

Comparison of treatment effects between animal experiments and clinical trials: systematic review

Pablo Perel¹, Ian Roberts¹, Emily Sena², Philippa Wheble², Catherine Briscoe², Peter Sandercock², Malcolm Macleod², Luciano E Mignini³, Pradeep Jayaram⁴, Khalid S Khan⁴

EDITORIAL by Hackam

¹Crash Trials Coordinating Centre, London School of Hygiene and Tropical Medicine, London WC1E 7HT

²Clinical Neurosciences, University of Edinburgh

³Centro Rosarino de Estudios Perinatales, WHO Collaborative Centre in Maternal and Child Health, Rosario 2000, Argentina

⁴Division of Reproductive and Child Health, Birmingham Women's Hospital, University of Birmingham

Correspondence to: P Perel
pablo.perel@lshtm.ac.uk

BMJ 2007;334:197-200

doi: 10.1136/bmj.39048.407928.BE

ABSTRACT

Objective To examine concordance between treatment effects in animal experiments and clinical trials.

Study design Systematic review.

Data sources Medline, Embase, SIGLE, NTIS, Science Citation Index, CAB, BIOSIS.

Study selection Animal studies for interventions with unambiguous evidence of a treatment effect (benefit or harm) in clinical trials: head injury, antifibrinolytics in haemorrhage, thrombolysis in acute ischaemic stroke, tirlizad in acute ischaemic stroke, antenatal corticosteroids to prevent neonatal respiratory distress syndrome, and bisphosphonates in the prevention and treatment of osteoporosis.

Review methods Data were extracted on study design, allocation concealment, number of randomised animals, type of model, intervention, and outcome.

Results Corticosteroids did not show any benefit in clinical trials of treatment for head injury but did show a benefit in animal models (pooled odds ratio for adverse functional outcome 0.58, 95% confidence interval 0.41 to 0.83). Antifibrinolytics reduced bleeding in clinical trials but the data were inconclusive in animal models. Thrombolysis improved outcome in patients with ischaemic stroke. In animal models, tissue plasminogen activator reduced infarct volume by 24% (95% confidence interval 20% to 28%) and improved neurobehavioural scores by 23% (17% to 29%). Tirlizad was associated with a worse outcome in patients with ischaemic stroke. In animal models, tirlizad reduced infarct volume by 29% (21% to 37%) and improved neurobehavioural scores by 48% (29% to 67%). Antenatal corticosteroids reduced respiratory distress and mortality in neonates whereas in animal models respiratory distress was reduced but the effect on mortality was inconclusive (odds ratio 4.2, 95% confidence interval 0.85 to 20.9). Bisphosphonates increased bone mineral density in patients with osteoporosis. In animal models the bisphosphonate alendronate increased bone mineral density compared with placebo by 11.0% (95% confidence interval 9.2% to 12.9%) in the combined results for the hip region. The corresponding treatment effect in the lumbar spine was 8.5% (5.8% to 11.2%) and in the combined results for the forearms (baboons only) was 1.7% (−1.4% to 4.7%).

Conclusions Discordance between animal and human studies may be due to bias or to the failure of animal models to mimic clinical disease adequately.

INTRODUCTION

Before clinical trials are carried out, the safety and effectiveness of new drugs are usually tested in animal models.¹ Some believe, however, that the results from

animal experiments are not applicable to humans because of biological differences between the species and the type of animal model used.² In this paper we compared treatment effects from systematic reviews of clinical trials with a systematic review of corresponding animal experiments.³⁻⁵

METHODS

We identified six interventions with evidence of a treatment effect (benefit or harm) in systematic reviews of clinical trials: corticosteroids in head injury,⁶⁻⁸ antifibrinolytics in haemorrhage,⁹ thrombolysis in stroke,^{10 11} tirlizad in stroke,¹² antenatal corticosteroids to prevent neonatal respiratory distress syndrome,¹³ and bisphosphonates to treat osteoporosis.¹⁴ We then systematically searched for randomised and non-randomised controlled studies of the interventions in animal models (see bmj.com). For dichotomous measures we estimated odds ratios and confidence intervals and for continuous measures we estimated the effect size (see bmj.com). We calculated pooled odds ratios, effect sizes, and 95% confidence intervals using a random effects model. Heterogeneity was examined using the I^2 statistic.¹⁵ We investigated the possibility of publication bias using a funnel plot and statistical methods.¹⁶

RESULTS

The quality of the experiments for each intervention was poor (see bmj.com).

