CHAPTER 14

Critically Evaluating Animal Research

Andrew Knight
Professor of Animal Welfare and Ethics, Centre for Animal Welfare,
University of Winchester, United Kingdom

1 A Growing Tradition of Laboratory Animal Use

Researchers have sought to understand the mechanisms of human health and disease, for as long as the latter has existed. Serious interest in the structure and functioning of the human body has been evident at least since the ancient Greeks. However, the investigations of Greek physicians into human anatomy and physiology were greatly hampered by social taboos about dissecting human corpses (von Staden, 1989). But non-human animals (hereinafter referred to as animals), were not so revered or feared. Some dissected their corpses, while others, such as Alcmaeon of Croton (sixth–fifth century, BCE), practiced surgical or other invasive procedures on the living (Court, 2005; Maehle and Tröhler, 1990), and conducted some of the first animal experiments ever recorded.

Almost two millennia passed before such social dogmas were seriously questioned. The Renaissance heralded a new era of scientific inquiry, during which Flemish physician and surgeon Vesalius (1514–1564) began to source human cadavers for dissection illegally. He discovered that a number of anatomical structures believed to exist, following animal dissections, were unexpectedly absent in humans. His highly accurate anatomical descriptions challenged the authoritative texts of classical authors (O'Malley, 1964).

Throughout the seventeenth century the spirit of scientific inquiry grew and with it, experimentation on living animals. Some surgical investigations and demonstrations that predated anesthesia were infamously cruel and caused widespread social controversy. However, French philosopher, René Descartes (1596–1650), famously rebutted such critiques, claiming that animals were merely mindless automata, i.e., "machine-like" (Descartes, 1989); their cries were of no greater moral consequence than the squeals of a poorly-oiled machine.

Nevertheless, by the end of the seventeenth century, the question of animal suffering and the acceptability of such procedures had become an increasingly

prominent moral and social concern (Maehle and Tröhler, 1990). Jeremy Bentham (1748–1832), famously asked, "The question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?" (Bentham, 1823, Chapter 17, footnote). And his concerns have been echoed by many others since.

By the beginning of the nineteenth century, a revolution had begun within medicine. Growing awareness of the poor effectiveness of many traditional therapies led to investigations focused on understanding disease etiology (causation) and pathogenesis (progression), with the intention of increasing diagnostic and prognostic accuracy and treatment efficacy. The use of animals as investigative models increased in the second half of the nineteenth century, often in highly-invasive research and still predating most forms of anesthesia or analgesia. Increasing social unease about such research led to widespread opposition in Europe, and especially Britain, where organizations, such as the National Anti-Vivisection Society (NAVS), founded 1875, (NAVS, 2012) and the British Union for the Abolition of Vivisection, founded 1898, (now Cruelty Free International, n.d.), were established to campaign against it. The Cruelty to Animals Act (1876) entered into force, becoming the first legislation to regulate animal experiments (Franco, 2013).

In the latter part of the twentieth century, social concerns about animal suffering continued to grow, accompanied by a seemingly inexorable rise in animal experimentation. Currently, the most accurate evidence-based estimates of global laboratory animal use describe the year 2005. Approximately 126.9 million non-human vertebrates were used worldwide in that year (Knight, 2008a; Taylor et al., 2008). Driven by increased development and use of genetically-modified animals (Ormandy, Schuppli and Weary, 2009), and by large-scale chemical-testing programs (Knight, 2011), laboratory animal use has steadily increased in most developed countries, ever since.

The single largest category of research conducted today is fundamental biological research, much of which has no obvious application. The European Union (EU) is the world's largest region that publishes comprehensive analyses of its laboratory animal use. At the time of writing, the most recent published figures describe animal use in the 27 Member States of the EU in 2011 (with one state reporting for 2010). Within this period, 46.1% of the 11.5 million animals were used for this purpose. However, barring 1.6% of animals used for education and training, most of the remaining 52.3% were used in attempts to advance public health—for research, development, or toxicity testing; for quality control of products and devices for human or veterinary medicine and dentistry; or for disease diagnosis and other purposes (European Commission, 2013). Most of these animals would have been used in attempts to advance human, rather than animal, health.

2 Effectiveness of Laboratory Animal Use

Combined, this represents an enormous commitment of animal, scientific, personnel, and financial resources, ostensibly dedicated primarily to the advancement of human health. But how effective has all this research been?

Advocates of such research have regularly claimed it is essential for preventing, curing, or alleviating human diseases (e.g., Brom, 2002; Festing, 2004); and further, that the greatest achievements of medicine have only occurred through the use of animals (e.g., Pawlik, 1998). However, those who champion such claims frequently have careers dependent on such research. Furthermore, 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). In recent years, systematic reviews have become widely utilized to investigate a broad range of clinical and other research questions. Their aims are to retrieve as much high-quality evidence as possible, relevant to the research question, and to minimize bias during the selection, analysis, and reporting of results. Any conclusions reached should, accordingly, be as close as possible to biological, physical, chemical, or other truths.

