Evidence-based Decision Making: Session 9

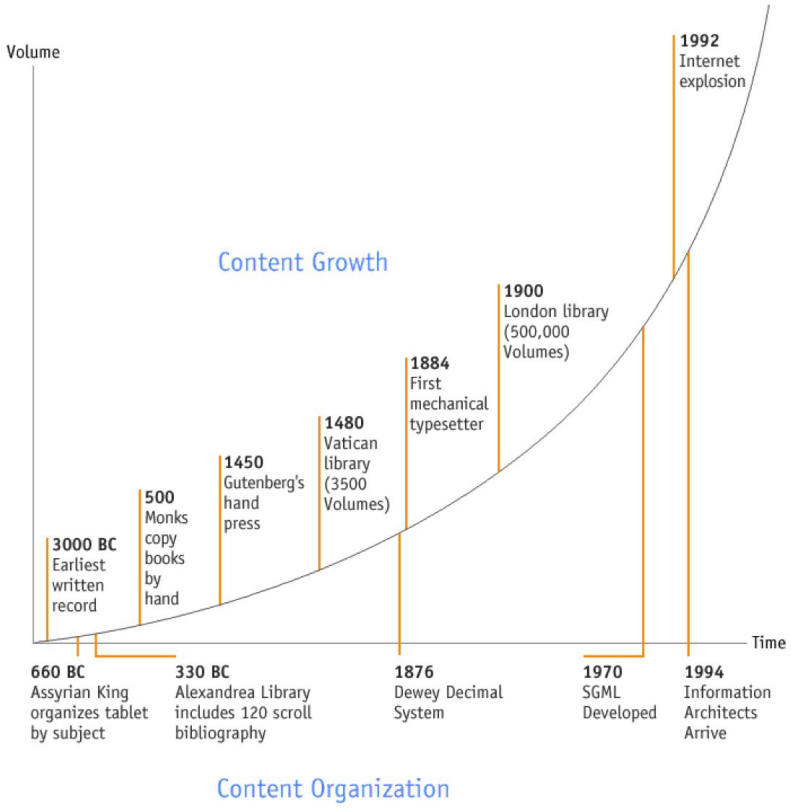
Rui Mata, FS 2021

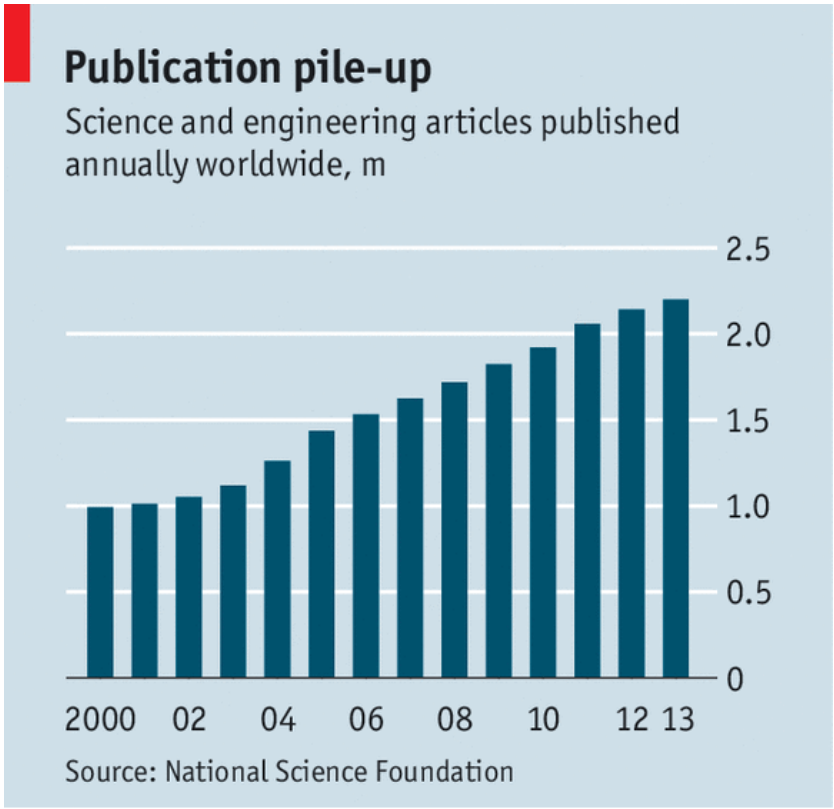
Goals

- understand the relevance of research synthesis and be able to sketch its brief history
- define key terms associated with research synthesis (i.e., systematic review, meta-analysis, protocol)
- consider limitations of research syntheses

