Weighing the Costs and Benefits of Animal Experiments

Andrew Knight

Oxford Centre for Animal Ethics, Oxford, UK

Summary

Most regulations governing animal experimentation require that the harms expected to be incurred by animal subjects should be balanced against the likely benefits of the project. Too often, however, expected human benefits are based on unrealistic assumptions. To critically assess the human clinical and toxicological utility of animal experimentation, the published literature was comprehensively surveyed to locate relevant systematic reviews. In only two of 20 reviews – one of which was contentious – did the authors conclude that animal models were either significantly useful in contributing to the development of human clinical interventions or substantially consistent with clinical outcomes. Included were reviews examining the clinical utility of invasive chimpanzee experiments, of highly cited animal experiments published in leading scientific journals, and of animal experiments approved by ethics committees at least partly on the basis of specific claims that they were likely to lead to concrete advances in human healthcare. Seven additional reviews failed to demonstrate reliable predictivity of human toxicities such as carcinogenicity and teratogenicity. Results in animal models were frequently equivocal or inconsistent with human outcomes. When considering costs and benefits overall, one cannot reasonably conclude that the human benefits exceed the costs incurred by animals subjected to scientific procedures. On the contrary, the evidence indicates that actual human benefit is rarely – if ever – sufficient to justify such costs. Despite this, deficiencies in the implementation of regulatory and policy requirements to replace, reduce, and refine animal use remain marked and widespread. A range of policy initiatives are warranted to address these deficiencies.

Keywords: 3Rs, systematic review, animal model, animal experiment, bioethics, animal ethics

1 Introduction

The core ethical principle underpinning animal experimentation regulation and policy is that the likely benefits of such research must outweigh its expected costs. Although considerable financial and human collateral costs do exist, the main costs are borne by the animals subjected to such research. Although such research may be directed at yielding benefits for animal species or the environment, the overwhelming majority is intended for human benefit, whether through the advancement of knowledge, through the development or toxicity testing of clinical interventions and consumer or industrial products, or through educational applications (Knight, 2011).

This utilitarian cost:benefit analysis fundamentally underpins regulations governing animal experimentation. *Directive 2010/63/EU on the protection of animals used for scientific purposes*, with which EU Member States must comply by 2013, asserts that it is *"essential, both on moral and scientific grounds, to ensure that each use of an animal is carefully evaluated as to the scientific or educational validity, usefulness and relevance of the expected result of that use. The likely harm to the animal should be balanced against the expected benefits of the project."* (EU, 2010)

The Canadian *Council on Animal Care guidelines on animal use protocol review* similarly state: *"Approval of a protocol does not guarantee that a benefit will be realized, but does mean that there will be a cost imposed on the animals. The (Animal Care Committee) must be convinced therefore of the need for animal use, and that the expected benefit will outweigh the cost."* (CCAC, 1997)

However, widespread contemporary reliance on animal models during biomedical research and toxicity testing depends heavily on assumptions of human utility and, in particular, reasonable predictivity for human outcomes, which are rarely scrutinized. Yet numerous cases of discordance between animal and human outcomes (Knight, 2011) question the veracity of such assumptions.

2 Systematic reviews of human utility

Arguments that rest on such cases often include relatively small numbers of studies, the selection of which may be subject to bias. To provide more robust conclusions, systematic reviews of the human clinical or toxicological utility of large numbers of animal experiments are necessary.

Systematic reviews may seek answers to a wide variety of clinical or research questions, including those relating to the human clinical or toxicological utility of animal models. The conclusions of systematic reviews are strongly evidence-based and provide a very high degree of reliability, for several reasons. Ideally, they comprehensively and systematically search the published scientific literature for all relevant data relating to

the research question(s), such as individual animal studies. They may screen search results for quality. Studies of low methodological quality, for example, may be excluded. Finally, all relevant data is collated and analyzed, sometimes statistically (meta-analysis), to provide answers about the research question(s) (Hooijmans et al., 2010).

