Ethics and Controversies in Animal Subjects Research and Impact on Clinical Decision-Making

Michael Balls, MA, DPhil[a,*], Jarrod Bailey, PhD[b]

KEYWORDS

• Alternative methods • Animal experimentation • Ethical acceptability
• Scientific justification

KEY POINTS

- Animal research and testing are justified, by those who accept them, by weighing the balance between harm to the animal subjects with consequent benefits to humans, but the harm is often underappreciated, while claimed human benefit is exaggerated or absent.
- The translation of animal research to human biology and benefit is poor, because one species cannot serve as a reliable model for any other, due to significant and intractable genetic and other biological differences. This leads to human harm from poor understanding of diseases, harmful new drugs, and the terminated development of drugs that might have been sufficiently safe in humans.
- Variability within species means that data are difficult to extrapolate between individuals of the same species, and a full understanding of human diseases and drug responses must incorporate human variability.
- A shift away from animal research and testing toward human-focused new-approach methodologies is urgently needed, for an acceptable level of translation to human biology and greater and more reliable clinical benefits.
- Greater awareness of these issues among professionals in all biomedical fields is crucial. This includes the field of anesthesiology, as it affects the preclinical knowledge base and, ultimately, clinical decision-making.

[a] Faculty of Medicine and Health Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK; [b] Research and Regulatory Affairs, Physicians Committee for Responsible Medicine (PCRM), 5100 Wisconsin Avenue, Northwest, Suite 400, Washington, DC 20016-4131, USA
* Corresponding author.
E-mail address: michael.balls@btopenworld.com

Anesthesiology Clin 42 (2024) 593–606
https://doi.org/10.1016/j.anclin.2024.03.003 **anesthesiology.theclinics.com**

INTRODUCTION

Each year, about 200 million laboratory animals worldwide are subjected to scientific procedures that would be considered immoral in human beings, but which often cause them to suffer pain, distress, lasting harm, and death. Most of the animals are rodents, but an estimated 200,000 are non-human primates and 160,000 are dogs. The aim of this research is ostensibly to increase understanding of human biology and diseases, and risks to human health, with a view to their avoidance, treatment, and management.[1] Here, we consider whether the use of animals as surrogates for humans can be considered scientifically justifiable and ethically acceptable.

THE SCIENTIFIC METHOD

Since the seventeenth century, the scientific method has been the basis for the acquisition of scientific knowledge. This can lead to *hypotheses*, tentative propositions which are tested by experimentation. The accumulation of evidence can lead to the adjustment or abandonment of hypotheses, but sometimes, as confidence in the probability that a hypothesis is correct increases, a *theory* can be proposed to explain what is now known. As with hypotheses, the validity of theories must be regularly and vigorously reevaluated, as more evidence is accumulated. This aspect of critical evaluation is rarely applied rigorously to the use of animals in research, and, if it were, a considerable amount of animal use in science would not be permitted.

ETHICS AND BIOMEDICAL RESEARCH

Ethics includes the consideration of what individuals and organizations must/must not do or may/may not do in particular circumstances, that is, those circumstances under which certain actions are obligatory/prohibited or permitted/forbidden. To address these ethical obligations requires making judgments about goodness, morality, freedom, and responsibility, and whether actions are right or wrong, or just or unjust.

We all have to deal with both personal ethics and societal ethics, including professional ethics. There can be clashes between personal and professional ethics, which can cause a moral conflict. Some people avoid this by recognizing what is expected of them in terms of personal ethics, whilst accepting that they are also subject to codes of ethics in their professional lives. For example, doctors and lawyers must work according to standards laid down by the professional bodies without whose recognition they cannot practice.

There are no specific codes of ethics or conduct about how scientists should behave, though the Three Rs (*reduction, refinement, and replacement*) concept proposed by Russell and Burch[2] in 1959 provides guidance on how humane biomedical research should be conducted. On the whole, however, this depends on the personal ethics of the scientists, especially when they are involved in blue skies research or basic research aimed at increasing knowledge which might somehow be applied to particular human diseases in as yet unspecified ways. By contrast, there are laws and regulations which require the use of animal studies, as, for example, in the testing of chemicals and chemical products for potential toxic effects, and, in the case of pharmaceuticals and vaccines, also for efficacy. Here, scientists can, and do, justify their actions by referring to the requirements of their employers and the regulatory authorities, rather than to scientific evidence of their predictive relevance and reliability.

There are also laws and regulations designed to offer some protection to animals in laboratories, which ban or limit the use of certain procedures and regulate others,

including breeding, housing, and the use of anesthesia and analgesia. They also require a weighing of the balance between the likely benefits to humans and the potential harm caused to animals.

Ethics provides the essential link between science and society, in terms of what is done in the name of science and its value and moral acceptability. The trust that the public places in doctors and scientists is much higher than that for other professionals, but that trust should be earned and maintained, not assumed. The ethical evaluation of animal experimentation is particularly important, since the application of the results obtained may mean that while some humans may benefit, others may be exposed to hazardous chemicals, products, or treatments, which may cause serious adverse effects, and even death.

