

Meta-Analysis:

An Introduction



Elizabeth Moorman Kim, PhD SRM Statistics and Research Methodology Series April 22, 2011

Nebraska Lincoln

OVERVIEW

• What is meta-analysis?



• When is it appropriate to do a meta-analysis?

- What are the steps for conducting a metaanalysis?
 - Problem formation and literature search
 - Coding and analysis
 - Interpretation and presentation of results



WHAT IS META-ANALYSIS?

• Literature review

Theoretical review

- Research synthesis (i.e., research review, systematic review)
 - *Meta-analysis*: quantitative procedures for combining results



WHEN TO DO A META-ANALYSIS?

"High parent involvement in the child's education w[as] associated with school success in terms of school records of achievement."

~Miliotis, Sesma, & Masten, 1999, p. 111

"The findings suggest that parental involvement does not independently improve children's learning." ~Domina, 2005, p. 223

"Parental involvement in 3rd grade had a significant direct effect on achievement in 3rd grade."

~Englund, Luckner, Whaley, & Egeland, 2004, p.723

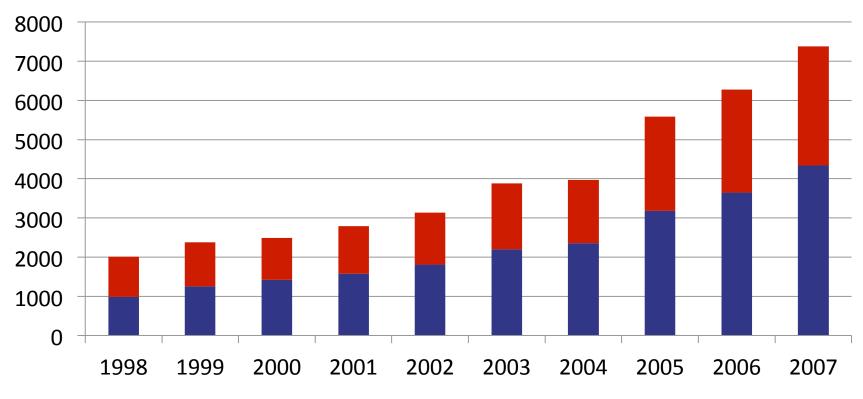
"The results of the between-child analyses suggested that higher parent involvement...is unrelated to average achievement across elementary school."

~El Nokali, Bachman, & Votruba-Drzal, 2010, p. 1001

What should we conclude from these findings?



Citations for the Years 1998 to 2007

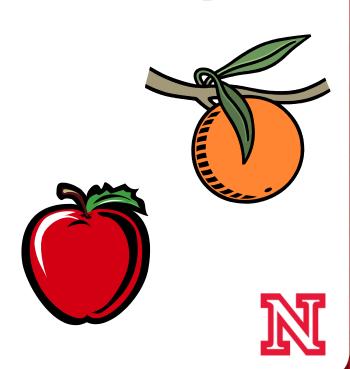


- Meta-analysis
- Research synthesis, systematic review, research review, or literature review



WHEN TO DO A META-ANALYSIS?

- Body of empirical research
- Producing quantitative findings
 - Utilizes data typically reported in research reports
- Comparable conceptually
- Similar statistical forms
 - Apples and oranges



STEPS IN META-ANALYSIS

- 1. Specifying the Problem
- 2. Searching the Literature

- 3. Report Coding
- 4. Data Analysis

5. Interpretation and Presentation of Results





1. SPECIFYING THE PROBLEM

- Consider it carefully
 - Breadth vs. Specificity
 - Does not need to be highly detailed early on
 - Research literature
 - Independent variables
 - Dependent variables
 - Type of relationship



