance **plus** a constant that represents the variability across the population effects.

$$w_i = \frac{1}{se_i^2 + \hat{\nu}_\theta}$$

This is the random effects variance component.

36

How To Estimate the Random Effects Variance Component

- The random effects variance component is based on Q.
- The formula is:

$$\hat{v}_\theta = \frac{Q_T - (k-1)}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)}$$

Calculation of the Random Effects Variance Component

Study	ES	w	w*ES	w*ES^2	w^2
1	-0.33	11.91	-3.93	1.30	141.73
2	0.32	28.57	9.14	2.93	816.30
3	0.39	58.82	22.94	8.95	3460.26
4	0.31	29.41	9.12	2.83	865.07
5	0.17	13.89	2.36	0.40	192.90
6	0.64	8.55	5.47	3.50	73.05
7	-0.33	9.80	-3.24	1.07	96.12
8	0.15	10.75	1.61	0.24	115.63
9	-0.02	83.33	-1.67	0.03	6944.39
10	0.00	14.93	0.00	0.00	222.76
		269.96	41.82	21.24	12928.21

- Calculate a new variable that is the w squared.
- Sum new variable.

Calculation of the Random Effects Variance Component

- The total Q for this data was 14.76
- k is the number of effect sizes (10)
- The sum of w = 269.96
- The sum of w² = 12,928.21

$$\hat{v}_\theta = \frac{Q_T - (k-1)}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)} = \frac{14.76 - (10-1)}{269.96 - \frac{12,928.21}{269.96}} = \frac{5.76}{269.96 - 47.89} = \underline{0.026}$$

Rerun Analysis with New Inverse Variance Weight

- Add the random effects variance component to the variance associated with each ES.

$$w_i = \frac{1}{se_i^2 + \hat{v}_\theta}$$

- Calculate a new weight.
- Rerun analysis.
- **Congratulations! You have just performed a very complex statistical analysis.**

Random Effects Variance Component for the Analog to the ANOVA and Regression Analysis

- The Q between or Q residual replaces the Q total in the formula.
- Denominator gets a little more complex and relies on matrix algebra. However, the logic is the same.
- SPSS macros perform the calculation for you.

SPSS Macro Output with Random Effects Variance Component

```

----- Homogeneity Analysis -----
              Q          df          p
Model         104.9704      3.0000      .0000
Residual      424.6276     34.0000      .0000

----- Regression Coefficients -----
              B          SE    -95% CI  +95% CI      Z      P      Beta
Constant    -.7782     .0925    -.9595  -.5970   -8.4170 .0000 .0000
RANDOM       .0786     .0215     .0364  .1207    3.6548 .0003 .1696
TXVAR1      .5065     .0753     .3590  .6541    6.7285 .0000 .2933
TXVAR2      .1641     .0231     .1188  .2094    7.1036 .0000 .3298

```

```

----- Estimated Random Effects Variance Component -----
v = .04715

Not included in above model which is a fixed effects model
Random effects variance component based on the residual Q. Add this
value to each ES variance (SE squared) and recalculate w. Rerun analysis
with the new w.

```

Comparison of Random Effect with Fixed Effect Results

- The biggest difference you will notice is in the significance levels and confidence intervals.
 - Confidence intervals will get bigger.
 - Effects that were significant under a fixed effect model may no longer be significant.
- Random effects models are therefore more conservative.

Review of Meta-Analytic Data Analysis

- Transformations, Adjustments and Outliers
- The Inverse Variance Weight
- The Mean Effect Size and Associated Statistics
- Homogeneity Analysis
- Fixed Effects Analysis of Heterogeneous Distributions
 - Fixed Effects Analog to the one-way ANOVA
 - Fixed Effects Regression Analysis
- Random Effects Analysis of Heterogeneous Distributions
 - Mean Random Effects ES and Associated Statistics
 - Random Effects Analog to the one-way ANOVA
 - Random Effects Regression Analysis

Interpreting Effect Size Results

- Cohen's "Rules-of-Thumb"
 - standardized mean difference effect size
 - small = 0.20
 - medium = 0.50
 - large = 0.80
 - correlation coefficient
 - small = 0.10
 - medium = 0.25
 - large = 0.40
 - odds-ratio
 - small = 1.50
 - medium = 2.50
 - large = 4.30
- These do not take into account the context of the intervention
- They do correspond to the distribution of effects across meta-analyses found by Lipsey and Wilson (1993)

1

Interpreting Effect Size Results

- Rules-of-Thumb do not take into account the context of the intervention
 - a "small" effect may be highly meaningful for an intervention that requires few resources and imposes little on the participants
 - small effects may be more meaningful for serious and fairly intractable problems
- Cohen's Rules-of-Thumb do, however, correspond to the distribution of effects across meta-analyses found by Lipsey and Wilson (1993)

2

Translation of Effect Sizes

- Original metric
- Success Rates (Rosenthal and Rubin's BESD)
 - Proportion of "successes" in the treatment and comparison groups assuming an overall success rate of 50%
 - Can be adapted to alternative overall success rates
- Example using the sex offender data
 - Assuming a comparison group recidivism rate of 15%, the effect size of 0.45 for the cognitive-behavioral treatments translates into a recidivism rate for the treatment group of 7%

3

Methodological Adequacy of Research Base

- Findings must be interpreted within the bounds of the methodological quality of the research base synthesized.
- Studies often cannot simply be grouped into "good" and "bad" studies.
- Some methodological weaknesses may bias the overall findings, others may merely add "noise" to the distribution.

4

Confounding of Study Features

- Relative comparisons of effect sizes across studies are inherently correlational!
- Important study features are often confounding, obscuring the interpretive meaning of observed differences
- If the confounding is not severe and you have a sufficient number of studies, you can model "out" the influence of method features to clarify substantive differences

5

Concluding Comments

- Meta-analysis is a replicable and defensible method of synthesizing findings across studies
- Meta-analysis often points out gaps in the research literature, providing a solid foundation for the next generation of research on that topic
- Meta-analysis illustrates the importance of replication
- Meta-analysis facilitates generalization of the knowledge gain through individual evaluations

6