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. Baker et al. (2014), for example, examined human neurological disease, which has been extensively studied in animal models, resulting in relatively few human treatments (Cheeran et al., 2009; Vesterinen et al., 2010). Similarly, despite reports of the efficacy of more than 1,000 treatments in animal models of multiple sclerosis (MS), very few treatments have progressed to the marketplace (Vesterinen et al., 2010). This usually indicates failures of efficacy or safety concerns in humans. And, despite the widespread use of animal models within stroke research, virtually no interventions described as effective in animal models have proven similarly effective in human patients (Cheeran et al., 2009). There are many other examples.

Several studies have sought to determine the maximal human clinical utility that may be achieved by animal models, by examining chimpanzee experiments, given that chimpanzees are our closest relatives (Knight, 2007); by examining experiments approved by ethics committees on the basis of explicit claims of likely human healthcare benefits (Lindl, Völkel and Kolar, 2005); or by examining highly-cited animal experiments published in leading scientific journals (Hackam and Redelmeier, 2006). Hackam and Redelmeier, for example, located 76 animal experiments, each of which had been cited well over 500 times and published in one of the world's seven top scientific journals when ranked by journal impact factor. Hence, these experiments represented some of the most important and scientifically-interesting animal research published at the time. In only 28 cases (36.8%), animal results were later replicated in humans. Most animal research is neither highly cited nor published in world-leading journals, and successful translation to humans is far lower.

3 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* (externally-derived) 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 defenses 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 esophagus 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 behavioral 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, behavioral, or cognitive characteristics (for further discussion see Herrmann, 2019, Chapter 1; Jayne and See, 2019, Chapter 21).

4 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.

Kilkenny et al. (2009) conducted one of the largest and most comprehensive systematic surveys to date, assessing the experimental design, statistical analysis, and reporting of published animal experiments. 271 papers were examined, which included 72 studies using mice, 86 using non-human primates, and 113 using rats. Most (99%; 269/271) of these papers were published between 2003 and 2005. They covered a wide variety of experimental fields, were published in a comprehensive range of journals, and were funded by leading grant agencies within the United Kingdom and the United States. However, only 59% of these studies clearly stated the hypothesis or objective of the study and the number and characteristics of the animals used. Details, such as animal strain, sex, age, and weight, are all scientifically important and can potentially influence results (Alfaro, 2005; GV-Solas, 1985; Obrink and Rehbinder, 2000). Nevertheless, in many cases these details 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, for example, when researchers sympathetically select animals they consider weaker, to be used as controls, rather than test animals. The introduction of such confounding factors (in this case, variable animal fitness), can potentially bias results (in this case, selection bias has occurred). Accordingly, randomized selection of animals for treatment groups is mandated, to ensure that outcome differences are most likely due to treatment effects (Festing and Altman, 2002; Festing et al., 2002). Haphazard selection does not give sufficient certainty that results are truly random, so a systematic approach is necessary, such as the use of a random number generator (Kilkenny et al., 2009). Nevertheless, despite its well-acknowledged importance, randomized allocation of animals to test groups was reported in only 12% of these studies.

Another crucial feature of good experimental design concerns the assessment of outcomes. Where qualitative assessments occur, which involve assessor judgements, it is similarly crucial that assessors do not know (are *blinded* to) the treatment (or lack thereof) of the animals assessed, lest such knowledge

subtly affect their judgement (Festing and Altman, 2002). Because, as Cochrane (1972) noted, "When humans have to make observations there is always the possibility of bias," even unintentional bias. Nevertheless, only 14% (5/35) of all papers in the survey by Kilkenny et al. (2009) that reported qualitative assessment of outcomes, also reported the use of blinding.

Many factors can affect experimental outcomes, so the incorporation of measures to minimize sources of bias is crucial to ensuring the reliability of research results. And yet, 87% of papers, examined by Kilkenny and colleagues, failed to report randomization during animal selection; and 86% failed to report blinded assessment of outcomes. Additionally, only 70% of the publications that used statistical methods described their methods and presented the results with a measure of error or variability. More recently, similar results were found in an even larger study. Vogt et al. (2016) determined the prevalence of seven basic measures against bias (i.e., allocation concealment, blinding, randomization, sample size calculation, inclusion/exclusion criteria, primary outcome variable, and statistical analysis plan), within 1,277 experimental applications approved by Swiss authorities in 2008, 2010, and 2012 and within 50 subsequent publications. Measures against bias were reported at very low rates, both in experimental applications (2%–19%) and in subsequent publications (0%– 34%).