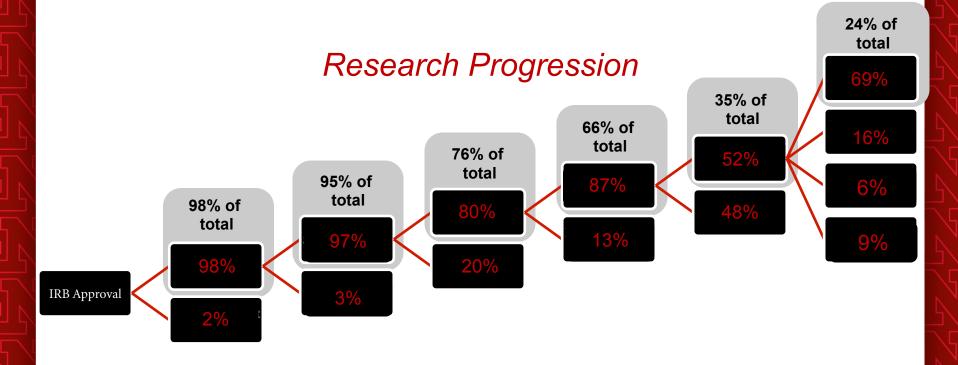


How should I conduct my search?

- What outlets should I use for gathering studies?
- What criteria should I use in my search?

- How do I go about retrieving the studies?
- How do I manage the yielded reports?

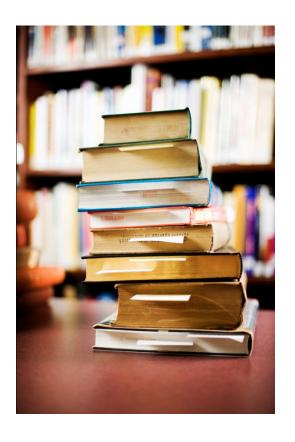




From "Finding the missing science: The fate of studies submitted for review by a human subjects committee," by H. Cooper, K. DeNeve, & K. Charlton, 1997, *Psychological Methods*, 2, 448-449.

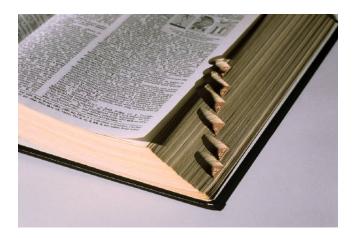


- Approaches
 - Contact
 - Mass Solicitation
 - Conference Presentations
 - Journals
 - References
 - Reference Databases

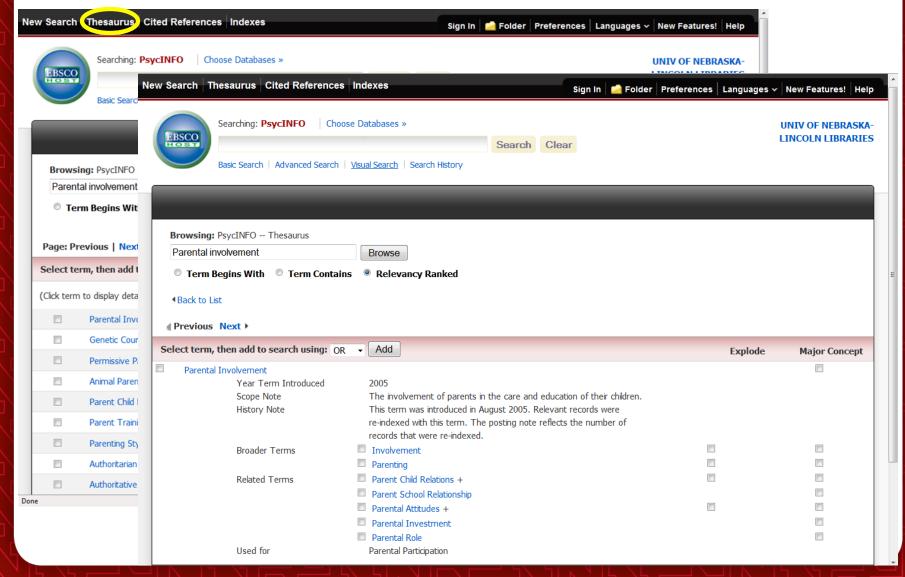




- Reference Databases
 - Which ones should I use?
 - Consider published and unpublished
 - How do I decide on my search terms?
 - Thesaurus







- Eligibility criteria
 - Distinguishing Features
 - Respondents
 - Variables
 - Design
 - Culture/Language
 - Time
 - Type of Publication



