6
The Costs of Animal Experiments

> The important thing is not to stop questioning.
>
> —Albert Einstein

A major glitch in drug development

Despite our tremendous prevention power, the fact remains that drugs are important tools in the arsenal of modern medical science. To produce new drugs, we need research. This involves applied research, that is, research directly intended to produce a new treatment. Basic, or more exploratory, research is also utilized to help direct applied research. To approve a drug for the market, regulatory requirements usually dictate at least two major stages of safety and efficacy testing. The preclinical stage includes the use of in vitro and/or animal experiments to assess whether a drug is a viable candidate for further clinical investigation based on safety and efficacy evaluations. The clinical stage is broken down into three phases. Phase 1 typically involves a small group of healthy human volunteers to test the safety of a compound. Phases 2 and 3 usually include larger groups of volunteers in controlled clinical trials to test for both safety and efficacy of the potential treatment against the targeted disease or condition. Post-marketing studies are also often required to monitor the safety of a product once in use.

The USA leads the world in the amount of resources directed toward biomedical research.[1] It spends an estimated $100–$120 billion on research annually. Pharmaceutical industries are the largest contributors to biomedical research spending with the publicly funded granting agency, the US National Institutes of Health (NIH), being the second largest contributor, funding approximately $31 billion per year of

A. Akhtar, *Animals and Public Health*

research.[2] Money spent on biomedical research is growing so fast that it outpaces growth of the gross domestic product in the USA.[3]

Despite the impressive amount of money being spent on biomedical research, the USA lags behind 41 countries in life expectancy.[4] Clearly something is amiss. Much of the reason why the USA lags in longevity is the relatively low priority it gives to disease prevention in comparison with many other developed nations. However, a major glitch in the drug development world has also been increasingly noted. In 2006 an article in the *Journal of the American Medical Association* (*JAMA*) reported, 'While investment in basic research in the United States doubled from 1993 to 2003, the number of therapeutics entering the clinic has actually declined.'[5] New compounds entering phase 1 trials today have about an 8 percent chance of reaching the market.[6] Many drug candidates that enter later phases of the drug development process are also falling by the wayside. A recent analysis revealed that in phase 3 trials the failure rate is now 50 percent.[7] Overall, 92 percent of drugs that pass preclinical tests fail to make it to the market because they are proved to be ineffective and/or unsafe in people.

From 1996 to 1999, 157 new drugs were approved in the USA. A decade later, from 2006 to 2009, only 74 new drugs were approved. Of all these approved drugs, not one of them, according to a recent report, was a cure or a meaningful novel treatment for a host of serious diseases.[8] This has led many to voice concerns about the stagnation in production of useful treatments—and this concern is nothing new.[9] Memorial Sloan-Kettering colon cancer specialist Leonard Saltz lamented the lack of cancer treatment breakthroughs when he said that despite all the hype and excitement about pricey new cancer drugs, by far the most important colon cancer drug remains a 50-year-old chemotherapeutic drug called 5-FU.[10]

There are several potential reasons offered for the reduced number of treatment approvals, including higher regulatory hurdles, longer and more expensive clinical trials and less flexibility in pricing.[11] Perhaps the most salient reason, however, is that noted by the Institute of Medicine (IOM). In June 2000 the IOM conducted a clinical research round table to discuss the state of medical research.[12] The IOM pointed out a 'disconnection between the promise of basic science and the delivery of better health'. In essence, basic biomedical research is generally not efficiently leading to therapies, despite our significant investment of money, time and other resources. It was reported in *JAMA* that because of a doubling of the NIH's budget in recent years as well as major new advances in basic research, many have made the assumption that progress was being

made that would result in improved human health.[13] The report then goes on to state that this assumption has been an illusion. Both John Ioannidis from Tufts University School of Medicine in Boston and an article in *Drug Discovery Today* echoed this sentiment and commented that while basic sciences are believed to have made major progress, this has not resulted in the same level of progress in understanding the clinical basis of diseases or in developing novel effective treatments.[14]

In summary, while the pace of basic biomedical research has been rapid, it has not translated effectively to new therapies that have a measurable impact on our health.[15] Something is not working. Why is our tremendous investment in biomedical research not returning on its promise? Two investigators looked at the overall lack of successful development of drugs to treat a host of central nervous system (CNS) disorders.[16] In recent years, only 9 percent of CNS compounds that enter phase 1 clinical trials survive launch. The investigators concluded that one of the main reasons for this high failure rate is that animal models are a far from perfect predictor of drug efficacy in humans:

> The increasingly high failure rates of CNS compounds in human trials has demonstrated that this success in animal models is no guarantee. No animal model is a perfect mimic of human disease. Animal models can serve as models of disease mechanisms, but not of the disease itself . . . Failure rates in clinical development attest to the disparity.

Over the years, much of biomedical research has moved away from more directly studying human physiological mechanisms and diseases and instead has focused on creating and studying models of diseases and mechanisms in animals. There is now a growing recognition that there is an incongruity between understanding mechanisms in animals and understanding an actual human disease.[17] Investigators from the Department of Clinical Neurosciences at the University of Edinburgh noted that while the mechanisms of stroke in animals are well understood, this has not translated to positive results in humans.[18] More than 350 interventions have published efficacy in animal stroke models, of which around 100 have been tested and proven ineffective in human stroke studies.[19] Thus, as illustrated by this one example, understanding disease mechanisms in animals, whether by creating animal models of diseases or though basic physiological research, is not successfully leading to new therapies. 'The failure of neuroprotective drugs in clinical trials,' commented one publication on stroke studies, 'represents a major challenge to the doctrine that animals provide a scientifically valid model for human stroke.'[20]