In 2004 Pound and colleagues noted that clinicians and the public often consider it axiomatic that animal research has contributed to human clinical knowledge, based on anecdotal evidence or unsupported claims. These constitute an inadequate form of evidence, they asserted, for such a controversial area of research, particularly given increasing competition for scarce research resources. Hence, they called for systematic reviews examining the human clinical utility of animal models. They commenced by examining six existing reviews, which did not demonstrate the clinical utility expected of the experiments in question. Similar calls for systematic reviews examining the human clinical utility of animal models have since been made by others (Nuffield Council on Bioethics, 2005).

Since then, at least 27 systematic reviews examining the utility of animal experiments in advancing human clinical outcomes (20) or in deriving human toxicity classifications (7) have been described (Knight, 2007a, 2008a, 2011). Three different approaches, each of which sought to determine the maximum human clinical utility achievable by animal models, are of particular interest.

Experiments expected to lead to medical advances

Lindl and colleagues (2005) examined animal experiments conducted at three German universities between 1991 and 1993. The studies had been approved by animal ethics committees at least partly on the basis of specific claims by researchers that the experiments might lead to concrete advances toward the cure of human diseases. Experiments were included only where they had achieved publication in biomedical journals and where previous, related research had confirmed the hypotheses of the researchers.

For 17 experiments, citations were analyzed over a period of at least 12 years. Citation frequencies and types of citing papers were recorded: whether reviews, animal-based, *in vitro*, or clinical studies. Of the 1,183 citations, only 8.2% (97) were in clinical publications, and of these, only 0.3% of all citations (four publications) demonstrated a direct correlation between the results of animal experiments and human outcomes. Even in these four cases, however, the hypotheses that had been verified successfully in animal experiments failed when applied to humans. None of these 17 experiments led to any new therapies or demonstrated any beneficial clinical impact during the period examined. Extension of the study period from 2005 to March 2011 yielded similar results (Lindl and Voelkel, 2011).

Highly cited animal studies

Hackam and Redelmeier (2006) noted that highly cited animal experiments are more likely to be subsequently tested in clinical trials. Accordingly, they searched for experiments with more than 500 citations published in seven leading scientific journals.

Of 76 animal studies with a median citation count of 889 (range: 639-2,233), only 36.8% (28/76) were replicated in randomized human trials, while 18.4% (14/76) were contradicted by randomized trials, and 44.7% (34/76) had not translated to clinical trials, despite allowing a median of 14 years for potential translation. Ultimately, only 10.5% (8/76) of these medical interventions were subsequently approved for use in patients. And even in these cases, human benefit cannot be assumed because adverse reactions to approved interventions are not uncommon. Indeed, they have been assessed as the 4th-6th leading cause of death (based on a 95% confidence interval) in US hospitals (Lazarou and Pomeranz, 1998).

Furthermore, the selective focusing on positive animal data while ignoring negative results (optimism bias) is one of several cited factors that may have increased the likelihood of translation beyond that scientifically merited. As Hackam (2007) stated, rigorous meta-analysis of all relevant animal experimental data probably would significantly decrease the translation rate to clinical trials.

Disturbingly, only 48.7% (37/76) of these highly-cited animal studies published in leading scientific journals were of good methodological quality. Few included random allocation of animals to treatment groups or blinded assessment of outcomes, although it is well established that studies lacking randomization or blinding often over-estimate the magnitude of treatment effects (Poignet et al., 1992; Aronowski et al., 1996; Marshall et al., 2000). Accordingly, Hackam and Redelmeier cautioned patients and physicians about extrapolating the findings of even highly-cited animal research to the care of human disease.

Chimpanzee studies

Chimpanzees are the species most closely related to humans, and consequently, most likely to be generally predictive of human outcomes when used in biomedical research. Accordingly, in 2005 I conducted a citation analysis examining the human clinical utility of invasive chimpanzee studies (Knight, 2007b, 2008b).

