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

Background and Purpose—As a research community, we have failed to demonstrate that drugs which show substantial efficacy in animal models of cerebral ischemia can also improve outcome in human stroke.

Summary of Review—Accumulating evidence suggests this may be due, at least in part, to problems in the design, conduct and reporting of animal experiments which create a systematic bias resulting in the overstatement of neuroprotective efficacy.

Conclusions—Here, we set out a series of measures to reduce bias in the design, conduct and reporting of animal experiments modeling human stroke. **(*Stroke*. 2009;40:e50-e52.)**

Key Words: animal models ■ basic science ■ drug trials ■ education ■ experimental ■ outcomes ■preventing bias ■ translational research

Downloaded from http://stroke.ahajournals.org/ by guest on January 1, 2017

Nearly 10 years after the first Stroke Therapy Academic Industry Roundtable participants established guidelines intended to support the translation of neuroprotective efficacy from bench to bedside,[1] there is still no clinically effective neuroprotective drug for stroke. One interpretation of this observation is that measures outlined in STAIR I have failed to deliver the promised improvements in drug development. However, a dispassionate analysis of data presented over the last 10 years suggests that the "STAIR hypothesis"—that improvements in animal experimental design will lead to improvements in translational efficiency—has yet to be adequately tested. Adhering to standards of conducting and reporting of experiments in order to reduce the confounding effects of bias and ensure adequate statistical power, as outlined below, will increase the confidence with which we can assess new data and maximize our chances of developing effective therapies.

The original STAIR proposal was that by paying due attention to experimental bias, to the breadth of physiological variables known to influence stroke outcome in patients, and by testing therapies in a range of model systems which might more faithfully reproduce the key facets of stroke pathophysiology, we would be able to translate what appeared to be clear evidence of neuroprotective efficacy in animals to the more heterogeneous circumstances of human stroke. Although we believe strongly that failure to adequately consider variables such as age, comorbidity, physiological status, and timing of drug administration contribute to the disparity between the results of animal models and clinical trials, they have been reviewed elsewhere[1,2] and are not the subject of this article.

Analyses of data supporting the efficacy of various neuroprotective strategies[3–5] have revealed that although many researchers adhere closely to the ethos of these guidelines, as a community we do not. A simple checklist derived from the STAIR guidelines to provide an overview of the range of data available for 1026 candidate therapies[4] revealed that only a handful came close to meeting the STAIR guidelines. A higher score against this checklist was accompanied by a marked reduction in effect size. This later trend could be seen clearly even within the data for individual drugs.[6] Moreover, studies which reported measures to avoid bias such as random allocation to treatment group, blinded induction of ischemia or the blinded assessment of outcome,[7,8] gave a markedly lower estimate of efficacy. Despite this, there has been some evidence of improvement in study quality, and the performance of animal stroke studies is substantially better than that for most other models of neurological disease.[5] And yet, the

Received May 14, 2008; final revision received July 16, 2008; accepted July 18, 2008.

From the Centre for Clinical Brain Sciences (M.R.M., E.S.S.), University of Edinburgh, UK; the Department of Neurology, University of Massachusetts Medical School (M.F.), Worcester, Mass, USA; the National Stroke Research Institute & University of Melbourne Department of Medicine (V.O., E.S.S., G.A.D., D.W.H.), Austin Health, Melbourne, Australia; the Charité Department for Experimental Neurology & Center for Stroke Research (U.D.), Berlin, Germany; the Stroke Trials Unit (P.M.W.B.), University of Nottingham, UK; the Acute Stroke Program (A.B.), Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK; the Department of Neurology (H.B.v.d.W.), Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands; the University of Colorado Denver (R.T.), Anschutz Medical Campus, Aurora, Colo, USA; and the Cerebrovascular Division (K.M.), Department of Medicine, National Cardiovascular Center, Osaka, Japan.

Correspondence to David Howells, National Stroke Research Institute, Austin Health, Studley Rd, Heidelberg, Victoria 3084, Australia. E-mail david.howells@unimelb.edu.au

© 2009 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org DOI: 10.1161/STROKEAHA.108.525386

majority of investigators still do not report whether or not they took measures to avoid bias.

Systematic reviews and meta-analyses of data from animal stroke studies suggest that these studies may be substantially distorted by experimental bias. Taken together, publications supporting the efficacy of NXY-059 include randomized data with allocation concealment and blinded outcome assessment, but most individual publications do not report these measures. Analyses of those data suggests that at least half of the reported 44% improvement in outcome could be attributed to experimental bias, specifically a failure to randomize the allocation to experimental group, a failure to conceal treatment group allocation from the surgeon or a failure to blind the assessment of outcome.[8] Similar observations have been made of the hypothermia literature, where nonrandomized studies and studies without blinded outcome assessment appear to give a relative overstatement of efficacy of 27% and 19%, respectively.[7] Despite the widely recognized importance of these aspects of study design, analyses conducted by the CAMARADES group suggest that only 36% of studies reported random allocation to treatment group, only 11% report allocation concealment and only 29% reported the blinded assessment of outcome.[5]

A related issue is the number of animals used in experiments. The probability of detecting a difference of a given size between groups is related to the number of animals in each group, the size of the difference and the variability in the outcome measure used. However, only 3% of studies identified in systematic reviews reported using a sample size calculation.[5] Importantly, if sample size calculations are based on falsely large estimates of effect size, studies will not be powered to detect real differences between treatment and control groups. Indeed, post hoc analysis suggests that most experimental stroke studies have only a 1 in 3 chance of detecting a 20% difference in outcome.

