



# How can we improve the pre-clinical development of drugs for stroke?

Emily Sena<sup>1</sup>, H. Bart van der Worp<sup>2</sup>, David Howells<sup>3</sup> and Malcolm Macleod<sup>1</sup>

<sup>1</sup> Department of Clinical Neurosciences, University of Edinburgh, UK

<sup>2</sup> Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre, Utrecht, The Netherlands

<sup>3</sup> National Stroke Research Institute and University of Melbourne Department of Medicine, Melbourne, Australia

**The development of stroke drugs has been characterized by success in animal studies and subsequent failure in clinical trials. Animal studies might have overstated efficacy, or clinical trials might have understated efficacy; in either case we need to better understand the reasons for failure. Techniques borrowed from clinical trials have recently allowed the impact of publication and study-quality biases on published estimates of efficacy in animal experiments to be described. On the basis of these data, we propose minimum standards for the range and quality of pre-clinical animal data. We believe the adoption of these standards will lead to improved effectiveness and efficiency in the selection of drugs for clinical trials in stroke and in the design of those trials.**

Stroke is the most significant neurological disease in developed and developing countries alike [1]; it kills around 5 million people a year and leaves many more to cope with the long-term consequences of neurological impairment. More than 2400 years ago, Hippocrates wrote that ‘it is impossible to remove a strong attack of apoplexy (stroke), and difficult to remove a weak one’; until the last decade very little had changed. The drug-development process has been characterized by success in animal studies and subsequent failure in clinical trials. This contradiction might be explained if animal studies had overstated efficacy, if clinical trials had understated efficacy or if the animal studies did not model the pathophysiology of human stroke with sufficient fidelity to be useful. Understanding which of these are responsible for past failures in translation might inform future strategies for drug development and translation in stroke.

Systematic review and meta-analysis are techniques developed for the analysis of data from clinical trials and have recently been applied to animal data in experimental stroke. This approach provides an unbiased summary of data from separate animal studies of a given treatment strategy. From these data, the scale of any publication or study-quality bias (which might have led to animal studies overstating efficacy) can be established. By defining the limits to efficacy in animals (e.g. time to treatment, dose), we can identify whether the clinical trials tested drug use outside of these limits, which might have led to the clinical trials understating efficacy.

Here we show that study-quality and publication bias have substantial effects on published estimates of drug efficacy in animal studies. Developing qualitative and quantitative techniques to account for these sources of bias will provide better evidence on which the selection of drugs for clinical development in stroke can be based.

## The pathophysiological basis of stroke treatments

We now understand a great deal about the pathophysiology of focal cerebral ischaemia in animal models of stroke [2], and this has spurred the investigation of the potential therapeutic effects of candidate drugs that either block these processes or restore blood flow [3]. Therapeutic approaches include re-instating blood flow (reperfusion), inhibiting pathways promoting neuronal death or augmenting endogenous protective mechanisms (neuroprotection) and promoting plasticity, repair and regeneration.

Only three interventions have been convincingly demonstrated to improve outcome in patients with ischaemic stroke. Thrombolysis with tissue plasminogen activator (tPA) is highly effective if given within 3 h of stroke onset [4], but the majority of patients do not reach hospital quickly enough to receive this treatment. Aspirin given within 48 h improves outcome much less dramatically [5], probably by the very early prevention of recurrent stroke, but it is a simple treatment from which many more patients can benefit. Finally, management within a dedicated stroke unit has a substantial beneficial effect on outcome [6], but it is not clear which components of stroke-unit care convey this benefit. The development of

## Glossary

**Control group:** we use this term to describe the reported untreated or vehicle-treated comparison group. Our recommendations do not relate to other untreated or vehicle-treated comparison groups used, for instance, for model development in pilot studies. Of course, each experiment should include appropriate vehicle-treated controls if meaningful comparisons are to be made.

**Random allocation to experimental group:** at the start of the experimental treatment, each animal has an equal chance of being allocated to any experimental group. Allocating animals in sequence, on the basis of odd or even numbers or days of the week, or choosing animals “at random” does not represent true randomization and might lead to systematic biases whereby animals with certain characteristics are more likely to be allocated to particular experimental groups.

**Allocation concealment:** the scientist performing the experiment does not know which treatment the animal is being given.

**Blinded outcome assessment:** the scientists measuring the outcome, be it infarct volume or neurobehavioural score, do not know to which treatment group the animals belong.

Corresponding author: Macleod, M. (malcolm@apoptosis.freeserve.co.uk). Available online xxxxxx.

these treatments has at best been based only loosely on data from animal experiments.

