CAWS Symposium 2017
Purdue University, West Lafayette, IN
May 18th, 2017

UNIVERSITÄT BERN

More than 3Rs

Improving the validity and reproducibility of animal research

Hanno Würbel

Division of Animal Welfare
Vetsuisse Faculty, University of Bern
Switzerland

Purdue University is an equal access/equal opportunity/affirmative action university
If you have trouble accessing this document because of a disability, please contact PVM IT at MIIT@purdue.edu

Harm-benefit analysis

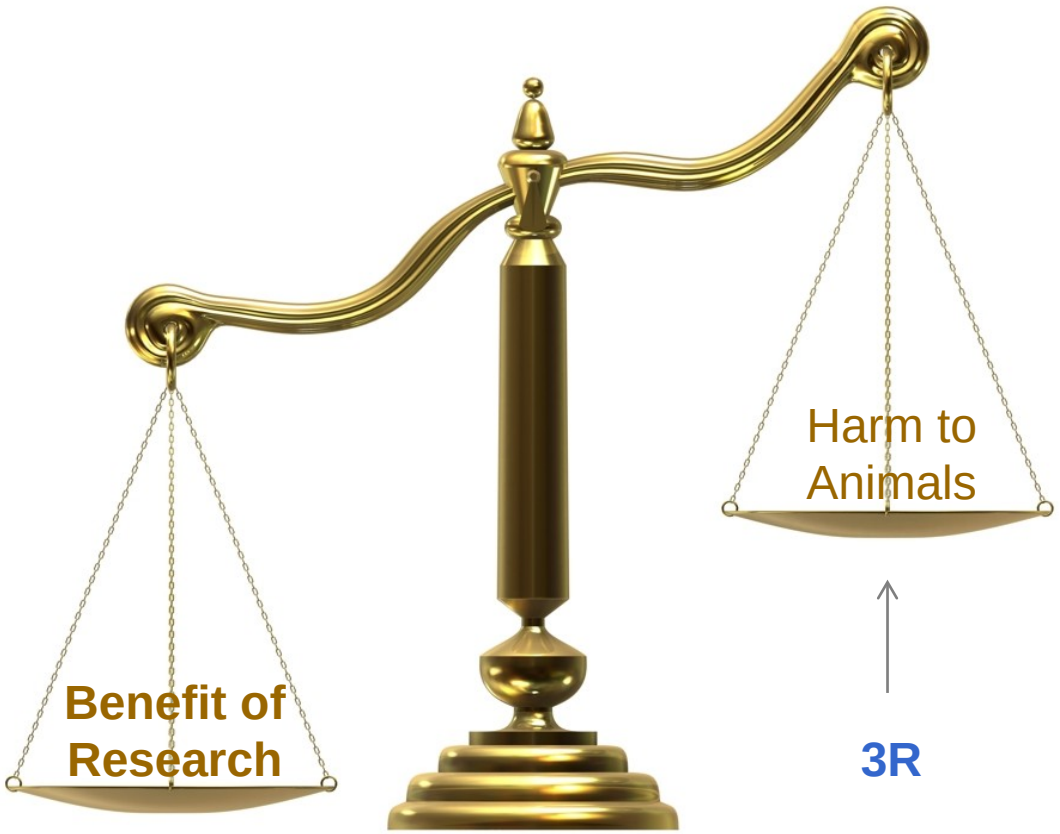
Principle of proportionality:

- Do the expected benefits outweigh the harms imposed on the animals?

Harm-benefit analysis

Preconditions:

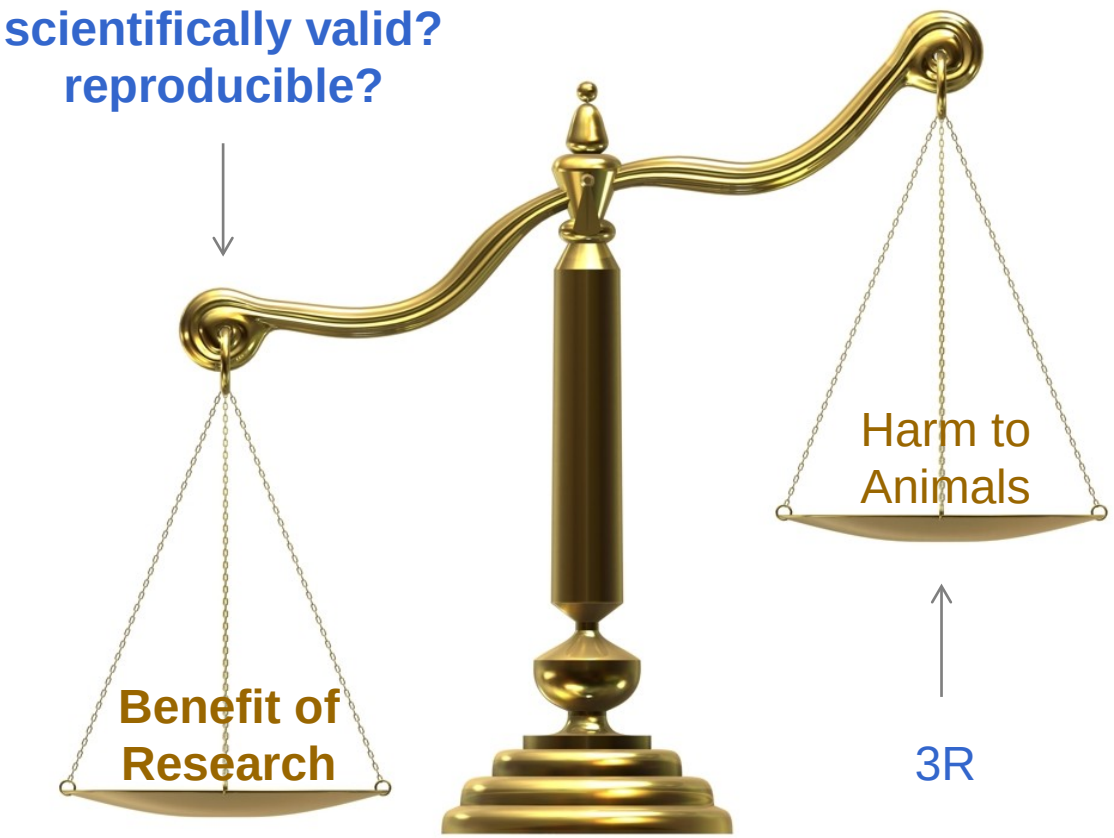
- Is the study **necessary** for achieving the expected benefit?



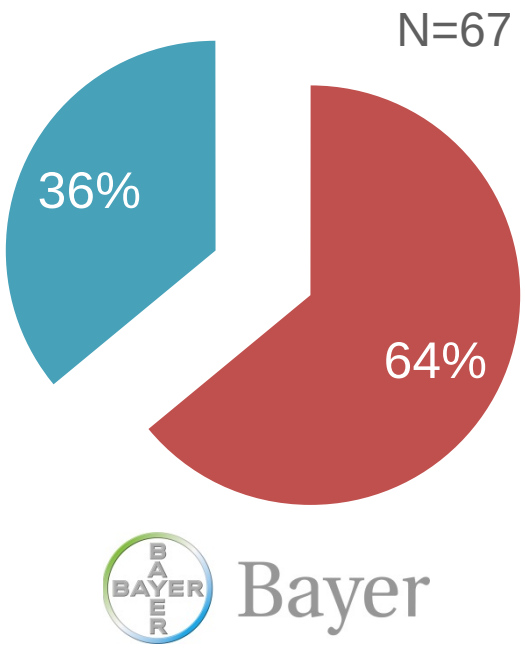
Harm-benefit analysis

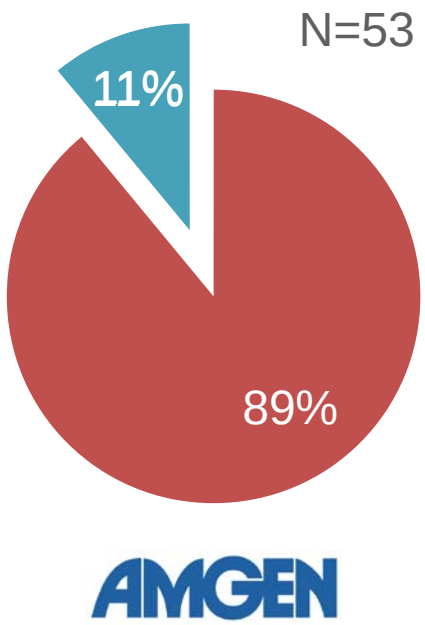
Preconditions:

- Is the study **necessary** for achieving the expected benefit?
- Is the study **suitable** for achieving the expected benefit?