Corticosteroids for traumatic head injury

Clinical trials of corticosteroids for head injury did not show any benefit and showed an increased risk of mortality.⁶ Seventeen reports were found in animal models of head injury.^{w1-w17} Three reported adequate allocation concealment. Two reported the effect of corticosteroids on mortality but an effect estimate could not be calculated owing to missing numbers in one experiment and death of all the animals in the other.

Seven experiments reported neurological outcomes. Neurological status was assessed by the grip test (see bmj.com) and neurological severity score. Four experiments reported the grip test: pooled odds ratio 0.58 (95% confidence interval 0.41 to 0.83; see bmj.com). No heterogeneity was found ($I^2=0\%$). Three experiments reported neurological severity scores but none reported the scores in each group.

Antifibrinolytics in haemorrhage

Clinical trials show that antifibrinolytics reduce blood loss during surgery.⁹ Eight reports were found

This is the abridged version of an article that was posted on bmj.com on 15 December 2006: <http://bmj.com/cgi/doi/10.1136/bmj.39048.407928.BE>

on antifibrinolytics in animal models.^{w18-w25} One did not give the number of animals. One reported mortality, five reported bleeding time, five reported blood loss, and one reported haemoglobin loss. None reported the method of allocation concealment. Four reported blinded outcome assessment but failed to describe the method. One assessed the effect of antifibrinolytics on mortality, but no deaths occurred. Five reported the effects on blood loss but only one had sufficient data to enable the effect estimate to be calculated. One reported a decrease in gastric bleeding of 0.09 ml (95% confidence interval 0.08 to 0.10). One reported a decrease in blood loss (average 523 ml) but failed to give the standard deviation. One reported a decrease in blood loss by 0.83 ml but did not give the number of controls.

Thrombolysis in acute ischaemic stroke

Thrombolysis with tissue plasminogen activator reduces death or dependency after stroke, despite an increase in intracranial haemorrhage.^{10 11} Overall, 113 reports were found on the effects of using tissue plasminogen activator or related agents for thrombolysis in animal models of stroke.^{w26-w138}

Infarct volume was reported in 212 comparisons, neurobehavioural scores in 84, and haemorrhage in 146. The funnel plot suggested an excess of studies overstating efficacy (see [bmj.com](#)). Tissue plasminogen activator reduced infarct volume by 24% (95% confidence interval 20% to 28%), improved neurobehavioural scores by 23% (17% to 29%), and increased the probability of haemorrhage (odds ratio 1.96, 95% confidence interval 1.63 to 2.35; see [bmj.com](#)). Heterogeneity was substantial for infarct volume ($I^2=78.2\%$) and neurobehavioural scores ($I^2=75.2\%$).

Tirilazad in acute ischaemic stroke

Tirilazad increases the risk of death and dependency in patients with stroke.¹² Eighteen reports were found of tirilazad in animal models of stroke.^{w139-w156} All 18 reported infarct volumes and eight reported neurobehavioural outcomes (see [bmj.com](#)). The funnel plot suggested an excess of small experiments overstating efficacy. Tirilazad reduced infarct volume by 29% (95% confidence interval 21% to 37%) and improved neurobehavioural scores by 48% (29% to 67%). Heterogeneity was substantial for infarct volume ($I^2=73.3\%$) and neurobehavioural scores ($I^2=58.1\%$).

Antenatal corticosteroids to prevent neonatal respiratory distress syndrome

Antenatal corticosteroids reduce respiratory distress syndrome and mortality in neonates.¹³ Fifty six controlled reports were found of corticosteroids in animal models of preterm delivery.^{w157-w212} Thirty two assessed intramuscular or subcutaneous injections in mothers, six assessed intraperitoneal instillation in mothers, and 18 assessed injections into the fetus. Only three experiments assessed the effect of antenatal corticosteroid injections on neonatal

respiratory distress syndrome, but the outcome was measured differently.

