Editorial

Preclinical studies of human disease: Time to take methodological quality seriously

Over the past three decades, mortality and morbidity associated with myocardial infarction have substantially reduced, mainly through the combination of early recanalization of occluded coronary arteries and antithrombotic therapy [1]. However, a wide range of animal studies has suggested that restoring blood flow may paradoxically also cause damage to the myocardium, a process that has been termed 'reperfusion injury.' In animal studies modelling reperfusion injury many treatment strategies have been shown to reduce reperfusion injury and to limit infarct size [2,3]. Unfortunately, the benefit of these strategies could not be replicated in the clinical trials that followed, casting doubt on the usefulness of these animal models of myocardial ischemia and reperfusion.

The disappointing translation of the results of animal studies to clinical trials is not unique to the field of myocardial reperfusion injury. In a review of animal studies published in 7 leading scientific journals (identified as such by their impact factor), only about a third of the studies translated at the level of human randomized trials, and just one tenth of the interventions were subsequently approved for use in patients [4]. This success rate of 10% may still be viewed as a success rather than as a failure, but the median citation count of these studies was 889, and in all likelihood less frequently cited animal research has a much smaller chance of translation to the clinic. This is illustrated by the fact that in animal models of acute ischemic stroke, over 500 treatments have been reported to improve outcome, but only aspirin and very early intravenous thrombolysis with alteplase have proved effective in patients, despite numerous clinical trials of other treatment strategies [5,6].

Reasons for translational failure have been discussed in several reviews and opinion articles [2,3,7–9]. This failure may in part be explained by shortcomings of clinical trials, where the trial design may not acknowledge for instance limitations of efficacy at longer delays to treatment seen in animal studies [9]. An alternative explanation for the disparity between the results of animal models and those of clinical trials is that the first may not reflect the disease in humans with sufficient fidelity. Most animal studies of myocardial reperfusion injury have been performed in small rodents even although it is known that the spatial and temporal development of myocardial infarction in these animals differs from that in humans. In addition, for reasons of costs and simplicity, most experiments have performed in very young and healthy rodents, whereas most patients with myocardial infarction are older and many have comorbid conditions such as hypertension or diabetes mellitus. Treatment of the animals has often been started before or at the onset of ischemia, which is not feasible in the clinic. Another remarkable finding is that most preclinical studies have been restricted to male animals, thereby limiting their generalizability to female patients. Finally, animal models of myocardial ischemia and reperfusion often differ from the clinical situation with regard to the mode of coronary occlusion and reopening, the duration of ischemia, and the duration of reperfusion [2,3,8].

Methodological shortcomings of animal studies are probably an equally important cause of translational failure, but have received much less attention in the cardiovascular literature. Components that are considered essential to the design of most clinical trials, such as randomisation, blinding, and sample size calculation, appear to be much less prevalent in animal research [9]. The presence and consequences of methodological flaws in animal studies have been tested most extensively in studies of neurological conditions, and of ischemic stroke in particular, probably because with this disease the gap between the laboratory and the clinic is both very large and well recognized. In systematic reviews of different interventions tested in animal models of acute ischemic stroke, other emergencies, Parkinson's disease, multiple sclerosis, or amyotrophic lateral sclerosis, generally about a third or less of the studies reported random allocation to the treatment group and even fewer studies reported concealment of treatment allocation or blinded outcome assessment [9–11]. A priori sample size calculations were reported in 0 to 3% of the studies. In one review of several treatment strategies for acute ischemic stroke, only one of 45 studies mentioned predefined inclusion and exclusion criteria, and in just 12 articles (27%) exclusion of animals from analysis was mentioned and substantiated; it is hardly plausible that in each of the other studies every single experiment went as smoothly as the investigators had planned [12]. Similar observations were made in a survey of 271 publications reporting research on live rats, mice, and non-human primates for various medical conditions, carried out in US and UK publicly funded institutions [13].

It appears self-evident that the quality of the design of an animal experiment will affect its scientific validity, but this has received little attention as a fruitful area for research in translational medicine. That evidence which does exist suggests that these issues are indeed crucial. In animal studies testing interventions in emergency medicine the odds of a positive result were more than three times as large if the publication did not report randomization or blinding as compared with publications that did report these methods [14]. In systematic reviews of treatments for acute ischemic stroke, the benefit of treatment was larger with lower study quality [15,16]. One review found large overstatements of the reduction in infarct volume in animal stroke studies without randomization or blinded outcome assessment when they were compared with randomized or blinded studies [16]. Indeed, even laboratory experiments measuring ligand binding at cannabinoid receptors can be confounded by study design, with studies which reported the use of the serine protease inhibitor PMSF giving significantly lower estimates of affinity than those which did not [17]. Much more robust observations have been made in reviews of studies in humans. Clinical trials in which authors did not report randomization, adequately concealed treatment allocation, or double blinding yielded larger estimates of treatment effects than trials in which these study quality issues were reported [18–20].

A final explanation for the apparent failure to replicate results of animal studies in clinical trials is publication bias. It is widely

Table 1
Aspects of study quality to be reported in the manuscript.

