Data quality and reproducibility in preclinical research

Malcolm Macleod

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

and

University of Edinburgh

Disclosures

- UK Commission for Human Medicines
- EMA Neurology SAG

- UK Animals in Science Committee

- Independent Statistical Standing Committee, CHDI Foundation
- Avilex Pharma Research Steering Group (on behalf of Wellcome Trust)

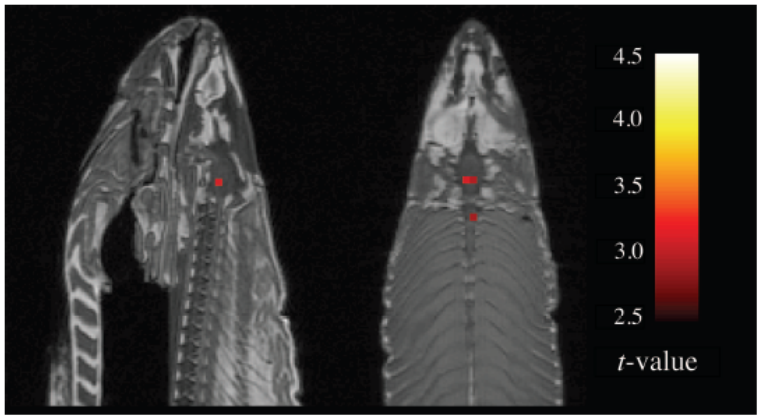
I am not in the office at the moment. Send any work to be translated.

Journal of Serendipitous and Unexpected Results

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

Craig M. Bennett [1]*, Abigail A. Baird [2], Michael B. Miller [1] and George L. Wolford [3]

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.

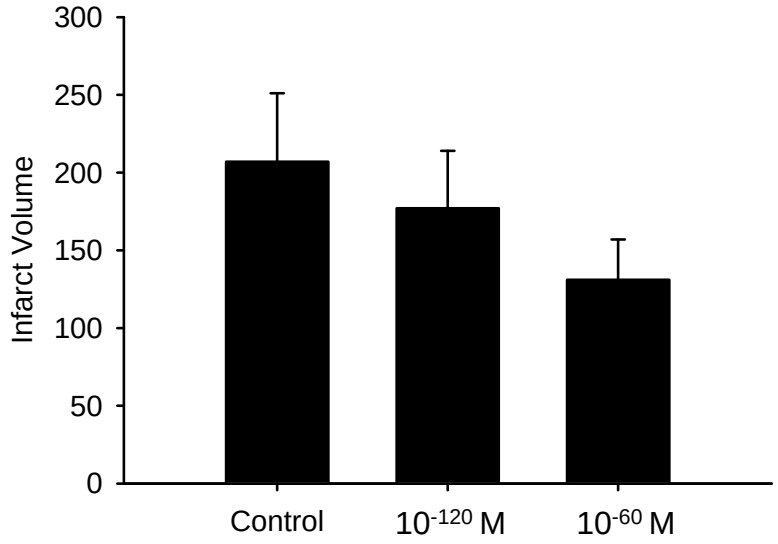
Several active voxels were observed in a cluster located within the salmon's brain cavity (see Fig. 1). The size of this cluster was 81 mm^3 with a cluster-level significance of $p = 0.001$.

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 Ignoble Prize for Neuroscience

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

W. Jonas[1], Y. Lin[2], A. Williams[2], F. Tortella[2], R. Tuma[3]
[1] Uniformed Services University of the Health Sciences, Bethesda, Maryland
[2] Walter Reed Army Institute of Research, Washington, D.C.
[3] Temple University, Philadelphia, PA



Typhoid fever

Moragues et al 1944

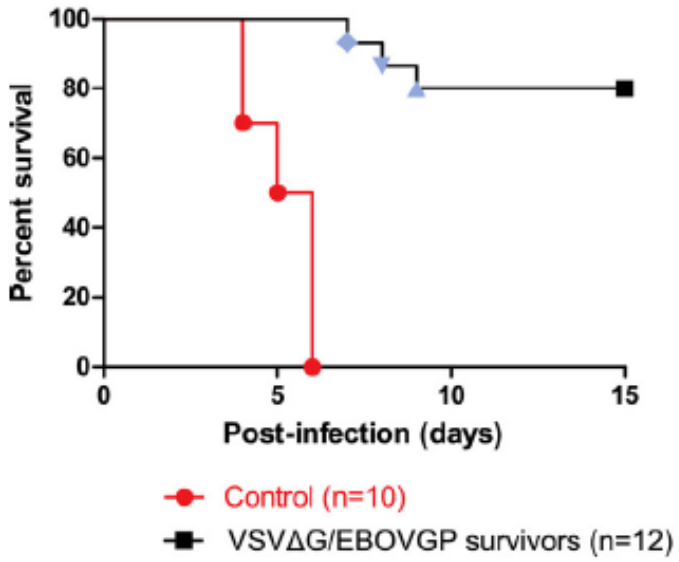
TABLE V

Effect of Penicillin on Murine Typhus in Mice, Experiment V
Moderate Dosate (Mouse Brain Intraperitoneally) Approaching the Minimal Lethal Dosage
Room temperature 65–72°F.

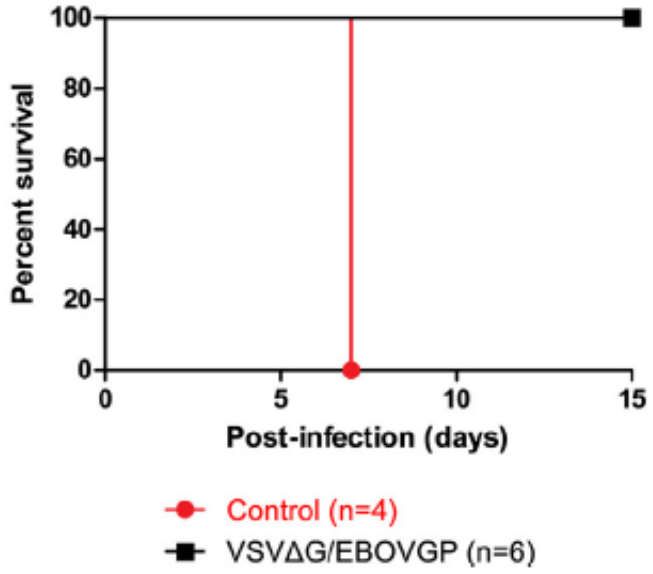
Mouse No.	Treatment begun	Dosage of penicillin	Route	Illness began	Survival period	Remarks
				days	*days*	
V C_1	Control			7	8	
V C_2	Control			6½	7	
V C_3	Control			6½	7	
V C_4	Control			6½	7¾	
V C_5	Control			7	7½	
V C_6	Control			6½	7½	
V P_1	7 hrs. after injection	930 units per day in divided doses at 9 a.m., 11 a.m., 1 p.m., 4 p.m., 7 p.m., 9 p.m., and 12 mid.	I.P.	No illness		Survived
V P_2			I.P.	No illness		Survived
V P_3			I.P.	No illness		Survived
V P_4			I.P.	No illness		Survived
V P_5			I.P.	No illness		Survived
V P_6			I.P.	No illness		Survived

VSV vaccine for EBOLA 2016

BALB/c mice, challenged at 12 months



Hartley Guinea Pig, challenged at 18 months

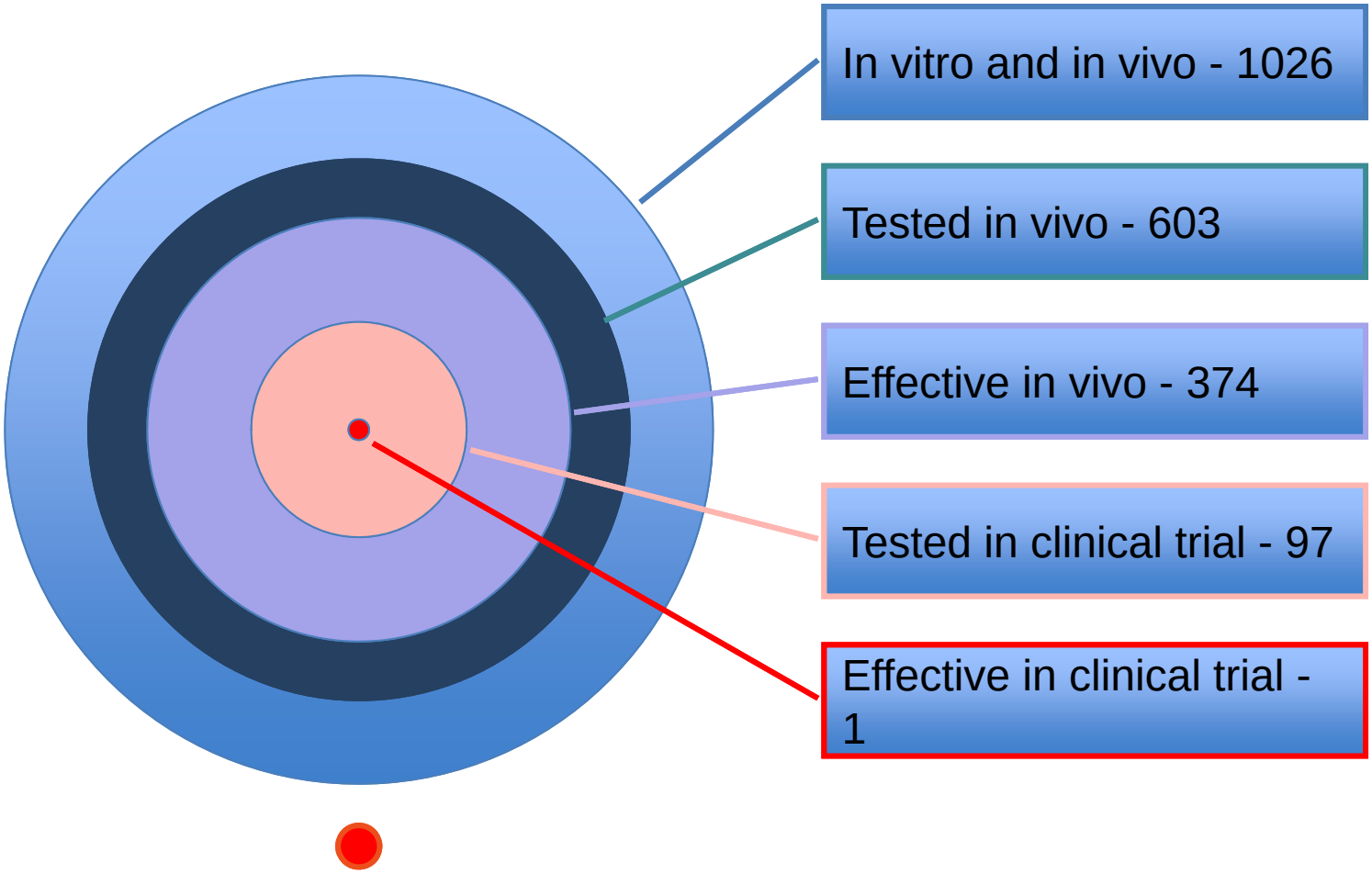


Definitions

Research: the process of producing “facts” (or rather predictions) which can be used by yourself or others to inform further research, clinical practice or other exploitation

Research Improvement Activity: Things done by stakeholders to increase the usefulness of research with which they are associated: either by the choice of research question or the certainty of the predictions made

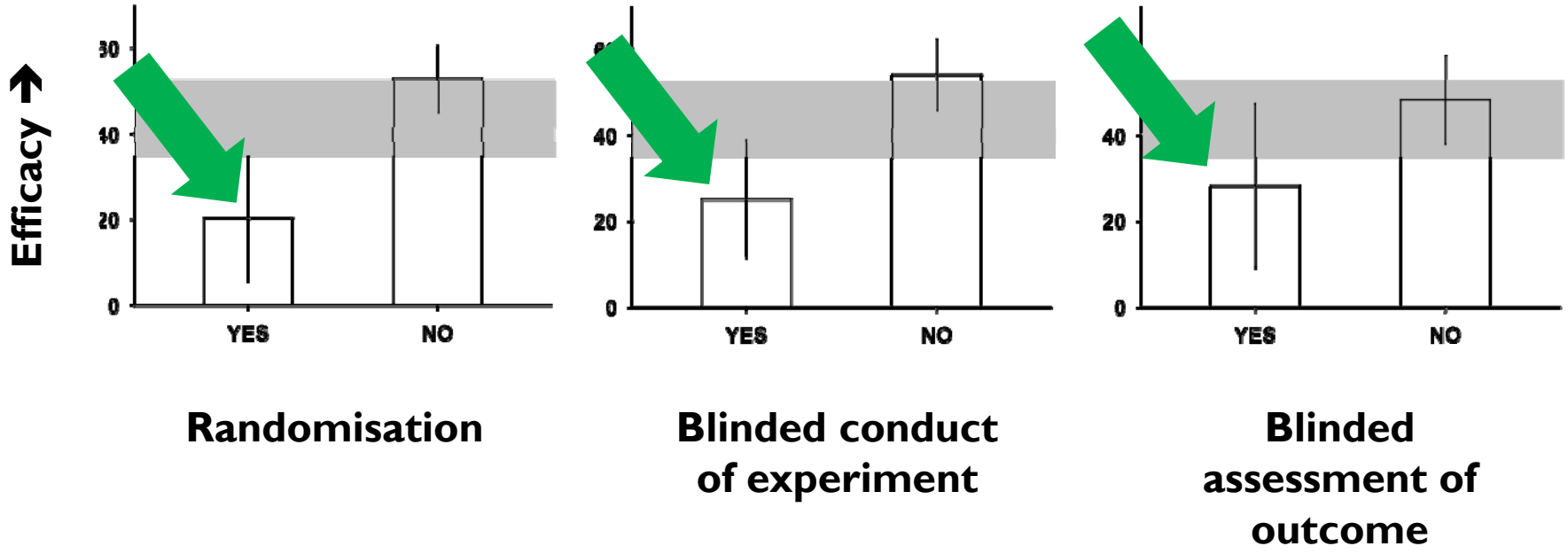
1026 interventions in experimental stroke



O'Collins et al, 2006

Risk of bias in animal studies

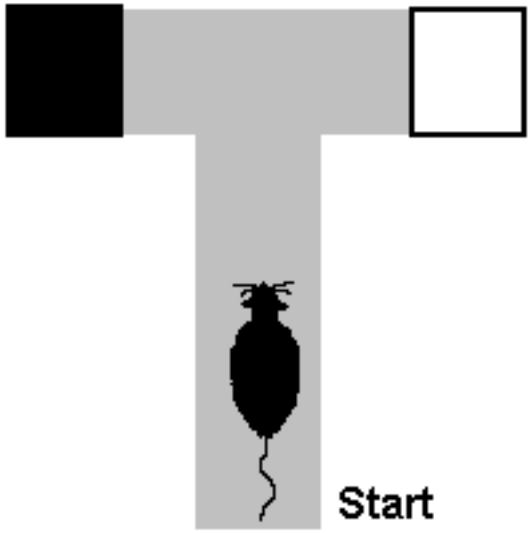
- Infarct Volume
  - 11 publications, 29 experiments, 408 animals
  - Improved outcome by 44% (35-53%)



Macleod et al, 2008

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"

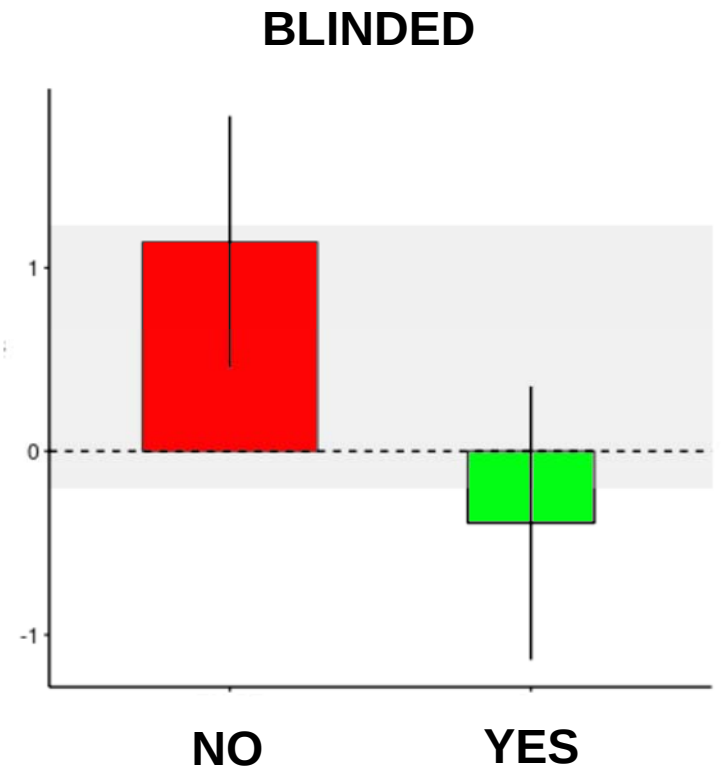


Group	Day 1	Day 2	Day 3	Day 4	Day 5
"Maze bright"	1.33	1.60	2.60	2.83	3.26
"Maze dull"	0.72	1.10	2.23	1.83	1.83
Δ	+0.60	+0.50	+0.37	+1.00	+1.43

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

Treating “depression” in animals with probiotics

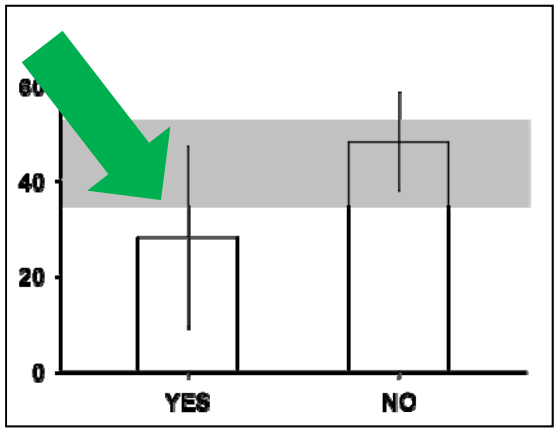
ROB item	Percent Reporting
Random Allocation to Group	25%
Blinded Assessment of Outcome	44%
Sample Size Calculation	6%
Reporting of Animal Exclusions	12%