Respiratory distress syndrome was reduced in the corticosteroid groups in all three experiments. In one experiment two of 15 calves in the corticosteroid group compared with nine controls developed respiratory distress syndrome ($P=0.01$). In another experiment the total (SD) lung capacity in newborn rabbits in the corticosteroid group was 1.8 (0.4) ml/g compared with 1.4 (0.4) ml/g in controls. In a third experiment, six of 12 monkeys in the corticosteroid group compared with 11 controls developed severe respiratory distress syndrome ($P=0.03$). Seven experiments reported the effects of corticosteroids on neonatal mortality: pooled odds ratio 4.2 (95% confidence interval 0.85 to 20.9). Heterogeneity was significant ($I^2=72.7\%$).

Bisphosphonates to treat osteoporosis

Bisphosphonates increase bone mineral density in postmenopausal women with osteoporosis.¹⁴ Sixteen controlled reports were found of bisphosphonates in animal models (two in baboons and 14 in rats).^{w213-w228} The effect of bisphosphonates on bone mineral density was reported in 11 experiments (see [bmj.com](#)). Eleven of 11 showed an increase in bone mineral density and six of six showed improvements in bone mass.

Meta-analysis showed that compared with placebo alendronate increased bone mineral density by 11.0% (95% confidence interval 9.2% to 12.9%) in the combined results for the hip region, by 8.5% (5.8% to 11.2%) for the lumbar spine, and by 1.7% (-1.4% to 4.7%) for the forearms (baboons only).

DISCUSSION

Concordance between animal studies and clinical studies varied for six interventions. Thrombolysis with tissue plasminogen activator was effective in animal models of stroke and the results agreed with the clinical trials. The animal studies were of poor quality, however, with evidence of publication bias. Our evidence for concordance may therefore be biased. We found over 100 experiments, totalling more than 3000 animals. The pooled result was therefore precise although not necessarily valid. The concordance may be explained by the large volume of evidence and the replication of similar designs in different animals and laboratories. Furthermore, tissue plasminogen activator was tested in older animals, those with comorbidities, and at a range of intervals after stroke onset, ensuring a reasonable match with patients in clinical trials. The results for bisphosphonates agreed between animal studies and clinical trials. We also found that antenatal corticosteroids reduced neonatal respiratory distress syndrome in animal studies and in clinical trials, although the data were sparse and we found no evidence of agreement for mortality.

The four experiments in our meta-analysis of corticosteroids in animal head injury models used the weight drop model.¹⁷ All had good allocation concealment and blinded outcome assessment. Taken together

they showed benefit. The experiments were, however, from one laboratory, had little evidence on adverse effects, and did not examine comorbidities. We also found a difference in results for tirlazad to treat stroke. The data from the animal studies suggested a benefit but the clinical trials showed no benefit and possible harm. The interval between stroke onset and treatment was, however, longer in the clinical studies (median five hours) than in the animal models (median 10 minutes). Some of the clinical trials recruited patients up to 24 hours after stroke onset. For antifibrinolytics in haemorrhage, clinical trials showed clear evidence of benefit despite the lack of any reliable data from animal models.

Animal studies are often carried out to learn about biological mechanisms and we cannot comment on the value of animal research. Although we tried to contact some authors, we analysed what was reported and cannot rule out unpublished data. Our systematic review does, however, provide insights into the limitations of animal models, including the extent to which they represent disease in humans.

Systematic reviews could facilitate the translation of research findings from animals to humans. The animal studies we included varied in methodological quality and sample sizes. Randomisation and blinding were rarely reported, which can have important implications as animal experiments carried out without either are five times more likely to report a positive treatment effect.¹⁸

In the systematic review of thrombolysis in acute ischaemic stroke we found strong evidence of publication bias. Assessment of publication bias was difficult in the other systematic reviews owing to the number of experiments. In most cases we pooled the results to provide estimates of efficacy, although accuracy is open to question owing to heterogeneity. Prospective registration of animal experiments might reduce publication bias. We asked the UK Home Office for details of animal experiments relevant to our study, however they were unable to provide them (see bmj.com).