- *Sample size calculation*: How the sample size was determined, and which assumptions were made.
- *Eligibility criteria*: Inclusion and exclusion criteria for enrolment.
- *Treatment allocation*: The method by which animals were allocated to experimental groups. If this allocation was by randomisation, the method of randomisation.
- *Allocation concealment*: The method to implement the allocation sequence, and if this sequence was concealed until assignment.
- *Blinding*: Whether the investigators and other persons involved were blinded to the treatment allocation, and at which points in time during the study.
- *Flow of animals*: Flow of animals through each stage of the study, with a specific attention to animals excluded from the analyses. Reasons for exclusion from the analyses.
- *Control of physiological variables*: Whether and which physiological parameters were monitored and controlled.
- *Control of study conduct*: Whether a third party controlled which parts of the conduct of the study.
- *Statistical methods*: Which statistical methods were used for which analysis.

Recommendations from Ref. [9], based on Refs. [22] and [24].

acknowledged that negative or neutral animal studies are published much less frequently, or in journals of lower impact, than positive studies, but the magnitude of the problem has only been tested with sufficient detail in studies of experimental stroke. A meta-analysis of 525 publications, included in systematic reviews of 16 interventions tested in animal studies of acute ischaemic stroke, suggested that publication bias accounts for at least one-third of the efficacy reported in systematic reviews of animal stroke studies. Of the 525 publications, only ten (2%) did not report at least one significant effect on either infarct volume or neurobehavioural score [21].

Although there is no direct evidence of a causal relationship, it is likely that the recurrent failure of apparently promising interventions to improve outcome in clinical trials has in part been caused by methodological flaws of preclinical studies, disparities between animal models and clinical trials, and publication bias favouring positive animal studies [9]. For this reason, we think that the design and reporting of each formal animal study testing the effectiveness of an intervention should be based on standards similar to those of clinical trials to ensure that decision making is based on high-quality and unbiased data. Aspects of study quality that should be reported in any manuscript are listed in the Table 1. To avoid the consequences of publication bias, we suggest that controlled animal studies should be registered in a way comparable to the registration of clinical trials, and that registration is referenced in publications. Several other recommendations have been published to improve the design, conduct, and analysis of animal experiments, and their reporting [8,9,22,23]. In 2003, a working group convened by the National Heart, Lung, and Blood Institute recommended the establishment of a system for rigorous preclinical testing of promising cardioprotective agents with clinical trial-like approaches, including blinding and randomization [7]. In compliance with this recommendation, a multicenter Consortium for preclinicAl assESsment of cARdioprotective therapies (CAESAR) has been established (http://www.nhlbi.nih.gov/meetings/workshops/horizons.htm). We hope that these strategies will help to reduce bias and to improve their reliability and reproducibility.

Disclosure Statement

Both authors have no conflict of interest.

References

[1] White HD, Chew DP. Acute myocardial infarction. Lancet 2008;372:570–84.
[2] Dirksen MT, Laarman GJ, Simoons ML, Duncker DJ. Reperfusion injury in humans: a review of clinical trials on reperfusion injury inhibitory strategies. Cardiovasc Res 2007;74:343–55.
[3] Yellon DM, Hausenloy DJ. Myocardial reperfusion injury. N Engl J Med 2007;357:1121–35.
[4] Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. JAMA 2006;296:1731–2.
[5] 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;59:467–77.
[6] Van der Worp HB, Van Gijn J. Clinical practice. Acute ischemic stroke. N Engl J Med 2007;357:572–9.
[7] Bolli R, Becker L, Gross G, Mentzer Jr R, Balshaw D, Lathrop DA. Myocardial protection at a crossroads: the need for translation into clinical therapy. Circ Res 2004;95:125–34.
[8] Hausenloy DJ, Baxter G, Bell R, Botker HE, Davidson SM, Downey J, et al. Translating novel strategies for cardioprotection: the Hatter Workshop Recommendations. Basic Res Cardiol 2010;105:677–86.
[9] Van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? PLoS Med 2010;7:e1000245.
[10] Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler 2010;16:1044–55.
[11] E.D. Rooke, H.M. Vesterinen, E.S. Sena, K.J. Egan, M.R. Macleod. Dopamine agonists in animal models of Parkinson's disease: A systematic review and meta-analysis. Parkinsonism Relat Disord 2011; epub ahead of publication.
[12] Van der Worp HB, de Haan P, Morrema E, Kalkman CJ. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. J Neurol 2005;252:1108–14.
[13] Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One 2009;4:e7824.
[14] Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med 2003;10:684–7.
[15] Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 2004;35:1203–8.
[16] 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;130:3063–74.
[17] McPartland JM, Glass M, Pertwee RG. Meta-analysis of cannabinoid ligand binding affinity and receptor distribution: interspecies differences. Br J Pharmacol 2007;152:583–93.
[18] Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002;359:614–8.
[19] Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002;359:696–700.
[20] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.
[21] Sena ES, Van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol 2010;8:e1000344.
[22] Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. Good laboratory practice: preventing introduction of bias at the bench. Stroke 2009;40:e50–2.
[23] Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol 2010;8:e1000412.
[24] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–94.

H. Bart van der Worp
Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands
Corresponding author at: Department of Neurology, G 03.228, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.
Tel.: + 31 88 7555555; fax: + 31 30 2542100.
E-mail address: h.b.vanderworp@umcutrecht.nl.

Malcolm R. Macleod
Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh, UK

14 December 2010