Data quality and reproducibility in preclinical research

Malcolm Macleod

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

University of Edinburgh

CAMARADES: Bringing evidence to translational medicine
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.
Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction

Craig M. Bennett, Abigail A. Baird, Michael B. Miller and George L. Wolford

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 mm³ 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. Jonas¹, Y. Lin², A. Williams², F. Tortella², R. Tuna³
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
<table>
<thead>
<tr>
<th>Mouse No.</th>
<th>Treatment begun</th>
<th>Dosage of penicillin</th>
<th>Route</th>
<th>Illness began</th>
<th>Survival period</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>V C₁</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V C₂</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V C₃</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V C₄</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V C₅</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V C₆</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V P₁</td>
<td>7 hrs. after injection</td>
<td>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 p.m.</td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
<tr>
<td>V P₂</td>
<td></td>
<td></td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
<tr>
<td>V P₃</td>
<td></td>
<td></td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
<tr>
<td>V P₄</td>
<td></td>
<td></td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
<tr>
<td>V P₅</td>
<td></td>
<td></td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
<tr>
<td>V P₆</td>
<td></td>
<td></td>
<td>I.P.</td>
<td>No illness</td>
<td></td>
<td>Survived</td>
</tr>
</tbody>
</table>

**TABLE V**

*Effect of Penicillin on Murine Typhus in Mice, Experiment V*

*Moderate Dosage (Mouse Brain Intraperitoneally) Approaching the Minimal Lethal Dosage*

Room temperature 65–72°F.
VSV vaccine for EBOLA
2016

BALB/c mice, challenged at 12 months

Hartley Guinea Pig, challenged at 18 months

CAMARADES: Bringing evidence to translational medicine
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

In vitro and in vivo - 1026
Tested in vivo - 603
Effective in vivo - 374
Tested in clinical trial - 97
Effective in clinical trial - 1

O’ Collins et al, 2006
CAMARADES: Bringing evidence to translational medicine
Risk of bias in animal studies

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

Macleod et al, 2008

CAMARADES: Bringing evidence to translational medicine
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”

<table>
<thead>
<tr>
<th>Group</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Maze bright”</td>
<td>1.33</td>
<td>1.60</td>
<td>2.60</td>
<td>2.83</td>
<td>3.26</td>
</tr>
<tr>
<td>“Maze dull”</td>
<td>0.72</td>
<td>1.10</td>
<td>2.23</td>
<td>1.83</td>
<td>1.83</td>
</tr>
<tr>
<td>Δ</td>
<td>+0.60</td>
<td>+0.50</td>
<td>+0.37</td>
<td>+1.00</td>
<td>+1.43</td>
</tr>
</tbody>
</table>

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

CAMARADES: Bringing evidence to translational medicine
# Treating “depression” in animals with probiotics

<table>
<thead>
<tr>
<th>ROB item</th>
<th>Percent Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Allocation to Group</td>
<td>25%</td>
</tr>
<tr>
<td>Blinded Assessment of Outcome</td>
<td>44%</td>
</tr>
<tr>
<td>Sample Size Calculation</td>
<td>6%</td>
</tr>
<tr>
<td>Reporting of Animal Exclusions</td>
<td>12%</td>
</tr>
</tbody>
</table>

Credit: Anthony Shek

[Graph showing the percentage of blinded assessment]
Evidence from various neuroscience domains…

CAMARADES: Bringing evidence to translational medicine
CAMARADES: Bringing evidence to translational medicine

The (polluted) research cycle
CAMARADES: Bringing evidence to translational medicine

RAE 1173

“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

CAMARADES: Bringing evidence to translational medicine
Different patterns of publication bias in different fields

<table>
<thead>
<tr>
<th></th>
<th>outcome</th>
<th>observed</th>
<th>corrected</th>
<th>Benefit/Harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease models</td>
<td>improvement</td>
<td>40%</td>
<td>30%</td>
<td>Less improvement</td>
</tr>
<tr>
<td>Toxicology model</td>
<td>harm</td>
<td>0.32</td>
<td>0.56</td>
<td>More harm</td>
</tr>
</tbody>
</table>

CAMARADES: Bringing evidence to translational medicine
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

<table>
<thead>
<tr>
<th>Number of animals per group</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of studies published</td>
<td>30%</td>
<td>54%</td>
<td>76%</td>
<td>86%</td>
</tr>
</tbody>
</table>

CAMARADES: Bringing evidence to translational medicine
How does that work?