As in all scientific research, the ethical value and acceptability of animal studies depend on high standards of experimental design and performance, no fabrication or biased selection of data, no falsification of outcomes, no misrepresentation of the work or the views of others, and no attempts to deceive anybody.

If these standards are not met in every aspect and at every stage, the work cannot be ethically acceptable, and human and economic resources and the lives of animals used will have been wasted.

Even if all the standards are met, there are important reasons why the ethical acceptability of research should be questioned. These include the interests, rights, and welfare of humans as well as animals, social issues and values, and conflicts of interest.

However, it must be recognized that we do not live in a world with only right or wrong, so informed, objective, and independent judgments are necessary, as complex cases and many different situations have to be considered. Nevertheless, in the rest of this article, we will consider some problems which seriously limit the scientific value and/or preclude the ethical acceptability of animal studies, and compelling reasons why they should be replaced as rapidly as possible to the benefit of humans and animals alike.

UNSOLVABLE PROBLEMS WHICH LIMIT THE VALUE OF ANIMAL STUDIES

A number of related and insoluble problems severely limit the applicability and translatability of knowledge gained from animal studies to human situations.[1–3]

The Nature and Use of Models

The use of mice and rats as models for humans began to be fashionable among scientists in the late 1800s, and animals gradually came to be regarded simply as research tools. By the 1930s, supplying the animals and associated equipment was very lucrative, and, in the 1980s, the development of techniques for the genetic manipulation of rodents opened up new commercial and scientific opportunities. Nevertheless, the unquestioning reliance on animal models as the default approach to understanding and preventing or treating human diseases, or testing chemicals and products for adverse effects, is fraught with danger—for humans as well as for animals.

In general, models are informative representations of what is being modeled. They can be physical models, as of a ship or plane, or abstract/conceptual models, such as mathematical models. Models seek to represent reality, whereas theories seek to explain reality.

Physical models are based on similarity to what is being modeled, with a selective emphasis on the features considered, which depends on what the modeler considers

to be relevant. In the biomedical context, overreliance on "similarity" can lead to drastic and disastrous consequences. Russell and Burch[2] warned of the *high-fidelity fallacy*—the assumption that, since other mammals are *generally similar* to humans, as they all have hearts, kidneys, livers, and lungs, and endocrine, immune, and nervous systems, they can provide relevant and reliable information about human diseases and the effects on humans of chemicals and chemical products, including pharmaceuticals. However, what are badly needed are *high discrimination models*, ideally not involving laboratory animals, with a very high *specific similarity* to what needs to be modeled.

The fact is that general similarity is not enough—the model should be as identical as possible to what is being modeled. The inescapable problem for biomedical research and testing is that, overwhelmingly, not enough is known about what is being modeled or about the model being used for there to be significant confidence that the information obtained from the animals can reliably and safely be translated to human situations. It is like looking into a black box to see what could happen in a different black box. That is irrational and unscientific, and therefore it cannot be considered to be ethical.

Species Differences

To a zoologist, the very idea that members of mammalian orders such as the Rodentia (guinea pigs, mice, and rats), Carnivora (dogs), Lagomorpha (rabbits), or subgroups of the order Primates, such as New World monkeys (marmosets), Old World monkeys (macaques), or other Hominini (chimpanzees) could serve as "models" of *Homo sapiens* would be ridiculous. Whatever may have been the features of their shared ancestors in the very distant past, and whatever superficial general similarities they may share, the fact is that there are major qualitative and quantitative differences in the structures and functions of the organs and systems of the various currently living mammalian species, because of the different ways in which they have evolved and adapted to their particular lifestyles and habitats. In particular, there are morphologic and physiologic differences between the cardiovascular, digestive, immune, nervous, and respiratory systems, which make it impossible to translate knowledge gained from laboratory animals to human situations.

Despite the vast investment of human and economic resources in the search for reliable animal models, very little progress is being made in relation to many serious human conditions, including Alzheimer's disease, amyotrophic lateral sclerosis (motor neuron disease), cancer, type 2 diabetes, human immunodeficiency virus (HIV) acquired immunodeficiency syndrome (AIDS), Huntington's disease, Parkinson's disease, rheumatoid arthritis, asthma, chronic obstructive pulmonary disease, multiple sclerosis, and sepsis (acute inflammatory disease). What tends to happen is that, as more becomes known about a current animal model, its comparability with the human condition becomes less and less certain, until the model has to be abandoned altogether. A related problem is that what is classified as one disease can actually involve a wide spectrum of conditions, which will differ fundamentally, both in themselves and in particular individuals. For example, *dementia* is an umbrella term for more than 200 different conditions, so it is inconceivable that one animal model, or a small number of models, could be appropriate for all forms of dementia or for all human patients.

