anzccart
HUMANE SCIENCE

CONFERENCE 2017

Maintaining social license in a changing world

QUEENSTOWN, NEW ZEALAND

Rydges Hotel
2-4 September 2017

Conference Proceedings 2017

EXPLORE | DISCOVER | SHARE

ROYAL SOCIETY TE APĀRANGI

c/o The Royal Society of New Zealand, PO Box 598 Wellington, 6140, New Zealand

ISBN: 978-1-877264-29-0

The Proceedings were edited by Mrs Gill Sutherland and Dr Marc Rands

Contents

Painful truths: what systematic reviews reveal about the utility of animal research

Andrew Knight
SAFE, PO Box 28110, Kelburn 6150, Wellington 6011, New Zealand

After standardising to match European Union (EU) definitions of animals and experimental procedures, it was estimated that 127 million living non-human vertebrates were used worldwide for scientific and educational purposes in 2005. This remains the most robust, evidence-based global estimate available (Knight 2008a; Taylor *et al.* 2008). The figures for Australia and New Zealand were 2.4 million and 261,000 respectively, making them the fourth and 28th-largest national users of laboratory animals in 2005 (Knight 2013).

The most recent figures at the time of writing described 2015. 9.9 million animals were used in Australia (HRA 2016), and 225,000 animals were used in New Zealand (MPI 2016), although these latter figures have not been standardised to match EU definitions. The Australian figure, for example, was increased by NSW counting 4.1 million native animals used in environmental studies which involved observation only (HRA 2016). Clearly, very large numbers of animals continue to be used within Australian and New Zealand research.

Additionally, animal research incurs other costs. The very substantial financial and scientific resources consumed by animal research are consequently unavailable to other fields, some of which – such as preventative healthcare or human clinical research – might well be expected to produce substantial public health benefits.

Ongoing societal approval for the use of these animals and research resources rests on the principle that the subsequent benefits are substantial, and represent the best use of limited research resources. However, the best available evidence indicates that much animal research fails to meet these standards.

Clinical and toxicological predictivity of animal research

A large number of systematic reviews of animal research have examined its utility for advancing human healthcare. Of 20 published systematic reviews examining human clinical utility located during a comprehensive search, animal models demonstrated significant potential to contribute toward clinical interventions in only two cases, one of which was contentious. 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).

Numerous additional reviews have since yielded similar results. Baker and colleagues (2014) examined human neurological diseases. Extensive animal studies have yielded relatively few human treatments (Cheeran *et al.* 2009; Vesterinen *et al.* 2010). Similarly, despite the efficacy of over 1,000 treatments in animal models of multiple sclerosis, very few have progressed to the marketplace (Vesterinen *et al.* 2010). This usually indicates concerns about human safety or efficacy. Numerous other examples exist (e.g. stroke studies: Cheeran *et al.* 2009).

Limitations of animal models

A variety of factors appear responsible for poor translation of animal outcomes into human patients. These limitations arise both from the animal models themselves, and from the ways in which they are used.

Fundamental biochemical differences result in interspecies differences in absorption, distribution, metabolism, and elimination pathways or rates, which may alter *toxico- or pharmacokinetics* (i.e. bodily distribution). *Toxico- and pharmacodynamics* (mechanisms of action and biological effects) may be similarly affected. Jointly these factors may alter organ systems that are impacted, and the nature and magnitude of those effects (Hartung 2008; Knight 2011).

Biological variability and predictability for diverse human populations are frequently compromised by restriction to single rodent strains, young animals, and single sexes. Common human co-morbidities and lifestyle risk factors are usually lacking (Hartung 2008; Knight 2011).

Additionally, many toxicity tests rely on *maximum tolerated doses* (above which acute toxicity-related effects preclude further dosing), and chronic dosing. Whilst maximising sensitivity to toxins, thereby minimising false negative results, these conditions can also overwhelm physiological defences effective at more environmentally realistic doses, resulting in false positive outcomes (Gold *et al.* 1998; Hartung 2008; Knight 2011).

Furthermore, animals used in laboratories commonly experience a significant array of stressors incurred during handling, restraint, and other routine laboratory procedures, and particularly, the stressful routes of dose administration common to toxicity tests. 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 can alter physiological, hormonal, and immune status, and even cognitive capacities and behavioural repertoires, in ways which are not always predictable (Balcombe *et al.* 2004; Balcombe 2006; Baldwin & Bekoff 2007).