The lack of sufficient success in utilizing animal experiments to yield new therapies is a fact not just in the field of stroke or basic research but also in applied research. Researchers from the Animal Bioscience and Biotechnology Lab from the US Department of Agriculture (USDA) provided a frank appraisal of the usefulness of animal experiments in predicting human outcomes and found that, on average, 'the extrapolated results from studies using tens of millions of animals fail to accurately predict human responses'.[21] Even the use of multiple species of animals frequently fails to predict efficacy in human trials.[22] In 2002 several leaders in the biotechnology and pharmaceutical industries published a paper outlining what they saw as the major problems underlying the drug development process. They concluded that the poor predictability of animal experiments is one of the major challenges facing the drug discovery community.[23] Based on these and a host of other reports, many in the health community are arriving at a harsh realization—we are failing to effectively discover new therapies in large part because of our focus on animal experimentation in biomedical research.[24]

Over the past few decades, evidence-based medicine has become the mantra of sound, scientifically based medical research and practice. We rely on evidence-based medicine in virtually every facet of health research and practice save one—the use of animal experimentation to inform human health. Animal experimentation has not been subjected to the kind of scrutiny it requires.[25] It is most often viewed as the default and 'gold standard' method of testing, yet it doesn't, with few exceptions later described, receive the critical examination needed to determine its relevance to human health.[26] As a result, there is a dearth of published, peer-reviewed evidence to support the usefulness of animal experimentation.[27] The lack of critical studies examining the relevance of animal experiments was reflected in a recent report from the Nuffield Council on Bioethics.[28] Instead of critical examination, anecdotal evidence or unsupported claims, which are inadequate forms of evidence for a scientific discipline, are substituted as justification for animal experiments.[29]

When animal model validity is discussed, it is usually in terms of the similarities between the model and the human condition it is intended to mimic. However, very infrequently is any formal validation of such models applied.[30] A review of the published literature revealed that even in cases when an animal model(s) is alleged to replicate a human condition, there were very few studies that formally evaluated the ability of these models to reproduce the human diseases in question.[31] In an article published in *Slate* magazine in 2006, entitled 'Of Mice and Men:

The Problems with Animal Testing', reporter Arthur Allen expressed this concern about the reliability of animal experiments to predict harmful adverse effects of drugs in humans: 'Surprisingly, although it is central to the legitimacy of animal testing, only a dozen or so scholars over the past 3 years have explored this question. The results, such as they are, have been somewhat discouraging.'[32] When we actually scrutinize animal experiments, we discover that they are far from the panacea we believe them to be. As a result, a growing number of scientists are questioning the relevance of animal experiments as they relate to human disease and their ability to lead us down the right path toward effective treatments to improve human health. Scientists have also highlighted several notable shortcomings with, and obstacles to, the use of animal experimentation to inform human health. These obstacles include the effect of the laboratory environment and other variables on animal physiology, and thus study outcomes; disparities between animal models of disease and human diseases; and species differences in physiology.

The many influences on animal experimental results

In 1995 Superman actor Christopher Reeve became quadriplegic after being thrown from a horse. He turned his tragedy into advocacy and galvanized the public and scientific community to invest in spinal cord injury research. Unfortunately, there was no substantial return on that investment during Reeve's lifetime. In 2004, following his death, *New Scientist* reported that in 2000,

> [Reeve] pointed out that it was an exciting time for the field—a time when he heard that researchers could cure a rat with a spinal cord injury. Sometimes, the actor said, he wished he were a rat . . . Following Reeves death this week, his rebuke seems as fitting as ever. While basic neuroscience research is booming, there are precious few treatments—let alone cures—for people with diseases of the brain and nervous system.[33]

Reeve's comment about curing spinal cord injury in rats was not far from the truth. Multiple neuroprotective agents have been successful in treating spinal cord injuries induced in animals in the laboratory. Yet they have all produced extremely disappointing results when tried in humans.[34] The clinical usefulness of one treatment being used in humans, methylprednisolone (MP), is hotly debated. The jury is still

out as to whether or not it causes any meaningful reduction in damage following spinal cord injuries in humans. In order to assess whether experiments on animals provided any clarity to the issue, several colleagues and I conducted a systematic review of all published animal experimental studies using MP to treat spinal cord injury and broke the results down by species.[35] The review found results differed between species and among strains within a species.

The question that then followed was: do we pool results from all tested species and experiments, or do we put our faith in the results from certain species and experiments we believe to be most predictive of human responses? If we choose the former, our answer on the usefulness of MP may depend on whether most of the animal experiments involved rats, which showed mostly negative results (i.e. the treatment was not effective), or cats and dogs, which showed mostly positive results (i.e. the treatment was effective). If instead we decide to put our faith in test results from species and experiments we believe to be most predictive of human responses, how do we know which species to choose, and which set of results do we decide are most applicable? The set of experiments conducted in rats or the ones using dogs and cats? But it doesn't stop there—do we trust the results from a certain *strain* of rat and not *another strain*? These are not questions to be taken lightly—answering them is critical to determining which animal experiments best predict human results. Unfortunately, situations in which we know in advance which species or which animal model is most predictive of human outcomes are exceedingly rare, if they exist at all.

