



Danish 3R-Center



# The 3Rs and the credibility of science: are they linked?

**Malcolm Macleod**

Collaborative Approach to Meta-Analysis and Review of  
Animal Data from Experimental Studies

and

University of Edinburgh

**CAMARADES: Bringing evidence to translational medicine**



# Disclosures



- UK Commission for Human Medicines
- EMA Neurology SAG
- UK Reproducibility Network
- Independent Statistical Standing Committee, CHDI Foundation
- Project co-ordinator, EQIPD IMI
- Minimum Standards Framework development group



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777364. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.



No entry for heavy  
goods vehicles.  
Residential site only  
←  
Nid wyf yn y swyddfa  
ar hyn o bryd. Anfonwch  
 unrhyw waith i'w gyfieithu.

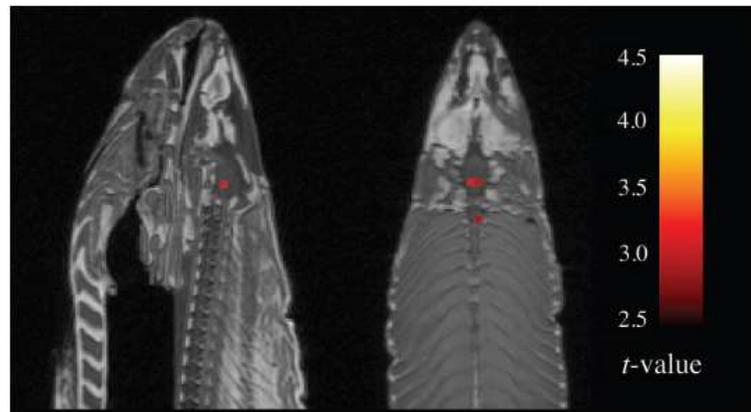
I am not in the  
office at the  
moment. Send  
any work to be  
translated.



## Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction

Craig M. Bennett<sup>1\*</sup>, Abigail A. Baird<sup>2</sup>, Michael B. Miller<sup>1</sup> and George L. Wolford<sup>3</sup>

One mature Atlantic Salmon (*Salmo salar*) participated in the fMRI study. The salmon measured approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning. It is not known if the salmon was male or female, but given the post-mortem state of the subject this was not thought to be a critical variable.



The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence, either socially inclusive or socially exclusive. The salmon was asked to determine which emotion the individual in the photo must have been experiencing.

Several active voxels were observed in a cluster located within the salmon's brain cavity (see Fig. 1). The size of this cluster was 81 mm<sup>3</sup> with a cluster-level significance of  $p = 0.001$ .

Either we have stumbled onto a rather amazing discovery in terms of post-mortem ichthyological cognition, or there is something a bit off with regard to our uncorrected statistical approach.



Winner of the 2012 Ig Nobel Prize for Neuroscience



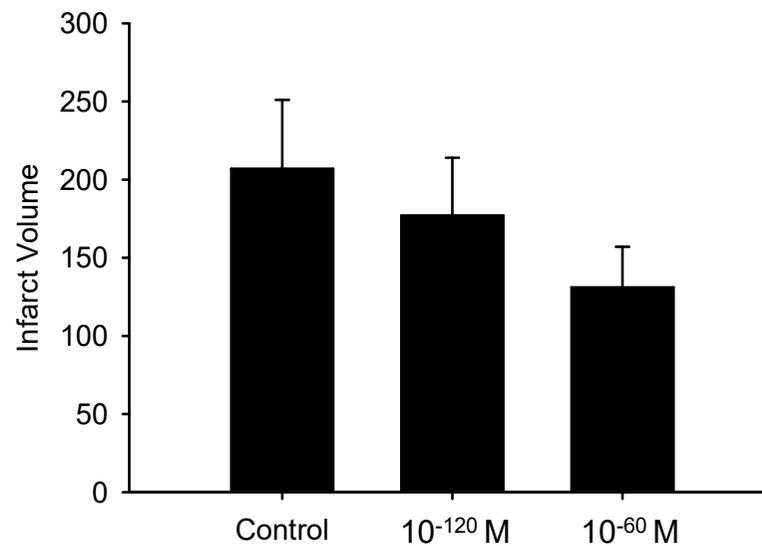
## Treatment of experimental stroke with low-dose glutamate and homeopathic *Arnica montana*\*

*W. Jonas*<sup>1</sup>, *Y. Lin*<sup>2</sup>, *A. Williams*<sup>2</sup>, *F. Tortella*<sup>2</sup>, *R. Tuma*<sup>3</sup>

<sup>1</sup> Uniformed Services University of the Health Sciences, Bethesda, Maryland

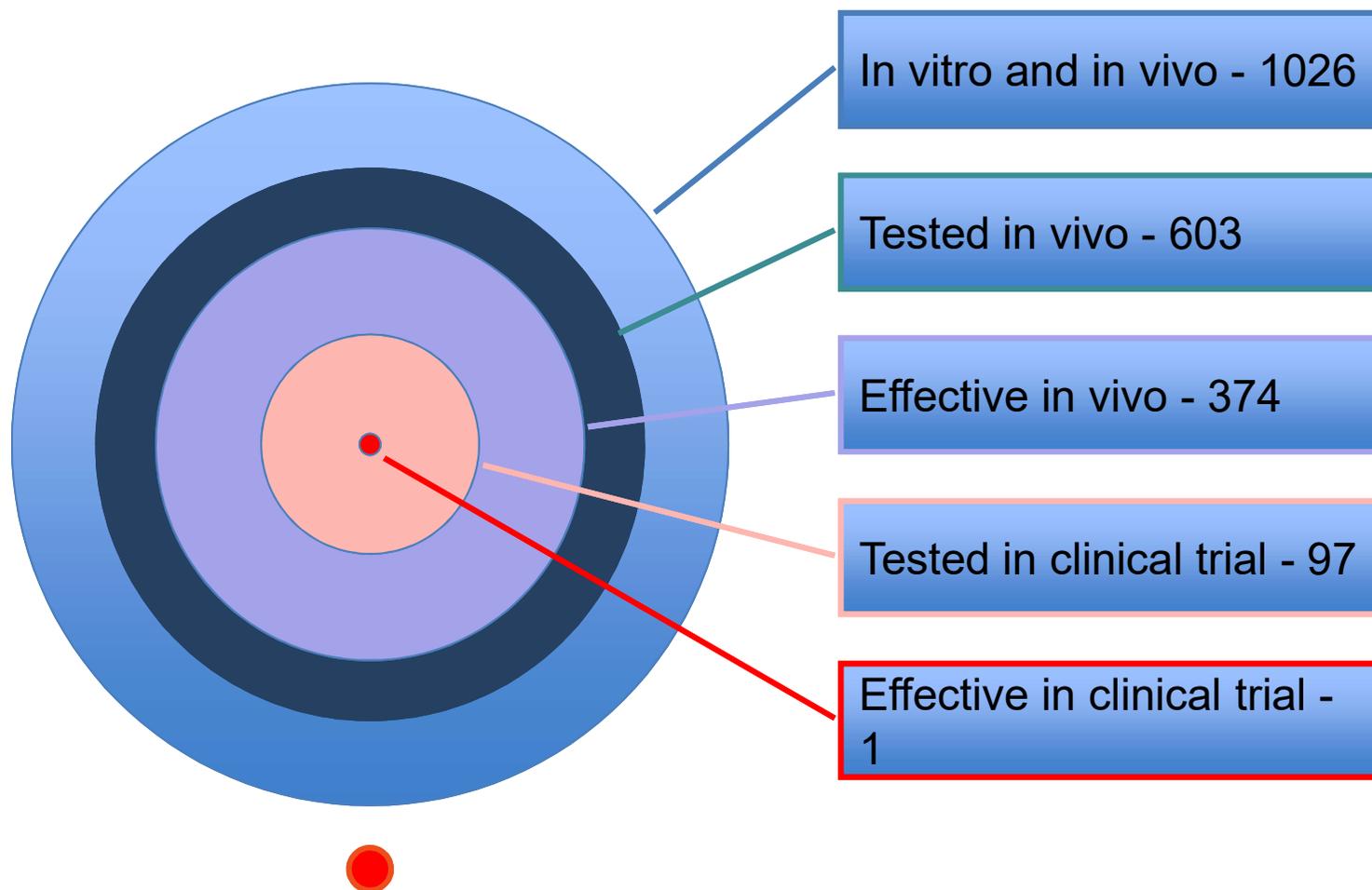
<sup>2</sup> Walter Reed Army Institute of Research, Washington, D.C.