Why research synthesis matters…

Information Explosion

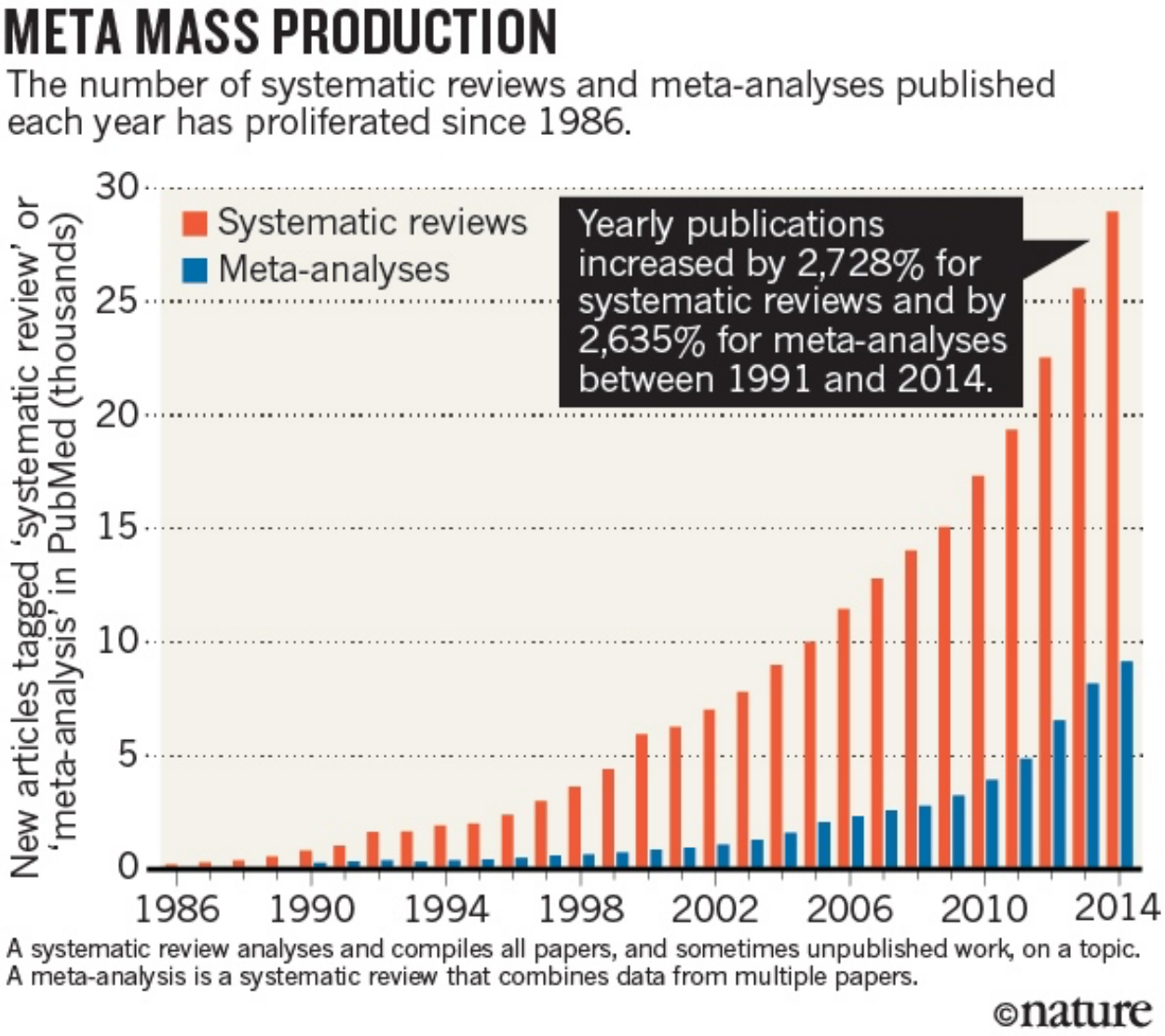




Economist.com

- rough estimates:
  - # of articles double every ~10 years
  - # of journals double every ~15 years

Why research synthesis matters…



https://www.nature.com/news/mass-production-of-review-articles-is-cause-for-concern-1.20617

A brief history of research synthesis

Pre-1970s

- narrative literature reviews
- vote counting methods
- some early forms of quantitative synthesis (medicine/vaccination: Pearson (1904); agriculture: Cochran (1937); physics: Birge (1932)

Post-1970s

- Origin of term “meta-analysis” (Glass, 1976)
- Textbooks: Hedges & Olkin (1985), Light & Pillemer (1984)
- EBM movement: Cochrane, Campbell
- Guidelines, guidelines, guidelines (CONSORT, PRISMA)…

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine, 100*(12), 579–582. http://doi.org/10.1258/jrsm.100.12.579

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. ***Evaluation & the Health Professions, 25***(1), 12–37.

A brief history of research synthesis

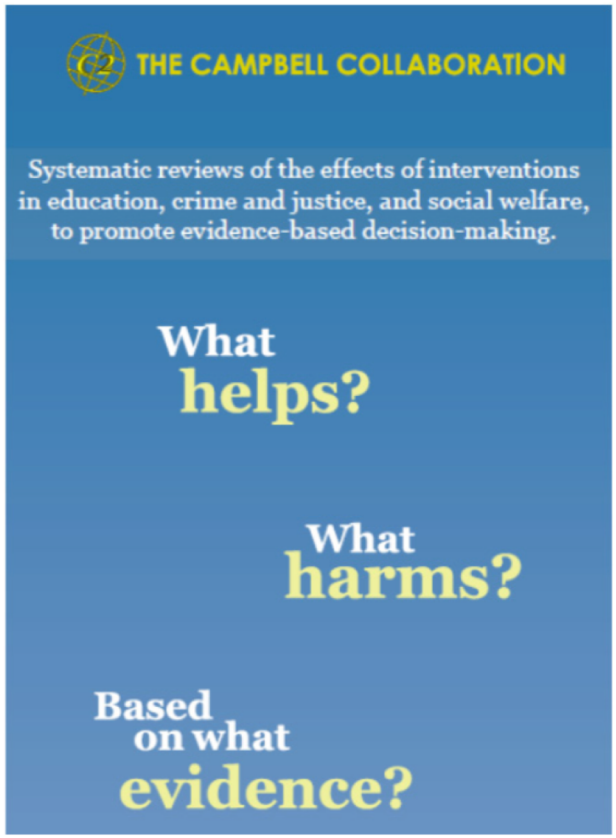
1996: QUOROM (QUality Of Reporting Of Meta-analyses);
see also (CONSORT, Consolidated Standards of Reporting Trials)

1999: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses)

2015: PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols)