My colleagues and I then conducted an investigation to explore the potential reasons for the wide variety of results between and among species and found that many factors in the experimental protocol affect study outcomes.[36] These include how animals are handled, housed, fed and tested, and what type of anesthesia is used during injury induction. For example, cage conditions were found to affect recovery from spinal cord injury in animals. Environmental conditions can influence neurogenesis, gene expression, signaling between nerves and behavioral responses, all of which can significantly impact the results of a study. Even more surprising was that the type of flooring on which an animal is tested or whether or not there are other animals in view of the tested animal can affect whether a drug shows a benefit or not. These unintended influences go beyond studies in spinal cord injury. In a study of a genetic mutation that causes defects in the aorta, the type of environment in which mice were housed affected whether they developed the defects.[37] Another study showed that even modest differences in

housing for just one month led to structural and biochemical differences in the brains of two groups of marmoset primates.[38]

Stress, housing environment and diet can all affect study outcomes.[39] These conditions and factors can affect study outcomes in ways that experimenters may not understand, be aware of or be able to control. Even routine laboratory procedures and conditions, such as blood collection, noise produced in the laboratory, cage components and handling by experimenters, can lead to pronounced and/or prolonged changes in genetic expression and stress-related physiologic markers.[40] Ventilation and ambient noise can produce stress and affect an animal's physiology. For example, noise levels of 90 dB (about the sound of a kitchen blender), which is not infrequent in the laboratory setting, have been found to increase heart rate and blood pressure, and to damage small blood vessels.[41] Even the time of day when animals are tested can give different results. In a study of mice, motor deficits (weakness) were evident only at one time at night.[42] Experiments performed on rats in the spring can generate very different results from those performed in the late fall.[43] Tests can be affected by many additional laboratory factors, including environmental humidity, cage density and within-cage order of testing.[44]

The Scientist acknowledges that the laboratory environment can influence the results of an experiment.[45] It reported that many of the underlying limitations associated with animal experiments involve the inherent nature of animal testing. The laboratory environment can have a significant effect on test results, as stress is a common factor in an animal's life in the laboratory. Jeffrey Mogil, a psychology researcher, also demonstrates that the very presence of a researcher alters behavior in mice, which could have an impact on study results.[46] Every procedure and every environmental element in a laboratory setting can, and likely does, influence what results a study produces. Unlike with humans in clinical trials, we can't tell animals to ignore one factor or another—we have little control over their reactions to different procedures and situations. Additionally, and most importantly, animals in laboratories have little to no control over their environments, to which they are exposed, on average, for the duration of their lives. Animals' lifelong exposure to the laboratory setting increases the likelihood that such settings will substantially affect their physiology in unpredictable ways.

For the above reasons, many have called for the standardization of laboratory settings and procedures.[47] The problem as it applies to animal testing is that there are simply too many variables to achieve true standardization. Many of these—most notably those that produce

significant stress, such as catching, restraining and blood collection—are unavoidable.[48] A study published in *Science* found that despite all attempts to standardize the environment across three laboratories, there were systematic differences in test results.[49] What's more, different mouse strains varied markedly in all behavioral tests, and for some tests the magnitude of genetic differences depended upon the specific testing laboratory. Controlling how animals react, whether physiologically or behaviorally, to the procedures and settings in laboratories is unattainable in any practical sense.

Ultimately, the attempt to standardize laboratory settings and procedures fails to address the fundamental issue, which is not to improve comparison between labs but to improve the predictive value of experiments to the human condition. As increasing numbers of studies reveal discrepancies between animal experimental and clinical trial results, many scientists are requesting that more rigorous methodologies and practices (in addition to standardized environmental settings) be applied to animal experiments in an effort to reduce the discrepancies. These more rigorous practices would include assurances of adequate study power, randomization and blinding, and minimization of bias in publications.[50] Yet, although a step in the right direction, the call for improved methodologies minimizes another, more important and unmodifiable limitation of animal experiments—the animals themselves. In the review of spinal cord experiments using MP previously described, subgroup assessment was conducted on only the animal experiments that were of the best quality (e.g. those that included blinding and randomization, and reporting of housing and handling procedures) and used the same dosing and regimen of MP treatment.[51] Despite this, study results still varied considerably, indicating that no matter how methodologically superior and standardized the experiments were, factors inherent in the use of animals accounted for some of the major differences in results.

Returning to stroke for a moment, many questions have been raised as to why more than 100 potential therapies failed to translate successfully from animal experiments to human trials. Acknowledging the failure of finding new, effective stroke treatments despite so many successes in animals, a set of guidelines was implemented by a stroke round table in 1999 to standardize and improve the applicability of stroke experiments in animals to humans.[52] One of the most promising stroke treatments later to emerge was NXY-059, which proved effective in animal experiments. In 2006, at the Joint World Congress for Stroke held in Cape Town, South Africa, news spread quickly that NXY-059 fell

victim to the same fate as so many prior drugs: it failed in clinical trials. It failed despite the fact that the set of animal experiments on this drug followed the guidelines set forth by the round table and was considered the poster child for the new experimental standards.[53] 'There's no doubt about the absence of an effect of [NYX-059], and that called into question the many other studies in stroke, and how good are the animal models?' said one of the clinical consultants to the trial.[54] Despite earnest attempts, standardization and improvement of animal experimental methodologies hasn't eliminated the substantial discrepancies between animal experiments and human results.

Incongruencies between animal models and human disease

In addition to the unpredictable influences of laboratory environments on animal experimental results, the lack of sufficient congruency between animal models and human disease is another frequent and significant obstacle. When we try to create stroke in animals, for example, we artificially create a disease that occurs naturally in people. The inability to reproduce the complexity of human diseases in animals is a crucial hindrance to their use.[55] Even if design and conduct of an animal experiment are sound and standardized, the translation of its results to the clinic may fail because of disparities between the animal experiments and the clinical trials.[56] In stroke research, these disparities include the presence of pre-existing diseases and conditions in humans, but not in animals, that affect the development of stroke, such as diabetes and atherosclerosis; use of additional medications to treat these risk factors in humans; and nuances in the pathology of the human disease that are absent or different in animal models. Other disparities cited include the use of young and male animals for diseases of the elderly or women.