Empirical evidence of bias from study design characteristics helped to improve the quality of clinical trials and might do the same for animal experiments. Standards for evidence based reporting might ensure that relevant aspects of experiment methodology are reported.¹⁹

Systematic reviews can provide insights into the limitations of animal models. In stroke, the time from the occlusive event to the start of treatment was similar in animal and human studies. In head injury, treatment was given within five minutes of injury in the animal models but up to eight hours after injury in the clinical trials. None of the animal experiments used models that mimic the complex situations after head injury. Comorbidities are relevant in people with stroke and people with head injuries, but they were examined only in stroke models.

Systematic reviews of animal experiments could promote closer collaboration between the research

WHAT IS ALREADY KNOWN ON THIS TOPIC

The relevance of animal models to human health is questioned because of differences between the species

WHAT THIS STUDY ADDS

Many studies in animal models are of poor methodological quality

Lack of concordance between animal experiments and clinical trials may be due to bias, random error, or the failure of animal models to adequately represent human disease

communities and encourage an iterative approach to improving the relevance of animal models to clinical trial design. When models do not represent the clinical context they could be adapted accordingly.

We thank Sir Iain Chalmers for his constructive comments on the manuscript.

Contributors: See bmj.com.

Funding: This work was funded by the National Health Service research and development health technology assessment programme. The views expressed in this publication are those of the authors and not necessarily those of the methodology programme, health technology assessment programme, or the Department of Health. Funding of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Competing interests: IR was an investigator in the corticosteroid randomisation after significant head injury trial. The trial was funded by the UK Medical Research Council. Pharmacia and Upjohn (Pfizer from 2003) provided the Medical Research Council with the methylprednisolone (free of charge) needed for the trial, a grant in aid for preparation of the placebo, and support for collaborators' meetings. PS is co-chief investigator of the third international stroke trial, testing intravenous recombinant tissue plasminogen activator in acute ischaemic stroke; the start-up phase (completed in 2005) of this trial was supported by Boehringer Ingelheim, the manufacturers of tissue plasminogen activator, a donation of drug and placebo for the first 300 patients. The current phase of the trial is supported by the Medical Research Council and the Health Foundation. None of the authors have any relevant competing financial interests.

Ethical approval: Not required.

A table showing the quality of animal experiments included in the systematic reviews is available at www.crash2.lshtm.ac.uk.

- 1 Sibbald WJ. An alternative pathway for preclinical research in fluid management. *Crit Care* 2000;4(suppl 2):S8-15.
- 2 Croce P. *Vivisection or science? An investigation into testing drugs and safeguarding health*. London: Zed Books, 1999.
- 3 Pound P, Ebrahim S. Supportive evidence is lacking for report on animal studies. *BMJ* 2002;325:1038.
- 4 Roberts I, Kwan I, Evans P, Haig S. Does animal experimentation inform human health care? Observations from a systematic review of international animal experiments on fluid resuscitation. *BMJ* 2002;324:474-6.
- 5 Pound P, Ebrahim S, Sandercock P, Bracken M, Roberts I. Where is the evidence that animal research benefits humans? *BMJ* 2004;328:514-7.
- 6 Alderson P, Roberts I. Corticosteroids for acute traumatic brain injury. *Cochrane Database Syst Rev* 2005;(1):CD000196. DOI:10.1002/14651858.CD000196.pub2.
- 7 The CRASH Trial Collaborators. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury—outcomes at 6 months. *Lancet* 2005;365:1957-9.
- 8 The CRASH Trial Collaborators. Effect of intravenous corticosteroids on death within 14 days in 10,008 adults with clinically significant head injury (MRC CRASH Trial): a randomised placebo-controlled trial. *Lancet* 2004;364:1321-8.
- 9 Henry DA, Moxey AJ, Carless PA, O'Connell D, McClelland B, Henderson KM, et al. Anti-fibrinolytic use for minimising perioperative allogeneic blood transfusion. *Cochrane Database Syst Rev* 1999;(4):CD001886. DOI:10.1002/14651858.CD001886.

