

A call for transparent reporting to optimize the predictive value of preclinical research

Story C. Landis¹, Susan G. Amara², Khusru Asadullah³, Chris P. Austin⁴, Robi Blumenstein⁵, Eileen W. Bradley⁶, Ronald G. Crystal⁷, Robert B. Darnell⁸, Robert J. Ferrante⁹, Howard Fillit¹⁰, Robert Finkelstein¹, Marc Fisher¹¹, Howard E. Gendelman¹², Robert M. Golub¹³, John L. Goudreau¹⁴, Robert A. Gross¹⁵, Amelie K. Gubitzi¹, Sharon E. Hesterlee¹⁶, David W. Howells¹⁷, John Huguenard¹⁸, Katrina Kelnar¹⁹, Walter Koroshetz¹, Dimitri Krainc²⁰, Stanley E. Lazic²¹, Michael S. Levine²², Malcolm R. Macleod²³, John M. McCall²⁴, Richard T. Moxley III²⁵, Kalyani Narasimhan²⁶, Linda J. Noble²⁷, Steve Perrin²⁸, John D. Porter¹, Oswald Steward²⁹, Ellis Unger³⁰, Ursula Utz¹ & Shai D. Silberberg¹

The US National Institute of Neurological Disorders and Stroke convened major stakeholders in June 2012 to discuss how to improve the methodological reporting of animal studies in grant applications and publications. The main workshop recommendation is that at a minimum studies should report on sample-size estimation, whether and how animals were randomized, whether investigators were blind to the treatment, and the handling of data. We recognize that achieving a meaningful improvement in the quality of reporting will require a concerted effort by investigators, reviewers, funding agencies and journal editors. Requiring better reporting of animal studies will raise awareness of the importance of rigorous study design to accelerate scientific progress.

Dissemination of knowledge is the engine that drives scientific progress. Because advances hinge primarily on previous observations, it is essential that studies are reported in sufficient detail to allow the scientific community, research funding agencies and disease advocacy organizations to evaluate the reliability of previous findings. Numerous publications have called attention to the lack of transparency in reporting, yet studies in the life sciences in general, and in animals in particular, still often lack adequate reporting on the design, conduct and analysis of the experiments. To develop a plan for addressing this critical issue, the US National Institute of Neurological Disorders and Stroke (NINDS) convened academic researchers and educators, reviewers, journal editors and representatives from funding agencies, disease advocacy communities and the pharmaceutical industry to discuss the causes of deficient reporting and how they can be addressed. The specific goal of the meeting was to develop recommendations for improving how the results of animal research are reported in manuscripts and grant applications. There was broad agreement that: (1) poor reporting, often associated with poor experimental design, is a significant issue across the life sciences; (2) a core set of research parameters exist that should be addressed when reporting the results of animal experiments; and (3) a concerted effort by all stakeholders, including funding agencies and journals, will be necessary to disseminate and implement best reporting

practices throughout the research community. Here we describe the impetus for the meeting and the specific recommendations that were generated.

Widespread deficiencies in methods reporting

In the life sciences, animals are used to elucidate normal biology, to improve understanding of disease pathogenesis, and to develop therapeutic interventions. Animal models are valuable, provided that experiments employing them are carefully designed, interpreted and reported. Several recent articles, commentaries and editorials highlight that inadequate experimental reporting can result in such studies being uninterpretable and difficult to reproduce^{1–8}. For instance, replication of spinal cord injury studies through an NINDS-funded program determined that many studies could not be replicated because of incomplete or inaccurate description of experimental design, especially how randomization of animals to the various test groups, group formulation and delineation of animal attrition and exclusion were addressed⁷. A review of 100 articles published in *Cancer Research* in 2010 revealed that only 28% of papers reported that animals were randomly allocated to treatment groups, just 2% of papers reported that observers were blinded to treatment, and none stated the methods used to determine the number of animals per group, a determination required to avoid false outcomes². In addition, analysis of several