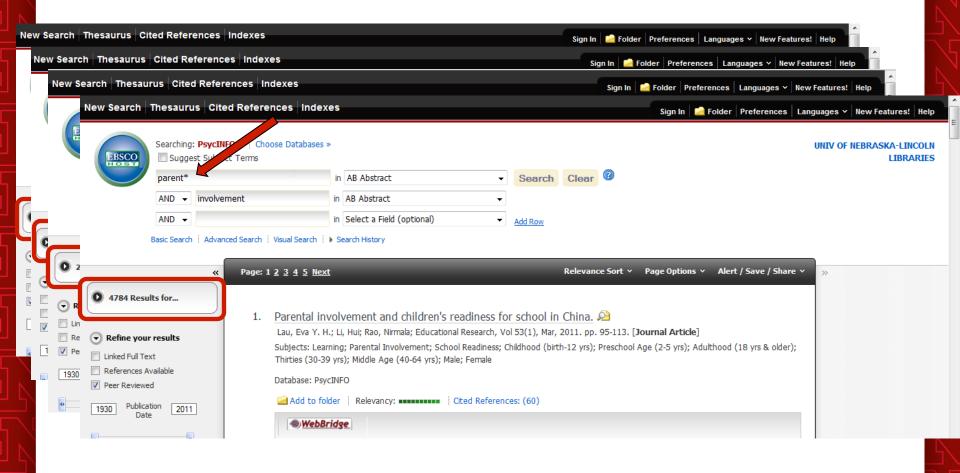


Limit your results			
Linked Full Text		References Available	
Publication Name		Publication Year from	to
Published Date from	Month Year: to Month Year:	Peer Reviewed	
		Publisher	
Publication Status	All first posting fully published	Publication Type	Peer Reviewed Journal Peer-Reviewed Status-Unknown All Books Authored Book
English		Language	English Finnish French Georgian
Age Groups	All Childhood (birth-12 yrs) Neonatal (birth-1 mo) Infancy (2-23 mo)	Population Group	All Human Animal Male
Intended Audience	All General Public Juvenile Psychology: Professional & Research	Document Type	Erratum/Correction Journal Article Letter Obituary
Book Type	Classic Book Conference Proceedings Handbook/Manual	Methodology	All All BRAIN IMAGING CLINICAL CASE STUDY EMPIRICAL STUDY
Classification Codes	2100 General Psychology 2140 History & Systems 2200 Psychometrics & Statistics & Methodology	Exclude Dissertations	

- Search terms
 - Operators
 - "AND" vs. "OR"
 - Truncation
 - Field



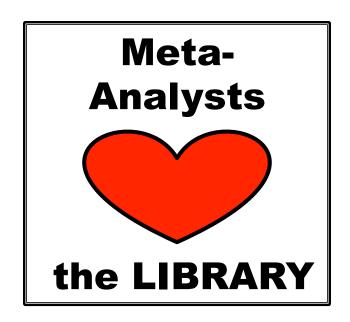




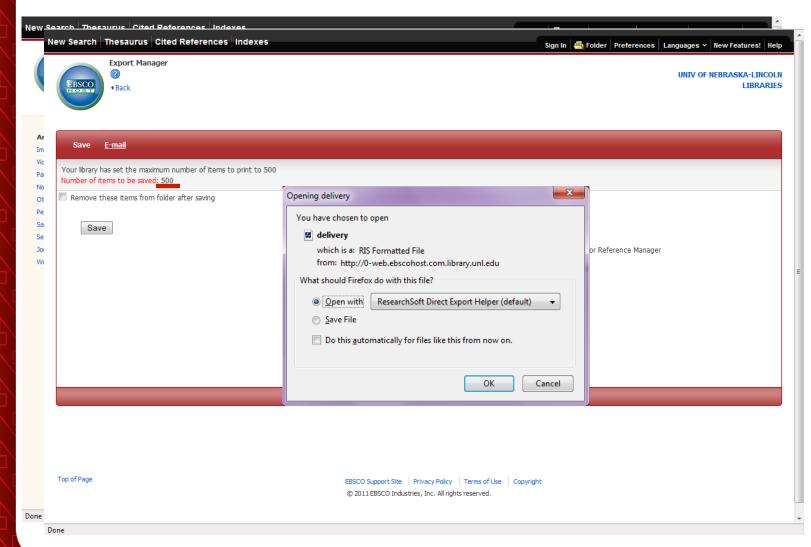


- Literature Retrieval
 - Library System
 - Inter-Library Loan
 - Personal Contact