As a result of the recognition of these discrepancies, several publications argue for the need to use animals who are matched in relative age and gender to the target humans, who are given the same medications as those given to human patients and who have also been altered to manifest the pre-existing conditions (and co-morbidities) that occur naturally in humans.[57] If we try to reproduce the pre-existing conditions in animals, we still face challenges regarding the inability to replicate their complexity. For example, stroke and heart disease are frequently a result of atherosclerosis. Most animals in laboratories don't naturally develop significant atherosclerosis, which is characterized by a narrowing of blood vessels by plaque build-up. In order to reproduce the effects

of atherosclerosis in animals, researchers ubiquitously clamp their blood vessels. Simply clamping blood vessels, however, does not replicate the elaborate pathology of atherosclerosis and the causes behind it. In attempting to reproduce the complexity of human diseases in animals, we need to reproduce the complex physiology of the predisposing diseases and conditions, which also proves difficult to accomplish. Thus we end up continuously chasing our own tails. Each time an animal model fails to successfully translate to humans, no shortage of reasons is proffered to explain what went wrong—poor methodology, lack of relevant pre-existing conditions and medications, wrong gender or age, and so on. Recognition of each potential difference between the animal model and the human disease creates a renewed effort to eliminate these differences. What is too often ignored is that these models are intrinsically lacking relevancy to the human diseases they are intended to reproduce.

As early as 1990, major discrepancies between animal models of stroke and stroke in humans were noted.[58] Several neuroscientists asserted that animal stroke models are severely simplistic in comparison with the human disease and labeled stroke animal models a failed paradigm, arguing instead for human-based research.[59] Given the continued failure of animal stroke experiments to unravel new, effective human treatments, and despite all attempts to improve their human relevancy, the sentiment expressed in 1990 remains salient today. Naturally occurring diseases are far more complex than what is produced when we alter a few mechanisms in an animal. Even with diseases for which there is great mechanistic understanding, there can still remain significant disparities between the animal models used and the human diseases being targeted for treatment.[60]

Consider animal models of Alzheimer's disease (AD). In humans, AD is characterized pathologically by the presence of several key features in the brain. A truly predictive animal model must reproduce the origins or etiology, the physiologic basis, the pathology and the symptoms or signs of the disease.[61] Experimenters have altered genes in mice to create models of AD. But herein lies the problem: each mouse model is different and no single mouse model shows all the pathologic features of AD.[62] Instead, each model displays bits and pieces of Alzheimer's and many display features not present in human AD. Consequently, these models often give conflicting results because they differ in regard to the signs that manifest and the causes behind these signs.

Substantial effort has been made to improve the relevancy of AD mouse models. Despite these attempts, these new mouse models still

fail to appropriately mimic what occurs in humans.[63] The lack of congruency between mouse models and the human disease may cause potential drugs to seem to be ineffective, while it's actually the mouse model that is to blame.[64] One of the key messages of the 2007 Inaugural Alzheimer's Drug Discovery Foundation Meeting was that the patient is currently the only true model of AD.[65] Existing animal models replicate various aspects of the disease but do not fully mimic the human condition, resulting in a low predictive value. The conference further concluded that using models with low predictive value provides little understanding of the pathophysiology (the physiology and functional changes) of a disease. One investigator commented that 'in reality, disease models usually model only certain aspects of clinical symptomatology, and because only rarely is the etiology of diseases well understood, the induction of the disease state in the model can differ from the clinical condition'.[66] In other words, because we rarely fully understand how and why a disease occurs in humans, when we try to replicate that disease in animals we are usually falling well short of the mark. We take a few observations from humans then try to recreate those observations in animals, and we end up relying on the animal models in place of understanding the full disease in humans. This illustrates a fundamental flaw in our use of animal experiments: we are usually studying models that are at best very incomplete or at worst contrary to the human disease. Either way, the models are incorrect.

David F Horrobin, an influential figure in drug development, commented on the obstacles the pharmaceutical industry faces in delivering new therapies and criticized assumptions made about the congruence of animal models of disease to human diseases.[67] For an animal model of disease to be congruent with the human disease, he argues, three conditions must be met:

1. we must fully understand the animal model;
2. we must fully understand the human disease; and
3. we must have examined the two cases and found them to be substantially congruent in all important respects.

Horrobin contends that these three conditions have not been fulfilled for any human disease. He asks, 'Does the use of animal models of disease take us any closer to understanding human disease? With rare exceptions, the answer to this question is likely to be negative.' He also criticizes assumptions made about in vitro tests for the same reasons above and argues that we need to get back to the human patient to truly understand human disease. Horrobin is correct in arguing for the need

to study human patients. However, as will be discussed later, in vitro tests, if using *human* cells and tissues (not cells from another species) and if used in concert with other human-based testing methods, are more likely to accurately predict human outcomes than animal tests.

Horrobin is not alone in observing the incongruency between what we are studying in animals and what we should be studying. It is extremely troubling that because of our focus on animal models we know far more about a vast array of diseases in animals in the laboratory and how to treat them in animals than we do in humans (recall Christopher Reeve's comment).[68] In 2004 *New Scientist* reported on sentiments about the state of neuroscience research expressed by Susan Fitzpatrick, former Associate Executive Director of the Miami Project to Cure Paralysis and current Vice-President of the James S. McDonnell Foundation:

> 'The biomedical model is failing,' says Susan Fitzpatrick.... Basic biomedical research relies heavily on animal models, especially rats and mice, but she thinks it may be necessary to rethink this approach if treatments for brain diseases are going to reach the patients who need them. Even if we know all there is to know about the animal model we don't necessarily know about the disease, Fitzpatrick says. 'The model becomes what we study, not the human disease.'[69]

This sentiment can be applied to most human diseases. Rather than spending our time trying to unravel the mysteries behind human diseases directly, we instead create artificial animal models in the laboratory and these become our focus of attention. The *New Scientist* article continues:

> 'Take brain cancer. The traditional model for studying brain cancer is to take human cancer cells, sometimes tissue-cultured into cell lines, and transplant them under the skin of an immunosuppressed mouse. This approach ignores the fact that cancer is a disease of context: as soon as you change the environment you will change those cells. Any agent you test is probably unlikely to be effective when you have a tumour in context,' Fitzpatrick says. 'It's a fundamental flaw. We need a fundamentally new approach.'