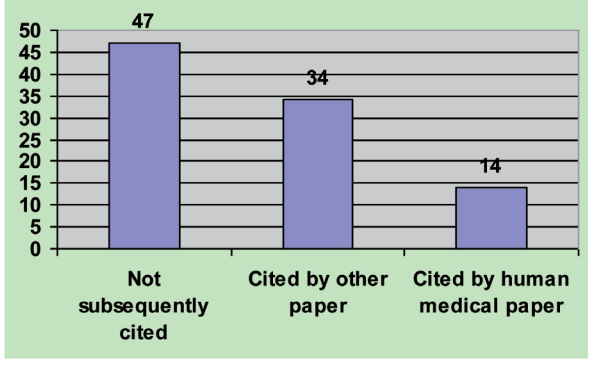

Fig. 1: Citations of 95 randomly-selected published chimpanzee studies

I examined a statistically significant sample of 95 studies randomly selected from a population of 749 papers describing invasive experiments on captive chimpanzees or their tissues, published in the peer-reviewed scientific literature between 1995 and 2004. Of these, 49.5% (47/95) were not cited by any subsequent papers (Fig. 1). Research of lesser value is not published at all; hence it appears that the majority of chimpanzee research generates data of questionable value, which makes little obvious contribution toward the advancement of biomedical knowledge. The publication year did not appear to affect this outcome substantially, as citation frequencies were similar across the decade, with more recent papers cited approximately as often as older papers.

A total of 35.8% (34/95) of these published chimpanzee studies were cited by 116 papers that clearly did not describe well developed methods for combating human diseases. Only 14.7% (14/95) of these chimpanzee experiments were cited by 27 papers with abstracts indicating well developed methods for combating human diseases. However, detailed examination of these medical papers revealed that *in vitro* studies, human clinical and epidemiological studies, molecular assays and methods, and genomic studies, contributed most to their development. In 63.0% (17/27) of these papers, the references were wide-ranging reviews of 26-300 (median 104) papers to which the cited chimpanzee study made a very small contribution. Duplication of human outcomes, inconsistency with human or other primate data, and several other causes resulted in the absence of any chimpanzee study able to demonstrate an essential contribution, or, in most cases, a significant contribution of any kind, toward the development of the medical method described.

Overall results

In only two of 20 systematic reviews examining clinical utility did the animal models appear substantially consistent with human outcomes. Due to a very limited sample size, one of these conclusions was contentious. Of seven systematic reviews examining toxicological utility, none demonstrated reliable predictivity of human outcomes such as carcinogenicity and teratogenicity (Knight 2007a, 2008a, 2011).

3 Methodological quality of animal studies

The poor human utility of animal models during toxicity assessment and the development of clinical interventions may be partly attributable to the poor methodological quality and statistical design of many animal studies. At least 11 systematic reviews have demonstrated substantial methodological flaws within many of the animal experiments examined, and none has demonstrated good methodological quality of a majority of experiments. Common deficiencies include lack of: sample size calculations, sufficient sample sizes, randomized allocation to treatment and control groups, blinded drug administration, blinded induction of injury (such as ischemia in the case of stroke models), blinded outcome assessment, and conflict of interest statements. Some studies also used anesthetics that may have altered experimental outcomes, and substantial variation was evident in the parameters assessed (Knight 2007a, 2008a, 2011).

Strategies designed to improve methodological quality would minimize the consumption of financial and scientific resources and animal lives in studies of questionable merit and quality. However, the poor human clinical or toxicological utility of many animal experiments is unlikely to result solely from methodological flaws. Several intrinsic characteristics of animal models also markedly limit their human predictivity (Knight 2007a, 2008a, 2011). As stated by Perel and colleagues (2007), such obstacles could be technically and theoretically insurmountable.

4 The Bateson *Review of Research Using Non-Human Primates*

Overestimation of human clinical utility during cost-benefit assessment of animal studies appears to be widespread. This was exemplified by the 2011 *Review of Research Using Non-Human Primates* (NHPs) by Bateson and colleagues. In 2006 a UK working group recommended that the major organizations funding NHP research should undertake a systematic review of its outcomes, given the advanced cognitive and social characteristics of NHPs, and the resultant ethical problems that arise when they are subjected to invasive research.

The Review Panel considered all NHP research funded by these organizations from 1997 to 2006, inclusive. It sought to assess the scientific importance of each research project, the probability of medical and public benefit, and the likelihood of animal suffering. It concluded that in many cases the use of NHPs was justifiable, although the Panel was concerned about approximately 9% of research programs, from which no clear scientific, medical, or social benefit had emerged. For several important reasons, however, the proportion of cases that were ethically justifiable was probably far lower than 91%.