Species differences are also important in relation to pharmacokinetics (drug absorption, distribution, metabolism, and excretion [ADME]), which can affect pharmacodynamics (the effects of drugs and their mechanisms at the system, organ, tissue, cellular, and molecular levels). In relation to metabolism, there are both interspecies and intraspecies differences in the presence, levels, and expression of

drug-metabolizing enzymes which confound the translation of data, not only from one species to another, but also from one individual to another within a species.

It is obvious that fundamental biological differences between species matter, but there is also increasing evidence to support these important differences and their implications. Comparative studies of gene complement and expression belie superficial claims of interspecies similarity. Defenders of animal research claim that since we are genetically very similar to non-human primates, biological processes in humans, the diseases they suffer from, and their responses to drugs and other chemical products must also be similar. But the evidence resoundingly shows otherwise. We now know why biological differences matter when using animals in research and why translation from animals to humans is, and can only ever be, very poor. Even when different species share identical or similar genes, their expression is affected by their location in the genome, the types of genetic material nearby, small differences in multiple genetic control mechanisms, and so on. There are immensely powerful and consequential differences in all aspects of gene expression between species that may affect the expression of hundreds or even thousands of genes, from chromosome and chromatin structure through RNA splicing and editing to post-translational modification. Many of these differences are associated with different manifestations of the major diseases on which little progress is being made.[4,5]

More and more scientists are now seeking new approaches to tackle persistent biomedical problems. It is heartening that many researchers are turning to human-focused methods in place of animal approaches. However, it is disappointing that some are content to tinker with the same 'tools' that have led to the problems in the first place. For example, developers of organ-on-a-chip technologies have often bemoaned difficulties in attracting funding, while funders were providing multi-million euro or US dollar support for projects such as the European Mouse Genome Mutagenesis Program, and the National Institutes of Health International Knockout Mouse Project. At the time of writing (February 2024), the UK government's Medical Research Council has just committed more than £20 million (US$ 25 million) "to investigate key disease areas using the mouse as the main model organism" (https://www.ukri.org/news/new-mouse-genetics-funding-to-tackle-human-ageing-dementia-and-diseases/).

Animal Variability

There is no such thing as *the* dog, *the* rhesus monkey, *the* mouse, or *the* rat, since there are immense variations within each species. No studies with a small group of animals can reflect what would happen in that species as a whole. No study in animals of one species or order could be used to model what might happen in another. If rats cannot be acceptable models for rats in general or for mice, how could they be acceptable models for humans?

Variations in rats and mice include many behavioral differences, such as sociability, performance in cognitive tasks, and responses to injury. Different strains also respond differently to drugs, both in their effects and toxicities. There are also sex and strain differences in many biological systems, conditions, and diseases, including neuroimmunity, obesity, diabetes, aging, the cardiovascular system, liver diseases, and cancer. Mice also differ in responses to pain and sensitivity to analgesia, due to genetic mechanisms which are 'large and heritable.'[6]

Another important example is that of genetic variability among different populations of conspecific macaques, which means they have a variety of physiologic and behavioral differences and can differ from each other as much as from another species. For example, differences in mobile or transposable DNA elements which can move around

the genome are known to affect gene complement and expression, and in turn the susceptibility to, and pathologies of, various diseases. Associated duplications of DNA sequences, which are a source of intraspecies genetic variation, affect susceptibility to diseases, the formation of tumors, responses of the immune system to infectious agents, and more. Interspecies and intraspecies variations in major histocompatibility complex or, in humans, human leukocyte antigen (HLA) genes, profoundly affect the immune system. In rhesus macaques, the equivalents are 'Mamu' genes, variations in which lead to significant variations in immune functions in macaques from different geographic regions, with implications for the extrapolation of data from research involving these animals to humans. One example is in the use of macaques in HIV/AIDS, polio, and malaria research, in which such subtle genetic variations led to major pathologic differences in those infections.[5]

Human Variability

Species differences become even more important in the light of the huge scale of human variation, which is affected by a very wide range of factors, including age, sex, race, occupation, lifestyle, and previous or concurrent illnesses or therapies, against a background of immense genetic variation.

There is no such thing as *the* human, and one human subpopulation cannot be a reliable model for all human subpopulations. There are major differences in susceptibility to diseases, and a drug which can help some patients may be lethal for others. It is not surprising that the results of early testing in humans, which normally involves young and healthy males, are unlikely to be indicative of what will happen in the wider population. There are particular problems with establishing the tolerance to some therapies in children and elderly people, where ADME may be different from that in the young adult. In addition, there can also be ethical problems, since children and individuals with certain illnesses could not be expected to give their informed consent. The use of juvenile or aging laboratory animals would be unlikely to give the clear indications needed.