Flaws of study design and conduct

Additionally, numerous recent studies and systematic reviews have confirmed the existence of significant methodological flaws, in most published animal experiments (e.g. Knight 2008b). Indeed, no systematic reviews have demonstrated that a majority of animal studies, when assessed against appropriate objective criteria, were of good methodological quality.

In particular, a number of design features must be included within animal experiments, to minimise the potential for bias. Hoojimans *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. Kilkenny and colleagues (2009) conducted one of the largest and most comprehensive systematic surveys to date, assessing the experimental design, statistical analysis and reporting of 271 published animal experiments. Some were funded by leading grant agencies within the United Kingdom and United States.

Details such as animal strain, sex, age and weight are all scientifically important and can potentially influence results (Obrink & Rehbinder 2000; Alfaro 2005). Nevertheless, in many cases these were omitted.

Knowledge of planned treatment (or lack thereof) is one of a number of factors that can unconsciously influence the assignment of animals to treatment groups. Accordingly, randomised selection of animals for treatment groups is mandated, to ensure that outcome differences are most likely due to treatment effects (Festing & Altman 2002; Festing *et al.* 2002). Nevertheless, such randomisation was reported in only 12% of these studies.

Another crucial feature of good experimental design concerns the assessment of outcomes. Where qualitative judgements occur, it is crucial that assessors are blinded to the treatment (or lack, thereof), of animals assessed – lest such knowledge subtly affects their judgement (Festing & Altman 2002). Nevertheless, only 14% of all papers that reported qualitative assessment of outcomes, also reported the use of blinding. More recently, similarly low rates of measures designed to minimise bias were found in an even larger study (Vogt *et al.* 2016).

Many factors can affect experimental outcomes, so the incorporation of measures to minimise sources of bias are crucial to ensuring the reliability of research results. Animal research reviews from the field of emergency medicine have demonstrated that estimates of treatment efficacy are significantly reduced in studies that incorporate mechanisms to reduce risks of bias (Bebarta *et al.* 2003; Macleod *et al.* 2008). Similar results have been found in numerous other studies. Animal studies incorporating the fewest measures to minimise bias tend to report the greatest effect sizes, demonstrating that such effects are not entirely real, and are partly due to bias (Macleod *et al.* 2005; Crossley *et al.* 2008; Vesterinen *et al.* 2010; Rooke *et al.* 2011; Hirst *et al.* 2014). The widespread failure to utilise mechanisms such as randomisation and blinding appears to result in false expectations of treatment efficacy, with the results that reported outcomes in animals often fail to translate into humans.

Another problem commonly observed by Kilkenny et al. (2009) concerned the transparency of reporting, and the robustness of statistical analysis. Almost 60% of surveyed publications were deficient in these areas. Most studies failed to provide sample sizes, or adequate justifications of these. And yet, studies using too many animals waste lives. Conversely, the results of underpowered studies (with insufficient experimental subjects) cannot be extrapolated to wider populations with sufficient certainty. Accordingly, power analyses or other simple calculations are widely used in human clinical trials, to ensure sufficient subjects (but few extras) are present, to be able to detect biologically important effects. The same principles should apply to animal studies (Dell *et al*. 2002; Festing & Altman 2002).

Improving research quality

In 2010 Kilkenny and colleagues proposed the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines. These comprised a checklist of 20 items, designed to provide minimum information on experimental variables such as the number and characteristics of animals used (such as species, strain, sex, and genetic background); housing and husbandry conditions; and the experimental, statistical, and analytical methods employed. Multiple measures to reduce bias were listed, including random allocation of animals to experimental groups, blinded outcome assessment, statistical justifications of sample sizes, and the reporting of animals excluded from analyses, exclusion criteria, and any investigator conflicts of interest.

Kilkenny et al. proposed that these items should be included within all scientific publications reporting animal research, thereby allowing critical assessment of methods and results. Other authors have proposed similar guidelines and checklists (e.g. Hoojimans et al. 2010).

The ARRIVE Guidelines have since been endorsed by over 1,000 research journals (including those published by the Nature Publishing Group, PLoS, and BioMed Central) (Reichlin *et al.* 2016). They have been 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).

