HOW TO READ, INTERPRET, AND APPLY THE RESULTS OF A META-ANALYSIS

A VERY BRIEF PRIMER

Kelly M. Babchishin

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

Carleton University

1

Upon completion of this educational activity learners should

1. be better able to discuss the strengths and weaknesses of meta-analysis within the article.
2. be better be able to understand the how the results are combined to better have more confidence in applying the results.
3. have an increased ability to better understand the variables and outcomes from the meta-analysis.
4. know how to critically weigh all that evidence in order to settle on the best, most evidence-based answer

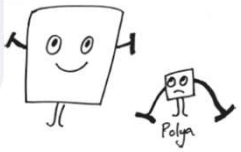
Carleton University

2

What is a Meta-Analysis?

- A statistical technique for combining individual effects reported in studies addressing the same research question

- Output:
  - Produces a summary effect
  - Measures the variability between the individual studies

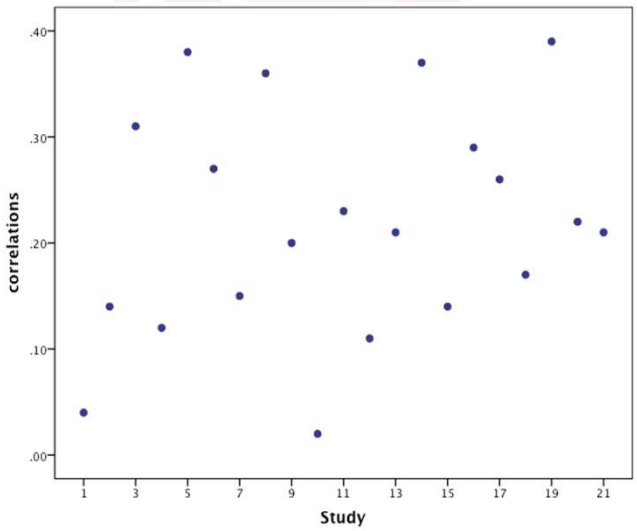


Why do we Need Meta-Analyses?

- Hypothetical research question:
  - Do professors grade more leniently as consumption of alcohol increases?

- Conventional lit review approach:
  - List all the studies on the topic, comment on differences between them

What would you conclude based on these studies?



Study	Correlation
1	.04
2	.14
3	.31*
4	.12
5	.38*
6	.27*
7	.15
8	.36*
9	.20
10	.02
11	.23
12	.11
13	.21
14	.37*
15	.14
16	.29*
17	.26*
18	.17
19	.39*
20	.22
21	.21

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

5

Carleton University

5

6

Why Should We Conduct Meta-Analyses?

- Objective test for consistency in findings
- Provides a more objective review compared to the narrative review
- More statistical power than a study
- Can look at moderator variables
- Useful tool to influence future research as well as inform policy
- Follows iterative approach to knowledge building

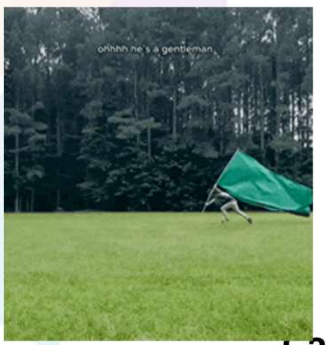
FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

7

Carleton University

7

- Lot of decisions rules
- Not always a 'correct' choice
- Can be coding intensive
- Conscientiousness is key

