Chapter 2

How to Avoid Bumping into the Translational Roadblock

Malcolm Macleod

Abstract

Translating neuroprotective efficacy from animal studies to clinical trials in humans has been fraught with difficulty. This failure might be because animal studies were falsely positive or clinical trials were falsely negative.

Here, I focus on the measures to improve the design, conduct, and reporting of animal studies to maximize both their internal and external validity. These include, but are not limited to, randomization, allocation concealment, blinded assessment of outcome, sample size calculation, and measures to avoid publication bias. In addition, I give a brief introduction to the systematic review and meta-analysis of data from animal experiments.

Key words: Systematic review, Meta-analysis, Publication bias, Study quality, Validity, Experimental design, Randomization, Allocation concealment, Blinded assessment of outcome, Sample size calculation

1. Introduction

> "…you will meet with several observations and experiments which, though communicated for true by candid authors or undistrusted eye-witnesses, or perhaps recommended by your own experience, may, upon further trial, disappoint your expectation, either not at all succeeding, or at least varying much from what you expected" (Robert Boyle 1693, *Concerning the Unsuccessfulness of Experiments*)

Valid experiments are those which give appropriate descriptions of some biological truth in a system being studied. *Internal validity* relates to the extent to which an experiment accurately describes what has happened in that model system. *External validity* relates to the extent to which the results from that model system can be generalized to predict what might happen, for instance, in a group of patients with the disease being modeled in response to the drug being tested.

Ulrich Dirnagl (ed.), *Rodent Models of Stroke*, Neuromethods, vol. 47,
DOI 10.1007/978-1-60761-750-1_2,

Stroke is one condition where, despite substantial efforts in the neuroscience community, translation of efficacy to humans has proved exceptionally difficult (1). The modeling of stroke in rodents usually involves measuring some outcome – be it a change in gene expression, an increase in protein phosphorylation, a volume of cerebral infarction or post stroke behavior – and it may also involve determining a change caused by an experimental intervention in that outcome. For our purposes, here we concentrate on experiments testing the efficacy of candidate stroke drugs, but the same considerations apply to all experiments where any outcome is measured.

1.1. The Validity of Individual Experiments

Following treatment, infarct volume and neurobehavioral score may improve, worsen, or be unchanged. We hope that results from our experiments will reflect "biological truth", but of course this is not always the case. Measurement error and biological variability mean that our sample of animals can never describe completely what the outcome would be across all animals. These are random errors, and while they reduce the precision of the estimate of biological effect, they are as likely to underestimate effects as they are to overestimate effects. The likely scale of this random error can be estimated in preliminary experiments, and we can then predict how closely the results of an experiment of given size reflect the population response. This allows us to estimate how large an experiment should be to have a reasonable (specified) probability of detecting a treatment effect of a given size.

There are also, however, sources of nonrandom error (bias), which cause the estimate of effect to be consistently understated or, more usually in this context, overstated. These sources of bias include selection bias, performance bias, ascertainment bias, and attrition bias (see also Chap. 19).

1.1.1. Selection Bias

The only difference between experimental groups should be the different treatments they receive. If there are consistent patterns in the way animals are allocated to groups, this may lead to additional, unwanted differences which might be responsible for any differences observed in the outcome (2, 3). For instance, if there is a standard allocation sequence, this may confound interpretation because the first animal from the cage is likely to be less anxious than the later animals. Furthermore, while it is expected that the effectiveness of any intervention may be modulated by known factors (drug dose, time to treatment, comorbidities), effectiveness may also be altered by unknown or unidentified factors. The resulting selection biases can only be avoided if treatment group allocation is truly an explicitly random process, with each animal having an equal chance of entering any of the experimental groups.

1.1.2. Performance Bias

Knowledge of treatment group allocation might alter the way in which ischemia is induced or how the animal is managed in the recovery phase, including decisions about whether and when animals should be euthanized. These differences are probably due to subconscious and inadvertent changes in the behavior or practice of investigators, and have been shown to be a potentially powerful source of bias (2, 3). They can be avoided if the person conducting the experiment is unaware of the treatment group allocation.

1.1.3. Ascertainment Bias

Knowledge of treatment group may also confound the measurement of outcome. It is only human nature that investigators wish their experiments to "work", and that they should desire that the intervention being tested should be found to have definite effects (negative or positive) rather than being neutral. This is partly because of the investment of time and energy in mounting the experiments; partly because the currency of science is publication and neutral studies are much less likely to be published; partly because the continuous funding of a project may depend on "positive" results and partly because labs which are able to reliably report "positive" findings may be more likely to attract commercial funding.

