Navigating preclinical research: directions from meta-research

Emily S Sena, PhD
Centre for Clinical Brain Sciences, University of Edinburgh

@drEmilySena
@camarades_

Disclosures

- **BMJ Open Science** (Editor-in-Chief)
  - I receive an honorarium for this role

- I applied and have received (& will continue) grant funding for this research

Me

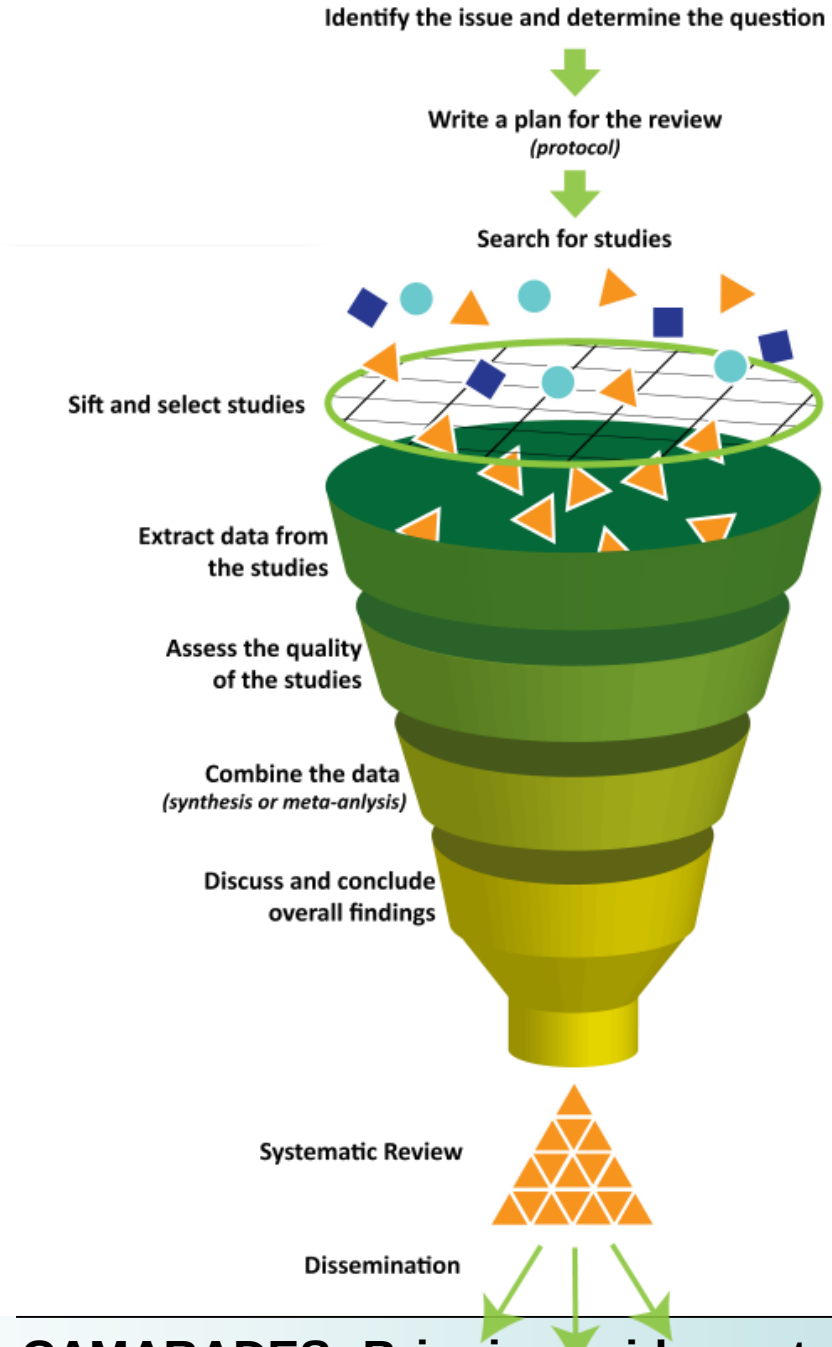
- **2005:** BSc (Hons) Pharmacology, University of Edinburgh
- **2010:** PhD *"Systematic review and meta-analysis of animal models of acute ischaemic stroke*" Universities of Edinburgh & Melbourne
- **2018**: Senior Non-Clinical Lecturer (stroke association)

Collaborative Approach to Meta Analysis and
·C·A·M·A·R·A·D·E·S·
Review of Animal Data from Experimental Studies

- Look systematically across the modelling of a range of conditions
- Data Repository
  - 30 Diseases/40 Projects/25,000 studies/over 400,000 animals

Concept

- Systematic review sets out to use a structured process to identify all data relevant to a specific research question.
- May be followed by meta-analysis, a statistical process that provides a summary estimate of the outcomes from a group of studies
- These outcomes from different groups of studies may be compared.


Identify the issue and determine the question
Write a plan for the review
(protocol)
Search for studies
Sift and select studies
Extract data from the studies
Assess the quality of the studies
Combine the data
(synthesis or meta-anlysis)
Discuss and conclude overall findings
Systematic Review
Dissemination

Why do we do meta-analysis of animal studies?

- Preclinical studies are often performed with the purpose of improving human health
- Used in preclinical research to:
  - assess the quality and range of evidence
  - identify gaps in the field
  - assess for publication bias
  - try to explain discrepancies between preclinical and clinical trial results
  - inform clinical trial design
- Fundamental differences:
  - Many small (10s) animal studies
  - Fewer large (100s/1000s) clinical trials

Research & Bias

- All research is susceptible to bias
- Systematic reviews and meta-analyses
  - Developing field of research with a number of different approaches to its conduct and reporting
  - provide empirical evidence to spur the field to improve the rigor (& reduce bias) of preclinical research
- Our worst bias is meta
  - being more aware of biases makes us more willing to assume that others' biases, and not ours, are responsible for our disagreement

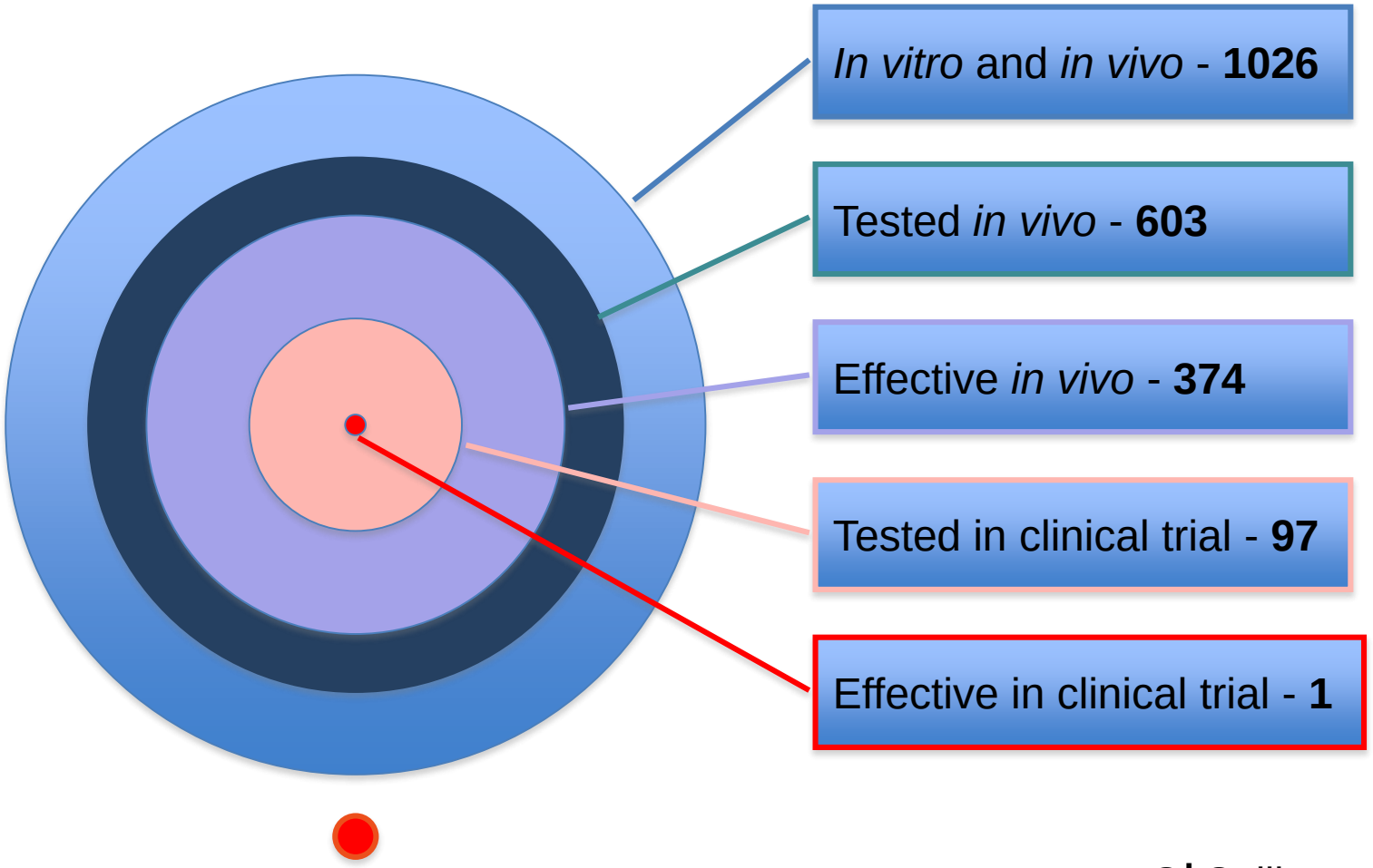
Outline

- Translational failure
- Internal validity
- External/construct validity
- Reporting bias
- Research improvement activities

Hypotheses

- In the life sciences there are perverse incentives (publication, funding, promotion) to produce positive results with little attention paid to their validity
- In the use of animal disease models, pressure to reduce the number of animals (cost, time, ethics, feasibility) results in studies either being underpowered or of unknown power
- These factors combine to compromise the utility of animal models and contribute to translational failure

1,026 interventions in experimental stroke

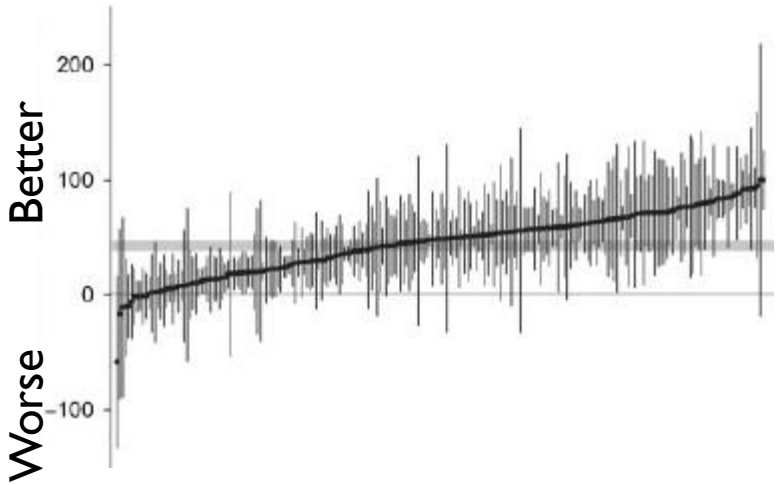


O'Collins et al, 2006

Animal data in stroke

- There are huge amounts of often confusing data
- Systematic review can help to make sense of it
- If you select extreme bits of the evidence you can "prove" either harm or substantial benefit
- Investigating the sources behind this variation may be helpful in translation

