

PRECLINICAL STUDIES

Extrapolating from Animals to Humans

John P. A. Ioannidis

Because of a variety of caveats, the safety and effectiveness of interventions in human subjects can only be speculated from animal studies. Careful synthesis of data from multiple animal studies is needed to begin to assess the likelihood of successful cross-species translation (Fay *et al.*, this issue).

A perusal of PubMed (search date: 21 July 2012) with the search term “animal” yields 4,993,350 papers—almost a quarter of the biomedical literature—a number that exceeds that obtained by searching with the term “patient” ($N = 4,337,985$ hits). Mice and rats are king in the biomedical literature ($N = 1,193,679$ and $N = 753,612$ hits, respectively), although these rodents are arguably distant relatives of *Homo sapiens sapiens*. Other animals are also well represented; for example, the search term “rabbit” yields 354,561 hits, and “rhesus monkey” yields 35,558 hits. The phrase “animal model” yields almost a half million papers ($N = 495,339$).

Although some research is performed purely for the sake of studying the physiology and pathophysiology of animals, the goal of the majority of animal studies is to gain knowledge and insights that are useful for understanding human biology, the response of humans to treatments or other interventions, or both. But how successful is this cross-species translation (Fig. 1)? In this issue, Fay *et al.* (1) present a careful synthesis of mortality data from 21 animal studies on the new recombinant protective antigen (rPA) vaccines and the already-licensed anthrax vaccine adsorbed (AVA).

ANIMALS VERSUS HUMANS

How well treatment effects observed in animals translate to human subjects may depend on the type of intervention, administration protocol, disease complexity, animal model, and other case-specific factors. There are strong opposing opinions among enthusiasts and skeptics about the relevance of animal data for humans. To simplify

matters, this discussion will set aside mechanistic research and focus on preclinical animal studies that attempt to predict whether specific interventions will have preventive or therapeutic effects in human subjects. Empirical evaluations that have assessed the performance of animal research in this regard (2, 3) have not been favorable: Limited concordance exists between treatment effects in preclinical animal experiments and clinical trials in human subjects.

A large systematic evaluation examined

the results of preclinical animal experiments for several interventions—corticosteroids for head injury or to prevent neonatal respiratory distress syndrome; antifibrinolytics for hemorrhage; tissue plasminogen activator (tPA) or tirilazad for acute ischemic stroke; and bisphosphonates for osteoporosis—for which there is unambiguous evidence of a treatment effect (benefit or harm) in clinical trials in humans (2). The results in animals were often opposite of those seen in humans; for example, in animal studies, corticosteroids had a therapeutic effect on head injury but increased mortality in newborns; antifibrinolytics did not reduce bleeding; and tirilazad improved treatment of ischemic stroke. Conversely, tPA was beneficial in treating ischemic stroke in both humans and animals, and bisphosphonates increased bone mineral density in both patients with osteoporosis and animal models.

Potential explanations for the failure of animal models to capture treatment effects in humans can be placed into two categories:

First, both the human and animal results are accurate, but human physiology and disease are not adequately captured by animal models. Second, the animal literature is susceptible to biases in the study design, to reporting biases that distort the published evidence, or both. Indeed, although the scientific literature related to human clinical trials suffers from biases (4), data from preclinical animal studies appear to be associated with even greater bias, for a variety of reasons discussed below.

The first category of animal data translation failures is difficult to overcome. If the animal model is not a good representation of human physiology or disease, there is little that can be done beyond identifying or creating a new, more suitable model—not a straightforward task. At a minimum, claims for effectiveness of interventions should be made only after the results are reproduced in different species and settings.



Fig. 1. Animal instincts. Half-horse, half-human, the mythological centaur Chiron (root, *chirurgos*: surgeon) taught all of the great heroes with skills related to human health (Jason, Peleus, Asklepios, and Achilles). Following in this tradition, most animal studies are performed to gain insights into human physiology, pathophysiology, and response to new therapies. Unfortunately, many other centaurs (like the female centaur shown here) were more biased and profligate.

Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine, and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA 94305, USA.

E-mail: jioannid@stanford.edu

The second category of translation failures, bias, is a common and more easily remediable cause of poor concordance between preclinical and clinical outcomes. Empirical studies (5–7) suggest that animal research often suffers from poor study design, and features of study quality correlate with the robustness of results obtained. For example, an evaluation of abstracts accepted to the Society for Academic Emergency Medicine appraised 290 animal studies with two or more experimental groups (5) and found that 194 studies were not randomized and 259 studies were not blinded; the non-randomized and nonblinded studies had 3.4- and 3.2-fold higher odds, respectively, of claiming a statistically significant outcome than did those that were randomized and blinded. In another empirical evaluation of 13 meta-analyses of experimental stroke (6) that described outcomes in 15,635 animals, studies with unblinded induction of ischemia and those that used healthy animals reported 13.1 and 11.5% higher effect sizes than blinded studies and studies of animals with stroke comorbidities, respectively. Another recent systematic review of animal model studies on stem cell treatment of stroke found larger benefits in nonrandomized than in randomized studies (7).

