



THE CAMPBELL COLLABORATION

What helps? What harms? Based on what evidence?

Food security and novel food solutions: implications
for food choice, safety, and waste: 2nd June
Newcastle

Establishing an evidence-based approach to food security

Gavin Stewart



@campbellreviews

@metanutter

Gavin.Stewart@Newcastle.ac.uk

Direction

- Why do we need evidence?
- What does our evidence look like?
- The Replication crisis
 - The dance of the p values
 - Publication bias
 - Reporting and researcher degrees of freedom
 - Novelty and theory
- Possible Solutions



Why do we need evidence?

- The challenge of feeding Nine billion people
 - No more land, climate change, increasing variability



Science 327, 812 (2010)

What does our evidence look like?

- **The replication crisis**



Schooler, J. W. (2014). "Metascience could rescue the 'replication crisis'". *Nature*. 515 (7525): 9.

Empirical evidence

Domain	Findings	Sources
Medicine	<p>Out of 49 highly cited papers, 45 claimed that studied therapy was effective. Of these studies, 16% were contradicted by subsequent studies, 16% had found stronger effects than did subsequent studies, 44% were replicated, and 24% remained largely unchallenged.</p> <p>11% of pre-clinical cancer studies were replicable</p>	<p>Ioannidis JA (13 July 2005). Contradicted and initially stronger effects in highly cited clinical research. <i>JAMA</i>. 294 (2): 218–228.</p> <p>Begley, CG., and Lee ME., (2012) Drug Development: Raise Standards for Preclinical Cancer Research, <i>Nature</i>. 483, 531–533.</p>
Psychology	<p>Out of 100 studies from high-ranking journals only 36% had significant findings (p value below .05) compared to 97% of the original studies. The mean effect size in the replications was approximately half the magnitude of the effects reported in the original studies.</p> <p>Questionable research practices (QRPs) have been identified as common in the field (majority of 2000 scientists confess to at least one of: e.g. selective reporting, p-hacking, nonpublication of data, post-hoc storytelling (framing exploratory analyses as confirmatory analyses), manipulation of outliers).</p>	<p>Collaboration, Open Science (2015). "Estimating the reproducibility of psychological science". <i>Science</i>. 349 (6251): aac4716.</p> <p>Leslie JK.; Loewenstein, GP, Drazen (2012). "Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling". <i>Psychological Science</i>. 23 (5): 524–532</p>

The dance of the P values

Strength of evidence
$P < 0.001$
$P < 0.01$
$P < 0.05$
$P \text{ } 0.05 \text{ to } ?$
$P > 0.1$

The classical P value: The probability of observing data at least as extreme as the actual data given infinite observations.... assuming the null hypothesis to be true



The dance of the P values

Strength of evidence	Significance language
$P < 0.001$	Very highly Significant
$P < 0.01$	Highly significant
$P < 0.05$	Significant
P 0.05 to ?	Approaching Significant
$P > 0.1$	Non-significant

The dance of the P values

Strength of evidence	Significance language	Suggests Truth
$P < 0.001$	Very highly Significant	There is definitely an effect
$P < 0.01$	Highly significant	There is an effect
$P < 0.05$	Significant	Most likely there is an effect
P 0.05 to ?	Approaching Significant	Almost? Probably? (but low power)
$P > 0.1$	Non-significant	No effect?

The dance of the P values

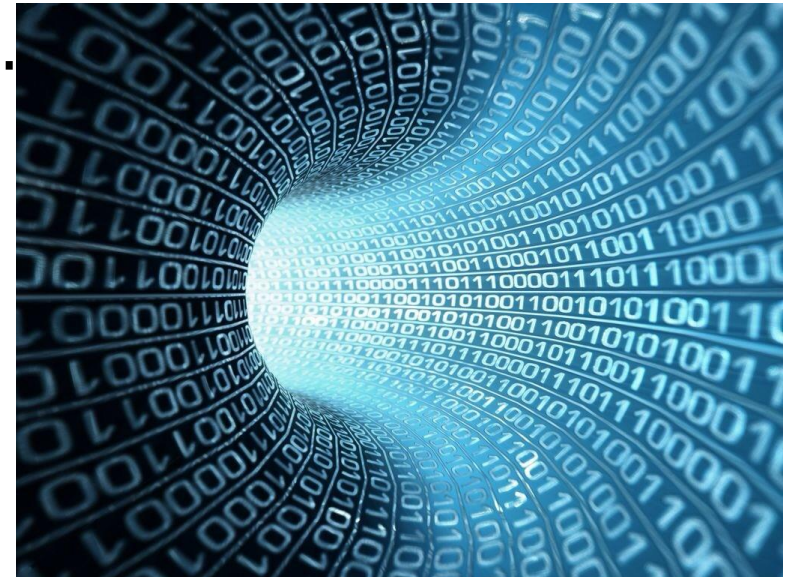
Strength of evidence	Significance language	Suggests Truth	Evokes emotion
$P < 0.001$	Very highly Significant	There is definitely an effect	Elation Exuberance Smugness?
$P < 0.01$	Highly significant	There is an effect	Dancing, Drinking
$P < 0.05$	Significant	Most likely there is an effect	Relief Cheerfulness
$P \text{ } 0.05 \text{ to } ?$	Approaching Significant	Almost? Probably? (but low power)	Frustration (if only)
$P > 0.1$	Non-significant	No effect?	Despair, depression