The importance of randomization and blinding when comparing two or more experimental groups has been highlighted by reviews of animal research in the field of emergency medicine, which have found that estimates of treatment efficacy were significantly reduced in studies that incorporated these mechanisms to reduce risks of bias (Bebarta, Luyten and Heard, 2003; Macleod et al., 2008). Similar results have been found in numerous other studies. In fact, studies 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). The widespread failure to utilize mechanisms, such as randomization and blinding, appears to result in false expectations of treatment efficacy and reported outcomes in animals often fail to translate into humans. Similar results were reported following a literature review by Holman, Head, Lanfear and Jennions (2015). They found that blind protocols are uncommon in the life sciences, and that non-blind studies tend to report more significant outcomes and higher effect sizes. They noted that: "Observer bias and other 'experimenter effects' occur when researchers' expectations influence study outcome. These biases are strongest when researchers expect a particular result, are measuring subjective variables, and have an incentive to produce data that confirm predictions. To minimize bias, it is good practice to work 'blind,' meaning that experimenters

are unaware of the identity or treatment group of their subjects while conducting research" (p. 1).

Another common problem 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 them. And yet, studies that use too many animals waste animal lives. Conversely, the results of underpowered studies (with insufficient numbers of 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 enough subjects (but not more) are present to detect biologically important effects. Indisputably, the same principles should apply to animal studies (Dell, Holleran and Ramakrishnan, 2002; Festing and Altman, 2002).

Unfortunately, methodological flaws appear to be prevalent even within animal research conducted at highly-ranked universities and published in leading journals. After studying 814 randomly-selected studies reporting primary research, 2,671 publications reporting drug efficacy in eight disease models, and 4,859 publications from five UK institutions ranked highest across six units of assessment in biomedical sciences, in the 2008 National Research Assessment Exercise, Macleod et al. (2015) reported that severe deficiencies of experimental design remain the norm. These deficiencies were prevalent in research conducted at leading UK research universities, in research funded by leading UK funding organizations, and in research reported in high-impact journals.

5 Evidence-based Research within Human Clinical Trials

The importance of sound experimental design, and, particularly, the necessity of incorporating factors designed to minimize bias risks have long been recognized within the field of human research. *The Consolidated Standards of Reporting Trials* (CONSORT) Statement for randomized controlled human clinical trials was one of the first guidelines developed to ensure the quality of human-based research. It provides an evidence-based, minimum set of recommendations, including a checklist of 25 recommended items that should be included when reporting randomized human trials (Moher, Schulz and Altman, 2001; Schulz, Altman and Moher, 2010). Since then, more than 90 guidelines have been developed for reporting different types of health research (see Altman et al., 2008; Simera et al., 2010; www.equator-network.org).

An increasing number of leading journals have, subsequently, requested that their authors comply with the CONSORT guidelines (Altman, 2005; Hopewell

et al., 2008). Organizations commending the use of such guidelines include, the Committee on Publication Ethics (n.d.); the Nuffield Council on Bioethics (2005); the Council of Science Editors (2018); and the International Committee of Medical Journal Editors (2015). Subsequent to the widespread endorsement of such guidelines, studies have indicated that the quality and transparency of reports on human clinical trials have improved (Plint et al., 2006; Kane, Wang and Garrard, 2007).

6 Application to Animal Studies

More recently, multiple attempts have been made to introduce similar standards within animal studies. In 2009, Kilkenny and colleagues observed that most biomedical journals provided little or no guidance about the reporting of animal research, other than the requirement to report ethical review of the proposed protocols. They noted the contrast between biomedical journals and those within other several research areas, particularly medical research, in this respect. Accordingly, 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.

Hooijmans et al. (2010) similarly proposed a *Gold Standard Publication Checklist* (GSPC), which includes 74 items designed to improve the quality of animal studies and to fully integrate 3Rs "(replacement, reduction and refinement)" methods and facilitate their incorporation within systematic reviews and meta-analyses. In 2014, Hooijmans and colleagues also proposed a *Risk of Bias* (RoB) tool to assess methodological quality and risk of bias within animal studies. The tool is based on the similar Cochrane RoB tool (Higgins et al., 2011), which was adjusted for particular aspects of bias that play a role in animal studies.

Other authors have proposed similar guidelines and checklists for the conduct and reporting of animal research. In 2009, Osborne and colleagues from

the Royal Society for the Prevention of Cruelty to Animals (UK) proposed a 12-point assessment scheme for scoring biomedical journals' policies on animal welfare and the 3Rs. And in 2015, Martins and Franco proposed their *Excellence in Editorial Mandatory Policies for Animal Research* (EXEMPLAR) scale, comprising four categories: regulatory compliance, quality of research and reporting of results, animal welfare and ethics, and criteria for the exclusion of papers.