The CAMARADES (Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies) initiative has conducted several large-scale investigations of preclinical studies performed with animal models for diverse conditions, and they consistently show hints of serious reporting bias (7–9). For example, one such empirical evaluation assessed the accumulated evidence on 16 interventions tested in animal models of acute ischemic stroke, a total of 525 unique scientific publications (8). Only one of the 16 stroke interventions that yielded positive therapeutic results in preclinical animal studies, tPA, functioned similarly in human subjects, and even this agent was effective only in selected patients and circumstances. Only ten of the 525 publications (2%) reported no statistically significant effect of the intervention on infarct volume, and only six (1.2%) did not report at least one statistically significant favorable finding. For all 16 interventions, regressions relating the effect size to the magnitude of the observed treatment effect found that studies with smaller numbers of animals showed more prominent therapeutic benefits than did studies with larger test-animal populations. This pattern is compatible with

serious publication bias or other selective reporting biases, such as selective outcome and analysis reporting. It is possible that animal studies are published only if they show that the tested treatment displays a therapeutic effect (traditional publication bias) or if they yield results that show that the treatment is effective, even if it is not (selective outcome and analysis reporting bias).

THE ANIMAL RULE

Because of these caveats, it is nearly impossible to rely on most animal data to predict whether or not an intervention will have a favorable clinical benefit–risk ratio in human subjects. However, a particularly difficult situation arises when testing interventions for diseases or exposures in which human experimentation is unethical or otherwise not feasible. In 2002, after years of deliberation and in response to the rising threat of bioterrorism, the U.S. Food and Drug Administration (FDA) formulated the so-called Animal Rule (10, 11), which offers the ability to license medical countermeasures for biological, chemical, and radiation threats on the basis of effectiveness data in multiple species of animals coupled with immunogenicity and safety data in animals and humans.

Between 2001 and 2011, more than \$50 billion was spent by the U.S. government on diverse aspects of biodefense, including therapeutic discovery and development, and the rationality of this resource allocation has been questioned (12). Regardless, with this magnitude of investment, one would predict that the Animal Rule has led to the licensing of dozens of countermeasures, including vaccines that protect against anthrax, plague, smallpox, viral encephalitis, or Ebola hemorrhagic fever, all of which are too rare to make clinical trials feasible.

However, only two licenses have been granted by the Animal Rule, and neither pertains to a vaccine. In fact, these two licensed countermeasures had been developed in the past, and their manufacturers used the Animal Rule to obtain formal licensing for new indications: Pyridostigmine bromide, which is used to treat myasthenia gravis, was newly approved for the management of exposure to Soman gas (a cholinesterase inhibitor), and hydroxocobalamin, which is already used to treat vitamin B12 deficiency, was newly approved as Cyanokit for treating (in much larger doses) cyanide toxicity (11). Still, a large number of other bioterrorism countermeasures exist—including several

vaccines—and despite criticisms (12) and intermittent lack of formal FDA approval, persuasive cases have been made for creating large national stockpiles of some of these countermeasures for use as investigational agents by consenting individuals in cases of emergency and by military or other personnel at risk of exposure (11).

LICENSING VACCINES

The lack of licensing of any vaccine countermeasures through the Animal Rule does not mean that research in this area has stalled. Several vaccines against bioterrorism organisms are available, and more are being developed for *Bacillus anthracis*, *Yersinia pestis*, viral encephalitis agents, and Ebola virus. But vaccine developers have not typically made use of the Animal Rule because the vaccines and countermeasures can be sold to and used by the U.S. government without FDA approval.

Fay *et al.* (1) present a careful synthesis of data from 21 animal studies on anthrax vaccines—new rPA vaccines and the already-licensed (since 1970) AVA. Studies on anthrax vaccines date back to 1881, with work by Louis Pasteur and colleagues, and cell-free vaccines for use in humans were developed in the 1950s. As early as 1962, a publication on the results of a placebo-controlled trial documented clinical efficacy of a cell-free vaccine (precursor of AVA) in human subjects (13); proof of clinical efficacy had required a study population of 1249 mill workers who were followed for 4 years to document 26 cases of anthrax. Such a trial is no longer feasible in the United States: According to the Mortality and Morbidity Weekly Report of the Centers for Disease Control, only three cases of human anthrax have been reported in the past 5 years.