Hypothermia: a systematic search identified 222 experiments in 3353 animals



Van der Worp *et al Brain* 2007

Translational failure

Improving the translational hit of experimental treatments in multiple sclerosis

Multiple Sclerosis
0(00) 1–12

DOI: 10.1177/1352458510379612
msj.sagepub.com

Hanna M. Ve
Charles ffren
Siddharthan

PAIN® 154 (2013) 917–926

IASP

PAIN®

ORIGINAL ARTICLE

Animal mo

Gillian L. Curri
Hanna M. Vest

Treatment of Intracerebral Hemorrhage in Animal Models: Meta-Analysis

Review

Cell PRESS

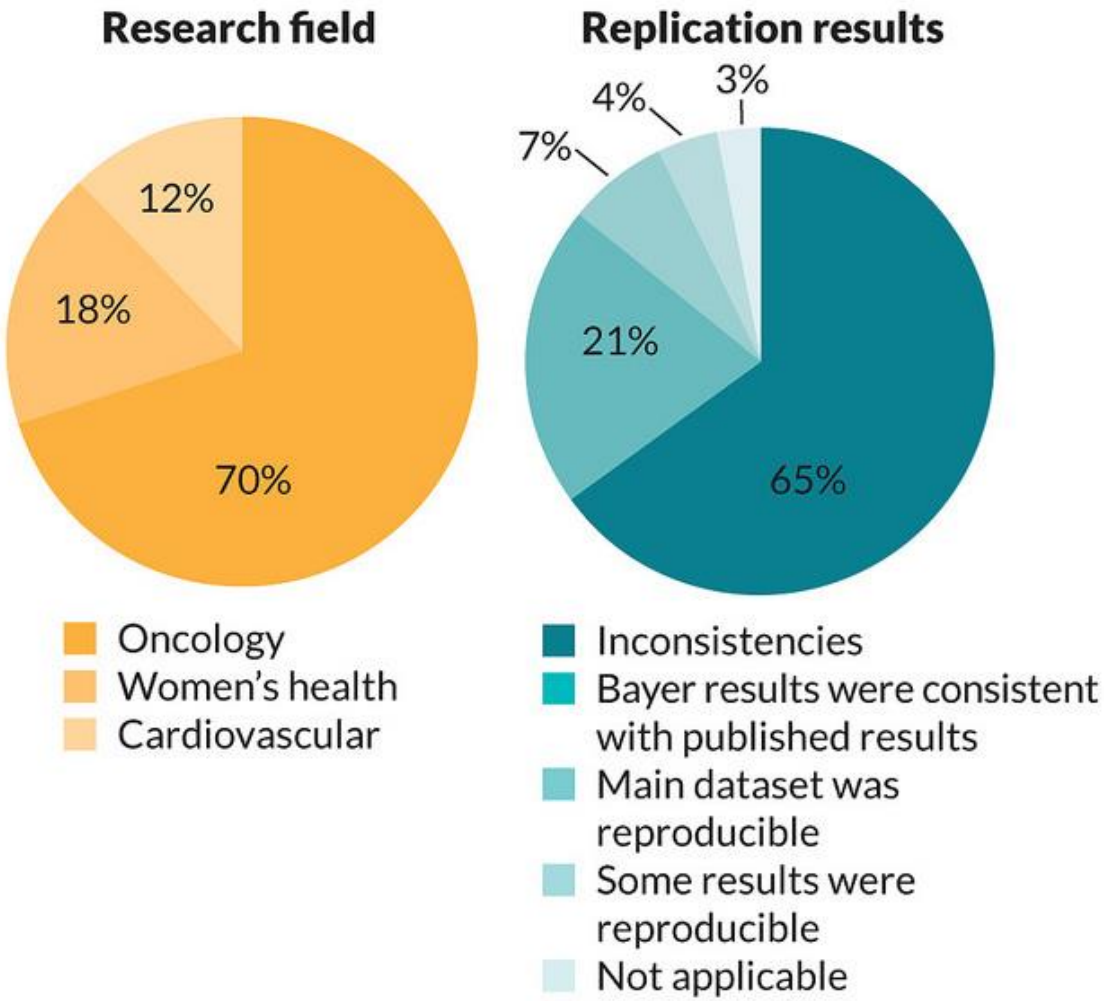
'Too much good news' – are Alzheimer mouse models trying to tell us how to prevent, not cure, Alzheimer's disease?

Kathleen R. Zahs[1,2,4] and Karen H. Ashe[1,2,3,5]

[1] N. Bud Grossman Center for Memory Research and Care
[2] Department of Neurology
[3] Department of Neuroscience
[4] Department of Integrative Biology and Physiology, University of Minnesota Medical School, Minneapolis, MN 55455, USA
[5] Geriatric Research Education Clinical Center, Minneapolis VA Medical Center, Minneapolis, MN 55417, USA

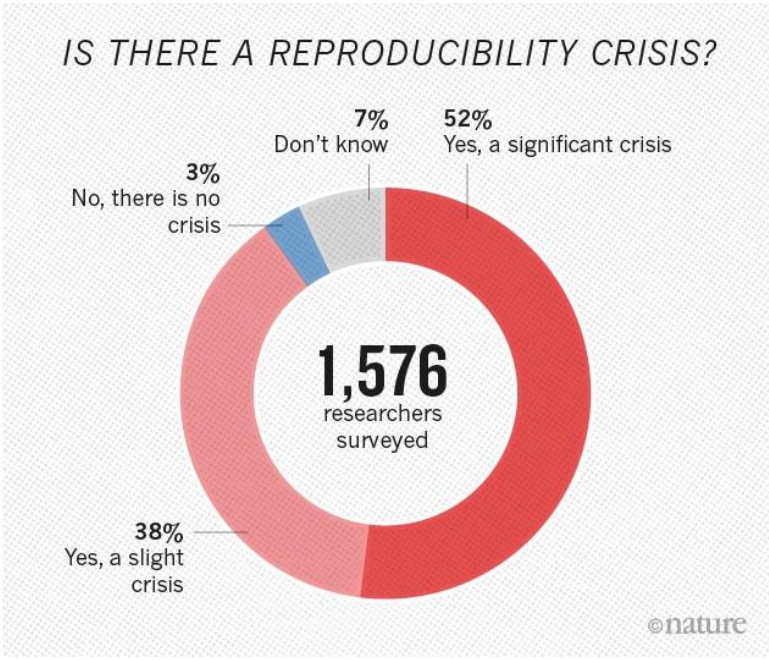
What happens when pharma tries to replicate academic findings?

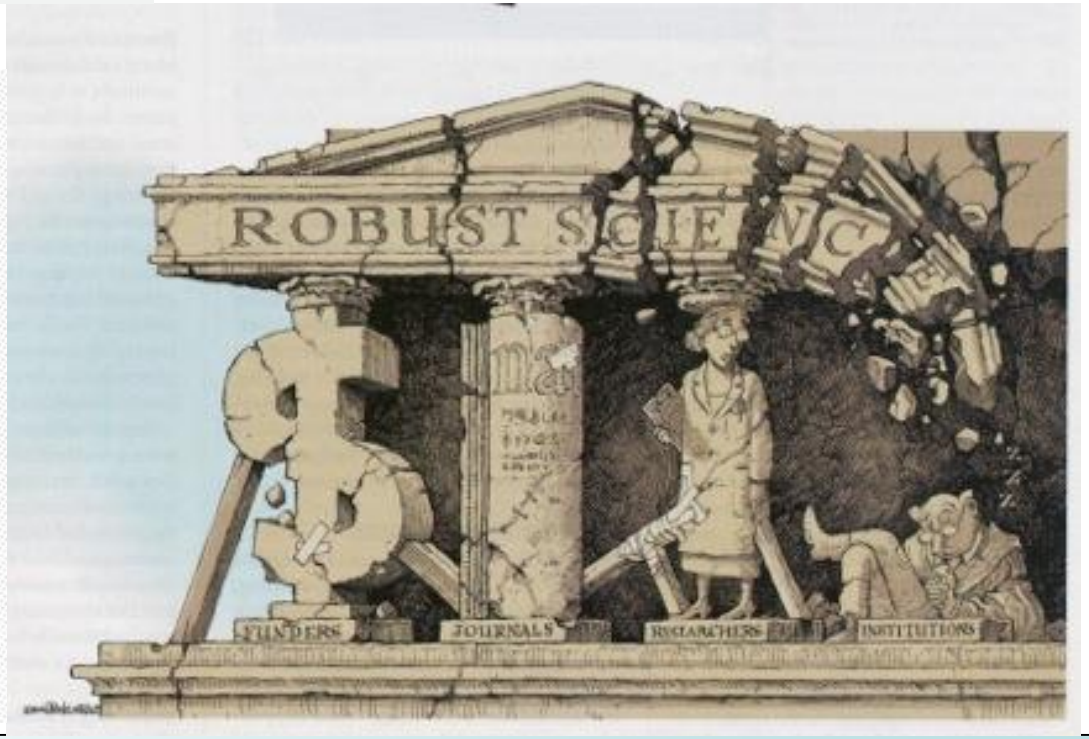
- Bayer, Berlin
- 67 in-house projects over 4 years



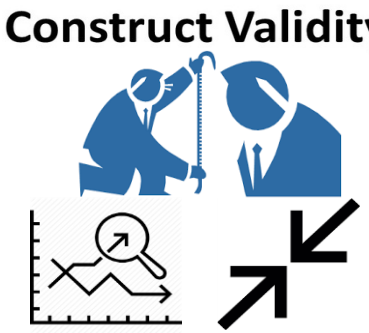
Prinz *et al*, Nature Reviews Drug Discovery, 2011

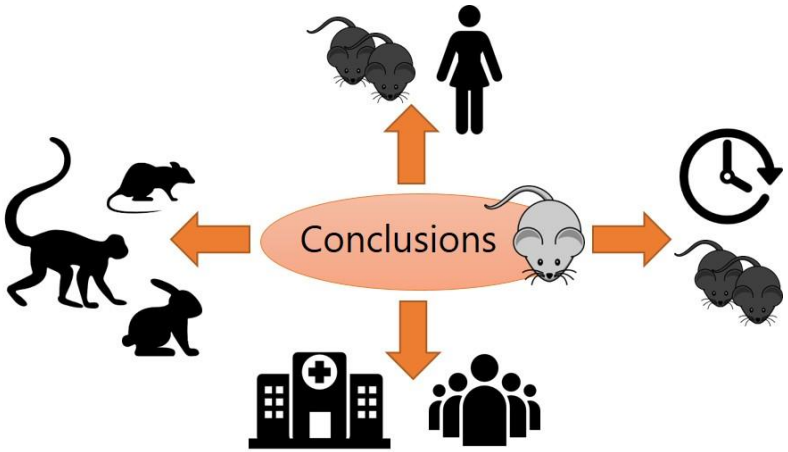
Reproducibility crisis





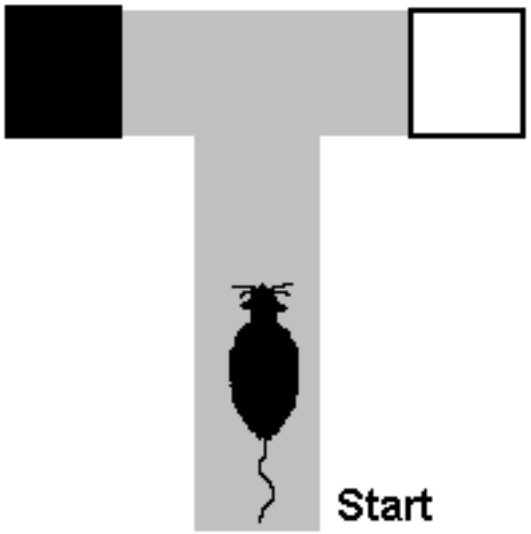
Some potential sources of bias





You can usually find what you're looking for …

- 12 graduate psychology students
- 5 day experiment: rats in T maze with dark arm alternating at random, and the dark arm always reinforced
- 2 groups – "Maze Bright" and "Maze dull"