¹National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20892, USA. ²Department of Neurobiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. ³Bayer HealthCare, 13342 Berlin, Germany. ⁴National Center for Advancing Translational Sciences, NIH, Rockville, Maryland 20854, USA. ⁵CHDI Management/CHDI Foundation, New York, New York 10001, USA. ⁶Center for Review, NIH, Bethesda, Maryland 20892, USA. ⁷Department of Genetic Medicine, Weill Cornell Medical College, New York, New York 10021, USA. ⁸Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA. ⁹Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ¹⁰Alzheimer's Drug Discovery Foundation, New York, New York 10019, USA. ¹¹Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01545, USA. ¹²Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. ¹³JAMA, Chicago, Illinois 60654, USA. ¹⁴Department of Neurology, Michigan State University, East Lansing, Michigan 48824, USA. ¹⁵Department of Neurology, University of Rochester Medical Center, Rochester, New York 14642, USA. ¹⁶Parent Project Muscular Dystrophy, Hackensack, New Jersey 07601, USA. ¹⁷The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg 3081, Australia. ¹⁸Neurology and Neurological Sciences and Cellular and Molecular Physiology, Stanford University, Stanford, California 94305, USA. ¹⁹Science Translational Medicine, AAAS, Washington DC 22201, USA. ²⁰Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²¹F. Hoffmann-La Roche, 4070 Basel, Switzerland. ²²Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, California 90095, USA. ²³Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ²⁴PharMac LLC, Boca Grande, Florida 33921, USA. ²⁵University of Rochester Medical Center, School of Medicine and Dentistry, Rochester, New York 14642, USA. ²⁶Nature Neuroscience, New York, New York 10013, USA. ²⁷Department of Neurological Surgery, University of California San Francisco, San Francisco, California 94143, USA. ²⁸ALS Therapy Development Institute, Cambridge, Massachusetts 02139, USA. ²⁹Reeve-Irvine Research Center, University of California Irvine, Irvine, California 92697, USA. ³⁰Office of New Drugs, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland 20993, USA.

hundred studies conducted in animal models of stroke, Parkinson's disease and multiple sclerosis also revealed deficiencies in reporting key methodological parameters that can introduce bias⁶. Similarly, a review of 76 high-impact (cited more than 500 times) animal studies showed that the publications lacked descriptions of crucial methodological information that would allow informed judgment about the findings⁹. These deficiencies in the reporting of animal study design, which are clearly widespread, raise the concern that the reviewers of these studies could not adequately identify potential limitations in the experimental design and/or data analysis, limiting the benefit of the findings.

Some poorly reported studies may in fact be well-designed and well-conducted, but analysis suggests that inadequate reporting correlates with overstated findings^{10–14}. Problems related to inadequate study design surfaced early in the stroke research community, as investigators tried to understand why multiple clinical trials based on positive results in animal studies ultimately failed. Part of the problem is, of course, that no animal model can fully reproduce all the features of human stroke. It also became clear, however, that many of the difficulties stemmed from a lack of methodological rigor in the preclinical studies that were not adequately reported¹⁵. For instance, a systematic review and meta-analysis of studies testing the efficacy of the free-radical scavenger NXY-059 in models of ischaemic stroke revealed that publications that included information on randomization, concealment of group allocation, or blinded assessment of outcomes reported significantly smaller effect sizes of NXY-059 in comparison to studies lacking this information¹⁰. In certain cases, a series of poorly designed studies, obscured by deficient reporting, may, in aggregate, serve erroneously as the scientific rationale for large, expensive and ultimately unsuccessful clinical trials. Such trials may unnecessarily expose patients to potentially harmful agents, prevent these patients from participating in other trials of possibly effective agents, and drain valuable resources and energy that might otherwise be more productively spent.