By contrast, at least 494 drugs have been reported to have efficacy in animal models of stroke [3]. The abject failure to translate this efficacy from bench to bedside (an attrition rate of 99% at the stage of clinical trial alone) is substantially worse than the 89% overall industry attrition rate from first use in man to regulatory approval reported by Kola and Landis [7], and this difference raises important questions about the effectiveness and efficiency of conventional approaches to drug discovery in stroke [8]. In short, the animal experiments are falsely positive, the clinical trials are falsely negative or the animal studies do not model human disease with sufficient fidelity to provide a useful guide for translation.

### A scientific approach to understanding failures in translation

Both laboratory scientists and clinical trialists have recognized translational failure in stroke as an area of major concern, and several recommendations, based on expert opinion, have attempted to set out the circumstances in which translation is most likely to be successful. Foremost among these are the series of recommendations from the Stroke Therapy Academic Industry Round Table (STAIR) [9]. A compound (NXY-059) reported to meet all of their recommendations recently joined the long list of drugs found to be ineffective in a clinical trial [10]. However, a systematic review of the animal data supporting the efficacy of NXY-059 has yet to be conducted, so the true range and quality of the supporting evidence is not known. It might be that the supporting animal data are not as robust as was once thought.

Systematic review and meta-analysis are simple techniques developed to provide summary information by combining results from clinical trials. The systematic nature of the approach aims to capture all relevant data. This is important because neutral studies, if published at all, are more likely to be published in journals of low impact, in a language other than English or in abstract only. Non-systematic reviews are therefore likely to overstate biological effects. Meta-analysis is a straightforward technique that produces an overall estimate, based on a weighted average, of the effect seen across different studies. For any such analysis it is highly unlikely that included studies are identical in every respect, and the heterogeneity between studies (akin to lumping oranges with giraffes) limits the usefulness of the overall estimate of efficacy. However, this heterogeneity also brings some benefit in that its explanation (statistically, by grouping studies according to criteria of interest) allows conclusions to be drawn about differences in efficacy between groups.

### Publication bias

If positive (rather than neutral) studies are more likely to be published [11] (and some journals explicitly favour positive studies), then any conclusions drawn from the published literature will overstate the magnitude of any effect seen. In a group of studies reporting the same phenomenon, the presence of a significant publication bias can be inferred from the relationship between the precision

of individual studies and the sizes of the effects seen. In this way, it can be shown that the animal data reporting the efficacy of FK506 [12] and of tPA [13] in models of stroke are confounded by a significant publication bias; the 'true' efficacy is likely to be lower than that reported from these analyses.

### Evidence that animal studies might be falsely positive

In clinical trials, it is a universally acknowledged truth that certain aspects of study design can introduce bias to the results of those trials, and this bias usually leads to the overestimation of drug efficacy [14–17]. However, it is less widely accepted that the same truth might apply to experiments in the basic sciences. Although there is very little research on which aspects of study design are most likely to generate bias or the magnitude of the bias thus introduced, several study-quality checklists have been proposed (Table 1). These checklists appear to have some validity; in our review of the efficacy of FK506, for instance, we found that studies that scored highly on the ten-item checklist developed by the Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Stroke (CAMARADES) gave substantially lower estimates of efficacy than low-scoring studies (see Figure I in Box 1) [12]. The prevalence of studies that fulfill each of the CAMARADES quality items in six systematic reviews published to date is shown in Table 2.

Such checklists have included items relating first to the range of circumstances under which efficacy has been shown and second to the characteristics that might act as a source of bias in individual experiments. Because the first relates to characteristics of a body of evidence and the second to individual experiments, we recommend that they be considered separately and suggest applying the modified CAMARADES checklists shown in Table 1 before starting any large clinical trial.

### Randomization, allocation concealment and blinded outcome assessment

In spite of the recognized importance of these aspects of study design (see Glossary) in clinical trials, they are seldom reported in animal studies; one systematic review of animal studies of focal cerebral ischaemia found that only 42% of studies reported random allocation to experimental group, 22% reported allocation concealment and 40% reported the blinded assessment of outcome [18]. In data from six systematic reviews, we found proportions of 36%, 11% and 29%, respectively (Table 2).

These factors do appear to be important in experimental stroke studies; for instance, in a systematic review of hypothermia in experimental stroke, non-randomized studies overstated the reduction in infarct volume by 27% and studies without blinded outcome assessment overstated efficacy by 19% (see Figure I in Box 2) when they were compared to randomized and blinded studies, respectively [19]. Similar effects have been reported in other systematic reviews [12,13,20–22].