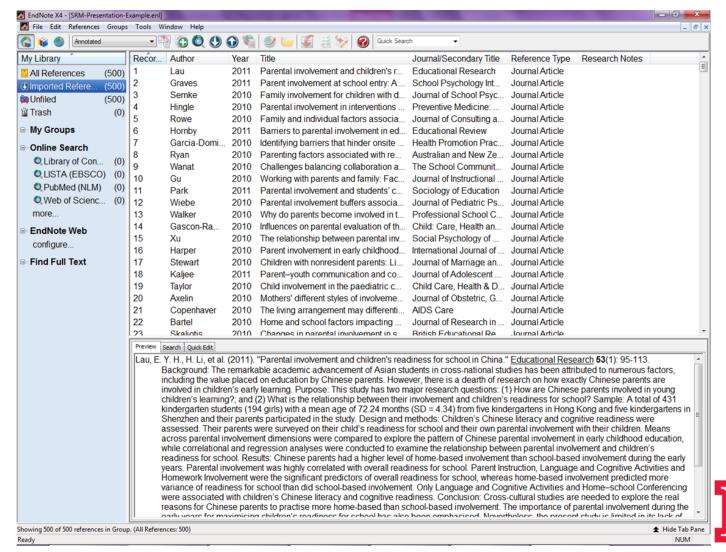
Maintain Records











- Developing a coding protocol
 - Planning
 - A priori
 - Review of literature
 - Frequency
 - Content
 - Study descriptors
 - Moderators
 - Effect sizes
 - Effect size statistic
 - Associated sample size



- Study Descriptors
 - Bibliographic Information
 - e.g., type of report, publication year, authors, etc.

Parental Involvement Research Synthesis

Study Descriptors		
Bibliographic Information	Code	
1. Record Number		
EndNote Record Number		
2. Study Authors		
Last Name, First Initial; etc.		
3. Year		
Publication Year		
4. Type of Report		
1 = Journal article		
2 = Book or book chapter		
3 = Government report		
4 = Conference paper	18	
9 = Other (please specify)		

- Study Descriptors
 - Setting
 - e.g., country, type of community (urban, rural), type of school (public, private), etc.

Setting Information	Code
5. Country	
1 = USA	
2 = Canada	
3 = Mexico	
9 = Other (please specify)	
6. Community type	
1 = Urban	
2 = Suburban	
3 = Rural	
4 = Heterogeneous	
9 = Other (please specify)	
7. Language research conducted in	
1 = English	
2 = Spanish	
9 = Other (please specify)	
8. Type of school	
1 = Public	
2 = Private	
9 = Other	



- Study Descriptors
 - Description of the Sample
 - e.g., age, ethnicity, gender, etc.

Description of Sample	Code
9. Number of children	
Specify number of children participating	
10. Grade(s) in school	
-1 = Preschool	
0 = Kindergarten	
1 = First grade	
12 = 12 th grade	
11. Socioeconomic status	
1 = Low SES	
2 = Low-middle SES	
3 = Middle SES	
4 = Middle-upper SES	
5 = Upper SES	
6 = Heterogeneous	
9 = Other (please specify)	
12. Gender	
% Male	
% Female	



- Study Descriptors
 - Research Design

• e.g., unit of assignment, type of assignment, etc.

e.g., unit of assignment, type of assignment, etc.			
Research Design	Code		
13. Unit of assignment			
1 = Individual children/parents			
2 = Small groups			
3 = Classroom			
4 = School			
5 = Region			
9 = Other (please specify)			
14. Type of assign			
1 = Completely randomized			
2 = Randomized block			
3 = Nonrandomized			
4 = Nonrandomized block			
9 = Other (please specify)			
15. Control group procedures			
1 = Business-as-usual			
2 = Attention placebo			
3 = Alternative intervention			
4 = Wait list/delayed intervention			
9 = Other (please specify)			



- Study Descriptors
 - Independent Variables

Indonondant Variables	Codo
Independent Variables	Code
16. Type of Involvement	
Home-based	
11 = Homework involvement	
12 = Talking about school	
13 = Reading together	
14 = Planning	
19 = Other home-based (please specify)	
School-based	
21 = Volunteering	
22 = Parent-teacher association	
23 = Parent-teacher conferences	
24 = Open houses	
29 = Other school-based (please specify)	
17. Reporter	
1 = Parent	
2 = Child	
3 = Teacher	
9 = Other (please specify)	