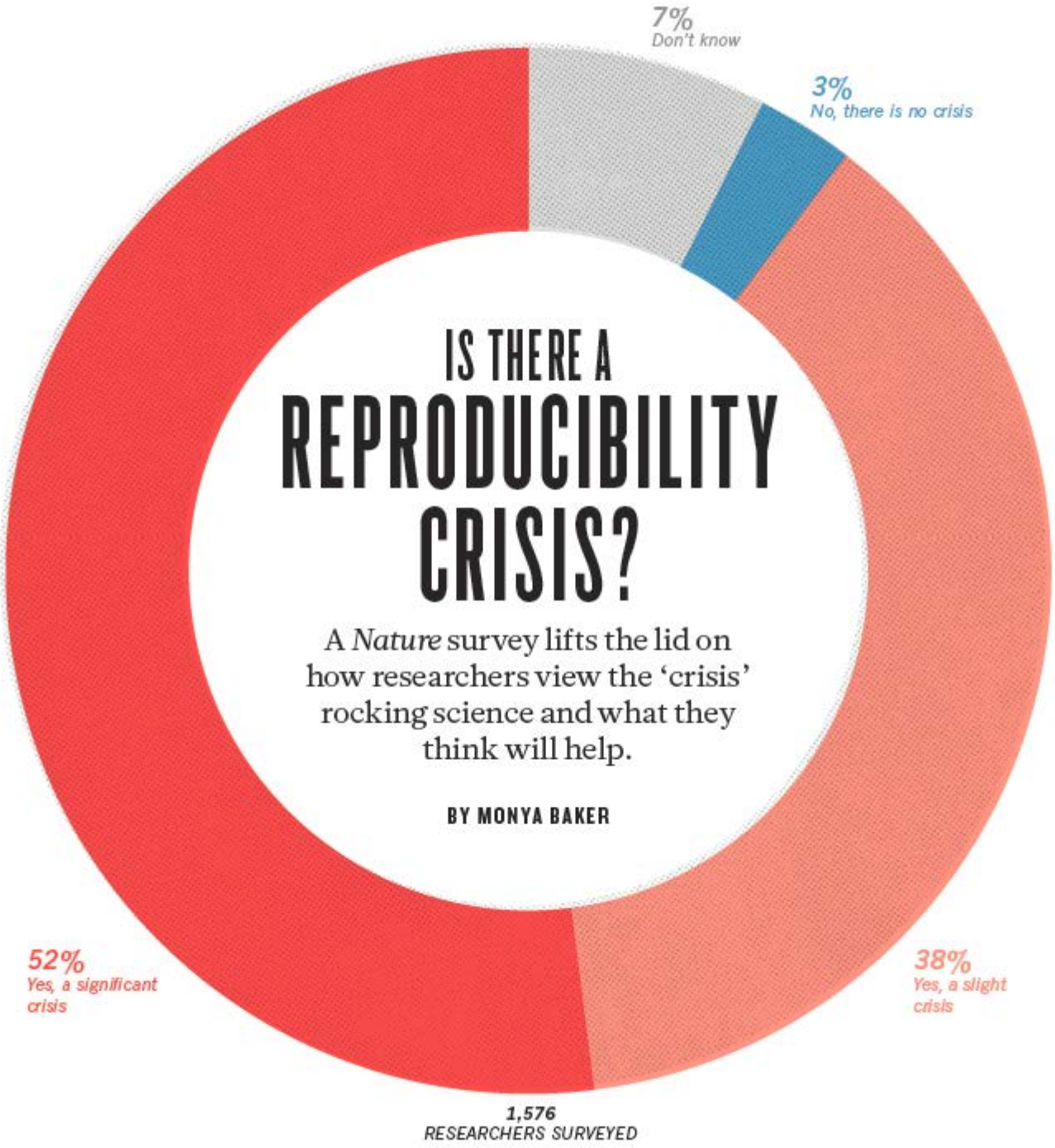


Spectacular failures to replicate pre-clinical studies





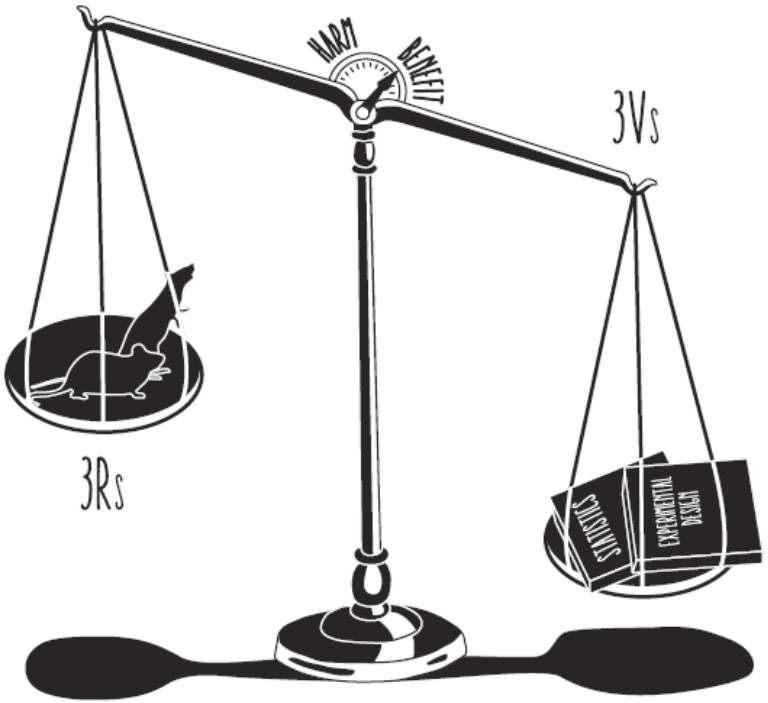
Prinz et al. 2011 **Nature Reviews Drug Discovery**
Begeley & Ellis 2012 **Nature**


7%
Don't know
3%
No, there is no crisis
IS THERE A
REPRODUCIBILITY
CRISIS?
A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help.
BY MONYA BAKER
52%
Yes, a significant crisis
38%
Yes, a slight crisis
1,576
RESEARCHERS SURVEYED

Potential problems with scientific validity

The 3Vs

- Poor construct validity → poorly validated animal models and outcomes
- Poor internal validity → poor experimental design and conduct
- Poor external validity → poor representation of target population



Würbel 2017 **Lab Animal**

Potential problems with scientific validity

The 3Vs

- Poor construct validity → poorly validated animal models and outcomes
- Poor internal validity → poor experimental design and conduct
- Poor external validity → poor representation of target population

Focus on Reproducibility

REVIEW

Introducing Therioepistemology: the study of how knowledge is gained from animal research

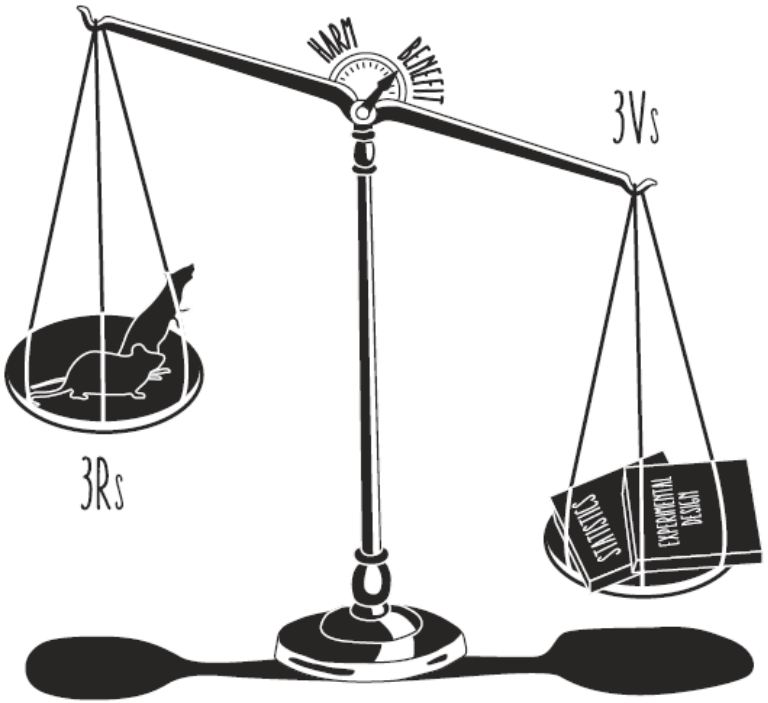
Joseph P Garner[1,2], Brianna N Gaskill[3], Elin M Weber[1], Jamie Ahloy-Dallaire[1] & Kathleen R Pritchett-Corning[4,5]

LabAnimal Volume 46, No. 4 | APRIL 2017 103-113

Potential problems with scientific validity

The 3Vs

- Poor construct validity → poorly validated animal models and outcomes
- Poor internal validity → poor experimental design and conduct
- Poor external validity → poor representation of target population



Würbel 2017 **Lab Animal**

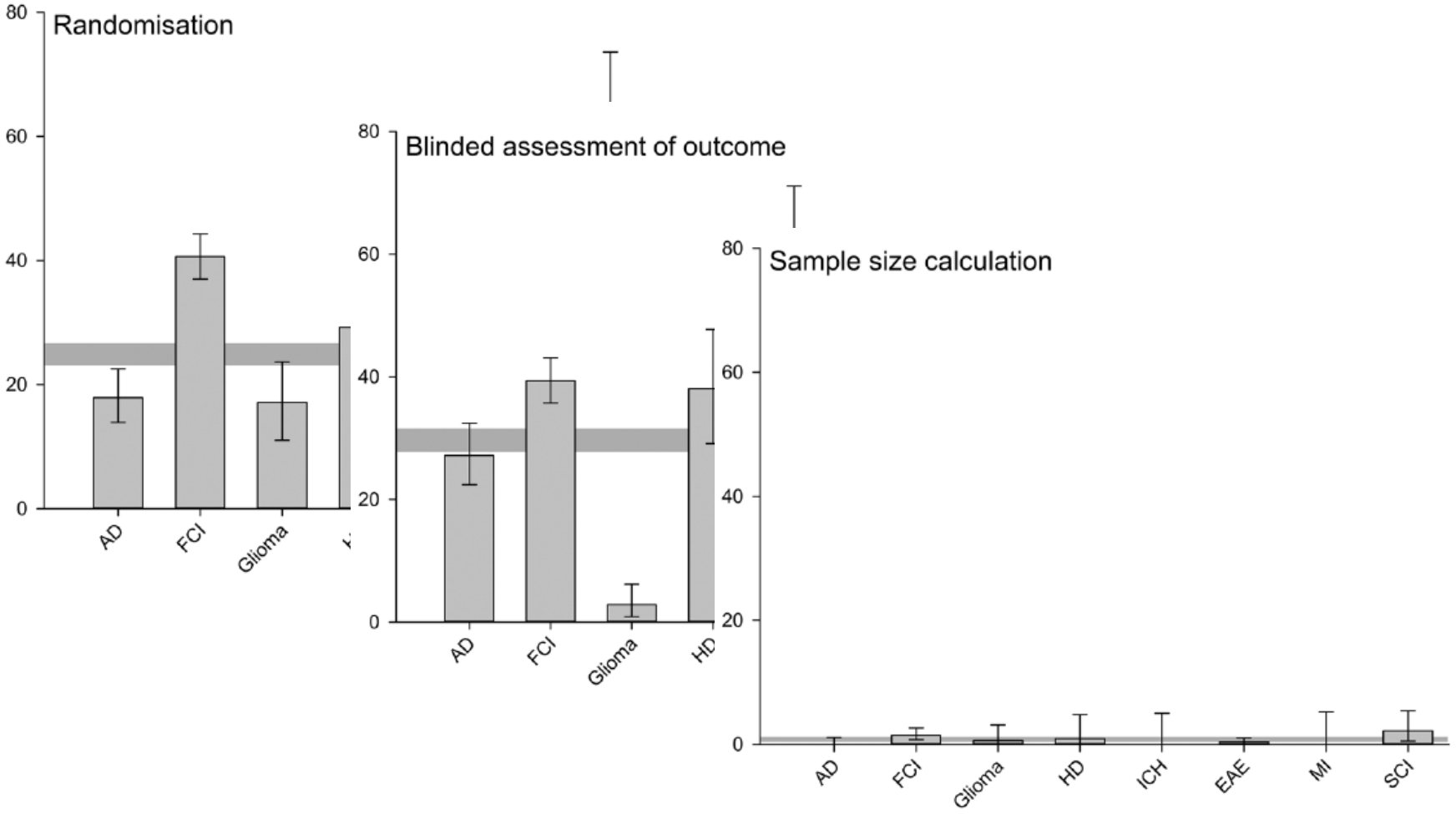
Internal validity

Measures to avoid risks of bias

- Randomisation → selection bias
- Blinding → detection bias
- Sample size calculation → statistical power
- Inclusion and exclusion criteria → attrition bias
- Definition of primary outcome variable → selective reporting
- Statistical analysis plan → analytical flexibility
- ...