A brief history of research synthesis

THE COCHRANE COLLABORATION



THE CAMPBELL COLLABORATION

www.cochrane.org

www.campbellcollaboration.org

1993

1999

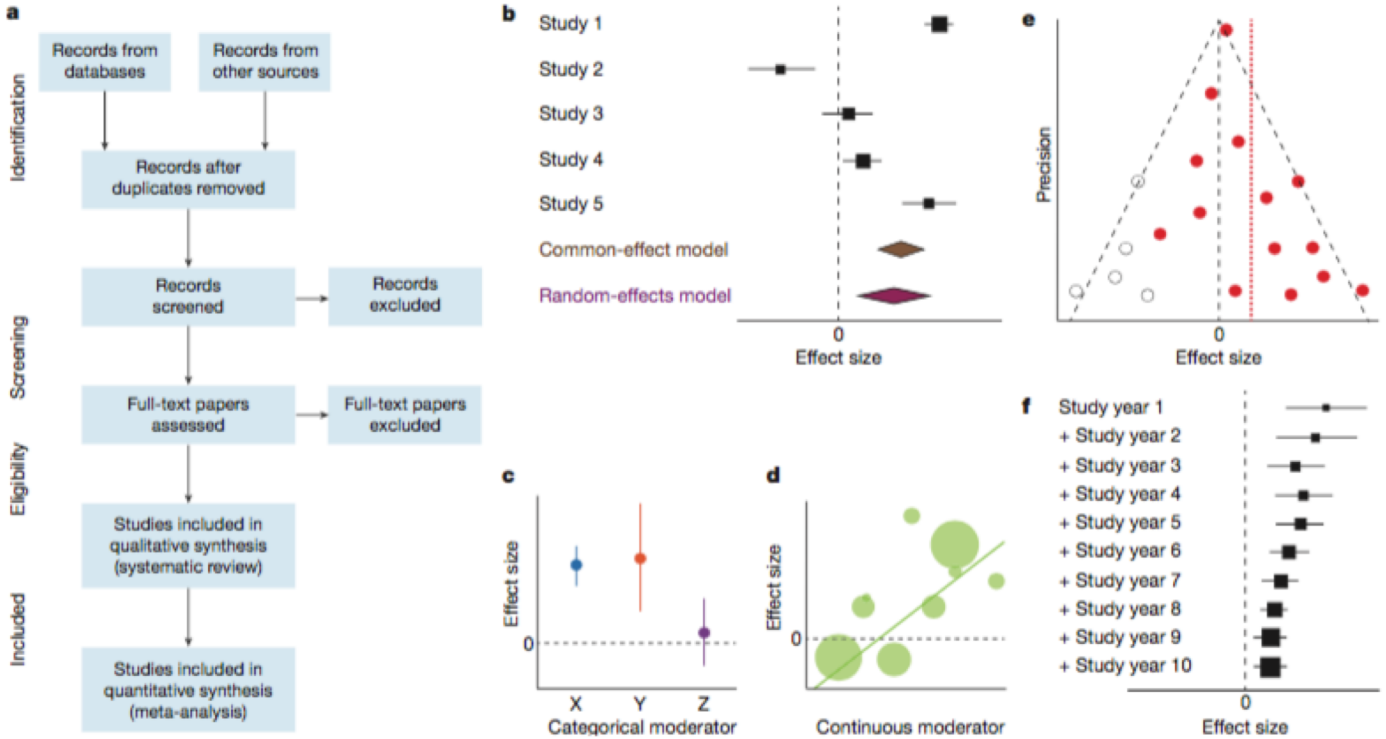
Definitions

Table 2 PRISMA-P terminology

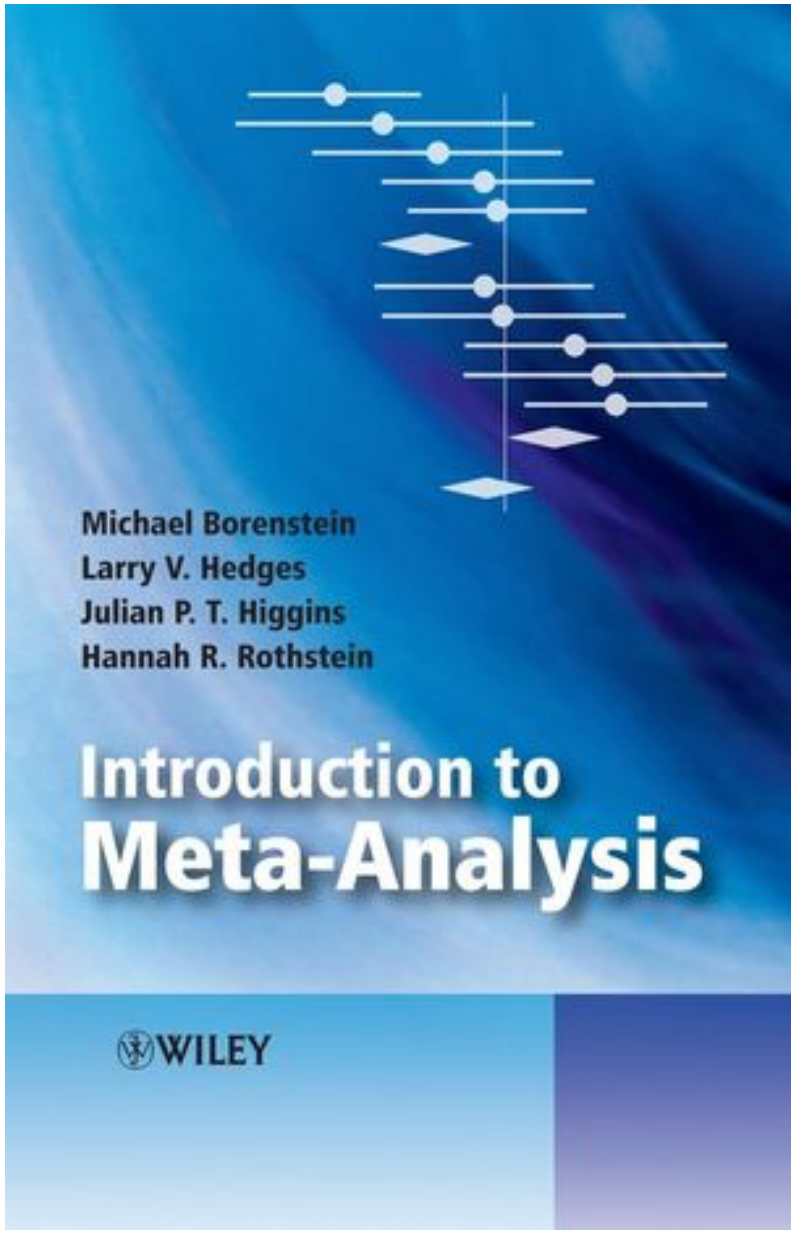
Term	Definition
Systematic review	A systematic review attempts to collate all relevant evidences that fits pre-specified eligibility criteria to answer a specific research question. It uses explicit, systematic methods to minimize bias in the identification, selection, synthesis, and summary of studies. When done well, this provides reliable findings from which conclusions can be drawn and decisions made [25,26]. The key characteristics of a systematic review are (a) a clearly stated set of objectives with an explicit, reproducible methodology; (b) a systematic search that attempts to identify all studies that would meet the eligibility criteria; (c) an assessment of the validity of the findings of the included studies (e.g., assessment of risk of bias and confidence in cumulative estimates); and (d) systematic presentation, and synthesis, of the characteristics and findings of the included studies
Meta-analysis	Meta-analysis is the use of statistical techniques to combine and summarize the results of multiple studies; they may or may be contained within a systematic review. By combining data from several studies, meta-analyses can provide more precise estimates of the effects of health care than those derived from the individual studies
Protocol	In the context of systematic reviews and meta-analyses, a protocol is a document that presents an explicit plan for a systematic review. The protocol details the rationale and *a priori* methodological and analytical approach of the review