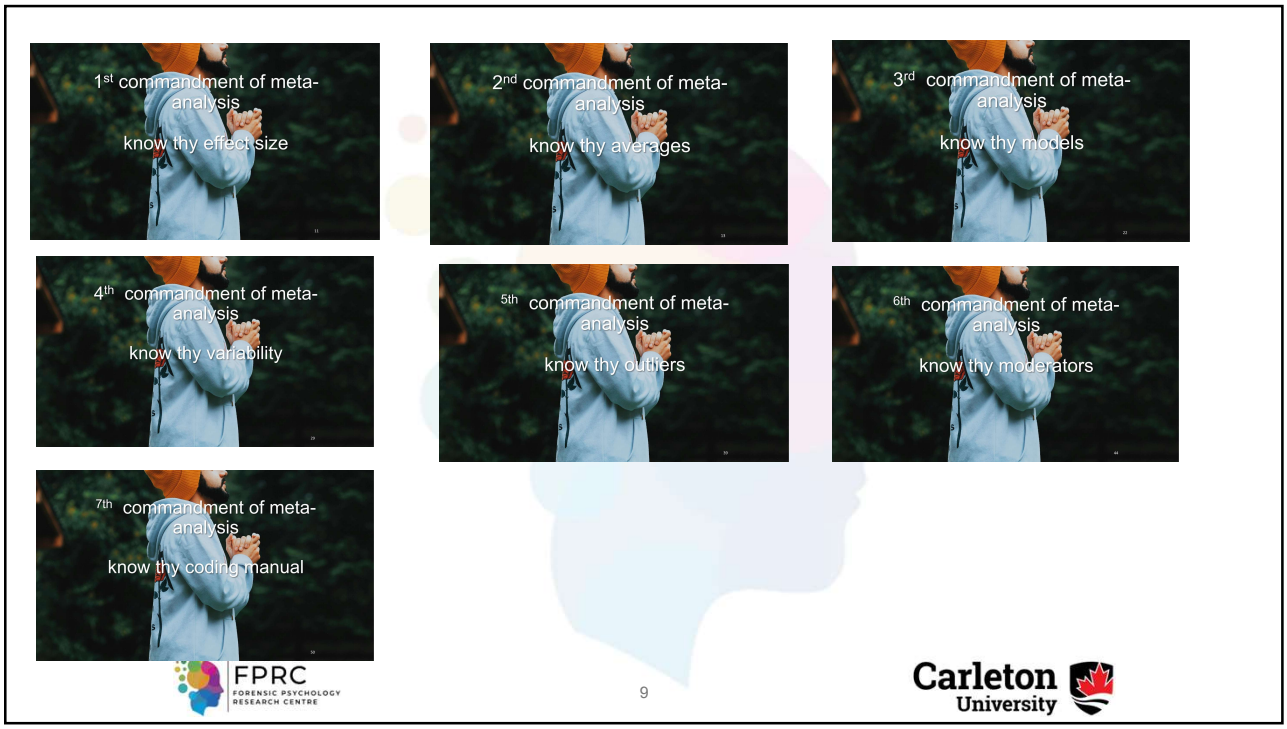


FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

8

Carleton University

8


1st commandment of meta-analysis
know thy effect size
2nd commandment of meta-analysis
know thy averages
3rd commandment of meta-analysis
know thy models
4th commandment of meta-analysis
know thy variability
5th commandment of meta-analysis
know thy outliers
6th commandment of meta-analysis
know thy moderators
7th commandment of meta-analysis
know thy coding manual
FPRC
FORENSIC PSYCHOLOGY RESEARCH CENTRE
9
Carleton University

9


1st commandment of meta-analysis
know thy effect size
10

10

Effect Size

- ES can also be an estimate of the value of ONE variable
  - What is the recidivism rate of men convicted of CSEM offences?
- But most describes MAGNITUDE of a relationship between TWO variables
- More complicated effect sizes: Relationship between 3+ variables…
- Should indicate ES (bonus if formula is provided)

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

11

Carleton University

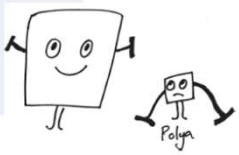
11

12

What is a Meta-Analysis?

- A statistical technique for combining individual effects reported in studies addressing the same research question

- Output:
  - Produces a summary effect
  - Measures the variability between the individual studies



Example 1:

- Relationship between the number of *Buffy the Vampire Slayer* TV episodes someone has watched & the number of Buffy references they make in a day

- 3 studies conducted
  r = .10
  r = .50
  r = .60

Simple Average

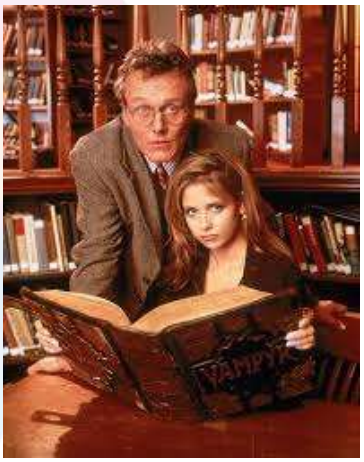
$$M_{ES} = \frac{\sum ES}{k}$$

= (0.10 + 0.50 + 0.60) / 3
= 1.20/3
= 0.40

The average effect size is 0.40

But What If……

- That third study is so darn good that we want to give it TWICE as much weight as the other studies?