For many years, human immune responses to infectious agents have been known to vary *substantially*. Common genetic variations in humans affect susceptibility to infection by severe acute respiratory syndrome coronavirus 2 (coronavirus disease 2019), influenza virus, respiratory syncytial virus, human immunodeficiency virus, human T-cell leukemia virus, human papilloma virus, hepatitis B and C viruses, herpes simplex virus, norovirus, rotavirus, parvovirus, and Epstein-Barr virus (EBV).[7] Variants of the HLA genes are associated with differences in EBV and rubella virus levels in infected individuals.[8] Even minor variations in gene or protein activity can elicit major effects: for instance, levels of the abundant protein, HSP90, powerfully affect the consequences of genetic variation in humans.[9]

This is also true for responses to drugs, both in terms of efficacy and adverse effects. The genes involved in pharmacokinetics and pharmacodynamics are highly variable. For example, 4 in every 5 patients carry a genetic variant that may have functional effects on drug efficacy for the 100 most prescribed medications in the United States, and the frequencies of these variations differ substantially in different human populations. This issue is so serious that many major medical institutions have genotyping protocols for pre-emptive pharmacogenetic testing, though only a very small proportion of drug-related genes can be covered.[10] It is not surprising that the extrapolation of data from another species to 'humans' collectively is unreliable and unrealistic. The aim must be to provide human-derived cell lines incorporating human biological diversity to facilitate the reliable modeling of human biology, human pathology, and human responses to xenobiotics.

OTHER ISSUES WITH SCIENTIFIC AND ETHICAL IMPLICATIONS

Standards and Bias

There has long been concern about the overall standard of medical research, much of which is flawed, cannot be repeated, and/or has only limited value or no use at all. There is also concern about the quality of animal studies, which is one of the reasons why clinical trials fail. A related problem is the failure of many biomedical scientists to understand statistical methods and to use them properly.

These problems can result from the poor design of studies; the poor conduct, collection, and analysis of data; and the inadequate reporting of what has been done and what can be concluded from it.

A particularly serious problem is that of bias,[1] since, despite the lack of systematic and objective reviews of its benefits and oft stated commitments in support of the Three Rs, animal experimentation remains the established norm in biomedical research. Bias can seriously prejudice the quality of the work and has serious implications. Bias can impact study design, data presentation, and interpretation for many reasons, including, but not limited to ensuring academic and research appointments and advancement, ongoing research funding, and, in some cases, personal financial gain. The use of animal data in publications in the scientific literature, in regulatory requirements for the testing of drugs and other chemicals and products, and even in the application of laws aimed at giving protection to laboratory animals represents inherent bias toward using animals without considering the alternatives. It should be a matter of concern that the world's two leading scientific journals, *Nature* (London) and *Science* (New York) both have a long history of accepting articles involving animal experiments without question, while paying scant recognition to the concept and use of replacement alternatives.

Harm Versus Benefit

In the United Kingdom, the *Animals (Scientific Procedures) Act 1986* requires the Home Secretary to "weigh the likely adverse effects on the animals concerned (in research) against the benefits likely to accrue as a result of the program of work," before animal research can be approved. However, little consideration has been given to precisely how this harm-benefit analysis (HBA) should be done.[3,11] In particular, relatively little attention has been paid to how human benefit should be evaluated. Likely benefit has tended to be assumed, based solely on the inherent seriousness of the medical problem to which the work was claimed to be relevant, alongside an underlying assumption that laboratory animals are sufficiently like humans for studies on them to be meaningful. The European Commission has tried to improve this process in recent years, for example, via new legislation (*Directive 2010/63/EU*, replacing *Directive 86/609/EEC*) that requires some retrospective assessments of *actual* harm and benefit following research programs, but it is too early to evaluate the impact of this.

Crucially, *human* harms are not (yet) factored into the HBA process.[12] These may be direct or indirect.

Direct harm can result when new drugs, considered to be sufficiently likely to be efficacious and safe after preclinical testing, lack efficacy and/or have serious adverse side effects when tested in clinical trials. Sometimes, this can result in the most extreme human harm—death. Two of many examples are particularly noteworthy, since they involved how the animal test data were used, rather than only considering the quality of the data themselves. First, the improved glucose tolerance brought about by benfluorex in an animal model of insulin resistance suggested a wider therapeutic application in man, to include the treatment of type 2 diabetes. But it turned

out that the preclinical studies were misleading. The drug was patented by a French pharmaceutical company between 1976 and 2009, during which time it was thought to have caused between 500 and 2000 deaths as a result of heart disease. Secondly, TGN412 was a humanized T-cell activator monoclonal antibody given to human volunteers at a dose based on data from tests in rats and macaques. Six of the eight volunteers rapidly developed multiorgan failure. They survived after 30 days in intensive care, but were warned to expect an early death. Why was it assumed that animal test data would be relevant for a product deliberately engineered to be human specific?

The classification and labeling of many types of chemical and chemical products (including pesticides) are also highly dependent on animal testing requirements and guidelines stipulated by national and international regulatory authorities. As for drugs, the relevance and reliability of the data are questionable, and the use of these data can lead to direct human harm as a result of exposure at home, work, or in the environment.

