An introduction to meta-analysis: History, methods, misconceptions, and recent developments

Wolfgang Viechtbauer
Maastricht University
http://www.wvbauer.com

The Information Explosion

- 2,300 biomedical journals in 1940
- now there are close to 25,000
- approximately 27,000+ RCTs per year
- similar growth in other disciplines
- rough estimates:
  - # of articles double every ~10 years
  - # of journals double every ~15 years

The Information Explosion

↓

1. finding relevant literature
2. accessing the literature
3. maintaining awareness of the literature
4. **reading and processing the information**

Is this a new problem?

> The individual scientist is being overloaded with scientific information […] and can no longer keep up with and assimilate all the information being produced
> Garvey & Griffith (1971)

How to Summarize the Results?

- traditionally:
  - narrative literature reviews
  - vote counting methods
  - combining tests of significance
- now:
  - systematic reviews
  - meta-analysis

Narrative Literature Review

- a description/summary of the current state of knowledge on a particular topic supported by empirical findings as well as the underlying theories and models
- possible problems:
  - unsystematic
  - subjective
  - intractable
- in essence scientifically unsound

Narrative Literature Review

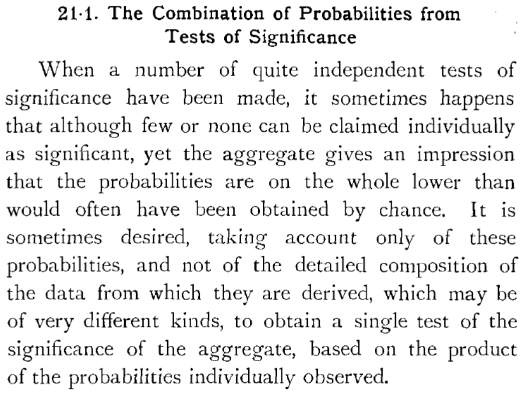
how replicable is the process of a person
reading dozens or even hundreds of
papers, thinking about them,
and then writing
down his or her
conclusions?

Vote Counting

- examine all relevant studies conducted
- categorize based on statistical significance
  - statistically significant (with $Trt > Ctrl$ or $r > 0$)
  - not statistically significant
  - statistically significant (with $Trt < Ctrl$ or $r < 0$)
- declare most frequent category the 'winner'
- inconsistent when power of studies is low (Hedges & Olkin, 1985): as $k \rightarrow \infty$, method fails to find a true effect or association

Combining Tests of Significance

- long history of methods for combining the results from independent significance tests (Tippett, 1931; Fisher, 1932; Pearson, 1933; Stouffer et al., 1949; Wilkinson, 1951; Mosteller & Bush, 1954; Good, 1955; Lipták, 1958; Lancaster, 1961; …)



21·1. The Combination of Probabilities from Tests of Significance

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. It is sometimes desired, taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate, based on the product of the probabilities individually observed.

Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.). London: Oliver and Boyd.

Fisher's Method

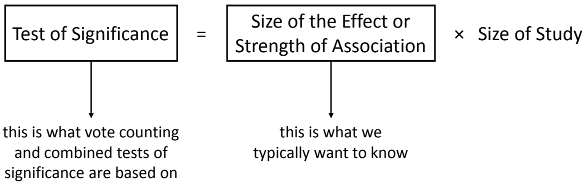
- if H_0 is true, then $p \sim \text{Uniform}(0, 1)$
- then we can show that $-2\ln[p] \sim \chi^2_2$
- now assume H_0 is true for $i = 1, \ldots, k$ tests
- then $-2\sum\ln[p_i] \sim \chi^2_{2k}$

Example

- want to know if x and y are correlated
- test $H_0: \rho = 0$ in three different studies
- $n = 50$ in all three studies
- find $r_1 = .33$, $r_2 = .25$, and $r_3 = .15$
- then $p_1 = .02$, $p_2 = .08$, and $p_3 = .30$
- so $-2\sum \ln[p_i] = 15.28$
- under a χ^2 distribution with $df = 6$, this yields a combined p-value of .018
- reject $H_0: \rho_1 = \rho_2 = \rho_3 = 0$

Combining Tests of Significance

- tests a fairly uninteresting null hypothesis
- uses little information from the studies
- also:



Test of Significance = Size of the Effect or Strength of Association × Size of Study

this is what vote counting and combined tests of significance are based on

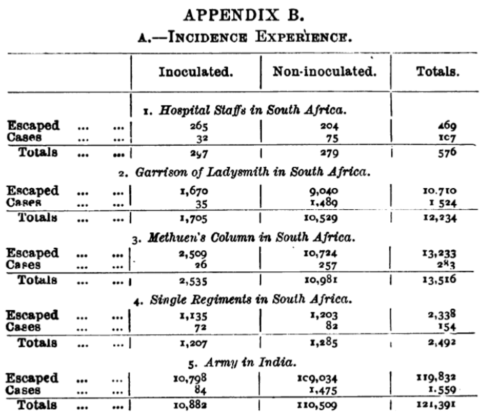
this is what we typically want to know

History/Origins of Meta-Analysis

- nice summary in Chalmers et al. (2002)
- Pearson (1904) – the earliest MA?
- Cochran et al. work in agriculture
- physics (Birge, 1932)
- origin of term “meta-analysis” (Glass, 1976)
- some early MAs in psychology
- Hedges & Olkin (1985), Light & Pillemer (1984)
- Cochrane and the EBM movement

