ANIMALS IN RESEARCH

Dr. Andrew Knight

FRIDAY, 21 MAY 2021

14.00 – 15.15 (CEST)

AK-01-02

CRITICALLY EVALUATING THE USE OF ANIMAL MODELS WITHIN RESEARCH (#129)

Prof. Andrew Knight

University of Winchester, Centre for Animal Welfare, Winchester, UK

Abstract

Widespread reliance on animal models of humans during preclinical research and toxicity testing assumes their reasonable predictivity for human outcomes. However, of 20 published systematic reviews examining human clinical utility located during a comprehensive literature search, animal models demonstrated significant potential to contribute toward clinical interventions in only two cases, one of which was contentious. Included were experiments expected by ethics committees to lead to medical advances, highly-cited experiments published in major journals, and chimpanzee experiments—the species most generally predictive of human outcomes. Seven additional reviews failed to demonstrate utility in reliably predicting human toxicological outcomes such as carcinogenicity and teratogenicity. Results in animal models were frequently equivocal, or inconsistent with human outcomes. Consequently, animal data may not be assessed as reliably predictive of human outcomes. Barriers to human predictivity appear to arise from: interspecies differences, stressful environments and protocols, chronic high dose rodent studies, flaws within research design, conduct and reporting, and the effects of publication bias.

Conflict of Interest

I do not declare any conflict of interest.

Animal model utility within human healthcare advancement

Advocates of animal models within human-focused research have regularly claimed such animal use is essential for preventing, curing, or alleviating human diseases (e.g., Brom, 2002; Festing, 2004); and even that the greatest achievements of medicine have only occurred through the use of animals (e.g., Pawlik, 1998). However, counter-narratives by others contest the contributions or necessity of such research for the advancement of medical progress (e.g., Greek and Greek, 2002). To support their argument, advocates on either side regularly cite cases in which animal and human outcomes are similar or different. However, only small numbers of experiments are normally included in such reviews, and their selection may be subject to bias. These are known as *narrative reviews*.

To provide more definitive conclusions, *systematic reviews* of the human clinical or toxicological utility of large numbers of animal experiments are necessary. A systematic review is “a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies” (Moher et al., 2009).

A large number of systematic reviews of animal experiments within various research fields have examined their utility for advancing human healthcare, and the results have not been good. Of 20 published systematic reviews examining human-clinical utility located during a comprehensive literature search, animal models demonstrated significant potential to contribute toward clinical interventions in only two cases, one of which was contentious. Included were experiments approved by ethics committees on the basis of claims that

medical advances were likely to result; highly-cited experiments published in leading journals; and chimpanzee experiments, utilizing the species most generally predictive of human outcomes. Seven additional reviews failed to demonstrate utility in reliably predicting human toxicological outcomes, including those associated with the greatest public health concerns, such as carcinogenicity and teratogenicity. Results in animal models were frequently equivocal or inconsistent with human outcomes (Knight, 2011). Since then, numerous additional reviews have yielded similar results.

Limitations of animal models

A variety of factors appear responsible for the poor rates of translation of outcomes from animal studies into human patients and consumers. These relate both to the animal models themselves and to the ways in which they are used. Fundamental biochemical differences between species may result in differences in absorption, distribution, metabolism, and elimination pathways or rates, which may alter *toxico-or pharmaco-kinetics* (i.e., bodily distribution). *Toxico- and pharmaco-dynamics* (mechanisms of action and biological effects) may also be altered. Jointly these factors may contribute to differences in organ systems affected and in the nature and magnitude of those effects (Hartung, 2008; Knight, 2011).

Further problems arise from the characteristics of the animals used. Biological variability and predictivity for humans are frequently compromised by restriction to single rodent strains, young animals, and single sexes, usually without concurrent human risk factors, such as common comorbidities, that can alter human responses to exogenous compounds (Hartung, 2008; Knight, 2011).