<sup>3</sup> Temple University, Philadelphia, PA





# 1026 interventions in experimental stroke



O' Collins et al, 2006

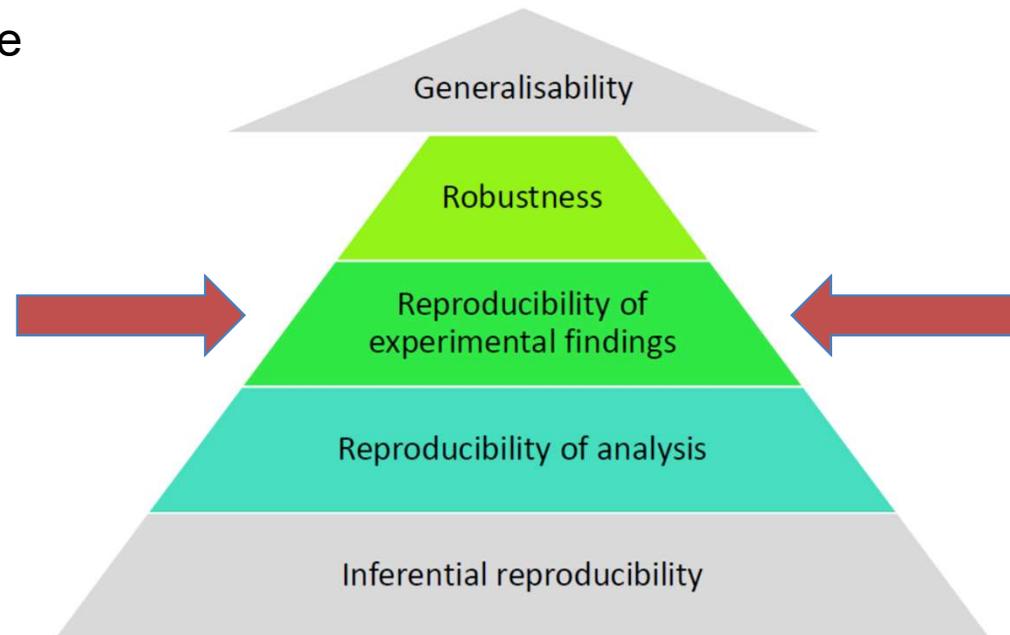


# Reproducibility and replication



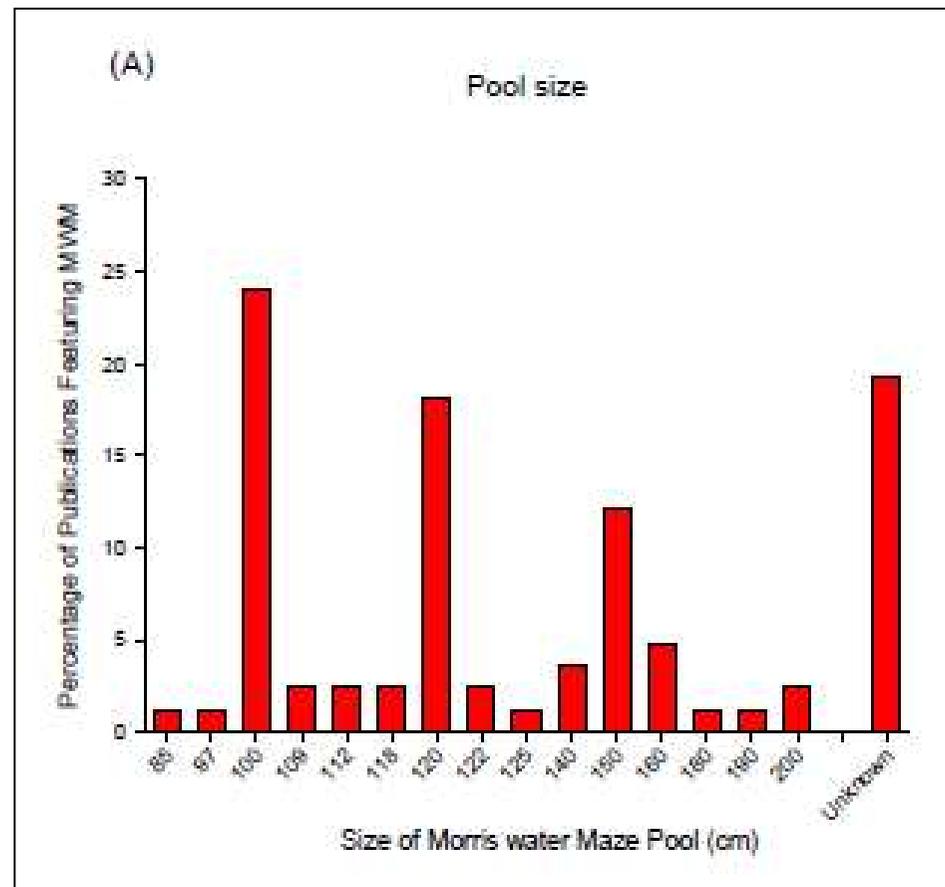
“Reproducibility” related to the re-analysis of existing data following the same analytical procedures.

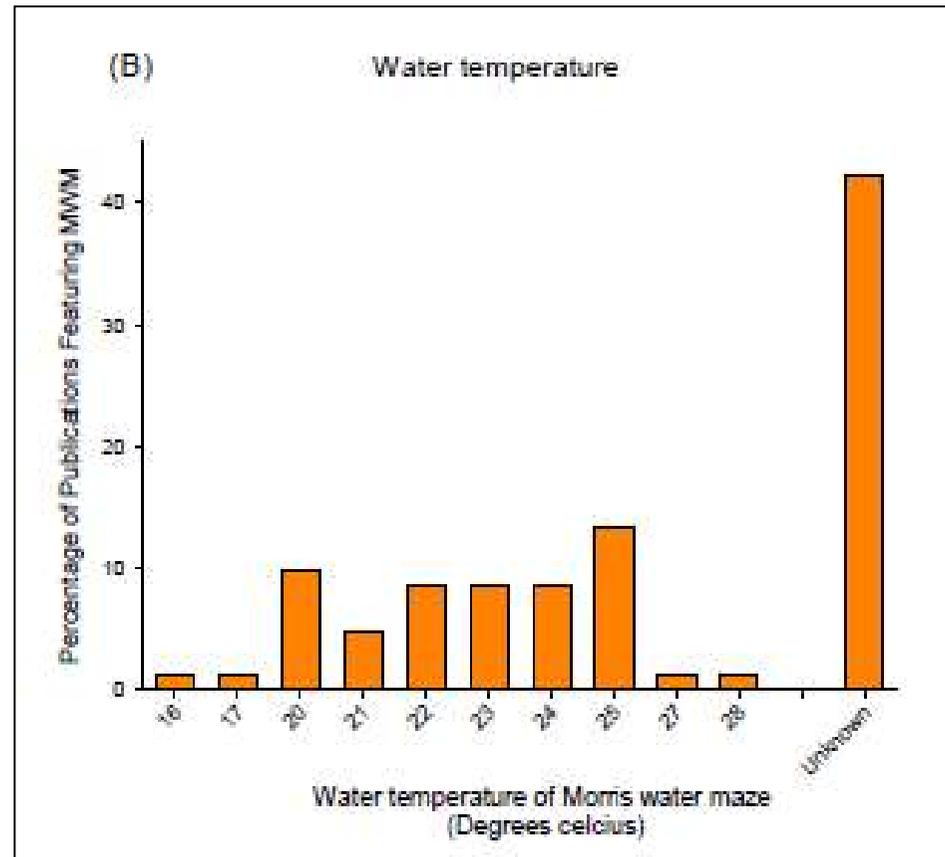
“Replication” was held to require the collection of new data, following the same methods.

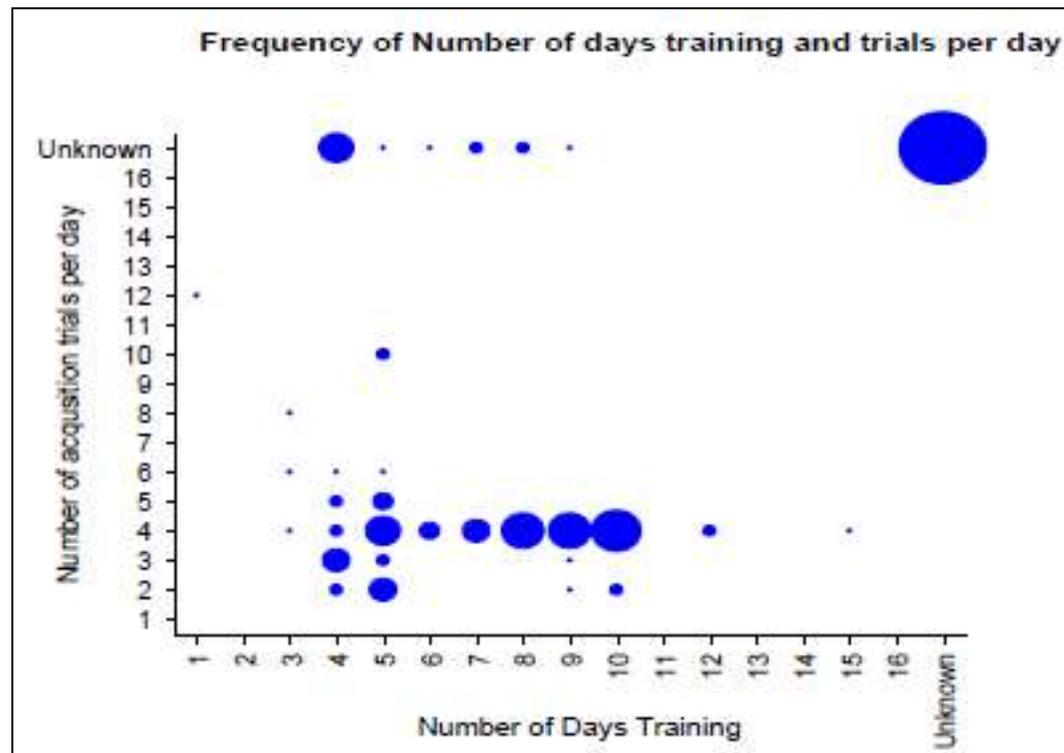




# Most studies are not described well enough to enable replication efforts









# Replication studies

1. **Retrospective** – Pharmaceutical companies sharing their historical experience when they have attempted replication

- Bayer 33% of 67
- Amgen 11% of 53

Selection bias (2 companies out of ?)

? Recall Bias



# Replication studies

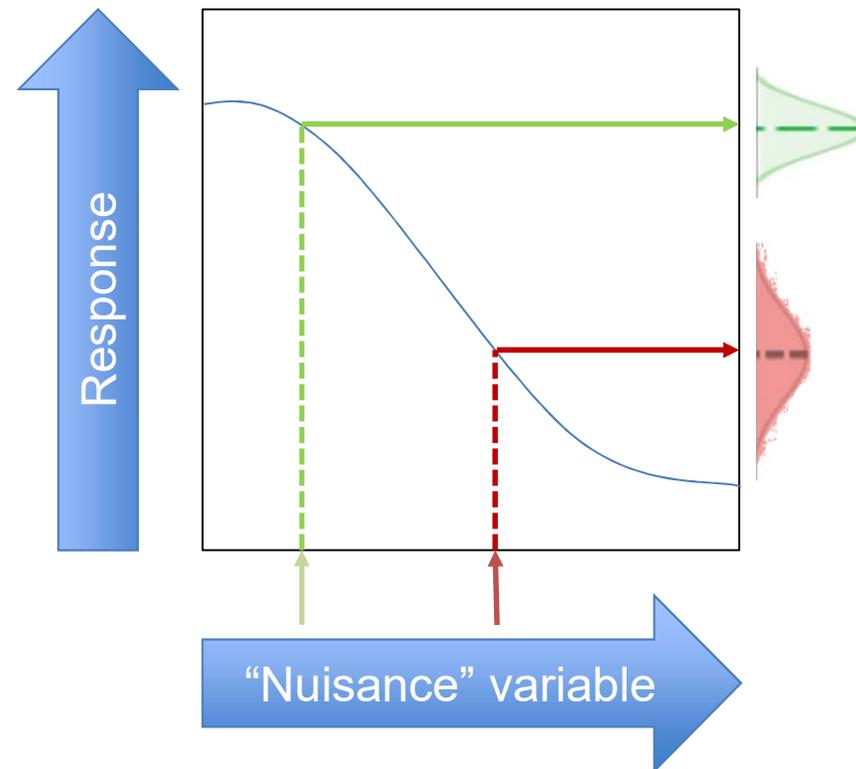
2. **Prospective** - Academic led, great attention given to faithfulness to original study design, adequate statistical power, preregistration

– Psychology	36% of 97	$ES_R=49\%$
– Cancer biology	40% of 10	
– Economics	61% of 18	$ES_R=66\%$
– Social sciences	62% of 21	$ES_R=54\%$

? Selection bias (how did they choose what to try to replicate?)