→ **The commands of good research practice**

Poor reporting of risks of bias in animal research



Macleod et al. 2016 **PLoS Biol**

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]

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

In the last decade the number of bioscience journals has increased enormously, with many filling specialised niches reflecting new disciplines and technologies. The emergence of open-access journals has revolutionised the publication process, maximising the availability of research data. Nevertheless, a wealth of evidence shows that across many areas, the reporting of biomedical research is often inadequate, leading to the view that even if the science is

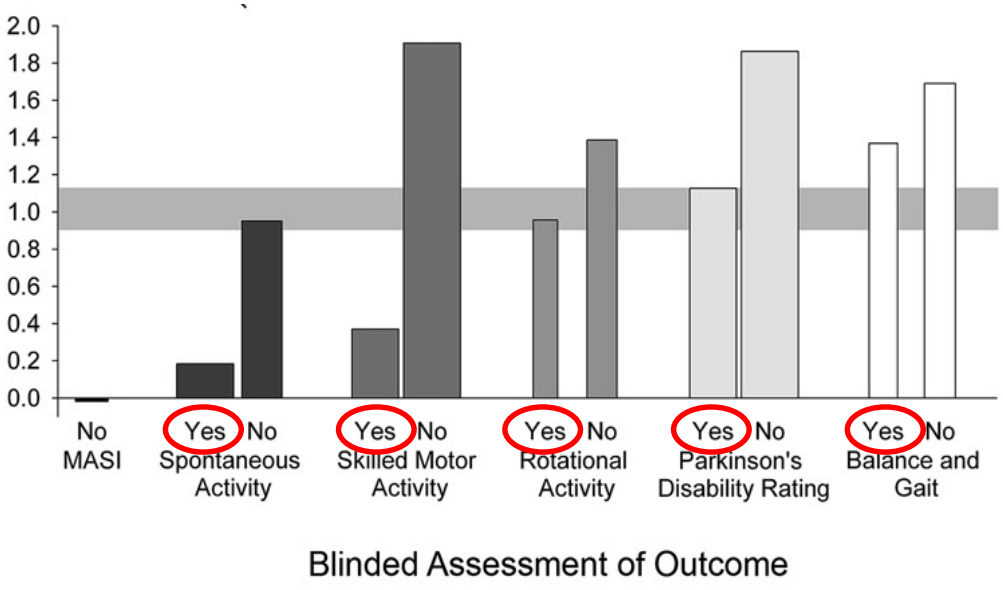
animals used (i.e., species/strain, sex, and age/weight). Most of the papers surveyed did not report using randomisation (87%) or blinding (86%) to reduce bias in animal selection and outcome assessment. Only 70% of the publications that used statistical methods fully described them and presented the results with a measure of precision or variability [5]. These findings are a cause for concern and are consistent with reviews of many research areas,

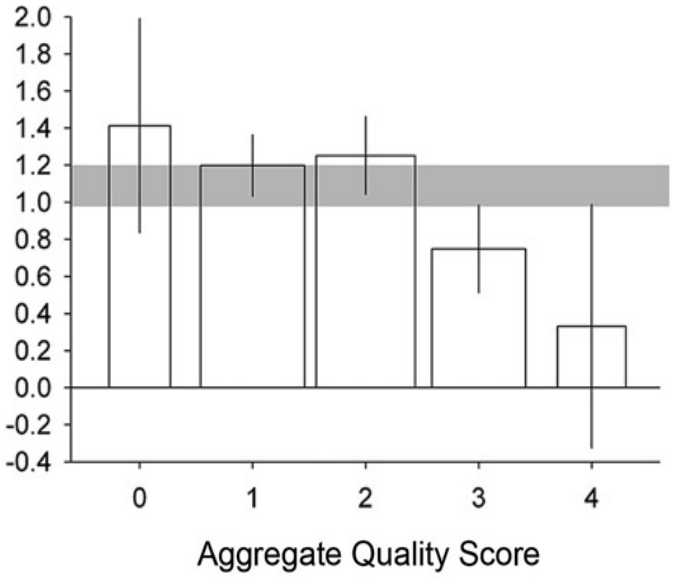
the study and the reliability and validity of the findings. There should also be enough information to allow the experiment to be repeated [23]. The problem therefore is how to ensure that all relevant information is included in research publications.

Using Reporting Guidelines Measurably Improves the Quality of Reporting

Poor reporting is associated with overstated effects

Efficacy of Dopamine agonists in animal models of Parkinson‘s Disease



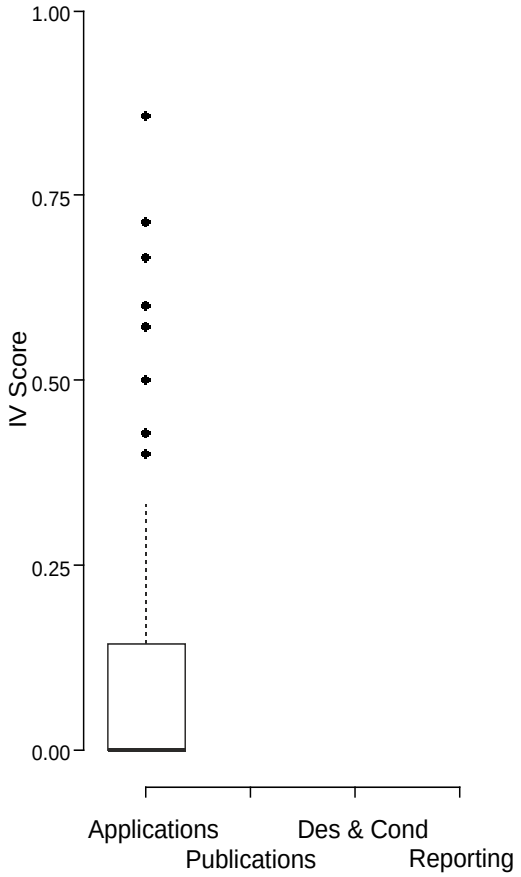


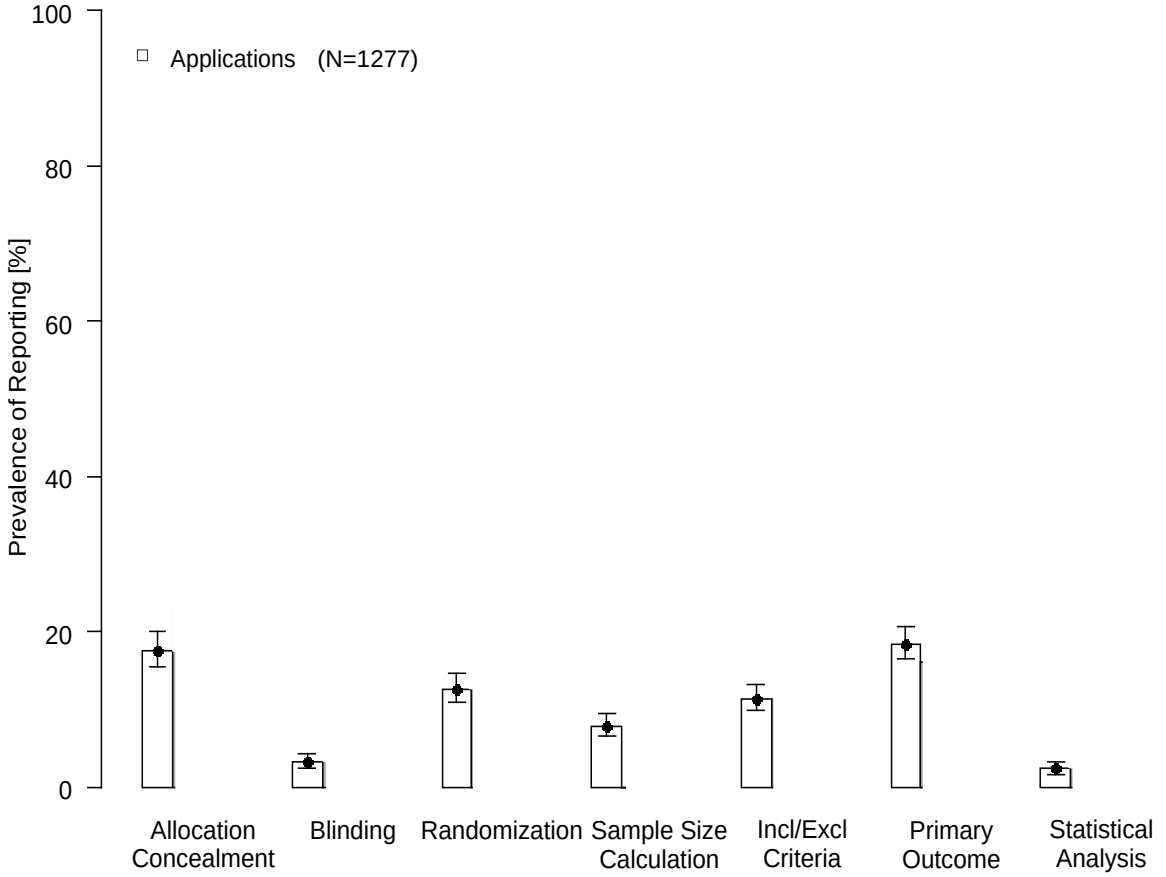
→ Studies reporting blinded outcome assessment found smaller treatment effects