Although the genetic differences between experimental animals are small, particularly in inbred strains, these data suggest that randomization is indeed important. Furthermore, the experimental procedures for inducing

**Table 1. Checklists relating to the range of different circumstances under which efficacy should be shown and the quality characteristics of individual studies**

Range of Evidence	S	A	C	U	P
Assessment of functional outcome	N	N	N	N	Y
Assessment of histological outcome	N	N	N	N	Y
Recovery of sensorimotor function	Y	N	N	N	N
Recovery of cognitive function	Y	N	N	N	N
Replicated in two laboratories	Y	N	N	N	Y
Tested in models of permanent and temporary occlusion	Y	N	N	N	Y
Tested in males and females	Y	N	N	N	Y
Tested in animals with hypertension or diabetes	N	N	Y	Y	Y
Behavioural outcome measured for at least one month	Y	N	N	N	N
Efficacy in primates	Y	N	N	N	N
Clinically appropriate route of administration used	Y	N	N	N	Y
Dose–response relationship investigated	N	Y	N	N	Y
Optimal time window of treatment investigated	N	Y	N	Y <sup>a</sup>	Y
Assessment of at least two outcomes	N	Y	N	Y	N
Assessment in acute phase	N	Y	N	Y	Y
Assessment in chronic phase	N	Y	N	Y	Y
<b>Quality checklist for individual experiments</b>					
Monitoring of blood pressure and blood gases	N	Y	N	Y	Y
Publication in peer-reviewed journal	N	N	Y	N	Y
Control of temperature	N	N	Y	N	Y
Randomization of treatment or control	Y	Y	Y	Y	Y
Allocation concealment	Y	N	Y	Y	Y
Blinded assessment of outcome	Y	Y	Y	Y	Y
Avoidance of anaesthetics with marked intrinsic neuroprotective properties	N	N	Y	N	Y
Sample-size calculation before start of experiment	N	N	Y	Y	Y
Statement of compliance with regulatory requirements	N	N	Y	N	Y
Statement regarding possible conflict of interest	N	N	Y	N	Y
<b>Total number of items</b>	<b>11</b>	<b>8</b>	<b>10</b>	<b>10</b>	<b>21</b>

Abbreviations: S – Stroke Therapy Academic Industry Roundtable (STAIR) [30]; A – Amsterdam criteria [26]; C – Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Stroke (CAMARADES) [21]; U – Utrecht criteria [18]; P – Proposed (the checklists recommended here).

<sup>a</sup>The Utrecht criterion is that there be a ‘clinically relevant time window for start of treatment’.

ischaemia are inherently vulnerable to complications (such as failure to obtain sufficient ischaemia, perioperative hypotension and hypoxaemia), and knowledge of the experimental group might subtly affect the management of these complications and therefore affect the outcome.

### Sample size and its calculation

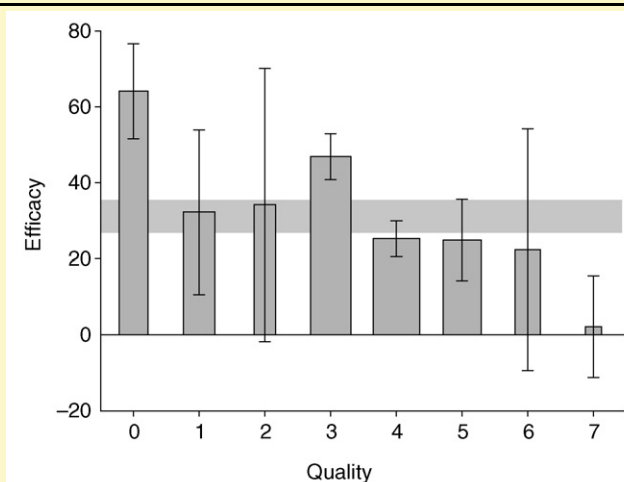
The size of an experiment is crucial – too small and the result will be imprecise, too large and the costs (both financial and in animal use) will be unnecessarily large. The likelihood of identifying a given difference between groups is related to the number of animals per group, the expected difference

### Box 1. The CAMARADES checklist

- Publication in peer-reviewed journal
- Statement of control of temperature
- Randomization of treatment or control
- Allocation concealment
- Blinded assessment of outcome
- Avoidance of anaesthetics with marked intrinsic properties
- Use of animals with hypertension or diabetes
- Sample-size calculation
- Statement of compliance with regulatory requirements
- Statement regarding possible conflict of interest

The CAMARADES ten-item study-quality checklist was derived partly from existing checklists regarding animal studies in experimental stroke, partly by inference from aspects known to be important for study quality in clinical trials and partly from knowledge of the potential synergistic effects of unintended hypothermia or of the anaesthetic used (principally a concern relating to the use of ketamine anaesthesia).