# Both studies may be correct Reaction norms (Voelkl 2016)





# Lifespan in worms



Source of variation	Developmental Rate	Fertility
Genetic	83.1%	63.3%
Between labs	8.3%	7.9%
Within labs	3.8%	5.6%
Individual	4.8%	23.3%

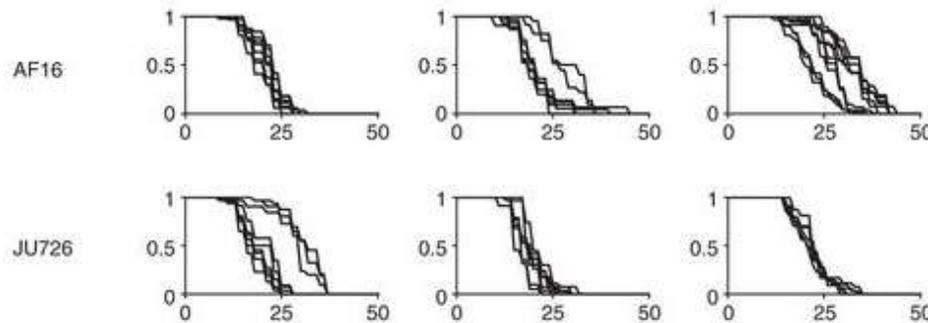
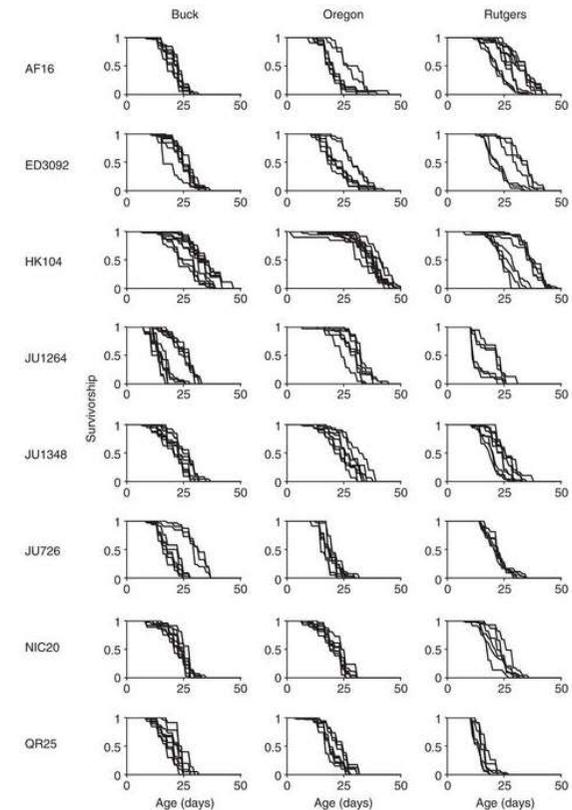


Figure 3: Variation in longevity within labs for each replicate plate for eight natural isolates of *C. briggsae*.



Lucanic et al Nature Comms 2017



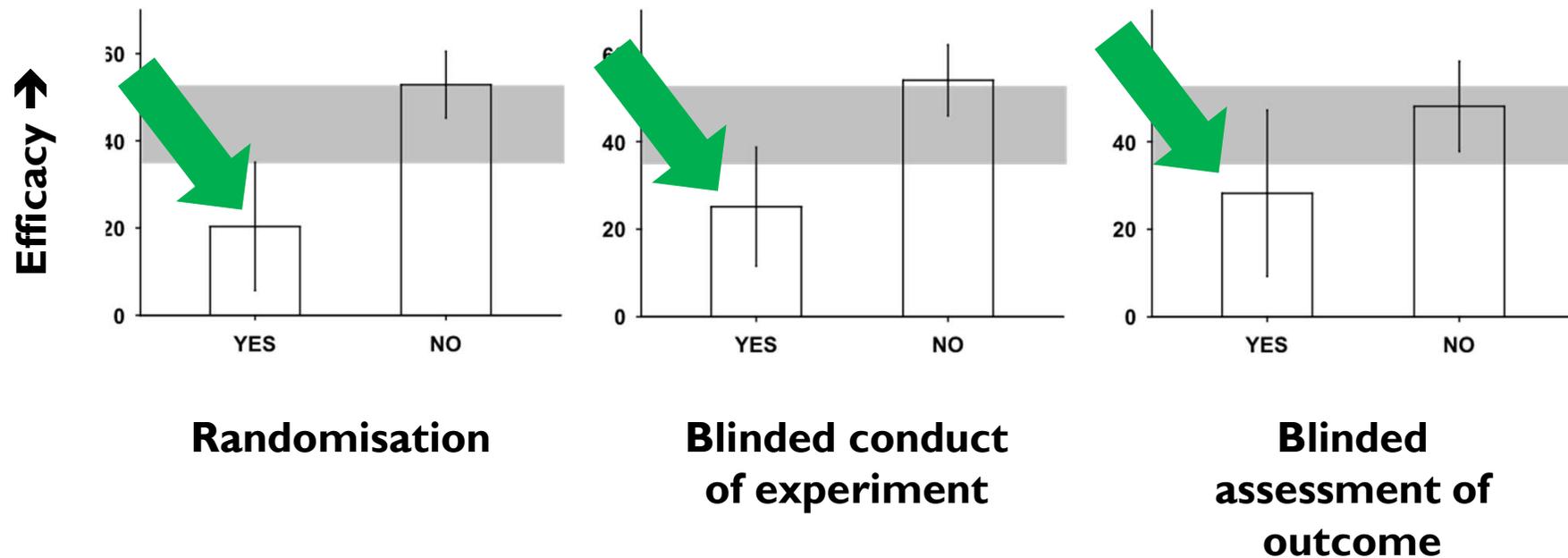
Experiments which are  
at high risk of bias



# Risk of bias in animal studies



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



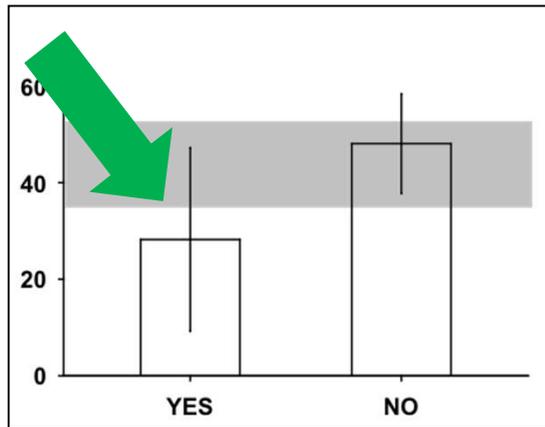
Macleod et al, 2008



# Evidence from various neuroscience domains ...



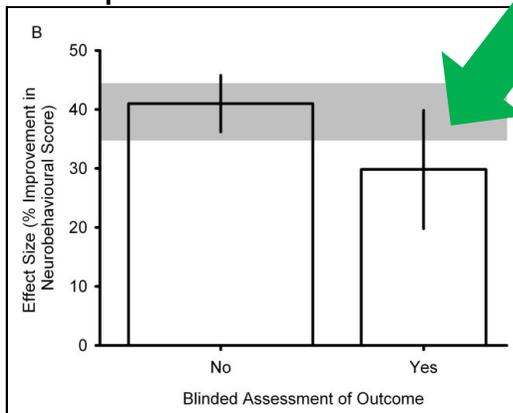
### Stroke



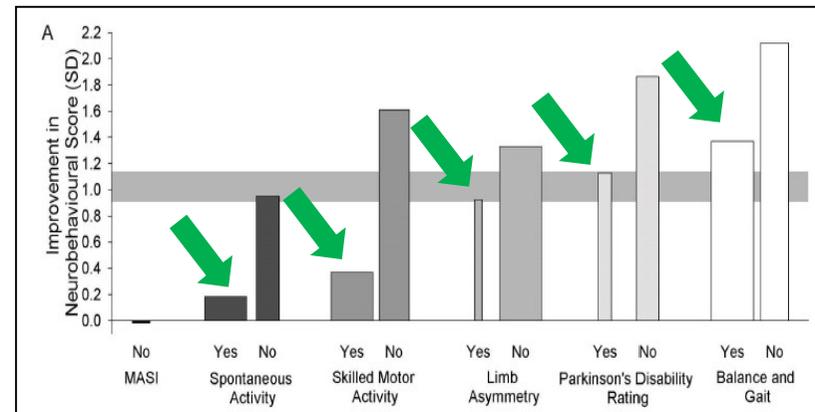
### Alzheimer's disease



### Multiple Sclerosis



### Parkinson's disease

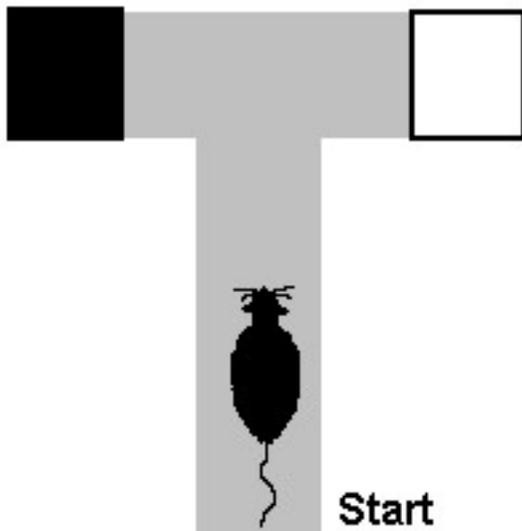




# You can usually find what you're looking for ...



- 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
$\Delta$	+0.60	+0.50	+0.37	+1.00	+1.43