First, the Panel appears to have considered animal experiments that contribute to scientific advancements in general to be of similar importance to those providing medical benefits. However, all experiments that advance knowledge arguably contribute to the advancement of science. Therefore, they would all be ethically justifiable, if they were cost-free. But of course they are not. The very substantial costs incurred by invasive NHP research include animal welfare costs, the consumption of considerable financial and scientific resources and even potentially adverse impacts on patients and consumers when human results prove different from those in NHPs. Therefore, experiments that simply contribute to the advancement of science, without providing any specific, credible medical benefit – which comprise a substantial proportion of the 91% condoned by the Review Panel – should not have been assessed as ethically justifiable in the absence of exceptional, additional justification.

The Panel stated that, *"In most cases… little direct evidence was available of actual medical benefit in the form of changes in clinical practice or new treatments."* And, *"This contrasts with the emphatic public statements about the medical benefits of NHP research made by some of the funding bodies and by grant applicants."*

They recommended that: *"In their public engagement, the funders and researchers should avoid overstating and generalizing the medical benefit of NHP research, since this cannot be substantiated in many cases."* And, *"It is important that the justifications offered for research projects are soundly based and demonstrable. Health benefits should only be claimed when their potential is real."*

5 Laboratory animal impacts

The costs to animals who participate in invasive research protocols are also widely underestimated. A wide variety of stressors have the potential to cause significant stress, fear, and sometimes distress in laboratory animals. These may be associated with the capture of wild-sourced species to supply laboratories or breeding centers, transportation (which may be prolonged for some animals), laboratory housing and environments, and both routine and invasive laboratory procedures (Knight, 2011).

Stress caused by wild capture, transportation, and invasive procedures is relatively well acknowledged. Less recognized, however, is the stress caused by routine laboratory procedures such as handling, blood sampling, and gavaging (the insertion of an esophageal tube for the oral administration of test compounds, which is not uncommon during toxicity studies), as well as by standardized laboratory housing. However, a large body of published evidence indicates that such routine laboratory procedures may result in profound, statistically significant distortions in a range of physiological parameters, including cardiovascular parameters and serum concentrations of glucose and various hormones. Similarly, standardized laboratory housing may markedly alter behavior, with behavioral stereotypies and increased aggression developing over time. Alterations in neuroanatomical parameters, and even cognitive capacities, also have been documented (Balcombe et al., 2004; Balcombe, 2006; Baldwin and Bekoff, 2007).

In addition to creating significant animal welfare and ethical problems, such procedures and environments may distort a wide range of experimental outcomes, including those dependent on accurate determination of physiological, behavioral, or cognitive parameters.

The Bateson Review described previously provides a recent example in which the costs to the research animals were systematically underestimated. Of 31 neuroscience studies examined, for example, half were assessed as having imposed a high welfare impact on the animals used. Disturbingly high though this is, unfortunately it appears to have been an underestimate. Studies involving the surgical creation of experimental lesions (that is, tissue or organ damage) were not assessed as having a high welfare impact, unless "significant and lasting impairments to the monkeys' welfare" resulted. Therefore, surgical or other procedures causing significant damage that was not prolonged and healed within a normal post-operative timeframe would probably have received a moderate impact rating, even though such procedures may severely impact animal welfare.

Additionally, studies conducted under terminal anesthesia received a low welfare impact rating. This may have been technically correct from a strict welfare perspective, if minimal suffering resulted. However, it makes no allowance at all for the serious moral issues raised by killing. Laboratory animals are not simply expendable. Their lives have intrinsic value, and the loss of those lives is morally significant. The killing of these primates should not have been so lightly dismissed simply because they occurred under anesthesia. The intrinsic value of these animals' lives should have been included in the moral calculus.