7 Poor Compliance of Animal Studies

Such guidelines provide indisputable benefits in ensuring the reporting of methodological quality, reliability of results, and incorporation of the 3R principles of animal research. 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.

Noting that, "Despite reports of over 1,000 treatments effective in animal models of multiple sclerosis (MS), very few treatments have so far made it to the marketplace following initial development in disease-related animal models (Vesterinen et al., 2010)," Baker et al. (2014) investigated the general adequacy of reporting within animal studies of MS. They uncovered significant inadequacies within the reporting of experimental design, including the selection of appropriate statistical analyses and the application of key points in the ARRIVE guidelines. They observed that the ARRIVE guidelines are not being implemented by authors, reviewers, and journal editors (Baker and Amor, 2012; Landis et al., 2012; Schwarz, Iglhaut and Becker, 2012).

Despite their very widespread publication and endorsement, lack of awareness of such guidelines appears to remain a major problem. After surveying all registered *in vivo* researchers in Switzerland recently, Reichlin et al. (2016) reported that among 302 self-selected participants, 56.3% did not know of the ARRIVE Guidelines. A total of 1,891 researchers were surveyed, but only 302 (16%) returned fully-completed questionnaires and, hence, were not excluded. Even among those whose latest paper was published in a journal that had endorsed the ARRIVE guidelines, 51% had never heard of them.

The failure of biomedical journals to insist on compliance with quality control standards is partly to blame. After surveying 236 biomedical journals' policies on animal research, Osborne et al. (2009) found no mention of animal use, within author guidelines or elsewhere, in 35% of journals studied. In 18% of the journals, animals were mentioned, but no perceptible guidelines were provided; and most of the remaining journals scored poorly, with 37% scoring three or fewer points out of 12 equally weighted items within their quality checklist. Martins and Franco (2015) examined 170 journals that publish studies on animal models of three human diseases, namely Amyotrophic Lateral Sclerosis (ALS, also known as Motor Neuron Disease); Type-1 Diabetes; and Tuberculosis. Their results were broadly similar to the results of a survey by Osborne et al.'s. (2009), when assessing studies using their EXEMPLAR scale. They noted that, "little progress found regarding in-house policies on the ethical treatment of animals is worrisome" (p. 325).

8 Improving Study Quality

A range of measures are strongly warranted to increase the implementation of the 3R principles, the methodological quality of animal research, and 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 anesthetics 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.

Where journal space constraints limit the description of methodological details, these should be included in supplementary online databases, which are now widely available (Kilkenny et al., 2009). This would also facilitate the transfer of alternative technologies, such as the development of new alternative methods, between institutions (Gruber and Hartung, 2004).

To enable animal researchers and technicians to meet the necessary standards, training and continuing professional development in 3R methodologies and the design, conduct, and reporting of animal research should be compulsory. The existing lack of focus on replacement methods (in favor 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 licensing 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 over-estimations of treatment efficacy and partly explains the widespread failures in humans of treatments apparently efficacious in animals.

Within the field of human studies, clinical trial registers allow researchers to learn about existing and prior clinical trials, including those with negative outcomes, before results are formally published. A similar international initiative to register animal studies and their results is warranted (Hooijmans et al., 2014).

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. If these measures were to be successfully implemented throughout the broad field of animal research, then we may be able to predict treatment effects accurately within the animal species under study. However, interspecies differences will 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 these effects. Such factors, which reflect the intrinsic complexity of living organisms, will continue to pose barriers to extrapolation to humans that will remain insurmountable, in many cases.

9 Impacts on Laboratory Animals

Human patients are far from the only victims of poorly conducted, poorly predictive, animal research. A wide variety of stressors have the potential to

cause significant stress, fear, and possibly distress in laboratory animals. These stressors may be associated with the capture of wild-sourced species, such as primates, to supply laboratories or breeding centers; with transportation, which may be prolonged for some animals; with laboratory housing and environments; and with both routine and invasive laboratory procedures (see Knight, 2011). An *invasive* procedure is an intervention that interferes with bodily integrity through puncture, incision, or insertion of an instrument or foreign material, as in surgical and some experimental procedures (Knight, 2011).

A large minority of all procedures are markedly invasive. These include procedures resulting in death (whether or not the animals are conscious); surgical procedures (excluding very minor operative procedures); major physiological challenges; and the production of genetically-modified animals. Few regions report procedural invasiveness, but Canada does. From 1996–2008 inclusively, the proportion of markedly invasive procedures reported in Canada ranged between approximately 29%–44% (Canadian Council on Animal Care, 2009). These procedures were defined by the Canadian Council on Animal Care (2009) as resulting in moderate to severe stress or discomfort (Category D); or in severe pain near, at, or above the pain tolerance threshold of unanesthetized conscious animals (Category E) compared to procedures resulting in little or no discomfort or stress (Category B) or minor stress or pain of short duration (Category C).

