

Animal experiments and clinical trials: systematic review

Do animal experiments have any tangible benefits for assessing the value of a drug for clinical use? **Kristina Fister** describes a recent paper that attempted to find out

Kristina Fister
associate editor
BMJ
kfister@bmj.com

This month's paper is "Comparison of treatment effects between animal experiments and clinical trials: systematic review" by P Perel and colleagues (*BMJ* 2007;334:197-200). You can read it by going to studentbmj.com and clicking on the link.

What is the value of animal research in human clinical medicine? This month's paper offers a new methodological approach to investigate this question. It seems that although some animal studies find results similar to those found in human trials, other animal studies may give different, sometimes even opposite, results.

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, tirilazad in acute ischaemic stroke, antenatal corticosteroids to prevent neonatal respiratory distress syndrome, and bisphosphonates to treat 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% (20% to 28%) and improved neurobehavioural scores by 23% (17% to 29%). Tirilazad was associated with a worse outcome in patients with ischaemic stroke. In animal models, tirilazad 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, 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% (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 because of bias or the failure of animal models to mimic clinical disease adequately.

Why do the study?

Few people would choose to support animal research, which is arguably closer to cruel than to humane, if they didn't believe it was necessary for the advancement of human medicine. Yet little scientific evidence has informed the debate on how useful animal research for human medicine really is.

In experiments, researchers look at how animals react to drugs or procedures. The results are then used to predict how humans might react to the same drugs or procedures, and clinical trials are designed based on these predictions. But how accurate are they? And how do we go about finding that out?

What did the authors do?

Previous researchers used different methods to assess the value of animal studies: historical analyses, critiques of animal models, surveys of clinicians, and citation analyses. However, the authors of this month's paper thought of another way to do this.

They chose six well defined clinical topics on which we have unambiguous evidence of benefit and harm. For example, we know that giving corticosteroids to people with traumatic brain injury doesn't make people better and in fact increases people's risk of dying. These conclusions have previously been reached in a systematic review of available clinical trials, which the authors appropriately reference. To assess the value of animal studies, the authors wanted to see how well these results corresponded to those from a systematic review of animal studies on the same topic. By directly comparing rigorously collected, appraised, and pooled evidence from animal studies with the corresponding evidence from systematic reviews of human trials, the authors presented a novel, robust approach to providing the evidence for the value of animal research.

Only a couple of dozen systematic reviews of animal studies have been done to date. But it has been suggested that animal experiments are not put to good enough use if results from different studies answering the same research question are not combined. If we do systematic reviews for clinical trials, then we should also do them for animal studies to see what is the current best evidence for a given research question. This might avoid doing a number of unnecessary animal studies.

The authors of this month's paper did six systematic reviews of animal studies and reported them all in one paper. This is unusual. It means that in critically appraising this paper we should check how each of

these six reviews was done. This is not entirely possible because not all details important for critical appraisal could be reported in one paper reporting on so much research. What's more, the findings were then compared with results from corresponding systematic reviews of human trials.

Let's glance at the quality of the six reviews. In all, the authors clearly defined the research question. They searched multiple databases (Medline, Embase, SIGLE, NTIS, Science Citation Index, CAB, and BIOSIS) for relevant articles published in any language. They checked reference lists of the included articles for additional relevant studies. Furthermore, the authors assessed the quality of the included studies and when they were unsure of something, they contacted the authors of a particular study to clarify it. Many stages of the process, including extraction of data, were separately done by two independent reviewers. This is all good methodology.

In all but one review the authors were able to use a meta-analysis to calculate pooled treatment effects. This was lucky. Studies included in systematic reviews are often too heterogeneous, too different from one another, for their results to be pooled in this way. Here, however, the authors were able to do so for at least some outcomes in five out of six reviews. This made it easier to compare the results of reviews of animal studies with the results of human trials.

What was found?

Of the six interventions studied, three showed concordance in animal and human studies and three showed discordance. Let's briefly look at each of these interventions.