• Study Descriptors

• Dependent Variables

Dependent variables	
Dependent Variables	Code
18. Child Academic Adjustment	
Grades	
111 = Grades: Combined Subjects	
112 = Grades: Language Arts	
113 = Grades: Math	
119 = Grades: Other	
Standardized Test Scores	
121 = Test Scores: Combined	
122 = Test Scores: Language Arts	
123 = Test Scores: Math	
129 = Test Scores: Other	
Academic Competence	
131 = Academic Competence: Combined	
132 = Academic Competence: Language Arts	
133 = Academic Competence: Math	
139 = Academic Competence: Other	
19. Reporter	
1 = Records	
2 = Parent	
3 = Child	
4 = Teacher	
9 = Other (please specify)	



Effect Size Coding

Standardized Mean Difference

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_{pooled}}$$

 \bar{x}_1 = observed mean of group 1

 \overline{x}_2 = observed mean of group 2

 S_{pooled} = pooled standard deviation

Bias-Corrected

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

N = total sample size(Hedges, 1981)

Sampling Variance

$$v_i = \frac{n_1 + n_2}{n_1 n_2} + \frac{(ES'_{sm})^2}{2(n_1 + n_2)}$$

 n_1 = sample size in group 1

 n_2 = sample size in group 2



Effect Size Information	Code
17. Treatment Group Mean	
18. Treatment Group Standard Deviation	
19. Treatment Group Sample Size	
20. Control Group Mean	
21. Control Group Standard Deviation	
22. Control Group Sample Size	
23. Effect Size (calculated)	



Effect Size Coding

Correlation Coefficient

$$r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

 σ_{xy}^2 = covariance between x and y

 $\sigma_{\rm r}$ = standard deviation of x

 σ_{v} = standard deviation of y

Z-transformation

$$ES_{z_r} = \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right]$$

r = correlation coefficient

In = natural logarithm

Sampling Variance

$$v_{Z_r} = \frac{1}{N - 3}$$

N = total sample size



Correlation Coefficient Information	Code	
26. Correlation Coefficient		
27. Correlation Sample Size		



- Coder Training
 - Choosing coders
 - Elaborated coding protocol
 - Independent coding, group comparison and discussion
 - Double coding





- Data Preparation
 - Converting effect sizes to a common metric
 - Direction of effect
 - Multiple effect sizes
 - Random selection
 - Underrepresented areas
 - Aggregation
 - Statistical artifacts (Hunter & Schmidt, 2004)
 - Sampling error
 - Measurement unreliability
 - Range restriction
 - Dichotomization





Model Selection

$$ES_i = \theta_i + \varepsilon_i$$

Fixed effects model

i.e.,
$$\theta_1 = \theta_2 = \dots = \theta_k$$

- Fixed effects with moderators
 - Differences due to presence of moderators
- Random-effects
 - Differences due to random heterogeneity
- Mixed-effects
 - Differences due to presence of moderators and random heterogeneity

Model Selection

Model	Moderators present	Random heterogeneity
Fixed effects	No	No
Fixed effects with moderators	Yes	No
Random effects	No	Yes
Mixed effects	Yes	Yes

From Roberts, B. W., Kuncel, N. R., Viechtbauer, W, & Bogg, T. (2007). Meta-analysis in personality psychology: A primer. (Ch. 36, pp. 652-672). In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology.* New York, NY, USA: Guilford Press.