These items seem to have some validity; for instance, when publications reporting the efficacy of the macrolide immunosuppressant FK506 in experimental stroke are grouped according the number of study-quality items met, there is an inverse relationship such that high-quality studies give low estimates of efficacy, and vice versa (Figure 1).



**Figure 1.** Point estimates and 95% confidence intervals (CIs) of effect size for studies of FK506 in experimental stroke are grouped by study-quality score. The thickness of each bar reflects the number of comparisons contributing to that data point. The 95% CI for the global estimate is again shown as a grey band. Reproduced, with permission, from Journal of Cerebral Blood Flow and Metabolism.

**Table 2. Number and proportion of publications reporting individual components of the ten-item CAMARADES study-quality checklist in systematic reviews of nicotinamide [21], melatonin [20], FK506 [12], tirilazad [22], tPA [13] and hypothermia [19]**

	Nicotinamide		Melatonin		FK506		Tirilazad		tPA		Hypothermia		Overall	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Number of publications	14		13		29		18		113		101		288	
Publication in peer-reviewed journal	11	79	12	92	24	83	17	94	98	87	86	85	248	86
Statement of control of temperature	9	64	11	85	22	76	17	94	69	61	101	100	229	80
Randomization to treatment or control	3	21	4	31	6	21	12	67	42	37	36	36	103	36
Allocation concealment	2	14	2	15	1	3	1	6	23	20	4	4	33	11
Blinded assessment of outcome	3	21	4	31	2	7	13	72	24	21	38	38	84	29
Avoidance of anaesthetics with marked intrinsic properties	10	71	11	85	24	83	15	83	76	67	84	83	220	76
Use of animals with hypertension or diabetes	2	14	0	0	3	10	1	6	10	9	14	14	30	10
Sample-size calculation	0	0	0	0	0	0	0	0	8	7	0	0	8	3
Statement of compliance with regulatory requirements	10	71	10	77	15	52	11	61	69	61	48	48	163	57
Statement regarding possible conflict of interest	0	0	0	0	0	0	0	0	2	2	65	64	67	23

The colours indicate where the proportion of studies meeting that criteria are less than 25% (red), 25%–50% (pink), 50%–75% (light green) and more than 75% (green). Publications systematically perform poorly on randomization, allocation concealment, blinded assessment of outcome, use of animals with comorbidity, sample-size calculation and the statement of possible conflicts of interest.

between groups and the expected variance. Planning sample size is an important (although sometimes overlooked) routine aspect of study design in clinical trials. The reporting of a sample-size calculation also provides some assurance that sample size has not been increased incrementally in the light of ongoing analyses; such an approach substantially increases the risk of falsely concluding that an observed difference is real. However, in spite of these good reasons for reporting sample-size calculations in experimental stroke studies, only 3% of studies identified in systematic reviews made such reports (Table 2).

The biological similarities between individual experimental animals (compared to human subjects) mean that animal experiments can indeed be smaller than clinical trials. However, the difference in observed variance is modest (in animal studies, the difference is at best half of that seen in clinical trials) and does not justify the much smaller size of animal studies. Most animal studies are therefore substantially underpowered given the observed efficacy and variance.

One consequence of this is the prediction that many studies will falsely conclude that interventions are without efficacy, when in fact the study was simply underpowered. The resources (time, money and animals) used in those studies will therefore have been wasted. Importantly, these underpowered neutral studies do not feature highly in published work, although they must exist: this provides further evidence for publication bias [23]. In short, robust conclusions cannot be drawn from underpowered studies. The ethical and regulatory drive to minimize the size of individual animal experiments must therefore be tempered by a recognition that studies need to be large enough to provide useful information.

#### *Anaesthetics with intrinsic neuroprotective activity*

Most middle-cerebral-artery occlusion models require anaesthesia, and the choice of anaesthetic appears to

influence the efficacy of candidate drugs. Ketamine is a non-competitive antagonist at the phencyclidine binding site of the NMDA receptor, and ketamine anaesthesia enhances the neuroprotective efficacy of nicotinamide [24]. This was also found in a systematic review of nicotinamide [21] and reported in systematic reviews of FK506 [12], tirilazad [22] and tPA [13]. Similarly, the protective effect of hypothermia is enhanced by the use of phenobarbital or chloral hydrate anaesthesia [19]. By contrast, these reviews suggest that studies that use isoflurane systematically under-report efficacy when they are compared with those using halothane. We recommend that the use of anaesthetics that are known to have neuroprotective properties be avoided and that novel treatment strategies be tested under a range of anaesthetic conditions.