PRISMA-P Group, Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. ***Systematic Reviews, 4(***1), e1000326–9. http://doi.org/10.1186/2046-4053-4-1

Happy 40th, meta-analysis!



Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. ***Nature, 555***, 175. http://doi.org/10.1038/nature25753


Michael Borenstein
Larry V. Hedges
Julian P. T. Higgins
Hannah R. Rothstein
Introduction to
Meta-Analysis
WILEY

Effect sizes

Effect size can mean different things (and be calculated in different ways), it can refer to either a treatment effect (e.g., the effect of drug vs. no drug on some outcome), or a single group summary (e.g., average correlation between two variables in a population), or a generic statistic (e.g., the average value of one variable in the population). The actual calculations to compute an effect size differ by type of data and study design.

Bornstein et al. (2009) give a roadmap of formulas and examples for different effect sizes as do others

Table 3.1 Roadmap of formulas in subsequent chapters.

Effect sizes based on means (Chapter 4)
Raw (unstandardized) mean difference (*D*)
Based on studies with independent groups
Based on studies with matched groups or pre-post designs
Standardized mean difference (*d* or *g*)
Based on studies with independent groups
Based on studies with matched groups or pre-post designs
Response ratios (*R*)
Based on studies with independent groups
Effect sizes based on binary data (Chapter 5)
Risk ratio (*RR*)
Based on studies with independent groups
Odds ratio (*OR*)
Based on studies with independent groups
Risk difference (*RD*)
Based on studies with independent groups
Effect sizes based on correlational data (Chapter 6)
Correlation (*r*)
Based on studies with one group

Meta-analysis (based on mean differences)

1 Calculate an effect size and its precision (variance) for each study

We can estimate the standardized mean difference (δ) from studies that used two independent groups as

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}. \quad (4.18)$$

In the numerator, $\bar{X}_1$ and $\bar{X}_2$ are the sample means in the two groups. In the denominator S_{within} is the within-groups standard deviation, pooled across groups,

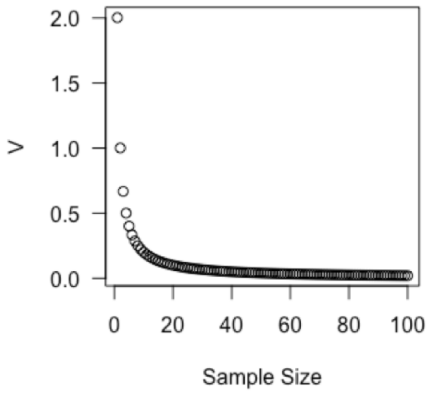
$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \quad (4.19)$$

where n_1 and n_2 are the sample sizes in the two groups, and S_1 and S_2 are the standard deviations in the two groups. The reason that we pool the two sample estimates of the standard deviation is that even if we assume that the underlying population standard deviations are the same (that is $\sigma_1 = \sigma_2 = \sigma$), it is unlikely that the sample estimates S_1 and S_2 will be identical. By pooling the two estimates of the standard deviation, we obtain a more accurate estimate of their common value.

The variance of d is given (to a very good approximation) by

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}. \quad (4.20)$$

In this equation the first term on the right of the equals sign reflects uncertainty in the estimate of the mean difference (the numerator in (4.18)), and the second reflects uncertainty in the estimate of S_{within} (the denominator in (4.18)).