Pearson (1904)

- maybe the earliest MA
- examined effectiveness of a vaccine against typhoid based on studies conducted among soldiers in the British Empire
- outcomes: infection and mortality
- data summarized in terms of 2 × 2 tables



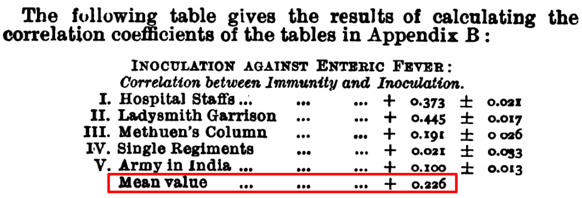
APPENDIX B.

A.—INCIDENCE EXPERIENCE.

	Inoculated.	Non-inoculated.	Totals.
1. *Hospital Staffs in South Africa.*			
Escaped	265	204	469
Cases	32	75	107
Totals	297	279	576
2. *Garrison of Ladysmith in South Africa.*			
Escaped	1,670	9,040	10.710
Cases	35	1,489	1 524
Totals	1,705	10,529	12,234
3. *Methuen's Column in South Africa.*			
Escaped	2,509	10,724	13,233
Cases	26	257	283
Totals	2,535	10,981	13.516
4. *Single Regiments in South Africa.*			
Escaped	1,135	1,203	2,338
Cases	72	82	154
Totals	1,207	1,285	2,492
5. *Army in India.*			
Escaped	10,798	109,034	119,832
Cases	84	1,475	1.559
Totals	10,882	110,509	121,391

Pearson (1904)

- calculated tetrachoric correlation between the two variables and averaged the results



The following table gives the results of calculating the correlation coefficients of the tables in Appendix B :

INOCULATION AGAINST ENTERIC FEVER:

Correlation between Immunity and Inoculation.

I. Hospital Staffs ...	...	...	+	0.373	±	0.021
II. Ladysmith Garrison	...	...	+	0.445	±	0.017
III. Methuen's Column	...	...	+	0.191	±	0 026
IV. Single Regiments	...	...	+	0.021	±	0.033
V. Army in India ...	...	...	+	0.100	±	0.013
Mean value ...	...	...	+	0.226		

Agricultural Research

- “Agricultural experiments on the same factor or group of factors are usually carried out at a number of places and repeated over a number of years. […] The agricultural experimenter is thus frequently confronted with the results of a set of experiments on the same problem, and has the task of analysing and summarizing these.” (Yates & Cochran, 1938)

Agricultural Research

- laid out methods that are still in use today (Cochran, 1937, 1943, 1954; Cochran & Carroll, 1953; Rao, Kaplan, & Cochran, 1981; Yates & Cochran, 1938)
- some core ideas:
  - estimates not equally precise (different variances)
  - compute weighted average of the estimates, with weights inversely proportional to the variances
  - estimates may be more variable than one would expect given their variances (→ heterogeneity)

Physics

- “Let us suppose that a given constant has been measured in several different ways, […] it seems quite legitimate to combine by means of least squares the various results, i.e., to weight them according to their stated probable errors, and to derive the probable error in the final weighted average by the usual formulas.” (Birge, 1932)

Glass (1976)

- based on his presidential address at the 1976 Annual Meeting of the AERA
- “Meta-analysis refers to the […] statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.”

Glass (1976)

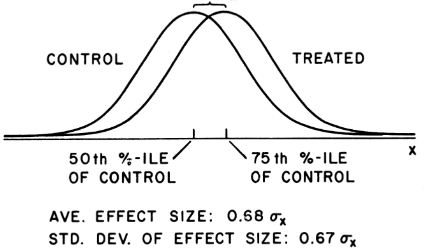


Figure 1. Normal curves illustrating the aggregate effect of psychotherapy in relation to untreated control groups. (Data based on 833 effect size measures from 375 studies, representing about 40,000 treated and untreated subjects.)