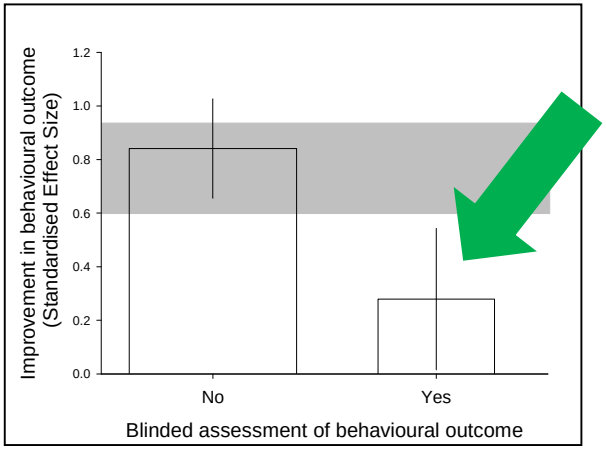
Credit: Anthony Shek

Evidence from various neuroscience domains …

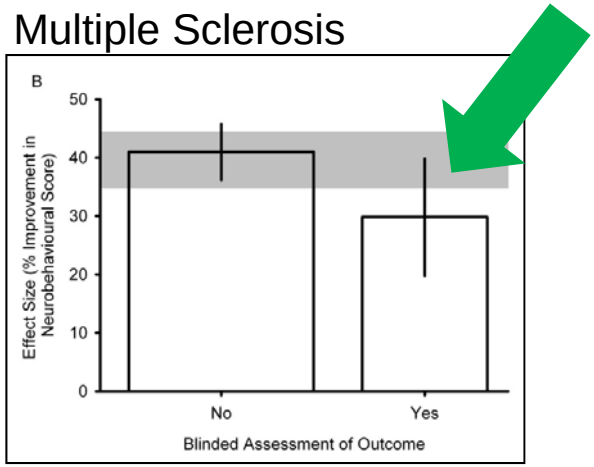
Stroke



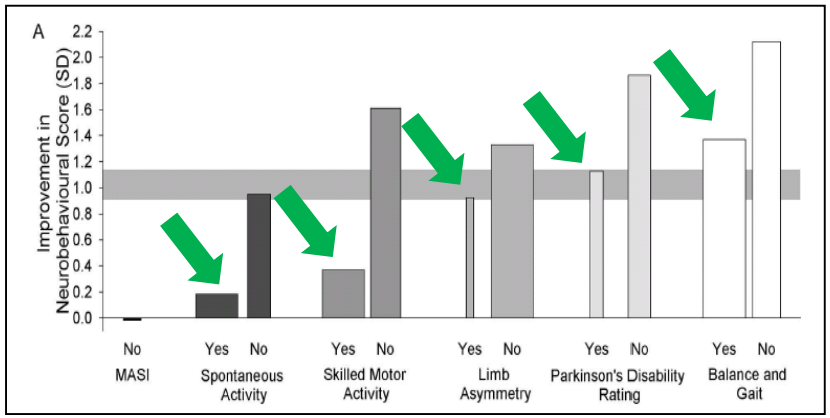
Alzheimer´s disease



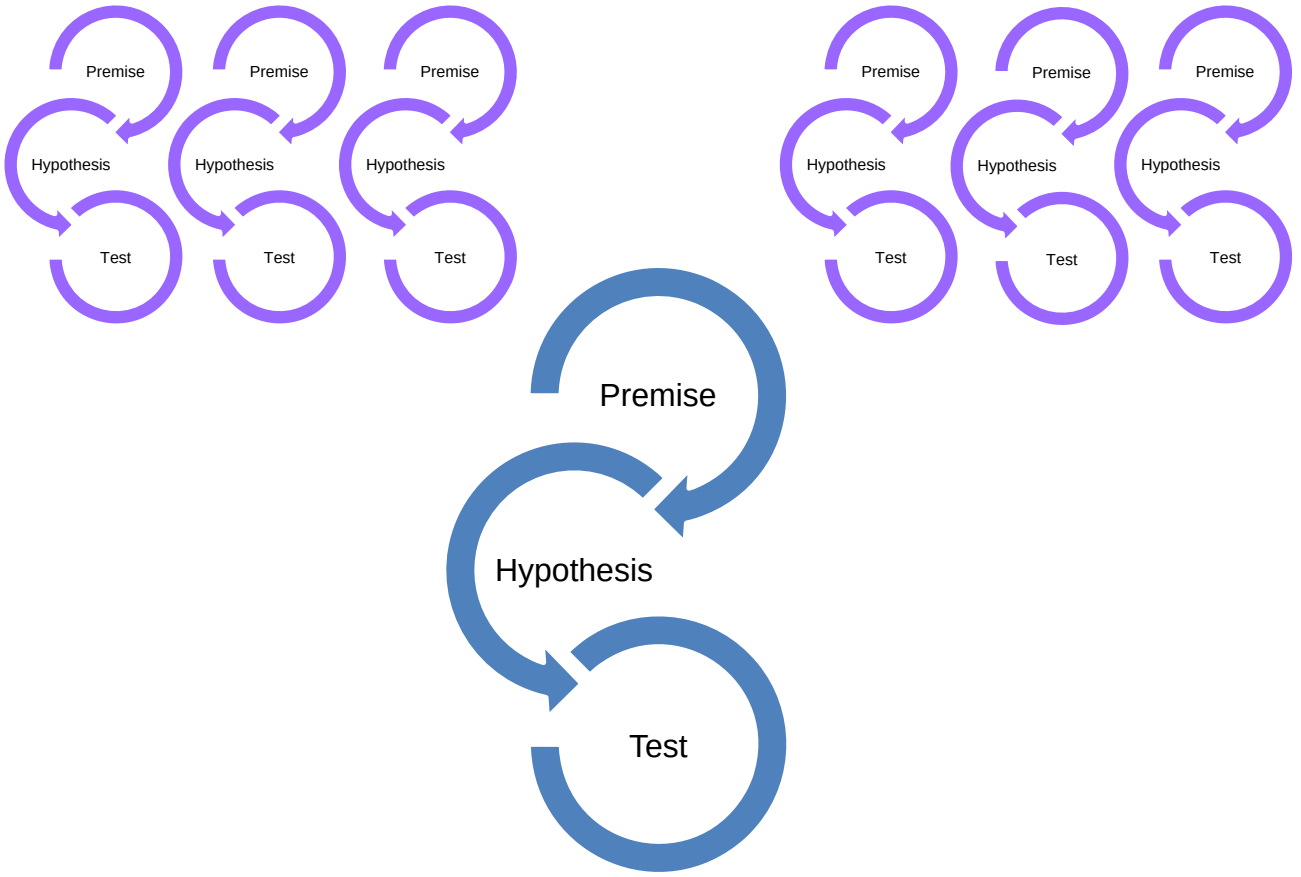
Multiple Sclerosis



Parkinson´s disease



The (polluted) research cycle



The scale of the problem

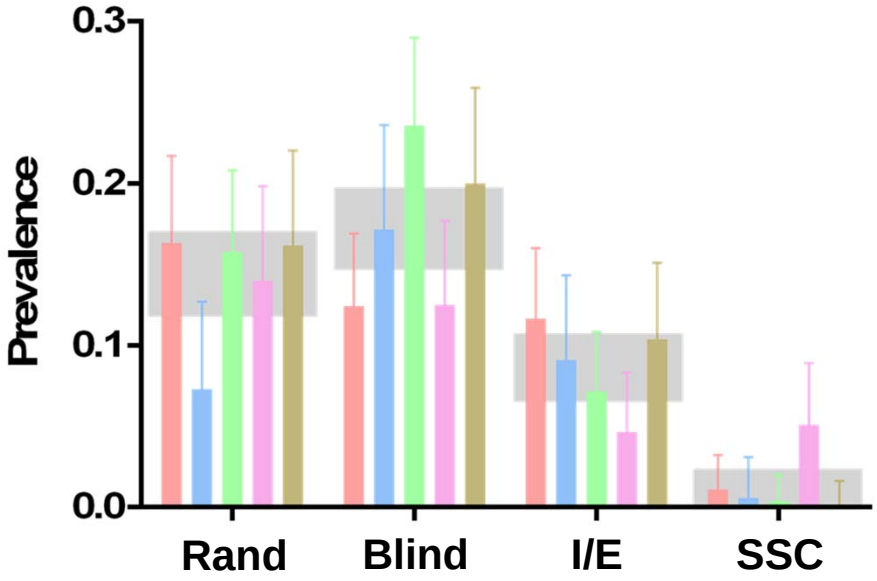
RAE 1173

rae2008
Research Assessment Exercise

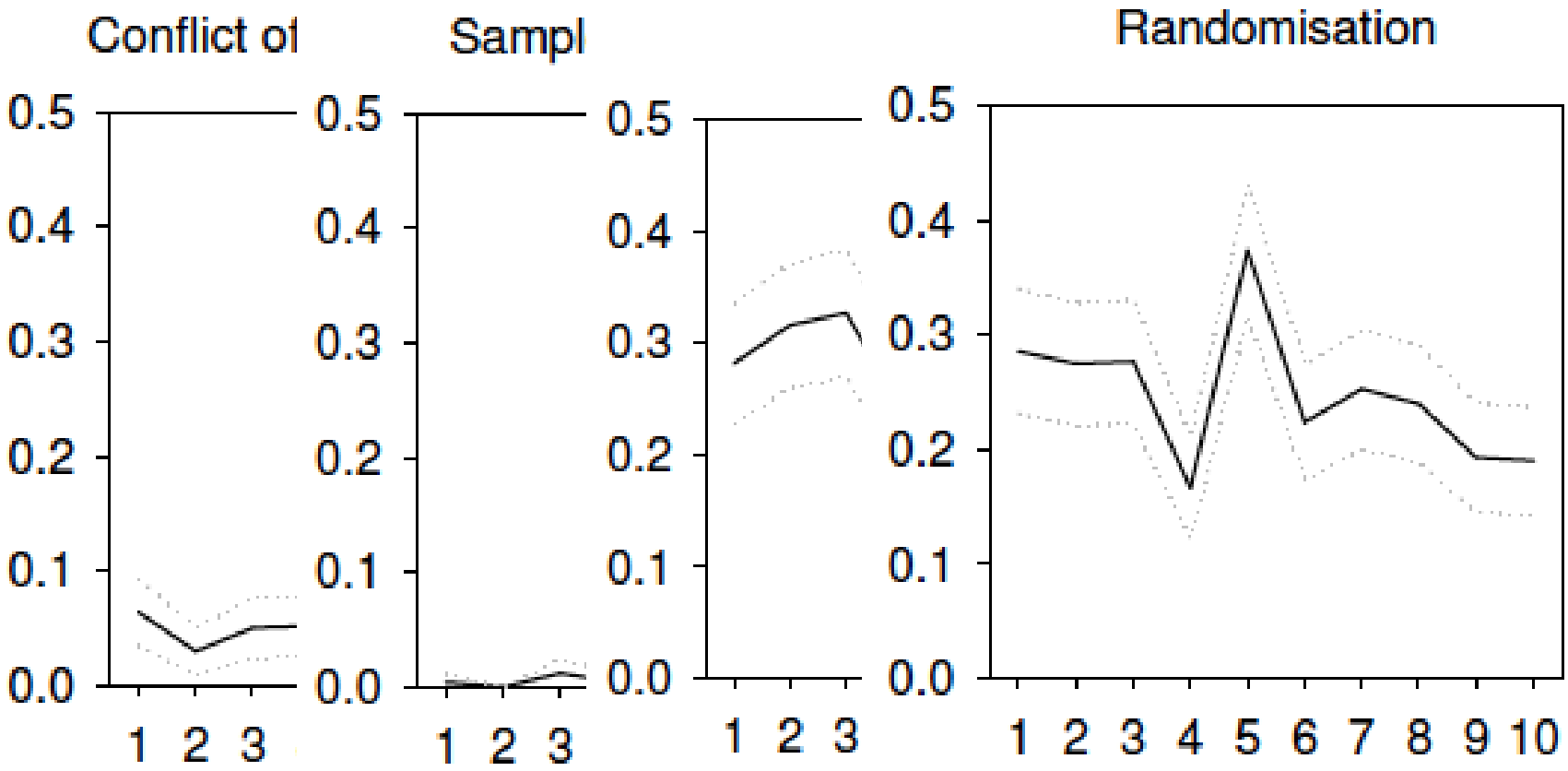
"an outstanding contribution to the internationally excellent position of the UK in biomedical science and clinical/translational research."

"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

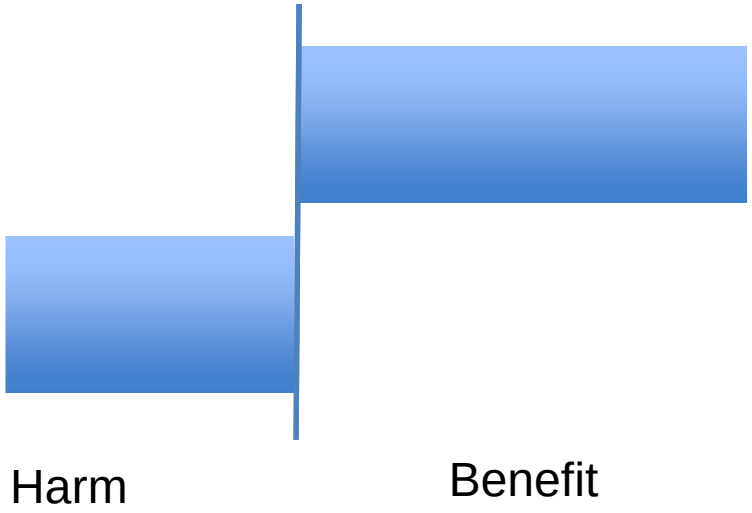


Reporting of risk bias items by decile of journal impact factor



Different patterns of publication bias in different fields

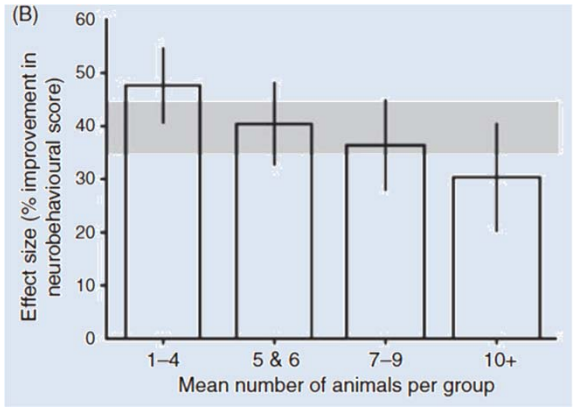
	outcome	observed	corrected	
Disease models	improvement	40%	30%	Less improvement
Toxicology model	harm	0.32	0.56	More harm

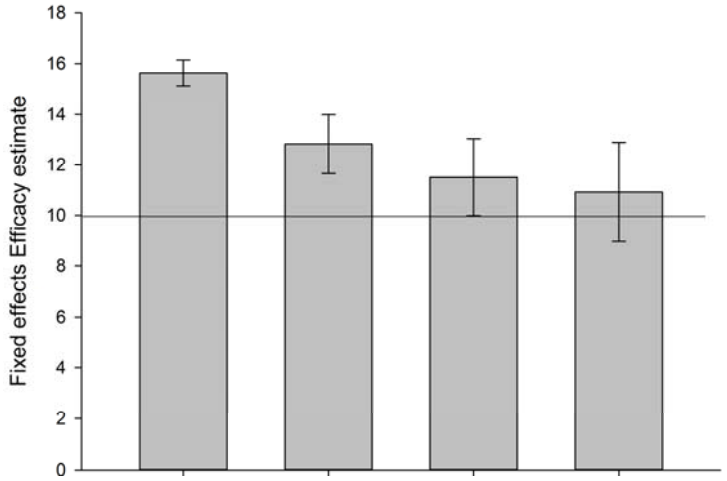


Small group sizes and publication bias conspire together

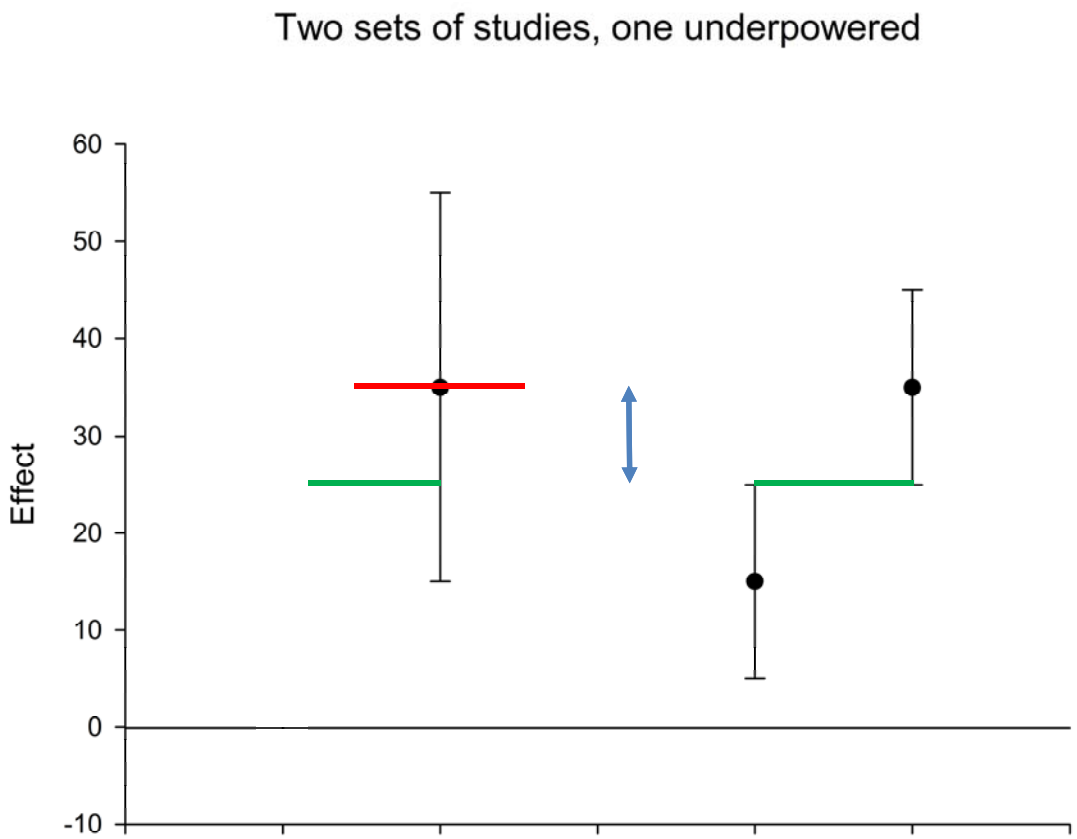
Simulation: 1000 studies
Complete publication bias (anything p>0.05 unpublished)
True effect size 10, SD 10