The effect size will often be a standardised value that represents the magnitude of the effect; the variance of the effect size captures the precision of the estimate and will be largely a function of the sample size (see figure)

Bornstein et al. (2009, Chapter 4)

Meta-analysis

2 Calculate a weighted average of the effect sizes across studies

In its simplest form, the weight is a function of the precision (variance) associated with each study

$$W_i = \frac{1}{V_{Y_i}},$$

The overall effect size across studies is obtained by averaging the studies in a weighted form

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}, \quad (11.3)$$

that is, the sum of the products $W_i Y_i$ (effect size multiplied by weight) divided by the sum of the weights.

The variance of the summary effect is estimated as the reciprocal of the sum of the weights, or

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i}, \quad (11.4)$$

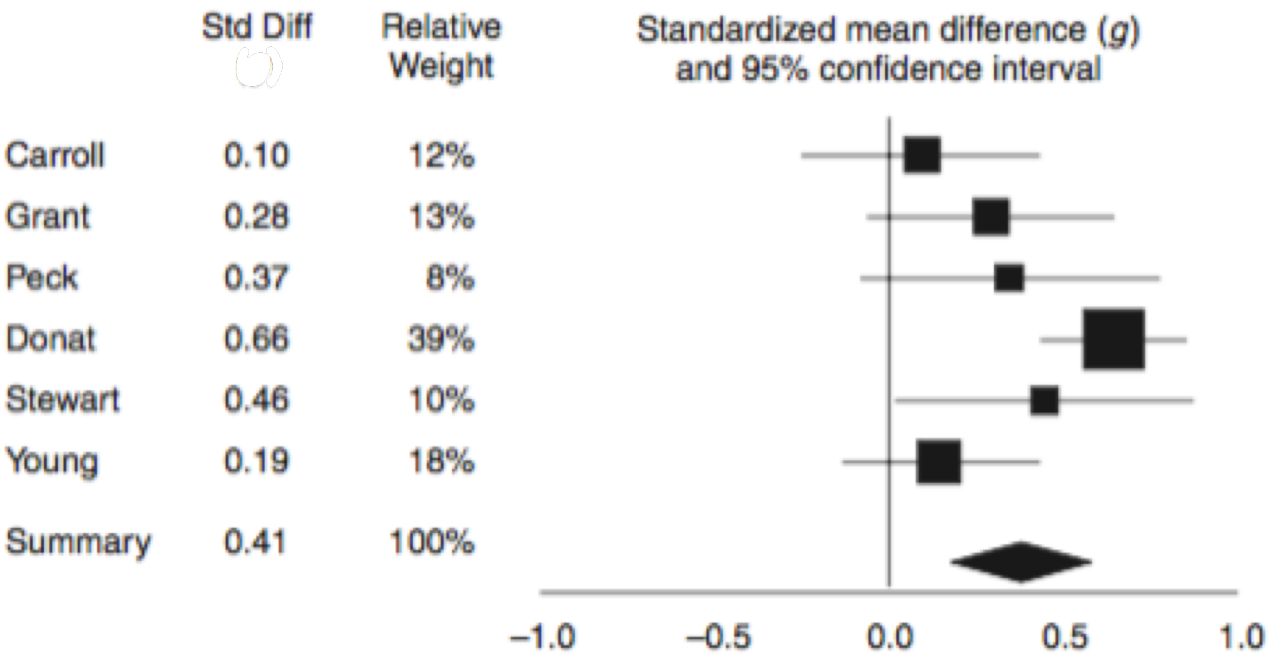
and the estimated standard error of the summary effect is then the square root of the variance,

$$SE_M = \sqrt{V_M}. \quad (11.5)$$

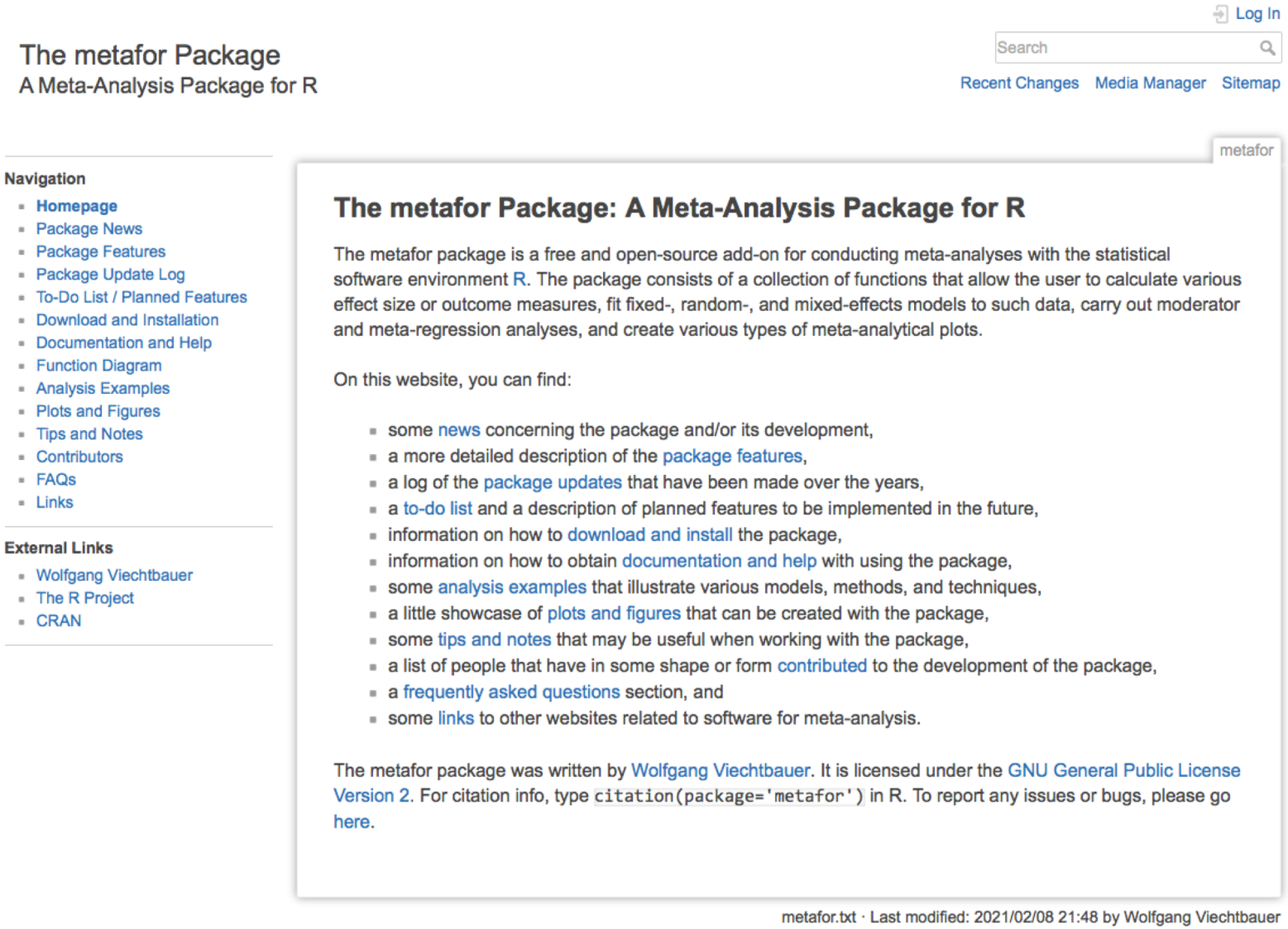
There are (slightly) more complex ways of aggregating studies that consider not only each study's precision but also between-study variance but the logic of weighted aggregation is the same.