1.1.4. Attrition Bias

Rodent stroke models rarely report the number of animals excluded from the final analysis. Animals might be excluded for a number of reasons, including a judgment that cerebral blood flow was not sufficiently reduced; that the induced neurobehavioral deficit was not sufficiently severe; or the animal may die or be euthanized prior to the assessment of the outcome which is being reported.

1.2. The Validity of Research Summaries

The decision to take a compound with promising animal data forward to clinical trial is usually based on a summary of the available animal data. The process of producing such research summaries is itself subject to the sources of bias that may present an overoptimistic view of the prospects for success. These biases include publication bias, false-positive bias, and selective citation bias.

1.2.1. Publication Bias

Where experiments are performed but remain unpublished, they are not usually available for inclusion in research summaries. If these unpublished experiments differ substantially in their outcomes from published work, conclusions drawn from published experiments may not reflect adequately the efficacy of the intervention being summarized. Indeed, analysis of data derived from systematic reviews of stroke studies curated on the CAMARADES database suggests that at least 15% of experiments remain

unpublished, and that this results – even following rigorous systematic review – in an overstatement of efficacy of 30% (7).

1.2.2. False Positive Bias

Where most studies are underpowered (as is the case for most reported rodent stroke experiments), the value of positive findings is devalued. The positive predictive value of a study in a given field – setting aside for the moment the issues of internal validity outlined above – is determined not just by the Type I error α, but also by the Type II error β and the true proportion of positive studies (4). Where roughly half of all studies are positive, where the power is roughly 20%, and where the Type I error is 5% (as is the case with rodent models of ischemia), the predictive value of positive studies is only 67%. That is, only two out of three statistically "positive" studies are truly positive.

1.2.3. Selective Citation Bias

Stroke patients are not rodents, and any assumptions of the translation of efficacy from rodents to man requires a leap of faith. However, it is important to systematically explore the limits to efficacy in animals and to make an informed judgment as to whether these are likely to affect efficacy adversely in man. Selective citation of individual studies supporting efficacy at low doses or late time points or in hypertensive animals are no substitute for a systematic summary of all experiments reporting efficacy at low doses or late time points or in hypertensive animals. Clinical trials generally require the recruitment of large numbers of patients, and industry usually seeks evidence for efficacy in as wide a potential market as possible. Because of these complimentary pressures, clinical trials have often recruited patients at extended time windows or with substantial comorbidities (typical of stroke patients) or where drug concentration at the site of their action is a fraction of that obtained in animal models. It is little surprise that so many of these trials in stroke have been neutral.

2. Materials

The implementation of Good Laboratory Practice in the context of middle cerebral artery occlusion requires no specific materials other than foresight, planning, and a computer with internet access (see also Chaps. 17, 18, 19).

3. Methods

3.1. Before the Experiment …

3.1.1. Systematic Review of Existing Data

It is probably unethical and certainly inefficient to unknowingly repeat work that is previously conducted by other groups explicitly unless for the purposes of replication, pilot studies or as a positive control. You should therefore conduct a brief systematic review of the existing published work to determine whether your hypothesis has been tested previously and what additional value your experiments can contribute. A simple approach to systematic review is given later.

3.1.2. Pilot Studies

Before conducting the planned experiments, it is important to show that you can reproducibly induce an infarct of the desired severity and behavioral deficit and to demonstrate that you can show a treatment effect where one is known beyond doubt to exist. Hypothermia is commonly used as "positive control". This approach will also provide pilot data on the variance observed in the model in your own hands that are important for sample size calculations (see below).

3.1.3. Sample Size Calculation

An assessment should be made of the size of the effect which the experiment should be able to detect. This is a matter of judgment; in some cases a 10% difference in outcome might be considered biologically significant, while other experiments may set out to detect only much larger differences. For a simple two-group comparison, sample size calculations are reasonably straightforward and are enabled in most standard statistical packages. For more complex studies where statistical testing uses the analysis of variance, specific online tools are available (see (5) for an excellent online resource). Importantly, the estimate of variance used in these sample size calculations should be derived from your own pilot experiments rather than the lowest variance achieved in the best hands in your group. The sample size calculation can also make allowances for the proportion of animals which it is expected will be excluded from the final analysis (see also Chap. 18).