The dance of the P values

Strength of evidence	Significance language	Suggests Truth	Evokes emotion	Implications
$P < 0.001$	Very highly Significant	There is definitely an effect	Elation Exuberance Smugness?	Nobel Prize Tenure Research Grant
$P < 0.01$	Highly significant	There is an effect	Dancing, Drinking	**** publication PhD
$P < 0.05$	Significant	Most likely there is an effect	Relief Cheerfulness	*** publication
$P \text{ } 0.05 \text{ to } ?$	Approaching Significant	Almost? Probably? (but low power)	Frustration (if only)	Stress leave counselling
$P > 0.1$	Non-significant	No effect?	Despair, depression	Reconsider life goals

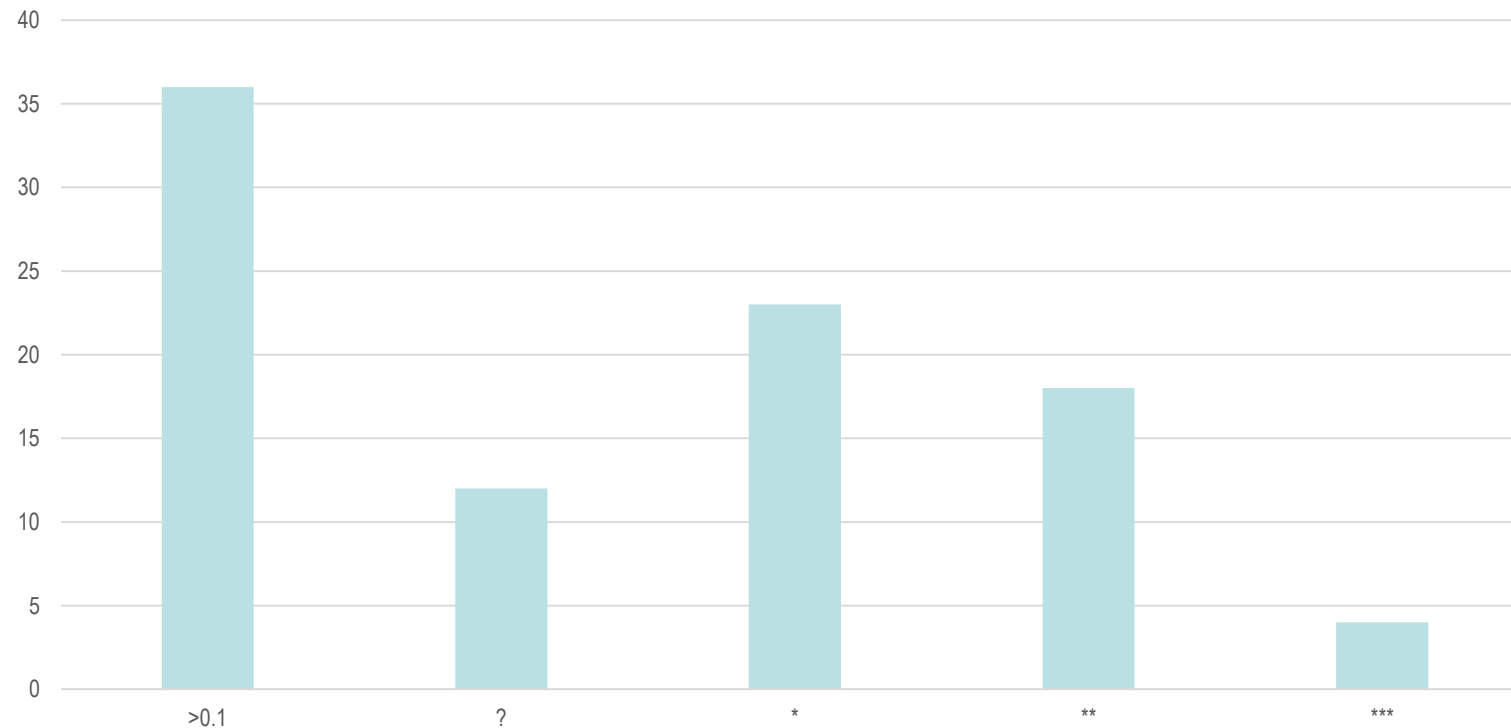
The Dance of the P values

- If P values are meaningful and represent the truth they should replicate...
- Lets run a simulation to see if they do...