Bornstein et al. (2009, Chapter 11)

Meta-analysis



Meta-analysis in practice…



http://www.metafor-project.org/

A long history of critique of research synthesis

- Garbage-in, garbage out ("mega-silliness", Eysenck, 1978)

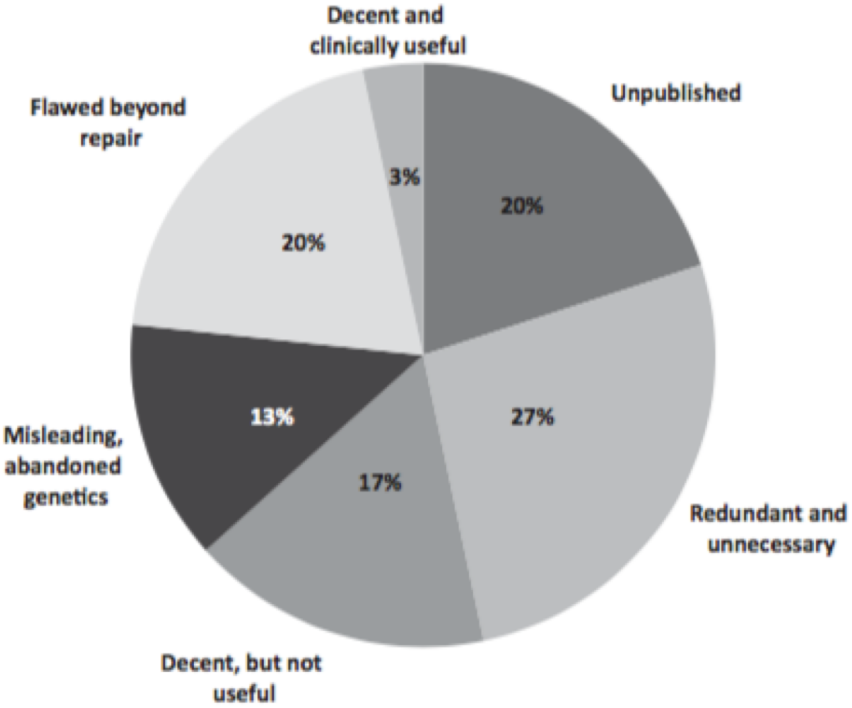
- "Instead of promoting evidence-based medicine and health care, these instruments often serve mostly as easily produced publishable units or marketing tools" (Ioannidis, 2016)

A long history of critique of research synthesis

Policy Points:

- Currently, there is massive production of unnecessary, misleading, and conflicted systematic reviews and meta-analyses. Instead of promoting evidence-based medicine and health care, these instruments often serve mostly as easily produced publishable units or marketing tools.
- Suboptimal systematic reviews and meta-analyses can be harmful given the major prestige and influence these types of studies have acquired.
- The publication of systematic reviews and meta-analyses should be realigned to remove biases and vested interests and to integrate them better with the primary production of evidence.

