NASEM Question:

What are best practices for addressing publication bias?

The potential for publication (reporting) bias!

'*the failure to publish the results of a study on the basis of the direction or strength of the study findings*' – DeVito & Goldarce (2018)

Why?

File drawer problem.

Time?

Motivation?

Sponsorship / Funding.

Slow publication.

Not limited to nutrition research

Question: What does publication bias look like?

'Results indicate that published studies yield larger effect sizes than unpublished (d = 0.18): Polanin et al (2016)'

Prevalence of assessing for publication bias

Table 1
Publication Bias Procedures Used in Included Meta-Analyses and the Proportion of Studies Using Them That Found Evidence of Bias

Procedure used	Proportion of meta-analyses	Positive hit rate for technique
Comparison of published/unpublished studies effect sizes	19 (21%)	35%
Visual examination of funnel plot	5 (5%)	0%
Fail-safe *N*	20 (22%)	35%
Orwin's fail-safe *N*	7 (8%)	14%
Begg and Mazumdar's (1994) rank-correlation	4 (4%)	50%
Egger's regression	5 (5%)	50%
Trim and fill	22 (24%)	64%
Other, less broadly used procedures	13 (14%)	15%
No publication bias procedure	27 (30%)	N/A
One procedure	38 (42%)	53%
Two procedures	23 (25%)	30%
More than two procedures	3 (3%)	0%

Across 91 meta-analyses **70% demonstrated some effort** to evaluate publication bias

41% reported some evidence of publication bias.

Publication Bias in Psychological Science: Prevalence, Methods for Identifying and Controlling, and Implications for the Use of Meta-Analyses

Christopher J. Ferguson
Texas A&M International University

Michael T. Brannick
University of South Florida

How often is publication bias assessed for in systematic reviews meta-analyses?

43% of health services delivery systematic reviews and meta-analyses **mention** publication bias.

		Assessed Publication bias		
Factor	**All (n = 200) [n (%)]**	**Yes (n = 19) [n (%)]**	**No (n = 181) [n (%)]**	**Univariable OR (95% CI)**
Being an intervention review (versus association review)	100 (50%)	14 (74%)	86 (48%)	3.09 (1.07–8.95)
Number of included studies (≥10)	157 (79%)	17 (90%)	140 (77%)	2.49 (0.55–11.22)
Meta-analysis included	43 (22%)	18 (95%)	25 (14%)	112.32 (14.35–879.03)
Included only RCT and controlled trials [a]	36 (18%)	7 (37%)	29 (16%)	3.06 (1.11–8.42)
Searched grey/unpublished literature	103 (52%)	6 (32%)	97 (54%)	0.40 (0.15–1.10)
Quality assessment performed	157 (79%)	18 (95%)	139 (77%)	5.44 (0.71–41.96)
Authors reported using GRADE	23 (12%)	2 (11%)	21 (12%)	0.90 (0.19–4.16)
Authors reported using systematic review guideline	73 (37%)	14 (74%)	59 (33%)	5.79 (1.99–16.84)
Journal impact factor in the year 2016 [median (IQR)]	3.00 (2.26,5.10)	3.85 (2.73,5.76)	2.94 (2.14,4.98)	1.09 (1.004–1.18)
Journal endorses systematic review guideline (as of the year 2018)	140 (70%)	10 (53%)	130 (72%)	0.44 (0.17–1.34)

Table. Elements Extracted From Each of 324 Systematic Reviews and Meta-analyses (continued)

Element	No. (%)
Publication bias formally evaluated	
Yes	257 (79.3)
No	67 (20.7)
Publication bias found	
Yes	26 (8.0)
Probable	1 (0.3)
No	37 (11.4)
Unspecified	260 (80.2)
Reporting guidelines used[a]	
PRISMA	133 (41.0)
MOOSE	8 (2.5)
CONSORT	5 (1.5)
AHRQ	2 (0.6)
NHS CRD	1 (0.3)
STREGA	1 (0.3)
None mentioned	179 (55.2)

Including unpublished studies in meta-analysis ‘grey-literature’

Should unpublished data be included?

Cook et al (JAMA, 1993): 78% of methodologists suggest unpublished data should definitely or probably be included, with around 47% of editors agreeing

Cochrane – ‘The inclusion of data from unpublished studies can itself introduce bias. ’

Of 203 reviews 36% of studies did not describe any attempt to obtain unpublished studies. Of those that sought it out, 89% found some! (Ziai et al, Search for unpublished data by systematic reviewers: an audit)

Will unpublished data stay hidden forever?

the publishing landscape has transformed

Should systematic reviews and meta-analyses include data from preprints?

Elisa Brietzke, [1,2,3] Fabiano A. Gomes, [1,3,4] Fernando Gerchman, [5,6] and Rafael C. R. Freire [1,2,3]

Using preprints in evidence synthesis: Commentary on experience during the COVID-19 pandemic

Barbara Clyne,[a,b,*] Kieran A. Walsh,[b] Eamon O'Murchu,[b] Melissa K. Sharp,[a] Laura Comber,[b] Kirsty K O' Brien,[b] Susan M. Smith,[a] Patricia Harrington,[b] Michelle O'Neill,[b] Conor Teljeur,[b] and Máirín Ryan[b,c]

PRISMA 2020 flow diagram template for systematic reviews. The new design is adapted from flow diagrams proposed by Boers [55], Mayo-Wilson et al. [56] and Stovold et al. [57] The boxes in grey should only be completed if applicable; otherwise they should be removed from the flow diagram. Note that a "report" could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report or any other document providing relevant information

What not to use… Fail Safe N.

Fail-safe N determines the number of studies with a significant effect that would need to be added in order for the meta-analytic effect to be non-significant.

Discouraged!

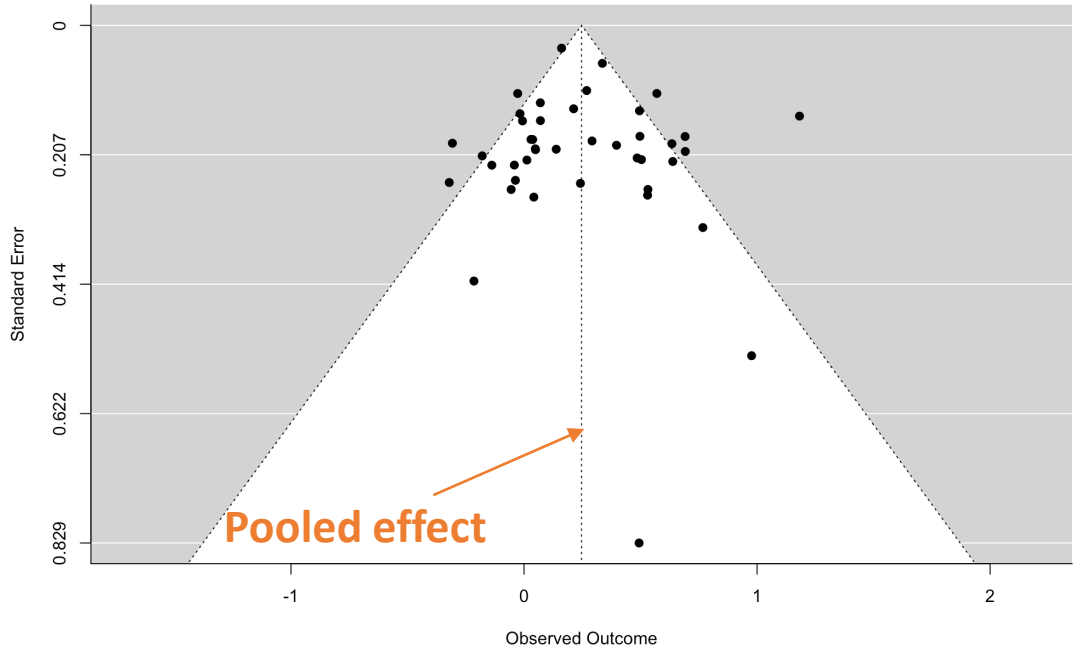
```
Fail-safe N Calculation Using the Rosenthal Approach

Observed Significance Level: <.0001
Target Significance Level:   0.05

Fail-safe N: 1399
```

1. It does not tell you whether there is bias.
2. Greater bias can lead to a greater Fail-Safe N.
3. Hypotheses that would appear to be false have otherwise obtained very large values of FSN.

Not recommended by Cochrane

Funnel plots

A scatter plot of the study effect size (intervention effect) against some measure of precision (sample size) of the study.