Lost in translation: Species differences

Even when we think we have created an animal model that adequately mimics a human disease, interspecies differences come into play. In spinal cord injury, drug test results vary according to which

species, and even which strain within a species, is used, largely because of numerous inter-species and inter-strain differences in neurophysiology, anatomy and behavior.[70] For example, the micropathology of spinal cord injury, injury repair mechanisms and recovery from injury vary greatly between different strains of rats and mice.[71] Surprisingly, even rats from the same strain but purchased from different suppliers produce different test results.[72] In one study, responses to 12 different behavioral measures on pain sensitivity, which is often used as a marker of spinal cord injury severity and recovery, varied among 11 strains of mice, with no clear-cut patterns that allowed prediction of how each strain would respond.[73] Each of these and numerous other differences influenced how the animals responded not only to spinal cord injury but also to any potential therapy being tested. A drug might help one strain of mice recover but not another.

There has been considerable enthusiasm for using mice as human disease models because of their ostensible genetic similarity with humans and because their entire genome has been mapped.[74] Mice have been extensively studied and, other than rats, are the most common animals used in experimentation. Scientists have modified their genes and created a host of new mouse strains designed to mimic a range of human diseases. Arguably, we know more about mouse physiology than we do about any other species, even humans. But do we know enough? In 2006, researchers reported in the journal *Science* the discovery that mice normally have more than one thymus gland.[75] Before this discovery, the predominant scientific view was that mice possessed only one. Since the thymus affects immune system function, experimenters had for decades been removing the one murine thymus gland of which they were aware, believing that they then created immune-deficient mice. Now we know the results of over half a century of research in immunodeficiency in thymectomized (thymus gland removed) mice were likely misleading. 'From the immunological point of view,' commented the study co-authors, 'a regular second thymus in mice raises important questions about previous studies using thymectomized mice.'

A 2006 report in the *Proceedings of the National Academy of Sciences* revealed that the internal structure of the human pancreas—including the insulin-producing Islet cells and surrounding cellular architecture—differs markedly from the experimental rodent models used for more than three decades.[76] Furthermore, these differences in architecture result in distinct differences in pancreatic function between mice and humans. The authors concluded that we cannot rely on mice and rat studies and that researchers must focus on human pancreatic cells and

tissues. Perhaps most shockingly, this simple description of human pancreatic structure instantly invalidated decades of mice and rat experiments that relied on the assumption of similar pancreatic structure and function between humans and these animals.

These two examples of inaccurate assumptions about mice and rats are just the tip of the iceberg. Moreover, discovering a second thymus or comparing the cellular anatomy of mice and humans are relatively simple investigations to conduct and simple answers to confirm. How many other false assumptions are made because of questions we don't even know how to ask, yet alone answer? An article published in *Drug Discovery World*, entitled 'The importance of using human-based models in gene and drug discovery', noted that 'Mice and humans have more than 95% of their genes in common, yet mice are not men (or women).'[77] University of Michigan evolutionary biologists Ben-Yang Liao and Jianzhi Zhang found that although mice share most of their genome with humans, identical genes may behave very differently between the two species.[78] They compared human and mouse orthologs, which are genes in different species that evolved from a common ancestral gene. Normally, it is assumed that orthologs retain the same function in closely related species, such as mice and humans, during the course of evolution, and this assumption is a main basis for the use of animal models to study human biology.[79] Essential genes are those that, following loss of their function, reduce the fitness of an organism to zero. Liao and Zhang identified 120 human genes for which the mouse has an identical counterpart and discovered that 22 percent of the essential genes in humans are nonessential in mice. The authors concluded that 'it is possible that mouse models of a large number of human diseases will not yield sufficiently accurate information'. Commenting on this study, a scientist from the Dr Hadwen Trust, a medical research charity that funds the development of human-based testing methods, reflected, 'We have long been concerned that equivalent genes in humans and mice don't have the same functional effects. Millions of genetically modified mice are used as research "models" for human diseases every year but the relevance of this research to human patients is highly questionable.'[80]

A study at Massachusetts Institute of Technology demonstrated wide differences in the regulation of the same genes between the human and mouse liver.[81] Consistent phenotypes (observable physical or biochemical characteristics) are rarely obtained by modification of the same gene, even among different strains of mice.[82] Gene regulation can substantially differ among species and among individuals within a species

and may be as important as the presence or absence of a specific gene. The disruption of a gene in one strain of mice may be lethal, whereas disruption of the exact same gene in another may have no detectable phenotypic effect.[83] Such findings question the wisdom of extrapolating data that are obtained in mice to other species. 'If one mouse gene is so difficult to understand in a mouse context,' asks Horrobin 'and if the genome of a different inbred strain of mouse has so much impact on the consequences of that single gene's expression, how unlikely is it that genetically modified mice are going to provide insights into complex gene interactions in the . . . human species?'[84]

'Humanized' mice

Genetically engineered mice are extensively used in amyotrophic lateral sclerosis (Lou Gehrig's disease) experiments but they are increasingly found to be inaccurate models of the disease and their use has failed to result in any effective treatment.[85] Cystic fibrosis knockout mice (genetically engineered mice in which one or more genes have been turned off through a targeted mutation) don't display the bronchopulmonary signs that are characteristic of human cystic fibrosis.[86] Despite their genetic similarity, there are fundamental differences between tumor cells in mice and humans. For example, in comparison with human tumor cells, those in mice tend to grow much more rapidly and are much more dependent on the formation of new blood vessels.[87]