- 10 Mielke O, Wardlaw J, Liu M. Thrombolysis (different doses, routes of administration and agents) for acute ischaemic stroke. *Cochrane Database Syst Rev* 2004;(4):CD000514. DOI:10.1002/14651858.CD000514.pub2.
- 11 The ATLANTIS, ECASS, and NINDS rt-PA Study Group Investigators. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet* 2004;363:768-74.
- 12 The Tirilazad International Steering Committee. Tirilazad for acute ischaemic stroke. *Cochrane Database Syst Rev* 2001;(4):CD002087. DOI:10.1002/14651858.CD002087.
- 13 Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 2006;(3):CD004454. DOI:10.1002/14651858.CD004454.pub2
- 14 Cranney A, Wells G, Willan A, Griffith L, Zytaruk N, Robinson V, et al. Meta-analyses of therapies for postmenopausal osteoporosis. II. Meta-analysis of alendronate for the treatment of postmenopausal women. *Endocr Rev* 2002;23:508-16.
- 15 Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
- 16 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analyses detected by a simple graphical test. *BMJ* 1997;315:629-34.
- 17 Cernak I. Animal models of head trauma. *NeuroRx* 2005;2:410-22.
- 18 Bebarta B, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad Emerg Med* 2003;10:684-7.
- 19 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4. References w1-w228 are on bmj.com

Accepted: 7 November 2006

BMJ UPDATES

Acupuncture may speed up the active phase of labour in women whose membranes rupture before onset of labour

Research question

Can acupuncture help pregnant women whose membranes rupture at term but before labour has started?

Answer

Acupuncture may speed up the active phase of labour and reduce the need for oxytocin to augment labour.

Why did the authors do the study?

There's some evidence that acupuncture can augment labour for women giving birth at term. These authors wanted to find out whether it might specifically help women whose membranes rupture before labour is established. Pre-labour rupture of membranes puts women and their infants at risk of ascending infection, and labour is usually induced if it doesn't start spontaneously within a day or two.

What did they do?

One hundred healthy women with singleton pregnancies took part in a randomised controlled trial. All had spontaneous rupture of membranes at term but before the onset of labour. Forty eight had acupuncture while they were waiting. The remaining 52 did not. The acupuncture was individually tailored to encourage and augment labour. If labour did not start spontaneously within 24 hours of rupture of membranes, it was induced with prostaglandins or oxytocin.

The authors compared the two groups after excluding six women who had caesarean sections, two who had intact membranes, and one who received acupuncture by mistake. They were interested in the impact of acupuncture on speed of delivery, length of the active phase of labour, need for induction, and need for augmentation of labour with intravenous oxytocin.

What did they find?

Women who had acupuncture had a shorter active

phase of labour than controls (mean time 4.4 hours v 6.1 hours, mean difference 1.7 hours (95% CI 0.2 to 3.1), P=0.027), but they did not deliver any faster overall (29.1 hours from rupture of membranes v 32.7 hours). Women who had acupuncture were no less likely to be induced than controls, but they were less likely to need an oxytocin infusion to augment labour once it started (use of oxytocin for >2 hours, 9/43 v 21/48, P=0.018).

Among the women who needed induction of labour, those who had received acupuncture had a faster active phase than did controls (3.46 hours v 7.06 hours, P=0.002).

What does it mean?

This small and preliminary study suggests that acupuncture has detectable beneficial effects on labour among women with spontaneous rupture of membranes at term. But because there was no active control (such as sham acupuncture), it's hard to say whether the benefits were due to the acupuncture itself or to the 20 minutes of personal attention received during treatment. The findings need to be confirmed in bigger trials with a better control group and analysed by intention to treat. Exclusions left this trial with only 43 women in the acupuncture group and 48 controls. The authors did not do a power calculation, but it's likely their trial was underpowered. The negative results—on overall speed of delivery, for example—should be treated with caution.

Gaudernack et al. Acupuncture administered after spontaneous rupture of membranes at term significantly reduces the length of birth and use of oxytocin. A randomized controlled trial. *Acta Obstetrica et Gynecologica* 2006;85:1348-53

This summarises a paper that has been selected by [bmjupdates](http://bmjupdates.com). To register for [bmjupdates](http://bmjupdates.com) (free email about high quality new papers in your favourite subjects) go to <http://bmjupdates.com>