Rooke et al. 2011 **Parkinsonism & Related Disorders**

Reporting in applications for animal experiments in CH

Internal validity score and measures to avoid risks of bias

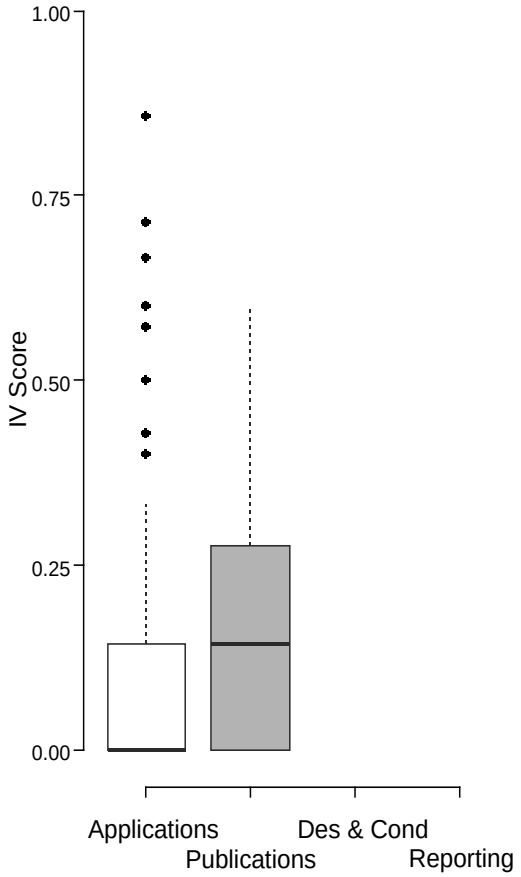


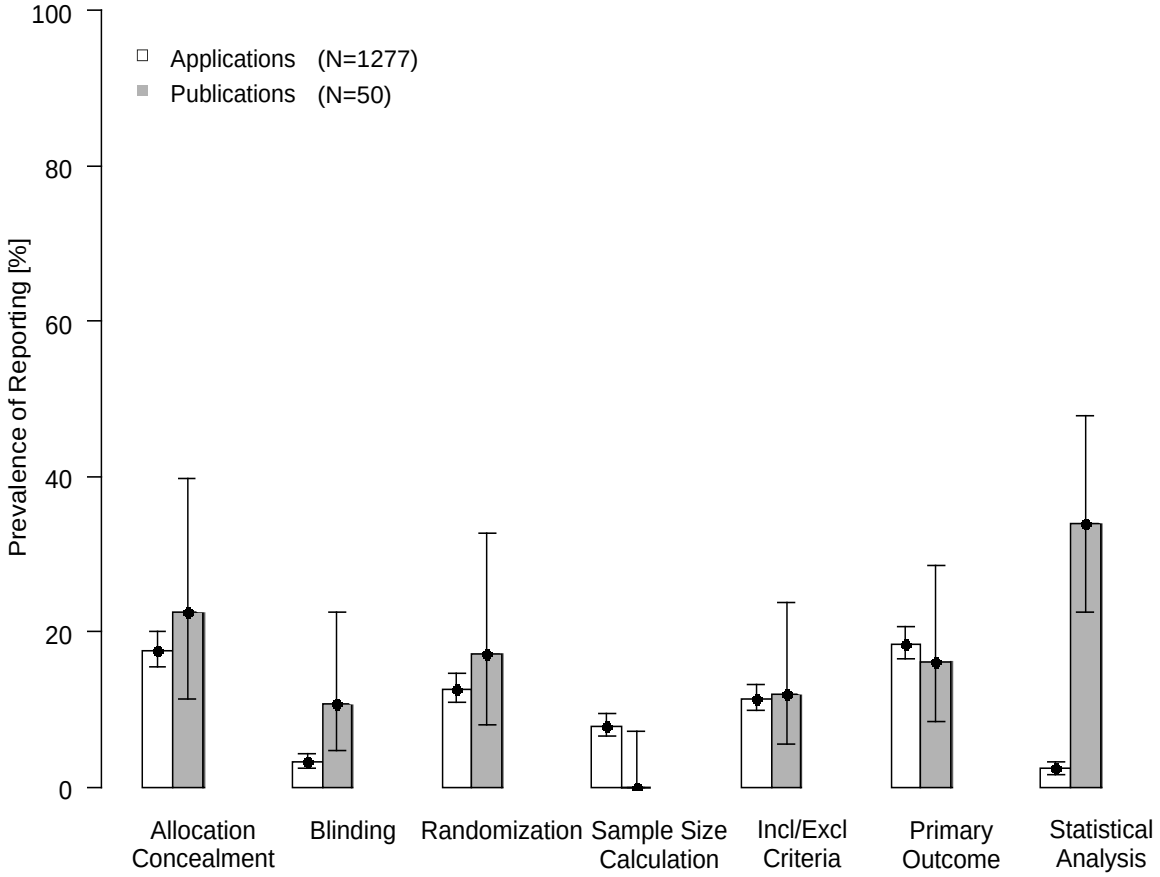


Vogt et al. 2016 **PLoS Biol**

Reporting in publications of animal experiments in CH

Internal validity score and measures to avoid risks of bias





Vogt et al. 2016 **PLoS Biol**

Conclusions

- Results confirm poor reporting of measures against risks of bias
- Authorities approving animal experiments and editors publishing papers lack information to assess risks of bias

Vogt et al. 2016 **PLoS Biol**

Use and reporting of measures against bias in CH

Online Survey

- Total sample (project leaders, experimenters): N=1891
- Return rate: 28% (N=530)
- Fully completed questionnaires: N=302 (16% of total sample)

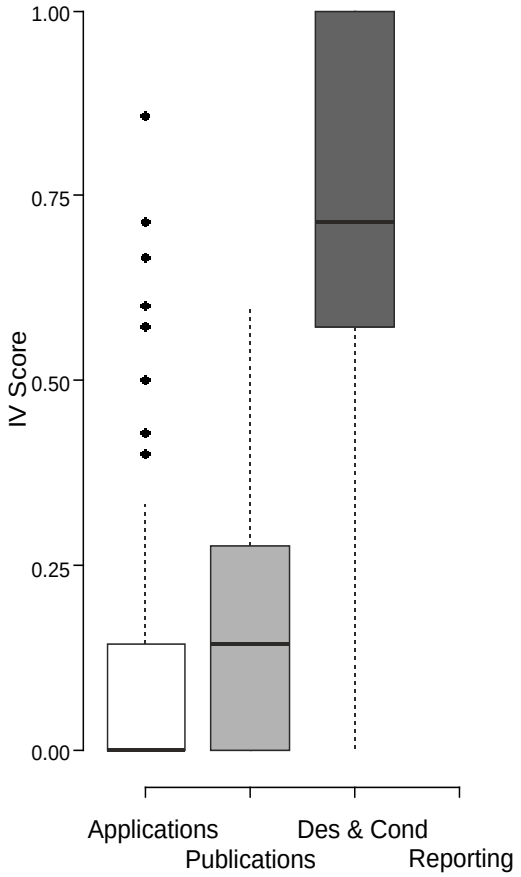
→ "*what measures against bias do you generally use in your research*?"

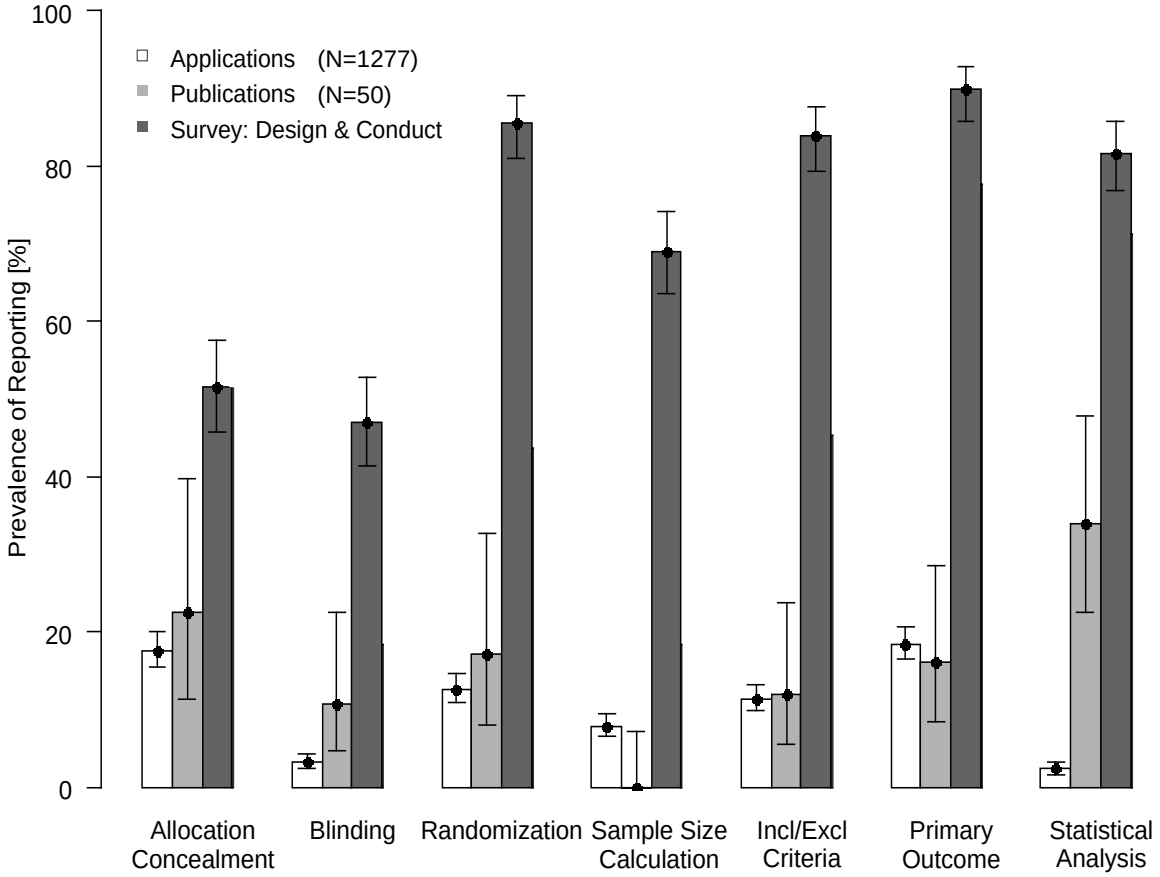
→ "*which of these measures have you reported in your latest publication*?"