Stanford University immunologist Mark Davis blames some of our limited understanding of the human immune system on our reliance on experimentation in mice.[88] As an example, he describes the results of tests using a type of protein to treat multiple sclerosis, an autoimmune disease: 'Injecting [myelin basic protein (MBP)] into mice causes a condition similar to multiple sclerosis, which can be prevented by doses of proteins that blunt the immune reaction to MBP. But clinical trials of these protective proteins were stopped because they made some people with multiple sclerosis worse.' A study published in *Science* in 2009 found that a crucial protein found in humans to regulate blood sugar is not found in mice, calling into question the relevance of the mouse model in the development of drugs to treat human diabetes, and suggesting that testing potential diabetes drugs in mice might give misleading results.[89] Even when the protein was expressed in genetically altered mice, it behaved differently than it does in humans. Genetic mouse models are poor substitutes for a number of other human conditions.[90] As we have seen with the multiple sclerosis trial example

given above, reliance on mouse models has led to direct human harm. In 2003 Élan Pharmaceuticals had to stop trials of an AD vaccine that had cured the disease in 'Alzheimer's mice' after the substance caused brain inflammation in humans.[91]

The more we look into their effectiveness, the more we discover that genetically engineered animal models aren't living up to their promise. Perhaps the major and immutable reason genetically modified animals will not solve the problems of animal experimentation translation to humans is the fact that the 'humanized' genes are still in non-human animals. When we introduce a 'humanized gene' into a mouse, that gene will be affected by all of the physiologic mechanisms that are unique to the animal. As aptly stated in *Slate* magazine, 'tinkering with a few genes doesn't make [mice] perfect stand-ins for people'.[92] Short of turning mice into human beings, no matter how we modify their DNA there will always be significant disparities between their physiology and ours.

Do non-human primates make good models?

Drug testing regulations often require the testing of a new agent in both rodent and non-rodent species. Non-human primates (NHPs) are widely used as the non-rodent species. Yet NHPs, despite their even closer evolutionary history and genetic make-up to that of humans, also make far from ideal stand-ins for human-based tests. In March of 2006, six healthy human volunteers were injected with small doses of TGN 1412, an experimental therapy for rheumatoid arthritis and multiple sclerosis, created by TeGenero. As described by *Slate*,

> Within minutes, the human test subjects were writhing on the floor in agony. The compound was designed to dampen the immune response, but it had supercharged theirs, unleashing a cascade of chemicals that sent all six to the hospital. Several of the men suffered permanent organ damage, and one man's head swelled up so horribly that British tabloids refer to the case as the 'elephant man trial'.[93]

What went wrong? Were there too few animal experiments conducted prior to the clinical trial? No, TGN 1412 was tested in mice, rabbits, rats and monkeys with no ill effects.[94] Were the animals used not the appropriate animals to use? The answer to this also appears to be no. TeGenero intentionally selected cynomolgus monkeys for

preclinical testing because they proved to best replicate a wide variety of mechanisms in humans specifically targeted by the drug.[95] Thus, not only were several different species used, but those deemed most relevant to humans were used. Did the problem then lie in the dose given to the test animals? Again the answer is no. Monkeys underwent repeat-dose toxicity studies and were actually administered 500× the dose given to the human volunteers for not less than four consecutive weeks.[96] Still, none of the monkeys manifested the ill effects that humans showed within minutes of receiving a minuscule amount of the test drug.

The problem with the TGN 1412 experiments is not that an inappropriate animal, dose or study design was used. The problem is that pharmaceutical research is now producing sophisticated, complex and nuanced molecules targeting very specific mechanisms in humans. Despite our close genetic relationship with NHPs, they are still not similar enough to make good models. In fact, humans are not always similar enough to other humans. We widely recognize that there are many differences in physiology and susceptibility to disease, and in effectiveness and side effects of treatments between individuals and groups within our own species. Hence, there is a growing interest in personalized medicine, in which treatments are tailored to individual patients. When clinical trials are conducted on a new blood pressure medication, for example, these, with rare exceptions, tend to include African-Americans, Hispanics, Asians and women because the results may vary between these groups. What works for a Caucasian male may not work for a Caucasian female or an Asian male. Scientists recognize the diversity in physiology within our own species, even among identical twins with the same genetic make-up. Twins display different susceptibility to diseases and genetic responses from one another and these responses become more disparate as the twins age.[97] If we can't reliably extrapolate from one identical twin to another, how can we expect to safely extrapolate results from completely different species to humans?

Our closest genetic cousins—chimpanzees—share about 95–96 percent of our genes but less of our DNA because of the tens of millions of differences in non-coding regions of our DNA. Many studies have demonstrated multiple disparities between chimpanzees and humans in DNA sequence, genetic insertion and deletion events, genetic expressions and post-translational modifications.[98] A recent study found a wide variety of both subtle and large-scale differences between chimpanzees and humans in cell death and DNA repair mechanisms.[99]

NHP models fail to reproduce key features of Parkinson's disease, both in function and in pathology.[100] Several therapies that appeared

promising in both NHP and rat models of the disease showed disappointing results and even higher incidence of adverse effects in humans.[101] NHPs are not good severe acute respiratory syndrome (SARS) models either, even though an enormous undertaking has been made to reproduce the disease in them.[102] Long-time SARS researcher Robert Hogan recently argued against the further use of NHPs given that so many groups, such as the Centers for Disease Control and Prevention and the Army Medical Research Institute of Infectious Diseases, have reported contradictory results with SARS testing in NHPs.[103] Chimpanzees have been widely used to develop vaccines against hepatitis C under the presumption that they closely resemble humans in their response to the virus, despite the fact that the supporting evidence to this claim is slim.[104] After decades of this line of investigation we still have not developed any hepatitis C vaccine that works well in humans.