#### *Range of evidence - animals with comorbidity*

The majority of stroke patients are elderly, and comorbidities such as diabetes and hypertension are common in such patients. By contrast, in animal experiments drugs are routinely tested on animals that are both young and healthy. For six drugs reviewed systematically, to date only 10% of publications included the modelling of efficacy in animals with high blood pressure or diabetes, and no experiments reported efficacy in aged animals (Table 2); where efficacy was reported in the context of comorbidity, it was substantially lower.

#### *Limitations to these analyses*

Two important factors limit the analysis of the impact of study design on outcome. First, such assessments are based on reported study quality, yet such reports might be incomplete because the authors did not consider them to be relevant, either because of editorial policy or limitations on space. In addition, definitions of – for example – randomization and blinding might vary across studies. We consider that knowledge of important aspects of study

### Box 2. Randomization, blinding and the estimate of efficacy

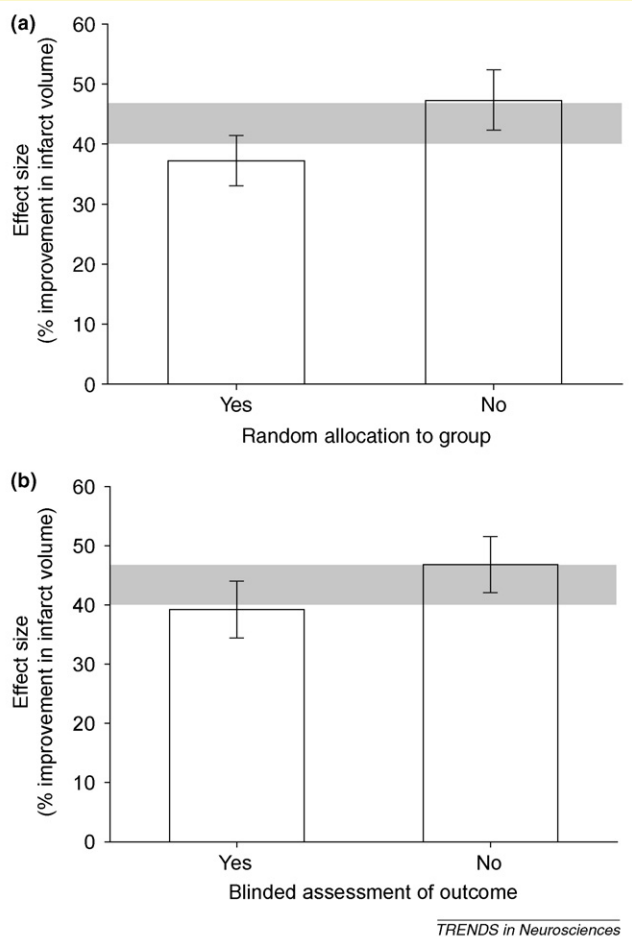
Bebarta and colleagues studied 290 abstracts that were presented at meetings of the Society for Academic Emergency Medicine between 1997 and 2001 and that described interventions in animal experiments (Table I). Overall, 87% of studies reported significant differences between experimental groups (i.e., they were 'positive'), and 13% of studies were neutral. Thirty-two per cent of studies reported randomization, and 11% reported the blinded assessment of outcome; these studies were much more likely to report neutral results than studies that did not meet these quality criteria.

#### Components of study quality

The frequencies with which individual quality items were scored over 288 publications identified in six systematic reviews are shown in Table 2. Which of these items are most important is the subject of continuing research, but univariate analysis for single interventions suggests an important role for randomization (Figure 1a) and for the blinded assessment of outcome (Figure 1b).

**Table I. Randomization, blinding and reported outcome**

	Randomization		Blinding	
	Yes	No	Yes	No
Neutral	22	16	9	29
Positive	72	180	22	230
% Neutral	23%	8%	29%	11%
Odds ratio (95% CI)	3.43 (1.71–6.91)		3.24 (1.36–7.72)	



**Figure 1.** Point estimates and 95% CI for studies of hypothermia in experimental stroke are grouped by (a) random allocation to group and (b) the blinded assessment of outcome. Grey bars represent the 95% CI of the overall estimate of efficacy. Non-randomized studies appear to overstate the improvement in infarct volume by 27%, and studies without blinded outcome assessment do so by 19%. Data from Ref. [19].

design is fundamental to the valid interpretation of reported findings and that studies should, as a minimum, report those characteristics outlined in Table 1.