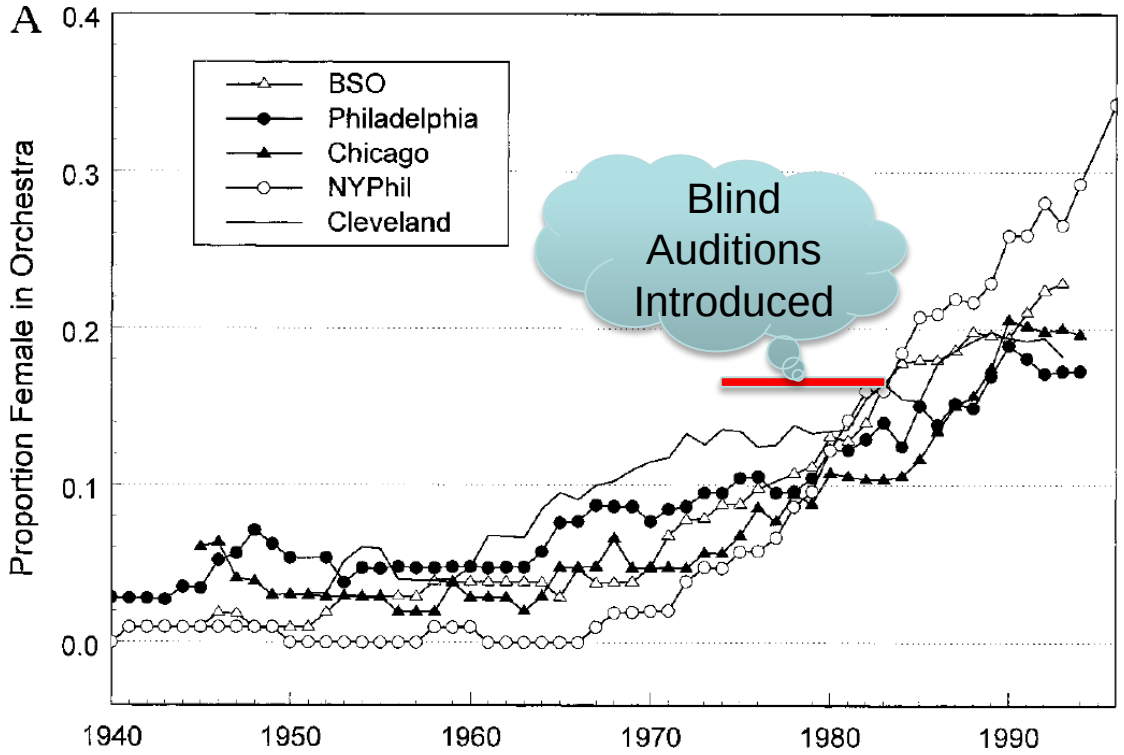
Group	Day 1	Day 2	Day 3	Day 4	Day 5
"Maze bright"	1.33	1.60	2.60	2.83	3.26
"Maze dull"	0.72	1.10	2.23	1.83	1.83
Δ	+0.60	+0.50	+0.37	+1.00	+1.43

Rosenthal and Fode (1963), Behav Sci 8, 183-9

Zubin Mehta

Conductor of the LA Symphony (1964-1978) and NY Philharmonic (1978-1990) credited with saying, ***"I just don't think women should be in an orchestra."***

www.curt-rice.com

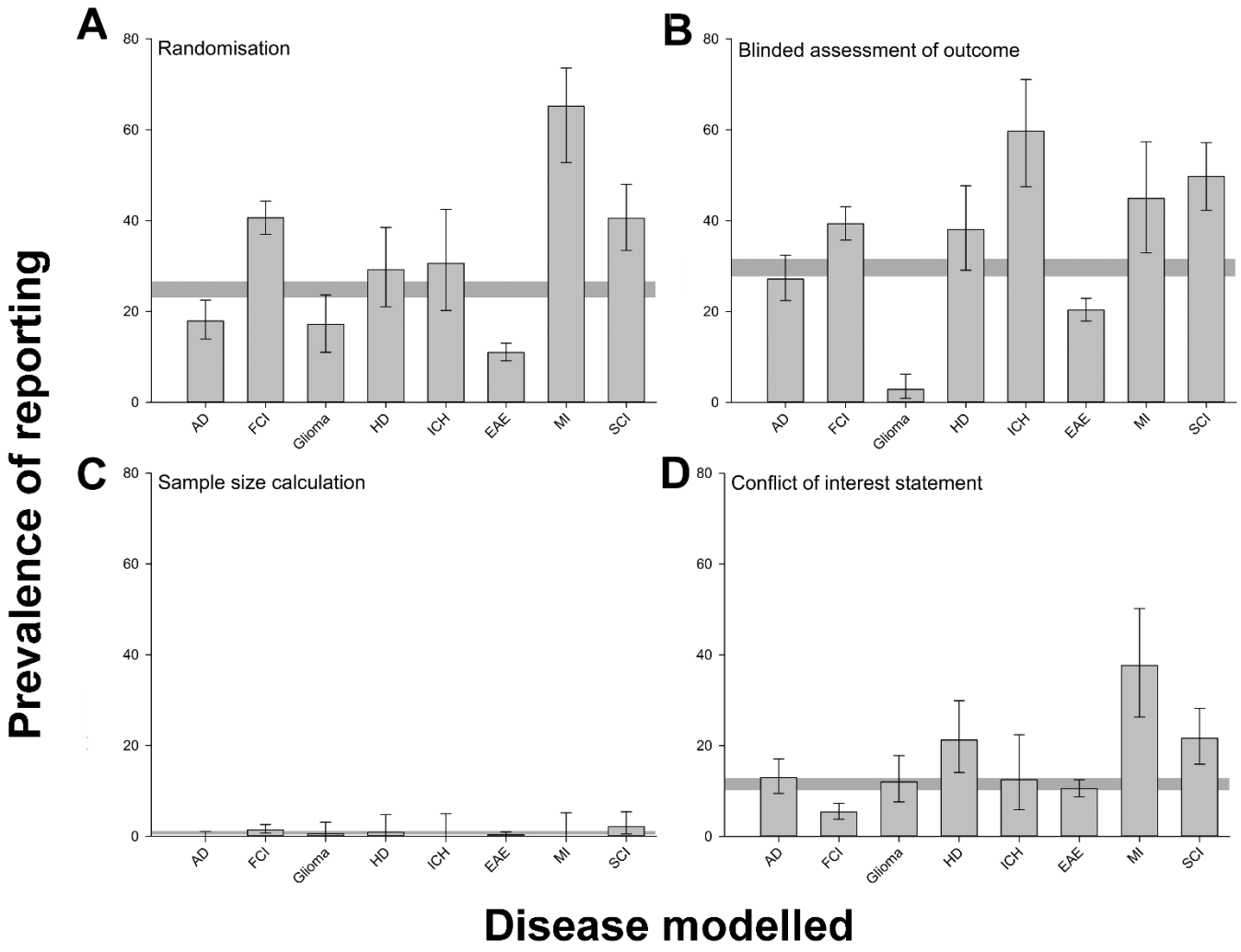


Goldin & Rouse (2000) American Economic Review 90: 715

Blind auditions explain ca. 30% of the increase in the female proportion of "new hires" at major symphony orchestras in the US