We know that corticosteroids don't make things better for people with head injury caused by trauma. Giving corticosteroids to these people only increases their risk of dying. However, four animal studies showed that corticosteroids made neurological outcomes better for mice after head injury, and injured rats' neurological functioning didn't improve with steroids. Animal studies didn't have enough data to give conclusive answers on the steroids' impact on mortality.

Giving antifibrinolytics to people having surgery reduces blood loss. However, studies that assessed the effects of antifibrinolytics in animals didn't find much reliable data. Some reported modest reductions in blood loss with antifibrinolytics, but the studies lacked information such as standard deviation or the number of animals in groups, which precluded the researchers from reaching any firm conclusions based on animal studies alone.

In people with acute ischaemic stroke, thrombolysis with recombinant tissue plasminogen activator increases intracranial haemorrhage but reduces deaths and dependency. There were 113 animal studies that looked at this research question, and the results were similar to those from human trials: the treatment reduced the infarct volume, improved neurological and behavioural outcomes, and increased the risk of haemorrhage. Still, there was convincing evidence of publication bias and overstatement of efficacy. This would have cast some doubt on the results of animal studies alone.

People who had acute ischaemic stroke should not take tirilazad because it increases the risk of death and dependency. However, animal studies had shown differently: tirilazad reduced infarct volume and improved neurobehavioural scores. Again, although animal trials showed benefits, clinical trials showed harm. Some evidence of publication bias and overstatement of efficacy was found here as well, but it was less convincing than in the studies of thrombolysis.

Corticosteroids given to pregnant women who are at risk of preterm birth reduce the risk of respiratory distress syndrome and mortality in neonates. These findings are similar to those from animal studies, although animal studies didn't provide unequivocal evidence for the reduction in mortality.

Finally, we know that bisphosphonates increase bone mineral density in postmenopausal women with osteoporosis. This is in full concordance with results from animal studies: alendronate improved bone mineral density and bone mass in all of the included studies.

The vast majority of the animal studies included in the six systematic reviews were of poor methodological quality. Less than half of the studies described randomisation having been done to allocate subjects to study groups, and less than a fifth of the studies had adequate allocation concealment. Reporting of blinded assessment of outcomes ranged from only 5% of the studies included in a systematic review to more than 70% in one systematic review.

Why does this matter? Randomisation and blinding reduce bias. A previous study reported that animal experiments without randomisation or blinding compared with experiments with designs that incorporated these methods of reducing bias were five times more likely to report a positive treatment effect. Poor quality of the included animal studies means that bias must be considered as one of the possible explanations in interpreting the results of this month's paper.

What does the study mean?

In an area of research that lacks robust evidence, this paper contributes two main things. Firstly, it shows how systematic reviews of animal studies can be used to assess the value of animal studies for human medicine. And secondly, the authors gave concrete robust evidence on the value of the evidence from animal studies for six specific medical interventions.

Although some of the presented reviews of animal studies found similar results to the reviews of clinical trials, some came to much different conclusions. The question is why? Discordance could have been because of bias, such as that coming from the lack of randomisation, inadequate allocation concealment, or the failure of the research community to publish small negative studies. Consciously or unconsciously, basic researchers may be skewing the evidence towards overstating the benefits of interventions tested in animal studies. To help minimise publication bias, the authors propose that all animal studies should be prospectively registered, as is required for clinical trials since recently.

Another explanation is that some animal studies fail to mimic human disease well enough. For example, animal studies of tissue plasminogen activator (which showed good concordance with trials) recruited older animals with comorbidities, who received the tested drug in the time period after stroke which was comparable to that used in clinical trials. On the other hand, in animal studies looking at corticosteroids for head injury (which showed poor concordance), comorbidities were not examined although they may have been relevant. Also, the animals received the drug within five minutes from the head injury, while the corresponding interval in clinical trials lasted for up to eight hours.

Although animal studies are largely used to inform the design of clinical trials, they are also used for researching biological mechanisms. One of the limitations of this month's paper is that it didn't explore the value of animal studies for this area of research. We can't make generalisations based on only the six examples presented here. But we've seen a good way to better inform our opinions on animal research.