Indirect harm results, for example, from the nonavailability of drugs which might have been sufficiently efficacious and safe in humans if their development had not been stopped, due to lack of efficacy or adverse effects in animals. It is not possible to quantify this, but there are known cases where animal data could have stopped the development of drugs found to be effective in humans. For example, tamoxifen, which is effective for certain types of breast cancer and was subsequently found to induce liver tumors in rats. Gleevec caused severe liver damage in dogs but not in human cells in vitro, so clinical trials proceeded with no signs of serious liver toxicity in humans, when used to treat chronic myelogenous leukemia. Other examples of useful drugs nearly lost as a result of animal data are cyclosporine, aripiprazole, and esomeprazole.

The European Union Registration, Evaluation, Authorisation, and Restriction of Chemicals System

In 2007, the European Union (EU) introduced a new system for the Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), managed by the European Chemicals Agency (ECHA), based in Helsinki.

Because safety data are lacking for 86% of the chemicals that were on the EU market before the current system for testing new chemicals came into use in 1981, about 27,000 companies have to fill in knowledge gaps on the toxicity of and uses of about 30,000 chemicals. It has been estimated that if this involved testing according to the traditional test procedures, between 18 and 54 million laboratory animals would be needed. The REACH regulation states that one of the purposes of the system is to promote alternative methods, and that animal testing should only be used as a last resort. It also requires the European Commission to regularly review the test method regulations and update them, in order to replace, reduce, or refine animal testing. How this works in practice is a matter of great concern. The ECHA appears to do little to encourage the use of non-animal test methods and to be willing to accept data from animal tests when alternative methods are available.[13] That is not scientifically justifiable or ethically acceptable. An overriding question remains unanswered: Where is the evidence that the REACH system is contributing meaningfully to the greater protection of human beings and the environment?

Dosage and Exposure

Risk = hazard $\times$ exposure, so the assessment of risk to humans depends on the quality of both the qualitative and quantitative estimates of hazard and the likely route, scale, and frequency of exposure. This is particularly important in relation to major aspects of human toxicity, such as carcinogenicity, cardiovascular toxicity,

neurotoxicity, and reproductive toxicity, in all of which risk assessment is highly dependent on predictions based on the outcomes of toxicity tests in laboratory animals.

Ideally, both the hazard and exposure components of the equation should be based on realistic and reliable predictions which are realistically relevant to humans, but the uses of animal studies which have evolved over time continue to be the subjects of debate and controversy. For example,

a. The two-year *rodent bioassay* has long been regarded by some scientists and regulators as the gold standard for identifying chemical carcinogens. The treatment of the animals (usually rats) begins a short time after their birth and continues for 24 months. The test item is given daily on 7 days per week to 50 animals per group per sex, at three dose levels, one of which is usually the maximum tolerated dose (the highest dose which produces toxicity without causing death or a decrease in body weight of more than 10%). Critics say that the two underlying assumptions, that rodent carcinogens are human carcinogens and that the effects of higher doses are indicative of what would occur at lower doses, are both incorrect.[14] It has also been estimated that, if the number of animals per group were increased from 50 to 200, virtually all the chemicals tested would be deemed to be rodent carcinogens.
b. *Chemical toxicity* used to be classified according to the median lethal dose acute toxicity value, the dose which killed 50% of the animals in a test group. This led to situations where animals were given immense doses in order to provide a value, resulting in immense suffering, so limit tests were introduced, which did require dosing above certain levels. Three alternative procedures are now accepted by the Organisation for Economic Co-operation and Development, which still require acute effects but cause less suffering. The 3T3 neutral red uptake in vitro test has now been accepted for certain purposes, such as determining starting doses, and other promising non-animal tests await full regulatory approval.
c. Hartung[15] has pointed out that about 5,500 of the 30,000 chemicals covered by REACH would need to be tested. It is estimated that about 2.5% of them (138 substances) are true reproductive toxicants in humans, and the goal of the testing would be to identify them. Testing according to the current test guidelines, first in rats, then in rabbits or mice for chemicals found to be non-toxic, 116 of the true reproductive toxins would be identified, but 3,454 non-toxic chemicals would give false-positive results. These results might restrict the use of a large number of substances that are already widely used. He also pointed out that the tests for each substance require about 3,200 animals for a single test—a total of 17.6 million animals for 5,500 substances, which raises questions about ethical acceptability, as well as scientific and economic wisdom.
d. The *Draize eye irritation test* involves adding a small amount of the test item to the eye of a restrained rabbit, ensuring that it does not leak out, then observing for signs of irritation or other damage for up to 14 days. The test was controversial for scientific reasons, because of differences between human and rabbit eyes, the enormous variability of the results obtained, and the fact that no entry of a chemical into the human eye would involve leaving it there and waiting for days to see what would happen. The test was also criticized on the grounds of inhumanity, since, while many toxicologists would take steps to avoid high scores, others would adhere to the standard operating procedure, at whatever cost to the animals. Non-animal tests are now available, which can reduce the need for the animal test, though it is still performed.

e. A selection from about 2,500 *food additives* is deliberately added to processed food, to preserve, color, flavor, sweeten, or thicken what we eat. These additives must be assessed for safety. Some toxicologists and authorities would favor animal testing for long-term, repeated-dose toxicity, carcinogenicity, neurotoxicity, and reproductive toxicology, but the traditional methods for such testing would require exposure at doses far higher than humans would ever be likely to experience. That would be costly, scientifically questionable, and ethically unacceptable. One approach to this problem is to consider thresholds of toxicologic concern (TTC), that is, levels of human intake or exposure that are considered to be of negligible risk, when the structure of a chemical is known, but only limited toxicity data for it are available.[15] The key question is how the TTC value would be determined, which opens up the possibility of using new-approach methodologies.