A sizeable majority of all procedures utilize no anesthetics of any kind. Few regions report anesthetic usage, but Britain does. During two recent decades (1998–2009), the proportion of procedures conducted in the UK without anesthesia fluctuated between approximately 59%–69% (Home Office, 2010). For example, in 2009, at the end of this period, 66.7% of cases did not utilize any form of anesthesia. General anesthesia was provided throughout or at the end of terminal procedures in 9.5% of cases. In 17.1% of cases, general anesthesia with recovery was provided, and in 6.7% of cases, local anesthesia (Home Office, 2010).

To assess animal impacts further, it is helpful to know the frequency of analgesic (pain-killer) use, and the level of correlation between markedly invasive procedures and anesthetic or analgesic use (See Herrmann and Flecknell (2018) for a review of original animal research proposals). Painful or invasive procedures warrant anesthesia and/or analgesia. Animal welfare is adversely affected when animals undergoing such procedures are denied these; or conversely, when they are provided without sufficient need (due to their potential side effects), although this is rare in practice. It would also be helpful to study the prevalence of environmental enrichment and socialization opportunities. Unfortunately, such information remains largely unreported.

10 Conclusions

Animal research is a mechanism by which we seek to increase our understanding of the biological world. The major useful applications of this knowledge lie in the development of new therapies for combatting human diseases and in predicting the human toxicity of chemicals used for a wide range of purposes. As we have seen, however, the actual efficacy of animal research for these purposes is very low. This is due to a range of causes, some of which are, at least theoretically, amenable to change and some of which are not.

When formulating social policy pertaining to animal research, the social benefits realized are only part of the equation. The other major part that must be considered concerns the resources consumed by this research. 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, may well be expected to produce greater gains for public health. And as we have seen, the impacts on animals are also severe. 127 million living non-human vertebrates were used worldwide in 2005, the most recent year for which an evidence-based global estimate was available. Based on figures from countries, such as Canada and the UK, where these are published, a large minority of all procedures are markedly invasive; and a sizeable majority utilize (or at least report) no anesthetics of any kind.

The core ethical principle underpinning modern animal experimentation regulation and policy is that the likely benefits of such research must outweigh its expected costs. This utilitarian harm-benefit analysis underpins all fundamental regulation governing animal experimentation. For example, European Directive 2010/63/EU on the protection of animals used for scientific purposes, which directs such animal use in all EU Member States, 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" (European Parliament, 2010, p. 37).

When considering harms and benefits overall, one cannot reasonably conclude that the benefits accrued for human patients or consumers, or 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 maximize 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.

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), pp. 495–502.

Altman, D.G. (2005). Endorsement of the CONSORT Statement by High Impact Medical Journals: Survey of Instructions for Authors. *British Medical Journal*, 330, pp. 1056–1057.

Altman, D.G., I. Simera, J. Hoey, D. Moher and K. Schulz (2008). EQUATOR: Reporting Guidelines for Health Research. *Lancet,* 371, pp. 1149–1150.

Baker, D. and S. Amor (2012). Publication Guidelines for Refereeing and Reporting on Animal Use in Experimental Autoimmune Encephalomyelitis. *Journal of Neuroimmunology,* 242, pp. 78–83.

Baker, D., K. Lidster, A. Sottomayor and S. Amor (2014). Two Years Later: Journals are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Preclinical Animal Studies. *PLoS Biology,* 12(1), p. e1001756.

Balcombe, J. (2006). Laboratory Environments and Rodents' Behavioural Needs: A Review. *Laboratory Animals,* 40, pp. 217–235.

Balcombe, J., N. Barnard and C. Sandusky (2004). Laboratory Routines Cause Animal Stress. *Contemporary Topics in Laboratory Animals Science,* 43, pp. 42–51.

Baldwin, A. and M. Bekoff (2007). Too Stressed to Work. *New Scientist*, 194(2606), p. 24.

Bebarta, V., D. Luyten and K. Heard (2003). Emergency Medicine Animal Research: Does Use of Randomization and Blinding Affect the Results? *Academic Emergency Medicine,* 10(6), pp. 684–687.

Bentham, J. (1823). *Introduction to the principles of morals and legislation*, 2nd ed. Oxford: Clarendon Press.

Brom, F.W. (2002). Science and Society: Different Bioethical Approaches Towards Animal Experimentation. *Alternatives to Animal Experimentation,* 19, pp. 78–82.

Canadian Council on Animal Care (CCAC) (2009). *2008 CCAC Survey of Animal Use*. Ottawa: CCAC. [online] Available at: http://ccac.ca/Documents/Publications/Statistics/Survey_2008.pdf [Accessed 6 February 2011].