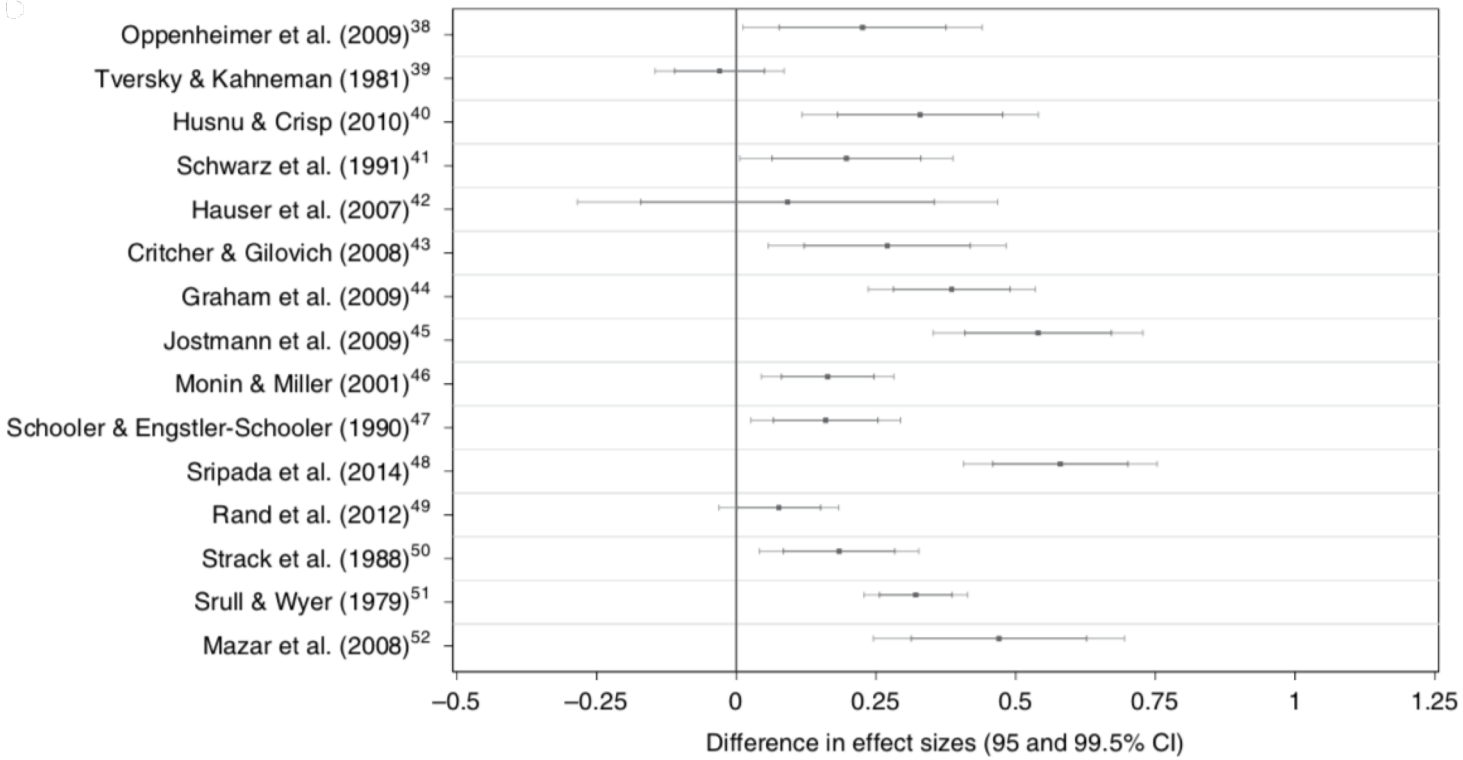
Figure 4. A Summary Overview of Currently Produced Meta-analyses



Ioannidis, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly, 94*(3), 485–514. http://doi.org/10.1111/1468-0009.12210

Limitations of Research Synthesis

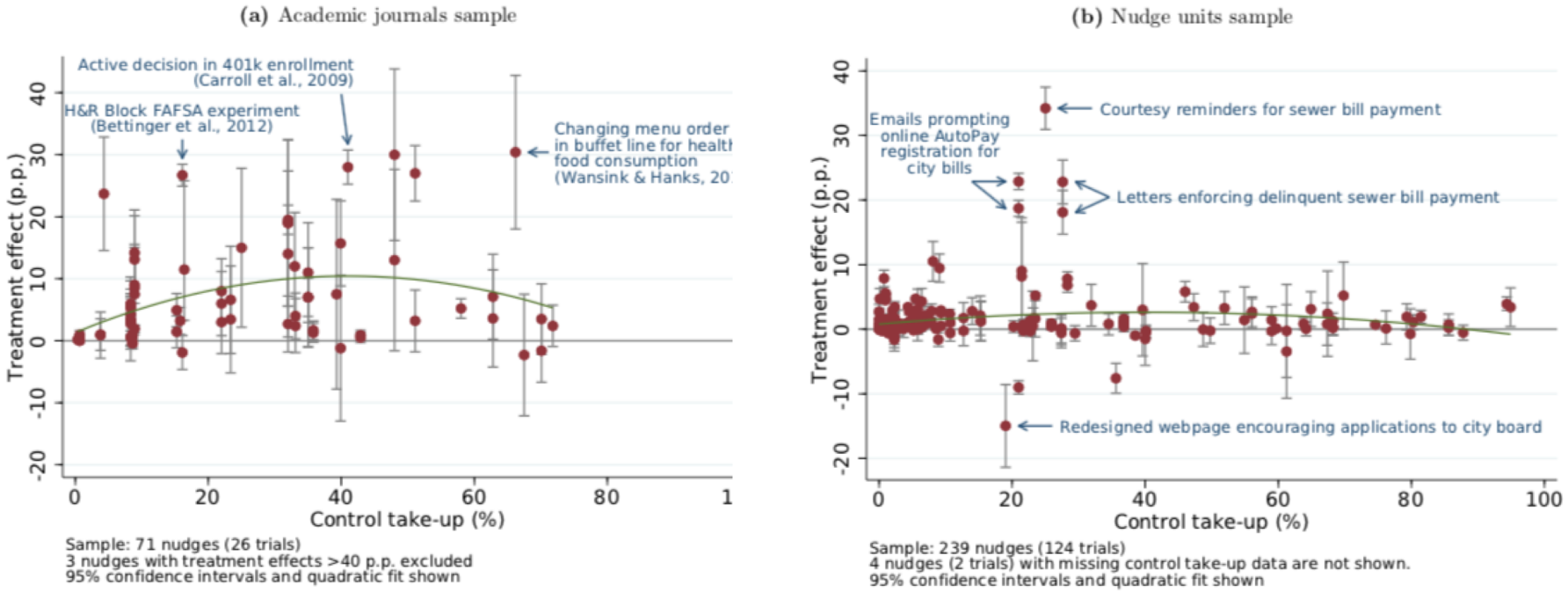
Difference between meta-analytic and replication studies effect size for 15 studies



"We find that meta-analytic effect sizes are significantly different from replication effect sizes for 12 out of the 15 meta-replication pairs. These differences are systematic and, on average, meta-analytic effect sizes are almost three times as large as replication effect sizes. We also implement three methods of correcting meta-analysis for bias, but these methods do not substantively improve the meta-analytic results." - these findings suggest that meta-analysis may not be able to adjust inflated effect sizes that arise from publication bias/selective reporting.