Glass (1976)

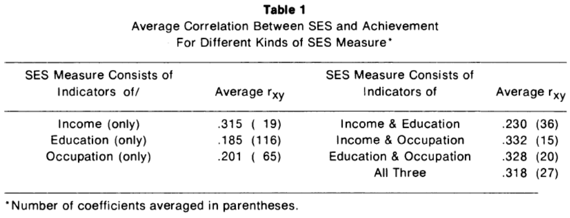


Table 1
Average Correlation Between SES and Achievement
For Different Kinds of SES Measure*

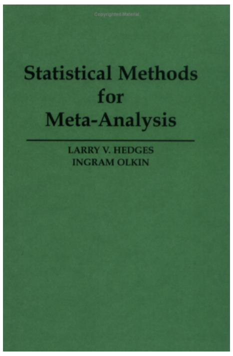
SES Measure Consists of Indicators of/	Average r_{xy}	SES Measure Consists of Indicators of	Average r_{xy}
Income (only)	.315 (19)	Income & Education	.230 (36)
Education (only)	.185 (116)	Income & Occupation	.332 (15)
Occupation (only)	.201 (65)	Education & Occupation	.328 (20)
		All Three	.318 (27)

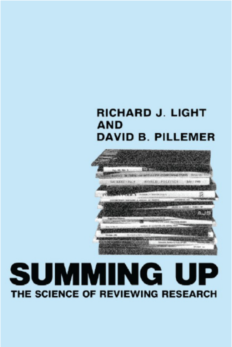
*Number of coefficients averaged in parentheses.

Early Meta-Analyses

- effects of interpersonal expectations on behavior (Rosenthal & Rubin, 1978)
- relation between class size and academic achievement (Glass & Smith, 1979)
- differential validity of employment tests for Black and White workers (Hunter et al., 1979)

Some Early Books (1985/1984)




RICHARD J. LIGHT AND DAVID B. PILLEMER
SUMMING UP
THE SCIENCE OF REVIEWING RESEARCH

But there were critics …

- “mega-silliness” (Eysenck, 1978)
- “meta-analysis/shmeta-analysis” (Shapiro, 1994)
- “statistical alchemy for the 21st century” (Feinstein, 1995)

Systematic Reviews

- research synthesis as a scientific process
- based on replicable and systematic methods that are meant to “limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic” (Last, 2001)
- methods should be made explicit
- not necessarily supported by quantitative methods (can also use qualitative methods)

Cochrane and the EBM Movement

- Cochrane (1972). *Effectiveness and efficiency: Random reflections on health services*.
- advocated the use of RCTs to form the evidence base for clinical decision making

THE COCHRANE COLLABORATION®
http://www.cochrane.org

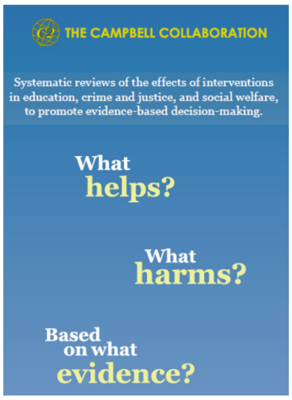
The Campbell Collaboration
http://www.campbellcollaboration.org

THE COCHRANE LIBRARY
Independent high-quality evidence for health care decision making

How do you know if one healthcare intervention works better than another, or if it will do more harm than good?

The Cochrane Library enables those involved with healthcare decisions to keep up-to-date with all the latest evidence

Go to www.thecochranelibrary.com to discover this essential resource today



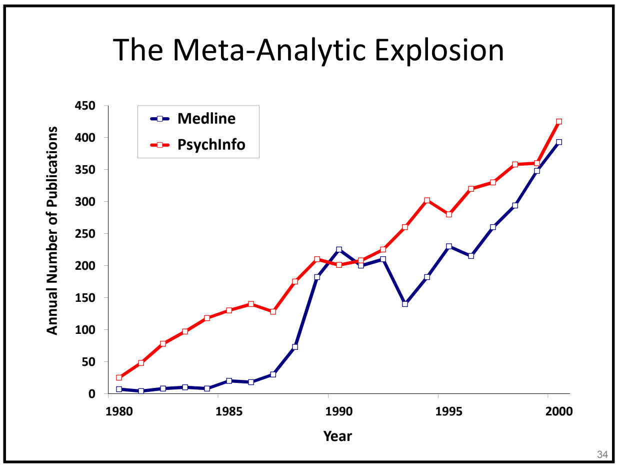
THE CAMPBELL COLLABORATION

Systematic reviews of the effects of interventions in education, crime and justice, and social welfare, to promote evidence-based decision-making.

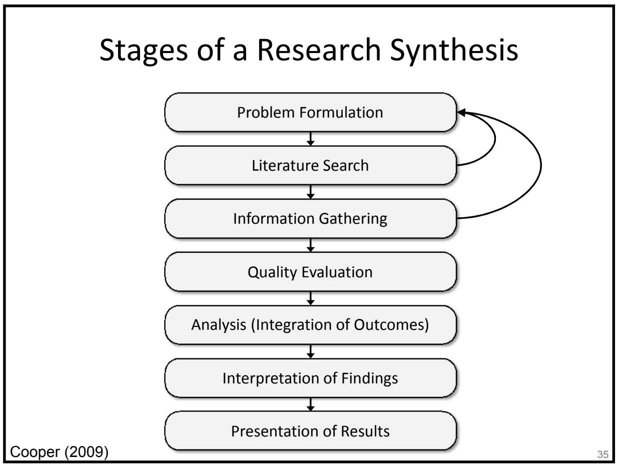
What helps?