6 Reforming policy

Accurate weighting of the probable harms incurred by laboratory animals against the likely benefits of such research is ethically necessary, expected by society, and required by applicable regulations. In practice, however, expected human benefits are frequently based on unrealistic assumptions, and the costs borne by laboratory animals are widely underestimated. A range of policy initiatives are warranted to address these deficiencies and to increase the implementation of the 3Rs generally (Knight, 2011). Key initiatives include:

Animals protected

Regulatory protection should be based on current scientific knowledge about neuroanatomical architecture, cognitive, psychological, and social characteristics, and consequent capacity for suffering in laboratory environments and protocols. Sufficient scientific evidence exists to warrant the protection of living vertebrates, including advanced larval forms and fetal developmental stages, as well as certain invertebrates such as cephalopods. In contrast, the US excludes mice, rats, birds, fish, reptiles, and amphibians from protection under the Animal Welfare Act, effectively eliminating well over 90 per cent of animals used in scientific procedures (USDA, 2005; Taylor et al., 2008).

Animals used to develop or maintain genetically modified (GM) strains, bred for organ or tissue harvesting, and bred or intended for laboratory use, including those killed when surplus to requirements, should similarly be protected, with neither their use nor their killing excluded from the ethical review and regulatory control afforded to other laboratory animals.

Species and procedures associated with high welfare risks

The advanced psychological and social characteristics of great apes render it impossible in practical terms to provide laboratory environments that satisfactorily meet their minimum psychological and behavioral requirements, which include family

preservation, ample opportunities for climbing, exploring, problem solving, and playing, as well as considerable space (Balls, 1995; DeGrazia, 1996; Smith and Boyd; 2002). Accordingly, their use should be prohibited, and remaining primate use very carefully scrutinized, as should other categories of animal use that pose particularly high welfare risks, such as terminal or surgical procedures, those involving major physiological challenges, and the production of GM animals. In particular, animal use should be prohibited where pain, suffering, or distress is likely to be severe or long-lasting.

Scrutiny of scientific animal use

In countries where this is not yet the norm, proposed experimental protocols should be subjected to independent ethical and scientific review prior to licensing. In all regions, the likely human benefits of scientific animal use should be scrutinized far more critically than is currently the norm and more accurately weighed against the animal, human, and financial costs incurred. Searches for replacement, reduction, and refinement methodologies should be thorough, and where scientifically suitable alternatives are identified, they should be used, rather than merely considered, as remains the current US requirement.

The societal values attached to laboratory animal lives and to the health and safety of patients and consumers are considerable, and the corresponding public interest is substantial. In recognition of the legitimacy of such public interest, and to facilitate critical review, proposed experimental protocols should also be made available for independent scientific and public scrutiny, with a mechanism provided for feedback to the ethical review committee.

Retrospective evaluation

To minimize unwarranted experimental duplication, and consistent with legitimate public interest, study results should be made publicly available in a timely fashion. To assess the degree to which experimental objectives were successfully met, the costs incurred by research animals, and to inform future research strategy and further experimental licensing decisions, retrospective evaluation of experiments should be mandatory where such experiments are considered likely to result in significant costs to laboratory animals or to public finances, or significant human benefits.

References

Aronowski, J., Strong, R., and Grotta, J. C. (1996). Treatment of experimental focal ischemia in rats with lubeluzole. *Neuropharmacology 35*, 689-693.

Balcombe, J., Barnard, N., and Sandusky, C. (2004). Laboratory routines cause animal stress. *Contemp. Top. Lab. Anim. Sci. 43*, 42-51.

Balcombe, J. (2006). Laboratory environments and rodents' behavioural needs: a review. *Lab. Anim. 40*, 217-235.

Baldwin, A. and Bekoff, M. (2007). Too stressed to work. *New Sci. 194*, 24.

Balls, M. (1995). Chimpanzee medical experiments: moral, legal, and scientific concerns. In Poor model man: Experimenting on chimpanzees: Proceedings of the first PACE (People Against Chimpanzee Experiments) conference on the use of chimpanzees in biomedical research. *ATLA 23*, 607-614.

Bateson, P., Johansen-Berg, H., Jones, D. K., et al. (2011). *Review of research using non-human primates*. London: Biotechnology and Biological Sciences Research Council, Medical Research Council and Wellcome Trust. www.bbsrc.ac.uk/web/FILES/Reviews/review-research-using-nhps.pdf (accessed 29.07.2011).