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

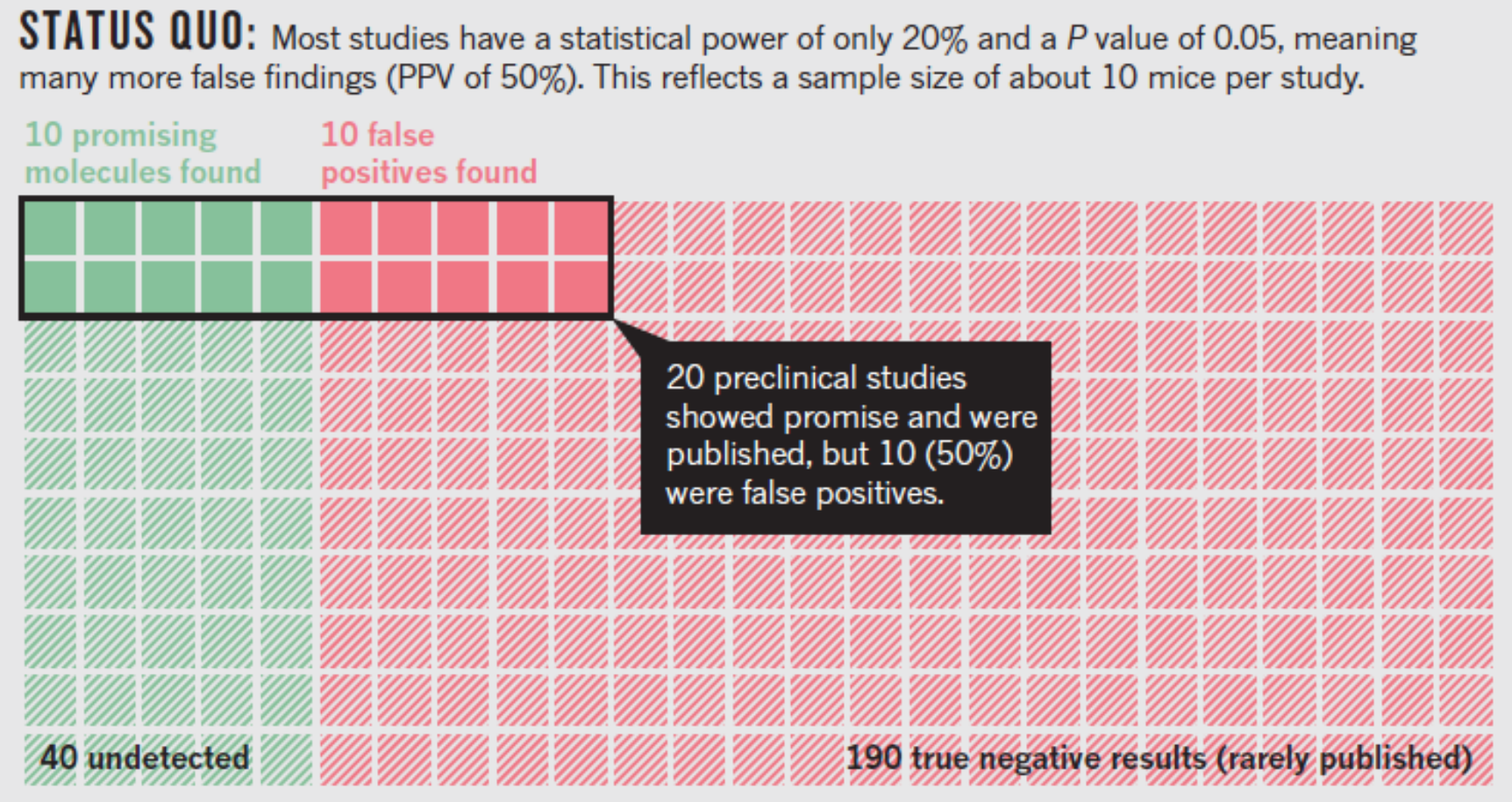




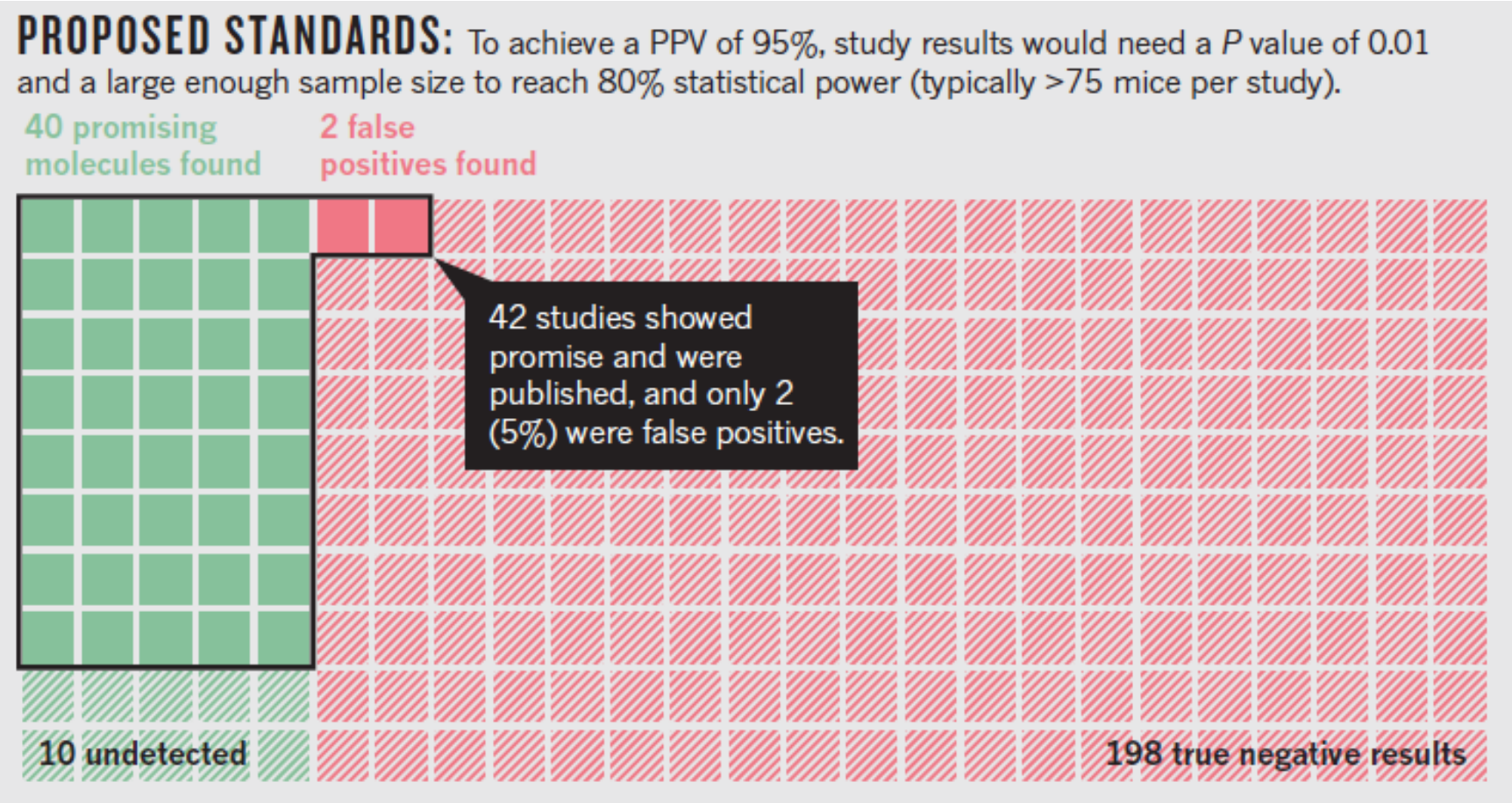
How does that work?



Say 250 studies ...



…with p<0.05, power @ 80%

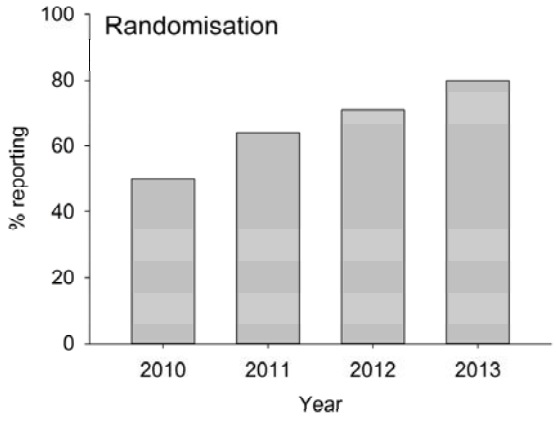


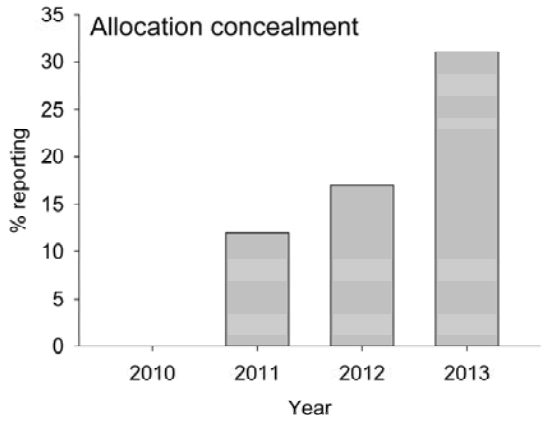
Comments, Opinions, and Reviews

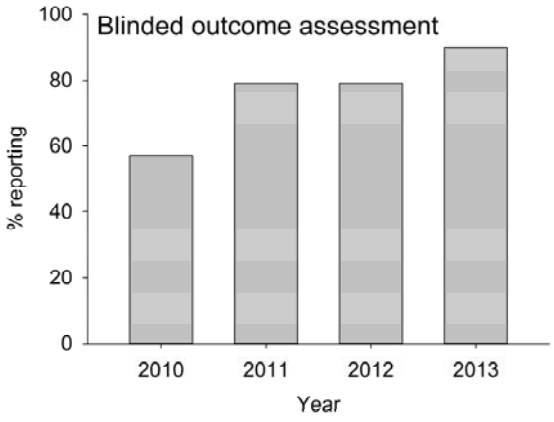
Good Laboratory Practice

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 the *Stroke* Basic Science Checklist, stratified by journal of publication

	Period 1* *n* (%)	Period 2* *n* (%)	Crude OR (95% CI)	*P*	Adjusted OR (95% CI)†	*P*†
Circulation	*n*=464	*n*=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
Circulation Research	*n*=303	*n*=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
Hypertension	*n*=485	*n*=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
Stroke	*n*=316	*n*=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
ATVB	*n*=476	*n*=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

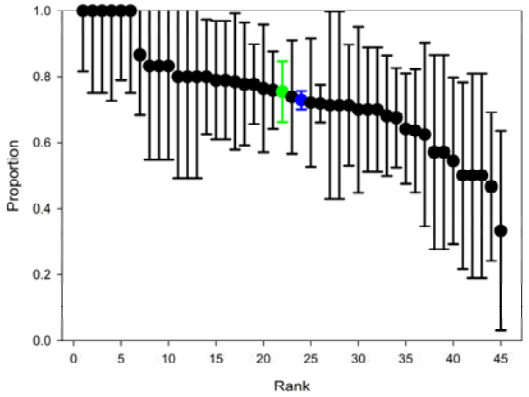
NR: not reported due to small number of events per predictor variable; OR: odds ratio

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

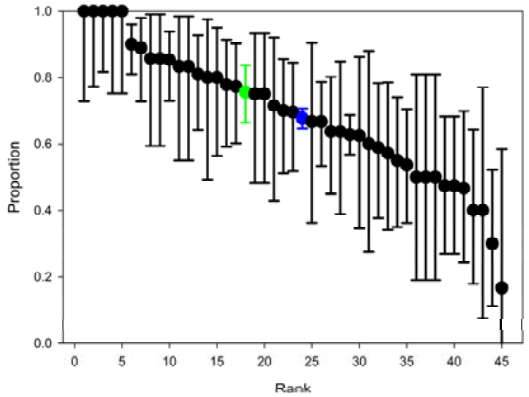
†Adjusted for cardiovascular disease studied and animal model used

Quality reporting by Journal MCAO 2014-16

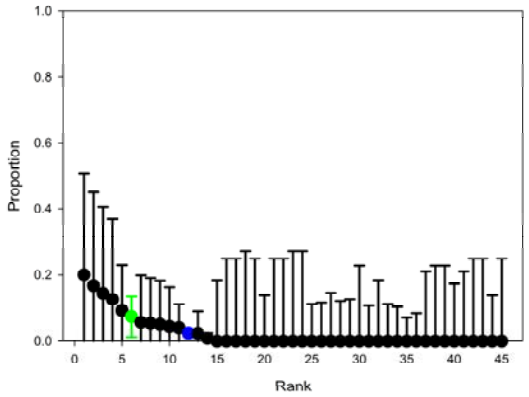
Randomisation



Blinding



Power calculation



Total in Blue
PLoS One in Green

Bahor et al Clinical Science 2017

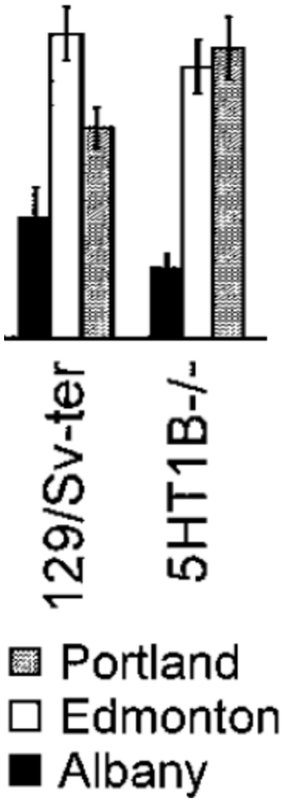
The replication difficulty ….

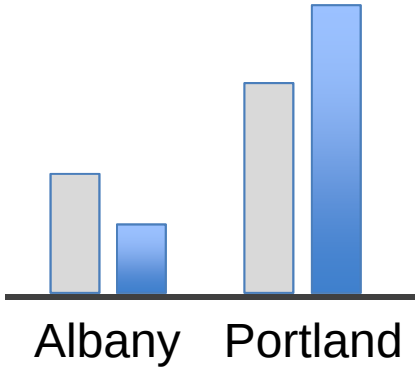
- Bayer: 53 of 67 findings did not replicate
- Amgen: 47 of 53 findings did not replicate
- Psychology:
  - positive findings in
    - 97% of original studies
    - 36% of replications
  - Mean effect size fell from 0.403 to 0.197
- Cancer Biology:
  - 3 of 5 did not replicate

What are the causes?

- ? Fraud
- ? False positive studies +/- dubious research practices
- ? Meta- (sectoral) problems like perverse incentives and publication bias
- ? True biological heterogeneity of observed effects

Crabbe (Science 1999)





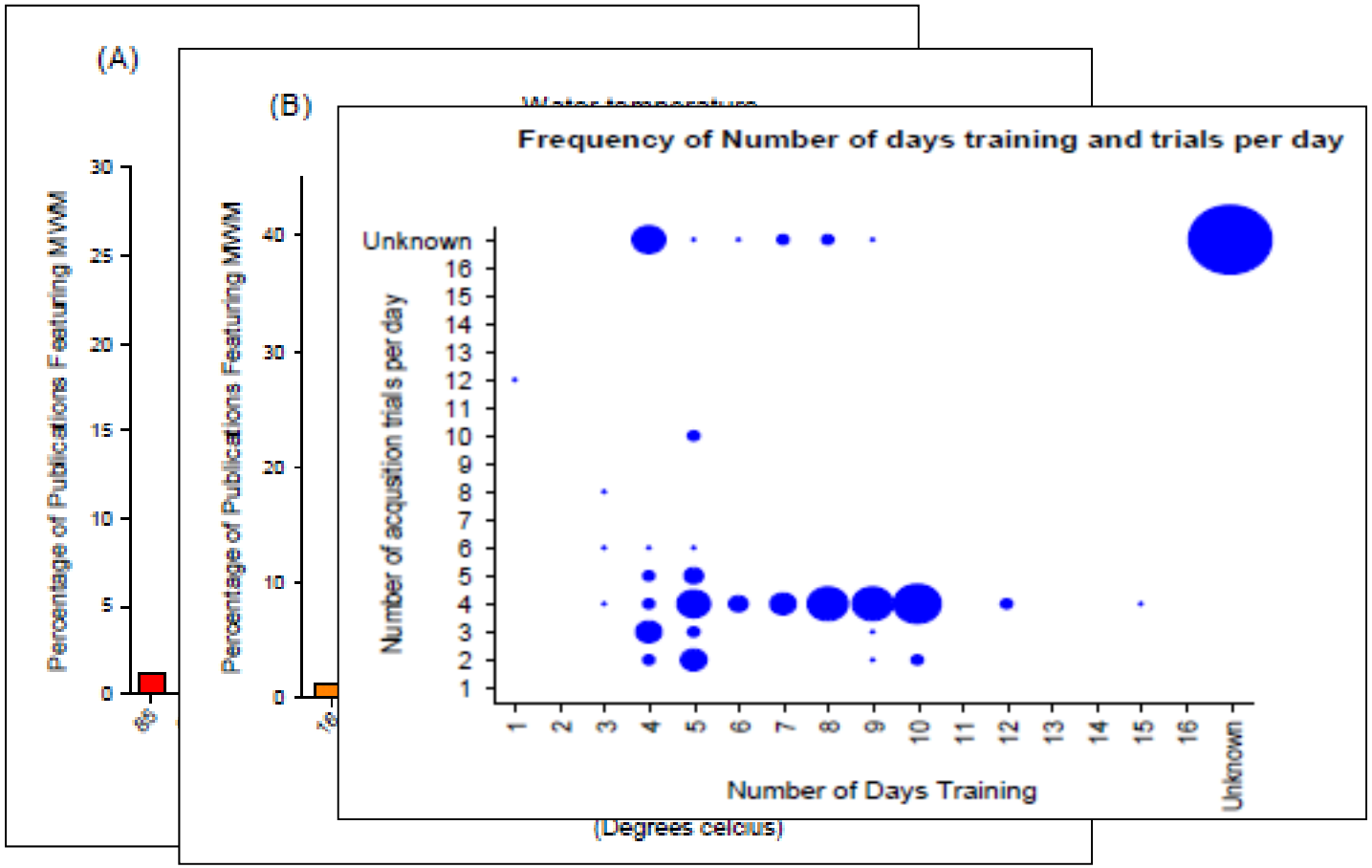
Task	Measure	Eight Genotypes	Three Sites	Two Sexes	Local vs Shipped	Genotype x Site	Genotype x Sex	Genotype x Ship	Multiple R^2
Open field	Distance in 15 min	.600	.157	----	----	.059	.045	----	.604
Open field	# vertical movements	.788	.281	.039	----	----	----	----	.772
Cocaine	Difference from Day 1	.338	.053	----	----	.086	----	----	.342
Plus maze	Total arm entries	.385	.327	----	----	.210	----	----	.660
Plus maze	Time in open arms	.082	.212	----	----	.066	----	----	.266
Water maze	Mean escape latency	.221	----	----	.026	----	----	----	.177
Alcohol preference	Alcohol consumed (g/kg)	.483	----	.043	----	----	----	----	.451
Body size	Weight (g)	.408	.204	.637	----	.071	.070	----	.698

Terms (Goodman et al)

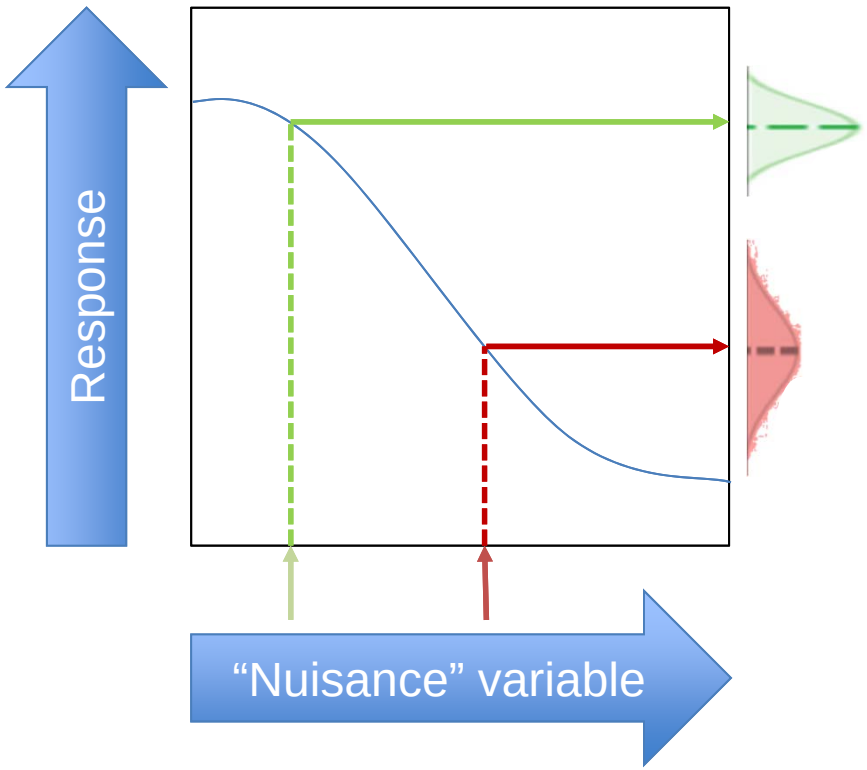
- **Methods reproducibility** is the ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results.
- **Results reproducibility** is the production of corroborating results in a new study, having followed the same experimental methods.
- **Inferential reproducibility** is the making of knowledge claims of similar strength from a study replication or reanalysis.
- **Robustness:** the stability of experimental conclusions to variations in either baseline assumptions or experimental procedures
- **Generalizability:** the persistence of an effect in settings different from and outside of an experimental framework.