What harms?

Based on what evidence?

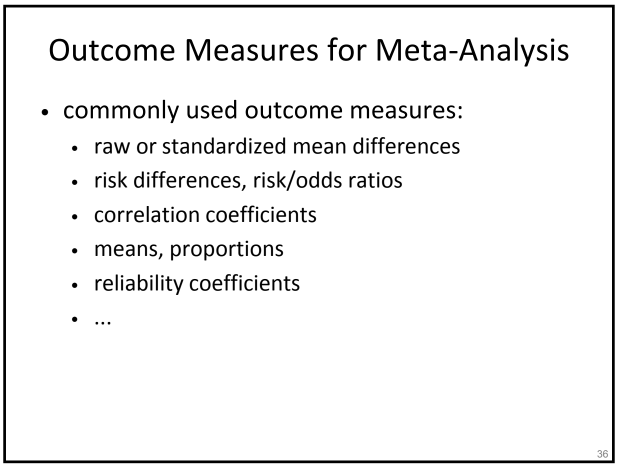


The Meta-Analytic Explosion



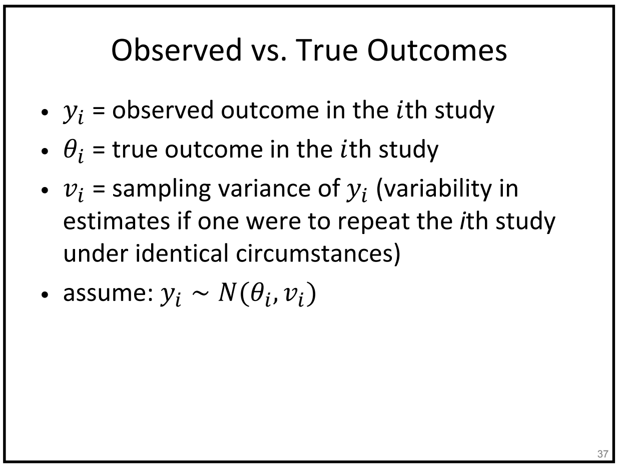
Stages of a Research Synthesis

Cooper (2009)



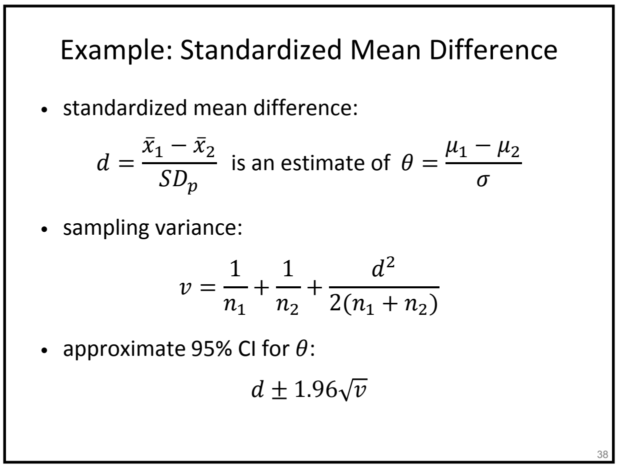
Outcome Measures for Meta-Analysis

- commonly used outcome measures:
  - raw or standardized mean differences
  - risk differences, risk/odds ratios
  - correlation coefficients
  - means, proportions
  - reliability coefficients
  - …



Observed vs. True Outcomes

- y_i = observed outcome in the ith study
- θ_i = true outcome in the ith study
- v_i = sampling variance of y_i (variability in estimates if one were to repeat the ith study under identical circumstances)
- assume: $y_i \sim N(\theta_i, v_i)$



Example: Standardized Mean Difference

- standardized mean difference:

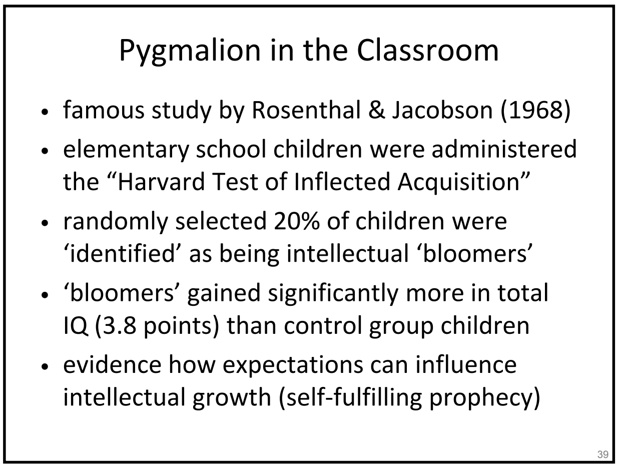
$$d = \frac{\bar{x}_1 - \bar{x}_2}{SD_p} \text{ is an estimate of } \theta = \frac{\mu_1 - \mu_2}{\sigma}$$

