# 2018 3Rs Prize HIGHLY COMMENDED

## Reproducibility of preclinical animal research improves with heterogeneity of study samples.


Principal Investigator:
Dr Bernhard Voelkl

Organisation:
University of Bern



Reference:
Voelkl B *et al.* (2018) Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol* 16(2): e2003693.

Funders:
European Research Council
Swiss Food Safety and Veterinary Office


## Background

- Single-laboratory studies conducted under highly controlled conditions are the widespread 'gold standard' in preclinical animal research, and standardisation to reduce the variability of results is seen as good laboratory practice.
- However, standardisation makes study populations more homogenous, and contrary to guaranteeing reproducibility, this can bias the data produced in single-laboratory studies. This means that a large number of animals are being wasted producing potentially unreliable data from which conclusions cannot be drawn.
- Multi-centre trials are the accepted and established approach in clinical research, where maintaining rigour is critical.
- The aim of this study was to assess how performing experiments in multiple laboratories affects the outcome of *in vivo* research, with the hypothesis that it improves the accuracy and reproducibility of the results.

## Summary


- Dr Bernhard Voelkl from the University of Bern and colleagues have used a database of existing experimental data (from the CAMARADES group, Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies) to run simulations based on 440 preclinical studies across 13 different interventions in animal models of stroke, myocardial infarction and breast cancer.
- Using this data, the team compared the accuracy of effect size estimates between single-laboratory and multi-laboratory study designs.
- The analysis considered the use of smaller and larger group sizes and was tested in a wide range of experimental setups across the disease areas.
- Single-laboratory studies generally failed to predict treatment effect accurately, resulting in an under- or overestimation of animal group sizes. Using more animals in single-laboratory studies often resulted in even less valid estimates of effect size, as this reduced variability between measurements but did not improve the accuracy of the measurements.
- By contrast, multi-laboratory designs including as few as two to four laboratories increased the accuracy of the prediction by up to 42% without a need for larger sample sizes.
- These findings demonstrate that despite common assumption, within-study standardisation is a major cause of poor reproducibility.


## 3Rs impacts

- Multi-laboratory study designs are highly effective in producing accurate and reproducible results without the need for larger numbers of animals being used.
- Reproducibility of results from preclinical animal research is affected by a range of factors including a lack of scientific rigor, low statistical power, analytical flexibility, and publication bias. Importantly, the findings of this study hold even where all other sources of poor reproducibility are avoided, illustrating the importance of heterogenous study samples.
- The key outcome of the paper has been included in the concept discussion of a multinational consortium EQIPD ("European Quality in Preclinical Data") funded by the European Commission.
- The authors are planning to develop ways to help researchers, funders and policy makers understand the impact of implementing multi-laboratory trials in animals.

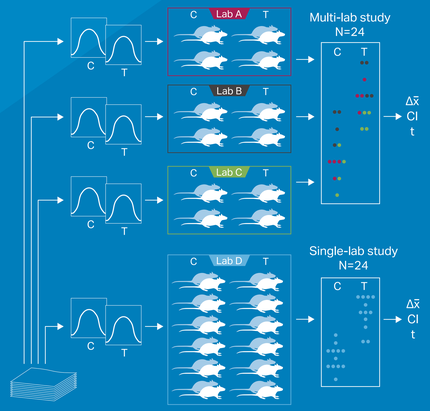


Figure 1: Sampling scheme for simulated single-lab and multi-lab studies. For a single-lab study, one original study is randomly selected from the study pool, and response values for control and treatment groups are generated by sampling from a Normal distribution with parameters as reported in the original study. For the multi-lab study, several original studies are selected, and values are sampled proportionate from the corresponding distributions. C, control group; SP, study pool; T, treatment group. Δx̄: the amount of change in the variable of interest; CI – confidence interval; t – t-distribution.

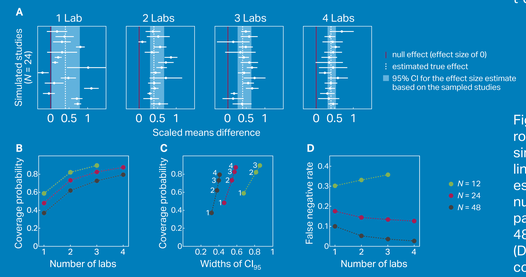


Figure 2: Results of resampling from studies on hypothermia in rodent models of stroke. (A) Forest plot of 15 randomly selected simulated studies for the 1-, 2-, 3-,and 4-lab scenario N = 24; dashed line: estimated true effect; shaded area: 95% CI for the effect size estimate based on the sampled studies. The red line indicates a null effect (effect size of 0). (B) pc plotted against the number of participating laboratories for N = 12 (green), N = 24 (red), and N = 48 (grey). (C) pc plotted against the average width of the 95% CI. (D) False negative rate plotted against number of laboratories. pc, coverage probability.