A core set of reporting standards

The large fraction of poorly reported animal studies and the empirical evidence of associated bias^{6,10–14,16–20}, defined broadly as the introduction of an unintentional difference between comparison groups, led various disease communities to adopt general^{12–23} and animal-model-specific^{6,24–26} reporting guidelines. However, for guidelines to be effective and broadly accepted by all stakeholders, they should be universal and focus on widely accepted core issues that are important for study evaluation. Therefore, based on available data, we recommend that, at minimum, authors of grant applications and scientific publications should report on randomization, blinding, sample-size estimation and the handling of all data (see below and Box 1).

Randomization and blinding

Choices made by investigators during the design, conduct and interpretation of experiments can introduce bias, resulting in false-positive results. Many have emphasized the importance of randomization and blinding as a means to reduce bias^{6,21–23,27}, yet inadequate reporting of these aspects of study design remains widespread in preclinical research. It is important to report whether the allocation, treatment and handling of animals were the same across study groups. The selection and source of control animals needs to be reported as well, including whether they are true littermates of the test groups. Best practices should also include reporting on the methods of animal randomization to the various experimental groups, as well as on random (or appropriately blocked) sample processing and collection of data. Attention to these details will avoid mistaking batch effects for treatment effects (for example, dividing samples from a large study into multiple lots, which are then processed separately). Investigators should also report on whether the individuals caring for the animals and conducting the experiments were blinded to the allocation sequence, blinded to group allocation and, whenever possible, whether the persons assessing, measuring or quantifying the experimental outcomes were blinded to the intervention.

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.

Sample-size estimation

Minimizing the use of animals in research is not only a requirement of funding agencies around the world but also an ethical obligation. It is unethical, however, to perform underpowered experiments with insufficient numbers of animals that have little prospect of detecting meaningful differences between groups. In addition, with smaller studies, the positive predictive value is lower, and false-positive results can ensue, leading to the needless use of animals in subsequent studies that build upon the incorrect results²⁸. Studies with an inadequate sample size may also provide false-negative results, where potentially important findings go undetected. For these reasons it is crucial to report how many animals were used per group and what statistical methods were used to determine this number.

Data handling

Common practices related to data handling that can also lead to false positives include interim data analysis²⁹, the *ad hoc* exclusion of data³⁰, retrospective primary end point selection³¹, pseudo replication³² and small effect sizes³³.

Interim data analysis

It is not uncommon for investigators to collect some data and perform an interim data analysis. If the results are statistically significant in favour of the working hypothesis, the study is terminated and a paper

is written. If the results look 'promising' but are not statistically significant, additional data are collected. This has been referred to as 'sampling to a foregone conclusion' and can lead to a high rate of false-positive findings^{29,30}. Therefore, sample size and rules for stopping data collection should be defined in advance and properly reported. Unplanned interim analyses, which can inflate false-positive outcomes and require unblinding of the allocation code, should be avoided. If there are interim analyses, however, these should be reported in the publication.

Ad hoc exclusion of data

Animal studies are often complex and outliers are not unusual. Decisions to include or exclude specific animals on the basis of outcomes (for example, state of health, dissimilarity to other data) have the potential to influence the study results. Thus, rules for inclusion and exclusion of data should be defined prospectively and reported. It is also important to report whether all animals that were entered into the experiment actually completed it, or whether they were removed, and if so, for what reason. Differential attrition between groups can introduce bias. For example, a treatment may appear effective if it kills off the weakest or most severely affected animals whose fates are then not reported. In addition, it is important to report whether any data were removed before analysis and the reasons for this data exclusion.

Retrospective primary end-point selection

It is well known that assessment of multiple end points, and/or assessment of a single end point at multiple time points, inflates the type-I error (false-positive results)³¹. Yet it is not uncommon for investigators to select a primary end point only after data analyses. False-positive conclusions arising from such practices can be avoided by specifying a primary end point before the study is undertaken, the time(s) at which the end point will be assessed, and the method(s) of analysis. Significant findings for secondary end points can and should be reported, but should be delineated as exploratory in nature. If multiple end points are to be assessed, then appropriate statistical corrections should be applied to control type-I error, such as Bonferroni corrections^{31,34}.