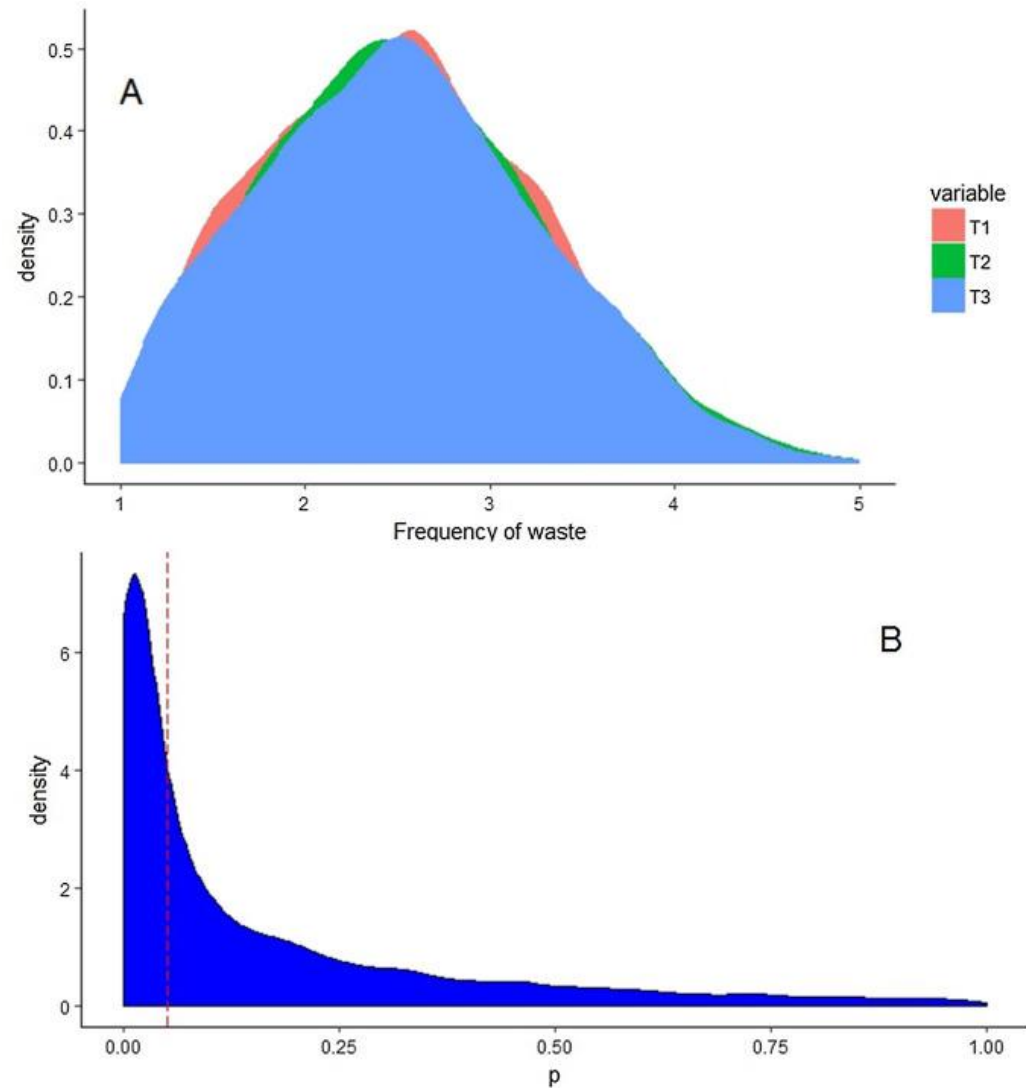


Dance of the P values

- P values do not replicate
- (Over)reliance on P values has serious consequences for the rigour of our science...