Pharmaceuticals

For therapeutic drugs, both animal and human data are available, because humans are deliberately exposed to them in clinical trials, after mandatory preclinical studies involving rodents and non-rodents (dogs or macaques).

An independent analysis of 12,728 drug development programs between 2011 and 2020, using the Biomedtracker database, found that the probability of progressing from a phase I clinical trial to US Food and Drug Administration approval was just 7.9%—a failure rate of more than 92%, for drugs entering clinical trials based on favorable animal data with regard to efficacy and toxicity/safety. Cardiovascular and oncology drugs were among the biggest failures, with a mean failure rate of 95%.[16] It has been estimated that 40% to 50% of these failures are due to poor efficacy, with another 30% due to unmanageable toxicity, and 10% to 15% due to poor pharmacokinetic properties.

Due to a general lack of stringent and comprehensive analyses of the human-predictive nature of preclinical (animal) tests of new human drugs, we set about conducting our own study. We analyzed an extensive toxicity data set for over 2,300 drugs, for which both animal and human data had been independently collected and classified. On the advice of eminent statisticians, we calculated likelihood ratios (LRs) for the data set, which included tissue-level effects and Medical Dictionary for Regulatory Activities level 1 to 4 biomedical observations. The resulting LRs showed that the absence of toxicity in dogs provided virtually no weight of evidence that adverse drug reactions (ADRs) would also be absent in humans. The LRs did suggest that the presence of toxic effects in dogs can sometimes provide evidential weight for a risk of potential ADRs in humans, but this was highly inconsistent, varying by over two orders of magnitude for different classes of compounds and their effects. We went on to compare the data from three other preclinical test species (rat, mouse, and rabbit) with the human data, and came to the same conclusion. In a third article, we presented further data from non-human primates, which supported our previous conclusions, and also showed, in particular, that test results indicating an absence of toxicity in one species provided no evidential weight with regard to toxicity in any other species, even when data from non-human primates and humans were compared. Subsequent published analyses from investigators in the pharmaceutical industry agreed with our conclusions. Following the publication of the authors' series of papers, we asked the British Home Office, the Association of the British Pharmaceutical Industry, the UK Medicines and Healthcare products Regulatory Agency, and the European Medicines Agency for meetings to discuss the implications of our studies, but none of them, nor even the National Centre for the Replacement,

Refinement and Reduction of Animals in Research, were prepared to meet us for worthwhile discussions.[17–20].

This situation makes it clear that new and directly human-relevant methods and approaches are needed, before new drugs can be satisfactorily evaluated in clinical trials. This must include recognition of the importance of human variation and its consequences in terms of drug efficacy and the induction of adverse effects.

One notable example of this need and of the capabilities of human-focused methods to much more effectively predict human adverse effects from new drugs is drug-induced liver injury (DILI), a leading cause of failed clinical trials and of post-marketing drug withdrawals. A recent (2023) paper from the organ-on-a-chip company, Emulate, reported the performance of its human liver chips with a blinded panel of 27 small molecule drugs of established hepatotoxicity or non-toxicity. The liver chips demonstrated 87% sensitivity at clinical concentrations, including 11 drugs that were responsible for 242 human deaths. Importantly, the preclinical animal test data had indicated an acceptable therapeutic window. The liver chips also showed a specificity of 100%. This approach showed significant reliability in predicting human DILI, as well as predicting which drugs should progress into clinical trials with no unacceptable safety risk. It was concluded that the routine use of these chips in drug development programs could be worth up to US$ 3 billion per annum to the pharmaceutical industry, or even an estimated US$ 24 billion, if used alongside four other chips modeling common toxicities associated with drug attrition.[21]

Anesthesiology

The scientific and ethical issues discussed in this article are no less important for those practising and using anesthetics than for professionals in any other biomedical discipline.

Anesthetic agents are often used in research procedures involving animals, and some important differences in the process of anesthesia underline the problem of species differences and why and how they matter.