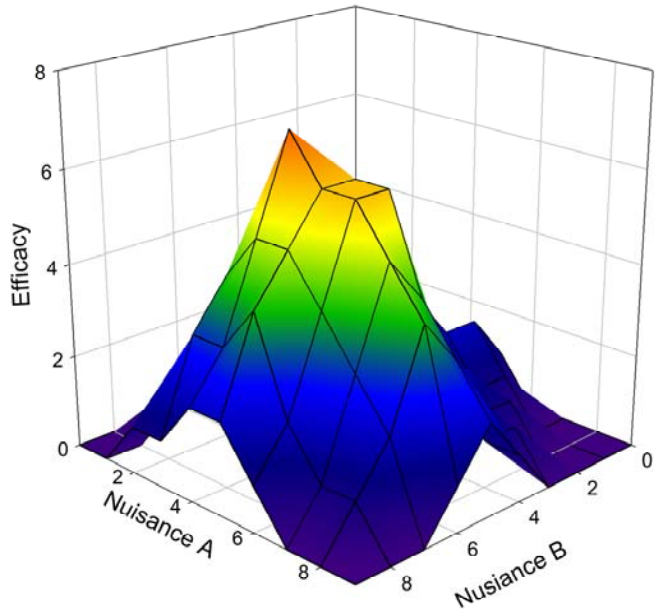
Methods reproducibility	Results reproducibility	Robustness
Exactly the same	The same methods	Variations in baseline assumptions and experimental procedures

Use of the Morris Water Maze



Reaction norms (Voelkl 2016)





Reflections

- Nuisance variables may be known or unknown
- Sampling the impact of nuisance variables without knowing what you are dealing with is only preliminary (i.e. “Do they exist?”)
- Investigating the impact of known and potential nuisance variables is science – coordinated, organised, stratified multi-centre studies

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%

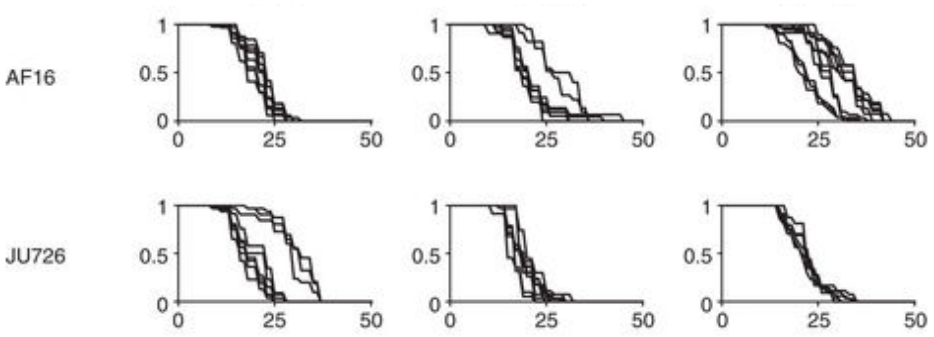
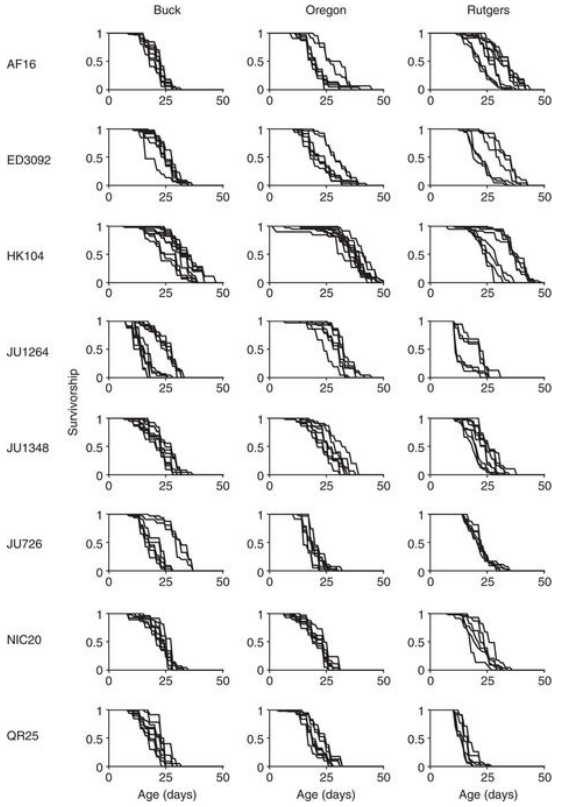
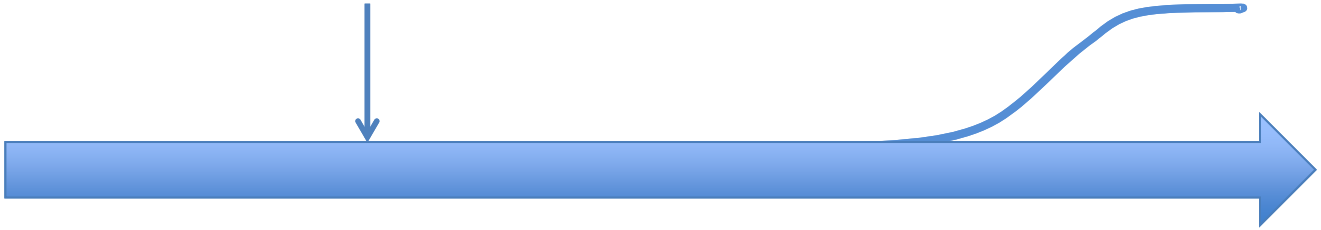


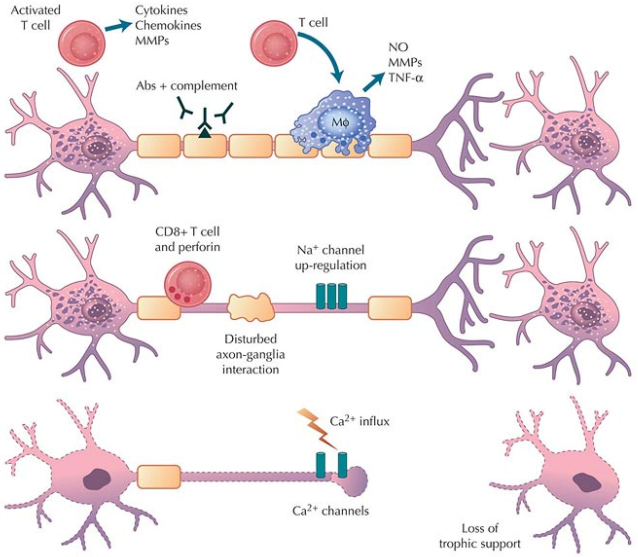
Figure 3: Variation in longevity within labs for each replicate plate for eight natural isolates of *C. briggsae*.



Lucanic et al Nature Comms 2017

Pathophysiology

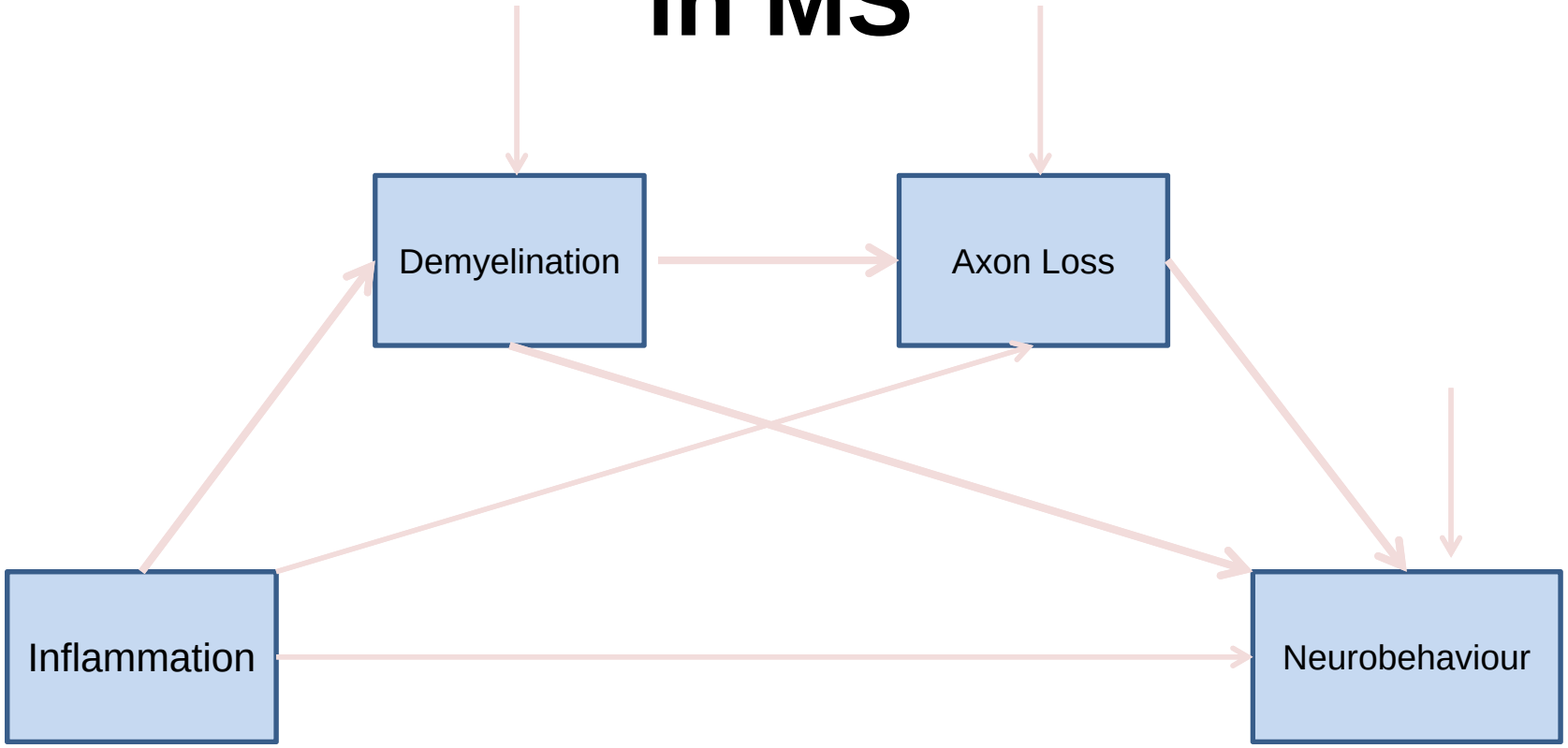




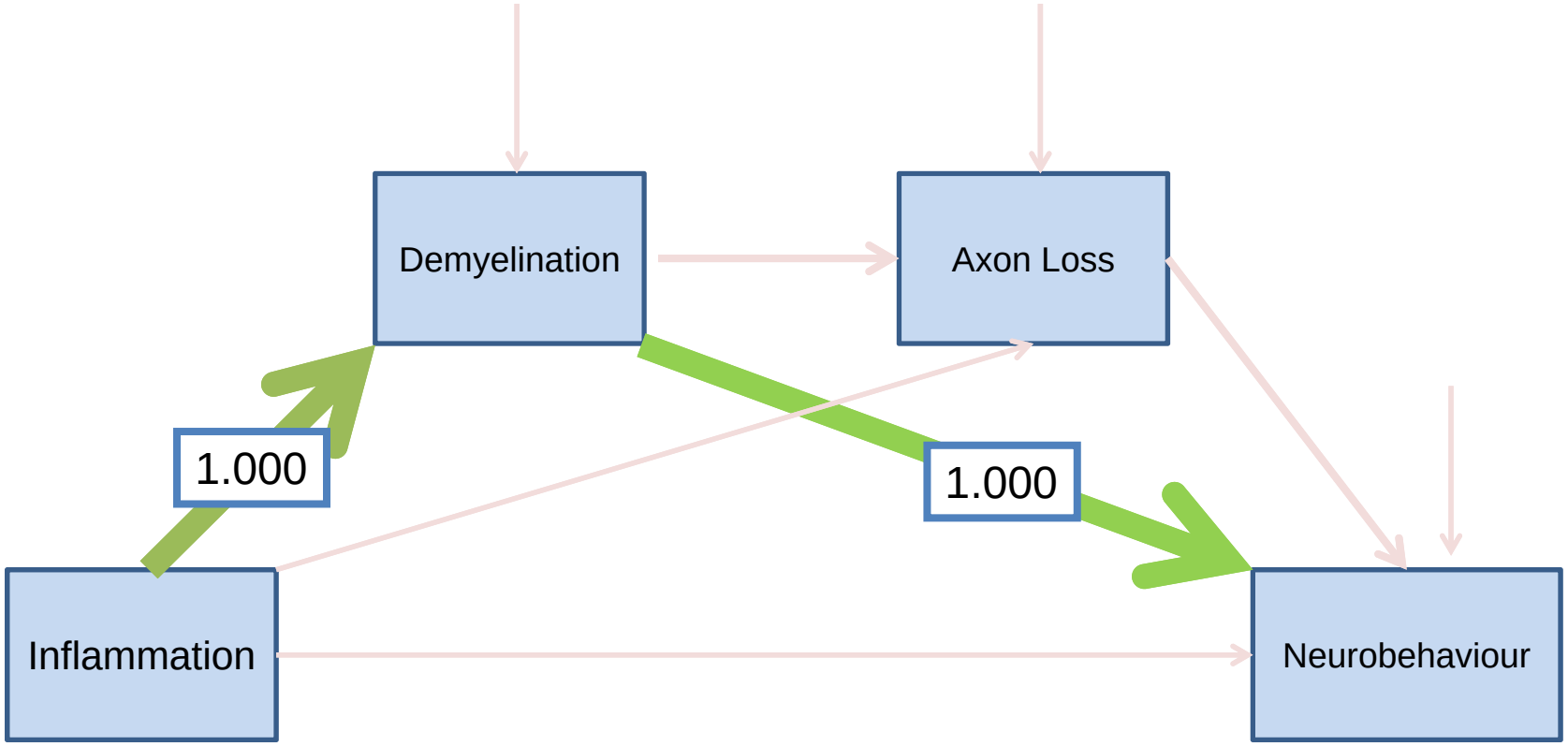
287 experiments identified in SyR reporting > 1 outcome domain

- inflammation
- demyelination
- axon loss
- neurobehaviour

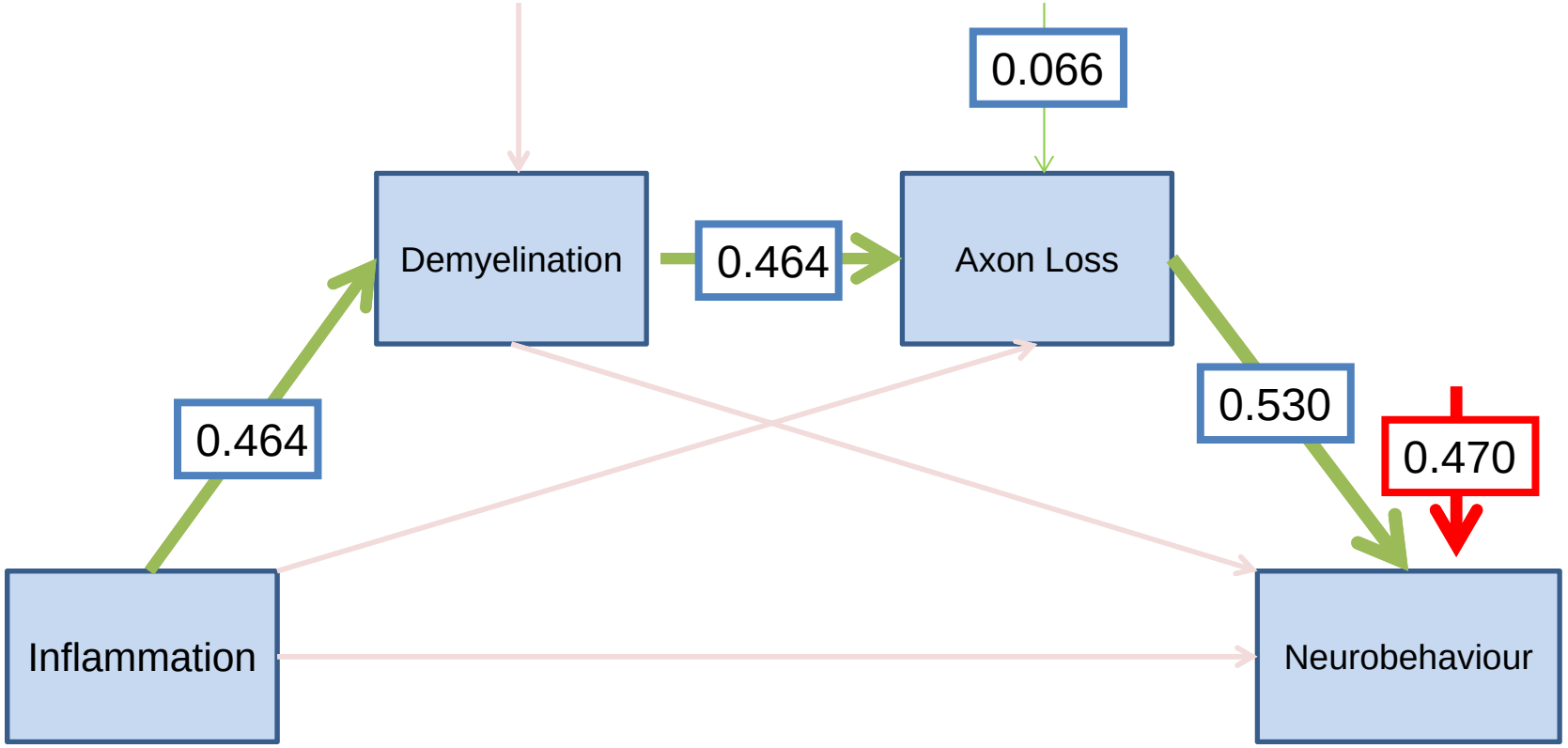
Drug efficacy in MS



Pre induction



Disease

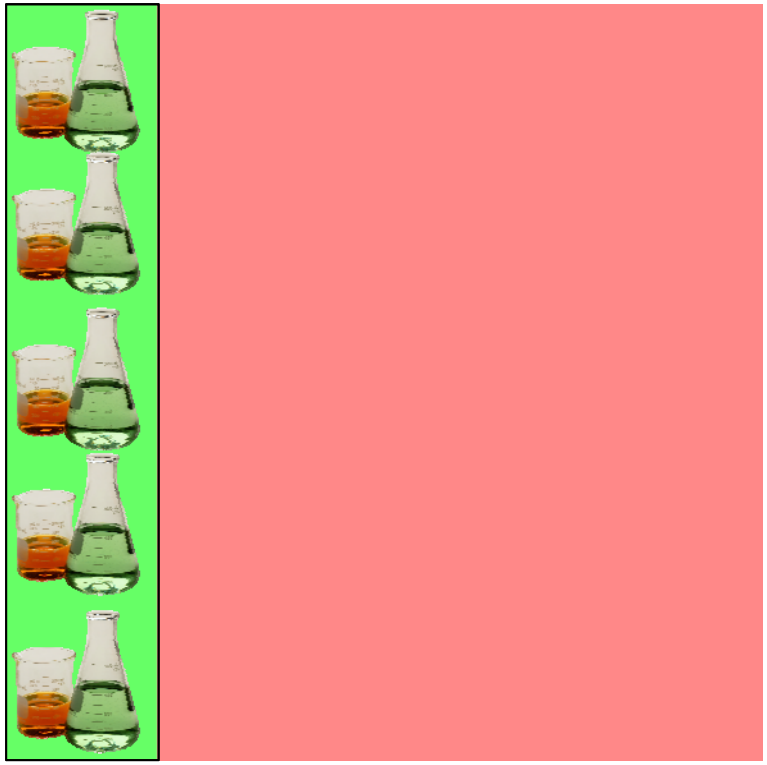

0.066
Demyelination
0.464
Axon Loss
0.530
0.470
0.464
Inflammation
Neurobehaviour

Why much published research is false ...