And in Food



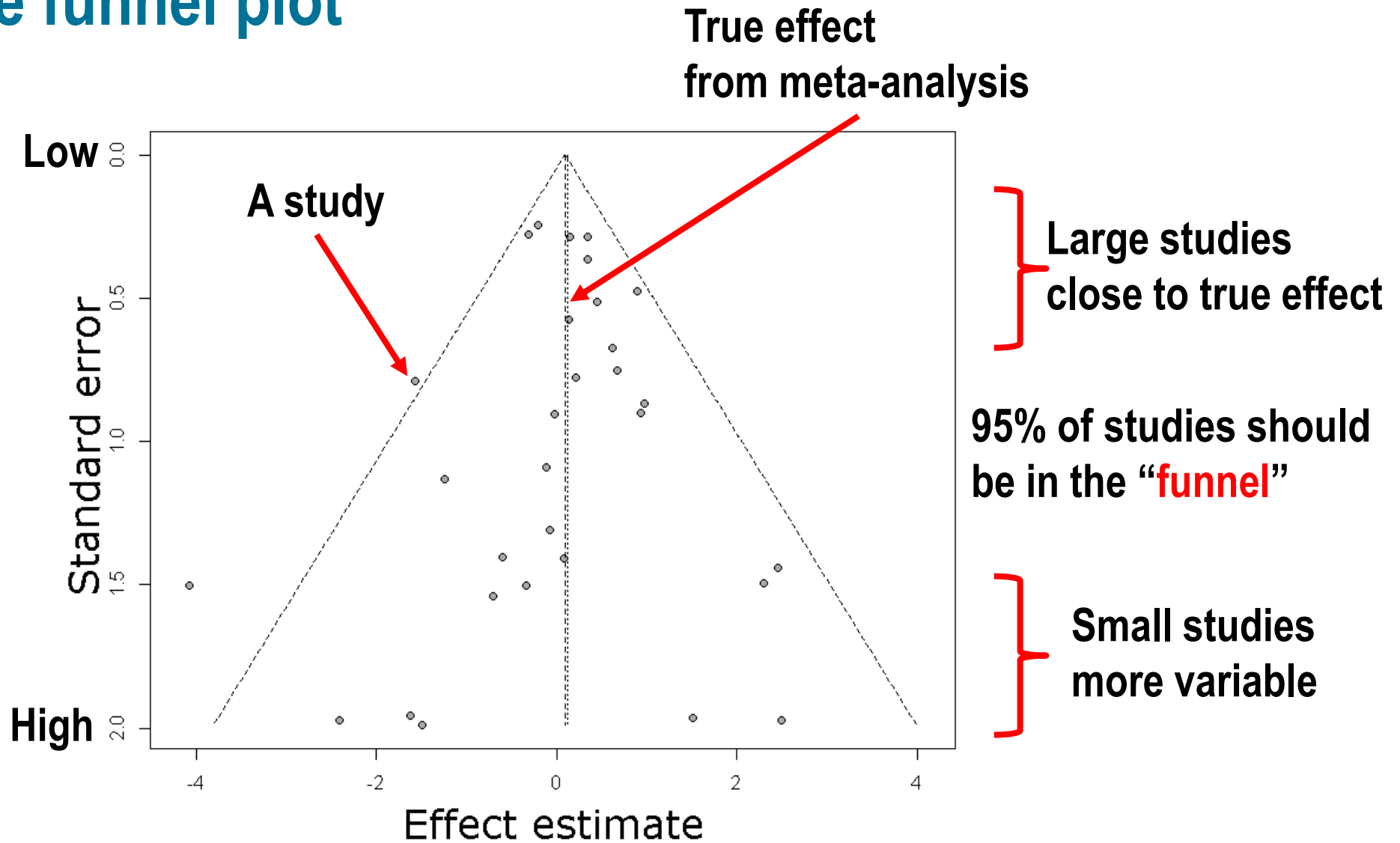
Grainger MJ, Stewart GB. **The jury is still out on social media as a tool for reducing food waste a response to Young et al. (2017).** *Resources, Conservation and Recycling* 2017, **122**, 407-410.

Publication bias

- Publication bias refers to bias that occurs when research found in the published literature is **systematically unrepresentative** of the population of studies (Rothstein et al., 2005)
- On average published studies have a larger mean effect size than unpublished studies, providing evidence for a publication bias (Lipsey and Wilson 1993)
- Also referred to as the **'file drawer' problem**:
 - “...journals are filled with the 5% of studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant (e.g. $p < 0.05$) results” (Rosenthal, 1979)
- Well-documented in different fields of research (biomedicine, public health, education, crime & justice, social welfare, ecology & evolution).

Rothstein, H. R., Sutton, A. J., & Borenstein, M. L. (Eds). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Hoboken, NJ: Wiley.