- sampling variance:

$$v = \frac{1}{n_1} + \frac{1}{n_2} + \frac{d^2}{2(n_1 + n_2)}$$

- approximate 95% CI for θ:

$$d \pm 1.96\sqrt{v}$$



Pygmalion in the Classroom

- famous study by Rosenthal & Jacobson (1968)
- elementary school children were administered the “Harvard Test of Inflected Acquisition”
- randomly selected 20% of children were ‘identified’ as being intellectual ‘bloomers’
- ‘bloomers’ gained significantly more in total IQ (3.8 points) than control group children
- evidence how expectations can influence intellectual growth (self-fulfilling prophecy)

Pygmalion in the Classroom

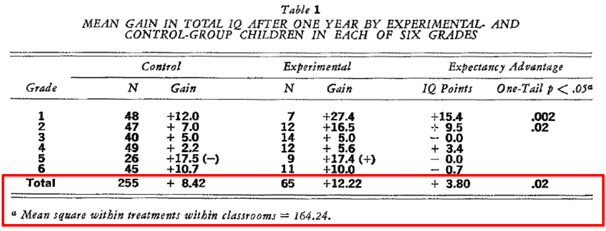


Table **1**

MEAN GAIN IN TOTAL IQ AFTER ONE YEAR BY EXPERIMENTAL- AND CONTROL-GROUP CHILDREN IN EACH OF SIX GRADES

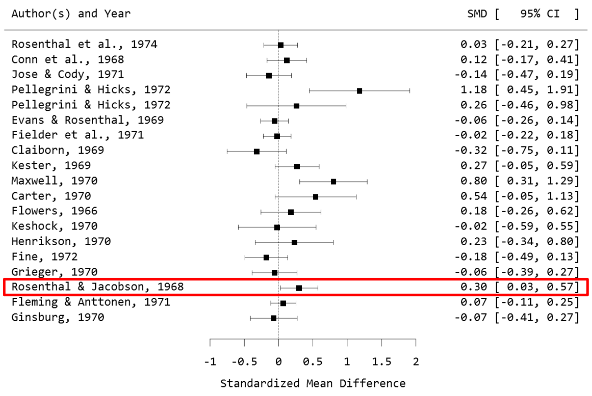
	Control		*Experimental*		*Expectancy Advantage*	
Grade	*N*	*Gain*	*N*	*Gain*	*IQ Points*	*One-Tail p < .05*[a]
1	48	+12.0	7	+27.4	+15.4	.002
2	47	+ 7.0	12	+16.5	+ 9.5	.02
3	40	+ 5.0	14	+ 5.0	− 0.0	
4	49	+ 2.2	12	+ 5.6	+ 3.4	
5	26	+17.5 (−)	9	+17.4 (+)	− 0.0	
6	45	+10.7	11	+10.0	− 0.7	
Total	255	+ 8.42	65	+12.22	+ 3.80	.02

[a] *Mean square within treatments within classrooms = 164.24.*

$$d = \frac{12.22 - 8.42}{\sqrt{164.24}} = 0.30 \qquad v = \frac{1}{255} + \frac{1}{65} + \frac{.30^2}{2(65 + 255)} = 0.019$$

Conflicting Results

- the study and its finding was readily accepted by some and harshly critiqued by others
- 14 years of additional research produced conflicting results
- Raudenbush (1984) conducted a meta-analysis of the existing evidence



Author(s) and Year	SMD [95% CI]
Rosenthal et al., 1974	0.03 [-0.21, 0.27]
Conn et al., 1968	0.12 [-0.17, 0.41]
Jose & Cody, 1971	-0.14 [-0.47, 0.19]
Pellegrini & Hicks, 1972	1.18 [0.45, 1.91]
Pellegrini & Hicks, 1972	0.26 [-0.46, 0.98]
Evans & Rosenthal, 1969	-0.06 [-0.26, 0.14]
Fielder et al., 1971	-0.02 [-0.22, 0.18]
Claiborn, 1969	-0.32 [-0.75, 0.11]
Kester, 1969	0.27 [-0.05, 0.59]
Maxwell, 1970	0.80 [0.31, 1.29]
Carter, 1970	0.54 [-0.05, 1.13]
Flowers, 1966	0.18 [-0.26, 0.62]
Keshock, 1970	-0.02 [-0.59, 0.55]
Henrikson, 1970	0.23 [-0.34, 0.80]
Fine, 1972	-0.18 [-0.49, 0.13]
Grieger, 1970	-0.06 [-0.39, 0.27]
Rosenthal & Jacobson, 1968	0.30 [0.03, 0.57]
Fleming & Anttonen, 1971	0.07 [-0.11, 0.25]
Ginsburg, 1970	-0.07 [-0.41, 0.27]