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



# The scale of the problem

RAE 1173

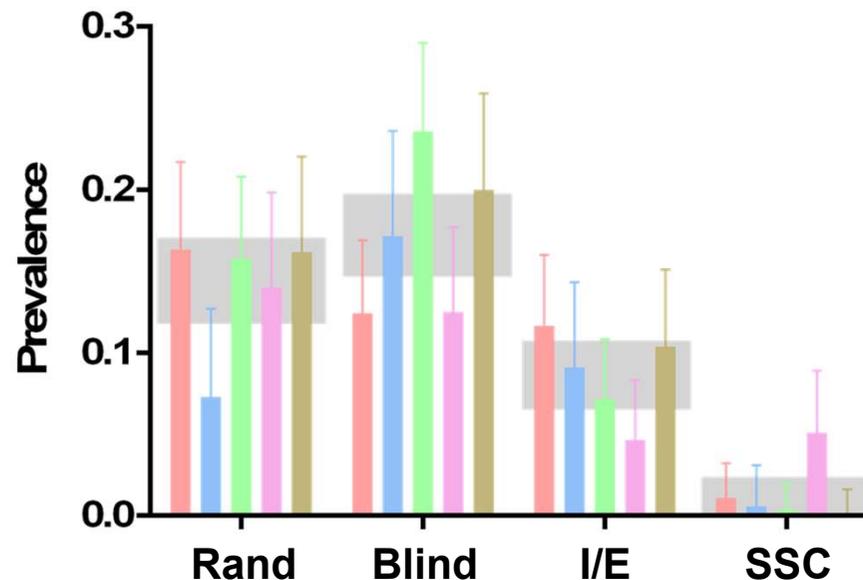


rae2008  
Research Assessment Exercise

“an outstanding contribution to the internationally excellent position of the UK in biomedical science and clinical/translational research.”

“impressed by the strength within the basic neurosciences that were returned ... particular in the areas of behavioural, cellular and molecular neuroscience”

1173 publications using non human animals, published in 2009 or 2010, from 5 leading UK universities





Experiments which are  
too homogenous  
Experiments at high risk  
of bias

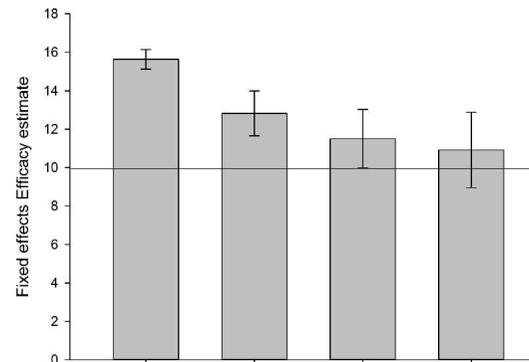
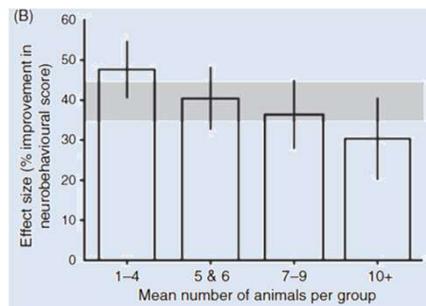


# Small group sizes and publication bias conspire together



Simulation: 1000 studies  
Complete publication bias (anything  $p > 0.05$  unpublished)  
True effect size 10, SD 10

Number of animals per group	5	10	15	20
% of studies published	30%	54%	76%	86%





ARTICLE

DOI: [10.1038/s41467-017-02765-w](https://doi.org/10.1038/s41467-017-02765-w)

OPEN

# Regulation of REM and Non-REM Sleep by Periaqueductal GABAergic Neurons

Franz Weber<sup>1,3</sup>, Johnny Phong Hoang Do<sup>1</sup>, Shinjae Chung<sup>1,3</sup>, Kevin T. Beier<sup>2</sup>, Mike Bikov<sup>1</sup>,  
Mohammad Saffari Doost<sup>1</sup> & Yang Dan<sup>1</sup> 



**Sample sizes.** For optogenetic activation experiments, cell-type-specific ablation experiments, and in vivo recordings (optrode recordings and calcium imaging), we continuously increased the number of animals until statistical significance was reached to support our conclusions.



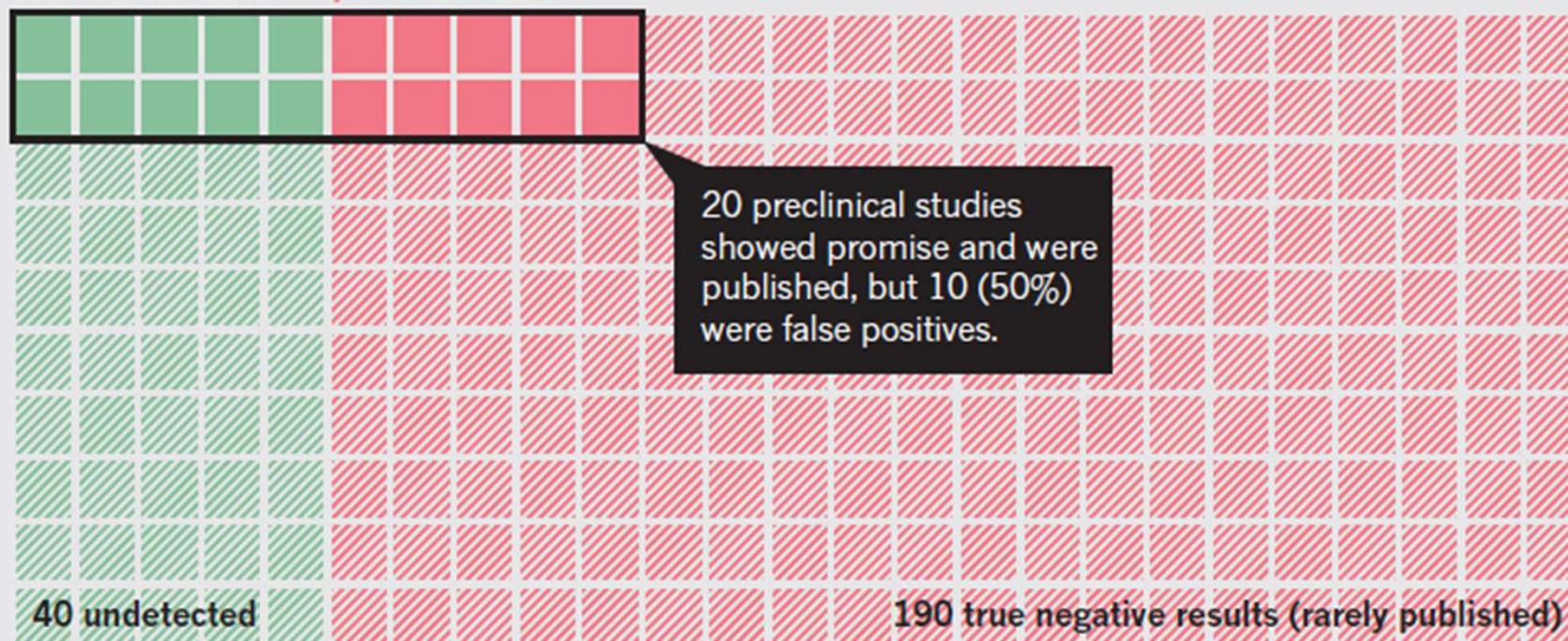
# Take 250 in vivo studies ...



**STATUS QUO:** Most studies have a statistical power of only 20% and a  $P$  value of 0.05, meaning many more false findings (PPV of 50%). This reflects a sample size of about 10 mice per study.

10 promising molecules found

10 false positives found



Macleod and Mogil, Nature, 2017



...with  $p < 0.01$ , power @ 80%



**PROPOSED STANDARDS:** To achieve a PPV of 95%, study results would need a  $P$  value of 0.01 and a large enough sample size to reach 80% statistical power (typically >75 mice per study).