Reichlin, Vogt & Würbel 2016 **PLoS ONE**

Use of measures against risks of bias in CH

Internal validity score and measures to avoid risks of bias

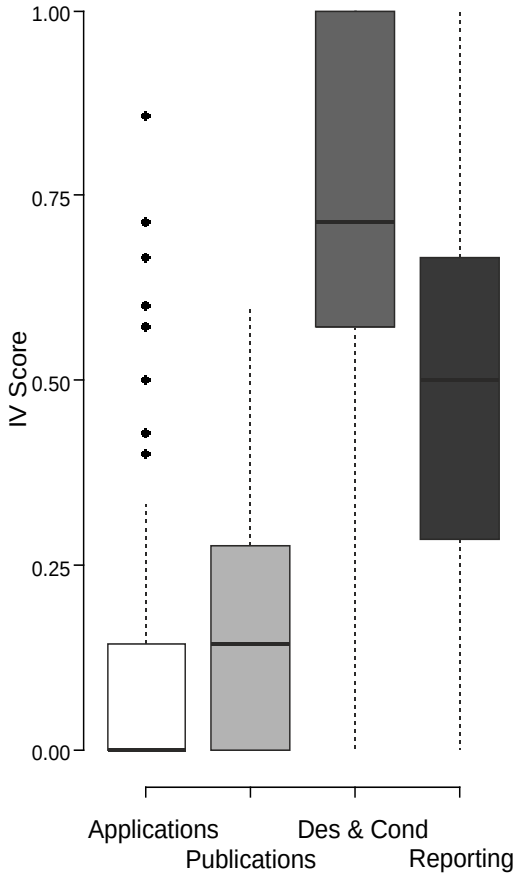


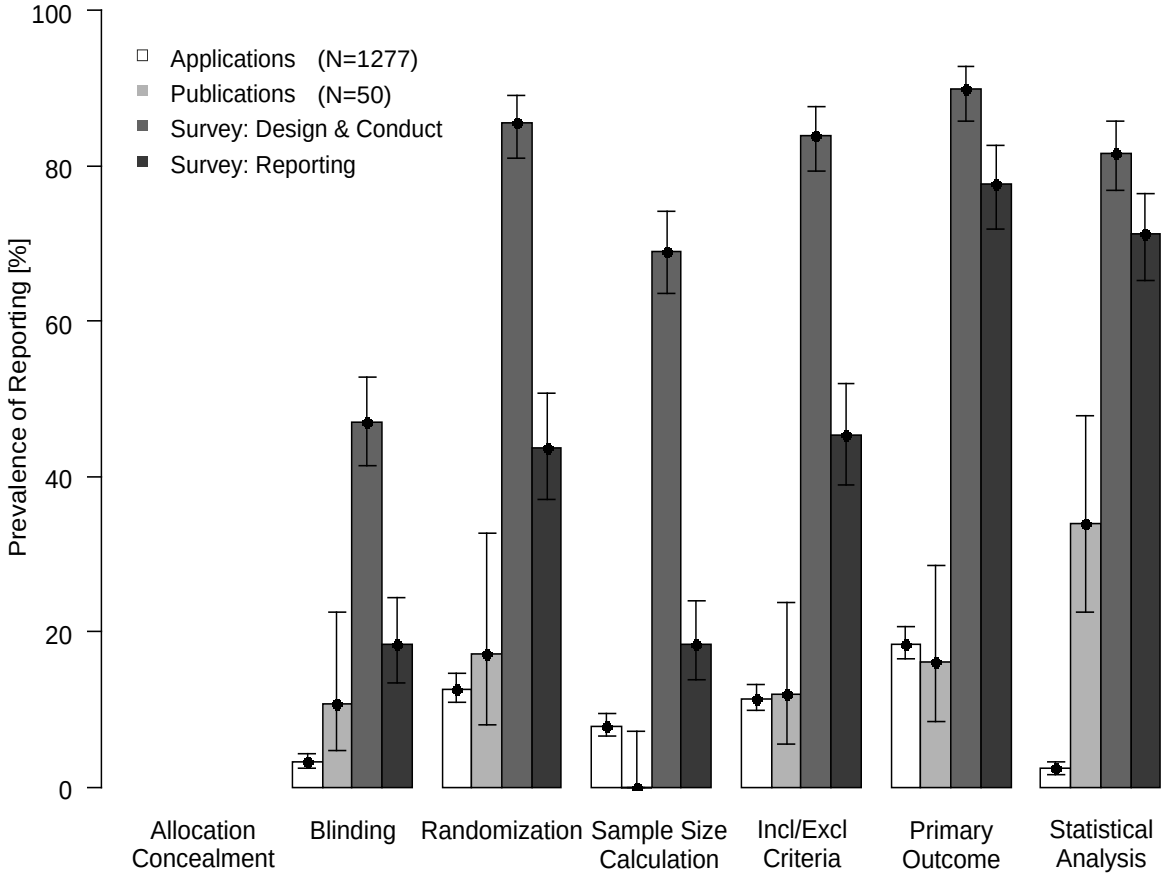


Reichlin, Vogt & Würbel 2016 **PLoS ONE**

Reporting of measures against risks of bias in CH

Internal validity score and measures to avoid risks of bias

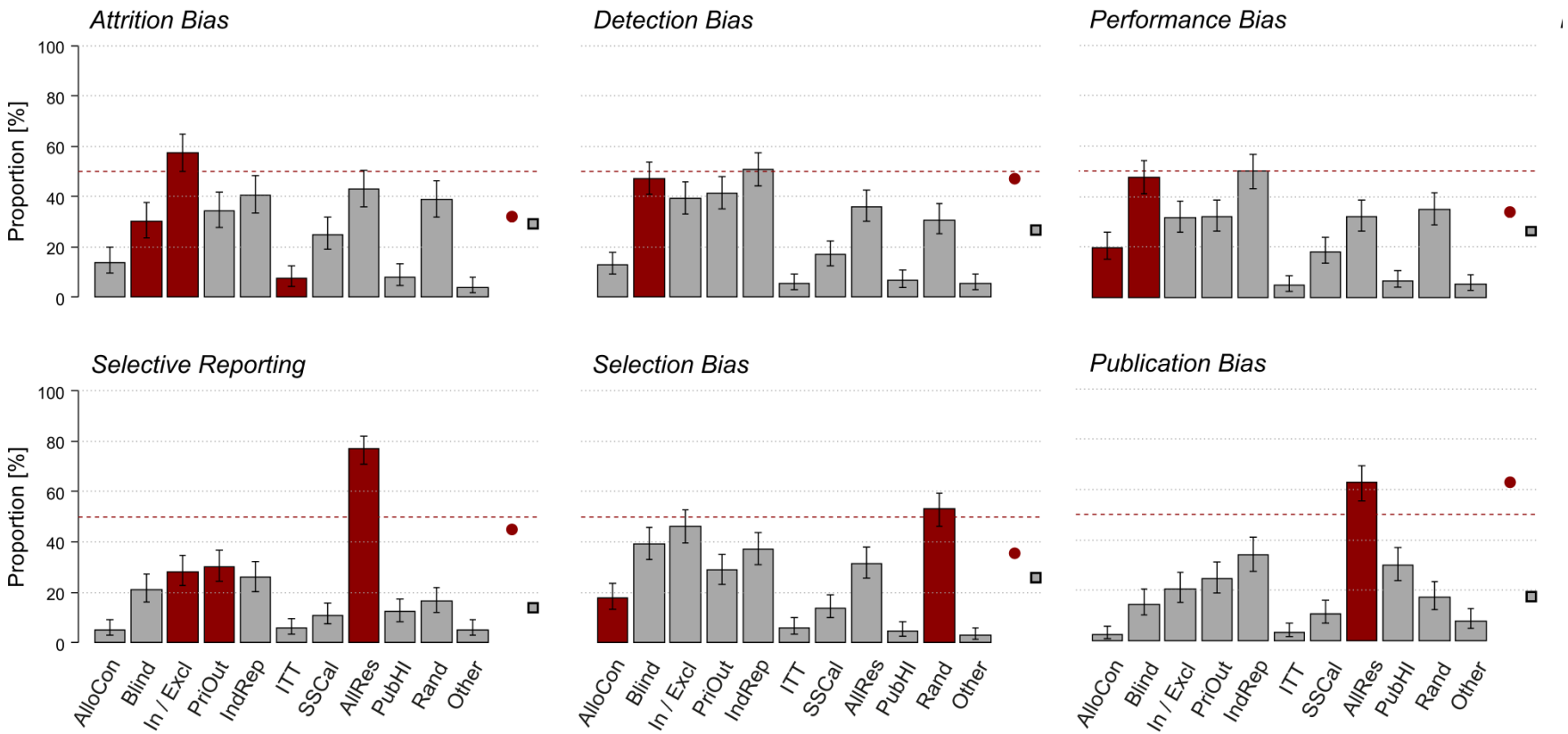




Reichlin, Vogt & Würbel 2016 **PLoS ONE**

Use of measures against risks of bias in CH

Participants' knowledge of measures to avoid risks of bias



Reichlin et al. 2016 **PLoS ONE**

Conclusions

B

- Results confirm poor reporting of measures against risks of bias
- Authorities approving animal experiments and editors publishing papers lack information to assess risks of bias