Two sets of studies, one underpowered
**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.

<table>
<thead>
<tr>
<th>10 promising molecules found</th>
<th>10 false positives found</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 undetected</td>
<td>190 true negative results (rarely published)</td>
</tr>
</tbody>
</table>

20 preclinical studies showed promise and were published, but 10 (50%) were false positives.

**CAMARADES:** Bringing evidence to translational medicine
...with p<0.05, 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
- 10 undetected

42 studies showed promise and were published, and only 2 (5%) were false positives.

198 true negative results

**CAMARADES:** Bringing evidence to translational medicine
Good Laboratory Practice
Preventing Introduction of Bias at the Bench

Malcolm R. Macleod; Marc Fisher; Victoria O’Collins; Emily S. Sena; Ulrich Dirmaigl;
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

CAMARADES: Bringing evidence to translational medicine
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

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

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

CAMARADES: Bringing evidence to translational medicine
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)

CAMARADES: Bringing evidence to translational medicine
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

Exactly the same

Results reproducibility

The same methods

Robustness

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

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Developmental Rate</th>
<th>Fertility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic</td>
<td>83.1%</td>
<td>63.3%</td>
</tr>
<tr>
<td>Between labs</td>
<td>8.3%</td>
<td>7.9%</td>
</tr>
<tr>
<td>Within labs</td>
<td>3.8%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Individual</td>
<td>4.8%</td>
<td>23.3%</td>
</tr>
</tbody>
</table>

Lucanic et al Nature Comms 2017

CAMARADES: Bringing evidence to translational medicine
287 experiments identified in SyR reporting > 1 outcome domain
inflammation
demyelination
axon loss
neurobehaviour
Drug efficacy in MS

Inflammation

Demyelination

Axon Loss

Neurobehaviour

CAMARADES: Bringing evidence to translational medicine
Pre induction

- Demyelination
- Axon Loss
- Inflammation
- Neurobehaviour

1.000

CAMARADES: Bringing evidence to translational medicine
CAMARADES: Bringing evidence to translational medicine

Disease

- Inflammation
- Axon Loss
- Demyelination
- Neurobehaviour

0.464
0.066
0.470
0.530
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

20% power, p<0.05

80% power, p<0.01

CAMARADES: Bringing evidence to translational medicine
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.

CAMARADES: Bringing evidence to translational medicine
Camarades: Bringing evidence to translational medicine

Protocols

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

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


4 pages, one year

4 pages, one year
PHISPS protocols

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

CAMARADES: Bringing evidence to translational medicine
Strategies to increase benefits from research

Consolidate through adding to standard care

Continuous improvement cycles

CAMARADES: Bringing evidence to translational medicine
<table>
<thead>
<tr>
<th>What performance do we aspire to?</th>
<th>“95% of UoE manuscripts describing animal research report randomisation where this would be appropriate”</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is our current performance?</td>
<td>Measure, eg with ML/TM [2009-10 = 8%]</td>
</tr>
</tbody>
</table>
| 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                                                                                      |
CAMARADES: Bringing evidence to translational medicine

Define target performance

Measure performance

Seek to improve performance

Measure performance

Did we succeed?
**CAMARADES**: Bringing evidence to translational medicine

- 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
CAMARADES: Bringing evidence to translational medicine

1. Define target performance
2. Measure performance
3. Seek to improve performance
4. Measure performance
5. Did we succeed?
6. Consolidate into standard practice
### Strategies to increase value

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Study reports comply with existing guidelines such as the ARRIVE guidelines, so that there is transparency in what was done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>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</td>
</tr>
<tr>
<td>Level 3</td>
<td>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</td>
</tr>
<tr>
<td>Level 4</td>
<td>The existence of a study is asserted through some system of registration, to address the issue of publication bias</td>
</tr>
<tr>
<td>Level 5</td>
<td>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</td>
</tr>
<tr>
<td>Level 6</td>
<td>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</td>
</tr>
<tr>
<td>Level 7</td>
<td>Where the in vivo data appear promising, to develop tools for multicentre animal studies to confirm effects in &quot;preclinical phase 3 studies&quot;</td>
</tr>
</tbody>
</table>

**CAMARADES: Bringing evidence to translational medicine**
Announcement: Reducing our irreproducibility

24 April 2013

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

CAMARADES: Bringing evidence to translational medicine
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


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).