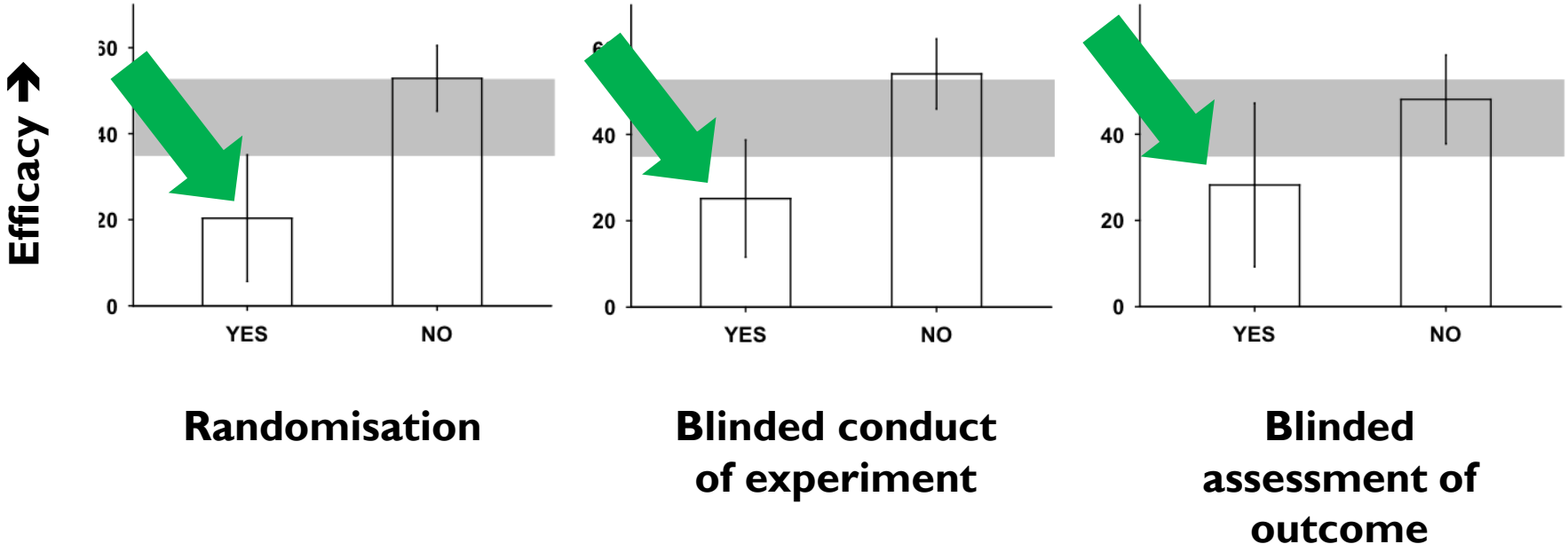
Non-random sample

- N=2671
- Most published *in vivo* research is at high risk of bias



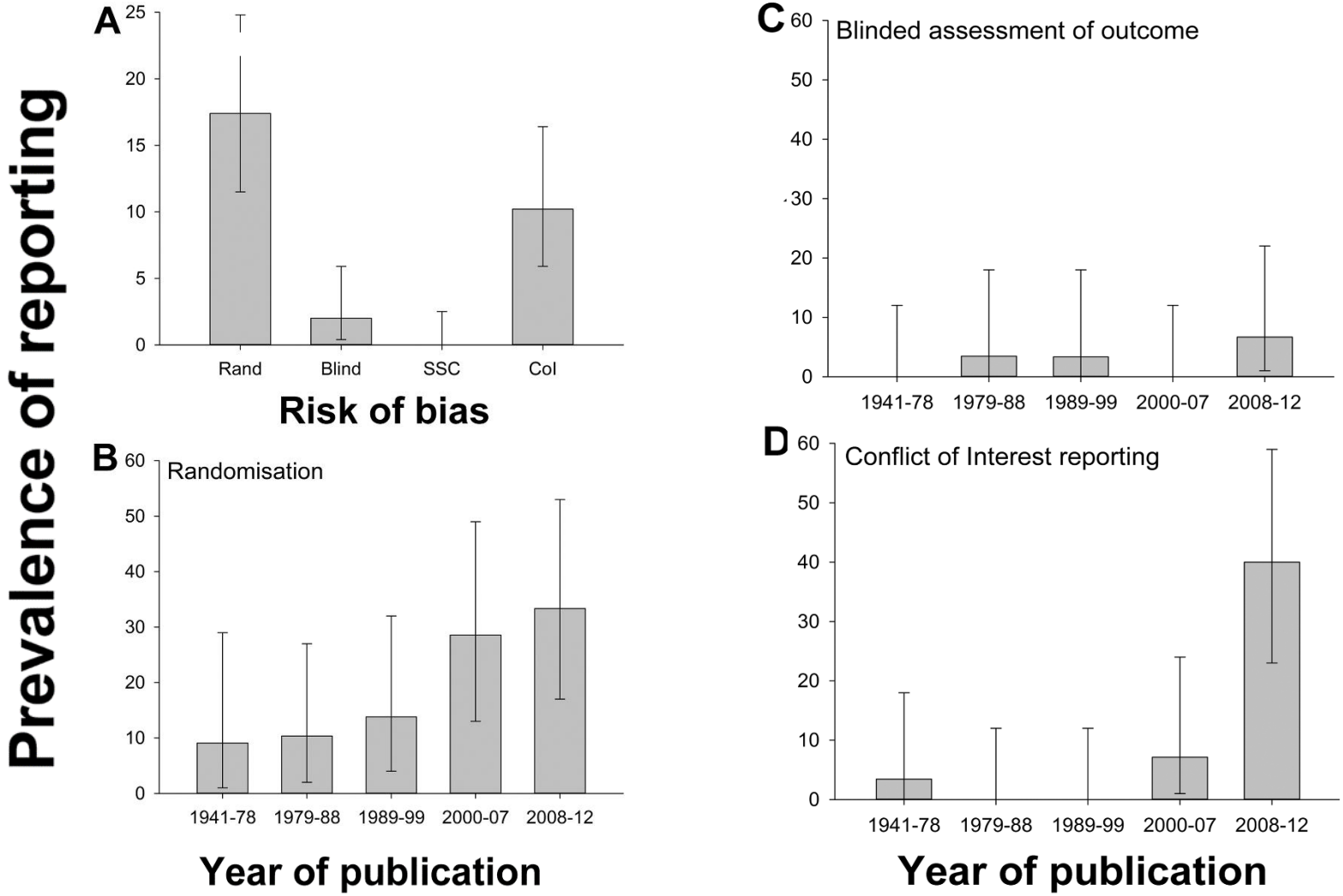
Internal Validity:
Lessons from NXY-059

- Infarct Volume
  - 11 publications, 29 experiments, 408 animals
  - Improved outcome by 44% (35-53%)