Second, pharmaceutical companies might hold large amounts of proprietary data that are not available for inclusion in systematic reviews. There are concerns that the protection of proprietary information might lead to a situation in which others re-test a hypothesis that has already been clearly confirmed or refuted. The collegiate and collaborative approach that characterizes academic research at its best runs completely counter to the competitive, market-driven approach adopted by industry at its worst. However, academic research is often not at its best and industry-sponsored research is rarely at its worst: in both cases there should be an expectation (perhaps enforced by the regulatory agencies) that data will be made available within, say, five years of the experiment's conclusion.

More important is the question of whether there are systematic differences between proprietary and publicly available data. It might be, for instance, that a more 'managed' approach to experiments in industry leads to greater uniformity of approach and a lower prevalence of potential sources of bias. However, given the importance of positive findings both to the individual and to the organization, it seems unlikely that such sources of bias, where present, are any less important than in published work.

We believe there is an urgent need for a systematic analysis of proprietary data to establish whether there are differences in the prevalence and impact of study quality compared with that for published data. For instance, a company might be prepared to allow unrestricted access to all preclinical data describing the efficacy of a candidate stroke drug that is no longer in development because of neutral or negative clinical-trial data. Access to such data might provide important insights that would inform both academic research and industry.

#### Evidence that clinical trials might be falsely negative

tPA is one of only three interventions known to be effective for human stroke, so the animal data for tPA can be taken to represent a 'gold standard' against which the animal data for other drugs can be judged. These animal data are characterized first by their quantity – more than 100 studies involving more than 3000 animals – and second by the observation that tPA was tested in a clinical context [4] under conditions (particularly the interval between stroke onset and the initiation of treatment) for which there was robust evidence for efficacy in animals [13]. This contrasts with the clinical trials of the 21-aminosteroid tirilazad, in which patients were recruited up to 24 h after stroke onset [25]; the median time to treatment in animal studies was 10 min [22].

For some drugs, animal data were not available (at least not in the published literature) when clinical trials were initiated. A systematic review of nimodipine in experimental stroke [26] showed that the earliest animal data were published four years after the publication of the earliest clinical-trial data. The failure of nimodipine to improve clinical outcome in stroke trials [27] cannot therefore be considered a failure of animal models of stroke. In

fact, meta-analysis showed concordance between the clinical-trial and animal-model data, in that nimodipine did not improve the outcome in animal models of stroke.

### Implications for experimental scientists

We believe there have been substantial improvements in the conduct and regulation of animal experiments over the past 50 years. These improvements have occurred through the desire of experimental scientists to continually improve the relevance, generalizability and precision of their work. The current focus on study quality should be seen in the context of this continuing quality improvement.

Certain aspects of study design in experimental stroke – particularly randomisation, allocation concealment and the blinded assessment of outcome – seem to influence how effective a drug appears to be. Although there is no direct evidence of a causal relationship, it seems likely that this is indeed the case. The relative importance of the various possible sources of bias is not yet known and is the subject of ongoing research; clearly, this is crucial information for scientists conducting such experiments. In the meantime, we recommend that every effort is made to randomize treatment allocation, conceal treatment allocation, blind the assessment of outcome, pre-specify study size based on a formal sample-size calculation and ensure full publication of neutral or negative data wherever possible.

Furthermore, we propose that the STAIR group might usefully revise their recommendations for pre-clinical drug development in stroke to include recommendations on (i) the range of circumstances under which efficacy should be demonstrated and (ii) the study quality required of individual experiments.

### Are animal experiments in stroke different?

It has been possible to measure the impact of potential sources of bias in preclinical studies of drug efficacy in experimental stroke because of the large amount of available data measuring outcome in a broadly similar way (i.e., as infarct volume). Given that the fundamental experimental approach is similar for studies exploring, for instance, stroke pathophysiology or stroke in transgenic animals, it seems highly likely that these experiments will also be susceptible to similar biases. Indeed, experimental approaches are similar across many of the life sciences, and it would be surprising if these sources of bias were not represented more broadly.