Although there are some interspecies commonalities and similarities, differences necessitate specific approaches to and/or requirements for preanesthetic assessment, preoperative fasting, preoperative and anesthetic medications, the monitoring and induction of anesthesia, recovery from anesthesia, and more.[22] Anesthetics also have highlighted variable effects between species, even for minor procedures in non-human primates.[23] Ketamine, for example, is known to reconfigure brain connectivity networks in rodents, with some similarities between rats and mice, but also some incomplete concordance and inconsistencies. Ketamine elicits variable antidepressive effects in rodents that are both species and strain dependent, alongside the effects of laboratory-related stresses/stressors. These anti-depressant effects also occur in humans, which has led to the use of ketamine as an off-label anti-depressant, with resultant concerns for teratological effects on developing fetuses. This research has revealed differences in the neurodevelopmental effects of ketamine between rodents and non-human primates, evidenced by species differences in necrosis and apoptosis that are underpinned by temporal and mechanistic differences in cell proliferation, neuronal differentiation, dendritic development, and synaptogenesis.[24,25]

One of the most obvious species differences is in perianesthetic mortality, which can be up to 100-fold higher in birds, cats, dogs, guinea-pigs, and horses than in humans. The benzodiazepines midazolam and diazepam are acknowledged as central nervous system depressants and muscle relaxants in humans, but their effects are highly species specific in animals: while some species are sedated, cats, dogs, and horses often experience excitement and dysphoria, and midazolam has

antinociceptive effects in sheep, unlike other species. Opiods are used across species, but may cause postoperative nausea and vomiting in humans, which are uncommon in animals, and which seem to be due to interspecies genetic variations in opioid receptors. Some anesthetic agents are used in animals but rarely/not in humans, for example, buprenorphine, butorphanol, urethane, and xylazine.[26]

Different anesthetic agents, and different doses and routes of administration of these agents, are known to differentially affect cortical electrophysiology. This has significant consequences for animals used in neurophysiological research, as it introduces bias in recordings and measurements of brain activity. Scientists need to consider their choice of anesthetic in their research, in order to minimize impact on the areas of the brain they are studying.[27]

A combination of anesthetic effects and species differences in brain connectivity, function, gene expression, and so forth make the translatability of non-human primate neuroscience to human neurology even poorer.[28]

THE WAY FORWARD—THE AUTHORS' OPINION

There is now a weight of multifactorial evidence that supports the phasing out of unreliable and misleading animal use in biomedical research and testing and the need to focus rigorously on *human* biology. This article briefly summarizes much of the evidence and the rationale supporting this.

There is increasing public opposition to the annual use of 200 million animals globally in research studies, each year, which causes much suffering and death. There is a lack of application of the scientific method and critical thinking about animal experimentation, and of constructive approaches to non-animal methods, in many fields, in many institutions, and by many scientists. Ethical acceptability and recognition of the true scale and depth of harm from animal research are overwhelmed by human self-interest. Ethical questions are raised by harm to humans as a consequence of false confidence in animal tests, leading to new chemicals and drugs that result in unacceptable adverse effects. Indirect costs result when research with animal models is misleading, and other avenues of investigation would have been more predictive and productive; leading, for example, to new drugs that would have been safe and effective in humans. Genetic and other biological differences mean that one species can *never* be used—ethics aside—as a model for any other. Human variability can only be factored into research and testing through the use of human-derived and human-specific investigative methods.

Research in anesthesiology must also be affected by interspecies and intraspecies variability in relation to the use and effects of anesthetics. This can affect the value of experimental data, making them more unreliable than they already are, as a result of species differences in what is being investigated.

All of the aforementioned negatively affects our knowledge of human biological systems and diseases being researched, and therefore our ability to understand them, and ultimately to do much about them, either preventively or therapeutically. We know the consequences of this—the how and the why of the failure of animal data to translate to human biology. We know that animal research is not delivering, but is instead confounding, misleading, and delaying. To still not *really* know much about so many diseases, well into the twenty-first century, is astounding. For example, we continue to struggle to get to understand, and to find effective new therapies or vaccines for dealing with, dementia, Parkinson's disease, HIV/AIDS, stroke, and various cancers. The attrition rate of pharmaceutical development is as great as it has ever been, mostly for reasons of poor efficacy and adverse reactions not predicted by animal tests.

Although opinions differ on whether the use of animals in laboratory experiments can be considered to be right or wrong, or acceptable or unacceptable, the crucial point is that they cannot be used to provide relevant, reliable, and applicable information on the nature and causation of human diseases or the risks to humans resulting from exposure to chemicals and drugs and other chemical products. Put simply, the current reliance of so much of biomedical research on the use of laboratory animals cannot be justified on scientific grounds, so it cannot be ethically acceptable. All concerned should act on the evidence, and phase in human-specific, patient-derived advanced cell and tissue culture methodologies as a matter of urgency, to more accurately model, reflect, and predict human biology, and treat human diseases.

CLINICS CARE POINTS

- Clinical care guidelines and advice (for example, developmental toxicology/teratology) that are mainly based on animal data should be treated with significant caution. This is due to the poor human relevance of animal data, underpinned by widespread interspecies biological differences.
- Human data to underpin clinical care advice and decisions should always be sought. Intraspecies variability may affect the applicability of general human data to some individuals (due to environmental, genetic, nutrition, lifestyle, and other factors), but human data will generally be more applicable than those from another species.