40 promising molecules found

2 false positives found

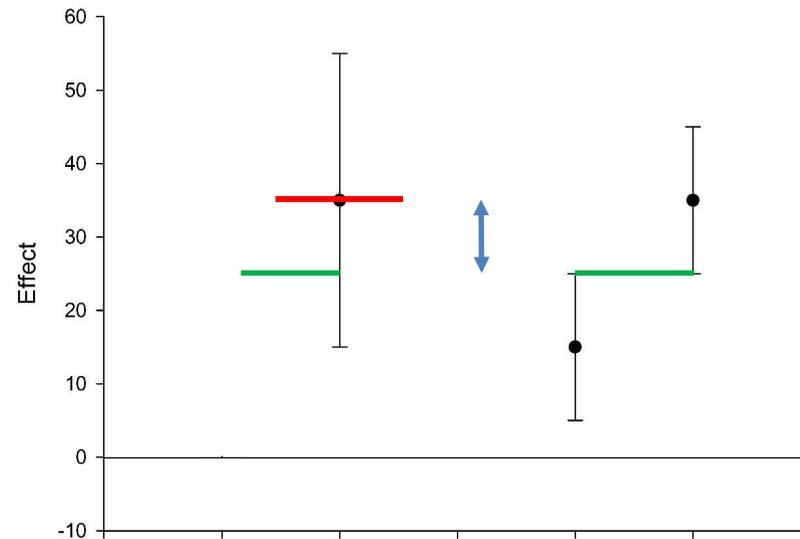


Macleod and Mogil, Nature, 2017



# How does that work?

Two sets of studies, one underpowered





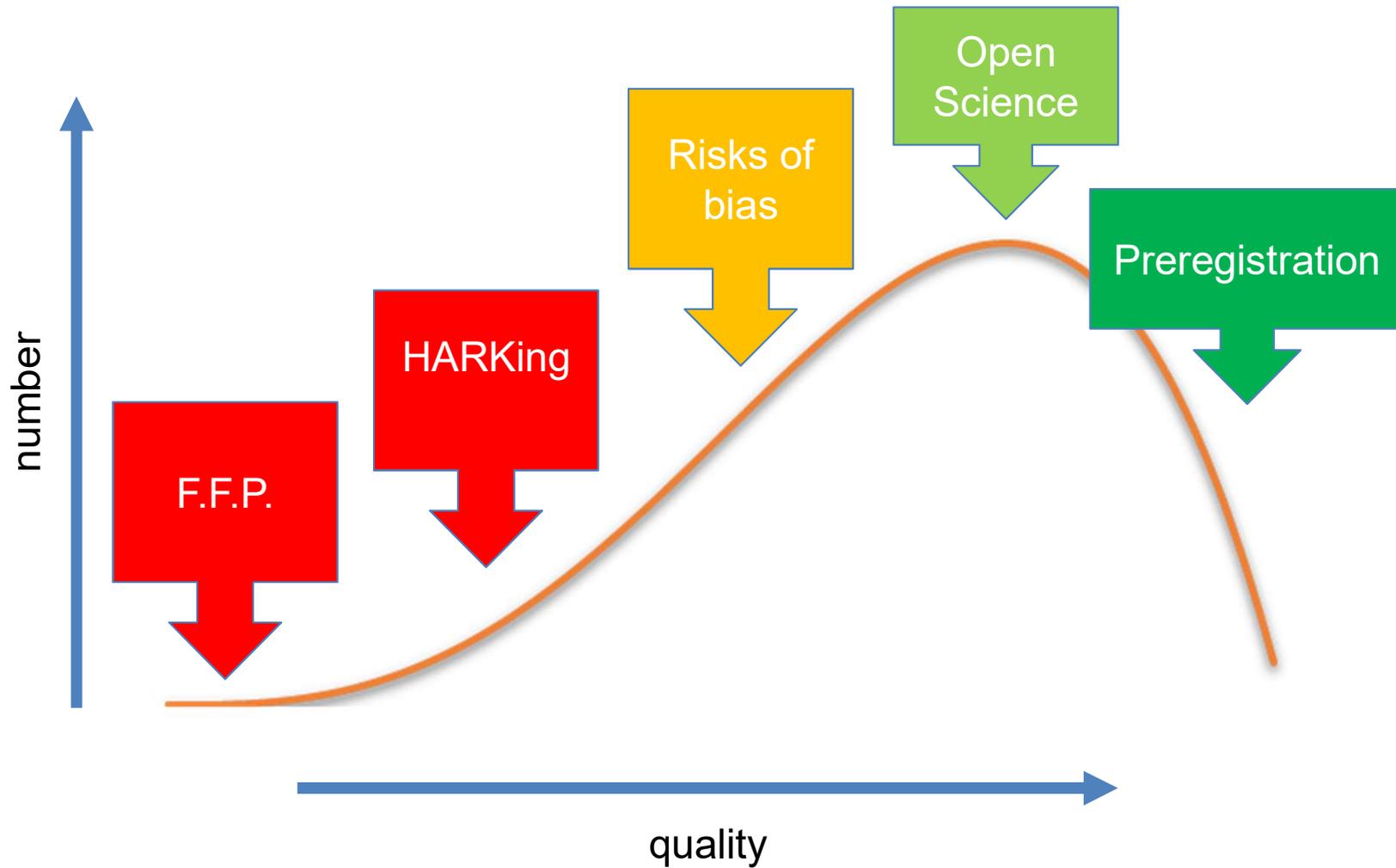
Experiments which are too homogenous

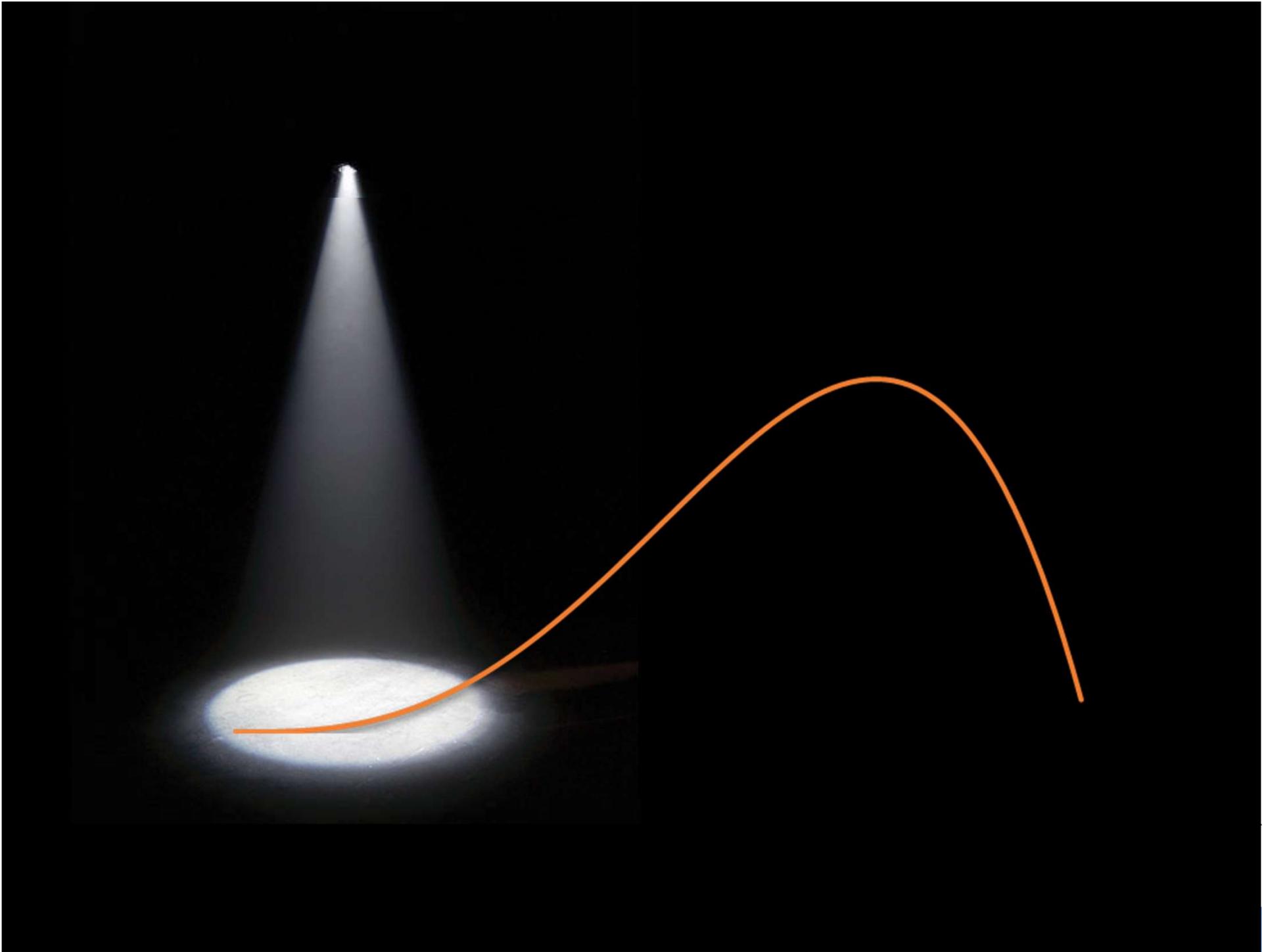
Experiments at high risk of bias

Experiments which are too small



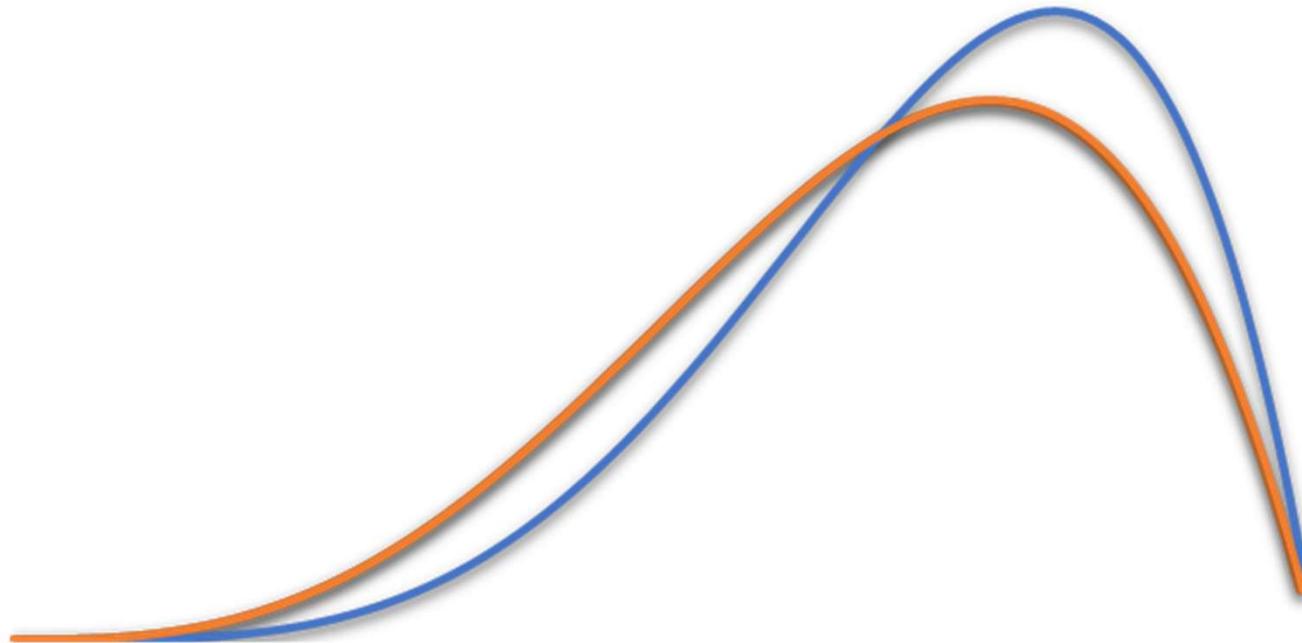
# Researchers are different ...