We have measured the reporting frequency of six key aspects of study quality in transgenic and pathophysiological experiments in stroke and in experiments modelling Parkinson's disease and multiple sclerosis. The reporting of study-quality attributes in these areas is at best as limited as it is in preclinical drug testing in stroke (Table 3). Although the presence of potential sources of bias in no way confirms the presence of bias itself, it seems prudent to recognize that this might be the case. Future research might usefully explore the impact of aspects of study quality and design in such experiments and whether measures might be taken to minimize the impact of such bias.

These issues of study quality are of particular relevance to the funding of research and to peer review and publication, and it might be that agreement among funders, regulatory agencies and journals on minimum quality standards (similar to the CONSORT (CONsolidated Standards Of Reporting Trials) statement for clinical trials [28]) might find favour.

### Implications for drug development in stroke

Those systematic reviews conducted to date have several implications for the successful translation of efficacy from bench to bedside. First, for most candidate neuroprotective drugs, a precise statement of efficacy and the limits to efficacy is not possible because too few experiments have been performed across too narrow a range of circumstances. Second, the quality of individual studies should be considered so that a judgement about the reliability of the contributing data can be made. Third, the possible presence of a publication bias should be explored. Finally, the circumstances in which efficacy is seen in animals should be a crucial factor in the design of human clinical trials. Indeed, a recent editorial in the British Medical Journal called for systematic review and meta-analysis to be an integral part of the drug-development process [29].

We have recently completed the first systematic review and meta-analysis of a candidate stroke treatment (hypothermia) conducted with the explicit purpose of informing the design of a human clinical trial [19]. That review has proved very helpful in guiding the target depth and time to initiation of hypothermia, and as a result it will be possible to design a clinical trial for testing the efficacy of hypothermia in humans under conditions in which efficacy has already been robustly demonstrated in animals.

**Table 3. Prevalence of selected quality characteristics in other experimental models**

	Number of publications	Randomisation (%)	Blinded assessment of outcome (%)	Sample-size calculation (%)	Statement of possible conflict of interest (%)
Transgenic stroke studies	157	n/a	3	0	2
Stroke pathophysiology studies	166	5	18	0	8
Parkinson's disease	118	12	15	0	1
Multiple sclerosis	183	2	11	0	5

Percentage of studies identified in systematic reviews of focal cerebral ischaemia in transgenic animals (Chan, L. (2007) Sources of bias in transgenic studies of stroke pathophysiology. CAMARADES Monograph No. 3 edn, available at [http://www.camarades.info/index\\_files/CAMARADES%20Monograph%20No.%203.pdf](http://www.camarades.info/index_files/CAMARADES%20Monograph%20No.%203.pdf)); of the efficacy of MK801 in cell culture, tissue slice, focal ischaemic and intrastrial excitotoxicity models (Wheble, P. (2007) A systematic approach to research synthesis for the pathophysiology of stroke. CAMARADES Monograph No. 7, available at [http://www.camarades.info/index\\_files/Page332.htm](http://www.camarades.info/index_files/Page332.htm)); of dopamine agonists in animal models of Parkinson's disease (Ferguson, E.D.M. (2007) Dopamine receptor agonist treatment of Parkinson's disease: a systematic review and meta-analysis. CAMARADES Monograph No. 6, available at [http://www.camarades.info/index\\_files/Page332.htm](http://www.camarades.info/index_files/Page332.htm)); and of proteins and peptides (Vesterinen, H. (2007) A systematic review and meta-analysis of the efficacy of proteins and peptides used in the treatment of experimental allergic encephalomyelitis. CAMARADES Monograph No. 4, available at [http://www.camarades.info/index\\_files/Page332.htm](http://www.camarades.info/index_files/Page332.htm)) and the role of Th17 cells (Hammar, M.L. (2007) Systematic review and meta-analysis on the role of Th17 cells in the pathophysiology of Experimental Allergic Encephalomyelitis. CAMARADES Monograph No. 5, available at [http://www.camarades.info/index\\_files/Page332.htm](http://www.camarades.info/index_files/Page332.htm)) in experimental allergic encephalomyelitis.

## Conclusions

Reports of the efficacy of candidate neuroprotective drugs in animal models of stroke are profoundly biased by aspects of study design. Conclusions drawn from individual publications or from narrative reviews cannot provide the basis for selecting drugs for clinical trial or for the design of those clinical trials. A sound judgement on efficacy, the limits to efficacy, the need for any further animal experiments and the design of any ensuing clinical trial can only be made on the basis of a systematic analysis of all available animal data; such an analysis must include the possible contribution of publication and study-quality bias to the observed efficacy.