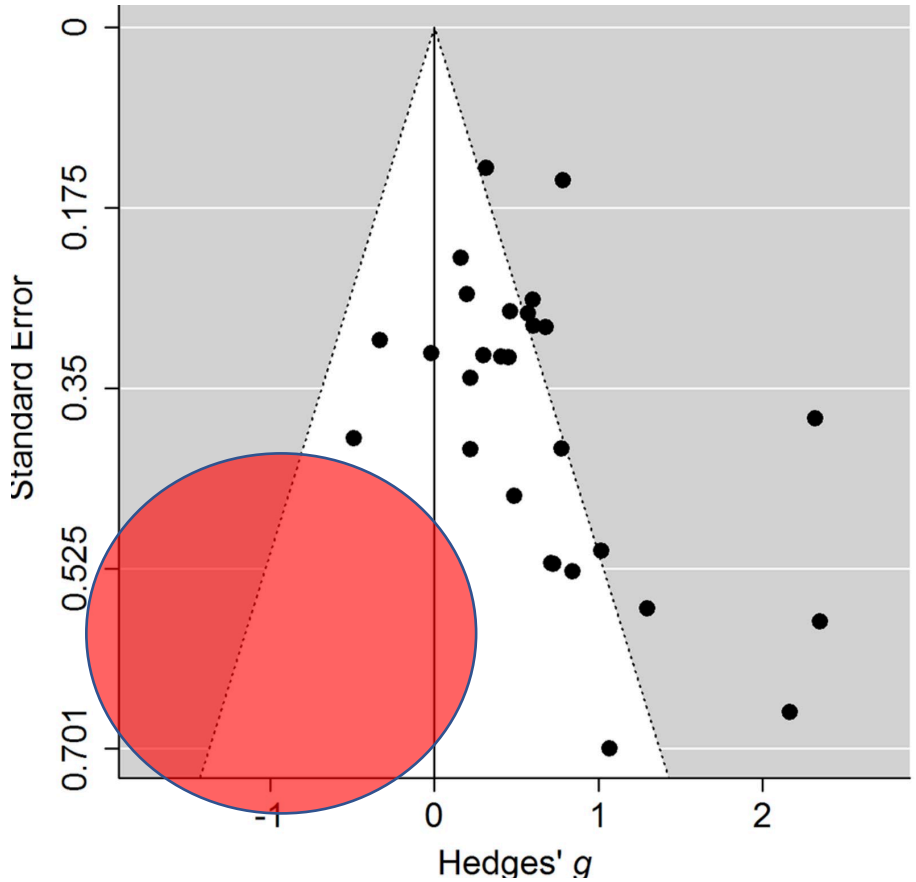


High powered studies

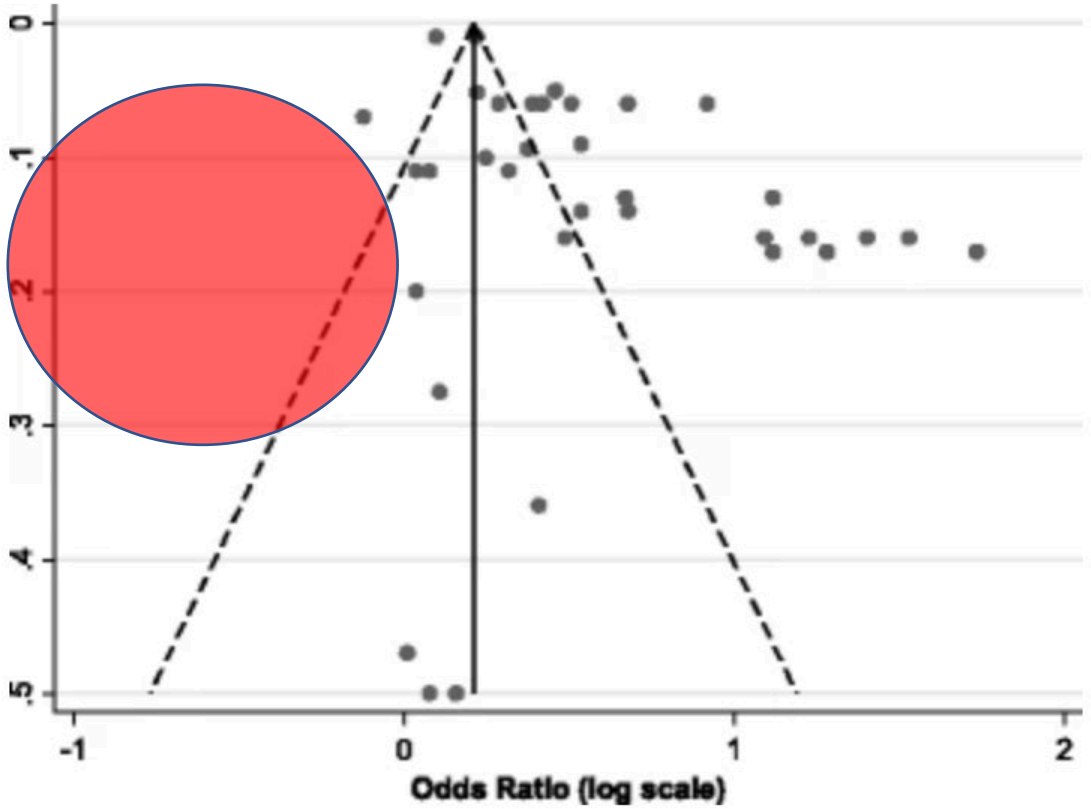
Low powered studies

Looking for symmetry!

Example from elsewhere



Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterols, and triglyceride



A systematic review, and meta-analysis of the impact of health-related claims on dietary choices

Funnel plots alone…..

Subjective! Can be difficult to judge especially when the number of studies is small

Used as evidence of publication bias … but actually ‘small study bias’, differences in study quality etc.

ELSEVIER

Journal of Clinical Epidemiology

Volume 58, Issue 9, September 2005, Pages 894-901

Original Article

In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias

Norma Terrin, Christopher H. Schmid, Joseph Lau

Asymmetry of the funnel plot, either visually interpreted or statistically tested, does not accurately predict publication bias

Inappropriate or misleading use of funnel plot tests may do more harm than good

BMJ. 2006 Sep 16; 333(7568): 597–600.