Pseudo replicates

When considering sample-size determination and experimental design, pseudo-replication issues need to be considered³². There is a clear, but often misunderstood or misrepresented, distinction between technical and biologic replicates. For example, in analysing effects of pollutants on reproductive health, multiple sampling from a litter, regardless of how many littermates are quantified, provides data from only a single biologic replicate. When biologic variation in response to some intervention is the variable of interest, as in many animal experiments, analysis of samples from multiple litters is essential. The unit of assessment is the smallest unit (animal, cage, litter) to which the intervention in question can be independently administered³⁵.

Small effect sizes

A statistically significant result does not provide information on the magnitude of the effect and thus does not necessarily mean that the effect is robust, which could account for the poor reproducibility of certain studies³⁶. Therefore, reporting whether results were substantiated by repetition, preferably under a range of conditions that demonstrate the robustness of the effect is encouraged. Also, reporting how often the particular experiment was performed as a means to control for a general tendency to publish only the best results would strengthen the validity of experimental results. To this end, carefully designed and powered animal studies should be budgeted for in the grant applications and funding agencies should consider supporting repetition studies where appropriate.

An important note about exploratory experiments

For the most part, these best practices do not apply to early-stage observational experiments searching for possible differences among experimental groups. Such exploratory testing is frequently conducted using a

small sample size, does not have a primary outcome and is often unblinded. However, because such experiments are likely to be subject to many of the limitations described above, they should be viewed as hypothesis-generating experiments and interpreted as such. Potential discoveries arising from the exploratory phase of the research should be supported by follow-up, hypothesis-testing experiments that take into consideration and adequately report on the core standards detailed above (Box 1).

The path to implementation

Improving the transparency and quality of reporting cannot be achieved by a single party, but will require cooperation among all stakeholders, including investigators, reviewers, funding agencies and journals. Calling upon investigators to provide key information about the design, execution and analysis of animal experiments described in grant applications and manuscripts and encouraging reviewers to consider these issues in their evaluations should, over time, increase both the quality and predictive value of preclinical research. Potential strategies for achieving this goal can be adopted from the clinical trials community, which also contended with poor reporting and associated bias. Evidence that clinical trials can yield biased results if they lack methodological rigor^{37–42} led to the development and implementation of the CONSORT guidelines for randomized clinical trials (among other guidelines), now adopted by many clinical journals and funding organizations. These guidelines require that authors report whether and how their studies were carried out blind and randomized, how sample size was determined, whether data are missing owing to attrition or exclusion, and supply information about other important experimental parameters^{43–45}. Importantly, the guidelines have improved the transparency of clinical study reporting in journals that have adopted them^{46–49}. Additional evidence for the power of such guidelines can be deduced from the observation that, although few animal studies report on randomization, blinding or sample-size determination, most describe compliance with animal regulations, which is required by journals^{6,9,10,50,51}.

As a first step, we recommend that funding organizations and journals provide reviewers with clear guidance about core features of animal study design (listed in Box 1). The goal is not to be prescriptive or proscriptive, but rather to delineate the minimum set of standards that should routinely be considered in evaluating the appropriateness of a study. Such guidance would make the task easier for reviewers of manuscripts and grant applications who volunteer their time and are often overextended. In addition, investigators and reviewers should be encouraged to consult published generic and model-specific guidelines for designing *in vivo* animal experiments^{6,21–27,52,53}. To assist reviewers, editors and funding organizations in making sure that applications and manuscripts contain sufficient information on the core reporting recommendation (Box 1), authors could be asked to append relevant information on a standardized form that accompanies the submission. This form could be as simple as a checkbox indicating the page on which the key reporting standard is addressed. Such a form is currently used by clinical research journals.