HIV/AIDS vaccine research using NHPs is probably one of the most notable failures of translation to humans. A lot of time and energy has been spent studying HIV in chimpanzees and other NHPs. In 2007, Alison Tonks, the associate editor of the *British Medical Journal* (*BMJ*), wrote about another failed HIV vaccine, gp120, and commented that important differences between monkey models and humans with HIV have misled researchers.[105] More than 85 HIV vaccines have failed in about 200 human trials following success in NHPs.[106] One of the most recent disappointments occurred in 2007 when a clinical trial testing a novel HIV vaccine developed by Merck (MRK-Ad5) was halted prematurely because it was actually found to increase the risk of HIV in certain groups of people.[107] MRK-Ad5, like all candidate HIV vaccines, was advanced into human trials after extensive preclinical experiments in NHPs.[108] The British newspaper the *Independent* summarized the incident as follows:

> One of the major conclusions to emerge from the failed clinical trial of the most promising prototype vaccine, manufactured by the drug company Merck, was that an important animal model used for more than a decade, testing HIV vaccines on monkeys before they are used on humans, does not in fact work.[109]

A recently published review found a paucity of evidence demonstrating successful translation of NHP research to human medicine in toxicology, stroke, AD, Parkinson's disease and infectious disease research.[110] It revealed that most data suggested experimentation on NHPs, including chimpanzees, to be irrelevant and unnecessary, to have

little or no predictive value and to be hazardous to human health. For example, the campaign to prescribe hormone replacement therapy in thousands of women to prevent heart disease and stroke was based in large part on experiments on NHPs. Hormone replacement therapy is now known to *increase* the risk of these diseases in women. The bottom line is that despite assumptions to the contrary, the evidence tells us that NHPs simply don't reliably make effective models of human diseases.

Toxicity testing in animals

Of all fields in medicine involving animal experimentation, none is getting as much scrutiny as toxicity and carcinogenicity testing. One of the most extensively used methods to predict the carcinogenicity of a substance is the costly and time-consuming two-year bioassay in which mice and rats are exposed to maximum tolerated doses of test chemicals for two years to determine whether the chemicals are carcinogenic. Health agencies in the USA and abroad have hailed this bioassay as the 'gold standard' in carcinogen identification.[111] These accolades appear premature as the human relevancy of this testing method is becoming increasingly dubious.[112] A growing body of evidence suggests that some chemicals produce cancer in mice and rats through species-specific mechanisms that are irrelevant to human physiology.[113] For example, male rats get bladder cancer from saccharin through a rodent-specific mechanism (humans lack the protein that is necessary for the development of cancer in rats).[114] Based on this understanding of the species' differences, the NIH dropped saccharin from its list of human carcinogens in 2000. Phenobarbital is carcinogenic in rats because it raises levels of thyroid-stimulating hormone (TSH), which triggers thyroid cancer cell development.[115] But it does not substantially raise TSH in humans, if at all, so our cancer risk from the drug is negligible.

The false-positive and false-negative results of the animal bioassay can be considerable. Ennever and Lave analyzed the data on known human carcinogens with the animal data for cancer predictability.[116] They found a disturbingly large proportion of incorrect predictions, 'potentially allowing widespread human exposure to misidentified chemicals'. An analysis of the data on 780 chemical agents listed in the International Agency for Research in Cancer database found the positive predictivity of the animal bioassay for a definite or probable human carcinogen to be only around 20 percent.[117] In addition to placing human lives at risk, the low predictability of this assay is costing us money and wasting time. Each assay requires up to millions of dollars and years

of planning.[118] In the meantime, as we continue to rely on this assay, there is a huge backlog of untested chemicals to which we are already exposing ourselves.[119]

Other toxicology and carcinogenicity tests that rely on animals are equally flawed. One study examined the toxicological profiles of 50 compounds in rodent and non-rodent (beagles and NHPs) species.[120] The study found poor correlation of target organ toxicity across species and concluded that 'simple extrapolation across species is unrealistic'. The study authors called for regulatory agencies to institute an evaluation of tests using animals as predictors of human adverse signs. In 1999 the Health and Environmental Science Institute examined the data on 150 compounds that had produced a variety of toxic effects in people.[121] It found that only 43 percent of the compounds produced similar effects in mice and rats and 63 percent did so in other animals. A reviewer of toxicology testing and regulations commented that

> compelled to act, regulators have chosen animal tests to forecast human cancer risks. To this end, animal data are filtered through a series of preconceived assumptions that are presumed to overcome a host of human/animal differences of biology, exposure, and statistics-differences that in reality are insurmountable.[122]

Recognizing the immense difficulty in predicting toxicity in one species based on the toxicity data from another is not new. As early as 1978, Fletcher found poor correlation between drug safety tests in animals and subsequent clinical experience with 45 major drugs, including anticancer agents, antibiotics, cardiac agents and neurological agents.[123] Fletcher's survey established that only 25 percent of the toxic effects observed in animals might be expected to occur in humans. Assessing three decades of data on the subject, toxicologist Ralph Heywood also found that the concordance between animals and humans is only 25 percent.[124] 'Toxicology,' he concluded, 'is a science without a scientific underpinning.'

'In retrospect,' Fletcher concluded in his 1978 report, 'it is a relatively simple matter to determine the correlation between animal and human studies, but prospectively it is difficult to know which particular toxic effects are likely to prove troublesome when it comes to giving the drug to man.'[125] And that's the catch: accurately predicting when the animal experimental results are relevant to humans is nearly impossible because of inter-species differences. We can always go (and have often gone) back after clinical trials have been conducted to assess whether

the animal experimental results correlated with the clinical results, but retrospective confirmation is not the purported reason for using animals in experimentation. They are intended to predict human results and inform human health care. If we find that the animal experimental results equated with the clinical results, then the research community hails the efficacy of the animal experiments. But when the animal and human results do not match, the proclaimed failure is said to be a result of flaws in experimental design, publication bias or use of young animals for a disease that occurs predominately in elderly humans. Rarely is the use of the animals themselves—not how they are used—questioned.