3.1.4. Inclusion and Exclusion Criteria

Before the experiment starts, you should also establish clear and unambiguous rules about any inclusion criteria which will be applied, for instance, a prespecified decrease in perfusion detected with laser–Doppler flowmetry, or the development of neurologic impairment of a given severity. You should also establish rules about the animals which will be excluded from the analysis (for instance, those which die spontaneously or are euthanized prior to the completion of the experiment). Crucially, these rules should be applied without knowledge of treatment group allocation.

Where the intervention being tested is delivered after the induction of ischemia, it is best if randomization (see below, and Chap. 17) occurs after the application of the inclusion criteria and is restricted to the pool of included animals.

3.2. During the Experiment …

3.2.1. Randomization

Animals should be allocated to an experimental group by randomization. Experience from clinical trials suggests that manual methods of random allocation (coin toss, allocation in sealed envelopes) are open to subversion by enthusiastic researchers, and so computerized methods are preferred. Random number tables or computerized random number generators (for instance the MS Excel command RAND()) allow a treatment allocation sequence to be generated, but it is important that the investigator have no prior knowledge of the allocation sequence (see blinding below). Alternatively, web-based randomization schedules are available at http://www.graphpad.com/quickcalcs/randomize1.cfm or http://www.randomization.com/. When using these, it is probably best to specify a blocked randomization design with block size set at twice the number of experimental groups (so a four-group experiment would have a block size of 8).

3.2.2. Allocation Concealment

As far as is possible, the investigator who is responsible for the induction, maintenance, and reversal of ischemia and for decisions regarding the care of (including the early killing of) experimental animals should be unaware of the experimental group to which an animal belongs. For instance, you could have a colleague prepare and administer a drug or intervention, or he could relabel or aliquot a drug which you have prepared. If this is done, it is better if the labeling is in the form of a reference number unique to each animal rather than simply relabeling A, B, C, etc. That is, it is better if the investigator does not know which animals belong to the same group, as they are likely to try to "guess" which group is which. This is much more difficult if group allocation is not known. Similarly, the assessment of outcome should be carried out by someone who is not aware of the treatment group allocation.

3.3. After the Experiment …

3.3.1. Was Your Study Big Enough?

If large numbers of animals were excluded from the study, or if the intervention was less effective than you had thought, it might be that your data will show the suggestion of an effect which is not statistically significant. Properly conducted pilot studies and sample size calculations based on reasonable estimates of efficacy will reduce this risk but will not remove it completely. If this becomes apparent after data have been analyzed, there is little that can be done except to repeat the experiment with adequate group sizes. However, before the data are analyzed, it is reasonable to add further animals to make up for exclusions as long as this can be done in a randomized fashion and with allocation concealment.

It is also possible to ask a colleague (usually a statistician) to examine unblinded data (sort the outcome by animal reference number and ask your colleague to send the treatment grouping for the concerned data, although not necessarily the treatment associated with each grouping) to carry out a preliminary analysis. The purpose of this would be to ask them to make recommendations as to whether (1) there are significant differences between groups; (2) there may be a difference, and given the estimated effect size and variance, a further *x* animals per group would be required to give power, for example 80%; or (3) the differences between groups are such that the number of animals required makes further experiments futile. Such an approach may also be used to conduct planned interim analyses, but if such multiple analyses are planned, the statistical methodology should take account of this.

3.3.2. Reporting Your Study

Your manuscript should describe how your work meets the principles outlined above, and specifically it should include the details of the number of animals excluded and the reason for their exclusion. It should be set out according to the principles outlined in the international consensus statement on Good Laboratory Practice in the modeling of focal cerebral ischemia (6).

Most academic institutions, funding organizations, and regulatory bodies require a commitment to publication and data sharing, and you should make every effort to have your work published. Indeed, this is a component of the Draft Amendment to EU directive 86/609 on the approximation of laws, regulations, and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Work should be published within 3 years of being completed, and if this is not possible, then details should be posted on Institutional websites. Outline details of unpublished work can also be submitted to CAMARADES (the Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies), where they will be made available to those wishing to conduct systematic reviews.