In addition to the measures proposed above, better dissemination of knowledge will be greatly facilitated by addressing publication bias, the phenomenon that few studies showing negative outcomes are published^{54–63}. Such deficiency in reporting contributes to needless repetition of similar studies by investigators unaware of earlier efforts^{60,61}. There is a widely accepted belief that the scientific community, promotions committees, funding agencies and journals favour positive outcomes, an impression that can lead to bias⁶⁴. Possible solutions include incentivizing investigators to publish negative outcomes, supporting studies of independent replication, encouraging journals to publish a greater number of studies reporting negative outcomes, creating a database for negative outcomes (analogous to <http://ClinicalTrials.gov/>), and linking the raw data to publications.

Change will not occur overnight. The importance of training scientists to properly design and adequately report animal studies cannot be overstated. Training and education focused on key features of experimental

design should be an ongoing process for both the novice and veteran involved in biomedical research. Attention to better study design reporting should be communicated at major meetings, brought to the attention of reviewers, editors and funders, required by the publishers of peer-review journals, and included in the training program of graduate and postdoctoral students. Furthermore, good mentorship is crucial for developing such skills and should be encouraged and rewarded. Rigorous experimental design and adequate reporting needs to be emphasized across the board and monitored in training grants awarded by the US National Institute of Health (NIH) and other funding agencies. Professional societies can also have an important role by highlighting this issue in their respective communities.

An important gatekeeper of quality remains the peer review of grant applications and journal manuscripts. We therefore call upon funding agencies and publishing groups to take actions to reinforce the importance of methodological rigor and reporting. NINDS has begun taking steps to promote best practices for preclinical therapy development studies. In 2011, a Notice was published in the NIH Guide encouraging the scientific community to address the issues described above in their grant applications, in describing both the project being proposed and the supporting data upon which it is based (<http://grants.nih.gov/grants/guide/notice-files/NOT-NS-11-023.html>). Points that should be considered in a well-designed study are listed on the NINDS website (http://www.ninds.nih.gov/funding/transparency_in_reporting_guidance.pdf). Furthermore, the reviewers of applications reviewed by the NINDS Scientific Review Branch are reminded of these issues and asked to pay careful attention to the scientific premise of the proposed projects.

We believe that improving how animal studies are reported will raise awareness of the importance of rigorous study design. Such increased awareness will accelerate both scientific progress and the development of new therapies.

Received 21 August; accepted 10 September 2012.