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
CAMARADES: Bringing evidence to translational medicine

NPG

2015

Pubmed “related citations” search

Non - NPG

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

May 2013

related citations search
CAMARADES: Bringing evidence to translational medicine

<table>
<thead>
<tr>
<th>NPG Publications (n=448)</th>
<th>Non NPG Publications (n=448)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publications</td>
<td></td>
</tr>
<tr>
<td>Before 01052013 (n=223)</td>
<td>After 01052013 (n=225)</td>
</tr>
<tr>
<td>Initial screen</td>
<td></td>
</tr>
<tr>
<td>Excluded: 3</td>
<td>Excluded: 1</td>
</tr>
<tr>
<td>Available for analysis</td>
<td></td>
</tr>
<tr>
<td>Analysis (n=220)</td>
<td>Analysis (n=224)</td>
</tr>
<tr>
<td>Exclusions</td>
<td></td>
</tr>
<tr>
<td>Excluded: 1</td>
<td>Excluded: nil</td>
</tr>
<tr>
<td>Final analysis set</td>
<td></td>
</tr>
<tr>
<td>n=219</td>
<td>n=224</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Types of experiment</th>
<th>In vivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>In vitro</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>144</td>
</tr>
<tr>
<td>42</td>
<td>148</td>
</tr>
<tr>
<td>30</td>
<td>104</td>
</tr>
<tr>
<td>77</td>
<td>112</td>
</tr>
</tbody>
</table>

| Compliance with          | In vivo                      |
| Landis 4 (in vivo)       |                              |
| 0/204                    |                              |
| 31/190                   |                              |
| 1/164                    |                              |
| 1/189                    |                              |
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

Randomisation

Blinding

Sample size calculation

Reporting exclusions

CAMARADES: Bringing evidence to translational medicine
In vitro experiments

Randomisation

Blinding

Power calculation

Reporting exclusions

CAMARADES: Bringing evidence to translational medicine
in vivo 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
CAMARADES: Bringing evidence to translational medicine

in vitro research, statistical reporting

(1) Exact  (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

(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?

CAMARADES: Bringing evidence to translational medicine
antibodies and cell culture details

(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 Checklist

Animal Research: Reporting In Vivo Experiments

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

CAMARADES: Bringing evidence to translational medicine

- Journals
- Funders
- Universities
- Learned societies
- Organisations
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

CAMARADES: Bringing evidence to translational medicine
### Results

**Sample size calculation**

**Primary:**
- 10% effect,
- \( \alpha = 0.05, \) \( \beta = 80\%
- \( n = 100/\text{group} \)

**Sub-items:**
- 20% effect
- \( \alpha = 0.0013, \) \( \beta = 80\%
- \( n = 200/\text{group} \)

---

**1689 submitted manuscripts randomised**

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>844</td>
<td>845</td>
</tr>
<tr>
<td>Sent for review</td>
<td>652</td>
<td>647</td>
</tr>
<tr>
<td>Accepted</td>
<td>322</td>
<td>340</td>
</tr>
<tr>
<td>Checklist completed</td>
<td>13</td>
<td>301</td>
</tr>
<tr>
<td>Full compliance</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

CAMARADES: Bringing evidence to translational medicine
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
• NC3Rs Experimental Design Assistant

https://www.nc3rs.org.uk/experimental-design-assistant-eda
CAMARADES: Bringing evidence to translational medicine

Protocol development and registration → Systematic search → Screen for inclusion → .pdf retrieval → Meta-data annotation → Outcome data extraction → Meta-analysis → Publication

- Living systematic search
- Machine Learning to aid Screening
- Automatic Annotation
- Data analysis app

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