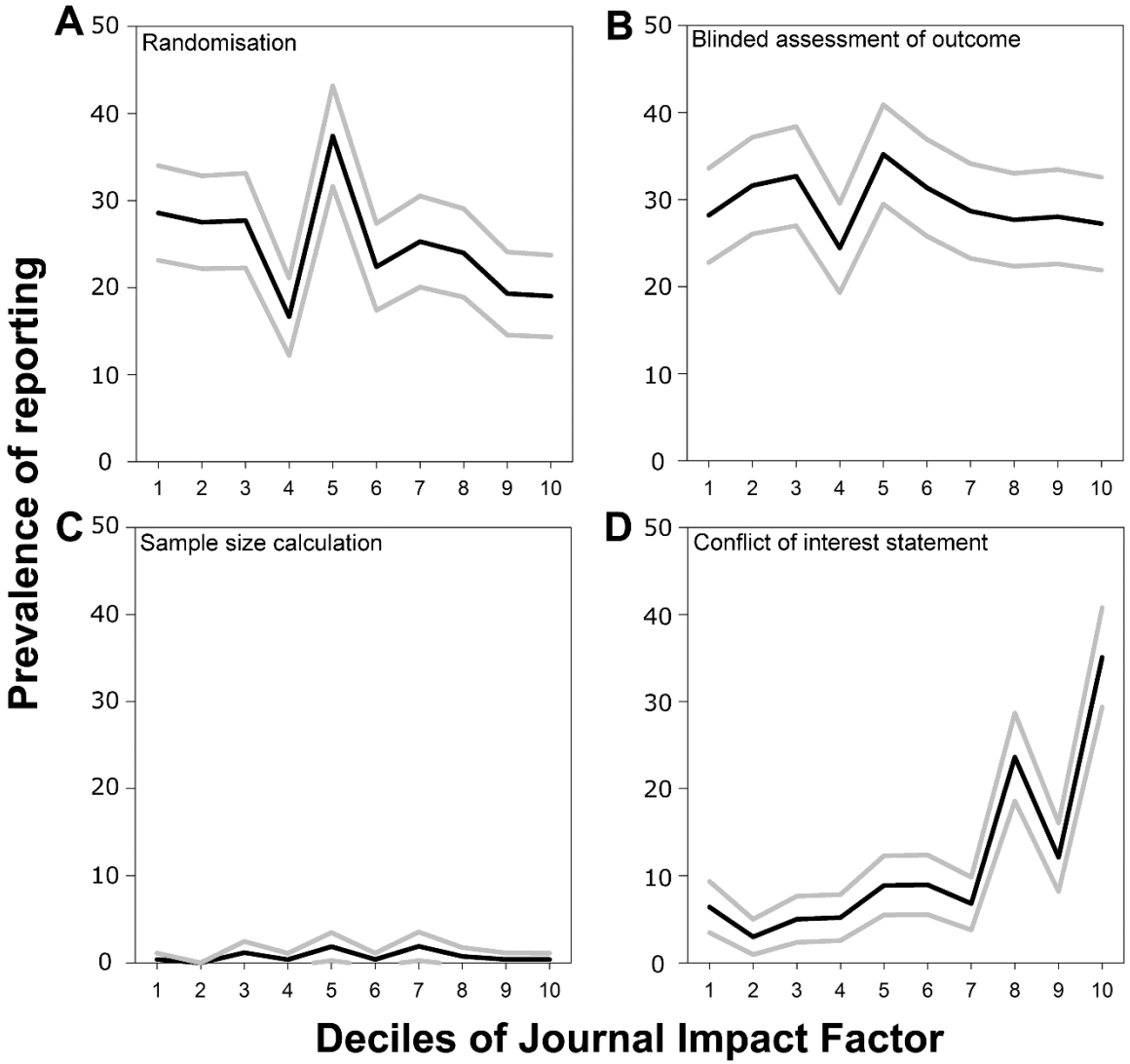


Macleod *et al*, 2008

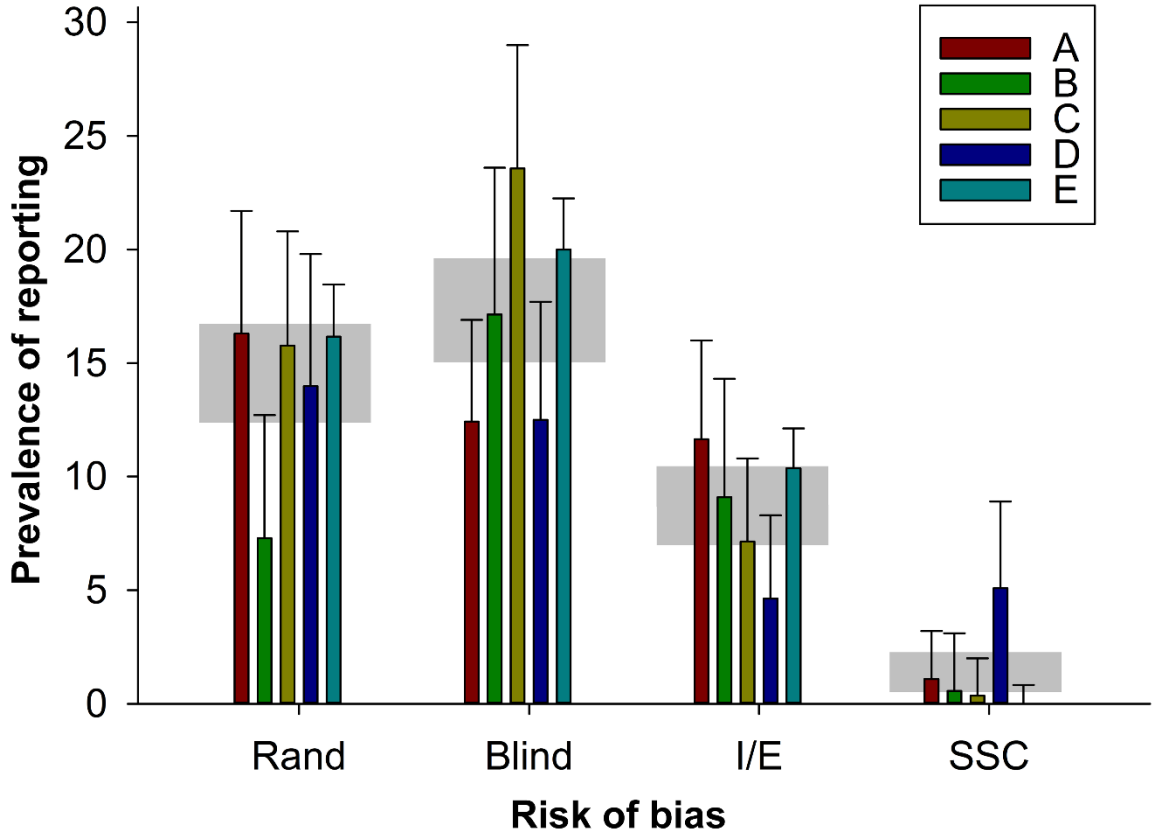
Random sample from PubMed



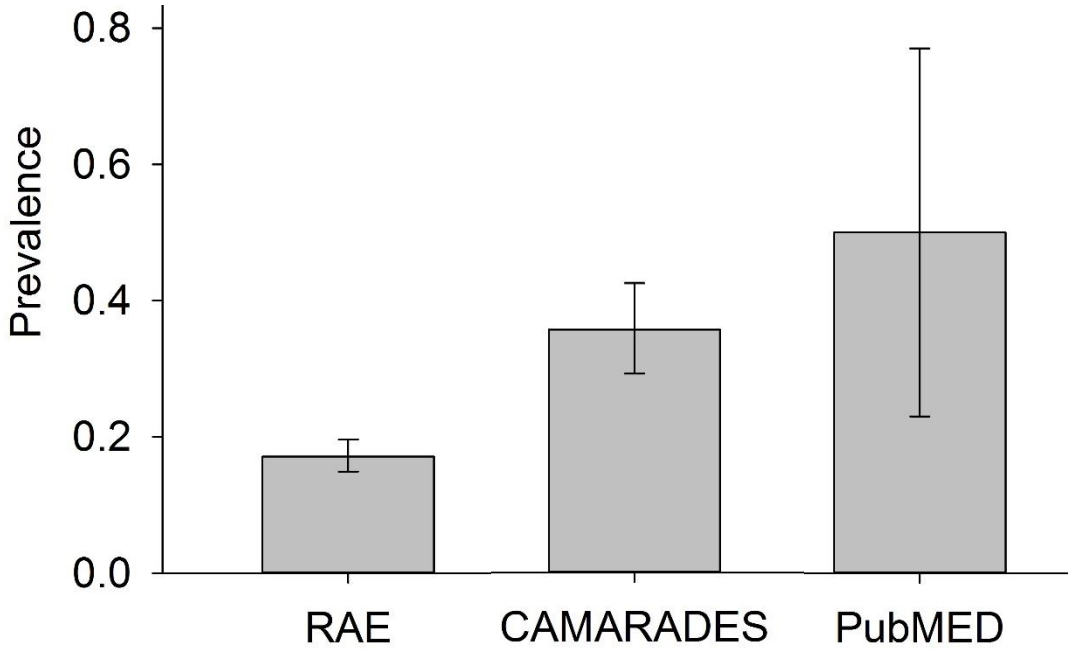
Good ‘quality’ journals



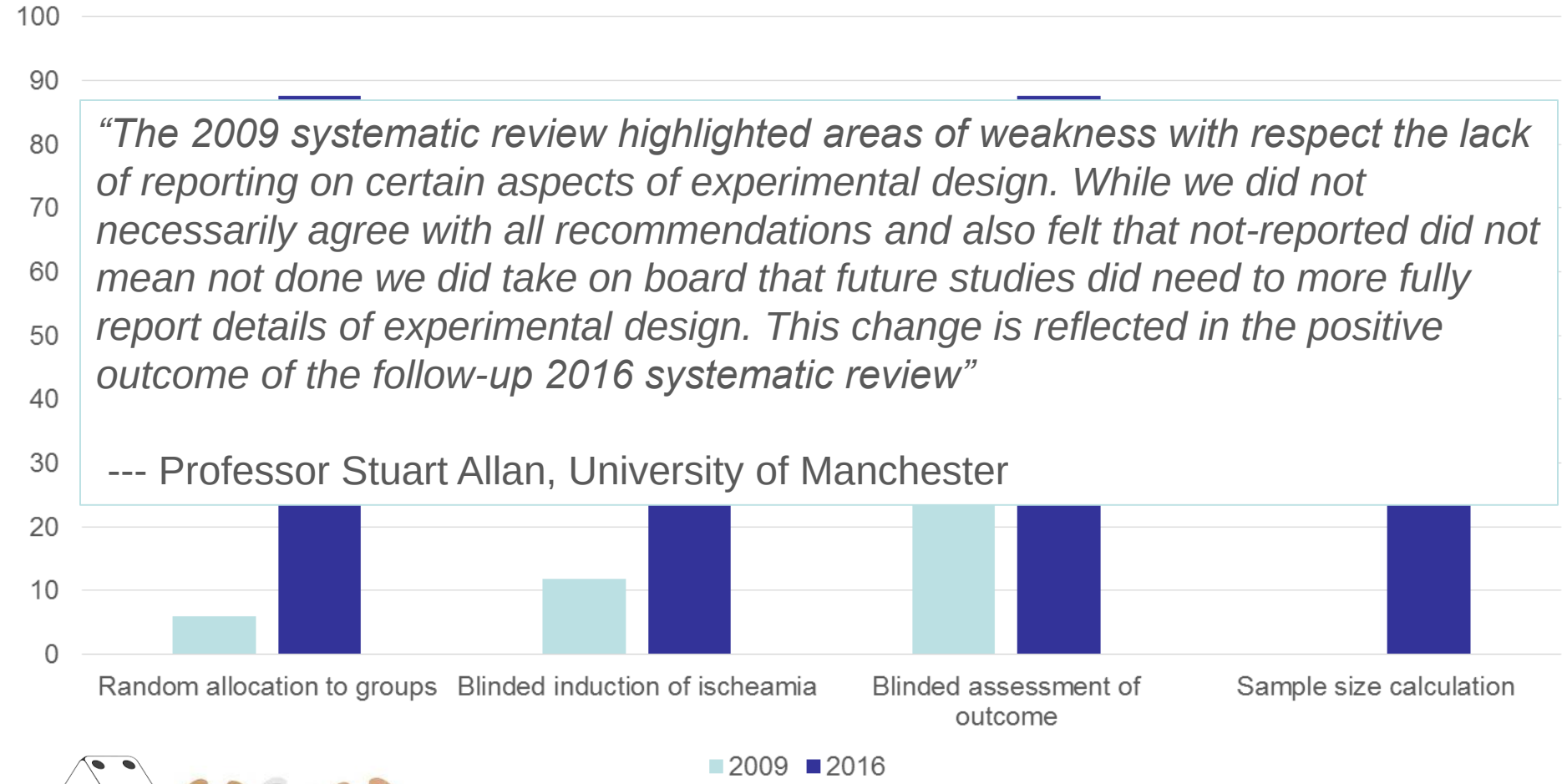
The ‘best’ institutions – RAE 1173



Reporting of randomisation across 3 datasets, 2009-10

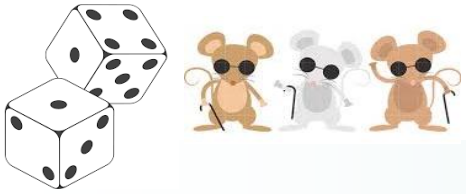


Things are improving



"The 2009 systematic review highlighted areas of weakness with respect the lack of reporting on certain aspects of experimental design. While we did not necessarily agree with all recommendations and also felt that not-reported did not mean not done we did take on board that future studies did need to more fully report details of experimental design. This change is reflected in the positive outcome of the follow-up 2016 systematic review"

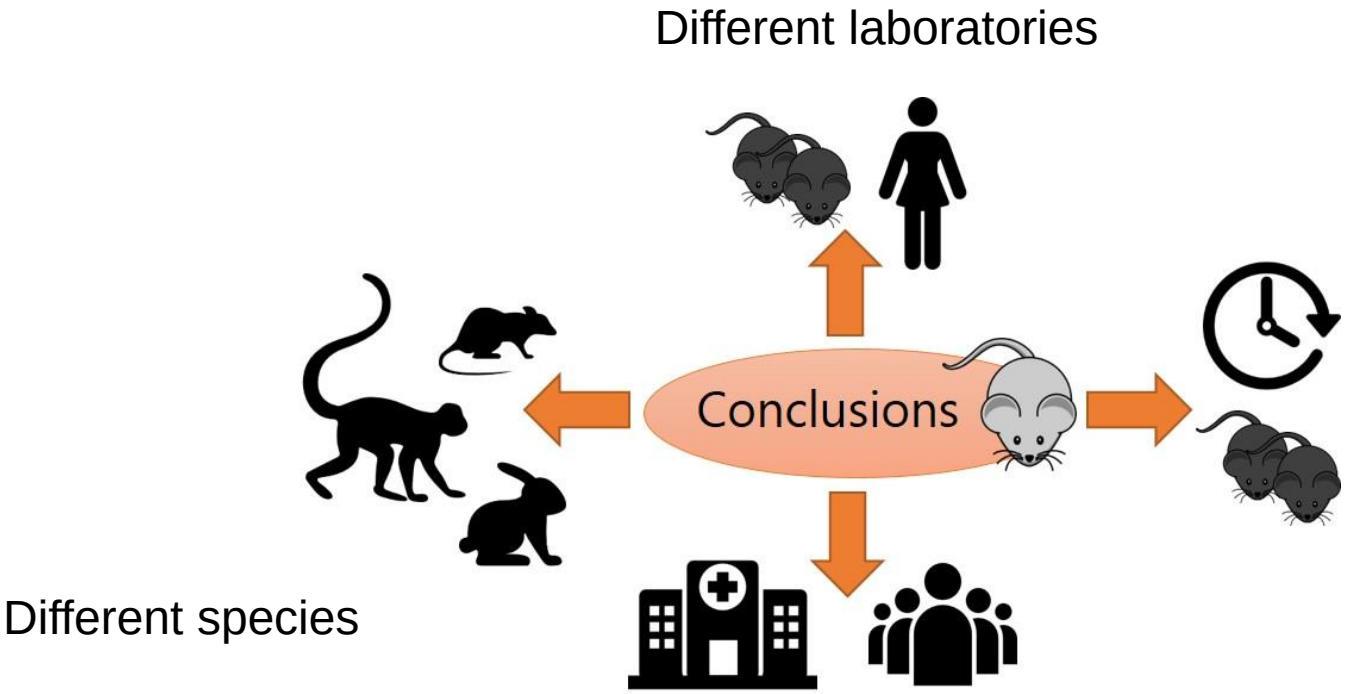
--- Professor Stuart Allan, University of Manchester



McCann SK, Cramond F, Macleod MR, Sena ES (2016). *Translational stroke research* **7**(5)**:** 395-406.

Why is external validity important?

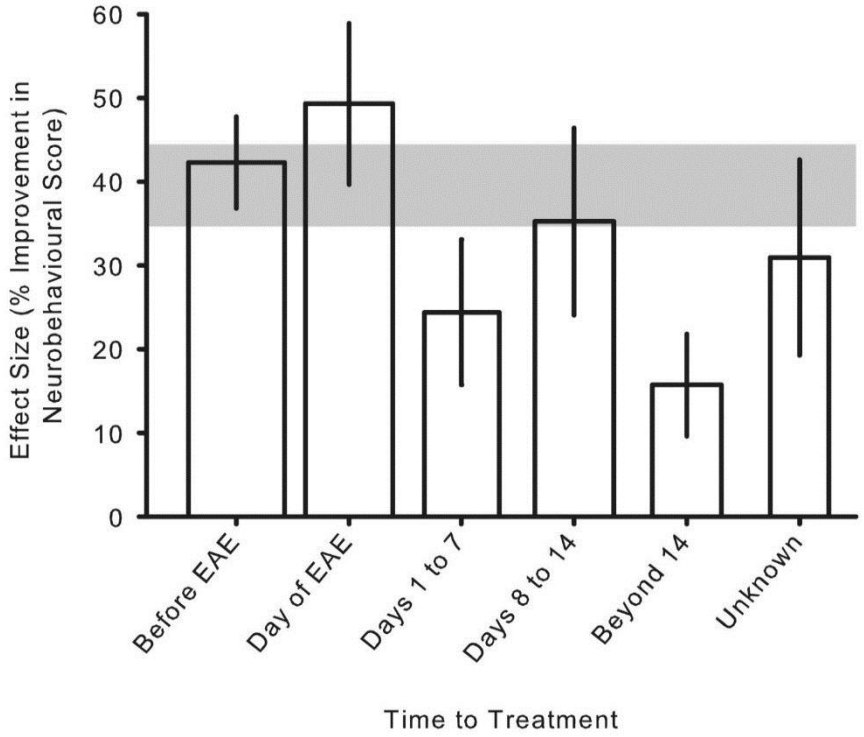
External validity influences our ability to reproduce effects across:



Time to Treatment in EAE

- Median: 0 days (IQR -11 to 4)
- 1% did not report time of administration

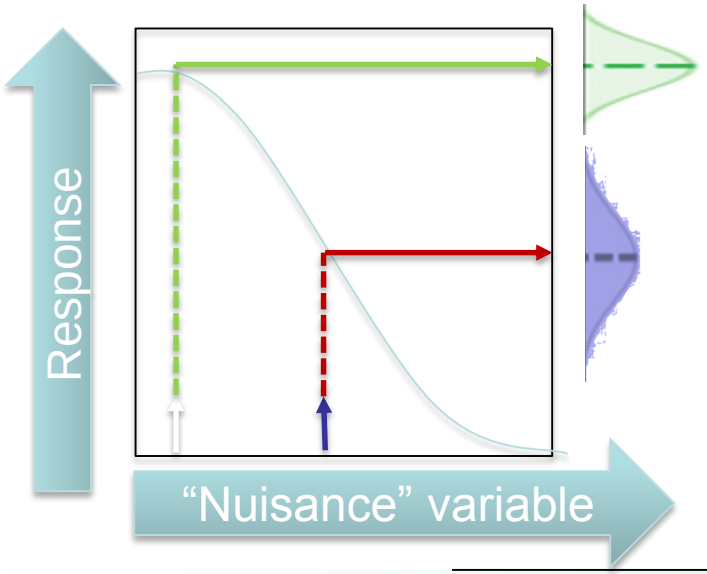
Before EAE	48%
Day of Induction	22%
After EAE	30%
Day of Symptom Onset	2%

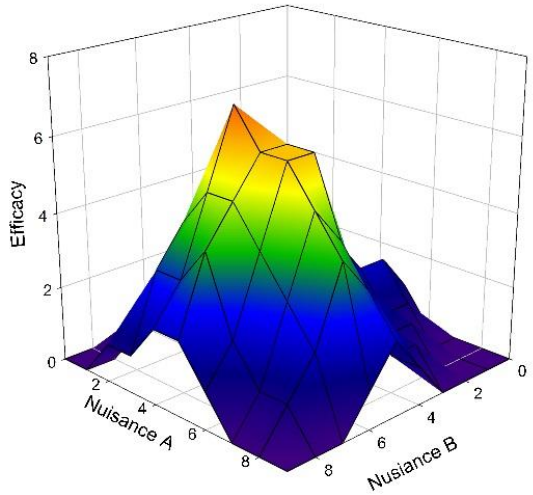


Vesterinen et al 2010 Multi Scler.

The standardisation fallacy

- Efforts to increase reproducibility by reducing variation by standardisation of:
  - lab environment
  - tests used
  - genetics of the animals
- Increases the risk of detecting effects with **low external validity** (or of missing effects with high external validity)



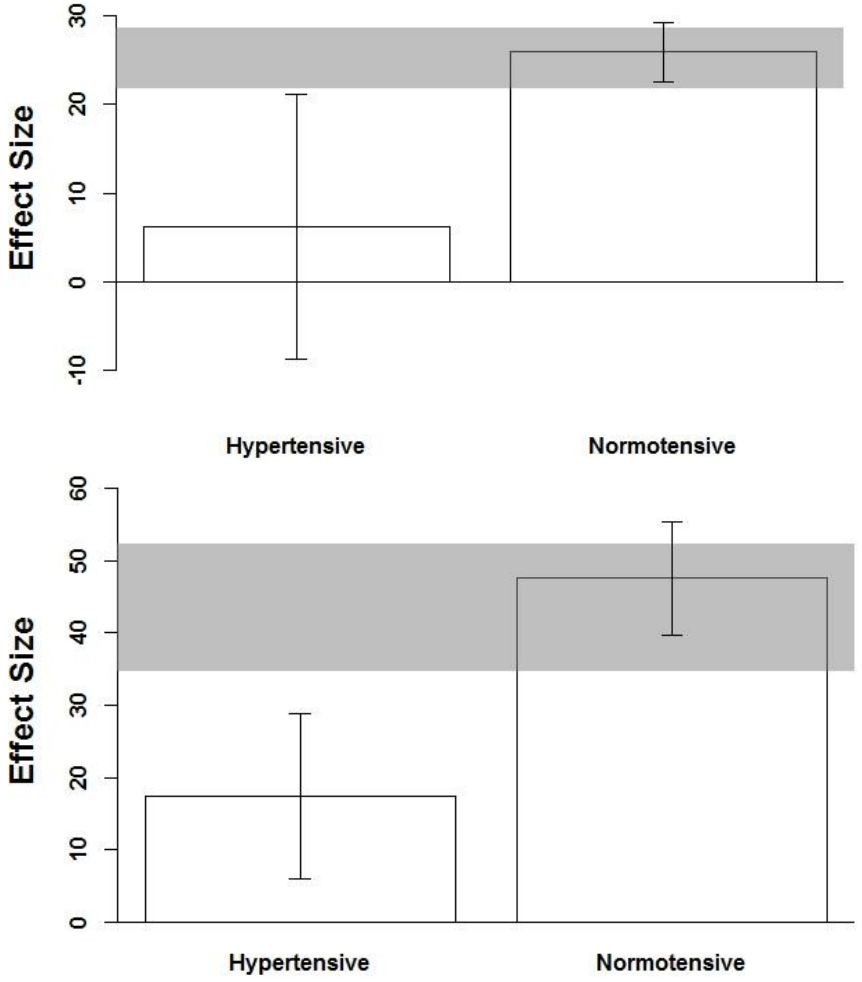


Wurbel, H. (2000) Nat. Genet
Voekl 2016 PLOS Biolgy

Construct validity

Hypertension in studies of experimental stroke

- High prevalence of hypertension in patients with stroke
- **tPA:**
  - substantially less effective in hypertensive animals
- **NXY-059:**
  - 7% of animal studies
  - 77% of patients in the (neutral) SAINT II study



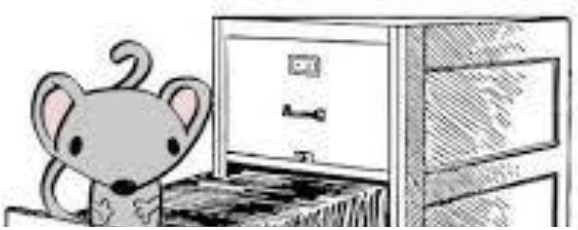
Construct Validity

Macleod *et al* Stroke 2009
Sena *et al JCBFM* 2010

The umbrella of reporting bias

Not all outcomes and *a priori* analyses are reported

- Publication bias
  - Neutral and negative studies
  - Time lag/remain unpublished
  - Less likely to be identified
- p-hacking
  - Selective analysis
  - Selective outcome repo

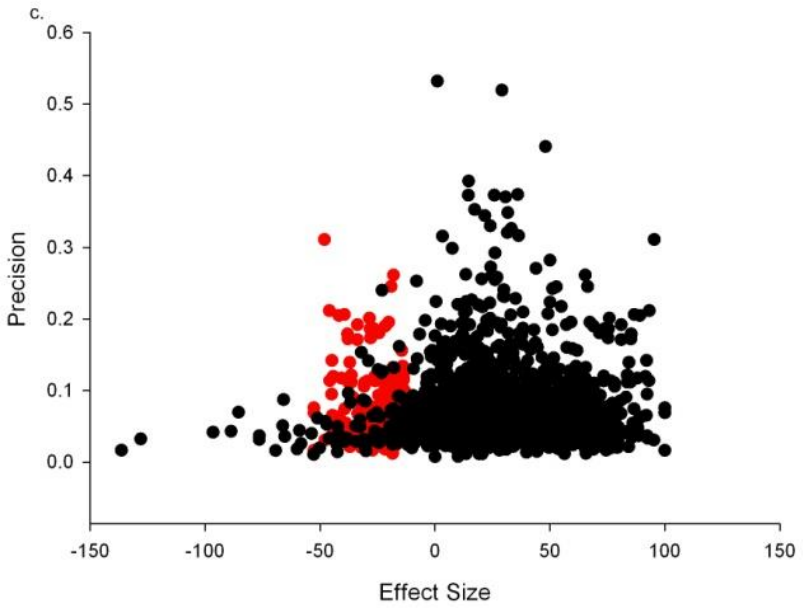


OPEN ACCESS Freely available online

PLoS BIOLOGY

Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy

Emily S. Sena[1,2,3], H. Bart van der Worp[4], Philip M. W. Bath[5], David W. Howells[2,3], Malcolm R. Macleod[1,6]*



- Overall efficacy was reduced from;
  - **32%** (95% CI 30 to 34%) to **26%** (95% CI 24 to 28%)
- 16% of experiments remain unpublished

Publication bias

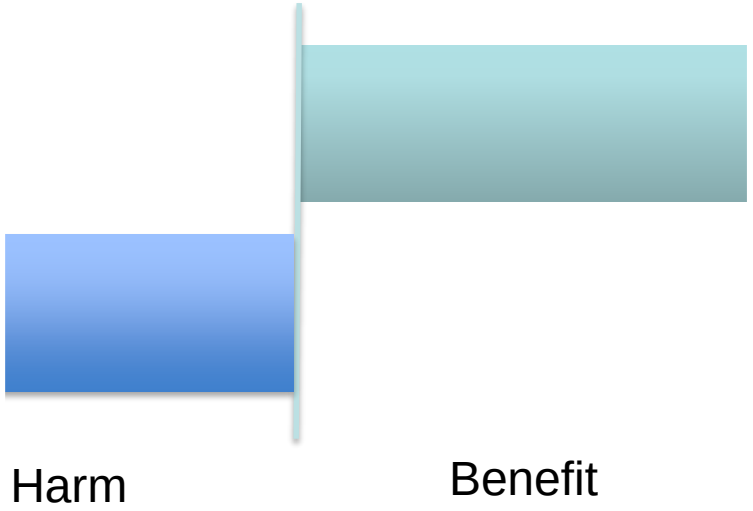
20%

- 32%

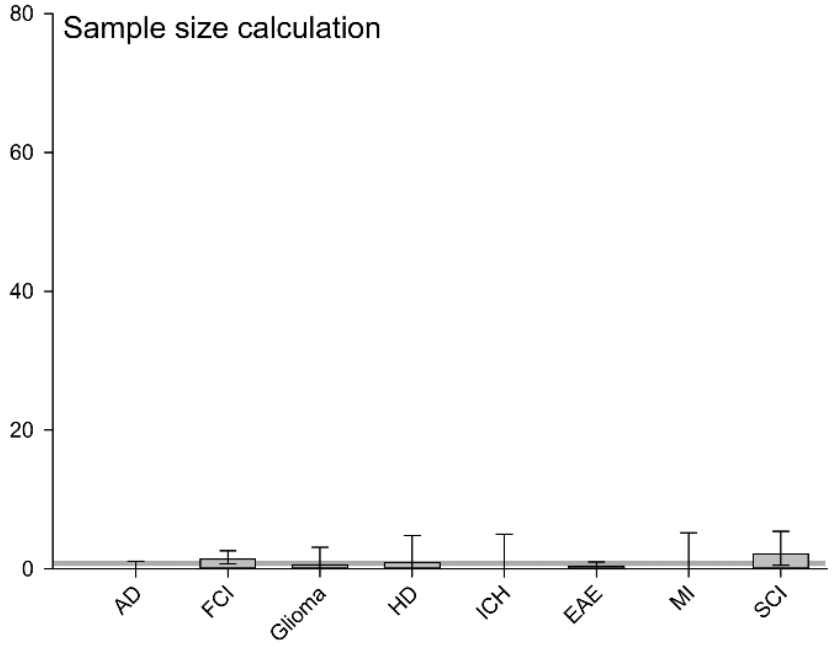
	n expts	Estimated unpublished	Reported efficacy	Corrected efficacy
Stroke – infarct volume	1359	214	31.3%	27.5%
EAE - neurobehaviour	1892	505	33.1%	15.0%
EAE – inflammation	818	14	38.2%	37.5%
EAE – demyelination	290	74	45.1%	30.5%
EAE – axon loss	170	46	54.8%	41.7%
AD – Water Maze	80	15	0.688 sd	0.498 sd
AD – plaque burden	632	154	0.999 sd	0.610 sd