- Begley, C. G. & Ellis, L. M. Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
- Hess, K. R. Statistical design considerations in animal studies published recently in *Cancer Research*. *Cancer Res.* **71**, 625 (2011).
- Kilkenny, C. *et al.* Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE* **4**, e7824 (2009).
- Moher, D., Simera, I., Schulz, K. F., Hoey, J. & Altman, D. G. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med.* **6**, 13 (2008).
- Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Rev. Drug Discov.* **10**, 712 (2011).
- The first report that many published studies cannot be reproduced by the pharmaceutical industry.**
- Sena, E., van der Worp, H. B., Howells, D. & Macleod, M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci.* **30**, 433–439 (2007).
- Steward, O., Popovich, P. G., Dietrich, W. D. & Kleitman, N. Replication and reproducibility in spinal cord injury research. *Exp. Neurol.* **233**, 597–605 (2012).
- van der Worp, H. B. & Macleod, M. R. Preclinical studies of human disease: time to take methodological quality seriously. *J. Mol. Cell. Cardiol.* **51**, 449–450 (2011).
- Hackam, D. G. & Redelmeier, D. A. Translation of research evidence from animals to humans. *J. Am. Med. Assoc.* **296**, 1727–1732 (2006).
- A study reporting that a large fraction of high-impact publications in highly reputable journals lack important information related to experimental design.**
- Macleod, M. R. *et al.* Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* **39**, 2824–2829 (2008).
- A study demonstrating that lack of reporting of key methodological parameters is associated with bias.**
- Bebarta, V., Luyten, D. & Heard, K. Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad. Emerg. Med.* **10**, 684–687 (2003).
- Crossley, N. A. *et al.* Empirical evidence of bias in the design of experimental stroke studies – A metaepidemiologic approach. *Stroke* **39**, 929–934 (2008).
- Rooke, E. D., Vesterinen, H. M., Sena, E. S., Egan, K. J. & Macleod, M. R. Dopamine agonists in animal models of Parkinson's disease: a systematic review and meta-analysis. *Parkinsonism Relat. Disord.* **17**, 313–320 (2011).
- Vesterinen, H. M. *et al.* Improving the translational hit of experimental treatments in multiple sclerosis. *Mult. Scler. J.* **16**, 1044–1055 (2010).
- Stroke Therapy Academic Industry Roundtable (STAIR). Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke* **30**, 2752–2758 (1999).
- Fanelli, D. "Positive" results increase down the hierarchy of the sciences. *PLoS ONE* **5**, e10068 (2010).
- Jerndal, M. *et al.* A systematic review and meta-analysis of erythropoietin in experimental stroke. *J. Cereb. Blood Flow Metab.* **30**, 961–968 (2010).
- Macleod, M. R., O'Collins, T., Horky, L. L., Howells, D. W. & Donnan, G. A. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. *J. Cereb. Blood Flow Metab.* **25**, 713–721 (2005).
- Sena, E. S. *et al.* Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis. *J. Cereb. Blood Flow Metab.* **30**, 1905–1913 (2010).
- Wheble, P. C. R., Sena, E. S. & Macleod, M. R. A systematic review and meta-analysis of the efficacy of piracetam and piracetam-like compounds in experimental stroke. *Cerebrovasc. Dis.* **25**, 5–11 (2008).
- Festing, M. F. & Altman, D. G. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J.* **43**, 244–258 (2002).
- Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
- van der Worp, H. B. *et al.* Can animal models of disease reliably inform human studies? *PLoS Med.* **7**, e1000245 (2010).
- Fisher, M. *et al.* Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke* **40**, 2244–2250 (2009).
- Ludolph, A. C. *et al.* Guidelines for preclinical animal research in ALS/MND: a consensus meeting. *Amyotroph. Lateral Scler.* **11**, 38–45 (2010).
- Shineman, D. W. *et al.* Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. *Alzheimers Res. Ther.* **3**, 28 (2011).
- Unger, E. F. All is not well in the world of translational research. *J. Am. Coll. Cardiol.* **50**, 738–740 (2007).
- Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
- Dienes, Z. Bayesian versus orthodox statistics: which side are you on? *Perspect. Psychol. Sci.* **6**, 274–290 (2011).
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
- Beal, K. G. & Khamis, H. J. A problem in statistical-analysis: simultaneous inference. *Condor* **93**, 1023–1025 (1991).
- Lazic, S. E. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci.* **11**, 5 (2010).
- Scott, S. *et al.* Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph. Lateral Scler.* **9**, 4–15 (2008).
- An enlightening analysis of how small sample sizes can lead to false-positive outcomes.**
- Proschan, M. A. & Wacziarg, M. A. Practical guidelines for multiplicity adjustment in clinical trials. *Control. Clin. Trials* **21**, 527–539 (2000).
- Festing, M. F. W. Design and statistical methods in studies using animal models of development. *ILAR J.* **47**, 5–14 (2006).
- Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **82**, 591–605 (2007).
- Chalmers, T. C., Celano, P., Sacks, H. S. & Smith, H. Bias in treatment assignment in controlled clinical-trials. *N. Engl. J. Med.* **309**, 1358–1361 (1983).
- Jüni, P., Altman, D. G. & Egger, M. Systematic reviews in health care - assessing the quality of controlled clinical trials. *Br. Med. J.* **323**, 42 (2001).
- Pildal, J. *et al.* Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int. J. Epidemiol.* **36**, 847–857 (2007).
- Pocock, S. J., Hughes, M. D. & Lee, R. J. Statistical problems in the reporting of clinical-trials. A survey of three medical journals. *N. Engl. J. Med.* **317**, 426–432 (1987).
- Schulz, K. F., Chalmers, I., Hayes, R. J. & Altman, D. G. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J. Am. Med. Assoc.* **273**, 408–412 (1995).
- Wood, L. *et al.* Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Br. Med. J.* **336**, 601–605 (2008).
- Moher, D. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *Br. Med. J.* **340**, c869 (2011).
- Moher, D., Schulz, K. F. & Altman, D. G. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* **357**, 1191–1194 (2001).
- Revision of guidelines by the CONSORT group to improve the reporting of randomized clinical trials.**
- Schulz, K. F., Altman, D. G. & Moher, D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med.* **7**, e1000251 (2010).
- Plint, A. C. *et al.* Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med. J. Aust.* **185**, 263–267 (2006).
- Kane, R. L., Wang, J. & Garrard, J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J. Clin. Epidemiol.* **60**, 241–249 (2007).
- Prady, S. L., Richmond, S. J., Morton, V. M. & Macpherson, H. A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE* **3**, e1577 (2008).
- Smith, B. A. *et al.* Quality of reporting randomized controlled trials (RCTs) in nursing literature: application of the consolidated standards reporting trials (CONSORT). *Nurs. Outlook* **56**, 31–37 (2008).
- Macleod, M. R., O'Collins, T., Howells, D. W. & Donnan, G. A. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* **35**, 1203–1208 (2004).