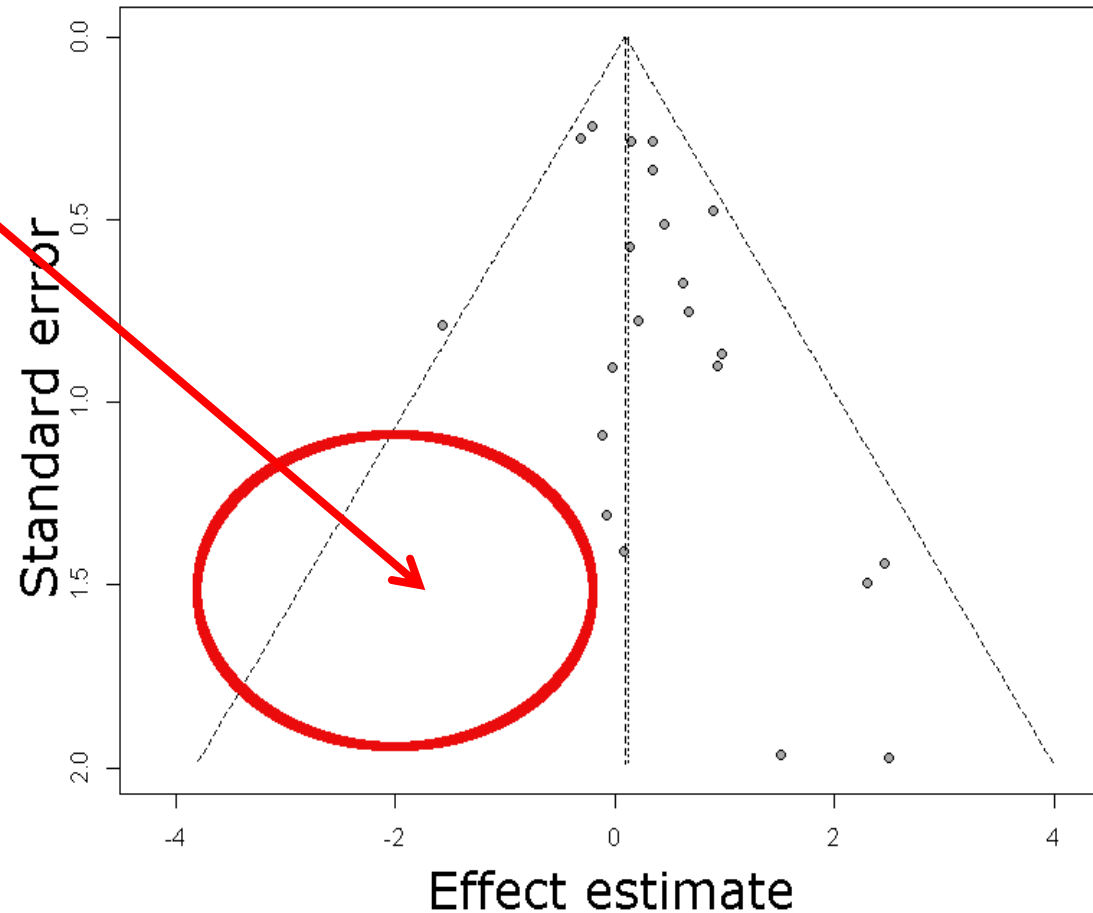
The funnel plot



Now with added publication bias

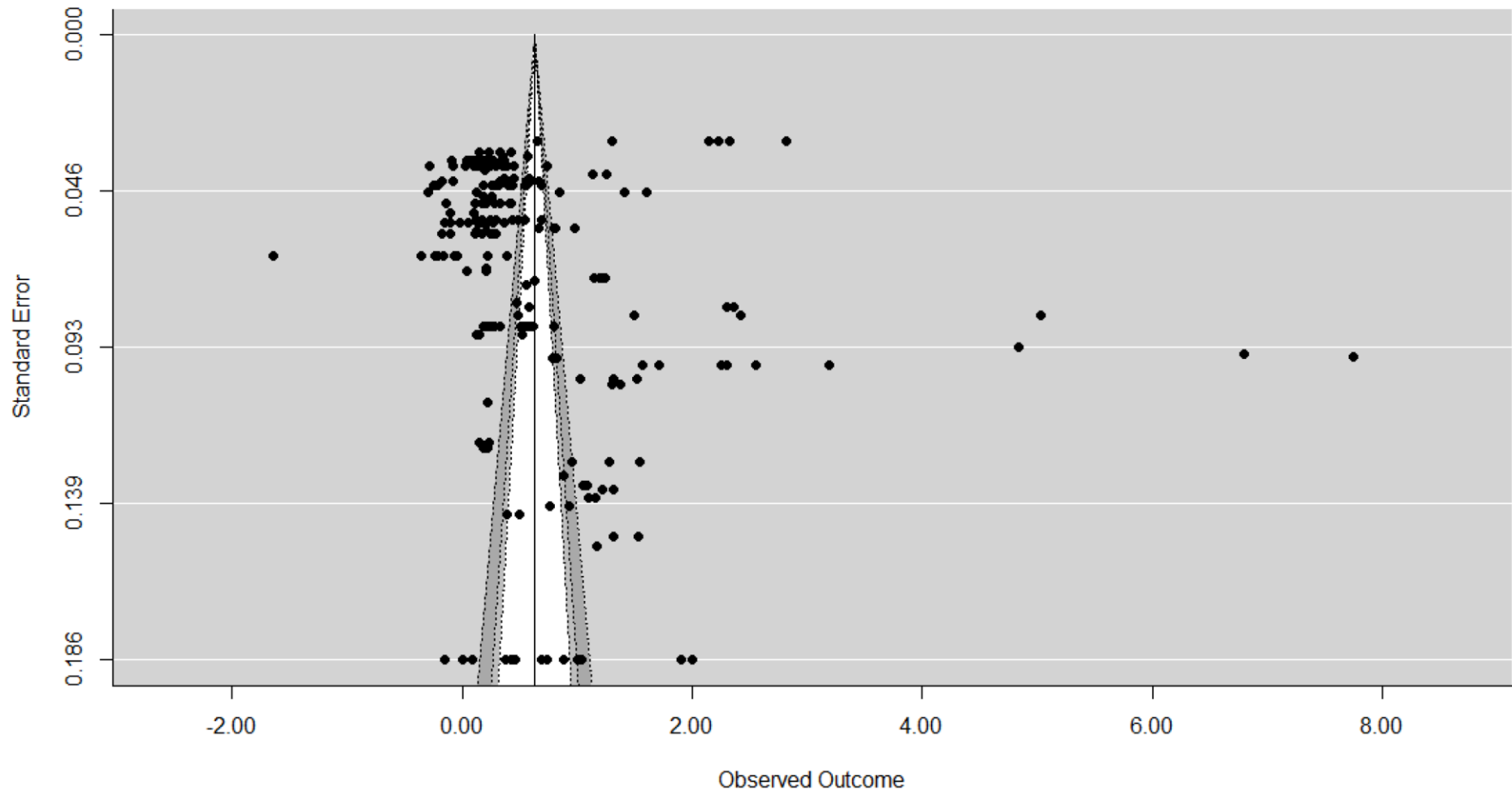
Studies missing from lower corner of funnel