Additional problems arise from the ways in which the animals are used. Many toxicity tests, for example, rely on *maximum tolerated doses* (above which acute, toxicity-related effects preclude further dosing), and chronic dosing. These factors maximize sensitivity to toxins, with the result that false negative results rarely occur. However, these conditions can also overwhelm the physiological defences that are effective at environmentally realistic doses, resulting in false positive outcomes. As a result, many compounds that would not normally be considered toxic are falsely indicated as such by animal tests; this substantially decreases the reliability and relevance of any positive result. Additionally, important human routes of exposure (e.g., inhaled) may differ from those tested in animals, requiring extrapolation between routes of exposure, as well as between species, introducing further uncertainty (Gold, Slone and Ames, 1998; Hartung, 2008; Knight, 2011).

Furthermore, animals used in laboratories commonly experience a significant array of stressors. These include stresses incurred during handling, restraint, and other routine laboratory procedures; and, in particular, the stressful routes of dose administration common to toxicity tests. Orogastric gavaging, for example, involves the insertion of a tube into the oesophagus for the forced administration of test compounds. Combined with environmental stressors (e.g., due to limited space and environmental enrichment) and social stressors (e.g., due to aggressive interactions between conspecifics), these represent a significant body of stressors. These stressors can alter physiological, hormonal, and immune statuses and even cognitive capacities and behavioural repertoires, in ways that are not always predictable (Balcombe, Barnard and Sandusky, 2004; Balcombe, 2006; Baldwin and Bekoff, 2007). The results may include alterations in the progression of diseases, in bodily responses to chemicals and test pharmaceuticals, and in a range of other scientific outcomes, such as those dependent on accurate determination of physiological, behavioural, or cognitive characteristics.

Methodological quality of animal studies

As if these were not problem enough, a sizeable body of recent studies and systematic reviews have confirmed the existence of significant methodological flaws, in most published animal experiments (e.g., Knight, 2008b). Indeed, to date, no systematic reviews appear to have been published in which a majority of animal studies, assessed against appropriate objective criteria, were found to have been of good methodological quality. In particular, a variety of design features must be included within animal experiments to minimize the potential for bias. Hooijmans et al. (2014) described 10 types of bias that have the potential to influence animal experimental results, which they grouped into selection bias, performance bias, detection bias, attrition bias, reporting bias, and other sources of bias. Many of these flaws are highly prevalent within animal studies. Common examples include lack of: statistical justification of sample sizes, statistically significant sample sizes, randomisation during treatment allocation, and blinding during outcome assessment. In numerous studies it has been demonstrated that those incorporating the fewest measures to minimize sources of bias, tended to report the greatest effect sizes (Crossley et al., 2008; Hirst et al., 2014; Macleod et al., 2005; Rooke et al., 2011; Vesterinen et al., 2010). Such alterations of effect sizes are artefacts, and are the results of flaws in experimental design, conduct or reporting.

In 2010, Kilkenny and colleagues proposed the *Animal Research: Reporting of In Vivo Experiments* (ARRIVE) guidelines. Prepared in consultation with scientists, statisticians, journal editors, and research funders, these guidelines comprise a checklist of 20 items, designed to provide minimum information on items such as: the number and specific characteristics of animals used (including species, strain, sex, and genetic background); housing and husbandry conditions; and the experimental, statistical, and analytical methods used. The latter points included measures to reduce bias, such as the random allocation of animals to experimental groups, blinded assessment of outcome measures, statistical justifications of sample sizes, reporting of animals excluded from analyses, exclusion criteria, and any investigator conflicts of interest. The intention was that these items should be included within all scientific publications reporting animal research, thereby allowing critical assessment of methods used and results obtained. Several similar guidelines have been published.

The ARRIVE guidelines of Kilkenny et al. (2010) have been published or endorsed by more than 1,000 research journals, including those published by the Nature Publishing Group, PLoS, and BioMed Central (Reichlin, Vogt and Würbel, 2016). They have been similarly endorsed by major UK funding agencies (including the Wellcome Trust, the Biotechnology and Biological Sciences Research Council, and the Medical Research Council); and they also form part of the US National Research Council Institute for Laboratory Animal Research guidelines (Baker et al., 2014). And yet, despite such widespread endorsement, a number of studies have demonstrated that compliance with such guidelines remains poor.