Equal-Effects Model

- assume $\theta_1 = \cdots = \theta_k \equiv \theta$
- then $y_i \sim N(\theta, v_i)$
- estimate θ with:

$$\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i}$$

where $w_i = 1/v_i$

Random-Effects Model

- but often true effects are not homogeneous (so-called Q-test can be used to test this)
- assume $\theta_i \sim N(\mu, \tau^2)$
- then $y_i \sim N(\mu, v_i + \tau^2)$
- estimate τ^2 and then μ with:

$$\hat{\mu} = \frac{\sum w_i y_i}{\sum w_i}$$

where $w_i = 1/(v_i + \hat{\tau}^2)$

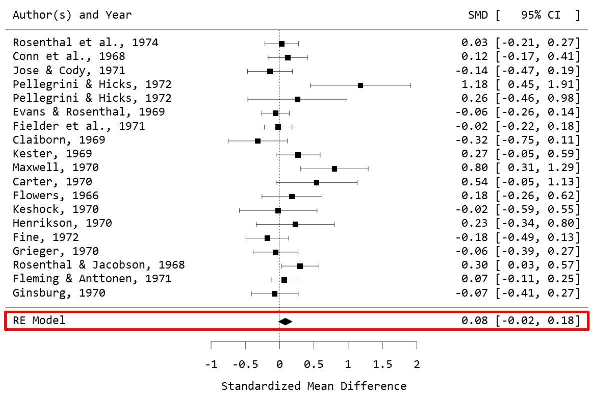
Results

- $Q = 35.83, df = 18, p = .007$
- reject homogeneity assumption
- we find:

$$\hat{\tau}^2 = 0.02$$
$$\hat{\mu} = 0.08$$

- 95% CI for μ:

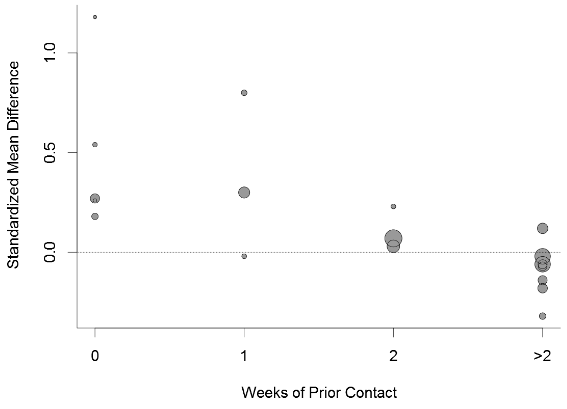
$$(-0.02, 0.18)$$



Author(s) and Year	SMD [95% CI]
Rosenthal et al., 1974	0.03 [-0.21, 0.27]
Conn et al., 1968	0.12 [-0.17, 0.41]
Jose & Cody, 1971	-0.14 [-0.47, 0.19]
Pellegrini & Hicks, 1972	1.18 [0.45, 1.91]
Pellegrini & Hicks, 1972	0.26 [-0.46, 0.98]
Evans & Rosenthal, 1969	-0.06 [-0.26, 0.14]
Fielder et al., 1971	-0.02 [-0.22, 0.18]
Claiborn, 1969	-0.32 [-0.75, 0.11]
Kester, 1969	0.27 [-0.05, 0.59]
Maxwell, 1970	0.80 [0.31, 1.29]
Carter, 1970	0.54 [-0.05, 1.13]
Flowers, 1966	0.18 [-0.26, 0.62]
Keshock, 1970	-0.02 [-0.59, 0.55]
Henrikson, 1970	0.23 [-0.34, 0.80]
Fine, 1972	-0.18 [-0.49, 0.13]
Grieger, 1970	-0.06 [-0.39, 0.27]
Rosenthal & Jacobson, 1968	0.30 [0.03, 0.57]
Fleming & Anttonen, 1971	0.07 [-0.11, 0.25]
Ginsburg, 1970	-0.07 [-0.41, 0.27]
RE Model	0.08 [-0.02, 0.18]