3.4. Increasing the Validity of Research Summaries

Systematic identification of relevant information: Narrative reviews are based on a selection of the available literature, often with the intention of supporting a particular point of view. Neutral studies are often published in journals of low impact; in languages other than English and in Conference Abstracts rather than full publications. As such they may not be identified in superficial PubMed searches. If you are really trying to identify all relevant experiments in the public domain which might have been conducted using your compound, you should search the literature systematically. The availability of online datasets and conference abstracts has made this a simpler, faster process.

Where to search: PubMed (www.pubmed.com), ISI Web of Science (http://apps.isiknowledge.com) and EMBASE (www.embase.com). Access to PubMed is free, and most institutions purchase access to ISI WoS and EMBASE. For PubMed, the search can be limited to publications describing animal experiments. For ISI WoS, ensure that you have checked to search Conference Proceedings as well as the Science Citation Index.

How to search: Firstly, you should define the characteristics of publications which you wish to include – for instance "studies reporting the efficacy of *compound* in animal models of focal cerebral ischemia, where the outcome is expressed as a change in infarct size or neurobehavioral score". Your search strategy should be designed to create two sets – firstly a set of all publications relating to your compound (using all known variations in the compound name), and secondly a set of publications relating to animal modeling of focal cerebral ischemia (we use {(stroke) OR (ischemia) OR (cerebrovascular) OR (middle cerebral artery) OR (MCA) OR (ACA) OR (anterior cerebral artery) OR (MCAO)}). Publications common to these sets are then identified and downloaded to your preferred reference management software.

Selecting studies for inclusion: The next step is to determine which publications meet your inclusion criteria, and to exclude the rest. This process can be time consuming, tedious, and is subject to error. These errors can be minimized if the task is carried out independently by two investigators, with disagreements resolved by discussion. The full text of the remaining articles should then be retrieved.

3.5. Special Considerations in Creating a Research Synthesis

Qualitative summarizing of identified publications is straightforward. It may also be possible to carry out a quantitative summary (using meta-analysis), perhaps stratified for various characteristics of interest.

3.5.1. The Range of Circumstances Under Which Efficacy Has Been Tested

Efficacy at early time points and in healthy young animals may not be seen at the later (clinically more relevant) time points or in animals with (clinically relevant) comorbidities. Establishing efficacy under these circumstances is a crucial prelude to clinical trials. Similarly, it is important to demonstrate efficacy at concentrations obtained in human brain (if this is known). As a first step, simply tabulating the details of included studies can offer substantial insights.

3.5.2. Study Quality

Given the known impact of the sources of bias described above, it is reasonable to establish the extent to which identified publications may be susceptible to bias. Key issues are randomization, allocation concealment, blinded assessment of outcome, and statement of a sample size calculation (see also Chaps. 17 and 18).

3.5.3. Estimates of Efficacy Under Various Circumstances

If there are sufficient data, it may be possible to explore the impact of, for instance, dose, or time, or comorbidity, using normalized mean difference meta-analysis. A detailed understanding of statistics is not required, as online tools are available. The CAMARADES group has developed online tools that allow investigators to enter the details of publications and reported outcomes, which produce summary information or the range and quality of evidence, and which will perform normalized mean difference meta-analysis. Investigators wishing to use these tools should register at www.camarades.info. In large datasets, it is possible to explore the possible presence of publication bias using funnel plotting, Egger regression or "Trim and Fill".

4. Notes

Excellent is the enemy of good, and very few investigators adhere to all the considerations described above. We all try to do the best we can, and as we seek continually to improve the quality of our animal experiments, we should also be candid in where we have fallen short of these standards, and how this might introduce bias to our conclusions.

References

1. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW (2006) 1, 026 experimental treatments in acute stroke. Ann Neurol 59:467–477
2. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA (2008) Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39:2824–2829
3. van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR (2007) Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain 130:3063–3074
4. Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2: e124
5. Lenth RV (2006–2009) Java Applets for Power and Sample Size [Computer software]. http://www.stat.uiowa.edu/~rlenth/Power. Accessed 31 March 2009
6. Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, Buchan A, van der Worp HB, Traystman R, Minematsu K, Donnan GA, Howells DW (2009) Good laboratory practice: preventing introduction of bias at the bench. Stroke 40: e50–e52
7. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. "PLOS_Biology_PubBias.pdf" PLoS Biol. 8(3):e1000344

http://www.springer.com/978-1-60761-749-5

Rodent Models of Stroke
Dirnagl, U. (Ed.)
2010, XII, 285 p. 49 illus., Hardcover
ISBN: 978-1-60761-749-5
A product of Humana Press