Weighted Average

- Let's count Study 3 twice! (double the weight)

$$= \frac{.10 + .50 + .60 + .60}{4}$$

$$= \frac{1(.10) + 1(.50) + 2(.60)}{4}$$

$$= \frac{1.8}{4} = 0.45$$

Compared to unweighted average of 0.40

Big Scary Formula: Average Weighted Effect Size

- Sum of all the weighted effect sizes, divided by the sum of the weights

$$M_{ES} = \frac{\sum W_i \times ES_i}{\sum W_i}$$

$$= \frac{1(.10) + 1(.50) + 2(.60)}{4}$$

19

Weights

- Variance: Measure of the precision of the effect size (used to calculate confidence intervals)
- Weighted by the inverse of the variance
- The smaller the variance, the more weight it gets
- Key determinant of small variance: Large sample size
  - But not the only determinant!
  - **Each ES has its own variance calculation**
  - **Should report the formula**

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

20

Carleton University

20

21

Fixed-effect vs Random-effects Models

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

22

Carleton University

22

Differences

- Conceptual
  - Random-effects meant to generalize beyond the studies included
- Computational
  - Both use the "big scary formula"
  - Fixed-effect ignores differences in effect sizes across studies (assumes it's sampling error)
  - Random-effects incorporates those differences into the error term (T^2): **Treats the variability as real and meaningful**

23

Practical Differences

- None, when the variability is consistent across samples!
- As variability increases:
  - Random-effects confidence intervals widen (harder to get significance)
  - Random-effects starts to weight studies more and more similarly

24

Which do we choose?

- Conceptually, random-effects almost always the correct approach (and more conservative)
- BUT – less stable when there are less than 30 studies in the analysis

25

Value in Both

- Ideally, similar conclusions from both models
- When they differ, give it some thought
  - Random-effects: when variability is high, small studies get too much weight
  - Fixed-effect: really large studies get too much weight
  - Small number of studies: random-effects unstable

26

Which do we Pay Attention to?

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

Carleton University

27

28

Variability Across Studies

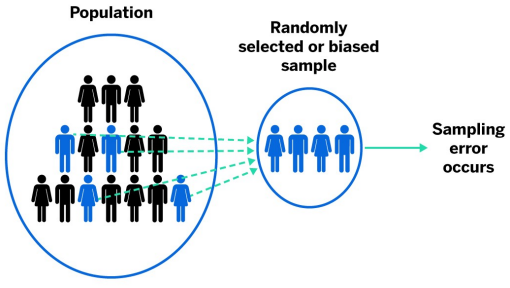
FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

Carleton University

29

Testing Variability Across Studies

- Are your effect sizes consistent across studies?
- Studies sampling the same true effect are not expected to produce the same effect size – Sampling error!



F FOR RES

Carleton University

30

Three Variability indicators

1. Q = Is there more variability than chance?
2. I^2 = How much variability?
3. Prediction intervals = What should we expect in a new study?

https://bmjopen.bmj.com/content/bmjopen/6/7/e010247.full.pdf

Cochrane's Q Statistic

- Null Hypothesis Significance Test (yes/no)
- Is the amount of variability across studies more than you would expect by chance?
- Adds a little something for each study to show how much that study differs from the average effect
  - More studies, larger Q value expected
  - Can't compare across analyses

I^2


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

- Effect size measure
- Percentage of overall variability that's beyond what you would expect by chance
- Heuristic: 25%, 50% and 75% are low, moderate, high variability (Higgins et al., 2003)

33

Q and I2 influenced by number of studies and size of studies

- With small (imprecise) studies, very different ES can yield an I^2 of 0/small Q
- I^2 /Q: with very large studies, even tiny between-study differences in effect size may result in a high I^2/large Q

34

Prediction intervals

- Estimates where the true effects are to be expected for 95% of similar (exchangeable) studies that might be conducted in the future
  - Based on the T^2
- Less dependent on *k* than I^2 /Q, but:
  - The PE will be imprecise if based on only a few studies and if these studies are small
    - T2 = akin to SD of the mean