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. ***Nature Human Behaviour***, *4*(4), 423–434. http://doi.org/10.1038/s41562-019-0787-z

Limitations of Research Synthesis

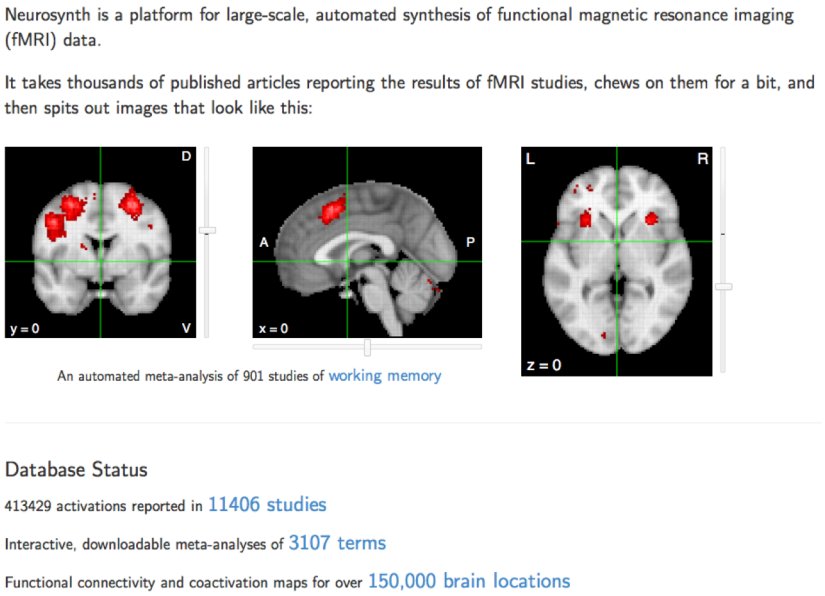


We assemble a unique data set including all trials run by two of the largest Nudge Units in the United States, including 126 RCTs covering over 23 million individuals. We compare these trials to a separate sample of nudge trials published in academic journals from two recent meta-analyses. In papers published in academic journals, the average impact of a nudge is very large – an 8.7 percentage point take-up increase over the control. In the Nudge Unit trials, the average impact is still sizable and highly statistically significant, but smaller at 1.4 percentage points. We show that a large share of the gap is accounted for by publication bias."

DellaVigna, S., & Linos, E. (2020). RCTs to Scale: Comprehensive Evidence from Two Nudge Units.

Automatization of Research Synthesis

Neurosynth



http://neurosynth.org

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. ***Nature Methods,*** ***8***(8), 665–670. http://doi.org/10.1038/nmeth.1635

Human Behaviour-Change Project

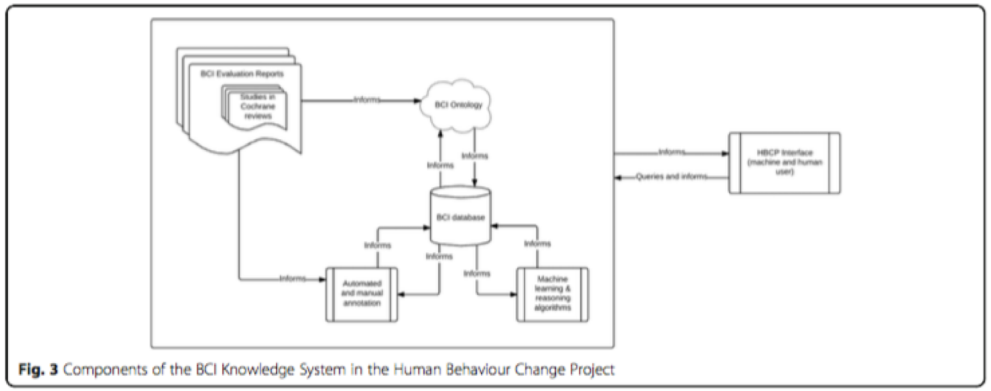


Fig. 3 Components of the BCI Knowledge System in the Human Behaviour Change Project

http://www.ucl.ac.uk/human-behaviour-change

Michie, S., Thomas, J., Johnston, M., Mac Aonghusa, P., Shawe-Taylor, J., Kelly, M. P., et al. (2017). The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation, ***Implementation Science***, 1–12. http://doi.org/10.1186/s13012-017-0641-5

Summary

- research synthesis can be helpful in dealing with information explosion and quantification of effects of interest
- the history of research synthesis is defined by a progressive standardisation through the development of terminology (i.e., systematic review, meta-analysis), guidelines (e.g., PRISMA), and procedures with the goal of increasing clarity, transparency, and reduce bias (e.g., transparent exclusion criteria, protocols)
- the key statistical ingredient of quantitative research synthesis is weighted aggregation in which the information from several estimates is aggregated as a function of the confidence in each study (precision)
- research synthesis is not a panacea and cannot provide accurate estimates of effects in the face of large reporting biases (publication bias & file-drawer problem)…