FURTHER READING

The authors have cited, and directed the reader toward, many of their own publications. This is for brevity and because their work contains many references to the work of others that are highly relevant to this article.

DISCLOSURE

Both authors have nothing to disclose.

FUNDING

Jarrod Bailey's contributions to this work were funded by the Physicians Committee for Responsible Medicine (PCRM), Washington, DC, USA.

REFERENCES

1. Pound P. Rat trap: the capture of medicine by animal research – and how to break free. Leicestershire, UK: Troubador Publishing Ltd; 2023.
2. Russell WMS, Burch RL. The principles of humane experimental technique. London, UK: Methuen; 1959. 238pp.
3. Balls M. It's time to reconsider the principles of humane experimental technique. Altern Lab Anim 2020;48:40–6.
4. Bailey J. Lessons from chimpanzee-based research on human disease: the implications of genetic differences. Altern Lab Anim 2011;39:527–40.
5. Bailey J. Monkey-based research on human disease: the implications of genetic differences. Altern Lab Anim 2014;42:287–317.
6. Smith JC. A review of strain and sex differences in response to pain and analgesia in mice. Comp Med 2019;69:490–500.

7. Kenney AD, Dowdle JA, Bozzacco L, et al. Human genetic determinants of viral diseases. Annu Rev Genet 2017;51:241–63.
8. Scepanovic P, Alanio C, Hammer C, et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. Genome Med 2018;10:59.
9. Karras GI, Yi S, Sahni N, et al. HSP90 shapes the consequences of human genetic variation. Cell 2017;168:856–66.e12.
10. Schärfe CPI, Tremmel R, Schwab M, et al. Genetic variation in human drug-related genes. Genome Med 2017;9:117.
11. Bateson P. When to experiment on animals. New Sci 1986;109:30–2.
12. Balls M. It's time to include harm to humans in harm-benefit analysis - but how to do it, that is the question. Altern Lab Anim 2021;49:182–96.
13. Taylor K. Ten years of REACH - An animal protection perspective. Altern Lab Anim 2018;46:347–73.
14. Goodman JI. Goodbye to the bioassay. Toxicol Res 2018;7:558–64.
15. EFSA SC, More SJ, Bampidis V, et al. Guidance on the use of the threshold of toxicological concern approach in food safety assessment. EFSA J 2019;17:e05708.
16. Thomas DW, Chancellor D, Micklus A, et al. Clinical development success rates and contributing factors 2011-2020. 2021. Available at: https://pharmaintelligence.informa.com/~/media/informa-shop-window/pharma/2021/files/reports/2021-clinical-development-success-rates-2011-2020-v17.pdf.
17. Bailey J, Thew M, Balls M. An analysis of the use of dogs in predicting human toxicology and drug safety. Altern Lab Anim 2013;41:335–50.
18. Bailey J, Thew M, Balls M. An analysis of the use of animal models in predicting human toxicology and drug safety. Altern Lab Anim 2014;42:181–99.
19. Bailey J, Thew M, Balls M. Predicting human drug toxicity and safety via animal tests: can any one species predict drug toxicity in any other, and do monkeys help? Altern Lab Anim 2015;43:393–403.
20. Bailey J, Balls M. Recent efforts to elucidate the scientific validity of animal-based drug tests by the pharmaceutical industry, pro-testing lobby groups, and animal welfare organisations. BMC Med Ethics 2019;20:16.
21. Ewart L, Apostolou A, Briggs SA, et al. Performance assessment and economic analysis of a human Liver-Chip for predictive toxicology. Commun Med 2022;2:154.
22. Kurdi MS, Ramaswamy AH. Anesthetizing animals: Similar to humans yet, peculiar. Anesth Essays Res 2015;9:298–303.
23. Lee VK, Flynt KS, Haag LM, et al. Comparison of the effects of ketamine, ketamine-medetomidine, and ketamine-midazolam on physiologic parameters and anesthesia-induced stress in rhesus (Macaca mulatta) and cynomolgus (Macaca fascicularis) macaques. J Am Assoc Lab Anim Sci 2010;49:57–63.
24. Cheung HM, Yew DTW. Effects of perinatal exposure to ketamine on the developing brain. Front Neurosci 2019;13:138.
25. Becker R, Braun U, Schwarz AJ, et al. Species-conserved reconfigurations of brain network topology induced by ketamine. Transl Psychiatry 2016;6:e786.
26. Carter J, Story DA. Veterinary and human anaesthesia: an overview of some parallels and contrasts. Anaesth Intensive Care 2013;41:710–8.
27. Sorrenti V, Cecchetto C, Maschietto M, et al. Understanding the effects of anesthesia on cortical electrophysiological recordings: a scoping review. Int J Mol Sci 2021;22:1286.
28. Bailey J, Taylor K. Non-human primates in neuroscience research: The case against its scientific necessity. Altern Lab Anim 2016;44:43–69.