• Standardized Mean Difference

Inverse Variance

$$w_i = \frac{2(n_{i1} + n_{i2})n_{i1}n_{i2}}{2(n_{i1} + n_{i2})^2 + n_{i1}n_{i2}d_i^2}$$

 n_{i1} = number of data points in group 1

 n_{i2} = number of data points in group 2

 d_i = effect size

Weighted Average Effect Size

$$d_{\bullet} = \frac{\sum_{i=1}^{k} d_i w_i}{\sum_{i=1}^{k} w_i}$$

k = number of independent samples

95% Confidence Interval

$$CI_{d.95\%} = d. \pm 1.96 \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}$$



Study	n_{i1}	n_{i2}	d_i	w_i	$d_i w_i$
1	111	117	0.06	56.93	3.42
2	64	63	0.14	31.67	4.43
3	55	51	0.25	26.26	6.56
4	200	194	0.11	98.33	10.82
5	285	290	0.08	143.62	11.49
6	120	129	0.30	61.48	18.44
7	71	75	0.16	36.36	5.82
Σ	906	919	1.10	454.65	60.98

95% Confidence Interval:

$$CI_{d.95\%} = d. \pm 1.96 \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}} = .13 \pm 1.96 \sqrt{\frac{1}{454.65}} = .13 \pm .09 = [.04, .22]$$



• Correlation Coefficient (*Z*-transformed)

Weighted Average Effect Size

$$z_{\bullet} = \frac{\sum_{i=1}^{k} (n_i - 3) z_i}{\sum_{i=1}^{k} (n_i - 3)}$$

 n_i = sample size for i^{th} independent sample

 $Z_i = z$ -transformed correlation

95% Confidence Interval

$$CI_{z.95\%} = z. \pm \frac{1.96}{\sqrt{\sum_{i=1}^{k} (n_i - 3)}}$$



Study	n_i	r_i	z_i	<i>n</i> _i -3	$(n_i$ -3) z_i
1	538	0.22	0.22	535	119.66
2	60	0.21	0.21	57	12.15
3	115	0.20	0.20	112	22.58
4	109	0.17	0.17	106	18.41
5	323	0.29	0.30	320	94.44
6	335	0.12	0.12	332	41.29
Σ	1480	1.21	1.23	1462	308.53

Weighted Average Effect Size:
$$z_{\bullet} = \frac{\sum_{i=1}^{k} (n_i - 3) z_i}{\sum_{i=1}^{k} (n_i - 3)} = \frac{308.53}{1462} = .21$$

95% Confidence Interval:

$$CI_{z.95\%} = z. \pm \frac{1.96}{\sqrt{\sum_{i=0.5\%}^{k} (n_i - 3)}} = .21 \pm \frac{1.96}{\sqrt{1462}} = .21 \pm .05 = [.16, .26]$$



Correlation Coefficient (untransformed)

Sampling Variance

$$v_r = \frac{\left(1 - r_i^2\right)^2}{n_i - 1}$$

 n_i = sample size for i^{th} independent sample

 \mathcal{V}_i = correlation coefficient

Weighted Average Effect Size

$$r_{\bullet} = \frac{\sum_{i=1}^{k} r_{i} w_{i}}{\sum_{i=1}^{k} w_{i}}$$

Inverse Variance

$$w_i = \frac{1}{v_r}$$

95% Confidence Interval

$$CI_{r.95\%} = r. \pm \frac{1.96}{\sqrt{\sum_{i=1}^{k} w_i}}$$



Study	n_i	r_i	ν_i	w_i	$r_i w_i$
1	538	0.22	0.00	593.01	130.46
2	60	0.21	0.02	64.57	13.56
3	115	0.20	0.01	123.58	24.58
4	109	0.17	0.01	114.68	19.72
5	323	0.29	0.00	382.33	109.67
6	335	0.12	0.00	344.47	42.63
Σ	1480	1.21	0.04	1622.65	340.62

$$CI_{r.95\%} = r. \pm \frac{1.96}{\sqrt{\sum_{i=0.5}^{k} w_i}} = .21 \pm \frac{1.96}{\sqrt{1622.65}} = .21 \pm .05 = [.16, .26]$$



- Homogeneity
 - Is the variance observed different than would be expected by sampling error only?
 - Q statistic
 - If so, what might moderate this effect?