Why much published research is false …

Assume:

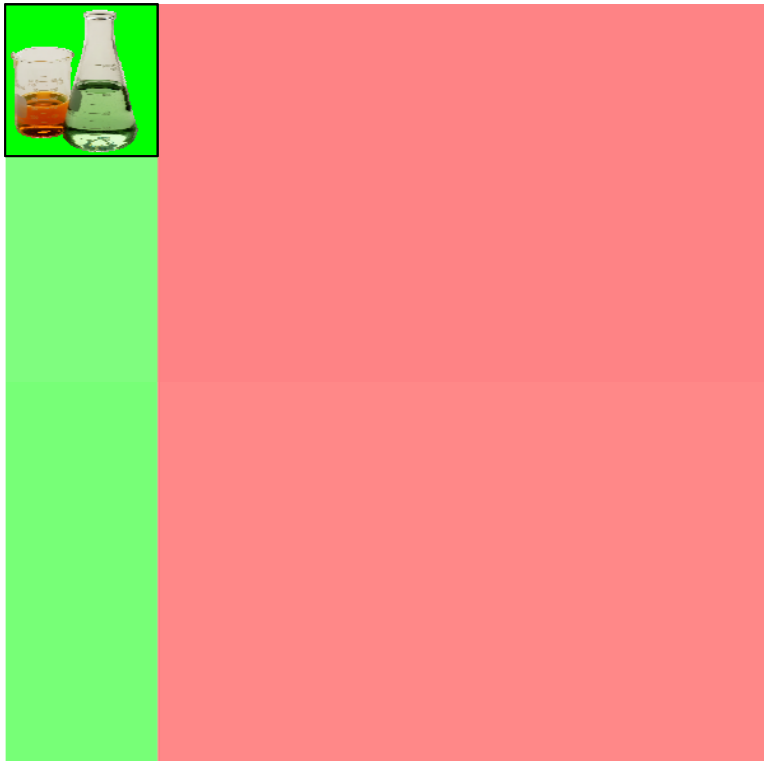
- 20% of hypotheses in a field are correct



Why much published research is false …

Assume:

- 20% of hypotheses in a field are correct
- Power to detect a biologically important effect of 20%



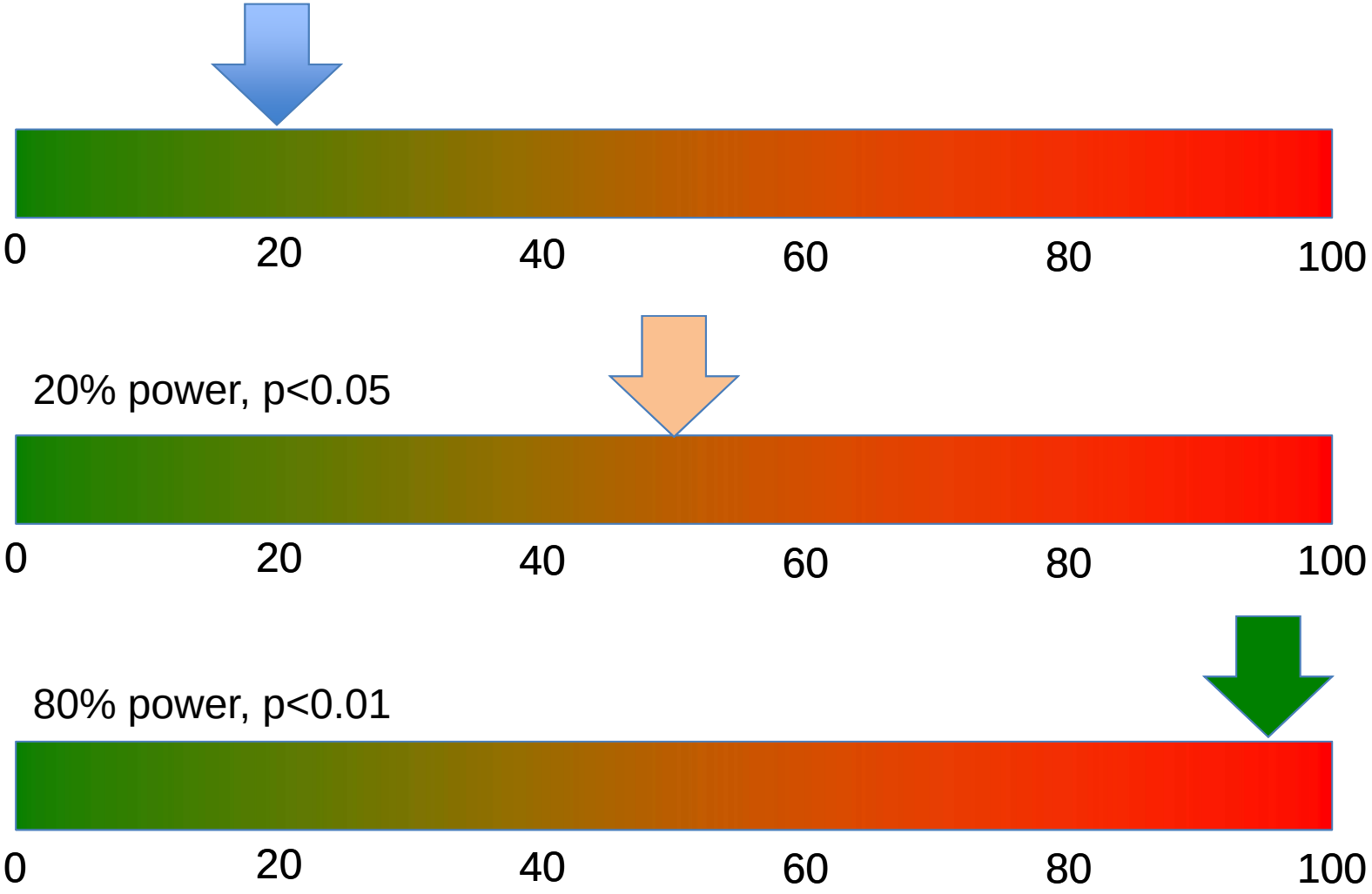
Why much published research is false …

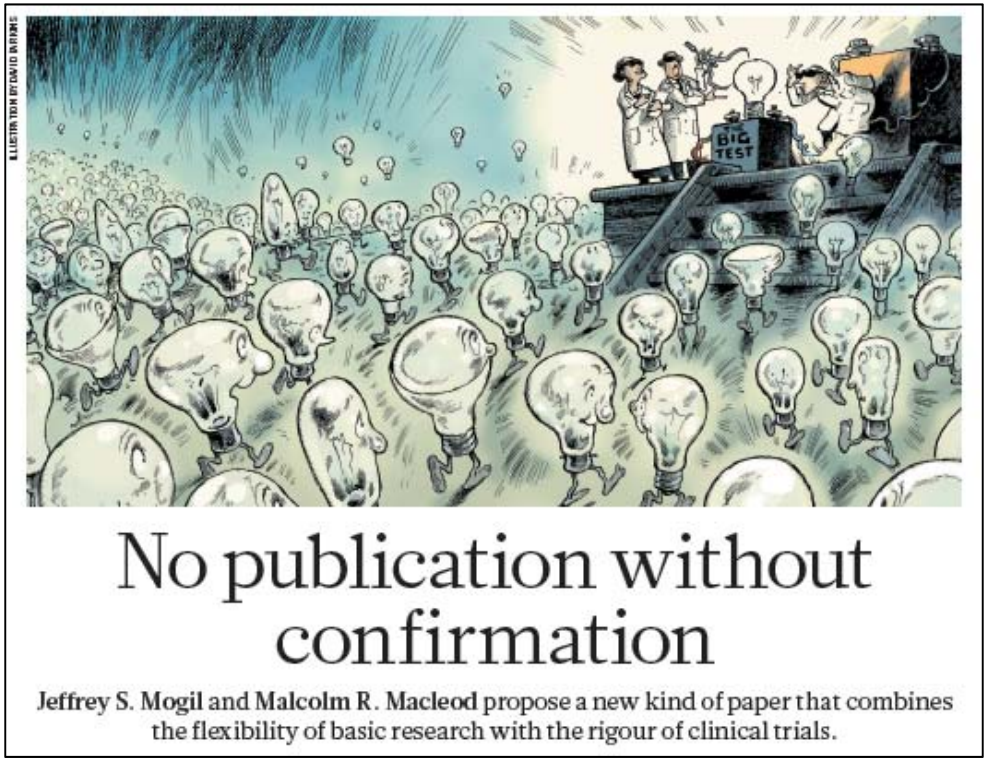
Assume:

- 20% of hypotheses in a field are correct
- Power to detect a biologically important effect of 20%
- Critical p threshold of 5%

Positive predictive value = 50%

Value of information added





No publication without confirmation

Jeffrey S. Mogil and **Malcolm R. Macleod** propose a new kind of paper that combines the flexibility of basic research with the rigour of clinical trials.

Develop hypothesis

Assert hypothesis

Test hypothesis

Protocols

EuroHYP-1
EudraCT-No 2012-002944-25

EuroHYP-1

EuroHYP-1: European multicentre, randomised, phase III clinical trial of therapeutic hypothermia plus best medical treatment versus best medical treatment alone for acute ischaemic stroke

189 pages, several years

Protocols

EuroHYP-1: European multicenter, randomized, phase III clinical trial of therapeutic hypothermia plus best medical treatment vs. best medical treatment alone for acute ischemic stroke

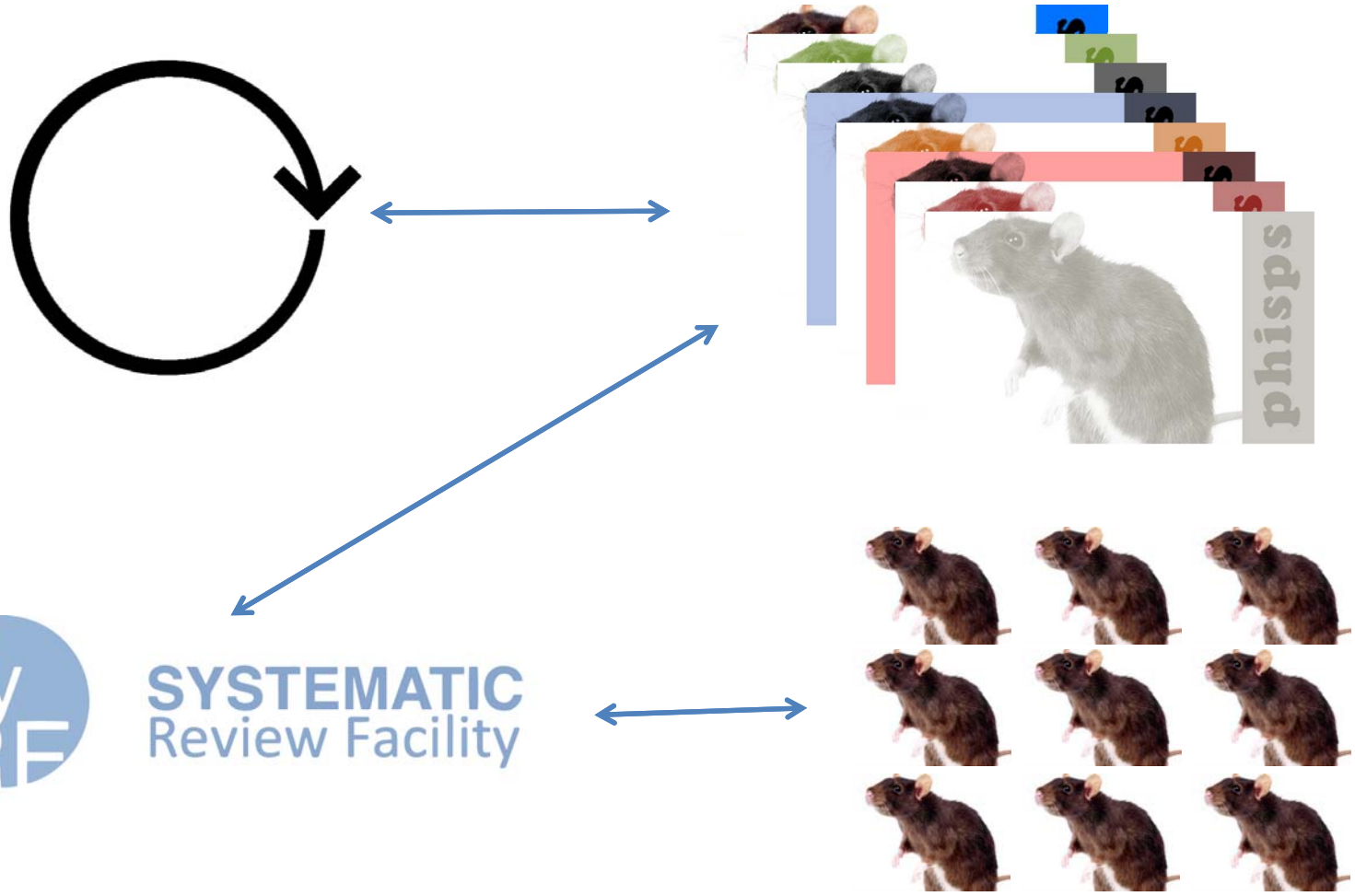
H. Bart van der Worp[1]*, Malcolm R. Macleod[2], Philip M. W. Bath[3], Jacques Demotes[4], Isabelle Durand-Zaleski[5], Bernd Gebhardt[6], Christian Gluud[7], Rainer Kollmar[8], Derk W. Krieger[9], Kennedy R. Lees[10], Carlos Molina[11], Joan Montaner[12], Risto O. Roine[13], Jesper Petersson[14], Dimitre Staykov[15], Istvan Szabo[16], Joanna M. Wardlaw[17], Stefan Schwab[15], and on behalf of the EuroHYP-1 investigators[†]

4 pages, one year

PHISPS protocols

- Population
- Hypothesis
- Intervention
- Sample size calculation
- Primary outcome measure
- Statistical analysis plan

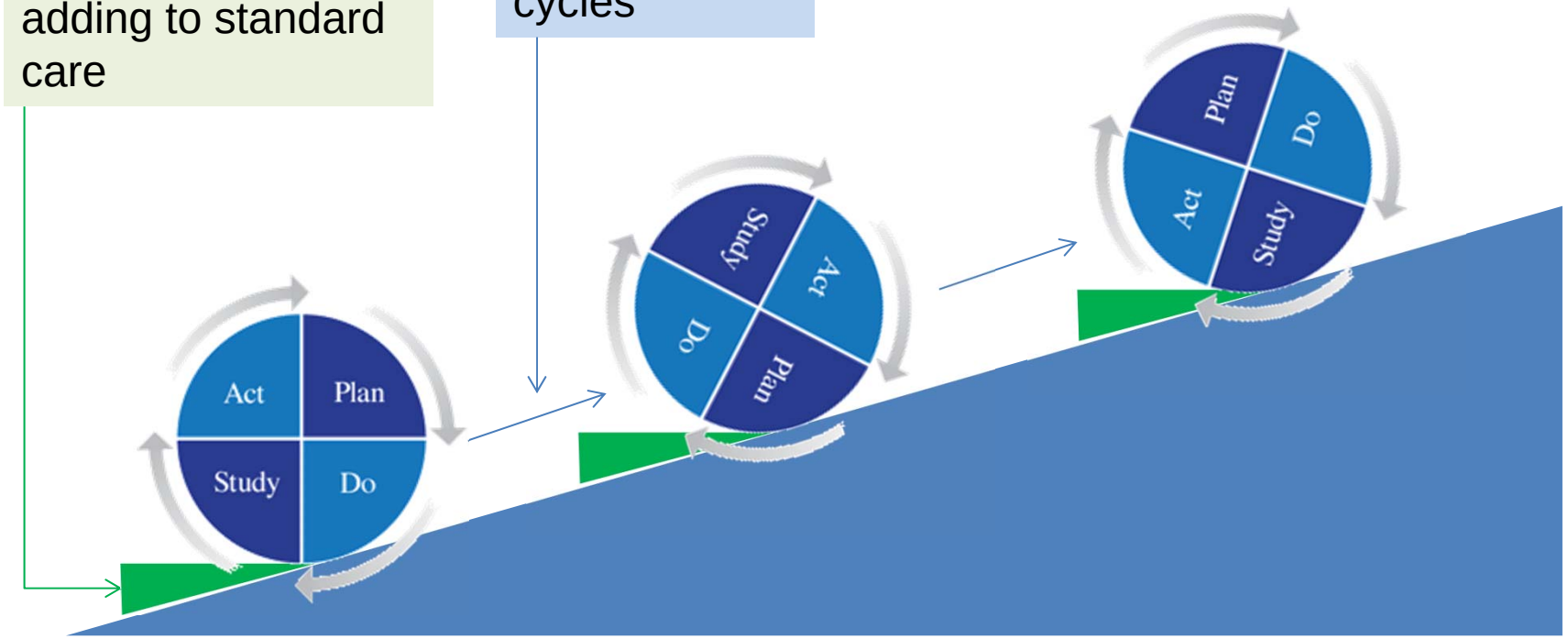
A proposal ...



Strategies to increase benefits from research

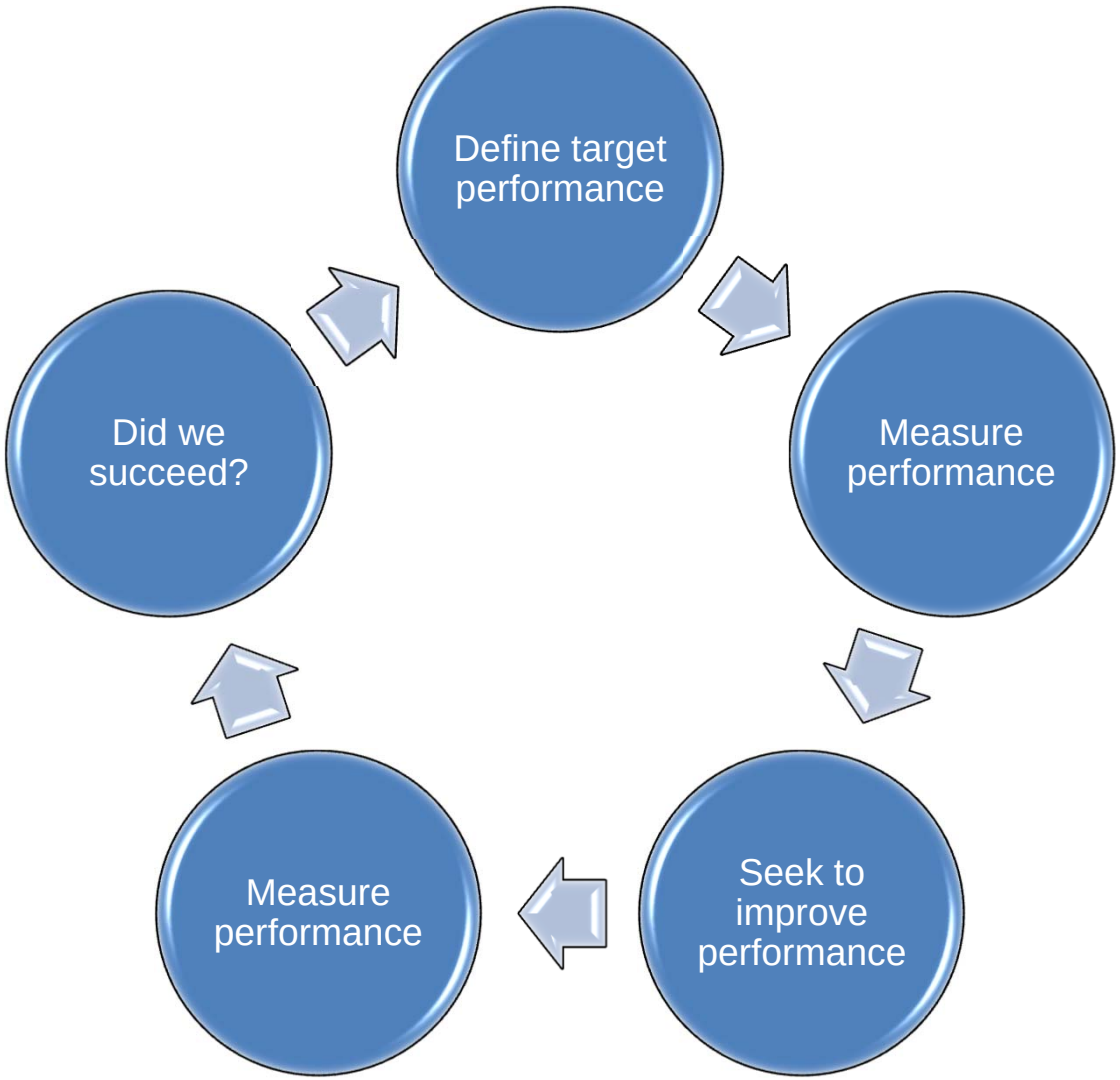
Consolidate through adding to standard care

Continuous improvement cycles



Research Improvement Activity

What performance do we aspire to?	"95% of UoE manuscripts describing animal research report randomisation where this would be appropriate"
What is our current performance?	Measure, eg with ML/TM [2009-10 = 8%]
What are we going to do about it?	• Education sessions for PhD/post Doc/ PIs • CPD for investigators • Highlighted component of AWERB review • Identified factor in resource allocation (open access publication funds, prioritisation of research resources)
Did that make a difference?	Measure, eg with ML/TM
Is performance now good enough?	Stick or twist


Define target performance
Measure performance
Seek to improve performance
Measure performance
Did we succeed?