PMCID: PMC1570006

doi: 10.1136/bmj.333.7568.597

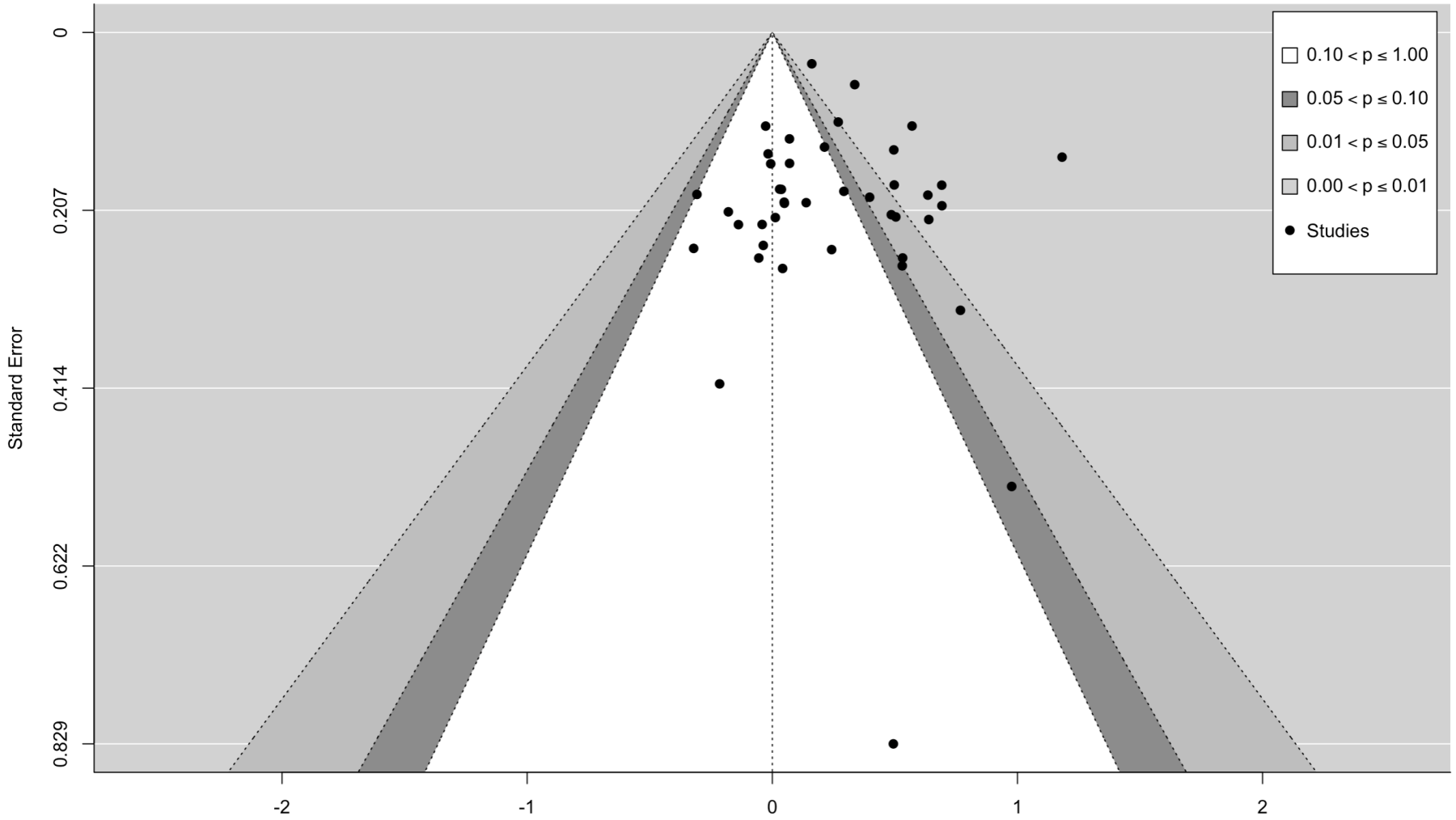
PMID: 16974018

Evidence based medicine

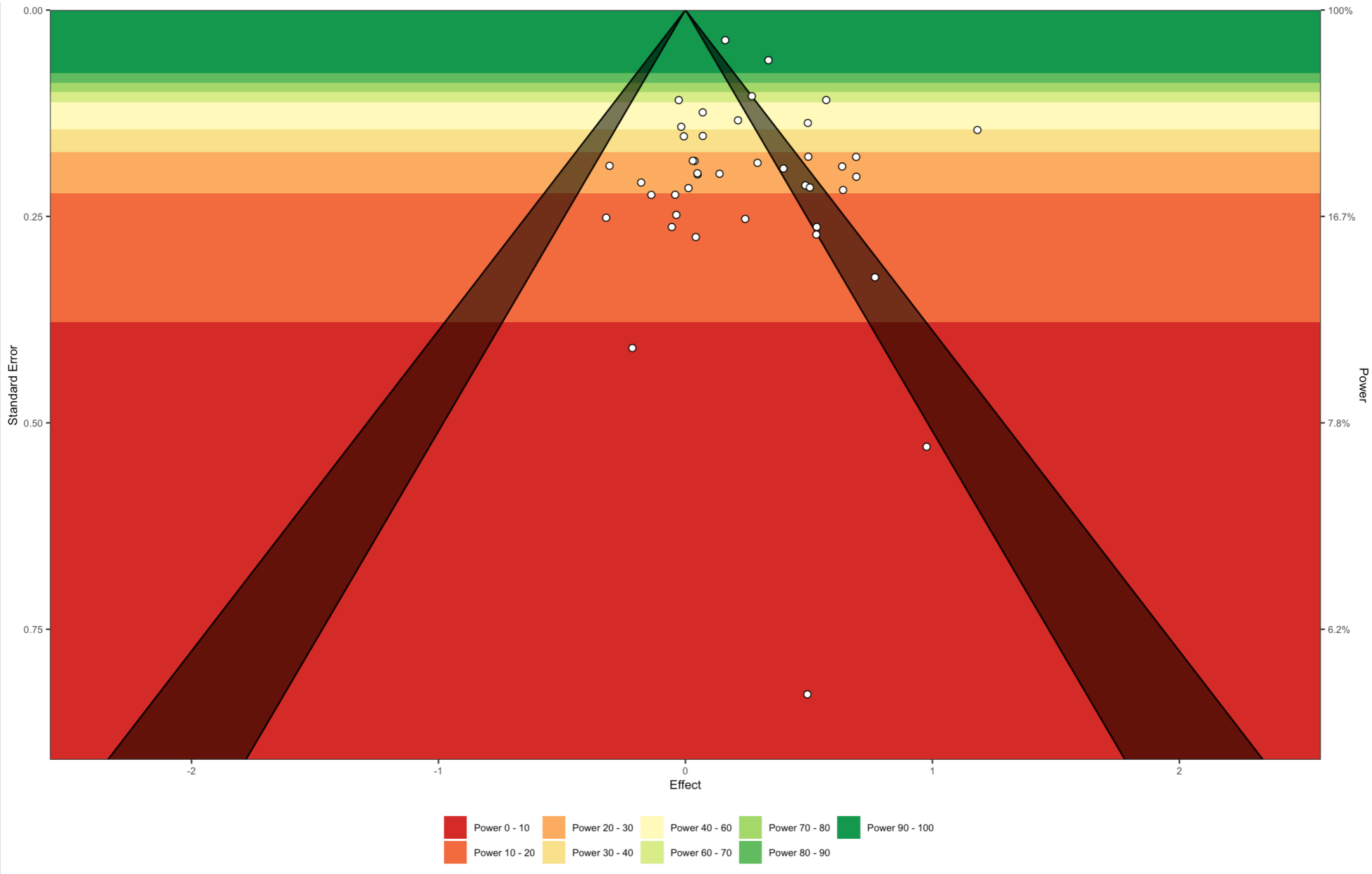
The case of the misleading funnel plot

Joseph Lau, professor,[1] John P A Ioannidis, professor,[2] Norma Terrin, associate professor,[1] Christopher H Schmid, professor of medicine,[1] and Ingram Olkin, professor[3]

Contour Enhanced Funnel plots



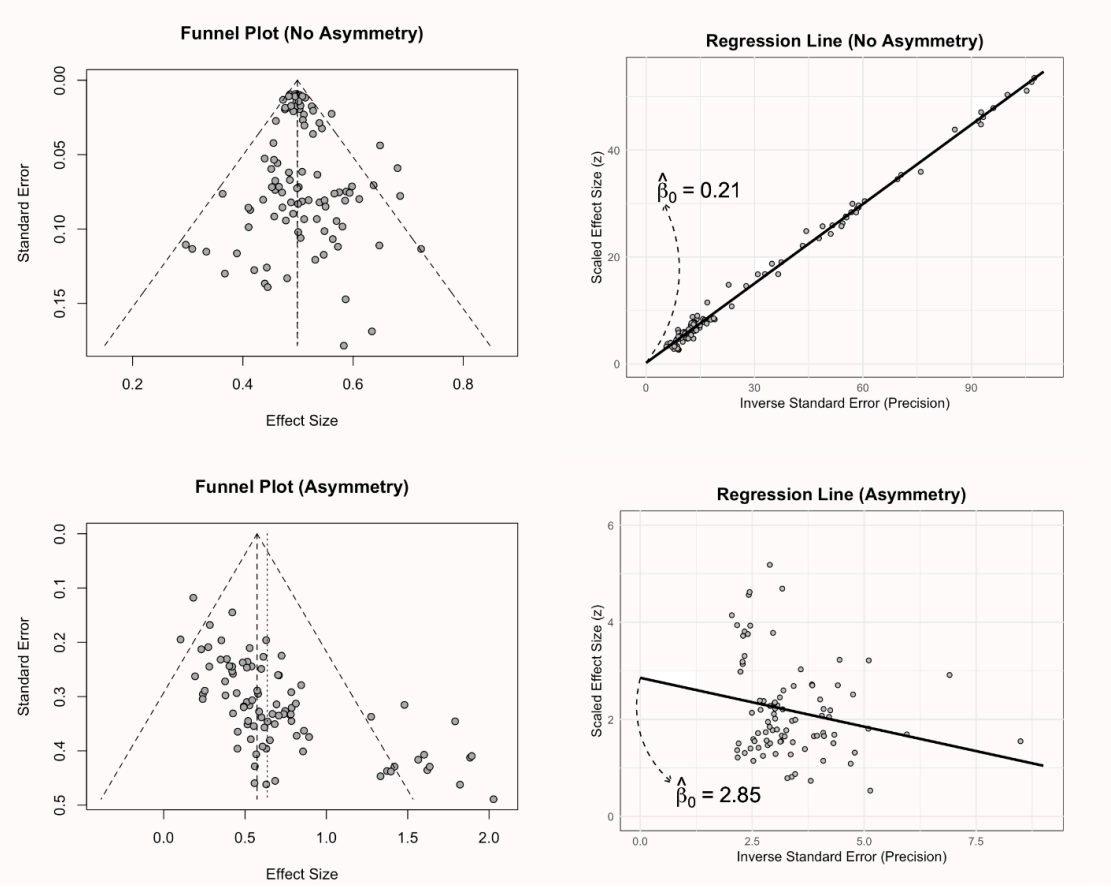
Sunset Funnel plot



Statistical assessment of publication (small study) bias based on asymmetrical funnel plots

Egger's test

Test of funnel plot asymmetry – regresses the scaled effect size against the precision. If there is no small study bias then the intercept should be 0



Papers

Bias in meta-analysis detected by a simple, graphical test

BMJ 1997 ; 315 doi: https://doi.org/10.1136/bmj.315.7109.629 (Published 13 September 1997)
Cite this as: *BMJ* 1997;315:629

PET/PEESE (Precision-effect test and precision-effect estimate with standard errors)

PET fits a linear regression line, then extrapolates to estimate the effect size of a hypothetical study with a standard error of zero

PEESE fits a quadratic regression line (using the squared standard error).