Canadian Council on Animal Care – CCAC (1997). CCAC guidelines on: animal use protocol review. www.ccac.ca/Documents/Standards/Guidelines/Protocol_Review.pdf (accessed 08.09.2011).

DeGrazia, D. (1996). *Taking animals seriously: Mental life and moral status*. Cambridge: Cambridge University Press.

European Union – EU (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official J. EU 276*, 33-79.

Hackam, D. G. and Redelmeier, D. A. (2006). Translation of research evidence from animals to humans. *J. Am. Med. Assoc. 296*, 1731-1732.

Hackam, D. G. (2007). Translating animal research into clinical benefit: poor methodological standards in animal studies mean that positive results may not translate to the clinical domain. *Br. Med. J. 334*, 163-164.

Hooijmans, C. R., Leenaars, M., and Ritskes-Hoitinga, M. (2010). A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three Rs, and to make systematic reviews more feasible. *ATLA 38*, 167-182.

Knight, A. (2007a). Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. *ATLA 35*, 641-659.

Knight, A. (2007b). The poor contribution of chimpanzee experiments to biomedical progress. *J. Appl. Anim. Welf. Sci. 10*, 281-308.

Knight, A. (2008a). Systematic reviews of animal experiments demonstrate poor contributions toward human healthcare. *Rev. Recent Clin. Trials 3*, 89-96.

Knight, A. (2008b). The beginning of the end for chimpanzee experiments? *Philos. Ethics Humanit. Med. 3*, 16.

Knight, A. (2011). *The Costs and Benefits of Animal Experiments*. Basingstoke, UK: Palgrave Macmillan.

Lazarou, J. and Pomeranz, B. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *J. Am. Med. Assoc. 279*, 1200-1205.

Lindl, T., Völkel, M., and Kolar, R. (2005). Tierversuche in der biomedizinischen Forschung. Eine Bestandsaufnahme der klinischen Relevanz von genehmigten Tierversuchsvorhaben (Animal experiments in biomedical research. An evaluation of the clinical relevance of approved animal experimental projects). *ALTEX 22*, 143-151.

Lindl, T. and Völkel, M. (2011). No clinical relevance of approved animal experiments after seventeen years. *ALTEX 28*, 242.

Marshall, J. W., Cross, A. J., Jackson, D. M., et al. (2000). Clomethiazole protects against hemineglect in a primate model of stroke. *Brain Res. Bull. 52*, 21-29.

Nuffield Council on Bioethics (2005). *The ethics of research involving animals*. London: Nuffield Council on Bioethics.

Perel, P., Roberts, I., and Sena, E. (2007). Comparison of treatment effects between animal experiments and clinical trials: systematic review. *Br. Med. J. 334,* 197-200.

Poignet, H., Nowicki, J. P., and Scatton, B. (1992). Lack of neuroprotective effect of some sigma ligands in a model of focal cerebral ischemia in the mouse. *Brain Res. 596*, 320-324.

Pound, P., Ebrahim, S., Sandercock, P., et al. (2004). Where is the evidence that animal research benefits humans? *Br. Med. J. 328*, 514-517.

Smith, J. A. and Boyd, K. M. (eds.) (2002). *The Boyd Group Papers on the use of non-human primates in research and testing*. Leicester, UK: The British Psychological Society.

Taylor, K., Gordon, N., Langley, G., et al. (2008). Estimates for worldwide laboratory animal use in 2005. *ATLA 36*, 327-342.

US Dept of Agriculture, Animal and Plant Health Inspection Service, Animal Care – USDA (2005). FY 2005 AWA Inspections. www.aphis.usda.gov/animal_welfare/downloads/awreports/awreport2005.pdf (accessed 10.03.2010).

Acknowledgements

My presentations at the Eighth World Congress on Alternatives and Animal Use in the Life Sciences were funded by the conference organizers. I am grateful for their support and the invitation to speak. Figure 1 is reprinted with permission of the publisher (Taylor & Francis Ltd, http://www.tandf.co.uk/journals), from: Knight, A. (2007b).

Correspondence to

Andrew Knight PhD, CertAW, MRCVS,
DipECAWBM-AWSEL, FOCAE
Oxford Centre for Animal Ethics
www.AnimalConsultants.org