Funnel is not symmetrical



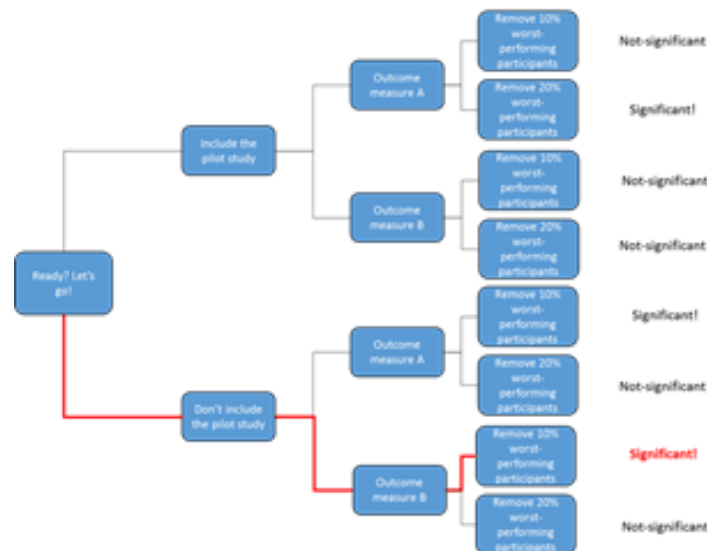
Sterne J *et al.* (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343, d4002.

In Food



Reporting and researcher degrees of freedom

- Do lots of things in different ways...and consciously or unconsciously introduce bias with selective reporting
- Develop an SEM with two different structures, split the data into male and female, analyse complete cases and imputed data...report only selected results (and worse selected methods)
- And just bad reporting of important information