Sample	*k*	*g*
Food Consumption	14	0.44• (-0.01, 0.89)
Hand Grip	13	0.56‡ (0.31, 0.81)
Impossible Anagrams	21	0.46‡ (0.23, 0.69)
Impossible Puzzles	16	0.79‡ (0.56, 1.02)
Possible Anagrams	12	0.24 (-0.07, 0.56)
Standardized Tests	13	0.30* (0.05, 0.54)
Stroop	16	0.24† (0.07, 0.41)
Working Memory	13	0.32* (0.08, 0.56)
Combined	116	0.43‡ (0.34, 0.52)

Table 7. Parameter estimates for PET, PEESE, and FAT.

Sample	PET	PEESE
Food Consumption	-0.21 (-2.35, 1.93)	-0.01 (-1.13, 1.11)
Hand Grip	-0.76 (-1.55, 0.04)	-0.11 (-0.54, 0.32)
Impossible Anagrams	0.04 (-0.66, 0.74)	0.15 (-0.22, 0.53)
Impossible Puzzles	-0.16 (-0.76, 0.43)	0.22 (-0.16, 0.60)
Possible Anagrams	-0.71 (-1.93, 0.51)	-0.23 (-0.83, 0.38)
Standardized Tests	0.27 (-0.85, 1.38)	0.27 (-0.37, 0.90)
Stroop	-0.27• (-0.58, 0.04)	-0.07 (-0.24, 0.11)
Working Memory	-0.15 (-1.20, 0.99)	0.09 (-0.47, 0.65)
Combined	-0.27* (-0.52, -0.01)	0.003 (-0.14, 0.15)

Can fail badly under **high heterogeneity**

Trim and Fill technique

Wolters Kluwer Medicine
Home Search Submit a Manuscript

Medicine (Baltimore). 2019 Jun; 98(23): e15987. PMCID: PMC6571372
Published online 2019 Jun 7. doi: 10.1097/MD.0000000000015987 PMID: 31169736

The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses

Linyu Shi, MS and Lifeng Lin, PhD*

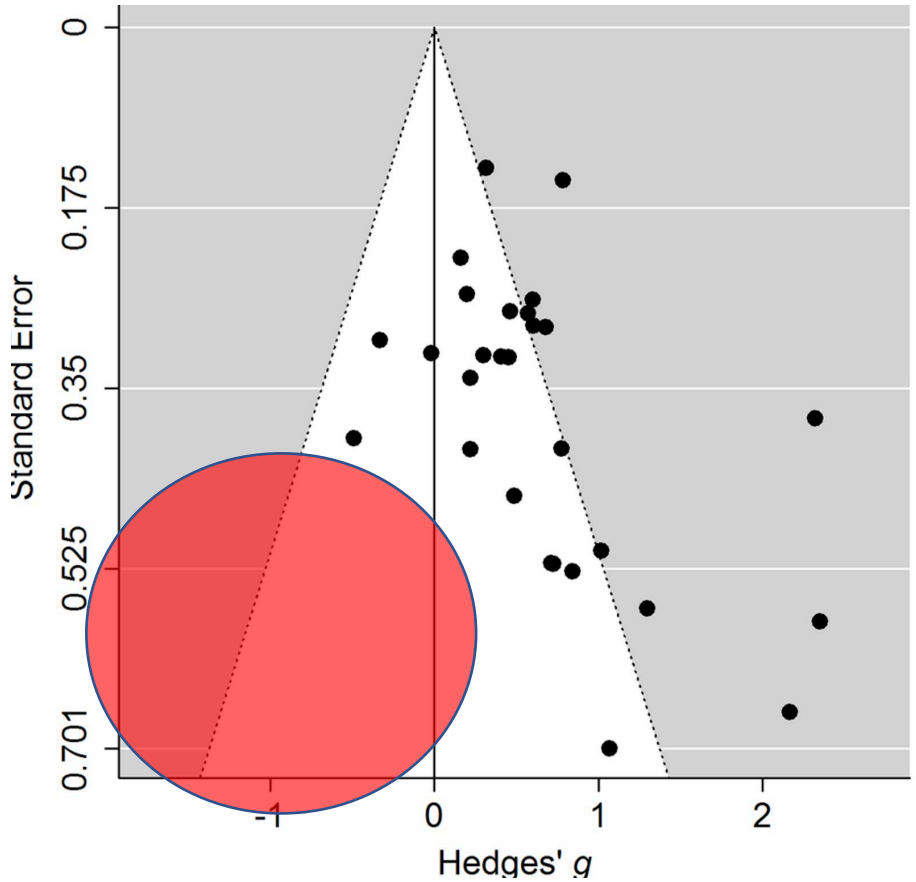
Seen as ‘intuitive’ (despite ‘non-trivial’ underlying statistics)

Studies with the most extreme effect sizes are **trimmed**

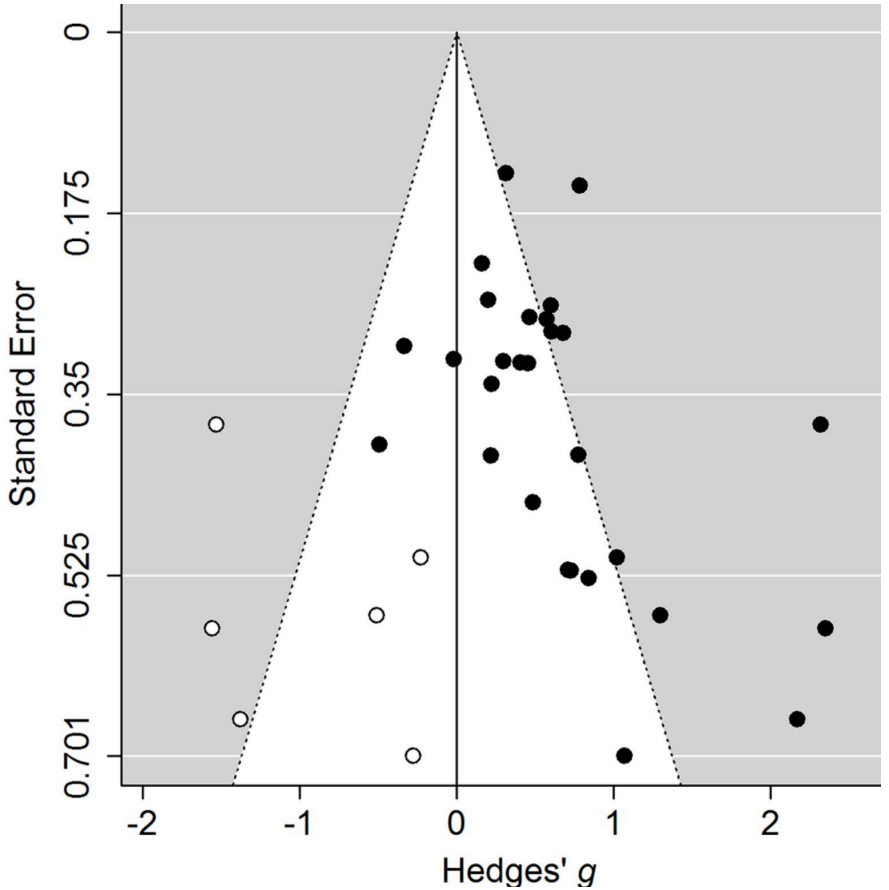
Then hypothetical effects are **filled** to create a more symmetrical funnel plot

‘Corrects for publication bias?’