Improving study quality

A range of measures are strongly warranted to increase the implementation of the 3Rs principles, the methodological quality of animal research, the reliability of results, and to overcome some of the barriers that currently prevent reliable extrapolation to human outcomes.

Compliance with each of the 3Rs and the ARRIVE guidelines and other best practice standards, during the design, conduct, and reporting of experiments, must become mandatory. Such standards should cover animal sourcing, housing, environmental enrichment, socialization opportunities, appropriate use of anaesthetics and analgesics, handling, non-invasive endpoints, and a range of measures designed to minimize sources of bias and to ensure methodological quality. Compliance with such standards should be a necessary condition for securing research funding and ethical approval; licensing of researchers, facilities,

and experimental protocols; and publication of subsequent results. Compliance would also facilitate subsequent systematic reviews.

To enable animal researchers and technicians to meet the necessary standards, training and continuing professional development in 3Rs methodologies, and the design, conduct, and reporting of animal research, should be compulsory. The existing lack of focus on replacement methods (in favour of refinement methods) must be addressed.

The adoption of measures such as these, would increase the reliability of research results and would facilitate their use within systematic reviews. Prior to designing any new animal study, researchers should conduct a systematic review to collate, appraise, and synthesize all existing, good-quality evidence relating to their research questions. Such systematic reviews should be similarly required by grant agencies, ethical review committees, other animal-experiment licencing bodies, and journals. Systematic reviews are studies in and of themselves. In recognition of their intrinsic value, and their necessity for informing further research, they should also be readily funded by grant agencies.

To ensure that all such evidence is publicly available, greater efforts must also be made by researchers and editors to publish negative results. Studies that fail to show a treatment effect are often considered less interesting and are, consequently, less likely to be published. The subsequent exclusion of such results from systematic reviews leads to overestimations of treatment efficacy and partly explains the widespread failures in humans of treatments apparently efficacious in animals.

Many of these measures will require cooperation and coordination between researchers, regulators, licensing bodies, ethical review committees, funding bodies, journals, and authors. And of course, the necessary willingness, among all parties, to change.

Conclusions

Global laboratory animal use for all purposes was estimated at 192 million in 2015 (Taylor et al. 2019). A great many of these animals suffer a wide range of serious harms, or are killed. Systematic reviews clearly indicate the resultant benefits for human healthcare are low. When considering harms and benefits overall, one cannot reasonably conclude that the benefits for human patients or consumers, or for those motivated by scientific curiosity or profit, exceed the harms incurred by animals subjected to scientific procedures. On the contrary, evidence indicates that actual human benefit is rarely, if ever, sufficient to justify such harms. And those harms are not limited to the many millions of animals used. Others potentially affected include patients and consumers. The social and ethical implications are profound, when consumers suffer serious toxic reactions to products assessed as safe in animal studies, or if patients with serious conditions are denied effective clinical interventions, partly because potentially more efficacious research fields are under-resourced (Knight, 2011).

A paradigm change in scientific animal use is clearly warranted. Instead of uncritically assuming the benefits of animal research, we must subject it to much more rigorous and critical evaluation. Where animal research continues to persist, a broad range of measures must be implemented to improve substantially its methodological quality and compliance with the 3Rs and to maximise the reliability of subsequent results (Knight, 2011). When such research fails to meet the harm-benefit standards expected by society, which underpin legislative instruments, such as Directive 2010/63/EU, then such research should cease; and the resources consumed by it directed into more promising and justifiable fields of research and healthcare.

Acknowledgment, references and further information

These notes were excerpted from: Knight A (2019). Critically evaluating animal research. In Herrmann K and Jayne K (Eds). *Animal Experimentation: Working Towards a Paradigm Change.* Leiden, The Netherlands: Brill. 321-340. https://brill.com/view/book/edcoll/9789004391192/BP000019.xml, accessed 07/04/21. Further information and references are available there.