In Food

- A systematic review of zinc biofortification



28 studies

Sample size reported for all

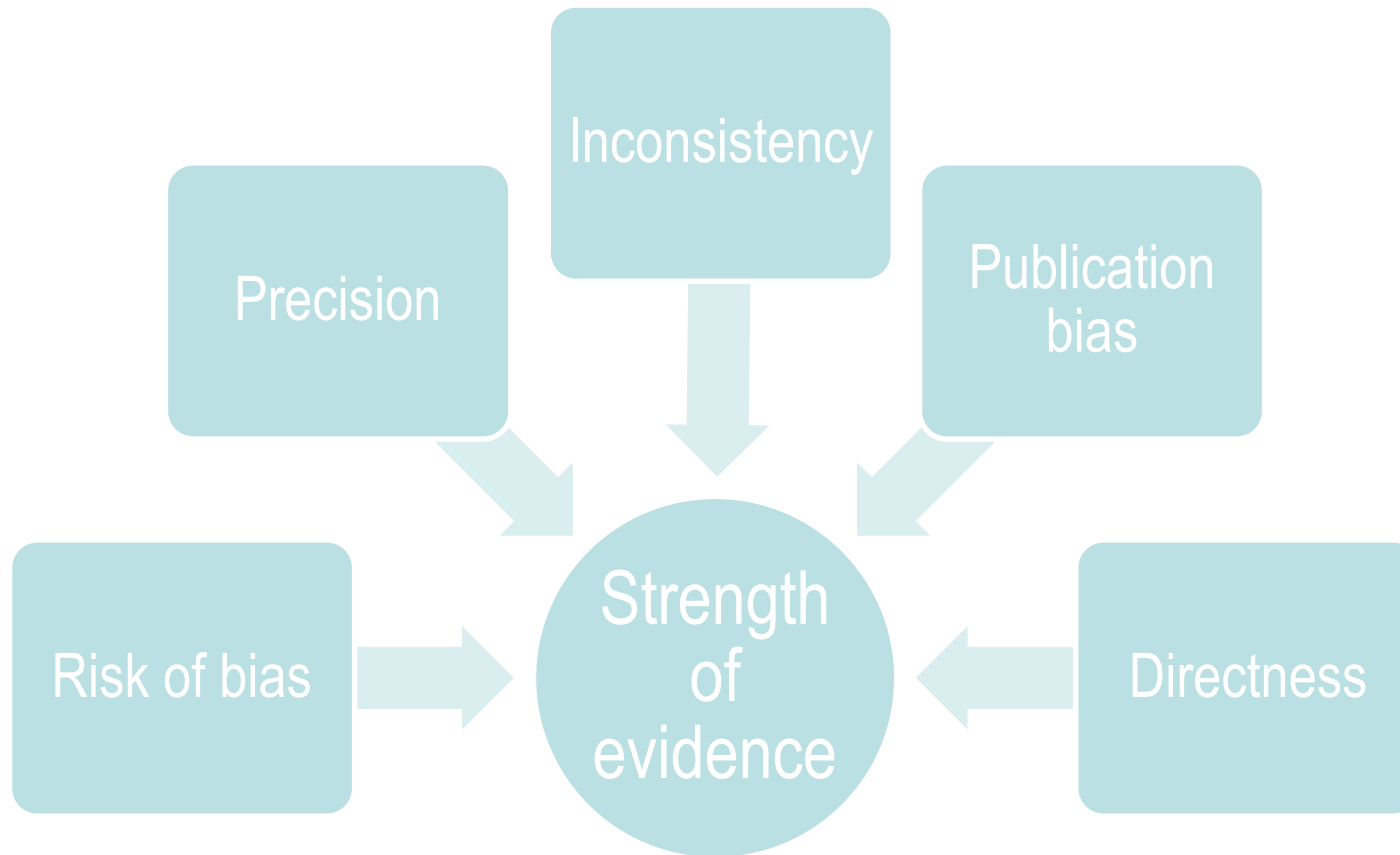
Only nine studies reported a measure of variance (standard error, standard deviation, confidence interval, variance)

Novelty and theory

- Good research must be novel with sound theoretical underpinnings?



Good research updates our belief about evidence



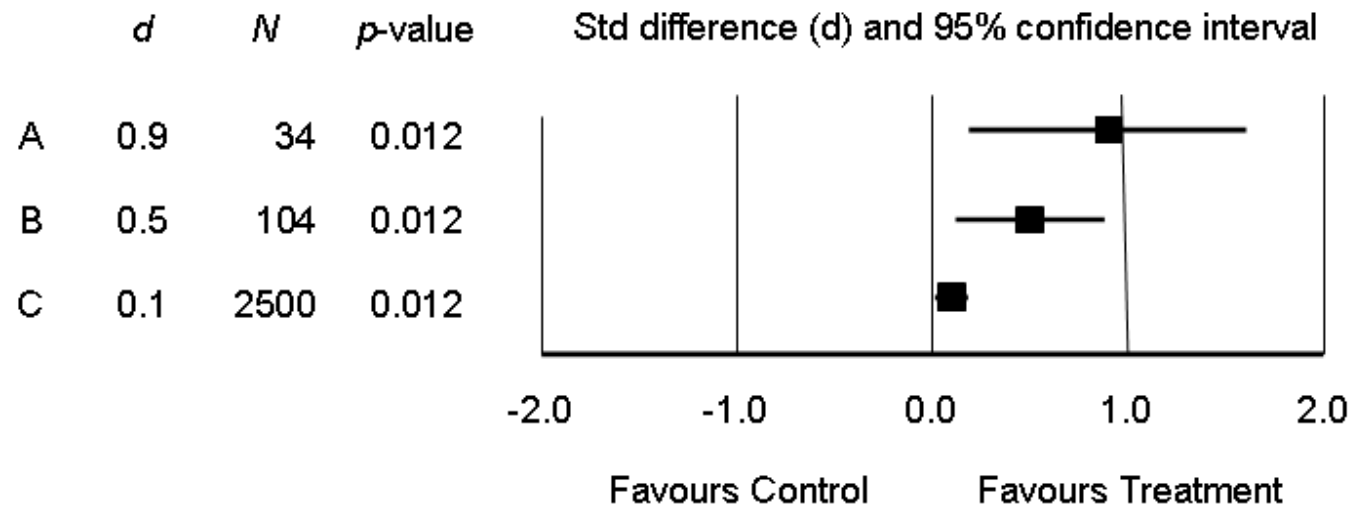
Stewart G, Higgins J, Schunnenman H, Meader N. (2015) The use of Bayesian Networks to assess the quality of evidence from research synthesis. *PLoS ONE* 10(4)