51. Macleod, M. R., O'Collins, T., Horky, L. L., Howells, D. W. & Donnan, G. A. Systematic review and meta-analysis of the efficacy of melatonin in experimental stroke. *J. Pineal Res.* **38**, 35–41 (2005).
52. Gallo, J. M. Pharmacokinetic/pharmacodynamic-driven drug development. *Mt. Sinai J. Med.* **77**, 381–388 (2010).
53. Moher, D. *et al.* Describing reporting guidelines for health research: a systematic review. *J. Clin. Epidemiol.* **64**, 718–742 (2011).
54. Callaham, M. L., Wears, R. L., Weber, E. J., Barton, C. & Young, G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *J. Am. Med. Assoc.* **280**, 254–257 (1998).
55. Dickersin, K. & Chalmers, I. Recognizing, investigation and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. *J. R. Soc. Med.* **104**, 532–538 (2011).
56. Fanelli, D. Negative results are disappearing from most disciplines and countries. *Scientometrics* **90**, 891–904 (2012).
57. Kyzas, P. A., Denaxa-Kyza, D. & Ioannidis, J. P. A. Almost all articles on cancer prognostic markers report statistically significant results. *Eur. J. Cancer* **43**, 2559–2579 (2007).
58. Liu, S. Dealing with publication bias in translational stroke research. *J. Exp. Stroke Transl. Med.* **2**, 16–21 (2009).
59. Rockwell, S., Kimler, B. E. & Moulder, J. E. Publishing negative results: the problem of publication bias. *Radiat. Res.* **165**, 623–625 (2006).
60. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 638–641 (1979).
61. Sterling, T. D. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *J. Am. Stat. Assoc.* **54**, 30–34 (1959).
62. Song, F. *et al.* Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess.* **14**, 1–220 (2010).
63. Sena, E. S., van der Worp, H. B., Bath, P. M. W., Howells, D. W. & Macleod, M. R. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* **8**, e1000344 (2010).
64. Fanelli, D. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE* **5**, e10271 (2010).

Acknowledgements Funded by NINDS.

Author Contributions R.F., A.K.G., S.C.L., J.D.P., S.D.S., U.U. and W.K. organized the workshop. R.B.D., S.E.L., S.C.L., M.R.M. and S.D.S. wrote the manuscript. All authors participated in the workshop and contributed to the editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.D.S. (silberbs@ninds.nih.gov).