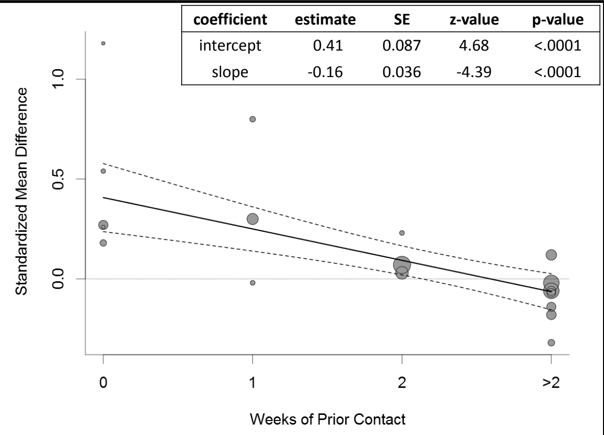
But that may not be the whole story …

- two phases to such experiments
  - first phase: induce expectation
  - second phase: test the expectancy hypothesis
- timing of first phase may be crucial
- if teachers had contact with children prior to expectancy induction, may not have an effect



Meta-Regression

- extension that allows inclusion of predictors (‘moderators’) in the models
- most general: mixed-effects meta-regression
- assume $\theta_i \sim N(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \tau^2)$
- estimate τ^2 (‘residual heterogeneity’) and the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$



coefficient	estimate	SE	z-value	p-value
intercept	0.41	0.087	4.68	<.0001
slope	-0.16	0.036	-4.39	<.0001

Some Misconceptions

- meta-analysis is objective
- need lots of studies for a meta-analysis
- meta-analysis is for synthesizing group differences or correlation coefficients
- a meta-analysis is a good first-year project for PhD students … well, maybe …

Some Recent Developments

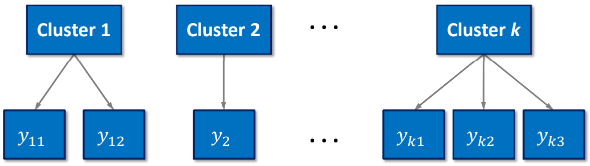
- better methods for inference
- quantification of heterogeneity
- methods for specific types of data
- publication bias
- multilevel/multivariate models
- network meta-analysis

Publication Bias

- affects all review methods (not a problem specific to meta-analysis!)
- in fact, due to meta-analysis:
  - increased awareness of publication bias
  - development of systematic methods to detect and address publication bias
  - increased emphasis on the importance of trial registries and pre-registration

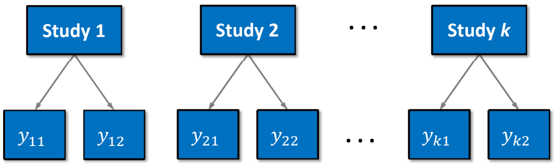
Multilevel Meta-Analytic Data

- multilevel structures can arise when we have multiple estimates for some higher clustering variable (paper, lab, research group, …)



Multiple (Correlated) Outcomes

- multivariate data also arise when multiple outcomes are measured within the studies



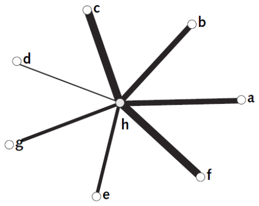
note: *not all studies have to measure all outcomes*

Network Meta-Analysis

- mixed treatment comparisons meta-analysis
- esp. relevant for examining treatment effects
- often there are multiple treatments available for the same condition/disease
- studies comparing the effectiveness of these treatments form a network of comparisons

Star-Shaped Networks

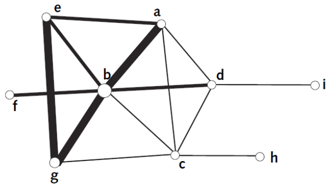
Second-generation antiepileptic drugs in partial epilepsy



a: levetiracetam, b: gabapentin, c: lamotrigine, d: oxcarbazepine, e: tiagabine, f: topiramate, g: zonisamide, h: placebo

Complex Networks

Chemotherapy regimens for ovarian cancer



a: platinum monotherapy, b: platinum-based combination,
c: taxane monotherapy, d: platinum + taxane-based combination,
e: nonplatinum/nontaxane monotherapy,
f: platinum-based combination (ip), g: nonplatinum/nontaxane combination, h: taxane-based combination,
i: platinum/taxane-based combination (ip)

Goals of a Network Meta-Analysis

- synthesize evidence provided by all studies and treatment comparisons in one model
- obtain indirect evidence about comparisons that have not been examined head-to-head
- establish hierarchy of treatment effectiveness
- ...

Multilevel/Multivariate/Network MA

- analysis conducted with more complex mixed-effects models (e.g., Berkey et al., 1998; Konstantopoulos, 2011; Lu & Ades, 2004; Lumley, 2002; Salanti et al., 2008; Senn et al., 2013; van Houwelingen et al., 2002)
- Bayesian methods popular for network MA

References

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. Statistics in Medicine, 17, 2537-2550.

Birge, R. T. (1932). The calculation of errors by the method of least squares. Physical Review, 40, 207-227.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. Evaluation and the Health Professions, 25, 12-37.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. Supplement to the Journal of the Royal Statistical Society, 4, 102-118.