These problems are not unique to the preclinical study of stroke. Clinical stroke trials have had problems with inadequate sample size[9] and have also failed to report whether they took measures to avoid bias.[10] Indeed Cochrane's observation that "when humans have to make observations there is always the possibility of bias"[11] was a lynchpin of the CONSORT (***Con***solidated ***S***tandards ***o***f ***R***eporting ***T***rials) initiative to improve the reporting, design, conduct, analysis, and interpretation of randomized controlled trials to inform decision-making in health care.[11,12] This initiative led to substantial improvements in the reporting and conduct of clinical trials.[13]

On the basis of the available evidence, it would now seem reasonable to suggest that preclinical testing in animal models of stroke, and indeed other models of disease, should adopt similar standards to ensure that decision-making is based on high quality unbiased data.[5,14] Adoption of such standards would have the added benefit of reducing wasteful use of financial and animal resources.

In general, studies should only be considered for publication if their Methods section includes a description of how they have addressed the standards below, or if authors make a cogent argument for why these standards are not relevant to their work. For these components of an article, citation of methods described in previous publications is not considered sufficient. These requirements should not preclude publication of important observational, pilot or hypothesis-generating data, but the conclusions of such studies should reflect their preliminary nature.

Animals

The precise species, strain, substrain and source of animals used should be stated. Where applicable (for instance in studies with genetically modified animals) the generation should also be given, as well as the details of the wild-type control group (for instance littermate, back cross etc).

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.

Inclusion and Exclusion Criteria

Where the severity of ischemia has to reach a certain threshold for inclusion (for instance a prespecified drop in perfusion detected with laser-Doppler flowmetry, or the development of neurological impairment of a given severity) this should be stated clearly. Usually, these criteria should be applied before the allocation to experimental groups. If a prespecified lesion size is required for inclusion this should be detailed, as well as the corresponding exclusion criteria.

Randomization

The manuscript should describe the method by which animals were allocated to experimental groups. If this allocation was by randomization, the method of randomization (coin toss, computer-generated randomization schedules) should be stated. Picking animals 'at random' from a cage is unlikely to provide adequate randomization. For comparisons between groups of genetically modified animals (transgenic, knock-out), the method of allocation to for instance sham operation or focal ischemia should be described.

Allocation Concealment

The method of allocation concealment should be described. Allocation is concealed if the investigator responsible for the induction, maintenance and reversal of ischemia and for decisions regarding the care of (including the early sacrifice of) experimental animals, has no knowledge of the experimental group to which an animal belongs. Allocation concealment might be achieved by having the experimental intervention administered by an independent investigator, or by having an independent investigator prepare drug individually and label it for each animal according to the randomization schedule as outlined above. These considerations also apply to comparisons between groups of genetically modified animals, and if phenotypic differences (eg, coat coloring) prevent allocation concealment this should be stated.

Downloaded from http://stroke.ahajournals.org/ by guest on January 1, 2017

Reporting of Animals Excluded From Analysis

All randomized animals (both overall and by treatment group) should be accounted for in the data presented. Some animals may, for very good reasons, be excluded from analysis, but the circumstances under which this exclusion will occur should be determined in advance, and any exclusion should occur without knowledge of the experimental group to which the animal belongs. The criteria for exclusion and the number of animals excluded should be reported.

Blinded Assessment of Outcome

The assessment of outcome is blinded if the investigator responsible for measuring infarct volume, for scoring neurobehavioural outcome or for determining any other outcome measures has no knowledge of the experimental group to which an animal belongs. The method of blinding the assessment of outcome should be described. Where phenotypic differences prevent the blinded assessment of for instance neurobehavioural outcome, this should be stated.

Reporting Potential Conflicts of Interest and Study Funding

Any relationship which could be perceived to introduce a potential conflict of interest, or the absence of such a relationship, should be disclosed in an acknowledgments section, along with information on study funding and for instance supply of drugs or of equipment.

We consider that these measures are of central importance to Good Laboratory Practice in the modeling of cerebral ischemia. Many groups already perform experiments to these high standards, and we hope that they will now report this in full, and that others follow their lead. Finally, we do not consider these requirements to represent a final or complete list of appropriate measures necessary to avoid bias. Future additions may be required as further evidence emerges and the experience of authors and reviewers evolves.

Disclosures

None.

References

1. STAIR. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*. 1999;30:2752–2758.
2. Grotta J. Why do all drugs work in animals but none in stroke patients? 2. Neuroprotective therapy. *J Intern Med*. 1995;237:89–94.
3. Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, Macleod M, Dirnagl U. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke*. 2008; 39:929–934.
4. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. *Ann Neurol*. 2006;59:467–477.
5. Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci*. 2007; 30:433–439.
6. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and metaanalysis of the efficacy of fk506 in experimental stroke. *J Cereb Blood Flow Metab*. 2005;25:713–721.
7. van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. *Brain*. 2007;130:3063–3074.
8. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke*. 2008;39: In press.
9. Weaver CS, Leonardi-Bee J, Bath-Hextall FJ, Bath PM. Sample size calculations in acute stroke trials: a systematic review of their reporting, characteristics, and relationship with outcome. *Stroke*. 2004;35:1216–1224.
10. Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke*. 1998;29: 2203–2210.
11. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T. The revised consort statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med*. 2001;134: 663–694.
12. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The consort statement. *JAMA*. 1996;276:637–639.
13. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the consort checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006;185:263–267.
14. Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab*. 2006;26:1465–1478.

Downloaded from http://stroke.ahajournals.org/ by guest on January 1, 2017

Stroke
Journal of the American Heart Association

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 and David W. Howells

Stroke. 2009;40:e50-e52; originally published online August 14, 2008;
doi: 10.1161/STROKEAHA.108.525386
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2008 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/40/3/e50

Downloaded from http://stroke.ahajournals.org/ by guest on January 1, 2017

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/