Efforts to minimize these sources of bias are an important priority for those conducting, funding, publishing and interpreting such experiments, and we and others are currently seeking to identify which aspects of study quality are of greatest impact and therefore represent priority areas for change\*. It seems highly likely that animal models of other neurological diseases will be susceptible to similar sources of bias and that a similar scientific approach to identifying the presence and impact of aspects of study design might lead to improvements in translational efficiency in these diseases.

## Acknowledgements

We are grateful to Peter Sandercock and Geoff Donnan for helpful discussions. The work described in this article has been supported by grants from the Scottish Chief Scientist Office, Mrs Dale's Neurology Fund, and the Australian NHMRC (National Health and Medical Research Council).

## References

- Lopez, A.D. *et al.* (2006) Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 367, 1747–1757
- Dirnagl, U. *et al.* (1999) Pathobiology of ischaemic stroke: an integrated view. *Trends Neurosci.* 22, 391–397
- O'Collins, V.E. *et al.* (2006) 1,026 experimental treatments in acute stroke. *Ann. Neurol.* 59, 467–477
- Hacke, W. *et al.* (2004) Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet* 363, 768–774
- Chen, Z.M. *et al.* (2000) Indications for early aspirin use in acute ischemic stroke: A combined analysis of 40 000 randomized patients from the chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups. *Stroke* 31, 1240–1249
- Langhorne, P. *et al.* (1993) Do stroke units save lives? *Lancet* 342, 395–398
- Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715
- Dirnagl, U. (2006) Bench to bedside: the quest for quality in experimental stroke research. *J. Cereb. Blood Flow Metab.* 26, 1465–1478
- Stroke Therapy Academic Industry Roundtable (1999) Recommendations for Standards Regarding Preclinical Neuroprotective and Restorative Drug Development. *Stroke* 30, 2752–2758
- [No authors listed] (2006) The bitterest pill. *Nature* 444, 532–533.
- Easterbrook, P.J. *et al.* (1991) Publication bias in clinical research. *Lancet* 337, 867–872
- Macleod, M.R. *et al.* (2005) Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. *J. Cereb. Blood Flow Metab.* 25, 713–721
- Perel, P. *et al.* (2007) Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ* 334, 197–199
- Schulz, K.F. and Grimes, D.A. (2005) Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 365, 1348–1353
- Schulz, K.F. and Grimes, D.A. (2002) Blinding in randomised trials: hiding who got what. *Lancet* 359, 696–700
- Schulz, K.F. *et al.* (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J. Am. Med. Assoc.* 273, 408–412
- Miettinen, O.S. (1983) The need for randomization in the study of intended effects. *Stat. Med.* 2, 267–271
- van der Worp, H.B. *et al.* (2005) Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. *J. Neurol.* 252, 1108–1114
- van der Worp, H.B. *et al.* (2007) Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. *Brain*. DOI: 10.1093/brain/awm083
- Macleod, M.R. *et al.* (2005) Systematic review and meta-analysis of the efficacy of melatonin in experimental stroke. *J. Pineal Res.* 38, 35–41
- Macleod, M.R. *et al.* (2004) Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* 35, 1203–1208
- Sena, E. *et al.* (2007) Systematic review and meta-analysis of the efficacy of tirilazad in experimental stroke. *Stroke* 38, 388–394
- Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychol. Bull.* 86, 638–641
- Chang, M.L. *et al.* (2002) Nicotinamide and ketamine reduce infarct volume and DNA fragmentation in rats after brain ischemia and reperfusion. *Neurosci. Lett.* 322, 137–140
- Tirilazad International Steering Committee (2000) Tirilazad mesylate in acute ischemic stroke: A systematic review. *Stroke* 31, 2257–2265
- Horn, J. *et al.* (2001) Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. *Stroke* 32, 2433–2438
- Horn, J. and Limburg, M. (2001) Calcium antagonists for ischemic stroke: a systematic review. *Stroke* 32, 570–576
- Moher, D. *et al.* (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357, 1191–1194
- Hackam, D.G. (2007) Translating animal research into clinical benefit. *BMJ* 334, 163–164
- [No authors listed], (1999) Recommendations for Standards Regarding Preclinical Neuroprotective and Restorative Drug Development. *Stroke* 30, 2752–2758.

\* A one day symposium addressing these issues – ‘Bench to bedside in acute stroke: Finding our way’ or ‘Lost in translation?’, organised by the British Neuroscience Association – will take place at the Royal Society of Edinburgh on 17th October 2007. Details from [events@bna.org.uk](mailto:events@bna.org.uk).