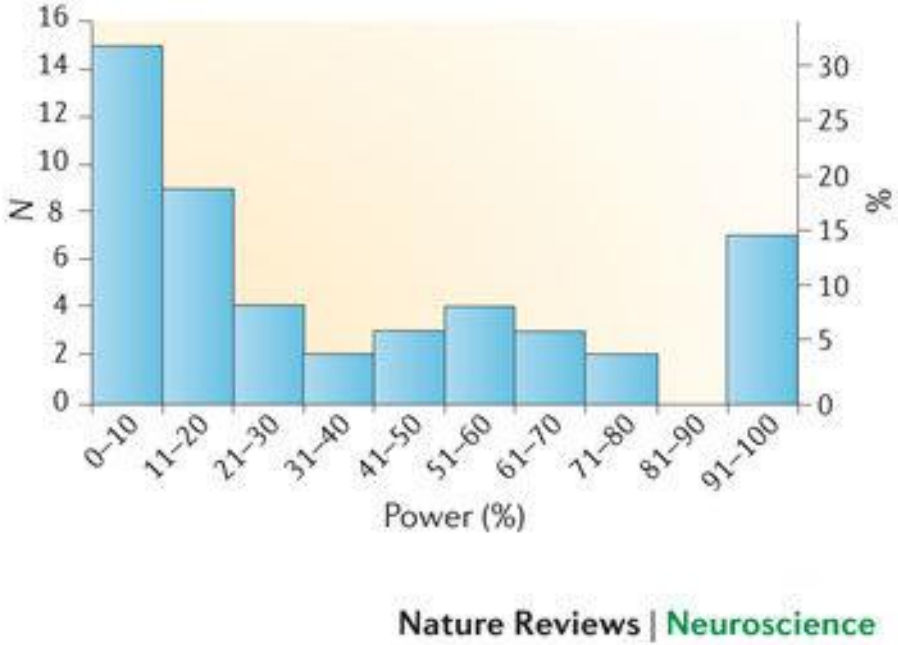
Different patterns of publication bias in different fields

	outcome	observed	corrected	
Disease models	improvement	40%	30%	Less improvement
Toxicology model	harm	0.32	0.56	More harm



Power problem…..

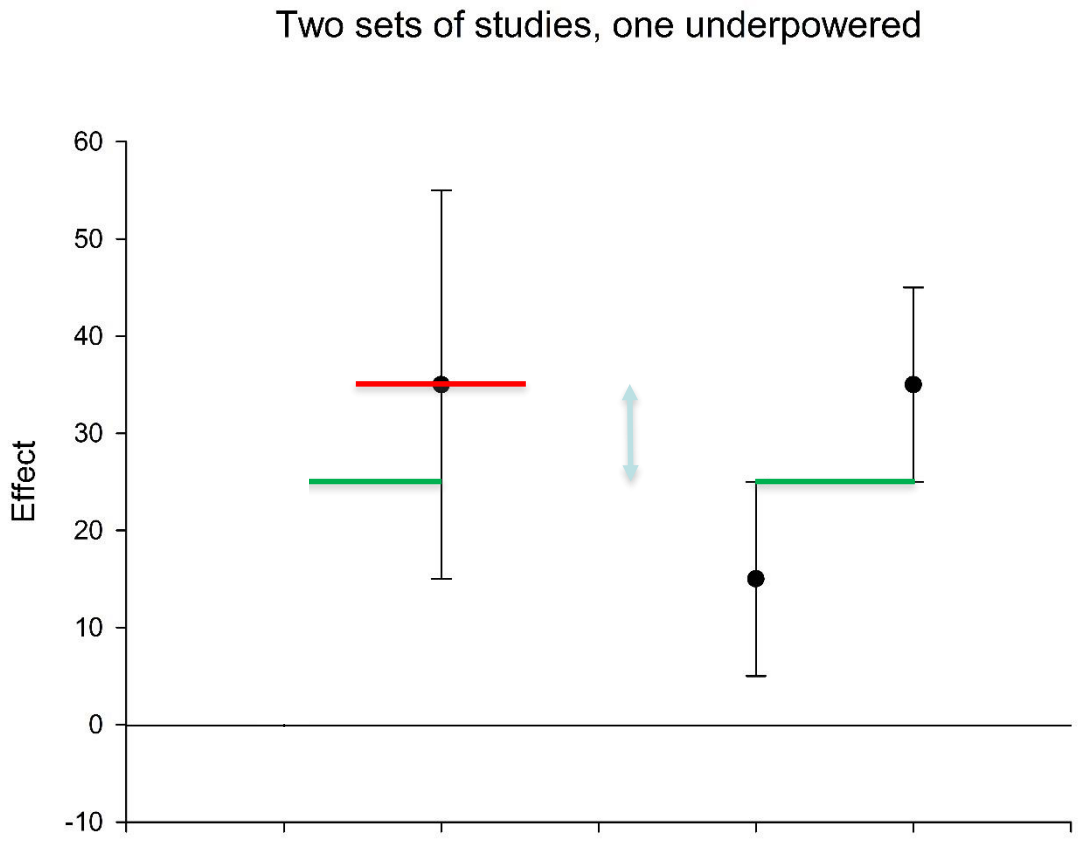




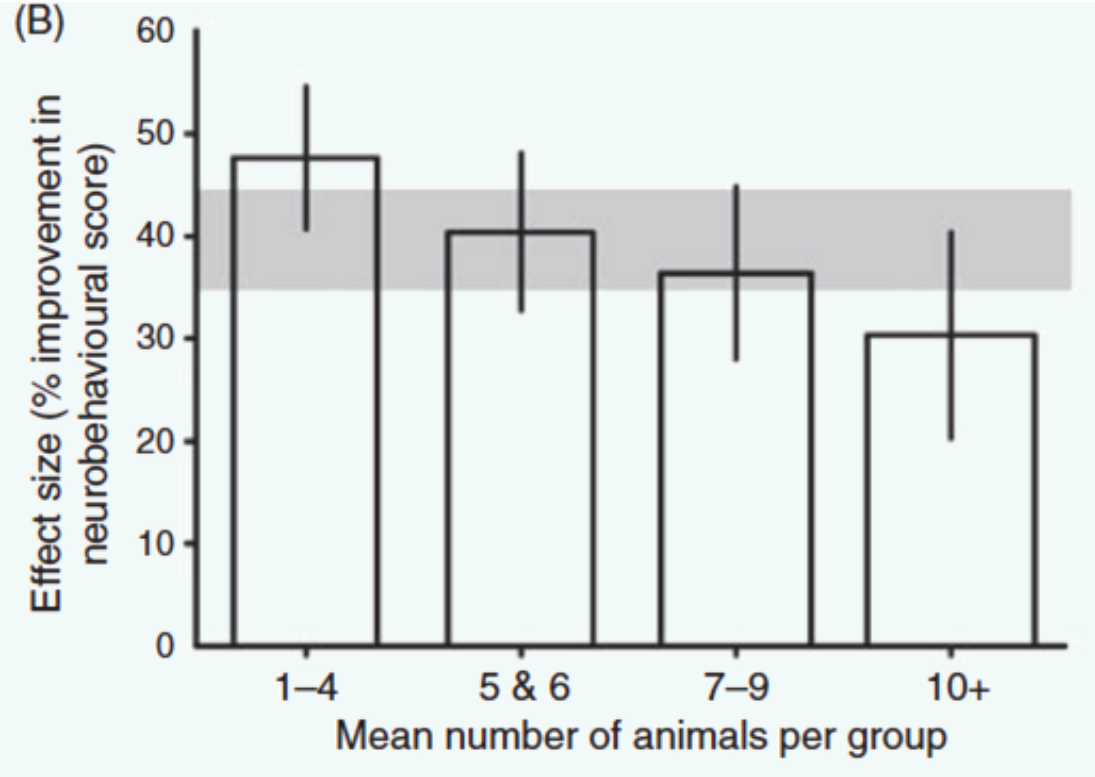
Macleod PLOS Biol 2015

Button Nat Rev Neuro 2013

Small group sizes and publication bias conspire together



Small group sizes and publication bias conspire together



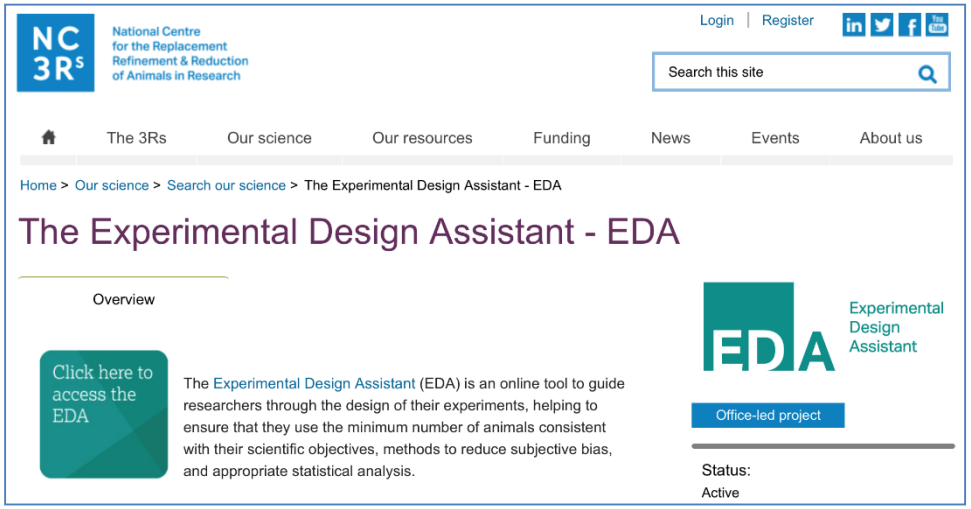
Vesterinen Mult Scler 2010

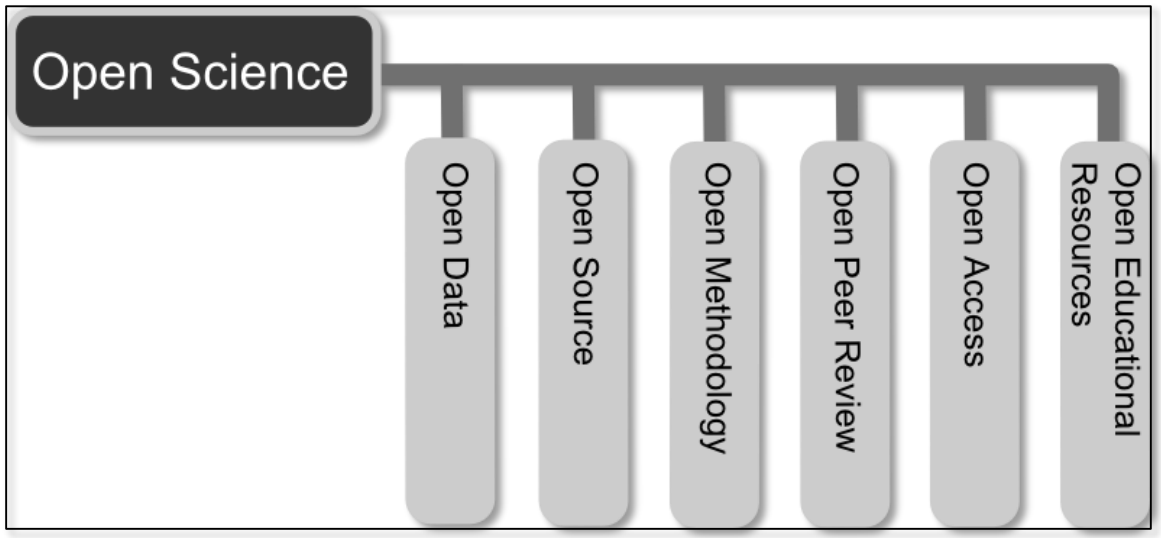
Research Improvement

- **Research Improvement Activity:** Things done by stakeholders to increase the usefulness of research with which they are associated
- How can we assess whether interventions improve research quality and reduce waste?
  - Observational studies
  - Randomised controlled trials

Improvement strategies

- Design
  - EDA
  - Multi-centre studies
  - Statistical input
- Transparency





Improvement strategies


ARRIVE

The ARRIVE Guidelines Checklist

Animal Research: Reporting In Vivo Experiments

Carol Kilkenny[1], William J Browne[2], Innes C Cuthill[3], Michael Emerson[4] and Douglas G Altman[5]

[1]The National Centre for the Replacement, Refinement and Reduction of Animals in Research, London, UK, [2]School of Veterinary Science, University of Bristol, Bristol, UK, [3]School of Biological Sciences, University of Bristol, Bristol, UK, [4]National Heart and Lung Institute, Imperial College London, UK, [5]Centre for Statistics in Medicine, University of Oxford, Oxford, UK.