The study by Fay *et al.* (1) is remarkable because of the meticulous way in which it tries to approach the challenges of meeting the four prerequisites of the Animal Rule to support the hypothesis that a vaccine shown to be effective in animals in producing both an immunological response and in aborting mortality also will prevent deaths in human subjects. Let us examine these prerequisites in the context of the Fay *et al.* data.

The first prerequisite is that “there is a reasonably well-understood pathophysiological mechanism of the toxicity of the substance and its prevention or substantial reduction by the product.” For an anthrax vaccine, this prerequisite seems to be met, because the pathophysiology of the disease

Table 1. Making animal research credible.

Goal	Actions
Minimize publication and selective-reporting biases	Consider preregistration of animal studies (especially experimental trials); promote work done by all-inclusive, publicly visible consortia.
Improve study design, conduct, and reporting	Use appropriate design and statistical methods; avoid over-interpretation; maximize transparency; fully document all anticipated, iterative, or exploratory steps in the research; use standard reporting guidelines [for example, ARRIVE (15)].
Make raw data, analyses, and protocols available	Allow other investigators to see the full workflow and to repeat the analyses, if need be, to verify results and integrate them with other parallel or future efforts.

has been well characterized for more than a century, and humoral immunity seems to play a clear role in protection against anthrax. Obviously, this does not mean that we know everything there is to know about anthrax disease and the human immune response to infection by *Bacillus anthracis*.

The second prerequisite is that “the effect is demonstrated in more than one animal species expected to react with a response predictive for humans.” Fay *et al.* summarized the results of 21 different experiments involving three different species—rabbits, cynomolgus macaques, and rhesus macaques. The disease manifestations in these species bear substantial similarity to those in humans. For example, rhesus macaques exhibit mediastinal, lymphatic, and pulmonary lesions, and rabbits exhibit fulminant systemic disease with necrotizing lymphadenitis, splenitis, pneumonia, and vasculitis. But differences also exist, such as in the survival response curves across different species. The authors are careful when discussing the extent of exchangeability of the survival response curves by highlighting the extent of the cross-species differences. Predictions for survival in humans vary from 54 to 84% depending on which animal model is extrapolated, and confidence intervals are substantial.

The third prerequisite is that “the animal study endpoint is clearly related to the desired benefit in humans, generally the enhancement of survival or prevention of major morbidity.” For anthrax, the endpoint of interest is survival, and the animal models can capture this adequately, with the caveats about exchangeability discussed above. For most other diseases and outcomes beyond death, juxtaposing such outcomes in animals and humans can be a delicate exercise; for example, in stroke it is difficult to find outcomes in animals that correspond to bet-

ter function and rehabilitation in human subjects (14).

Last, the Animal Rule stipulates that “the data or information on the kinetics and pharmacodynamics of the product or other relevant data or information, in animals and humans, allows selection of an effective dose in humans.” This is considered to be less of an issue for vaccines than for drugs, because on the basis of data collected from phase II and phase III trials in humans, it should be possible to select a reasonable dose to achieve immunogenicity in humans. However, this does not mean that the dosage will necessarily be the overall optimal one for humans. For most vaccines, many adverse events show dose-threshold and dose-responses, and identifying the minimal necessary dose to maximize the effectiveness-toxicity ratio is not straightforward, because it is unlikely that many different doses will be tested in large-scale studies.

For the Animal Rule pathway, when a vaccine is licensed, there may be very limited data on safety in humans; however, it is crucial to carefully record adverse events in humans after licensing. If toxicity signals emerge, manufacturers may need to revisit the chosen dose or mode and schedule of administration. Some medical interventions are eventually proven to be unsafe, even when their licensing package has included well-conducted human trials. It is unclear what the reversal rates might be for products that are licensed on the basis of animal data for effectiveness alone, but presumably they will be higher than those for products that are licensed on the basis of human trial data. Recording of adverse events in noncontrolled (post-licensing) settings leaves considerable uncertainty. Since the 1990s, researchers and regulators have endured the frustrating experience of long, unproductive debates surrounding

the safety of the AVA vaccine. Although the vaccine now appears to be safe, these debates represent an additional difficulty one must consider when evidence is limited or based on data that can be questioned with good or not-so-good intentions.

IMPROVING ANIMAL RESEARCH

Acknowledging the various unavoidable difficulties, lessons learned from careful animal work on vaccines for lethal, rare diseases also may be useful for improving research conducted in animals for common diseases (Table 1). Enhancing the quality of animal studies will directly improve a quarter of the biomedical literature and may also benefit much of the other three-quarters that have an interface with animal research. Efforts are needed to minimize publication and other selective-reporting biases. Study design, conduct, and reporting can be improved—for example, by using the Animals in Research: Reporting In Vivo Experiments (ARRIVE) guidelines (15).