# Research Improvement Strategy



**CAMARADES: Bringing evidence to translational medicine**



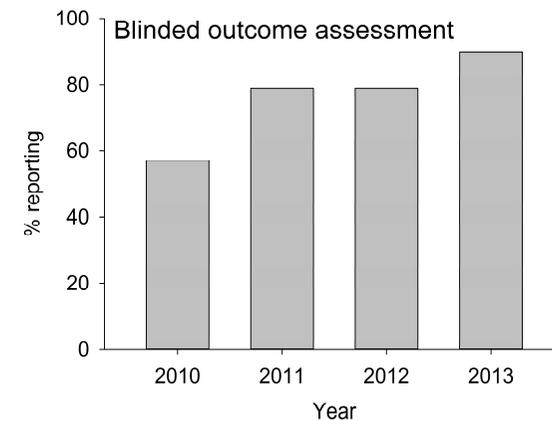
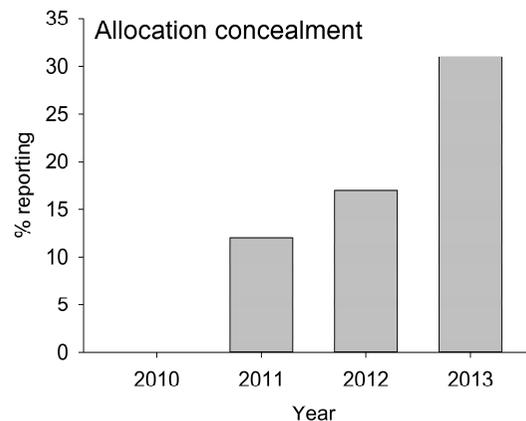
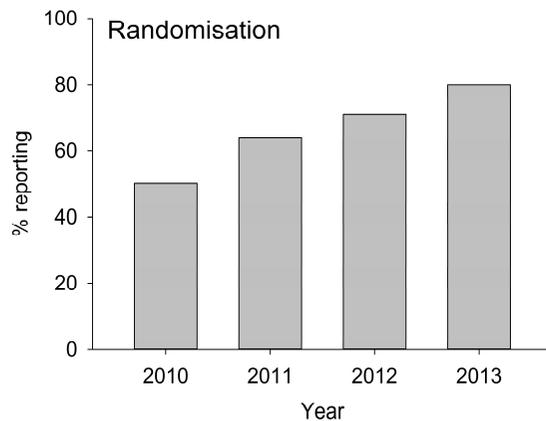
# Research Improvement at Journals



**Comments, Opinions, and Reviews**

**Good Laboratory Practice**  
**Preventing Introduction of Bias at the Bench**

Malcolm R. Macleod; Marc Fisher; Victoria O'Collins; Emily S. Sena; Ulrich Dirnagl;  
Philip M.W. Bath; Alistair Buchan; H. Bart van der Worp; Richard Traystman; Kazuo Minematsu;  
Geoffrey A. Donnan; David W. Howells



Minnerup et al, 2016



# Ramirez et al Circ Res 2017



Supplemental Table: Comparison of study design element implementation in preclinical studies before and after the implementation of the *Stroke* Basic Science Checklist, stratified by journal of publication

	Period 1* <i>n</i> (%)	Period 2* <i>n</i> (%)	Crude OR (95% CI)	<i>P</i>	Adjusted OR (95% CI) <sup>†</sup>	<i>P</i> <sup>†</sup>
<i>Circulation</i>	<i>n</i> =464	<i>n</i> =208				
Randomization	107 (23.1)	36 (17.3)	0.7 (0.5-1.1)	0.093	0.7 (0.4-1.1)	0.119
Blinding	169 (36.4)	59 (28.4)	0.7 (0.5-1.0)	0.042	0.7 (0.5-1.0)	0.043
Sample size estimation	7 (1.5)	5 (2.4)	1.6 (0.5-5.1)	0.422	NR	
Inclusion of both sexes	64 (13.8)	29 (13.9)	1.0 (0.6-1.6)	0.959	1.0 (0.6-1.6)	0.967
<i>Circulation Research</i>	<i>n</i> =303	<i>n</i> =183				
Randomization	35 (11.6)	29 (15.8)	1.4 (0.8-2.5)	0.176	1.4 (0.8-2.5)	0.261
Blinding	93 (30.7)	60 (32.8)	1.1 (0.7-1.6)	0.630	0.9 (0.6-1.4)	0.788
Sample size estimation	1 (0.3)	1 (0.3)	1.7 (0.1-26.7)	0.721	NR	
Inclusion of both sexes	57 (18.8)	33 (18.0)	0.9 (0.6-1.5)	0.830	1.0 (0.6-1.6)	0.937
<i>Hypertension</i>	<i>n</i> =485	<i>n</i> =375				
Randomization	104 (21.4)	81 (21.6)	1.0 (0.7-1.4)	0.956	1.2 (0.9-1.7)	0.298
Blinding	101 (20.8)	86 (22.9)	1.1 (0.8-1.6)	0.457	1.1 (0.8-1.5)	0.617
Sample size estimation	0 (0)	1 (0.3)	→∞ (0.0-∞)	0.946	NR	
Inclusion of both sexes	43 (8.9)	36 (9.6)	1.1 (0.7-1.7)	0.712	1.1 (0.7-1.7)	0.798
<i>Stroke</i>	<i>n</i> =316	<i>n</i> =185				
Randomization	120 (38.0)	119 (64.3)	2.9 (2.0-4.3)	<0.0001	3.2 (2.1-4.7)	<0.0001
Blinding	171 (54.1)	144 (77.8)	3.0 (2.0-4.5)	<0.0001	3.0 (2.0-4.5)	<0.0001
Sample size estimation	10 (3.2)	35 (18.9)	7.1 (3.4-14.8)	<0.0001	8.2 (3.7-18.4)	<0.0001
Inclusion of both sexes	15 (4.7)	20 (10.8)	2.4 (1.2-4.9)	0.012	2.4 (1.2-4.9)	<0.0001
<i>ATVB</i>	<i>n</i> =476	<i>n</i> =401				
Randomization	61 (12.8)	48 (12.0)	0.9 (0.6-1.4)	0.706	0.9 (0.6-1.4)	0.668
Blinding	130 (27.3)	97 (24.2)	0.8 (0.6-1.2)	0.293	0.7 (0.5-1.0)	0.026
Sample size estimation	2 (0.4)	10 (2.5)	6.1 (1.3-27.8)	0.021	NR	
Inclusion of both sexes	72 (15.1)	52 (13.0)	0.8 (0.6-1.2)	0.361	0.8 (0.6-1.3)	0.411

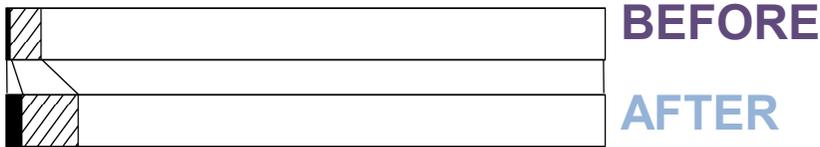
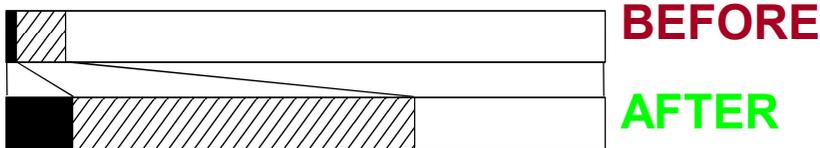
NR: not reported due to small number of events per predictor variable; OR: odds ratio

\*Periods 1 and 2 correspond to before and after the date of implementation of the 'Basic Science Checklist' by *Stroke*, respectively

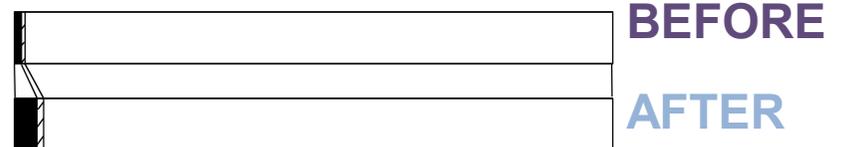
†Adjusted for cardiovascular disease studied and animal model used

# Impact of NPG checklist

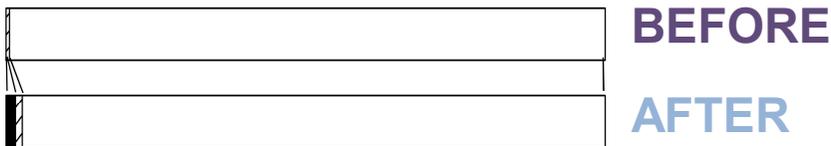
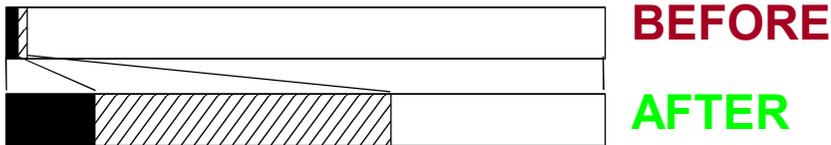
## Randomisation



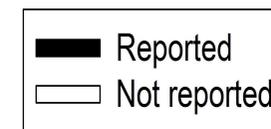
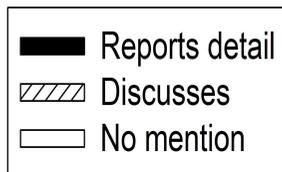
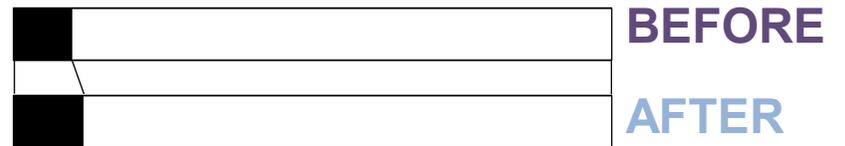
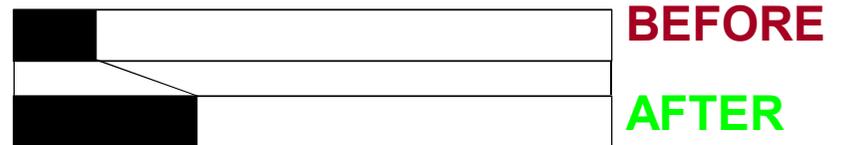
## Blinding



## Sample size calculation



## Reporting exclusions





Experiments which are too homogenous

Experiments at high risk of bias

Experiments which are too small

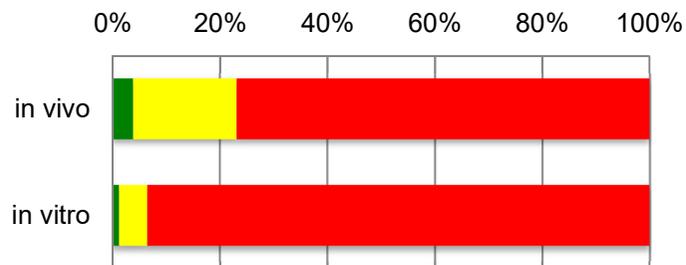
Journals committed to research improvement