Measure performance

Define “population”

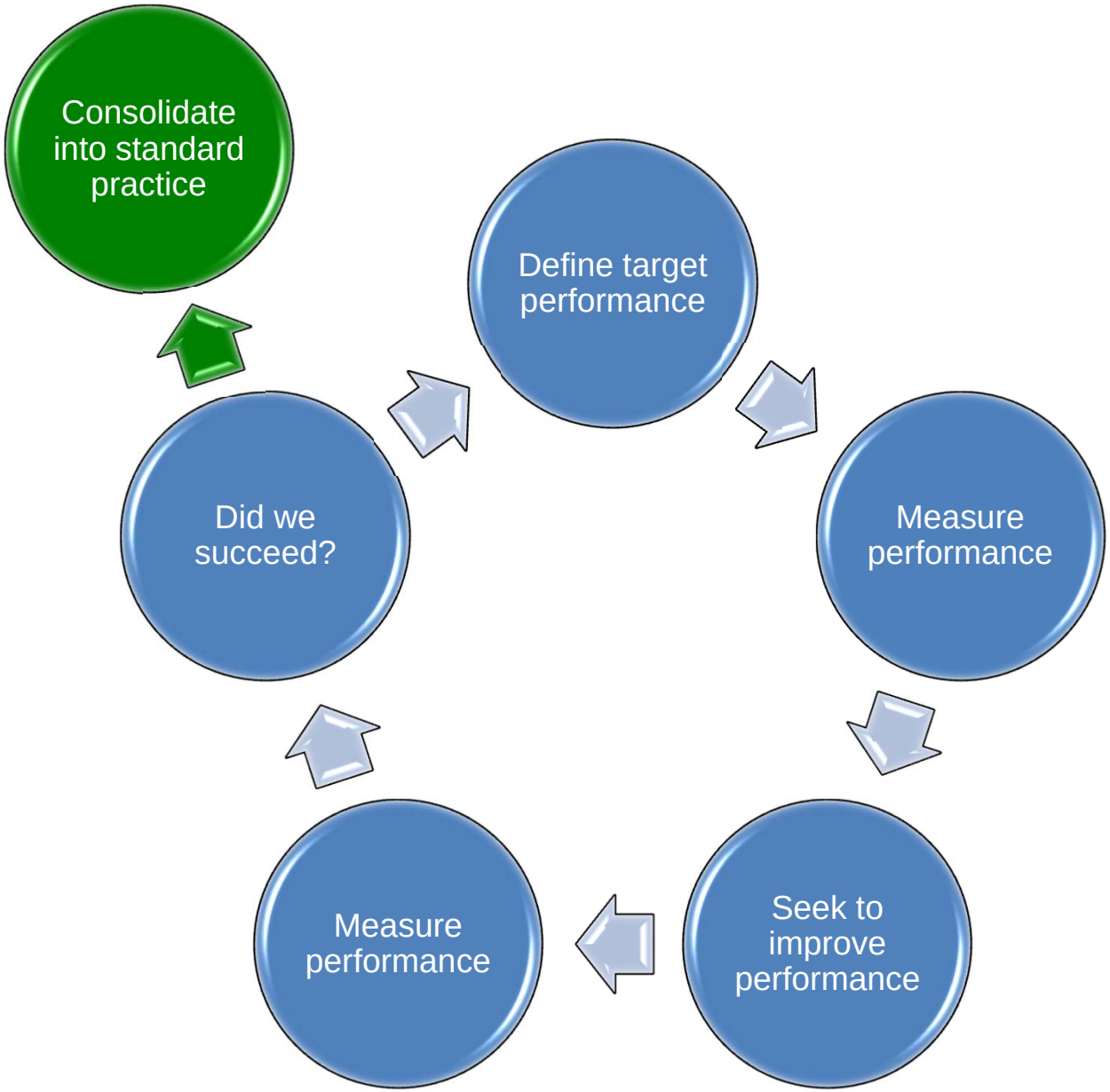
e.g. Research from Danish institutions reporting in vivo experiments

Identify research outputs

PubMed search by author affiliation	Title/Abstract screening to identify animal studies

Ascertain reporting of risks of bias

Retrieve full text	RegEx or Convoluted Neural Networks


Consolidate into standard practice
Define target performance
Measure performance
Seek to improve performance
Measure performance
Did we succeed?

Strategies to increase value

Level 1	Study reports comply with existing guidelines such as the ARRIVE guidelines, so that there is transparency in what was done
Level 2	Studies are conducted taking appropriate measures to reduce the risk of bias, such as randomisation, blinded conduct of the experiment and blinded assessment of outcome; and are planned on the basis of a coherent sample size calculation
Level 3	Study protocols, including statistical analysis plans, are determined in advanced and are archived such that research users can check where the study as executed deviated from the study as planned
Level 4	The existence of a study is asserted through some system of registration, to address the issue of publication bias
Level 5	The study is planned to have an appropriate positive predictive value, based on the likelihood of refuting the null hypothesis, the statistical power and the chosen Type 1 error; and this is asserted in advance, to avoid misinterpretation
Level 6	Formal strategies to assess the burden of evidence in favour of efficacy are developed, including but not limited to systematic review and meta-analysis of existing evidence and a GRADE-like approach to assess the strength of evidence
Level 7	Where the in vivo data appear promising, to develop tools for multicentre animal studies to confirm effects in "preclinical phase 3 studies"

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 496 > Issue 7446 > Editorial > Article

NATURE | EDITORIAL

Announcement: Reducing our irreproducibility

24 April 2013

PDF | Rights & Permissions

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects

Related stories

- If a job is worth doing, it

eveals
on Bia
s, PhD; Ge
l;
ematsu;
nature11556

BOX 1

A core set of reporting standards for rigorous study design

Randomization

- Animals should be assigned randomly to the various experimental groups, and the method of randomization reported.
- Data should be collected and processed randomly or appropriately blocked.

Blinding

- Allocation concealment: the investigator should be unaware of the group to which the next animal taken from a cage will be allocated.
- Blinded conduct of the experiment: animal caretakers and investigators conducting the experiments should be blinded to the allocation sequence.
- Blinded assessment of outcome: investigators assessing, measuring or quantifying experimental outcomes should be blinded to the intervention.

Sample-size estimation

- An appropriate sample size should be computed when the study is being designed and the statistical method of computation reported.
- Statistical methods that take into account multiple evaluations of the data should be used when an interim evaluation is carried out.

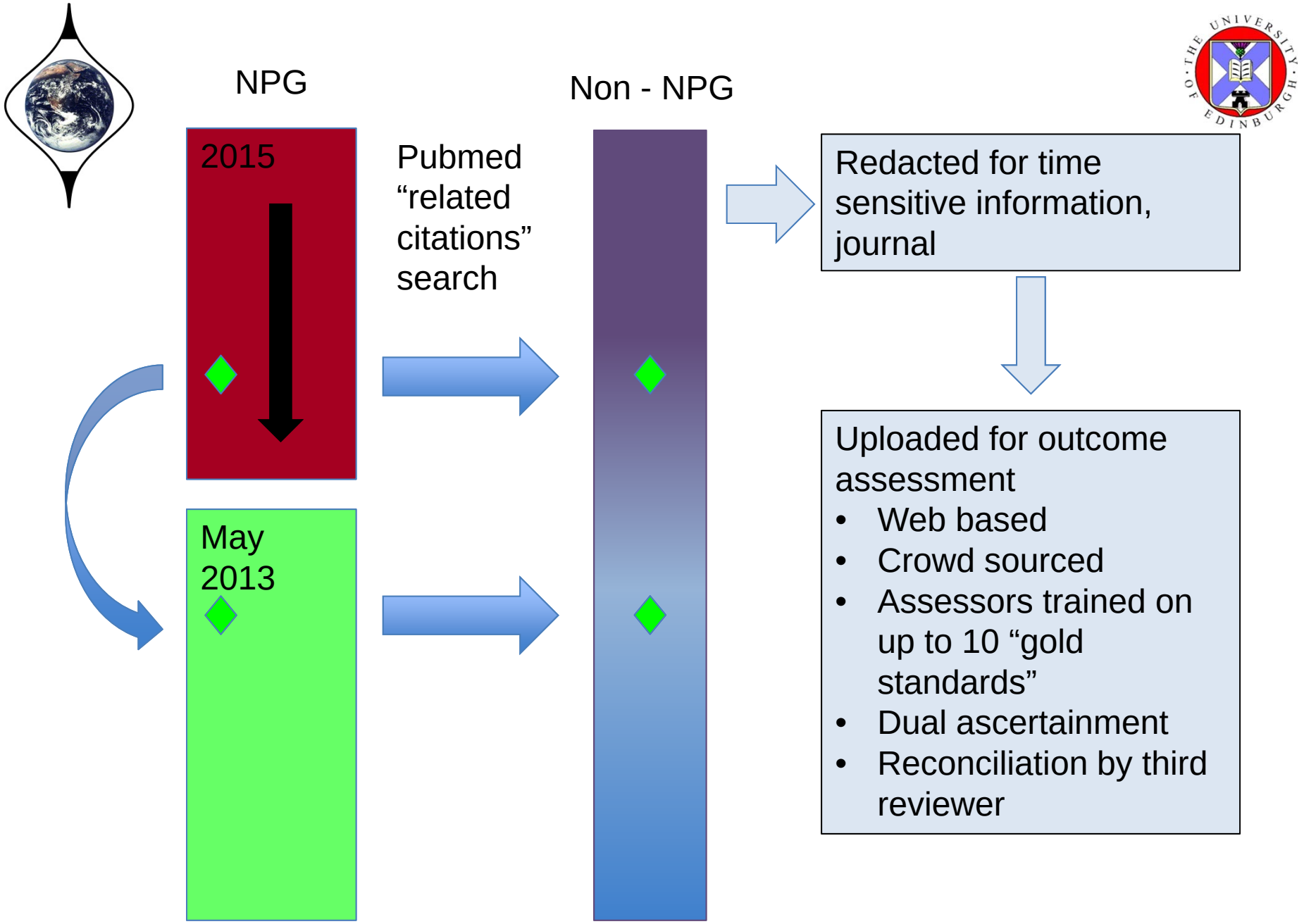
Data handling

- Rules for stopping data collection should be defined in advance.
- Criteria for inclusion and exclusion of data should be established prospectively.
- How outliers will be defined and handled should be decided when the experiment is being designed, and any data removed before analysis should be reported.
- The primary end point should be prospectively selected. If multiple end points are to be assessed, then appropriate statistical corrections should be applied.
- Investigators should report on data missing because of attrition or exclusion.
- Pseudo replicate issues need to be considered during study design and analysis.
- Investigators should report how often a particular experiment was performed and whether results were substantiated by repetition under a range of conditions.

NPQIP study

Objective: To determine whether a change in editorial policy, including the implementation of a checklist, has been associated with improved reporting of measures which might reduce the risk of bias
Design: Observational cohort study
Population Articles describing research in the life sciences published in Nature journals, submitted after May 1st 2013 and before November 1st 2014.
Intervention Mandatory completion of a checklist at the point of manuscript revision.
Comparators (1) Articles describing research in the life sciences published in Nature journals, submitted before May 2013; (2) Similar articles in other journals matched for date and topic.
Primary Outcome Change in proportion of Nature publications describing in vivo research published before and after May 2013 reporting the “Landis 4” items (randomisation, blinding, sample size calculation, exclusions).

Protocol: Cramond et al (2016) https://link.springer.com/article/10.1007/s11192-016-1964-8
Data Analysis Plan: Open Science Framework (June 2017) https://osf.io/mqet6/
Funding: Laura and John Arnold Foundation
Publication: http://www.biorxiv.org/content/early/2017/09/12/187245
Data: https://figshare.com/articles/NPQIP_final_analysis_set/5375275


NPG
Non - NPG
2015
Pubmed "related citations" search
May 2013
Redacted for time sensitive information, journal
Uploaded for outcome assessment
• Web based
• Crowd sourced
• Assessors trained on up to 10 "gold standards"
• Dual ascertainment
• Reconciliation by third reviewer

	NPG Publications (n=448)		Non NPG Publications (n=448)	
Publications	Before 01052013 (n=223)	After 01052013 (n=225)	Before 01052013 (n=202)	After 01052013 (n=246)
Initial screen	Excluded: 3	Excluded: 1	Excluded: 8	Excluded: 3
Available for analysis	Analysis (n=220)	Analysis (n=224)	Analysis (n=194)	Analysis (n=243)
Exclusions	Excluded: 1	Excluded: nil	Excluded: nil	Excluded: 1
Final analysis set	n=219	n=224	n=194	n=242
Types of experiment	In vivo 60 / 144 / 15 In vitro	In vivo 42 / 148 / 34 In vitro	In vivo 60 / 104 / 30 In vitro	In vivo 77 / 112 / 53 In vitro
Compliance with Landis 4 (in vivo)	0/204	31/190	1/164	1/189

CAMARADES: Bringing evidence to translational medicine

Compliance with “Landis 4”

NPG

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

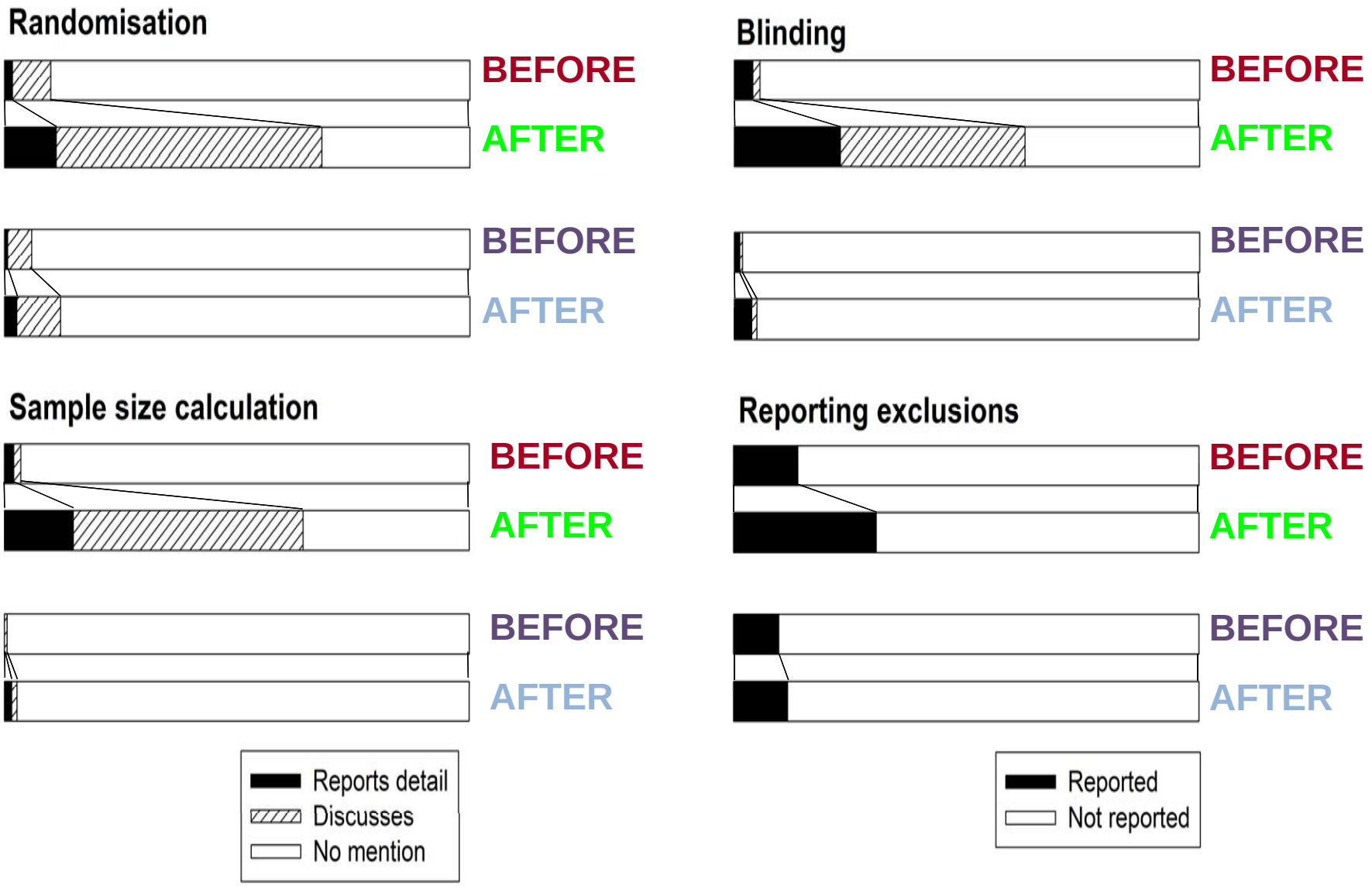
$p<10^{-8}$

Non NPG

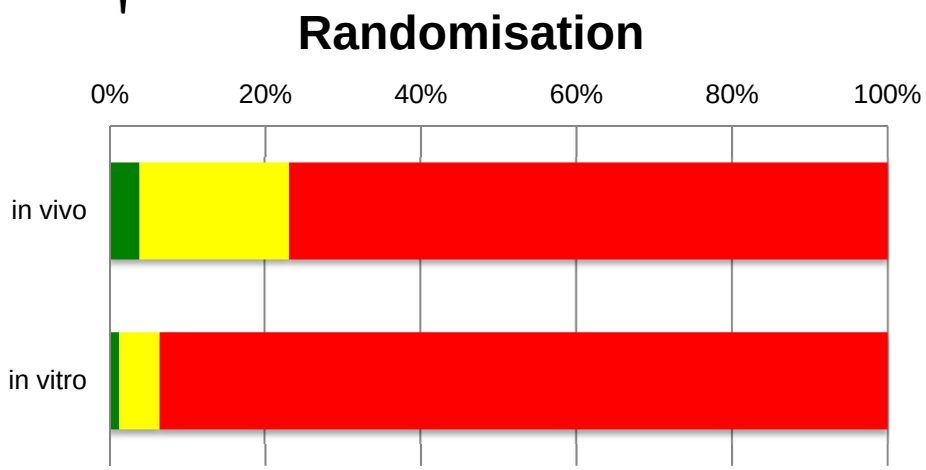
- Before: 1 of 164 (0.6%, 95% CI 0.1-4.2)
- After: 1 of 189 (0.5%, 95% CI 0.1-3.7)

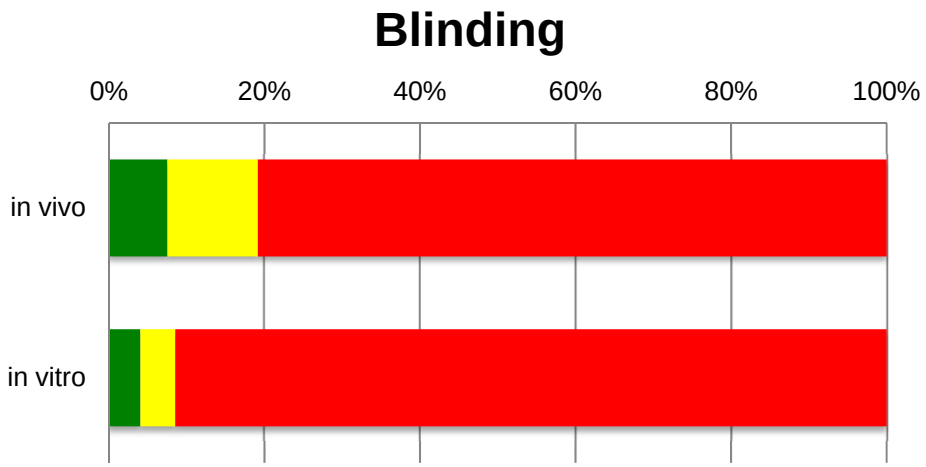
n.s.