Cheeran, B., L. Cohen, B. Dobkin, G. Ford, R. Greenwood, D. Howard, M. Husain, M. Macleod, R. Nudo, J. Rothwell and A. Rudd (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), pp. 97–107.

Cochrane, A.L. (1972). The History of the Measurement of Ill Health. *International Journal of Epidemiology,* 1, pp. 89–92.

Committee on Publication Ethics (COPE) (n.d.). *COPE Best Practice Guidelines for Journal Editors*. [online] Available at: http://publicationethics.org/files/u2/Best_Practice.pdf [Accessed 12 October 2016].

Council of Science Editors (CSE)—Editorial Policy Committee (2018). *CSE's White Paper on Promoting Integrity In Scientific Journal Publications, 2018 Update*. [online] Available at: https://www.councilscienceeditors.org/wp-content/uploads/CSE-White-Paper_2018-update-050618.pdf [Accessed 22 January 2010].

Court, W.E. (2005). Pharmacy from the Ancient World to 1100 AD. In: S. Anderson, ed., *Making Medicines: A Brief History of Pharmacy and Pharmaceuticals*. London: Pharmaceutical Press, pp. 21–36.

Crossley, N.A., E. Sena, J. Goehler, J. Horn, B. van der Worp, P.M. Bath, M. Macleod and U. Dirnagl (2008). Empirical Evidence of Bias in the Design of Experimental Stroke Studies. A Meta-epidemiologic Approach. *Stroke*, 39(3), pp. 929–934.

Cruelty Free International (n.d.). *About Us.* [online] Available at: https://www.crueltyfreeinternational.org/who-we-are/about-us [Accessed 7 May 2017].

Dell, R.B., S. Holleran and R. Ramakrishnan (2002). Sample Size Determination. *Institute for Laboratory Animal Research Journal,* 43(4), pp. 207–213.

Descartes, R. (1989). Animals are machines. In: T. Regan and P. Singer, eds., *Animal Rights and Human Obligations*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, pp. 13–19.

European Commission (EC) (2013). *Commission Staff Working Document.* Accompanying document to the seventh report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. Brussels: EC. [online]. Available at: http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm [Accessed 01 October 2016].

European Parliament (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 Journal of the European Communities*, L276, pp. 33–79. [online]

Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32010L0063 [Accessed 12 August 2017].

Festing, M.F. (2004). Is the Use of Animals in Biomedical Research Still Necessary in 2002? Unfortunately, "Yes." *Alternatives to Laboratory Animals,* 32(1B), pp. 733–739.

Festing, M.F. and D.G. Altman (2002). Guidelines for the Design and Statistical Analysis of Experiments Using Laboratory Animals. *ILAR Journal,* 43(3), pp. 244–258.

Festing, M.F., P. Overend, R. Gaines Das, M. Cortina-Borja and M. Berdoy (2002) The design of animal experiments: Reducing the use of animals in research through better experimental design. In: *Laboratory Animal Handbooks*, 14. London: Royal Society of Medicine Press Ltd., pp. 1–16.

Franco, N.H. (2013). Animal Experiments in Biomedical Research: A Historical Perspective. *Animals* 3(1), pp. 238–273.

Gold, L.S., T.H. Slone and B.N. Ames (1998). What Do Animal Cancer Tests Tell Us about Human Cancer Risk? Overview of Analyses of the Carcinogenic Potency Database. *Drug Metabolism Reviews,* 30, pp. 359–404.

Greek, C.R. and J.S. Greek (2002). *4th World Congress Point/Counterpoint: Is Animal Research Necessary in 2002?.* Los Angeles: Americans for Medical Advancement.

Gruber, F.P. and T. Hartung (2004). Alternatives to Animal Experimentation in Basic Research. *Alternatives to Animal Experimentation,* 21, pp. 3–31.

Hackam, D.G. and D.A. Redelmeier (2006). Translation of Research Evidence from Animals to Humans. *Journal of the American Medical Association*, 296, pp. 1731–1732.

Hartung, T. (2008). Food for Thought … On Animal Tests. *Alternatives to Animal Experimentation,* 25(1), pp. 3–9.

Herrmann, K. and Flecknell, P. A. (2018). Retrospective review of anesthetic and analgesic regimens used in animal research proposals. *Alternatives to Animal Experimentation preprint*, [online] Available at: https://www.altex.org/index.php/altex/article/view/780/1162 [Accessed 14 September 2018].

Higgins, J.P., D.G. Altman, P.C. Gøtzsche, P. Jüni, D. Moher, A.D. Oxman, J. Savović, K.F. Schulz, L. Weeks and J.A. Sterne (2011). The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials. *British Medical Journal*, 343, p. d5928.