Much animal research is iterative and exploratory, and it is not possible to lay out in advance and in detail the research agenda of all preclinical trials to be performed. This means that at a minimum, careful documentation and inclusion of all collected data (both published and unpublished) is essential. Fay *et al.* offer an example in this regard by providing supplementary data, *R* functions, and notes on how to apply them. Optimal documentation could allow broad access to raw data and the analytical codes (computational models, bioinformatics algorithms, and statistical methods) used to analyze datasets. Such practices will maximize transparency, allow integration of multiple studies on the same topic, and enhance trust in the results of animal research efforts.

REFERENCES AND NOTES

1. M. P. Fay, D. A. Follman, F. Lynn, J. M. Schiffer, G. V. Stark, R. Kohberger, C. P. Quinn, E. O. Nuzum, Anthrax vaccine-induced antibodies provide cross-species prediction of survival to aerosol challenge. *Sci. Transl. Med.* **4**, 151ra126 (2012).
2. P. Perel, I. Roberts, E. Sena, P. Wheble, C. Briscoe, P. Sandercock, M. Macleod, L. E. Mignini, P. Jayaram, K. S. Khan, Comparison of treatment effects between animal experiments and clinical trials: Systematic review. *BMJ* **334**, 197–199 (2007).
3. P. Pound, S. Ebrahim, P. Sandercock, M. B. Bracken, I. Roberts Reviewing Animal Trials Systematically (RATS) Group, Where is the evidence that animal research benefits humans? *BMJ* **328**, 514–517 (2004).
4. J. P. Ioannidis, Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
5. V. Bebarta, D. Luyten, K. Heard, Emergency medicine ani-

- mal research: Does use of randomization and blinding affect the results? *Acad. Emerg. Med.* **10**, 684–687 (2003).
6. N. A. Crossley, E. Sena, J. Goehler, J. Horn, B. van der Worp, P. M. Bath, M. Macleod, U. Dirnagl, Empirical evidence of bias in the design of experimental stroke studies: A meta-epidemiologic approach. *Stroke* **39**, 929–934 (2008).
 7. J. S. Lees, E. S. Sena, K. J. Egan, A. Antonic, S. A. Koblar, D. W. Howells, M. R. Macleod, Stem cell-based therapy for experimental stroke: A systematic review and meta-analysis. *Int. J. Stroke* 12 Jun 2012 (10.1111/j.1747-4949.2012.00797.x).
 8. E. S. Sena, H. B. van der Worp, P. M. Bath, D. W. Howells, M. R. Macleod, Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* **8**, e1000344 (2010).
 9. M. R. Macleod, T. O'Collins, D. W. Howells, G. A. Donnan, Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* **35**, 1203–1208 (2004).
 10. P. J. Snoy, Establishing efficacy of human products using animals: The US Food and Drug Administration's "animal rule." *Vet. Pathol.* **47**, 774–778 (2010).
 11. P. Aebersold, FDA experience with medical countermeasures under the Animal Rule. *Adv. Prev. Med.* **2012**, 507571 (2012).
 12. B. S. Levy, V. W. Sidel, Adverse health consequences of US Government responses to the 2001 terrorist attacks. *Lancet* **378**, 944–952 (2011).
 13. P. S. Brachman, H. Gold, S. A. Plotkin, F. R. Fekety, M. Werin, N. R. Ingraham, Field evaluation of a human anthrax vaccine. *Am. J. Public Health Nations Health* **52**, 632–645 (1962).
 14. B. Cheeran, L. Cohen, B. Dobkin, G. Ford, R. Greenwood, D. Howard, M. Husain, M. Macleod, R. Nudo, J. Rothwell, A. Rudd, J. Teo, N. Ward, S. Wolf, Cumberland Consensus Working Group, The future of restorative neurosciences in stroke: driving the translational research pipeline from basic science to rehabilitation of people after stroke. *Neurorehabil. Neural Repair* **23**, 97–107 (2009).
 15. C. Kilkenny, W. J. Browne, I. C. Cuthill, M. Emerson, D. G. Altman, Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).

Competing Interests: The author declares that he has no competing interests.

10.1126/scitranslmed.3004631

Citation: J. P. A. Ioannidis, Extrapolating from animals to humans. *Sci. Transl. Med.* **4**, 151ps15 (2012).

Science Translational Medicine

Extrapolating from Animals to Humans

John P. A. Ioannidis

Sci. Transl. Med., 4 (151), • DOI: 10.1126/scitranslmed.3004631

View the article online

<https://www.science.org/doi/10.1126/scitranslmed.3004631>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)