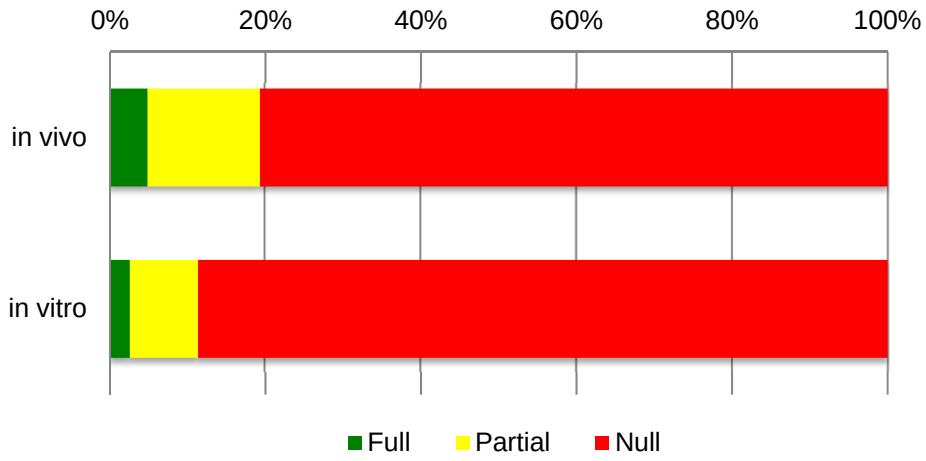
Individual risk of bias items

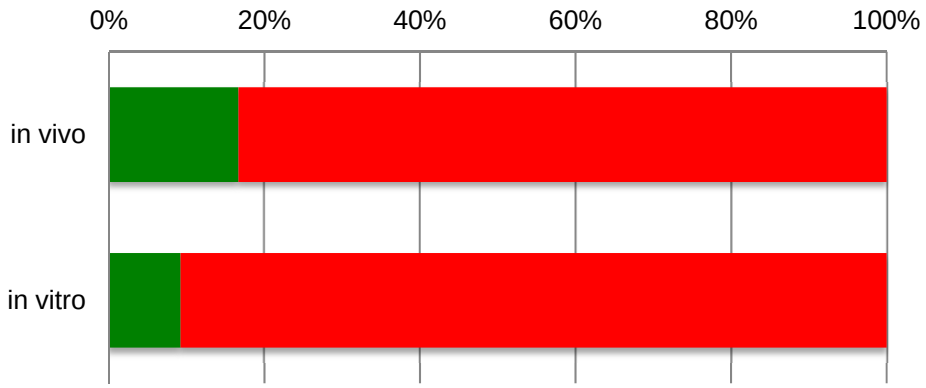


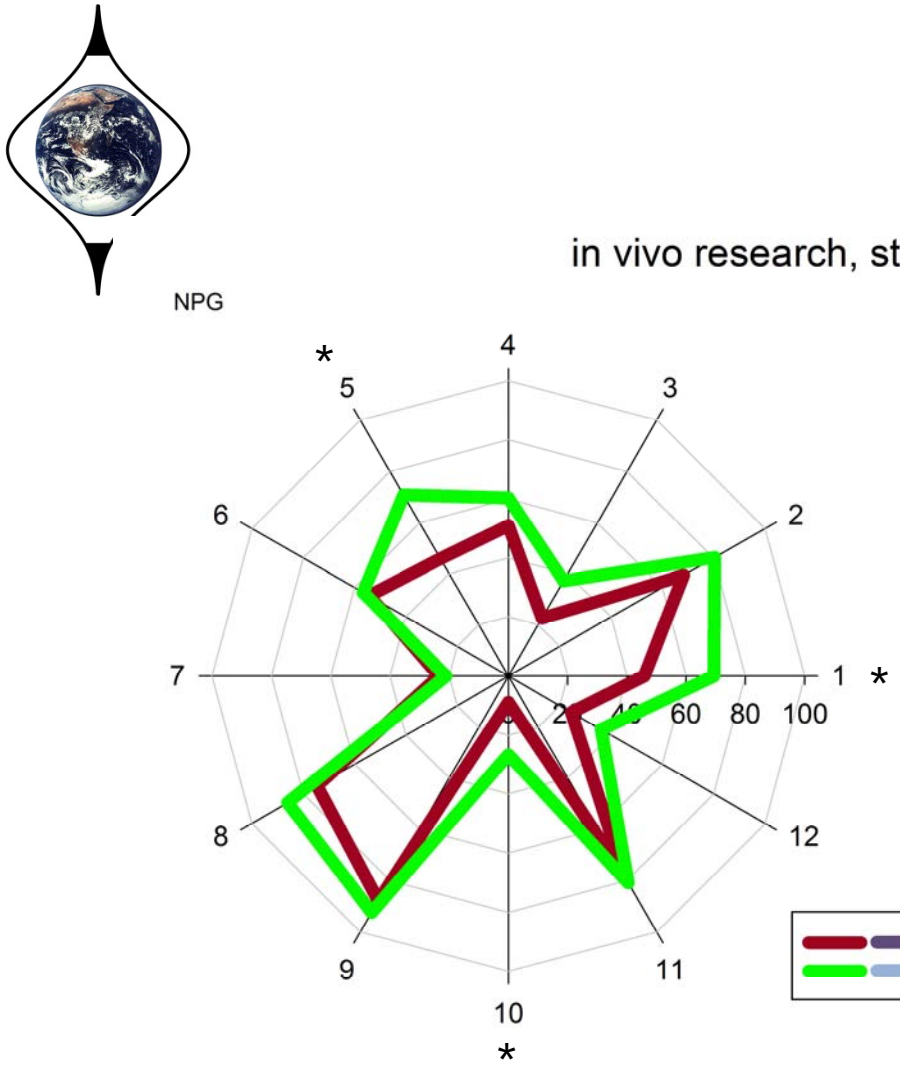
In vitro experiments











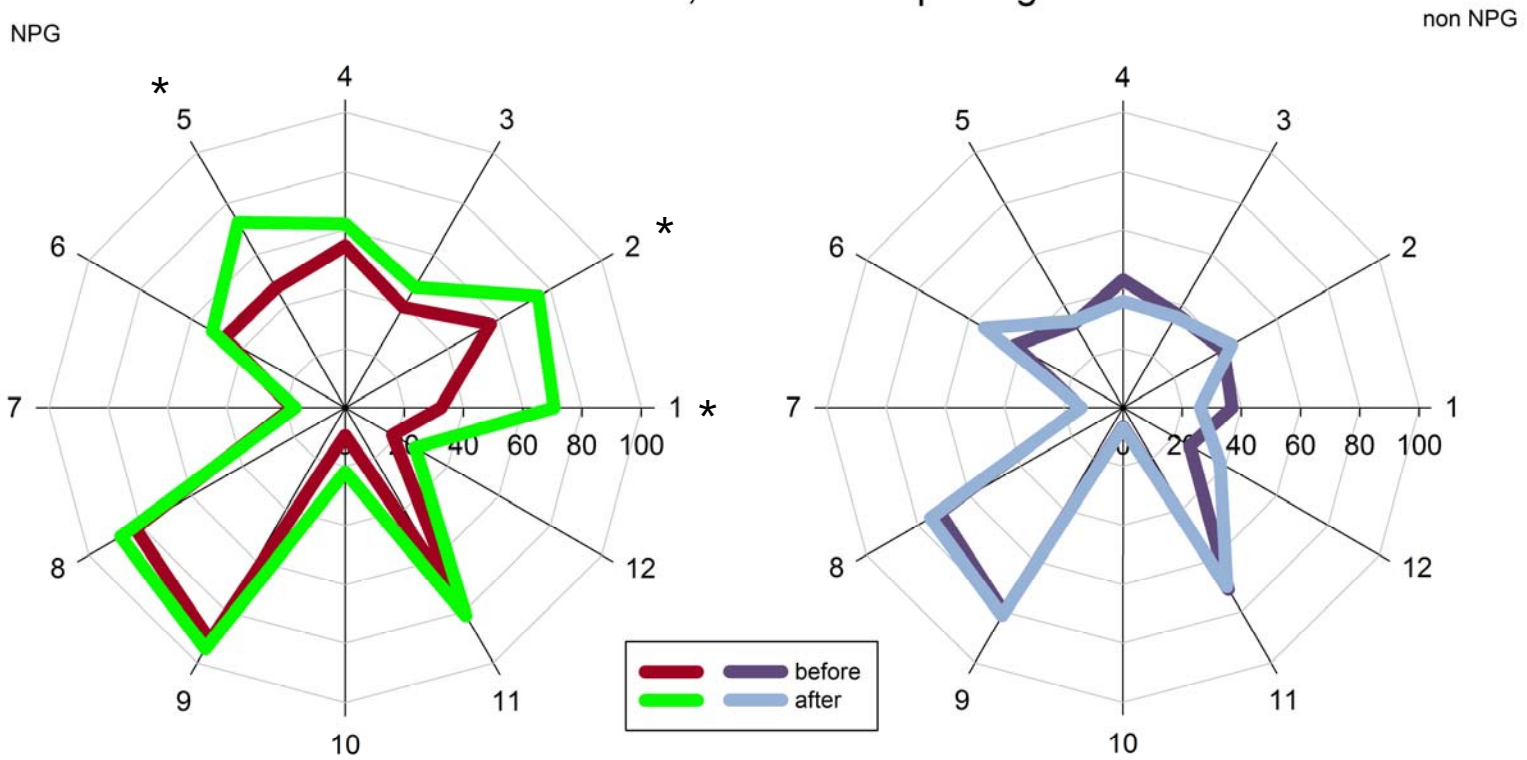
in vivo research, statistical reporting

NPG

non NPG

(1) Exact n (2) Technical or biological replicates (3) Number of times replicated (4) Test described if uncommon? (5) t-test defined as 1 or 2 sided?
(6) Correction for multiplicity (7) Reporting full statistics (8) Reporting of average (9) Definition of Error Bars (10) Testing of assumptions
(11) Reporting measures of variation (12) Variation < 2 fold

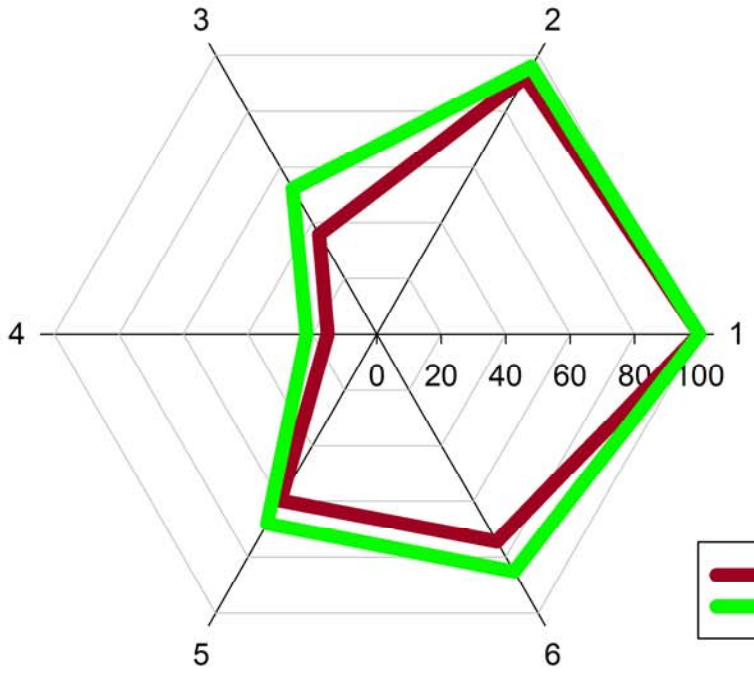
in vitro research, statistical reporting



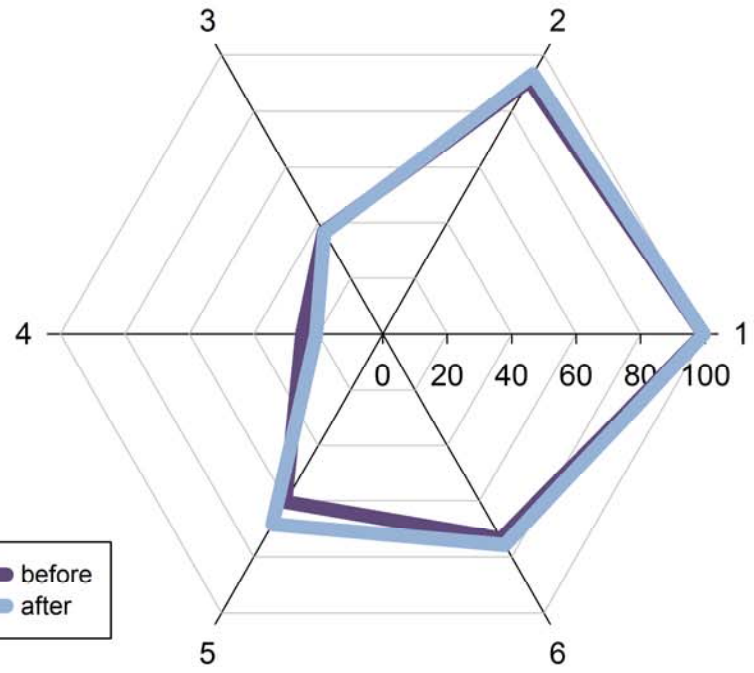
(1) Exact n (2) Technical or biological replicates (3) Number of times replicated (4) Test described if uncommon? (5) t-test defined as 1 or 2 sided?
(6) Correction for multiplicity (7) Reporting full statistics (8) Reporting of average (9) Definition of Error Bars (10) Testing of assumptions
(11) Reporting measures of variation (12) Variation < 2 fold

details of animal experiments

NPG



non NPG

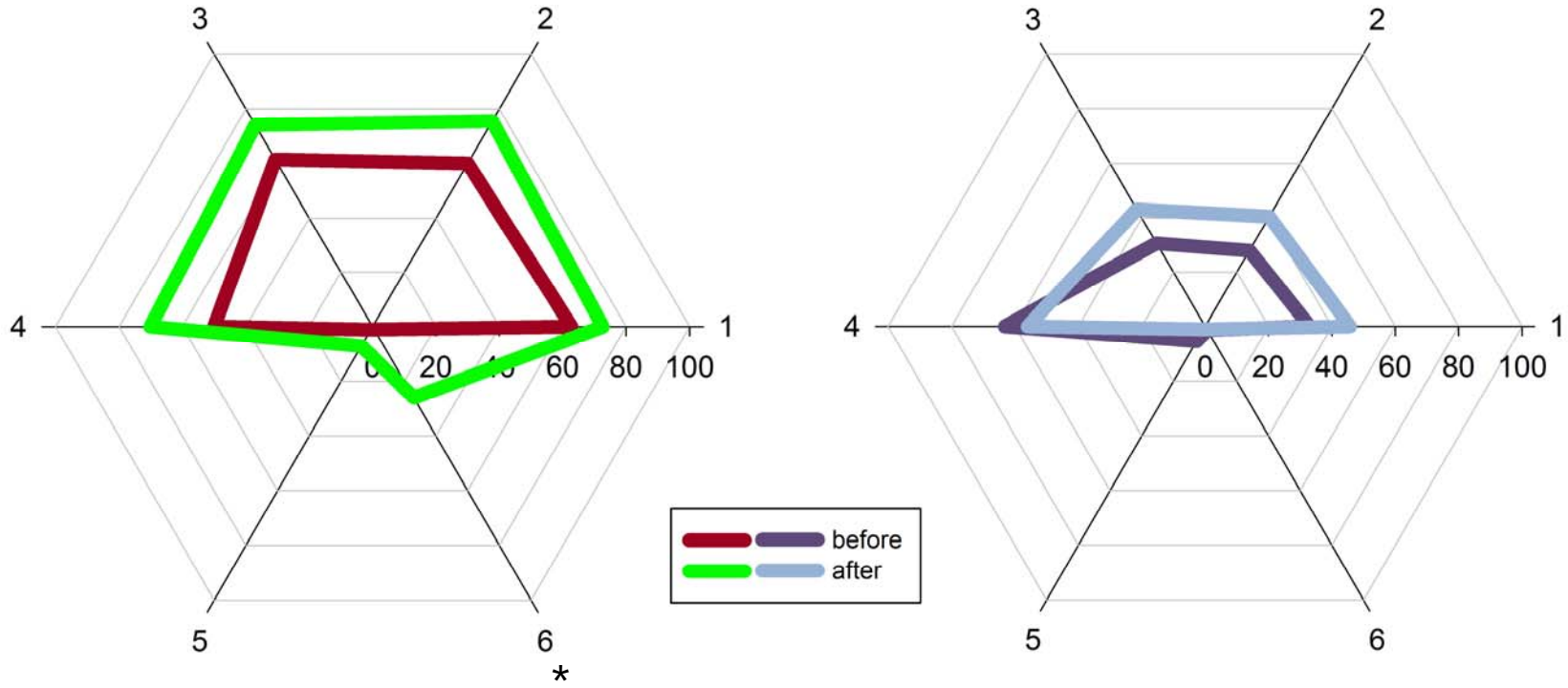


(1) Was the species reported? (2) Was the strain reported? (3) Was the sex reported? (4) Was exact age or weight given?
(5) Was ethical approval reported? (6) Ethical guidelines reported?

antibodies and cell culture details

NPG

non NPG



*

(1) reporting of antibodies used in In vivo experiments (2) reporting of antibodies used in In vitro experiments (3) Total antibody reporting where used (4) In vitro: cell line source (5) Recent authentication of cell line? (6) Recent mycoplasma testing?