35

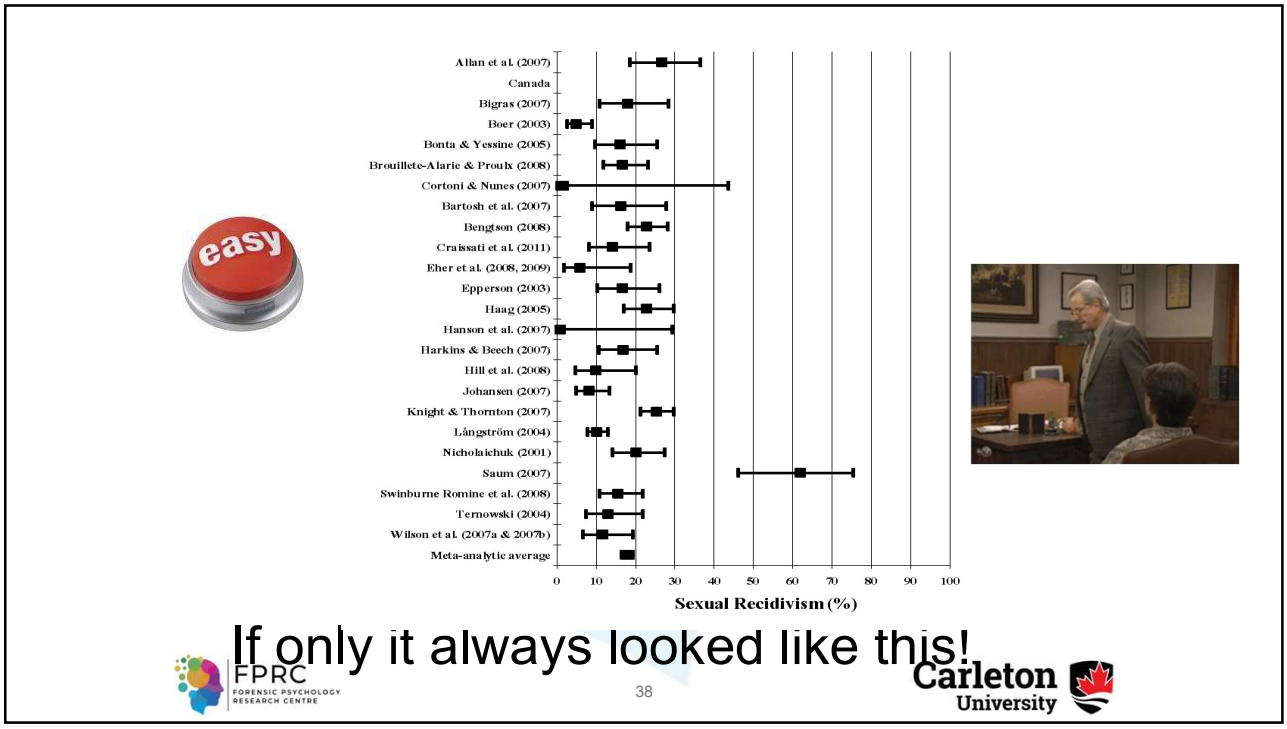
Testing Variability Across Studies

- Are your effect sizes consistent across studies?

36


5th commandment of meta-analysis
know thy outliers
37

37


easy
Allan et al. (2007)
Canada
Bigras (2007)
Boer (2003)
Bonta & Yessine (2005)
Brouillete-Alarie & Proulx (2008)
Cortoni & Nunes (2007)
Bartosh et al. (2007)
Bengtson (2008)
Craissati et al. (2011)
Eher et al. (2008, 2009)
Epperson (2003)
Haag (2005)
Hanson et al. (2007)
Harkins & Beech (2007)
Hill et al. (2008)
Johansen (2007)
Knight & Thornton (2007)
Långström (2004)
Nicholaichuk (2001)
Saum (2007)
Swinburne Romine et al. (2008)
Ternowski (2004)
Wilson et al. (2007a & 2007b)
Meta-analytic average
0 10 20 30 40 50 60 70 80 90 100
Sexual Recidivism (%)
If only it always looked like this!
FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE
38
Carleton University

38

Outliers

- General definition in research:
  - Extreme studies that may distort your results
  - Meaningful difference in findings if you include or exclude the outlier
- No solid conventions on how to identify outliers in meta-analysis
- Best practice: report outlier rules, and if present, present results with and without outlier

39

Extreme large sample size

- Outliers in different ways
- Extremely large sample will pull the meta average to its ES (think 1000 when all other studies are 50-100)
- Can reduce the weight to be 50% more than the next weight
- If large differences across Ns, authors should be discussing this

40

41

What if You See Lots of Variability?