Hirst, J.A., J. Howick, J.K. Aronson, N. Roberts, R. Perera, C. Koshiaris and C. Heneghan (2014). The Need for Randomization in Animal Trials: An Overview of Systematic Reviews. *PLoS ONE,* 9(6), p. e98856.

Holman, L., M.L. Head, R. Lanfear and M.D. Jennions (2015) Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLoS Biology,* 13(7), p. e1002190.

Home Office (2010) *Statistics of Scientific Procedures on Living Animals: Great Britain 2009*. London: The Stationery Office .

Hooijmans, C.R., M. Leenaars and M. Ritskes-Hoitinga (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, pp. 167–182.

Hooijmans, C.R., M.M. Rovers, R.B. de Vries, M. Leenaars, M. Ritskes-Hoitinga and M.W. Langendam (2014). SYRCLE's Risk of Bias Tool for Animal Studies. *BMC Medical Research Methodology,* 14(1), p. 43.

Hopewell, S., D.G. Altman, D. Moher and K.F. Schulz (2008) Endorsement of the CONSORT Statement by High Impact Factor Medical Journals: A Survey of Journal Editors and Journal "Instructions to Authors." *Trials,* 9, p. 20.

International Committee of Medical Journal Editors (2015). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. [online] Available at: http://www.icmje.org/icmje-recommendations.pdf [Accessed 12 October 2016].

Kane, R.L., J. Wang and J. Garrard (2007). Reporting in Randomized Clinical Trials Improved after Adoption of the CONSORT Statement. *Journal of Clinical Epidemiology,* 60, pp. 241–249.

Kilkenny, C., W.J. Browne, I.C. Cuthill, M. Emerson and D.G. Altman (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLoS Biology,* 8, p. e1000412.

Kilkenny, C., N. Parsons, E. Kadyszewski, M.F. Festing, I.C. Cuthill, D. Fry, J. Hutton and D.G. Altman (2009) Survey of the Quality of Experimental Design, Statistical Analysis, and Reporting of Research Using Animals. *PLoS ONE,* 4(11), p. e7824.

Knight, A. (2007). The Poor Contribution of Chimpanzee Experiments to Biomedical Progress. *Journal of Applied Animal Welfare Science,* 10(4), pp. 281–308.

Knight, A. (2008a). 127 Million Non-human Vertebrates Used Worldwide for Scientific Purposes in 2005. *Alternatives to Laboratory Animals,* 36, pp. 494–496.

Knight, A. (2008b). Systematic Reviews of Animal Experiments Demonstrate Poor Contributions to Human Healthcare. *Reviews on Recent Clinical Trials,* 3, pp. 89–96.

Knight, A. (2011). *The costs and benefits of animal experiments.* Houndmills, Basingstoke, UK: Palgrave Macmillan.

Landis, S.C., S.G. Amara, K. Asadullah, C.P. Austin, R. Blumenstein, E.W. Bradley, R.G. Crystal, R.B. Darnell, R.J. Ferrante, H. Fillit and R. Finkelstein (2012). A Call for Transparent Reporting to Optimize the Predictive Value of Preclinical Research. *Nature,* 490, pp. 187–191.

Lindl, T., M. Völkel and R. Kolar (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.] *Alternatives to Animal Experimentation* 22(3), pp. 143–151.

Macleod, M.R., A.L. McLean, A. Kyriakopoulou, S. Serghiou, A. de Wilde, N. Sherratt, T. Hirst, R. Hemblade, Z. Bahor, C. Nunes-Fonseca, A. Potluru, A. Thomson, J. Baginskaite, K. Egan, H. Vesterinen, G.L. Currie, L. Churilov, D.W. Howells and E.S. Sena (2015) Risk of Bias in Reports of *In Vivo* Research: A Focus for Improvement. *PLoS Biology,* 13(10), p. e1002273.

Macleod, M.R., T. O'Collins, L.L. Horky, D.W. Howells and G.A. Donnan (2005). Systematic Review and Meta-analysis of the Efficacy of FK506 in Experimental Stroke. *Journal of Cerebral Blood Flow and Metabolism,* 25(6), pp. 713–721.

Macleod, M.R., H.B. van der Worp, E.S. Sena, D.W. Howells, U. Dirnagl and G.A. Donnan (2008). Evidence for the Efficacy of NXY – 059 in Experimental Focal Cerebral Ischemia Is Confounded by Study Quality. *Stroke,* 39, pp. 2824–2829.

Maehle, A.H. and U. Tröhler (1990). Animal experimentation from antiquity to the end of the eighteenth century: Attitudes and arguments. In N.A. Rupke, ed., *Vivisection in Historical Perspective*. London: Routledge, pp. 14–47.