Trim and Fill in practice

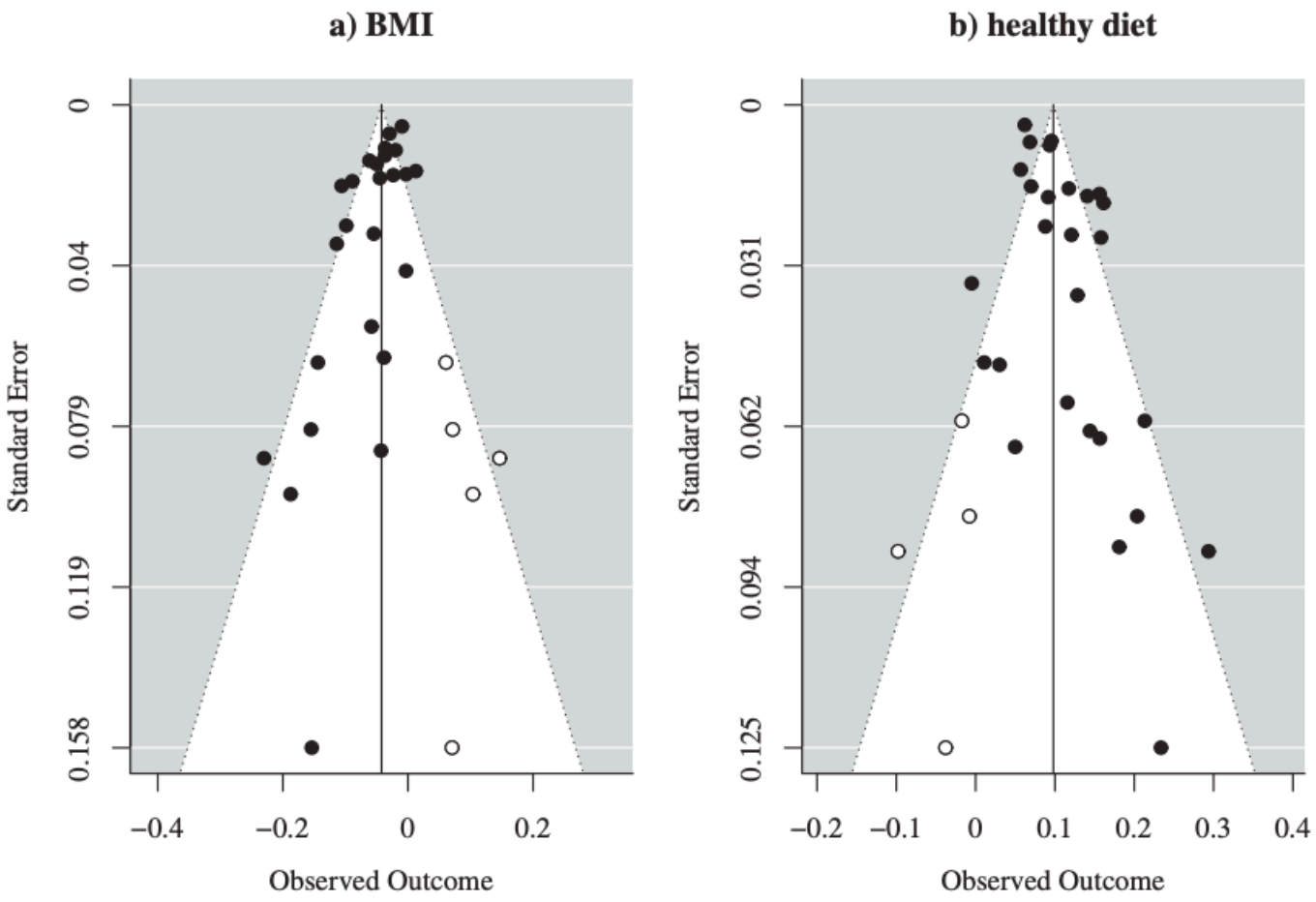


Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterols, and triglyceride



Matches the trimmed studies!

The frequency of family meals and nutritional health in children: a meta-analysis



Issues with Trim and Fill: Heterogeneity and Outliers

Having frequent family meals was significantly associated with a lower BMI (***r* = −0.05**, 95% CI [−0.06,–0.03]), a more healthy diet (*r* = 0.10, 95% CI [0.09, 0.12])

'Trim and fill analyses imputed **five** hypothetically missing studies for BMI, **four** studies for healthy diet.

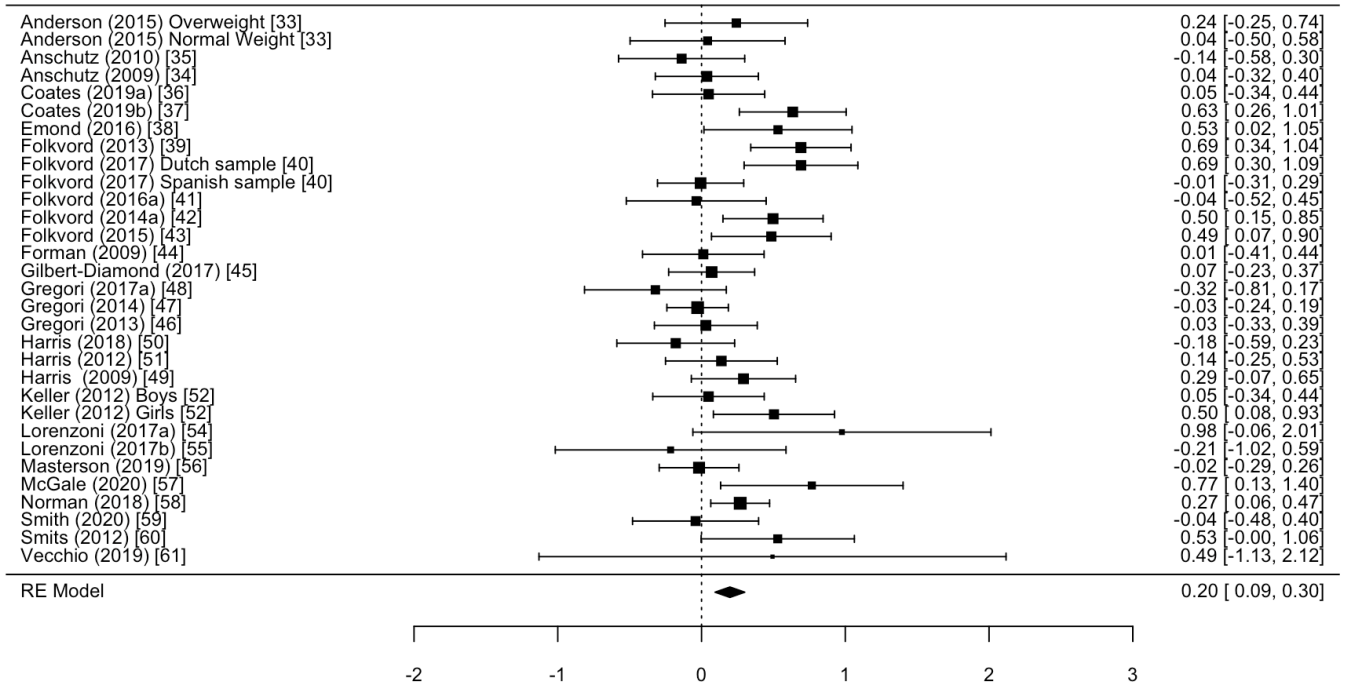
....

The adjusted effect sizes, taking publication bias into account, remained the same or were only slightly lower but still significant..

BMI: *r* = **−0.042**; 95% CI [−0.06, −0.03]; healthy diet: *r* = 0.10; 95% CI [0.08, 0.12];'

Moving beyond the funnel plot: Test of Excess Significance

Compares the expected number of significant effects against the observed number.
Expected number is based on the pooled effect and the power of the individual studies to detect that effect.



