Reduction of group size towards zero: reduced power and diminishing returns

Gillian Currie
Centre for Clinical Brain Sciences, University of Edinburgh
Power

• The **probability of finding an effect if that effect is truly present**

• In clinical studies this is conventionally taken as 80%.

• However, many clinical trials now aim for 90% power.
• 80% power in a study means that if you do an experiment 5 times, in 1 of those 5 the analysis will be falsely negative.

• As we’ll see power is fundamental to the usefulness of preclinical research
If 10 scientists carry out an experiment powered at 80%...
If 10 scientists carry out an experiment powered at 30%…

- Under powered experiments waste time, resources and animals lives.
Power calculations

- Power calculations allow researchers to estimate the number of animals required to identify a given effect size at a certain probability threshold.

- A study should only be conducted if beforehand a power calculation indicates that a clinically or scientifically relevant effect has a high chance of being detected if it exists.
Sample size

• **The smaller the studies conducted in a scientific field, the less likely the research findings are to be true.**
  
  – *Small sample size means smaller power*
  

• Before an experiment researchers should estimate the sample size necessary to detect an effect that they would consider to be biologically important.

• An experiment that is too small will falsely conclude results, whereas an experiment that is too large will waste animals.
Why use so few animals?

- Time and budget constraints
"Sample size was smaller than planned because I had been in grad school for 10 yrs & my advisor wanted me to graduate."
#overlyhonestmethods
Why use so few animals?

- Overzealous interpretation of guidelines

- Animal researchers are encouraged to reduce the number of experimental animals to a minimum.
  
  E.g. In the UK the need to use the **minimum number of animals** to obtain **valid** results is expressed in the Animals (Scientific Procedures) Act 1986.

- The UK NC3Rs Reduction definition

  Methods that minimise the number of animals used per experiment or study, either by enabling researchers to obtain comparable levels of information from fewer animals, or to obtain more information from the same number of animals, thereby avoiding further animal use.
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

Sample Size Calculation
The manuscript should describe how the size of the experiment was planned. If a sample size calculation was performed this should be reported in detail, including the expected difference between groups, the expected variance, the planned analysis method, the desired statistical power and the sample size thus calculated. For parametric data, variance should be reported as 95% confidence limits or standard deviations rather than as the standard error of the mean.

Stroke 2009;40;e50-e52; originally published online Aug 14, 2008
Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research

Carol Kilkenny¹*, William J. Browne², Innes C. Cuthill³, Michael Emerson⁴, Douglas G. Altman⁵

<table>
<thead>
<tr>
<th>Sample size</th>
<th>10</th>
<th>a. Specify the total number of animals used in each experiment, and the number of animals in each experimental group.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>b. Explain how the number of animals was arrived at. Provide details of any sample size calculation used.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Indicate the number of independent replications of each experiment, if relevant.</td>
</tr>
</tbody>
</table>

June 29, 2010
A call for transparent reporting to optimize the predictive value of preclinical research

Sample-size estimation

- An appropriate sample size should be computed when the study is being designed and the statistical method of computation reported.

Raising standards

Nature journals’ updated editorial policies aim to improve transparency and reproducibility.


Statistics and general methods

1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?

For animal studies, include a statement about sample size estimate even if no statistical methods were used.
Reporting of sample size calculations

- Few studies identified in systematic reviews reported using a sample size calculation.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample Size calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>3%</td>
</tr>
<tr>
<td>Experimental autoimmune encephalomyelitis</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
<td>0%</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>0%</td>
</tr>
<tr>
<td>Glioma</td>
<td>0%</td>
</tr>
<tr>
<td>Bone cancer pain</td>
<td>0%</td>
</tr>
</tbody>
</table>
Calculating sample size

- Several factors must be known or estimated:
  - **Effect size** (usually the difference between 2 groups)
  - **Population standard deviation** (for continuous data)
  - Desired **power** of the experiment to detect the postulated effect
  - **Significance level**

unique to the particular experiment

generally fixed by convention
We hypothesised that preclinical experiments are underpowered.

We set out to determine whether this is the case.
Methods

• In the CAMARADES data set we have a large number of reviews with effect size and variance for a number of disease models

• We used this data for indicative power calculations

• Median effect size
• Median variance
• Modelled as a simple two-tailed t-test
• Using STATA
Chances of wasting an animal

Number of animals per group

Percentage of animals wasted

% of animal wasted = (1 - power)
Indicative power calculations - infarct volume in stroke studies

Number of animals per group

Power

Effect size (% improvement)

Number of animals

Effect size (% improvement)

Power

CAMARADES: Bringing evidence to translational medicine
Multiple sclerosis


- Mean effect size 40%
- Median standard error 28%
- Control group n=5
- Treatment group n=8
- Typical study powered at 63%

CAMARADES: Bringing evidence to translational medicine
Multiple Sclerosis
Experiments are often underpowered

- Sample size calculations based on published data for effect size and variance suggest that experiments should be substantially larger than is reported in the literature.

- The typical focal ischemia study is powered at only 28% - that is, even if the effect being sought is present, there is a 72% chance that it will not be detected.

- The typical MS study is powered at 63%.
Power is also related to the effect size

- *The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.*


- Systematic reviews have shown that experiments in animals with co-morbidities (e.g. diabetes, hypertension) manifest smaller treatment effects.

- Therefore we underestimate the number of animals required to detect potentially important treatment effects.
Reducing variance?

- Variance may be reduced by using rat strains that produce more-consistent infarct sizes
  - such as the spontaneously hypertensive rat,

- Or by using models that provide more-consistent vascular lesions.
  - Such as photothrombosis

- However, if consistency comes at the expense of generalizability (for instance, through use of models with a very small ischaemic penumbra), these strategies will prove to be false economies.

Bringing rigour to translational medicine Howells et al. NATURE REVIEWS | NEUROLOGY VOLUME 10 (2014)
Conclusions

• Experiments are often underpowered.

• It is crucial that publications include a statement about method of calculation of sample size and justification of sample size in the manuscript they want to publish.

• All components of a sample size calculation such as effect size, one tailed/two tailed test and standard deviation should be reported in a manuscript sent for publication.
Acknowledgments

- Malcolm Macleod, Emily Sena, Zsanett Bahor, Robert Stewart, Jing Liao, Hanna Vesterinen, Kieren Egan

- IVSyRMAF supported by NC3Rs