Despite this, a number of studies have demonstrated that compliance with such guidelines remains poor (Baker & Amor 2012; Landis *et al.* 2012; Schwarz *et al.* 2012; Reichlin *et al.* 2016).

Compliance with each of the 3Rs, and with the ARRIVE guidelines and other best practice standards, during the design, conduct and reporting of experiments, should be mandatory. Standards should cover animal sourcing, housing, environmental enrichment, socialisation opportunities, appropriate use of anaesthetics and analgesics, handling, non-invasive endpoints, and a range of measures to minimise bias and ensure methodological quality. Full compliance should be necessary for securing research funding, ethical approval, licencing of researchers, facilities and experimental protocols, and publication of subsequent results.

Measures such as these would all increase the reliability of research results, and would facilitate their use within systematic reviews. It might allow us to accurately predict treatment effects within the animal species under study, and to address the current inability to reproduce many animal study results (Reichlin et al. 2016).

However, interspecies differences will still remain in absorption, distribution, metabolism, and elimination pathways or rates, resulting in differing toxico- or pharmaco- kinetics and -dynamics, and subsequently, differences in the organ systems affected, and in the nature and magnitude of those effects. Such factors, which reflect the intrinsic complexity of living organisms, will continue to pose barriers to extrapolation to humans, that may remain insurmountable, in many cases.

Conclusions

Animals are rarely responsible for human health or societal challenges, many of which are of our own making and preventable. Animal advocacy organisations such as SAFE, along with numerous animal ethicists (e.g. Regan 1987; Nobis 2011), do not consider it ethical to harm animals in our attempts to address these.

Nevertheless, millions of animal lives are annually consumed by animal research, along with very substantial research and financial resources, which are subsequently unavailable for human clinical or other research fields. Inaccurate human predictions resulting from poorly designed animal studies threaten patient and consumer safety, delay the development of efficacious clinical interventions, and deny potentially useful chemicals to society.

The essence of the scientific method is a willingness to engage in critical scrutiny - even of one's own practice. Instead of uncritically assuming the benefits of animal research, researchers should subject it to much more rigorous and critical evaluation. Poorly designed, conducted and reported animal research should never be considered acceptable. A broad range of measures should be implemented to substantially improve methodological quality and 3Rs compliance, and to maximise reliability of subsequent results (Knight 2011).

Social license to conduct animal research depends on ensuring that the societal benefits exceed its very substantial costs. Where such research fails to meet the harm-benefit standards expected by

society it should clearly cease, with resources directed into more promising and justifiable fields of research and healthcare.

References

Alfaro V (2005). Specification of laboratory animal use in scientific articles: Current low detail in the journal's instructions for authors and some proposals. *Methods and Findings in Experimental and Clinical Pharmacology, 27(7)*, 495–502.

Baker D and Amor S (2012). Publication guidelines for refereeing and reporting on animal use in experimental autoimmune encephalomyelitis. *Journal of Neuroimmunology, 242,* 78–83.

Baker D, Lidster K, Sottomayor A, and Amor S (2014). Two years later: journals are not yet enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies. *PLoS Biol, 12(1),* e1001756.

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

Balcombe J, Barnard N and Sandusky C (2004). Laboratory routines cause animal stress. *Contemporary Topics in Laboratory Animals Science, 43,* 42–51.

Baldwin A and Bekoff M (2007). Too stressed to work. *New Scientist, 194(2606),* 24.

Bebarta V, Luyten D and Heard K (2003). Emergency medicine animal research: does use of randomisation and blinding affect the results? *Academic Emergency Medicine, 10(6),* 684–687.

Cheeran B, Cohen L, Dobkin B, Ford G, Greenwood R, Howard D, Husain M, Macleod M, Nudo R, Rothwell J and Rudd A (2009). The future of restorative neurosciences in stroke: driving the translational research pipeline from basic science to rehabilitation of people after stroke. *Neurorehabilitation and Neural Repair, 23(2),* 97–107.

Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, Macleod M and Dirnagl U (2008). Empirical evidence of bias in the design of experimental stroke studies. A metaepidemiologic approach. *Stroke*, *39(3),* 929–934.

Dell RB, Holleran S and Ramakrishnan R (2002). Sample size determination. *ILAR Journal, 43(4),* 207–213.

Festing MF and Altman DG (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR Journal, 43(3),* 244–258.