Vogt et al. 2016 **PLoS Biol**

- Researchers overestimate scientific rigor of their own research
- Researchers lack problem awareness and knowledge to avoid risks of bias

Reichlin et al. 2016 **PLoS ONE**

Problems associated with lack of scientific rigor

- Inferential statistics applied to hypothesis-free fishing expeditions
- Searching for p-values < 0.05 (P-hacking)
- Hypothesizing After the Results are Known (HARKing)
- Selective reporting of outcomes
- Publishing “positive” results only (Publication bias)

Problems associated with lack of scientific rigor

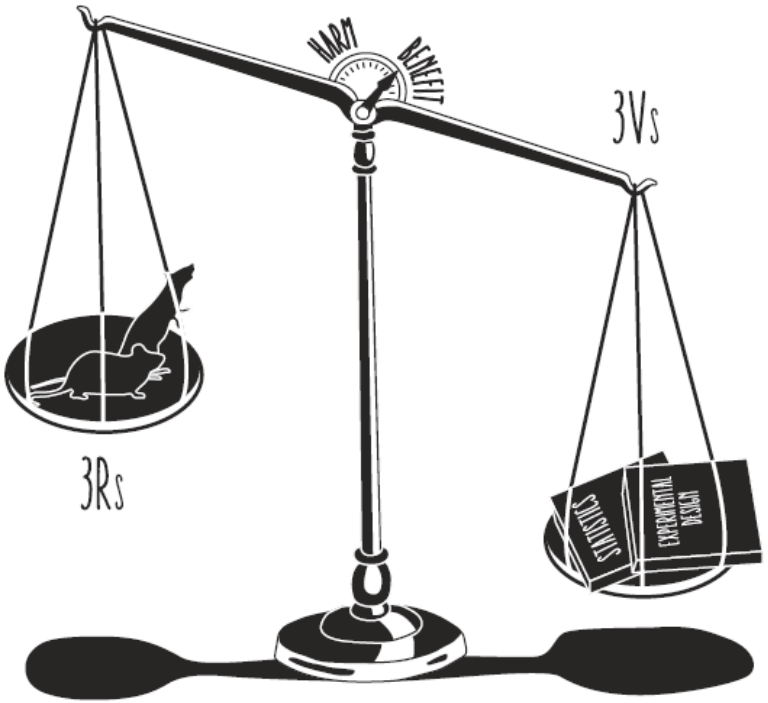
- Inferential statistics applied to hypothesis-free fishing expeditions
- Searching for p-values < 0.05 (P-hacking)
- Hypothesizing After the Results are Known (HARKing)
- Selective reporting of outcomes
- Publishing “positive” results only (Publication bias)

→ **Pre-registration of experimental design and statistical analysis plan**

Potential problems

The 3Vs

- Poor construct validity → poorly validated animal models and outcomes
- Poor internal validity → poor experimental design and conduct
- **Poor external validity → poor representation of target population**



Würbel 2017 **Lab Animal**

External validity and reproducibility

The birth of “reproducibility” as principle to distinguish facts from fiction

Robert Hooke setting up an air pump to replicate observations made by a Dutch scientist, witnessed by Royal Society members.

Rita Greer 2007 The Scientists
Source: **Wikimedia Commons**

Standardization and poor reproducibility

The standardization fallacy

«Standardization is the attempt to increase reproducibility at the expense of external validity.»

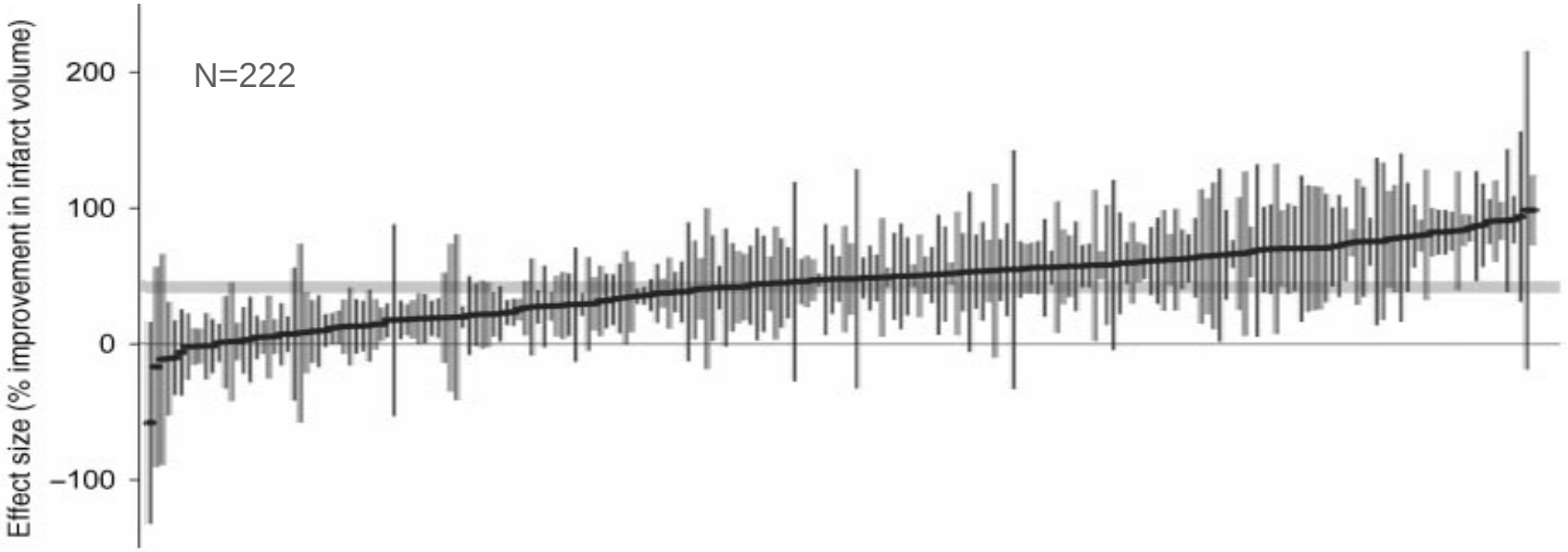
Würbel 2000 **Nat Genet**

«A highly standardized experiment supplies direct information only in respect of the narrow range of conditions achieved by standardization. ***Standardization, therefore, weakens rather than strengthens our ground for inferring a result****, when, as is the case in practice, these conditions are somewhat varied.»*

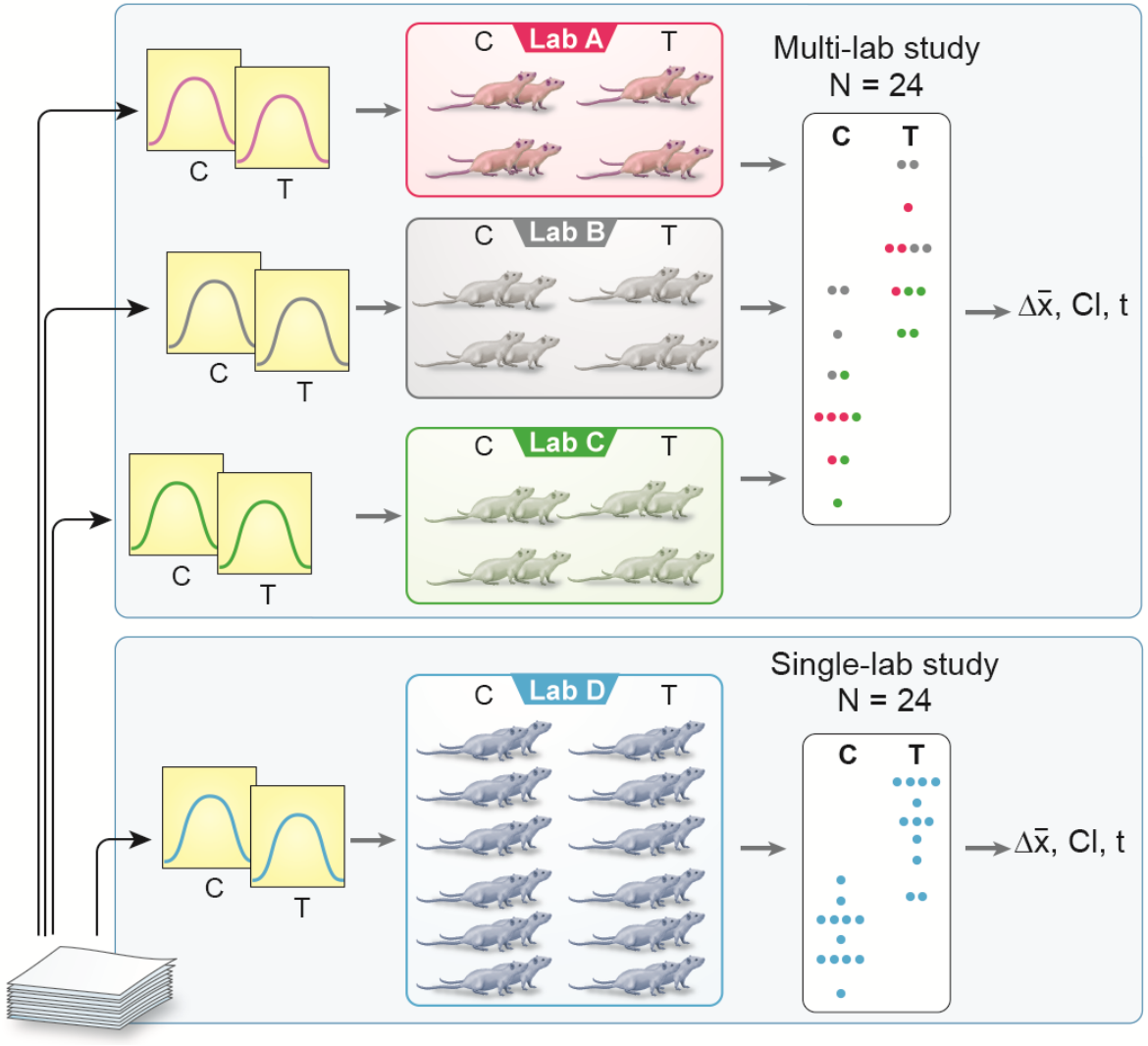
Fisher 1935 **The Design of Experiments**