Expected = sum of power

```
Observed Number of Significant Findings: 9 (out of 31)
Expected Number of Significant Findings: 7.9677
Observed Number / Expected Number:       1.1296

Estimated Power of Tests (based on theta = 0.1975)

   min      q1  median      q3     max
0.0635  0.1945  0.2502  0.2974  0.5254

Test of Excess Significance: p = 0.3357 (X^2 = 0.1800, df = 1)
Limit Estimate (theta_lim):  0.1150 (where p = 0.1)
```

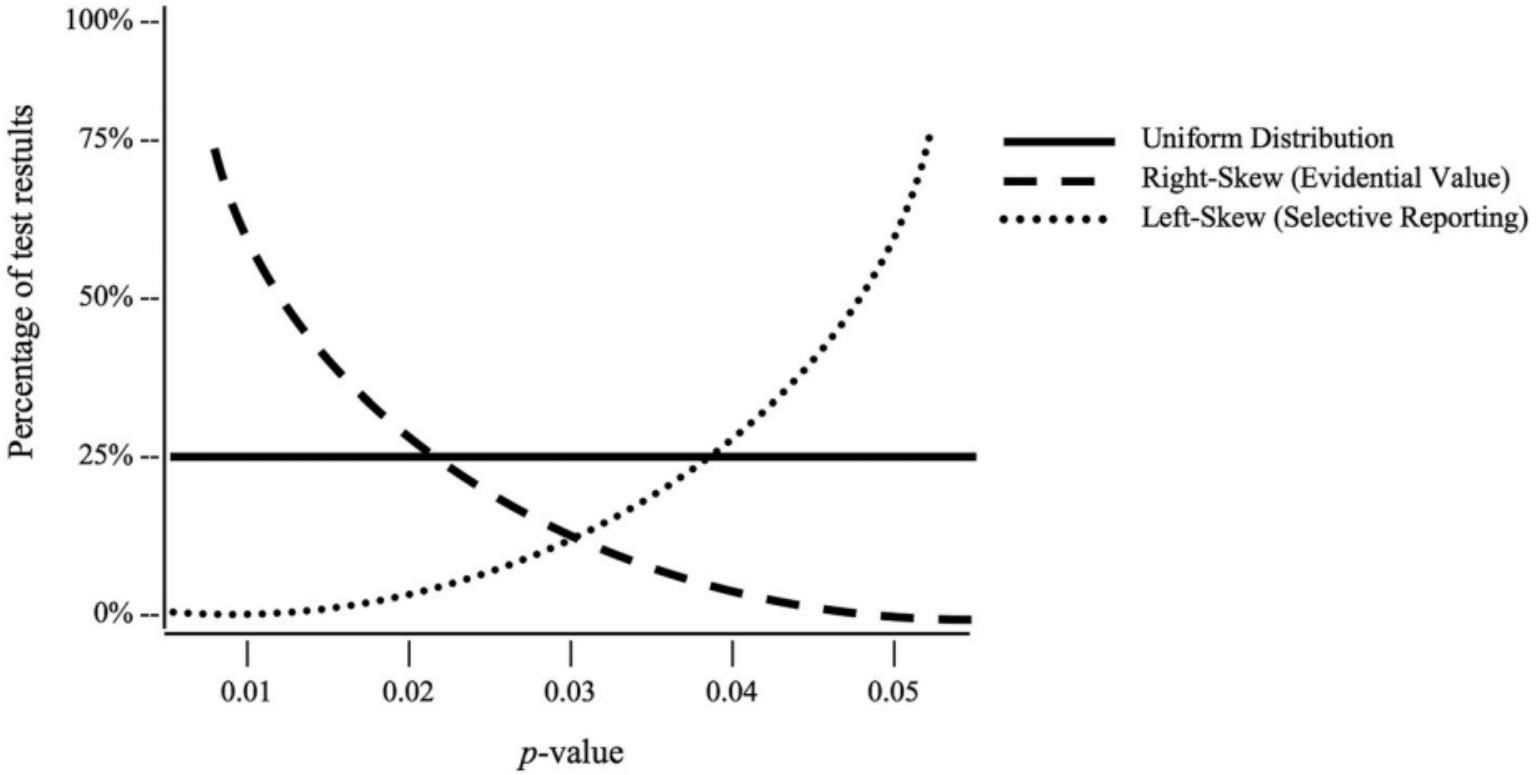
Less relevant for cohort studies where the power of studies tends to be much larger.

Not currently recommended by Cochrane

P-Curve (Simonsohn, Nelson and Simmons, 2014)

P-values are often used as a threshold for **importance.**

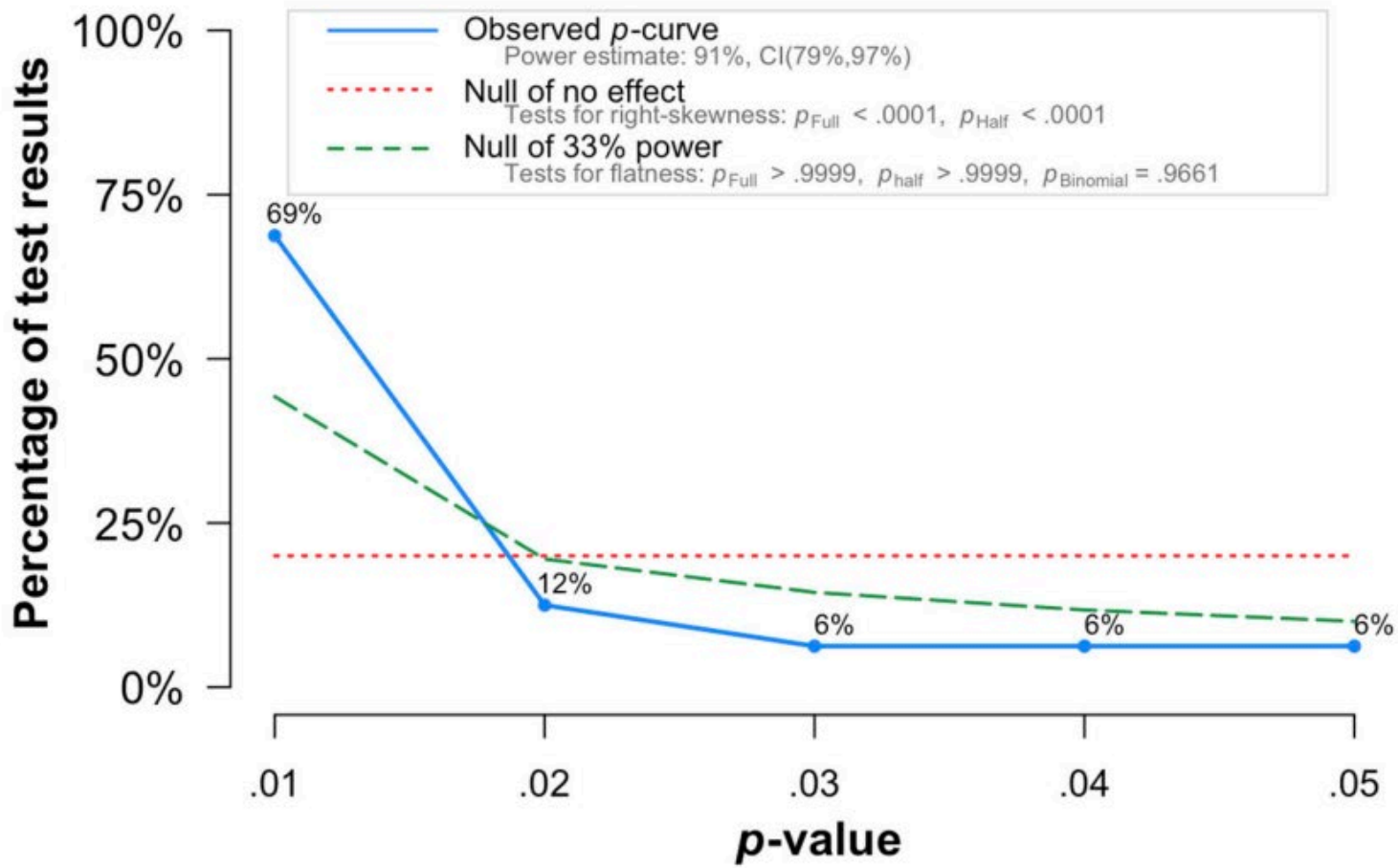
P-Curve focuses on the –values of the effects (not necessarily the effect sizes and variances). It doesn't infer the average effect, but rather considers the evidential value of the analyzed studies.