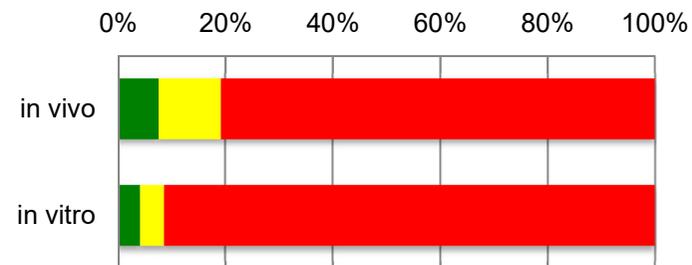
# In vitro experiments



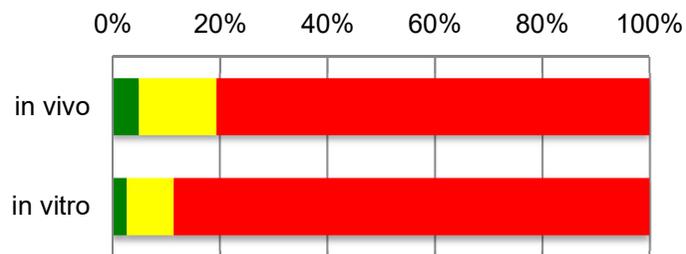
## Randomisation



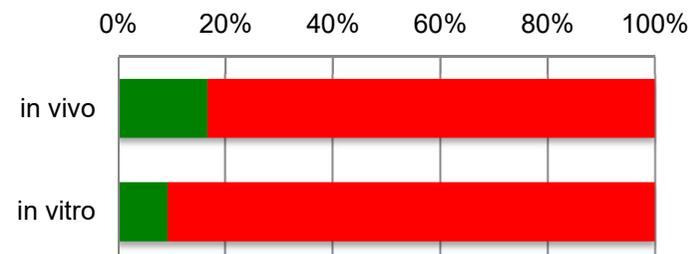
## Blinding



## Power calculation



## Reporting exclusions



■ Full ■ Partial ■ Null



Experiments which are too homogenous

Experiments at high risk of bias

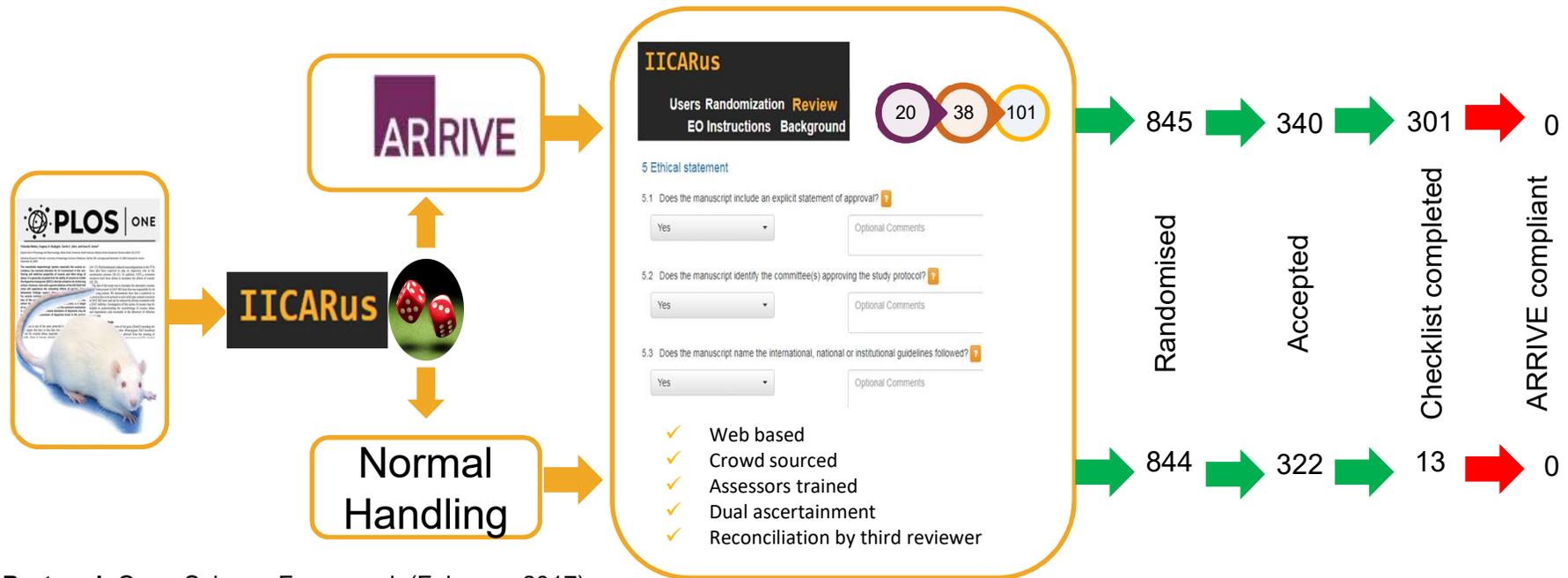
Experiments which are too small

Poorly conducted non animal alternatives

Journals committed to research improvement



# IICARUS (PI Sena)



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



Experiments which are too homogenous

Experiments at high risk of bias

Experiments which are too small

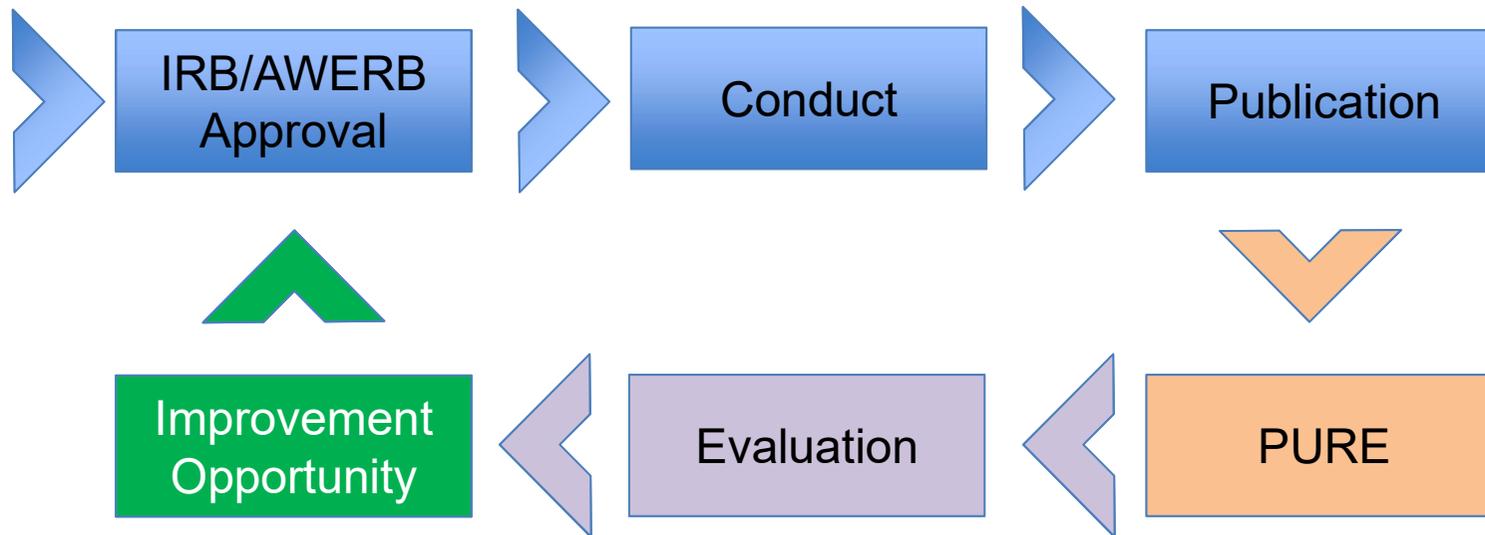
Poorly conducted non animal alternatives

Journals committed to research improvement

Using evidence to inform improvement

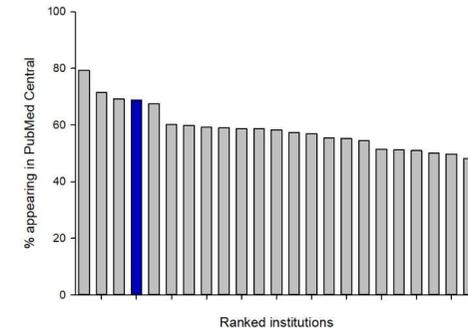
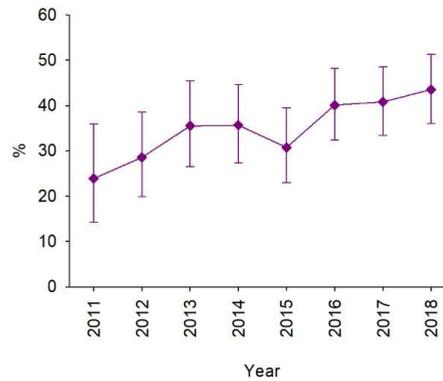
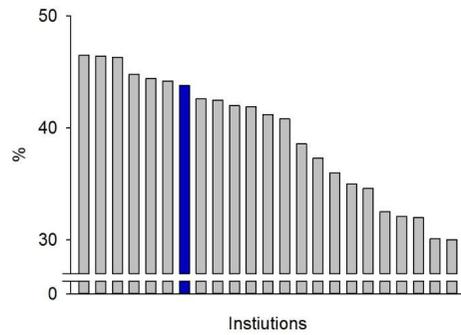


# Edinburgh Research Improvement Project

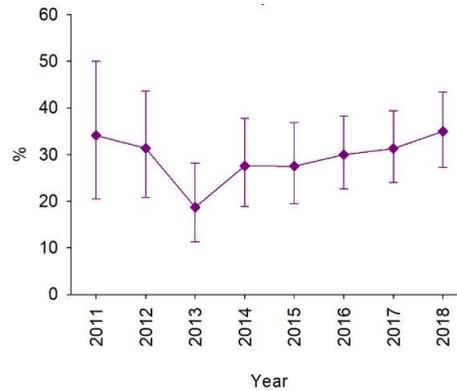
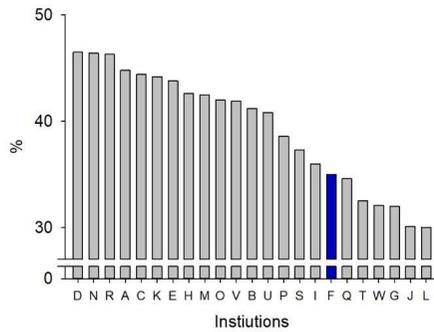




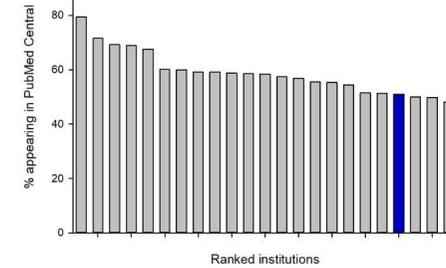
# Measuring institutional performance



Reporting of blinding by institution, 2018



% of PubMed animal studies in PubMed Central, 2018



# Preliminary (Draft) benchmarking of Wellcome Trust funded *in vivo* research

We present some preliminary benchmarking of Wellcome funded *in vivo* research. The methodologies used continue to be refined, and the performance reported is based on our best estimates, using the text mining and automation processes described. Future analyses will include reporting of randomisation and of power calculations, when the performance of these tools reaches the required threshold.