Cochran, W. G. (1943). The comparison of different scales of measurement for experimental results. Annals of Mathematical Statistics, 14, 205-216.

Cochran, W. G. (1954). The combination of estimates from different experiments. Biometrics, 10, 101-129.

Cochran, W. G. (1954). Some methods for strengthening the common χ2 tests. Biometrics, 10, 417-451.

Cochran, W. G., & Carroll, S. P. (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. Biometrics, 9, 447-459.

Cochrane, A. (1972). Effectiveness and efficiency: Random reflections on health service. The Nuffield Provincial Hospital Trust.

Cooper, H. M. (2009). Research synthesis and meta-analysis: A step-by-step approach (4th ed.). Thousand Oaks, CA: Sage.

Eysenck, H. J. (1978). An exercise in mega-silliness. American Psychologist, 33, 517.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. Journal of Clinical Epidemiology, 48, 71-79.

References

Fisher, R. A. (1932). Statistical methods for research workers (4th ed.). London: Oliver and Boyd.

Garvey, W. D., & Griffith, B. C. (1971). Scientific communication: Its role in the conduct of research and creation of knowledge. American Psychologist, 26, 349-362.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, 3-8.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of the relationship between class size and achievement. Educational Evaluation and Policy Analysis, 1, 2-16.

Good, I. J. (1955). On the weighted combination of significance tests. Journal of the Royal Statistical Society, Series B, 17, 264-265.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 86, 721-735.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. Research Synthesis Methods, 2, 61-76.

Lancaster, H. O. (1961). The combination of probabilities: An application of orthonormal functions. Australian Journal of Statistics, 3, 20-33.

Last, J. M. (2001). A dictionary of epidemiology. Oxford: Oxford University Press.

Light. R. J., & Pillemer, D. B. (1984). Summing up: The science of reviewing research. Cambridge, MA: Harvard University Press.

Lipták, T. (1958). On the combination of independent tests. A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei, 3, 171-197.

References

Lu, G., & Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine, 23, 3105-3124.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. Statistics in Medicine, 21, 2313-2324.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. Statistics in Medicine, 21, 2313-2324.

Mosteller, F., & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), Handbook of social psychology (Volume 1): Theory and methods (Vol. 1, pp. 289-334). Cambridge, MA: Addison-Wesley.

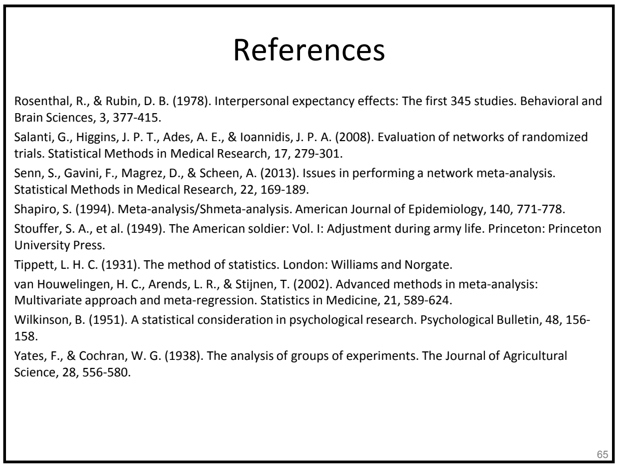
Pearson, K. (1904). Report on certain enteric fever inoculation statistics. British Medical Journal, 2, 1243-1246.

Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. Biometrika, 25, 379-410.

Rao, P. S. R. S., Kaplan, J., & Cochran, W. G. (1981). Estimators for the one-way random effects model with unequal error variances. Journal of the American Statistical Association, 76, 89-97.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. Journal of Educational Psychology, 76, 85-97.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. New York: Holt, Rinehart & Winston.



References

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. Behavioral and Brain Sciences, 3, 377-415.

Salanti, G., Higgins, J. P. T., Ades, A. E., & Ioannidis, J. P. A. (2008). Evaluation of networks of randomized trials. Statistical Methods in Medical Research, 17, 279-301.

Senn, S., Gavini, F., Magrez, D., & Scheen, A. (2013). Issues in performing a network meta-analysis. Statistical Methods in Medical Research, 22, 169-189.

Shapiro, S. (1994). Meta-analysis/Shmeta-analysis. American Journal of Epidemiology, 140, 771-778.

Stouffer, S. A., et al. (1949). The American soldier: Vol. I: Adjustment during army life. Princeton: Princeton University Press.

Tippett, L. H. C. (1931). The method of statistics. London: Williams and Norgate.

van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. Statistics in Medicine, 21, 589-624.

Wilkinson, B. (1951). A statistical consideration in psychological research. Psychological Bulletin, 48, 156-158.

Yates, F., & Cochran, W. G. (1938). The analysis of groups of experiments. The Journal of Agricultural Science, 28, 556-580.

Thank You!

Questions?