Association of Food and Nonalcoholic Beverage Marketing With Children and Adolescents' Eating Behaviors and Health

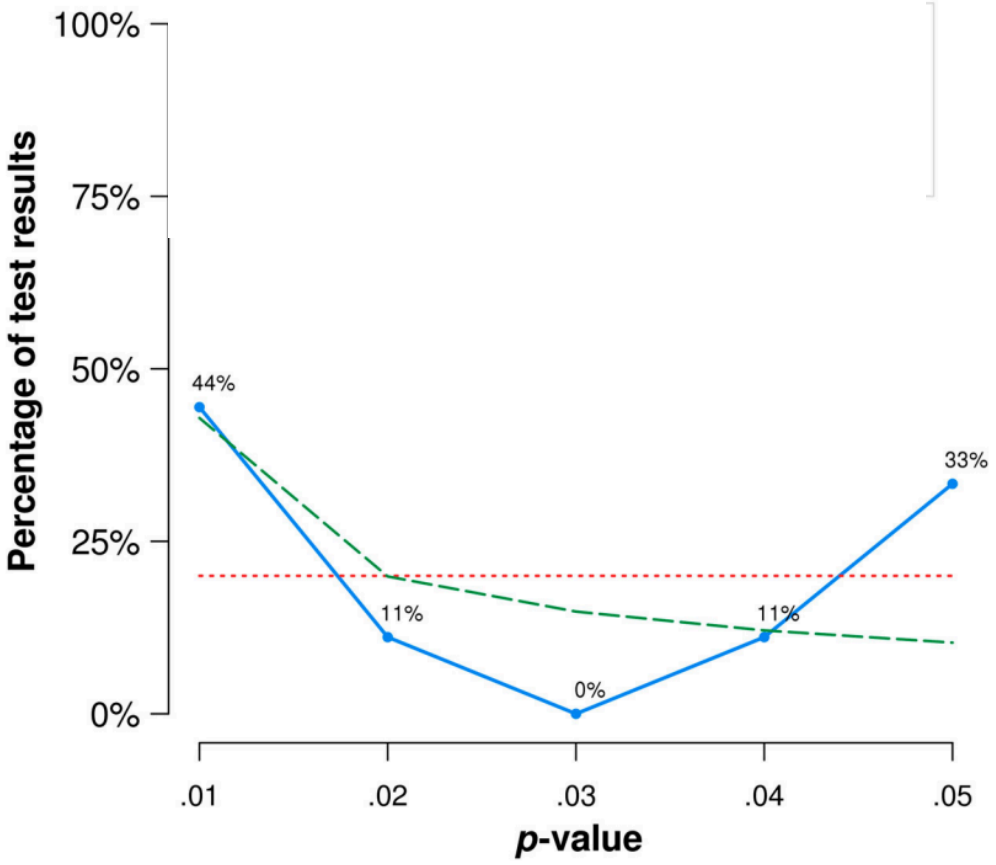
A Systematic Review and Meta-analysis

Emma Boyland, PhD,[1] Lauren McGale, PhD,[1,2] Michelle Maden, PhD,[3] Juliet Hounsome, PhD,[3] Angela Boland, PhD,[3] Kathryn Angus,[4] and Andrew Jones, PhD[1]



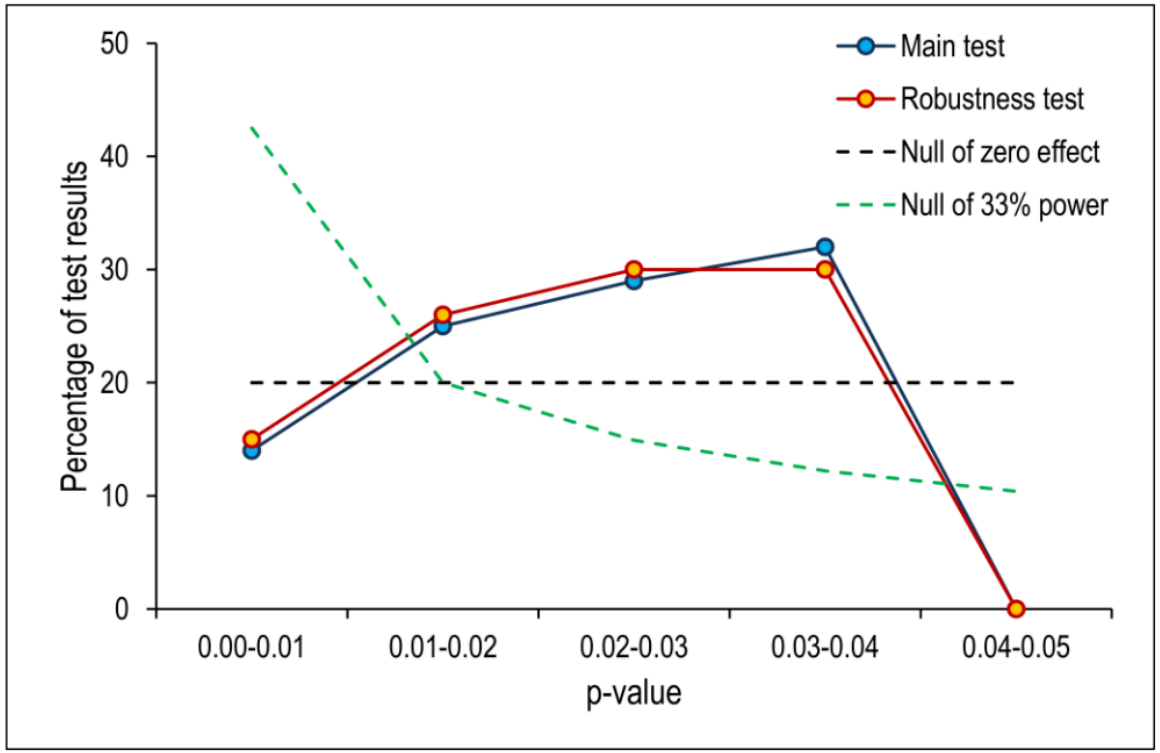
Quantifying the presence of evidential value and selective reporting in food-related inhibitory control training: a *p*-curve analysis

Kaylie A. Carbine[a] and Michael J. Larson[a,b]



The Bitter Truth About Sugar and Willpower: The Limited Evidential Value of the Glucose Model of Ego Depletion

Miguel A. Vadillo, Natalie Gold, and Magda Osman ✉ View all authors and affiliations



Using GOSH to identify impact of influential studies.

GOSH = Graphical Display of Study Heterogeneity

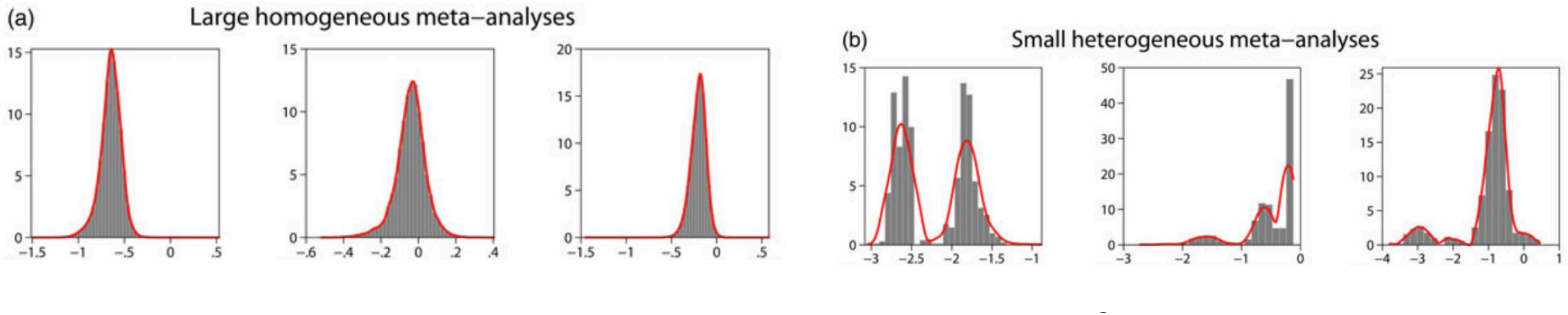
All subsets combinatorial meta-analysis.

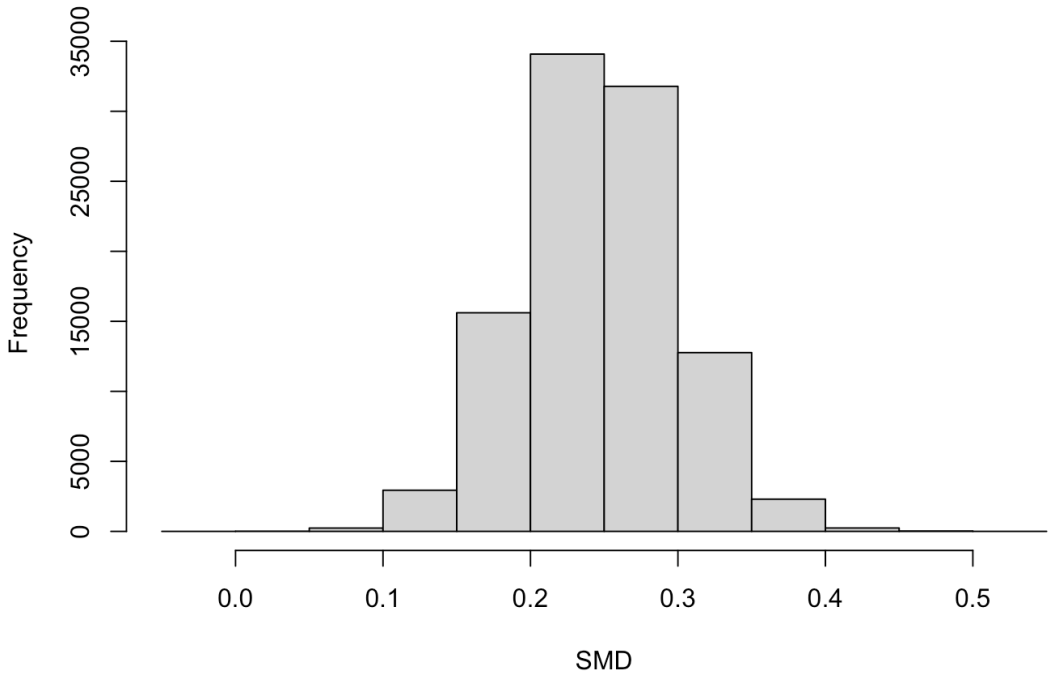
for a meta-analysis of $k \geq 2$ studies, there are $2k - 1$ potential subsets of studies