Performance measure: Proportion of *in vivo* research which reports whether or not the experiment was blinded.

Approach: To identify reports of *in vivo* research we trained an algorithm on a corpus of 5,000 dual screened publications to identify those describing primary animal research. We do not expect to have captured all *in vivo* research publications, and some identified publications will not describe *in vivo* research. We applied the algorithm to the PubMed Central corpus, extracting papers identified as describing *in vivo* research published from 2011 to 2018. We retrieved the PMIDs for these publications, and then matched this with the “Grant Agency” recorded in PubMed to identify the funding agency, where this was given.

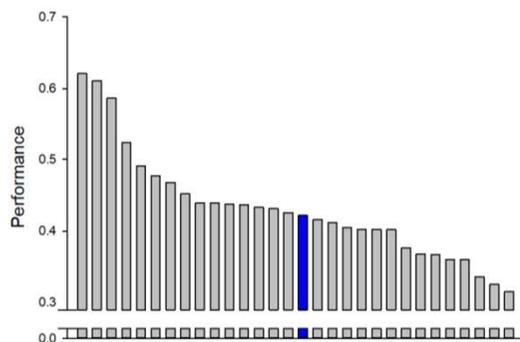
For funders with more than 500 records in this corpus we retrieved the full text of articles from PubMed Central, and analysed these using Regular Expressions (see <https://doi.org/10.1042/CS20160722> for approach). In the recent focal ischaemia literature this tool has a sensitivity of 0.99 and specificity of 0.77,

and so the tool will tend to overstate the prevalence of blinding. With 100 papers and an observed rate of 50%, the 95% bootstrapped confidence interval is around +/- 10%. We report the proportion of studies, by funder, which were scored as reporting blinding, for each year.

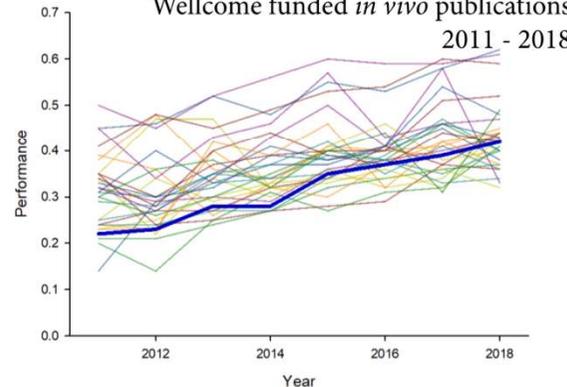
Because our intention is to support audit for improvement we present performance of the index Institute against other, unidentified funders. Wellcome data are based on analysis of 4858 manuscripts.

Figure 1 shows the reporting of blinding in 2018, with the Wellcome in blue and other Institutes in grey. Figure 2 shows the change in performance over 8 years, with the Wellcome performance emboldened. Finally (Figure 3), we show the change in performance over 8 years calculated through unweighted linear regression, again with the Wellcome in blue and other funders in grey.

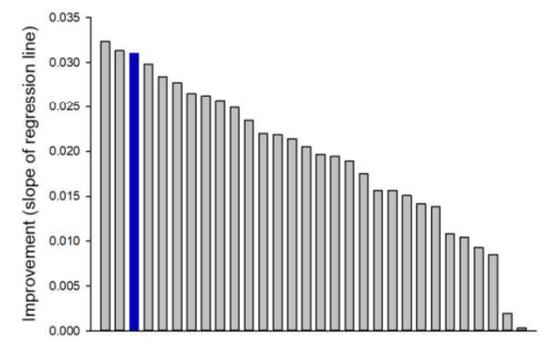
**Figure 1.** Reporting of blinding by institution 2018



**Figure 2.** Change in reporting of blinding in Wellcome funded *in vivo* publications 2011 - 2018



**Figure 3.** Change in reporting of blinding 2011 to 2018





Experiments which are too homogenous

Experiments at high risk of bias

Experiments which are too small

Poorly conducted non animal alternatives

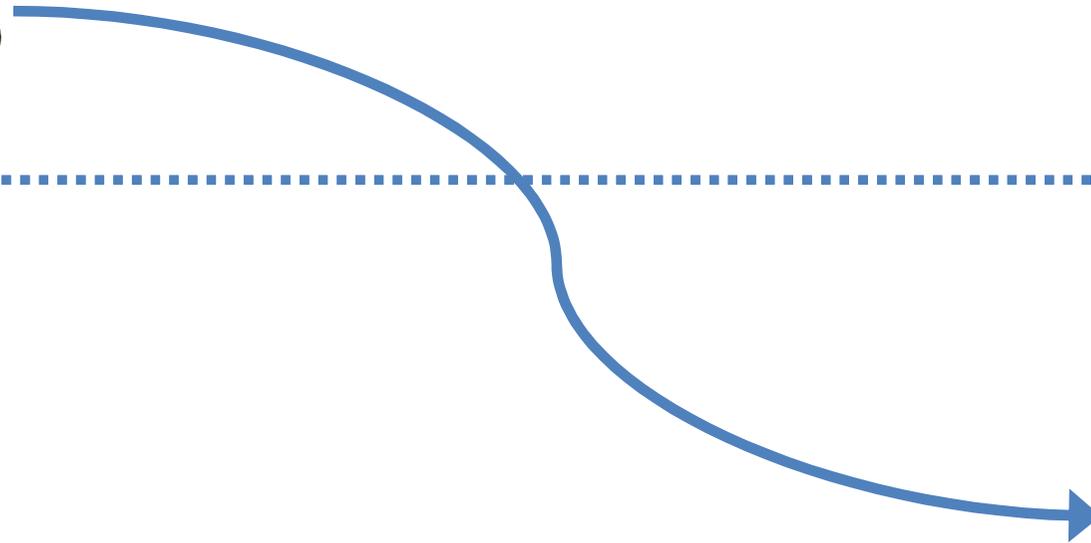
Journals committed to research improvement

Using evidence to inform improvement

Institutions and funders committed to research improvement



# Trans-lational research



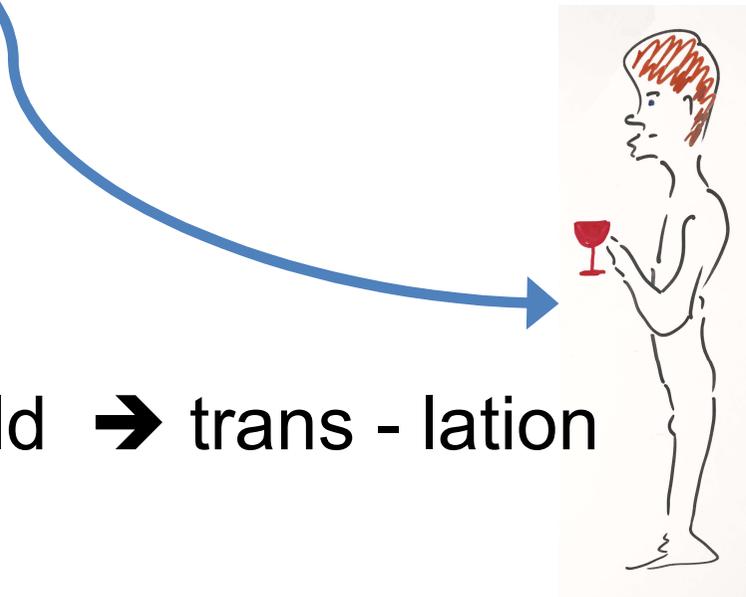
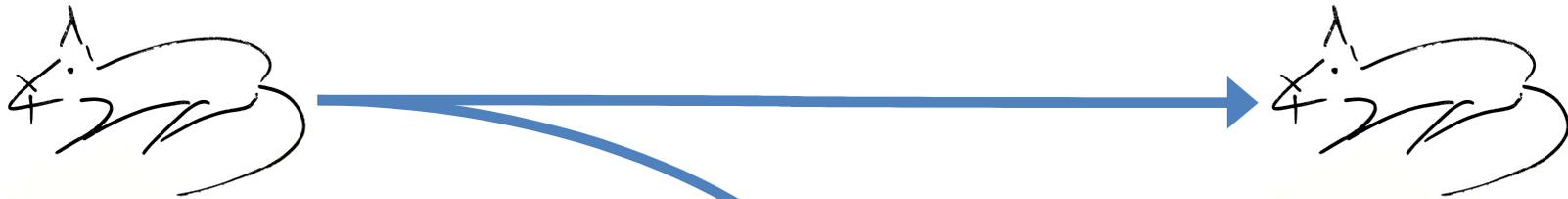


# Cis-lational research





If  $\Sigma(\text{knowledge}) < \text{threshold} \rightarrow \text{cis - lation}$



If  $\Sigma(\text{knowledge}) > \text{threshold} \rightarrow \text{trans - lation}$



Experiments which are too homogenous

Experiments at high risk of bias

Experiments which are too small

Poorly conducted non animal alternatives

Journals committed to research improvement

Using evidence to inform improvement

Institutions and funders committed to research improvement

Routine systematic evaluation of what is already known



# Biomedical research investment



- \$300bn globally, €50bn in Europe
- Glasziou and Chalmers claim 85% wasted
- Even if waste is only 50%, improvements which reduced that by 1% would free \$3bn globally, €500m in Europe, every year.
- Investing ~1% of research expenditure in improvement activity would go a long way



If you are planning a systematic review or meta-analysis of animal data, CAMARADES are here to help: [malcolm.macleod@ed.ac.uk](mailto:malcolm.macleod@ed.ac.uk)



innovative medicines initiative

The project leading to this application has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777364. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

**CAMARADES: Bringing evidence to translational medicine**