Martins, A.R. and N.H. Franco (2015). A Critical Look at Biomedical Journals' Policies on Animal Research by Use of a Novel Tool: The EXEMPLAR Scale. *Animals*, 5(2), pp. 315–331.

Moher, D., A. Liberati, J. Tetzlaff and D.G. Altman (2009). Preferred Reporting Items for Systematic Reviews and Meta-analyses: The PRISMA Statement. *PLoS Medicine,* 6, p. e1000097. [online] Available at: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097 [Accessed 26 June 2017].

Moher, D., K.F. Schulz and D.G. Altman (2001). The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-group Randomized Trials. *Lancet,* 357, pp. 1191–1194.

National Anti-Vivisection Society (NAVS) (2012). *The History of the NAVS*. [online] Available at: http://www.navs.org.uk/about_us/24/0/299/ [Accessed 07 May 2017].

Nuffield Council on Bioethics (2005). Discussion and recommendations. In: *The Ethics of Research Involving Animals*, Chapter 15, 313. [online] Available at: http://nuffieldbioethics.org/wp-content/uploads/The-ethics-of-research-involving-animals-full-report.pdf [Accessed 26 January 2010].

O'Malley, C.D. (1964). *Andreas Vesalius of Brussels*. Berkeley: University of California Press.

Obrink, K.J. and C. Rehbinder (2000). Animal Definition: A Necessity for the Validity of Animal Experiments?. *Laboratory Animals*, 34, pp. 121–130.

Ormandy, E.H., C.A. Schuppli and D.M. Weary (2009). Worldwide Trends in the Use of Animals in Research: The Contribution of Genetically-modified Animal Models. *Alternatives to laboratory animals,* 37, pp. 63–68.

Osborne, N.J., D. Payne and M.L. Newman (2009). Journal Editorial Policies, Animal Welfare, and the 3Rs. *American Journal of Bioethics.* 9, pp. 55–59.

Pawlik, W.W. (1998). Znaczenie Zwierzat w Badaniach Biomedycznych. [The significance of animals in biomedical research]. *Folia Medica Cracoviensia*, 39, pp. 175–182.

Plint, A.C., D. Moher, A. Morrison, K. Schulz, D.G. Altman, C. Hill and I. Gaboury (2006). Does the CONSORT Checklist Improve the Quality of Reports of Randomised Controlled Trials? A SYSTEMATIC REVIEW. *Medical Journal of Australia,* 185, pp. 263–267.

Reichlin, T.S., L. Vogt and H. Wurbel (2016). The Researchers' View of Scientific Rigor—Survey on the Conduct and Reporting of *In Vivo* Research. *PLoS ONE*, 11(12), p. e0165999.

Rooke, E.D., H.M. Vesterinen, E.S. Sena, K.J. Egan and M.R. Macleod (2011). Dopamine Agonists in Animal Models of Parkinson's Disease: A Systematic Review and Meta-analysis. *Parkinsonism Related Disorders,* 17(5), pp. 313–320.

Schulz, K.F., D.G. Altman and D. Moher (2010). CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *British Medical Journal,* 340, p. c332.

Schwarz, F., G. Iglhaut and J. Becker (2012). Quality Assessment of Reporting of Animal Studies on Pathogenesis and Treatment of Peri-implant Mucositis and Peri-implantitis. A Systematic Review Using the ARRIVE Guidelines. *Journal of Clinical Periodontology,* 39(12), pp. 63–72.

Simera, I., D. Moher, J. Hoey, K.F. Schulz and D.G. Altman (2010). A Catalogue of Reporting Guidelines for Health Research. *European Journal of Clinical Investigation,* 40, pp. 35–53.

Taylor, K., N. Gordon, G. Langley and W. Higgins (2008). Estimates for Worldwide Laboratory Animal Use in 2005. *Alternatives to Laboratory Animals*, 36(3), pp. 327–342.

Vesterinen, H.M., E.S. Sena, C. Ffrench-Constant, A. Williams, S. Chandran and M.R. Macleod (2010). Improving the Translational Hit of Experimental Treatments in Multiple Sclerosis. *Multiple Sclerosis,* 16(9), pp. 1044–1055.

Vogt, L., T.S. Reichlin, C. Nathues and H. Würbel (2016). Authorization of Animal Experiments Is Based on Confidence Rather Than Evidence of Scientific Rigor. *PLoS Biology,* 14(12), p. e2000598.

von Staden, H. (1989). *Herophilus: The art of medicine in early Alexandria*. Cambridge: Cambridge University Press.

Working Committee for the Biological Characterisation of Laboratory Animals (GV-Solas) (1985). Guidelines for Specification of Animals and Husbandry Methods When Reporting the Results of Animal Experiments. *Laboratory Animals,* 19, pp. 106–108.