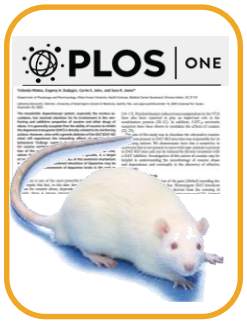
- Reporting
  - ARRIVE guidelines
  - Landis transparency guidelines
  - NPG publication policy
- Publication
  - Support new models (registered reports)
  - Encourage rapid publication anywhere, not vanity publishing in journals of the highest “impact”
- Careers
  - Appointment panels should look at the work, not where it was published
  - Emphasising rigour in grant award
  - CPD opportunities for scientists

Registered reports

- Peer review before data collection

- Emphasise the importance of the research question and the quality of methodology
- Designed to eliminate: low statistical power, selective reporting of results, and publication bias, while allowing complete flexibility to report serendipitous findings.

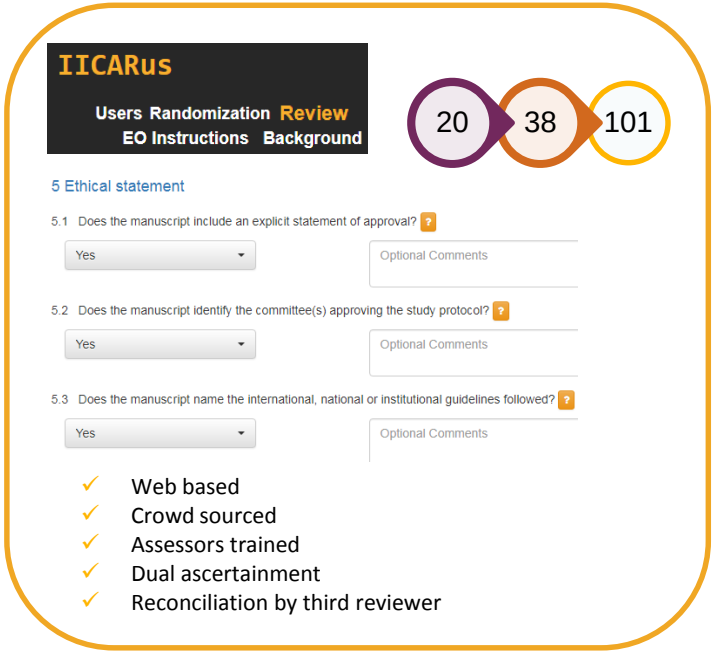
Impact of an Intervention to Improve Compliance With the ARRIVE Guidelines (IICARus)



IICARus

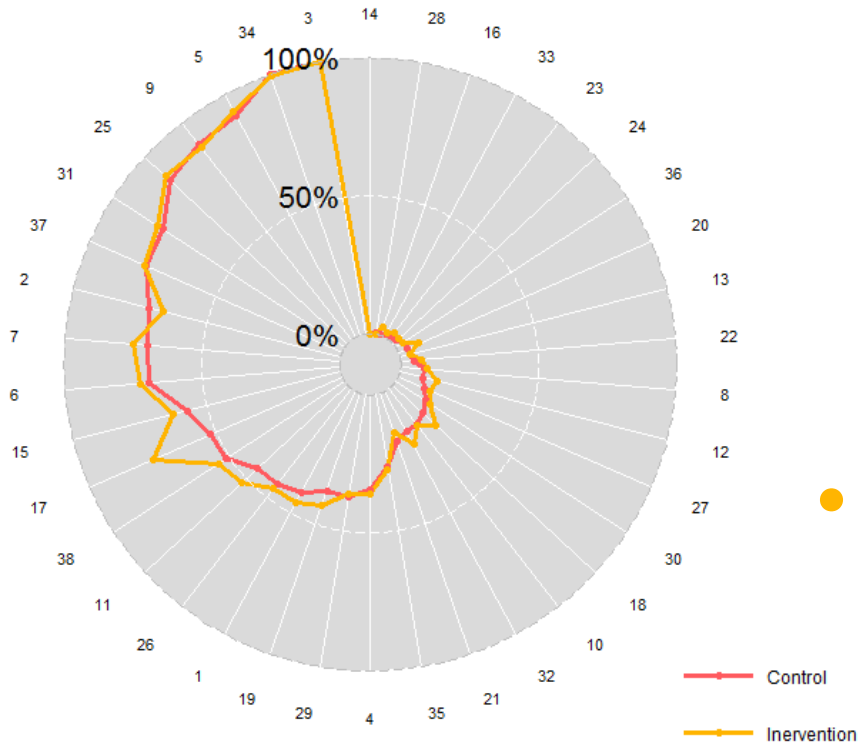
ARRIVE

Normal Handling



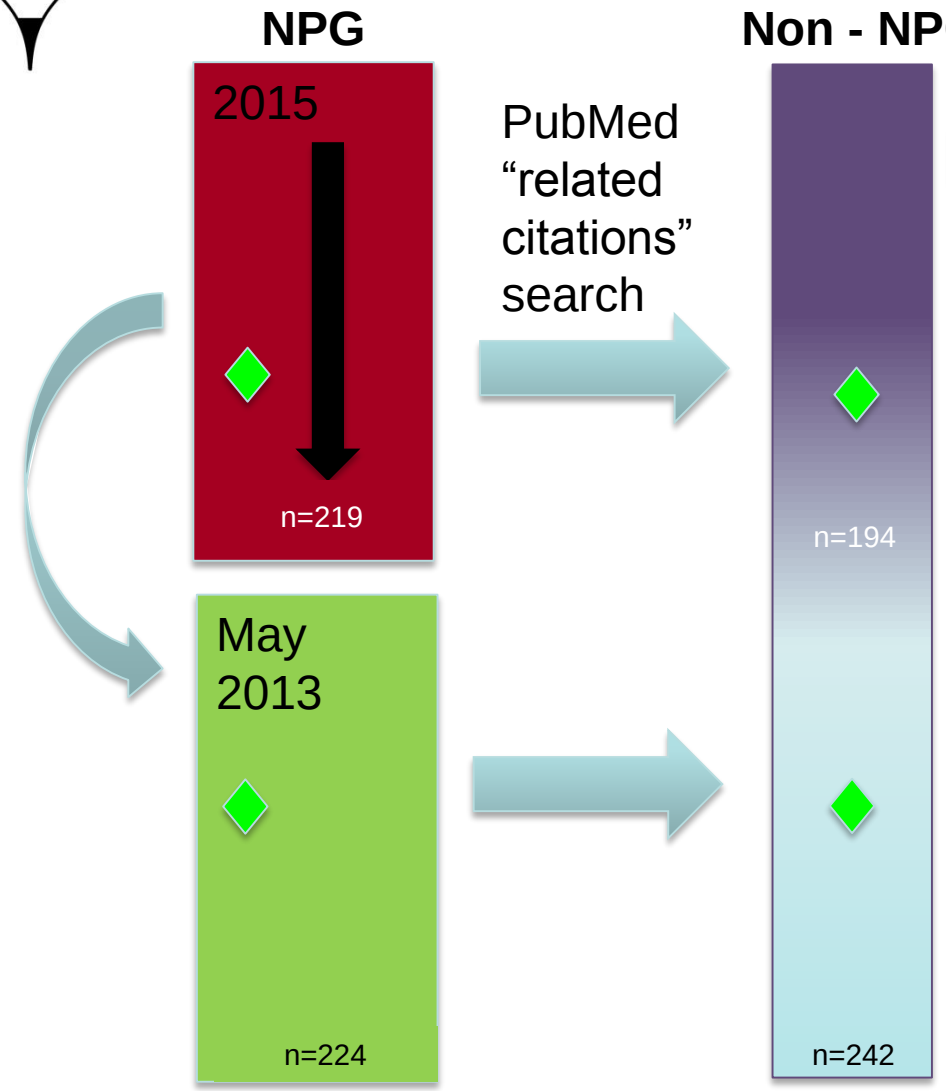
Protocol: Open Science Framework (February 2017)
Data Analysis Plan: Open Science Framework (September 2017)
Funding: MRC, NC3Rs, BBSRC & Wellcome Trust
Ethics: BMJ Ethics Committee

Primary outcome



- Control:
  - 100% compliance n=0 manuscripts
  - Median compliance 36.8% (29.7-42.1) of relevant items
- Intervention:
  - 100% Compliance n= 0
  - Median compliance 39.5% (31.6-44.7) of relevant items

Nature Publication Quality Improvement Project



Redacted for time sensitive information, journal

Uploaded for outcome assessment

- Web based, crowd sourced
- Assessors trained
- Dual ascertainment
- Reconciliation by third

Primary Outcome Change in proportion of NPG publications describing *in vivo* research published before and after May 2013 reporting the **"Landis 4"** items (randomisation, blinding, sample size calculation, exclusions).

Primary outcome measure

NPG

- Before: 0/204 (**0%,** 95% CI 0.0-2.3)
- After: 31/190 (**16.3%,** 95% CI 11.7-22.0)

$p<10^{-8}$

Non NPG

- Before: 1/164 (**0.6%,** 95% CI 0.1-4.2)
- After: 1/189 (**0.5%,** 95% CI 0.1-3.7)

n.s.

What we know

- *In vivo* studies which do not report simple measures to avoid bias give larger estimates of treatment effects
- Most do not report simple measures to reduce bias
  - Although improvements are happening
- We don't test interventions where efficacy has been shown in animals
- We generally working in silos with limited generalisability
- Our understanding/use of biological constructs must be improved
- Reporting biases are important and prevalent
- Most *in vivo* research is underpowered (or of unknown power)
- You cannot assume rigour, even in Journals of "impact"
- You can only find these things out by studying large numbers of studies
- Help is at hand but improvement strategies must be tested
- Any experimental design can be subverted; what's important is knowing how to recognise when this has happened

Thanks to.....