Festing MF, Overend P, Gaines Das R, Cortina-Borja M and Berdoy M (2002) The design of animal experiments: Reducing the use of animals in research through better experimental design. In: *Laboratory Animal Handbooks No 14*. London: Royal Society of Medicine Press Ltd. 1–16.

Gold LS, Slone TH and Ames BN (1998). What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metabolism Reviews, 30,* 359–404.

Hartung T (2008). Food for thought...on animal tests. *ALTEX, 25(1),* 3–9.

Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C and Heneghan C (2014). The need for randomization in animal trials: an overview of systematic reviews. *PLoS ONE, 9(6),* e98856.

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M and Langendam MW (2014). SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology 14(1),* 43.

Hooijmans CR, 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. *Alternatives to Laboratory Animals*, *38,* 167-182.

Humane Research Australia (HRA) (2016). Statistics of Animal Use in Research and Teaching in Australia. http://www.humaneresearch.org.au/statistics/, accessed 28 Jul. 2017.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M and Altman DG (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biology, 8,* e1000412.

Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J and Altman DG (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE, 4(11),* e7824.

Knight A (2008a). 127 million non-human vertebrates used worldwide for scientific purposes in 2005. *Alternatives to Laboratory Animals, 36,* 494–496.

Knight A (2008b). Systematic reviews of animal experiments demonstrate poor contributions to human healthcare. *Reviews on Recent Clinical Trials, 3,* 89–96.

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

Knight A (2013). The Australasian regulation of scientific animal use: a chimera of protection. In Sankoff P, White S and Black C (Eds.). *Animal Law in Australasia (2nd Edn.): Continuing the Dialogue.* Annandale, NSW, Australia: Federation Press. 264 – 288.

Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H and Finkelstein R (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature, 490,* 187–191.

Macleod MR, O'Collins T, Horky LL, Howells DW and Donnan GA (2005). Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. *Journal of Cerebral Blood Flow and Metabolism, 25(6),* 713–721.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U and Donnan GA (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischemia is confounded by study quality. *Stroke, 39,* 2824–2829.

Ministry for Primary Industries (MPI) (2016). *Statistics on the Use of Animals in Research, Testing and Teaching in New Zealand in 2015*. Wellington: MPI. https://mpigovtnz.cwp.govt.nz/document-vault/15346, accessed 28 Jul. 2017.

Nobis N (2011). The harmful, nontherapeutic use of animals in research is morally wrong. *The American journal of the medical sciences*, *342(4),* 297-304.

Obrink KJ and Rehbinder C (2000). Animal definition: a necessity for the validity of animal experiments? *Laboratory Animals, 34,* 121–130.

Reichlin TS, Vogt L and Wurbel H (2016). The researchers' view of scientific rigor—survey on the conduct and reporting of in vivo research. *PLoS ONE, 11(12),* e0165999.

Regan T (1987). The case for animal rights. In *Advances in Animal Welfare Science 1986/87*. Springer Netherlands. 179-189.

Rooke ED, Vesterinen HM, Sena ES, Egan KJ and Macleod MR (2011). Dopamine agonists in animal models of Parkinson's disease: a systematic review and meta-analysis. *Parkinsonism Related Disorders, 17(5),* 313–320.

Taylor K, Gordon N, Langley G and Higgins W (2008). Estimates for worldwide laboratory animal use in 2005. *Alternatives to Laboratory Animals*, *36,* 327–342.

Vesterinen HM, Sena ES, Ffrench-Constant C, Williams A, Chandran S and Macleod MR (2010). Improving the translational hit of experimental treatments in multiple sclerosis. *Multiple Sclerosis, 16(9),* 1044–1055.

Vogt L, Reichlin TS, Nathues C and Wurbel H (2016). Authorization of animal experiments is based on confidence rather than evidence of scientific rigor. *PLoS Biology, 14(12),* e2000598.

Andrew is Director of Research and Education for SAFE. He is also a European, American and RCVS-recognised Veterinary Specialist in animal welfare, a Professor of Animal Welfare and Ethics, and Founding Director of the Centre for Animal Welfare, at England's University of Winchester. He has over 80 academic publications and a series of YouTube videos on animal issues. These include an extensive series examining the contributions to human healthcare, veterinary and other education, of invasive procedures on animals, which formed the basis for his 2010 PhD and his subsequent book, The Costs and Benefits of Animal Experiments.