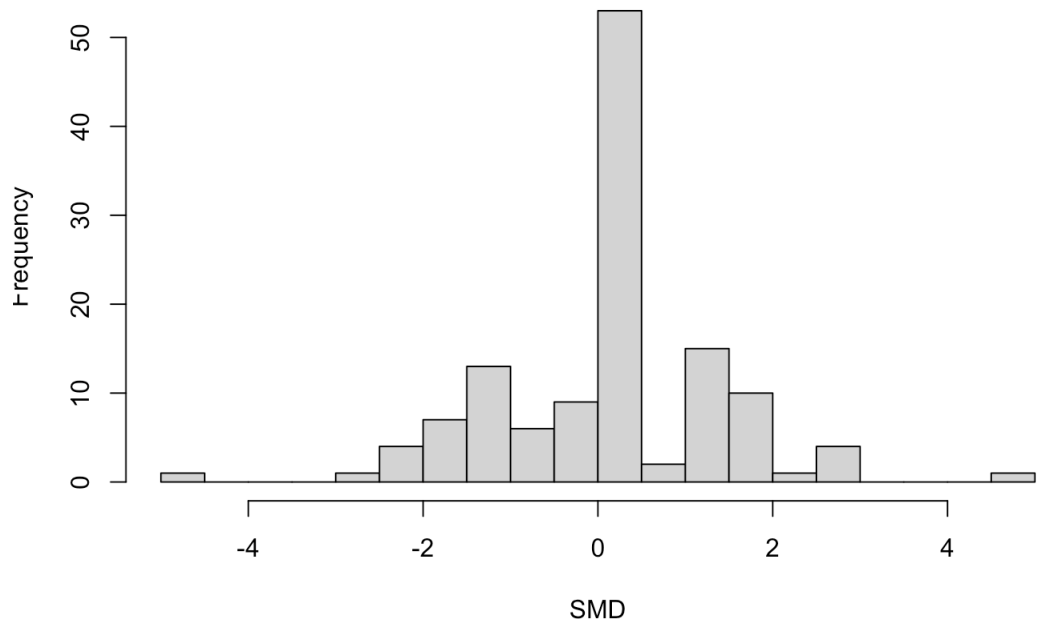
For a meta-analysis with 20 studies there are over 1 million possible combinations.

GOSH – a graphical display of study heterogeneity

Ingram Olkin,[a*†] Issa J. Dahabreh[b,c] and Thomas A. Trikalinos[c]





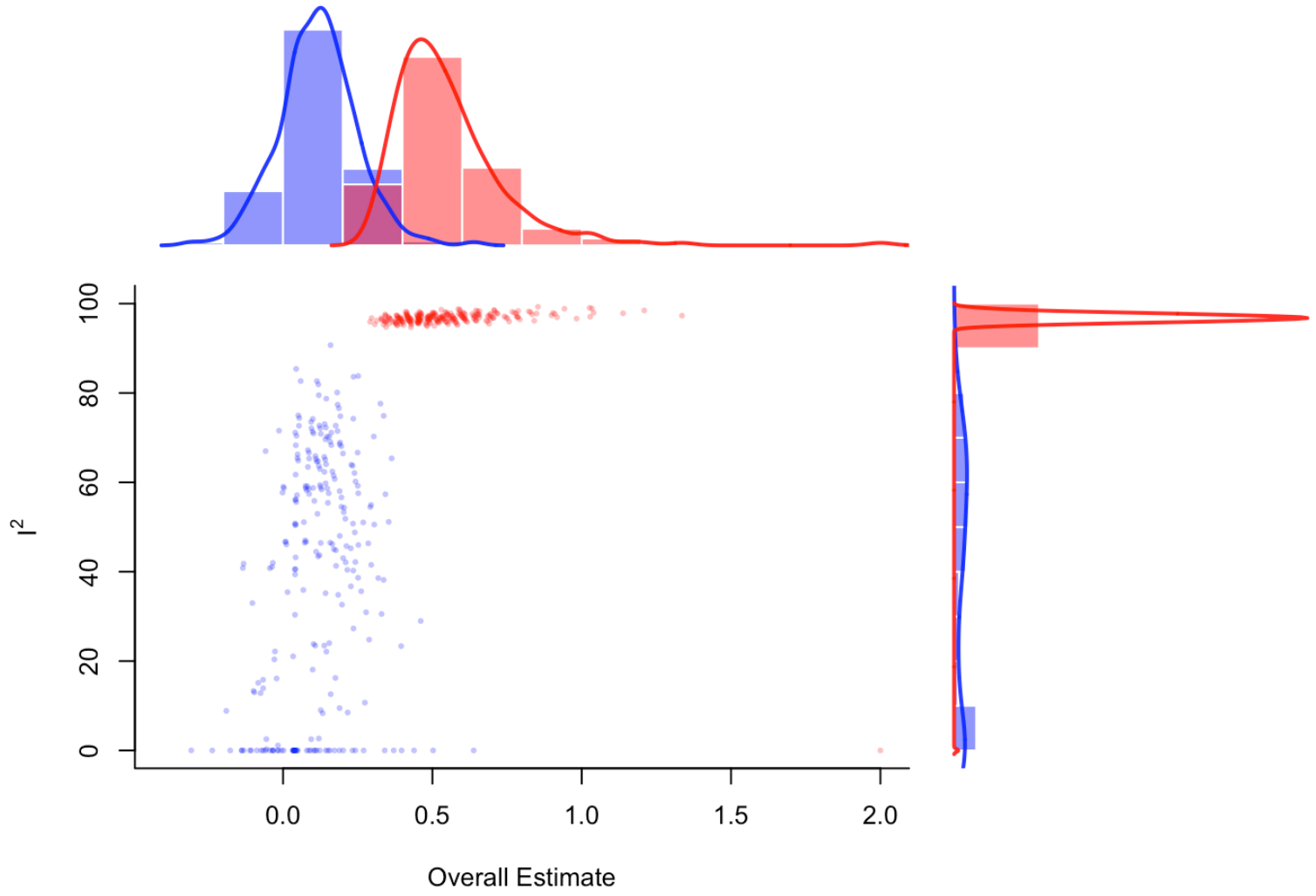


Association of Food and Nonalcoholic Beverage Marketing With Children and Adolescents' Eating Behaviors and Health

A Systematic Review and Meta-analysis

Emma Boyland, PhD,[1] Lauren McGale, PhD,[1,2] Michelle Maden, PhD,[3] Juliet Hounsome, PhD,[3] Angela Boland, PhD,[3] Kathryn Angus,[4] and Andrew Jones, PhD[1]

GOSH of influential study



One study with a large effect size can bias all models in which it is included, particularly in a smaller effect size.

However… might not be publication bias!

Recommendations

1. Check for unpublished research

2. Do publication bias tests (multiple)
   'we advocate that selection methods should be used less for obtaining a single estimate that purports to adjust for publication bias ex post and more for sensitivity analysis (McShane, 2016)'

3. Examine influential cases and what they might do to heterogeneity but also the meta-analytic effect

'Remember! The tests do not directly test for publication bias, but for a relationship between the observed effect sizes or outcomes and the chosen predictor. If such a relationship is present, then this usually implies asymmetry in the funnel plot, which in turn may be an indication of publication bias. However, it is important to keep in mind that there can be other reasons besides publication bias that could lead to asymmetry in the funnel plot.'

Happy to share any of the R code used here: a.j.jones@ljmu.ac.uk