- When
  - there is LOTS of variability in effect sizes
  - No single outlier
- We ask
  - What does this mean? Why is there so much variability?

- Meta: Does treatment reduce sexual recidivism?
  - Yes, but significant Q with large variability
  - What are some potential moderators?

FPRC FORENSIC PSYCHOLOGY RESEARCH CENTRE

42

Carleton University

42

Some Moderator Variables

- Methodological features
  - Year of study, country, published
- Features of sample
  - Average age, majority ethnicity, average prior offences, location, gender
- How IV is coded
  - Tx: Treatment length, treatment providers, treatment model, treatment quality
  - Predictors: coded from interview or file review?
- How DV is coded
  - Recidivism criteria/sources, follow-up length
- Study quality?
  - Recommended for treatment outcome research

What is the Measurement Property of Your Moderator?

- The type of moderator analysis is dependent on the measurement property of your moderators
  - Categorical moderators: *Between-level Q*
  - Continuous moderators: *Meta-regression*
  - *For each, you can use **fixed-effect** or random-effects*
- Should only pick moderators they have data for
  - If sample year, and there are three studies are 2018-2020 with one in 2000, should they be doing this moderator analysis?
  - Check descriptive tables

Publication bias

- Publication bias is one of the biggest threats to the validity of the meta-analysis

45

Solutions to Publication Bias

- Research/publication practices can reduce it
- Including unpublished studies in meta-analysis overcomes it
- Lots of different analyses that can be done
  - Moderator analysis (pub vs. not published)
  - Orwin's fail safe
  - Graphing

46

47

Coding manual

- Essential ingredient
  - Inclusion/exclusion criteria
  - Search strategy
  - Decision/coding rules
    - How to deal with overlapping samples
  - Appropriate effect sizes
- Lots of example coding manual on OSF
- OSF also includes searching strategies

FPRC Forensic Psychology Research Centre

48

Carleton University

48

Searching: Considerations

- Our experience: This is HARD and most meta-analyses not thorough
- Inclusion of unpublished literature?
- How many databases?
- Informal techniques (e.g., list-serves, contacting key authors)
- Keywords:
  - To identify the sample, the IV, the DV (or any combo)
  - Changes in language over time (e.g., person first)

49

Ex: People who commit sex offences

- sex* offen*, rape, rapist*, child molest*, pedophil*, paedophil*, exhibitionis*, sex* assault*, incest*, voyeur*, frotteur*, indecent exposure, sexual* devian*, paraphilia*, child porn*

- Sex* offen*

50

Inter-rater Reliability

- Good Quality
  - Minimum of 10 studies coded by 2 raters
  - Ideally strong inter-rater reliability
  - Moderators without good inter-rater: not particularly interpretable
- Better Quality
  - All studies double-coded
  - We promise: It's harder than it looks!

Principle of Independence of Observations

- Meta-analysis assumes each study it analyzes is from a **different** sample
- Including overlapping samples effectively counts it twice!
  - Reduces your variability
  - Narrows your confidence intervals
- Should discuss how they dealt with overlapping samples

Concluding Remarks

53

What Makes a Good Meta-Analysis? Considerations

- Clearly defined inclusion/exclusion criteria
- Comprehensive and well-documented search!
- Appropriate effect size (e.g., transformations if needed)
- Interrater reliability for variables coded
- Ideal: honoured independence of observations
- Analyze/discuss variability across findings
- Rationale for fixed vs random-effects
- Appropriate selection of moderators
- Appropriate analysis/interpretation of moderators
- Examined possible outliers or distorting effects

54

Limitations and Dangers of Meta-Analysis

- Not infallible!
  - Insufficient searching
  - “File drawer problem”
  - “Garbage in, garbage out”
  - Meta-analyzing too early
    - Can lead to unreliable conclusions
  - Coding manual missing or insufficient
  - Poor coding of effect sizes
  - Inappropriate analyses
  - Failed to consider moderators
    - Apples and oranges

55

Thank you.

Questions?

56