The ARRIVE guidelines

OPEN ACCESS Freely available online

PLOS BIOLOGY

Perspective

Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research

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

NC 3Rs National Centre for the Replacement Refinement & Reduction of Animals in Research

- Journals
- Funders
- Universities
- Learned societies
- Organisations

ARRIVE

The ARRIVE Guidelines Checklist

Animal Research: Reporting In Vivo Experiments

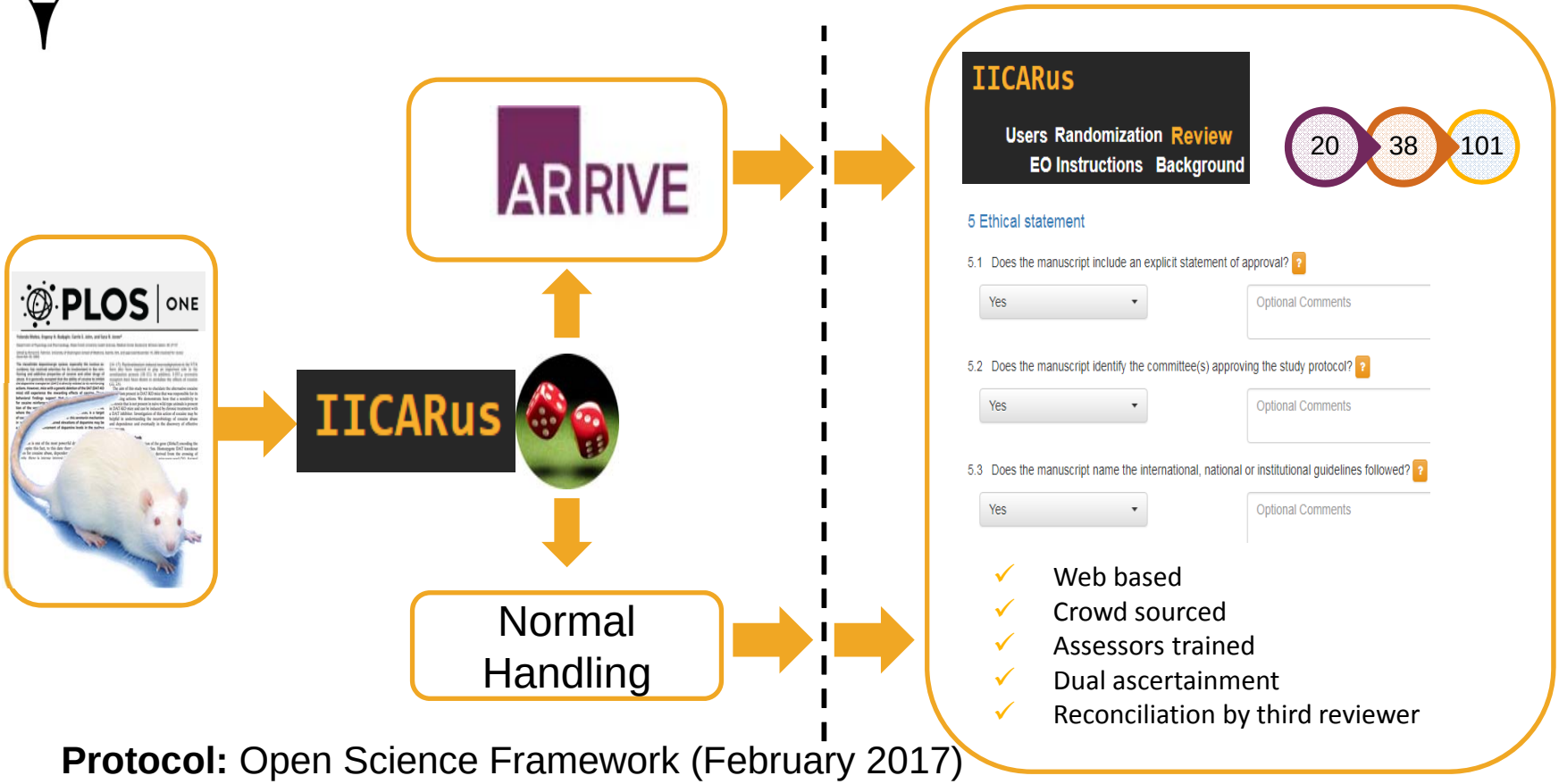
Carol Kilkenny[1], William J Browne[2], Innes C Cuthill[3], Michael Emerson[4] and Douglas G Altman[5]

[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.

	ITEM	RECOMMENDATION	Section/ Paragraph
Title	1	Provide as accurate and concise a description of the content of the article as possible.	
Abstract	2	Provide an accurate summary of the background, research objectives, including details of the species or strain of animal used, key methods, principal findings and conclusions of the study.	
INTRODUCTION			
Background	3	a. Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale. b. Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology.	
Objectives	4	Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.	
METHODS			
Ethical statement	5	Indicate the nature of the ethical review permissions, relevant licences (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.	
Study design	6	For each experiment, give brief details of the study design including: a. The number of experimental and control groups. b. Any steps taken to minimise the effects of subjective bias when allocating animals to treatment (e.g. randomisation procedure) and when assessing results (e.g. if done, describe who was blinded and when). c. The experimental unit (e.g. a single animal, group or cage of animals). A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out.	
Experimental procedures	7	For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. For example: a. How (e.g. drug formulation and dose, site and route of administration, anaesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s). b. When (e.g. time of day). c. Where (e.g. home cage, laboratory, water maze). d. Why (e.g. rationale for choice of specific anaesthetic, route of administration, drug dose used).	
Experimental animals	8	a. Provide details of the animals used, including species, strain, sex, developmental stage (e.g. mean or median age plus age range) and weight (e.g. mean or median weight plus weight range). b. Provide further relevant information such as the source of animals, international strain nomenclature, genetic modification status (e.g. knock-out or transgenic), genotype, health/immune status, drug or test naive, previous procedures, etc.	

The ARRIVE guidelines. Originally published in *PLoS Biology*, June 2010[1]

IICARUS: Study design



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

Results

Sample size calculation

Primary:

- 10% effect,
- α=0.05, β=80%
- n=100/group

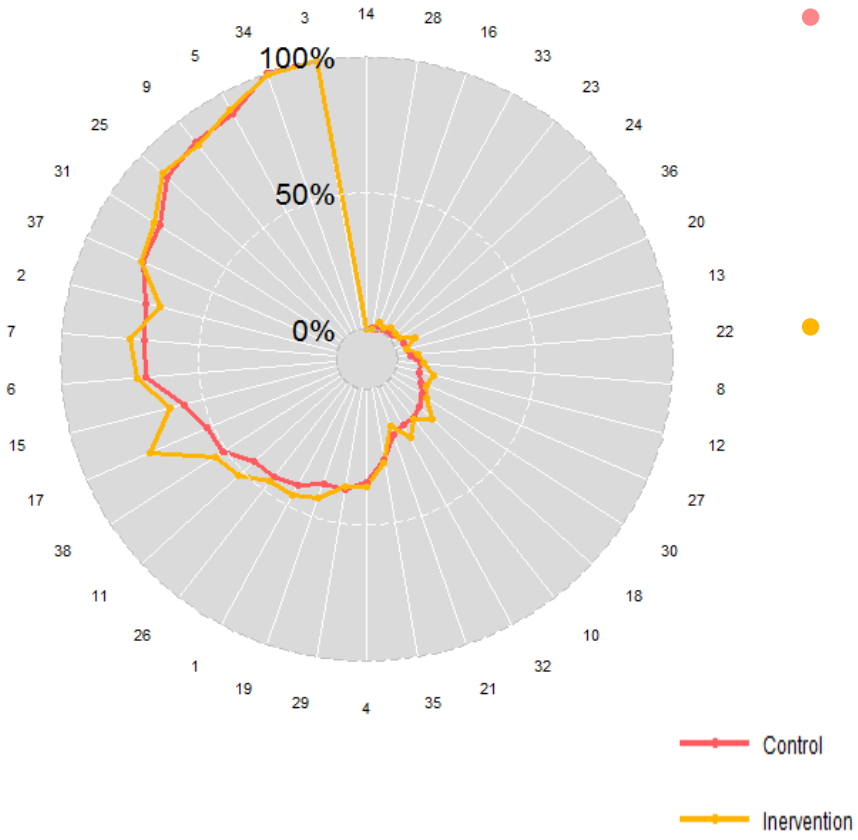
Sub-items:

- 20% effect
- α=0.0013, β=80%
- n=200/group

1689 submitted manuscripts randomised

Control	Intervention
844	845
Sent for review	Sent for review
652	647
Accepted	Accepted
322	340
Checklist completed	Checklist completed
13	301
Full compliance	Full compliance
0	0

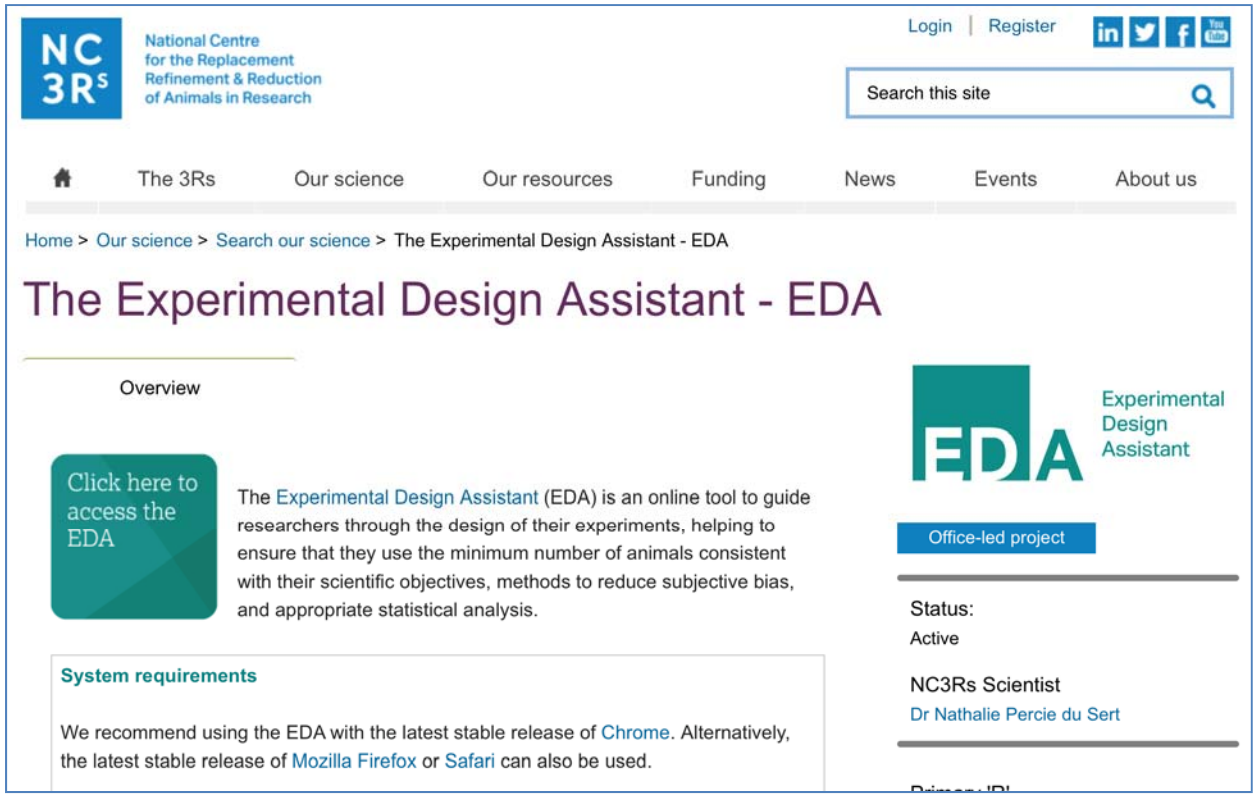
Primary outcome



- Control
  - Full compliance 0/322
  - Median compliance 36.8% (29.7-42.1) of relevant items
- Intervention
  - Full Compliance 0/340
  - Median compliance 39.5% (31.6-44.7) of relevant items

Resources

- NC3Rs Experimental Design Assistant

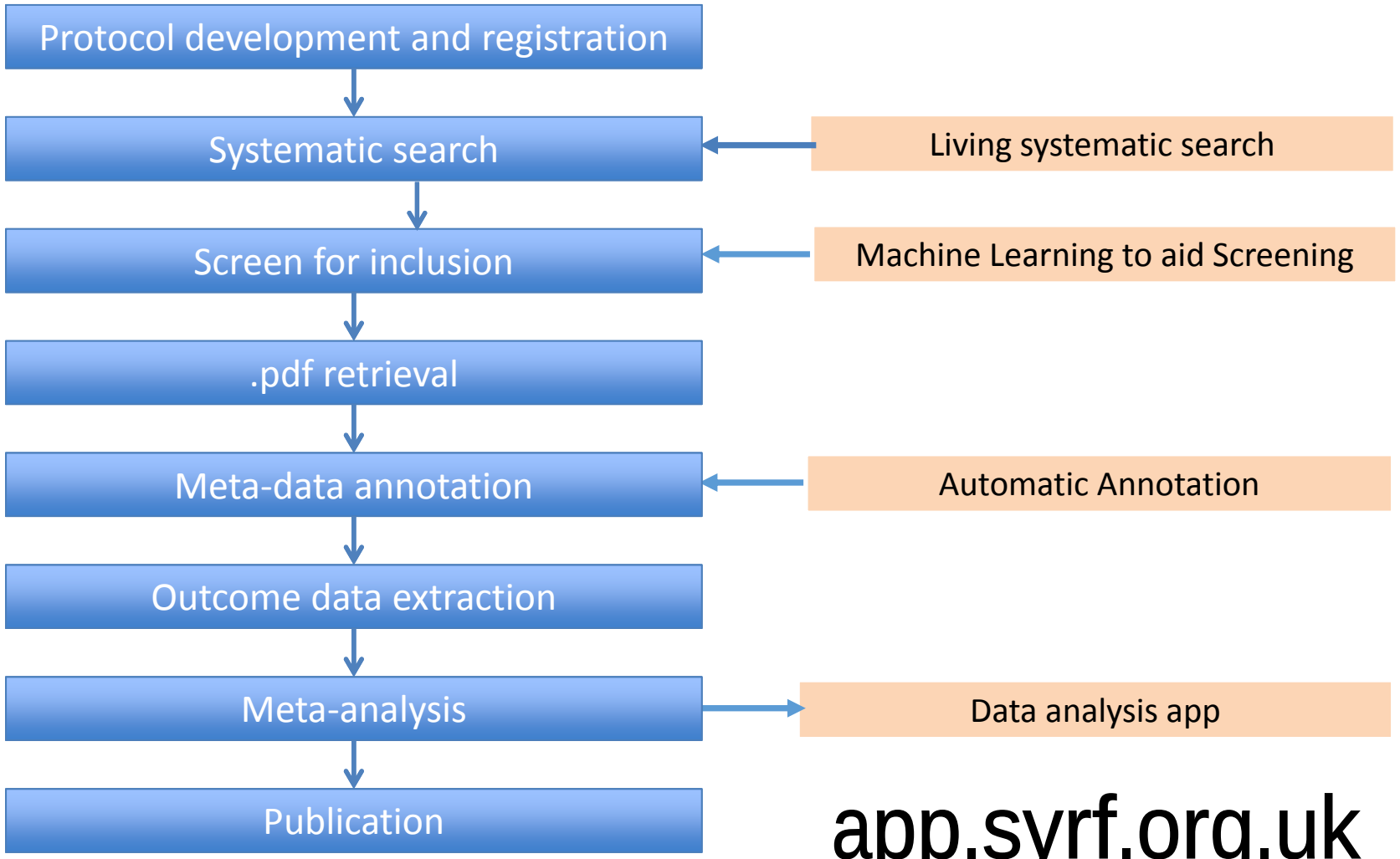


https://www.nc3rs.org.uk/experimental-design-assistant-eda

SYSTEMATIC Review Facility

SLIM Systematic Living Information Machine

THE UNIVERSITY OF EDINBURGH



app.syrf.org.uk

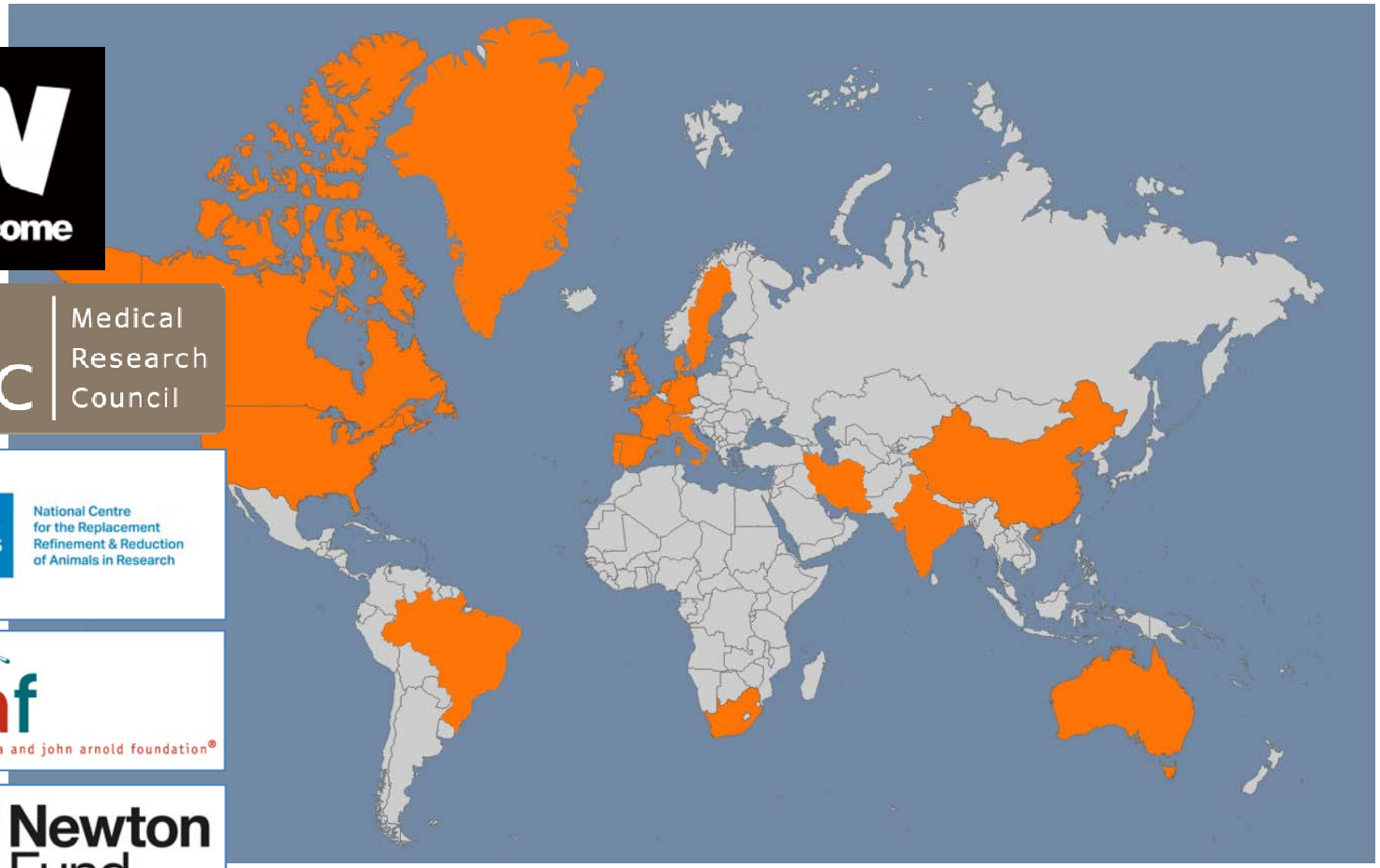
How scientists might approach their projects

- What parts of this are exploratory and what parts are hypothesis testing?
- What are the community standards in this field?
- Can I randomise, blind during the experiment, blind assessment of outcome?
- For tests of hypotheses, have I
  - Asserted my statistical analysis plan
  - Cemented my protocol where it can be checked
  - Described in advance my criteria for rejecting the null hypothesis
  - Got adequate statistical power to deliver a reasonable positive predictive value?

The future …

- Protocols/ registered reports as default
- Open access, open data as default
- Research Improvement Activities
  - Audit for improvement
    - Institutional culture
    - Institutional performance (and therefore audit tools)
  - Interventions at level of institution
    - Education
    - Policy changes
  - Interventions with other partners
    - Audit
    - Controlled trials

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



wellcome

MRC Medical Research Council

NC 3Rs National Centre for the Replacement Refinement & Reduction of Animals in Research

ljaf laura and john arnold foundation

Newton Fund

CAMARADES: Bringing evidence to translational medicine