Summary to date

- We're **BAD**
- Over(reliance on p values)
- Publication bias
- Selective and poor reporting
- Fail to consider cumulative evidence appropriately

Solutions 1: P values

- Report and interpret effect sizes and confidence intervals (they convey much more information than p values)

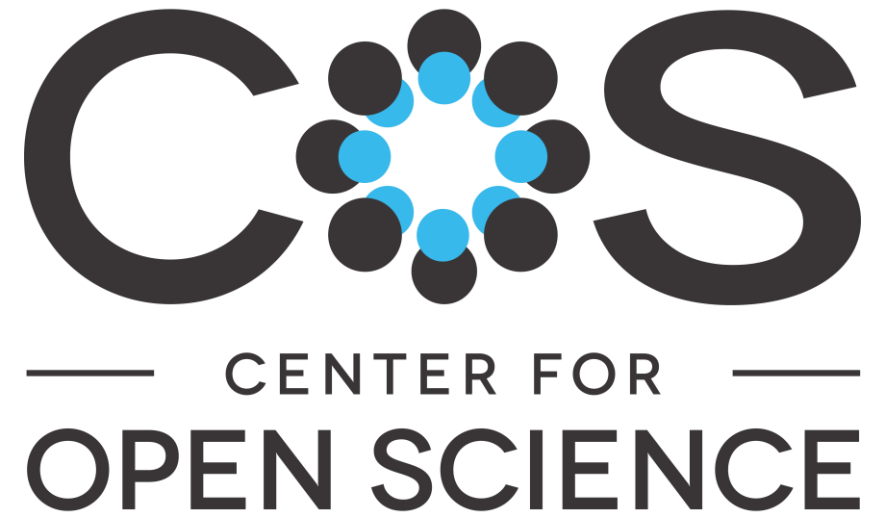


- Establish universal reporting guidelines to enforce this *cf* <https://www.equator-network.org>

- Some advocacy for banning p values altogether

Solution 2: Publication Bias

- Pre-registration
- TOP guidelines
 - Pre-registered
 - Open Data
 - Open Methods



Solution 3: selective and poor reporting

- See previous:
 - Less reliance on p values
 - Adherence to reporting guidelines
 - Pre-registration, open data, open methods

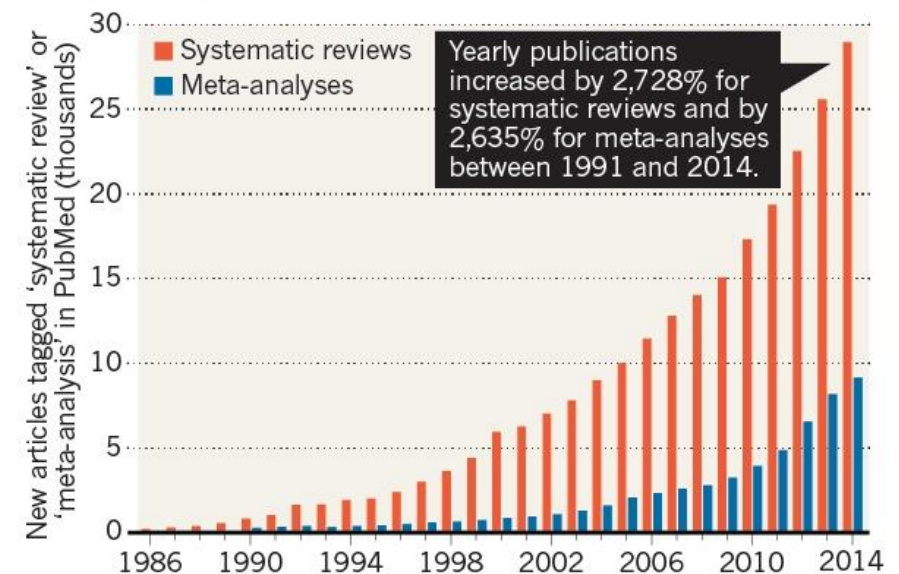
**What is
your n?**

Solution 4: considering the cumulative evidence

- More high quality evidence synthesis
- Strength of evidence rather than novelty
- Systems approach to funding
 - Informed by ES and informing ES
 - Common outcomes rather than novelty

META MASS PRODUCTION

The number of systematic reviews and meta-analyses published each year has proliferated since 1986.



A systematic review analyses and compiles all papers, and sometimes unpublished work, on a topic. A meta-analysis is a systematic review that combines data from multiple papers.

Acknowledgements

SRSM



*Society for Research
Synthesis Methodology*



<https://evidencesynthesisnewcastle.wordpress.com>