Standardized Mean Difference

$$Q = \sum_{i=1}^{k} w_i d_i^2 - \frac{\left(\sum_{i=1}^{k} w_i d_i\right)^2}{\sum_{i=1}^{k} w_i} = 11.04 - \frac{60.98^2}{454.65} = 11.04 - 8.18 = 2.86$$

Compare to chi-square with *k*-1 degrees of freedom

Study	n_{i1}	n_{i2}	d_i	${\it w}_i$	$d_i w_i$	$w_i d_i^2$
1	111	117	0.06	56.93	3.42	0.20
2	64	63	0.14	31.67	4.43	0.62
3	55	51	0.25	26.26	6.56	1.64
4	200	194	0.11	98.33	10.82	1.19
5	285	290	0.08	143.62	11.49	0.92
6	120	129	0.30	61.48	18.44	5.53
7	71	75	0.16	36.36	5.82	0.93
Σ	906	919	1.10	454.65	60.98	11.04
						_

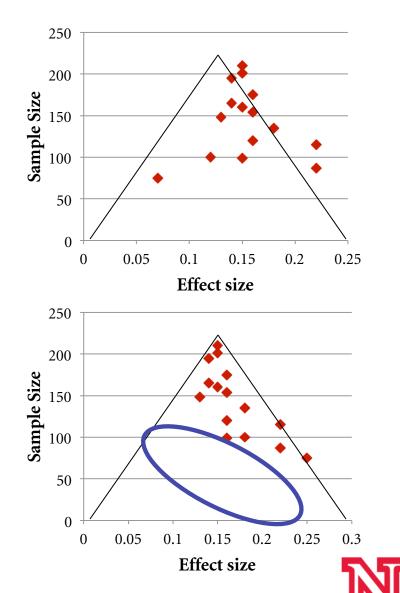
• Correlation Coefficient (*Z*-transformed)

$$Q = \sum_{i=1}^{k} (n_i - 3)z_i^2 - \frac{\left[\sum_{i=1}^{k} (n_i - 3)z_i\right]^2}{\sum_{i=1}^{k} (n_i - 3)} = 70.11 - \frac{308.53^2}{1462} = 70.11 - 65.11 = 5$$

Compare to chi-square with *k*-1 degrees of freedom

Study	n_i	r_i	z_i	n_i -3	$(n_i$ -3) z_i	$(n_i - 3)z_i^2$
1	538	0.22	0.22	535	119.66	26.76
2	60	0.21	0.21	57	12.15	2.59
3	115	0.20	0.20	112	22.58	4.55
4	109	0.17	0.17	106	18.41	3.20
5	323	0.29	0.30	320	94.44	27.87
6	335	0.12	0.12	332	41.29	5.14
Σ	1480	1.21	1.23	1462	308.53	70.11

- Publication Bias
 - File drawer problem
 - Funnel plot
 - Fail-safe N (Rosenthal, 1979)
 - Trim and fill (Duval & Tweedie, 2000)



- Programs
 - SAS
 - SPSS
 - STATA
 - R
 - Stand-alone programs
 - Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005)

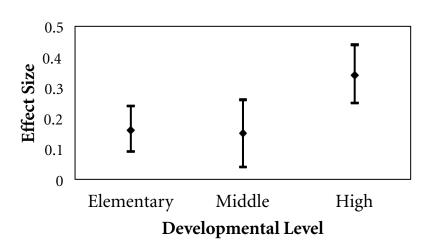


5. PRESENTATION OF RESULTS

Graphical Presentation

Error-Bar Chart

Stem and Leaf Display



• Meta-Analysis Reporting Standards (MARS; Publication Manual of the American Psychological Association (2010))

HOW DO I GET STARTED?

- Resources
 - Research team
 - Supplies



- Funding Mechanisms
 - IES Goal 1: Exploration
 - NIH R21: Exploratory/Developmental



REFERENCES

- Borenstein M, Hedges L, Higgins J, Rothstein H. (2005). *Comprehensive Meta-analysis Version 2*. Englewood NJ, USA: Biostat.
- Cooper, H. (2010). Research synthesis and meta-analysis. Thousand Oaks, CA, USA: Sage.
- Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, *2*, 448-449.
- Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.
- Hunter, J. E., & Schmidt, F. L. *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA, USA: Sage.
- Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, CA, USA: Sage.
- Roberts, B. W., Kuncel, N. R., Viechtbauer, W, & Bogg, T. (2007). Meta-analysis in personality psychology: A primer. (Ch. 36, pp. 652-672). In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology*. New York, NY, USA: Guilford Press.
- Rosenthal, R. (1979). The file-drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.





Thank You! Questions?

Elizabeth Moorman Kim, PhD emoorman2@unl.edu

Nebraska Lincoln