Simulation of multi-lab studies using real data

Effect of hypothermia on infarct volume in animal models of stroke



Simulation of multi-lab studies using real data

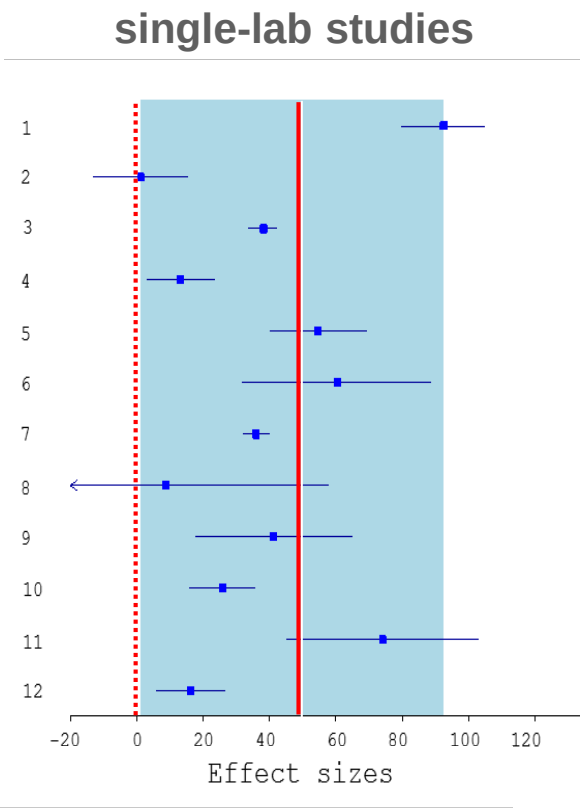


50 Studies

Voelkl et al. 2017 **unpublished results**

Simulation of single-lab versus multi-lab studies

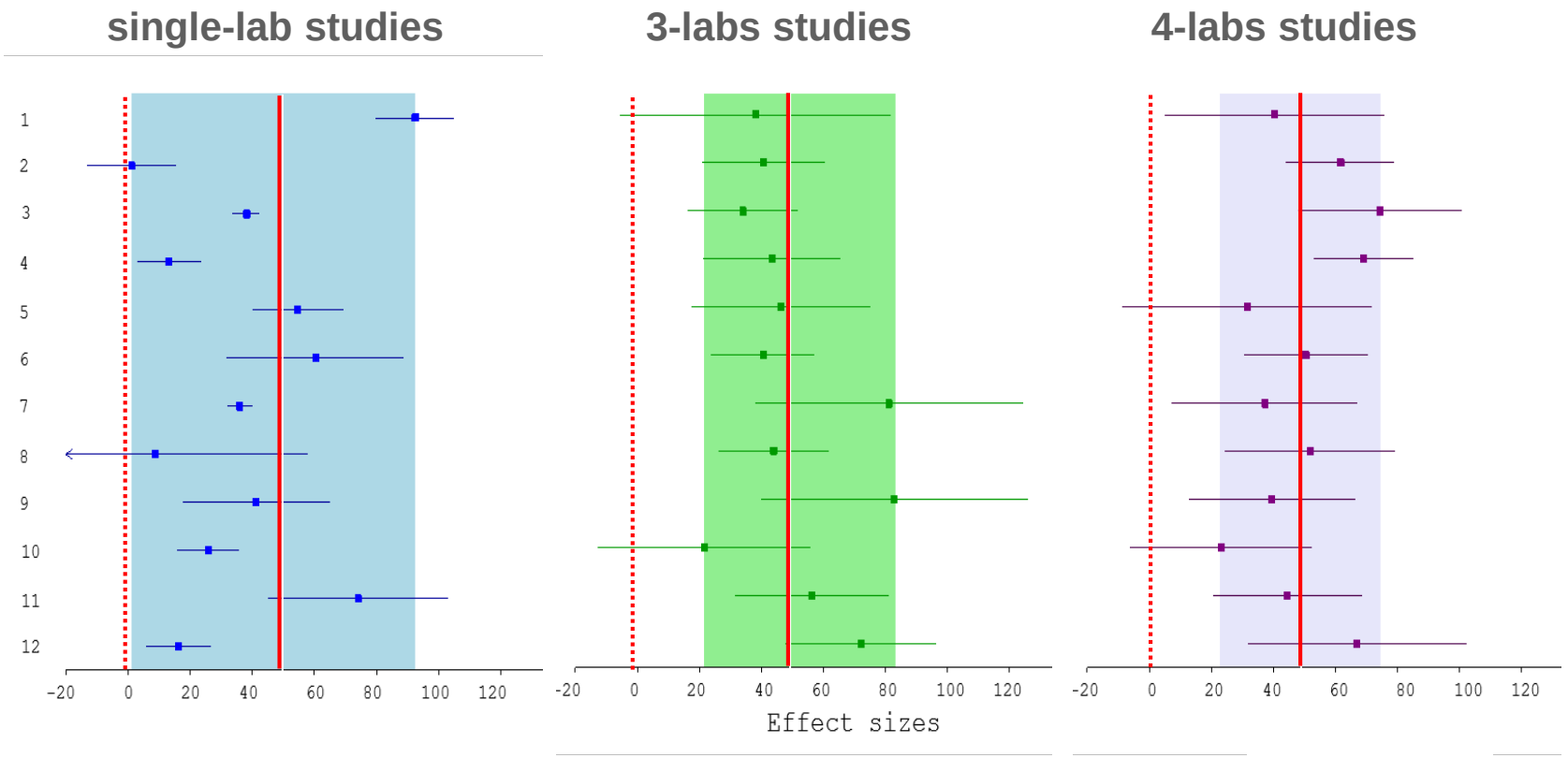
Effect of hypothermia on infarct volume in rodent models of stroke



Voelkl et al. 2017 **unpublished results**

Simulation of single-lab versus multi-lab studies

Effect of hypothermia on infarct volume in rodent models of stroke



Voelkl et al. 2017 **unpublished results**

Conclusions

- Results confirm poor reporting of measures against risks of bias
- Authorities approving animal experiments and editors publishing papers lack information to assess risks of bias
- Researchers overestimate scientific rigor of their own research
- Researchers lack problem awareness and knowledge to avoid risks of bias
- Reproducibility is dependent on external validity
- Standardized single-lab studies produce results of poor external validity and poor reproducibility
- Multi-lab studies (or other forms of heterogenization of study samples) will increase external validity and reproducibility

Acknowldgments

My lab

- Lucile Vogt
- Thomas Reichlin
- Bernhard Voelkl

Collaborators

- Christina Nathues (Bern)
- Emily Sena (Edinburgh)
- Malcolm Macleod (Edinburgh)
- Collaborative Approach to Meta Analysis and ·C·A·M·A·R·A·D·E·S· Review of Animal Data from Experimental Studies

Funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

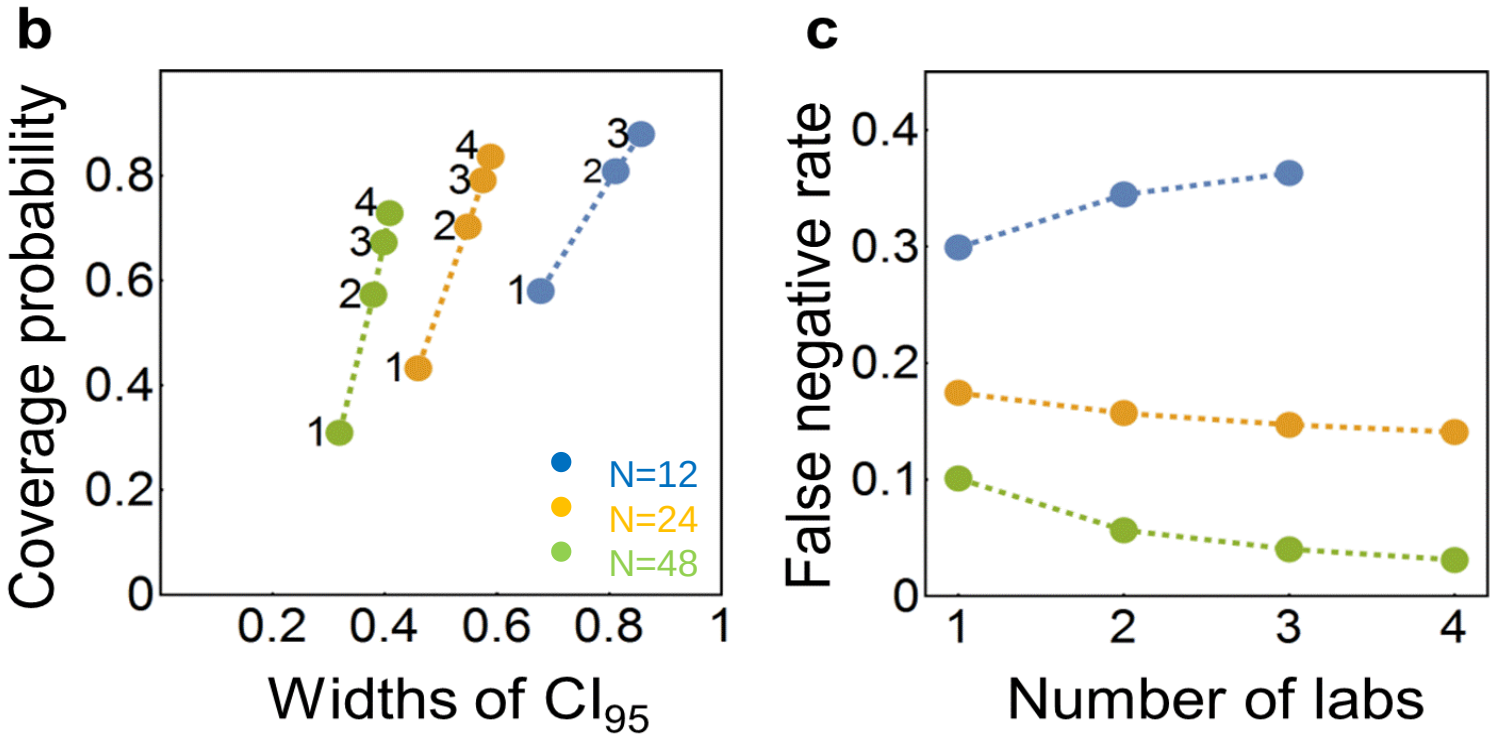
SEVENTH FRAMEWORK PROGRAMME

erc
European Research Council

UNIVERSITÄT BERN

Simulation of multi-lab studies using real data

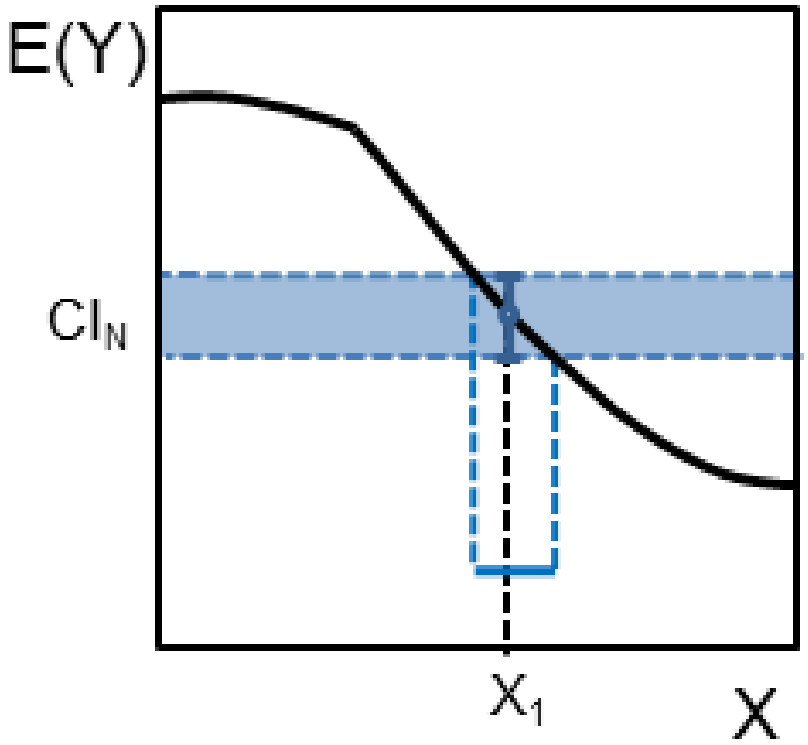
Coverage probability vs. false negative rate



Voelkl et al. 2016 **unpublished results**

Standardization and lack of replication

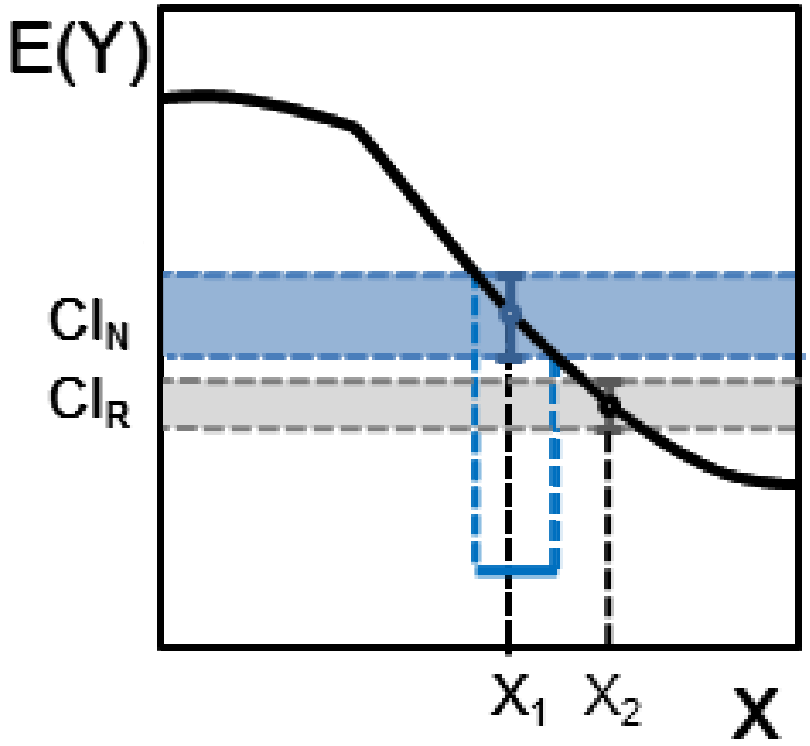
The reproducibility paradox



Voelkl & Würbel 2016 **Trends Pharmacol Sci**

Standardization and lack of replication

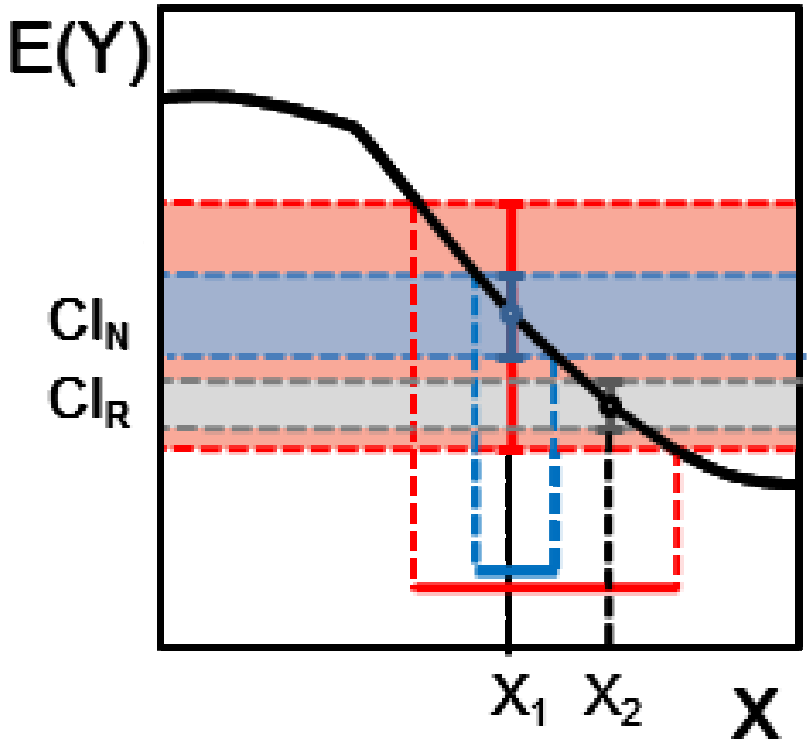
The reproducibility paradox



Voelkl & Würbel 2016 **Trends Pharmacol Sci**

Standardization and lack of replication

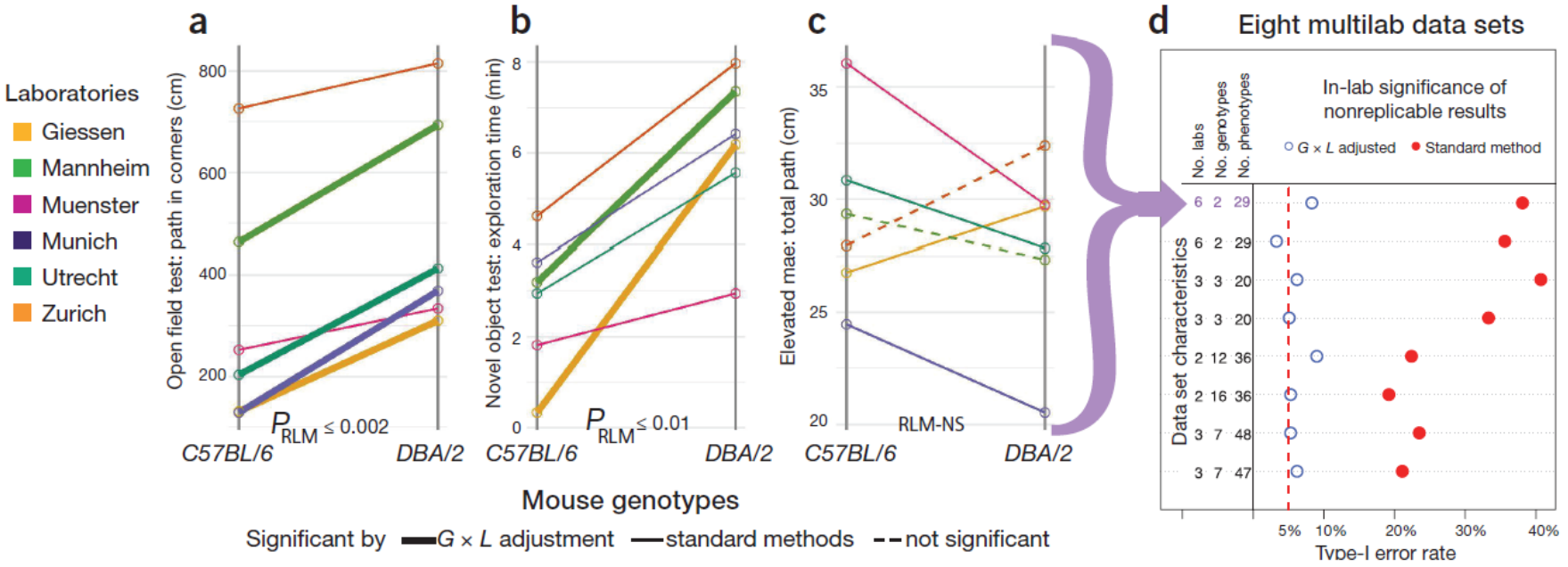
The reproducibility paradox



Voelkl & Würbel 2016 **Trends Pharmacol Sci**

Statistical correction for between-lab variation

Adjusting p-values based on data on genotype x laboratory interaction



Prototype web server for performing G × L adjustment:
https://stat.cs.tau.ac.il/gxl_app/

Kafkafi et al. 2017 **Nat Methods**