While most researchers admit the difficulty in extrapolating and applying information obtained from other species to humans, commonly proposed solutions to this colossal obstacle are far from helpful. Neyt et al. suggest that 'clearly profound differences may exist at the gross, microscopic and genetic level between humans and other mammals, and these differences must be appreciated before extrapolating the results of a given study to human clinical practice'.[126] Caution in extrapolating data from animals to humans is another common advice given.[127] In fact, 'appreciation of differences' and 'caution' about extrapolating results from animals to humans are now almost universally expressed in published reports on animal experimental results intended to inform human health. Yet, in reality, how does one take into account differences in drug metabolism, genetics, expression of diseases, anatomy, behavior, influences of laboratory environments, and species and strain-specific physiologic mechanisms and then discern what is applicable to humans and what is not? There is just no established formula or algorithm to do this. Many scientists have recently acknowledged that modeling human disease in animals is extremely problematic but have still argued for their use, instead, to study basic physiologic mechanisms.[128] But again, if we cannot predetermine what mechanisms in what species and what strain of species and in what caging system and even during what time of day are applicable to humans, then the usefulness of the experiments needs to be questioned.

As reviewed earlier, basic research using animals is not effectively leading to new therapies to improve human health, which is the ultimate goal of medical research. A 2003 *American Journal of Medicine* review of 101 of the most heralded basic science discoveries from 1979 to 1983 revealed how unreliable even the 'cream of the crop' basic science findings can be when transferred to human medicine.[129] Following the course of these 101 breakthrough discoveries for up to 20 years, the authors found that only 27 resulted in published randomized clinical

trials, only 5 were approved for human use and just 1 (a blood pressure drug) had a major clinical impact. The authors concluded, 'Even the most promising findings of basic research take a long time to translate into clinical experimentation, and adoption in clinical practice is rare.' Successful translation of basic research is, in fact, fairly uncommon.[130]

Of course, similarities in physiologic mechanisms exist across all species used in experiments and in humans. However, given the way medicine is practiced today, the differences between species appear to far outweigh the similarities and a growing body of evidence is attesting to this. The shortcomings of animal experiments for extrapolation to humans across a wide variety of fields are evident.[131] These include cancer, systemic sclerosis, osteomyelitis, asthma, Huntington's disease, Parkinson's disease, multiple sclerosis, fibromyalgia, alcohol addiction, sepsis (infection of the blood), shock and behavioral disease and psychiatric illness research.[132] Although only a few studies have systematically or critically reviewed whether animal experiments predict human outcomes, these are confirming the unreliability of animal experiments in a number of areas.[133]

Moving science forward

The argument that animal experiments are largely unreliable predictors of human disease mechanisms and health outcomes does not dismiss the fact that some animal experiments have proved successful. Statistically, it is inevitable that some animal experimental results will match human results. As Michael Bracken from the Yale School of Public Health stated, 'given the large number of animal studies conducted, it would be expected that some animal experiments do predict some human reactions'.[134] Based on these successful examples, one may argue that, despite the many limitations, animal experiments have provided useful information. While this assertion would certainly not be inaccurate, the question remains: is animal experimentation the *best* way to get the information we need *today?* The earliest telescopes gave us a glimpse of the universe around us, but they lacked the accuracy for us to target and discern the critical details that would allow us to arrive at a more comprehensive understanding of how the universe functions. Similarly, although animal experimentation may be one means by which we gain some understanding of physiologic and disease mechanisms, the details of these mechanisms that are human-specific and relevant to human health too frequently remain a mystery. Thus we are left with, at best, incomplete and, at worst, inaccurate information.

Even if animal experiments are causally related to the production of data relevant to human health, it does not follow that animal experiments are the only, or even the most efficient, way to obtain relevant data.[135] We are just starting to recognize how minor variations between species can substantially perturb study results. These are just the variations of which we are currently aware. They do not include the many differences between species and strains within a species that we have not yet discovered. These known and likely far more unknown differences render it extremely difficult to unravel and determine what results, if any, from an animal experiment can or cannot be applied to humans. The pivotal argument against using animals as models of disease or to study basic mechanisms is that it is impossible to know in advance which models and which mechanisms will show the same results as in humans. Evidence that some animal experiments accurately predict human results or provide useful information does not detract from the many costly and devastating failures or refute the underlying premise that extrapolation from animals to humans is highly tenuous.

It has been argued that some information obtained from animal experiments is better than no information.[136] This neglects several crucial points that illustrate how a little knowledge can be a bad thing, especially if it is dubious. As we have seen with some of the examples presented, many people have been directly, and often significantly, harmed because researchers were misled by the safety profile of a new drug based on animal experiments. A large number of people volunteering in clinical trials have put their lives at risk based on animal experimental results, which often turned out to be inapplicable to humans. A review in the *BMJ* expressed it thus: 'Biased or imprecise results from animal experiments may result in clinical trials of biologically inert or even harmful substances, thus exposing patients to unnecessary risk and wasting scarce research resources.'[137] We may already be exposing ourselves to numerous carcinogenic chemicals because animal tests were falsely negative. Thus, far from protecting us, animal experimentation often puts us at greater risk.

Furthermore, the indirect human harms caused by the opportunity costs may be substantial. An invalid disease model can lead the industry in the wrong direction, wasting time and significant investment.[138] Repeatedly, researchers have been lured down the wrong line of investigation because of information gleaned from animal experiments that later proved to be inaccurate, irrelevant or discordant with human biology. It's taken more than 25 years of failed HIV vaccine clinical trials for researchers to seriously question the usefulness of NHP HIV models,

and more than 30 years before we realized that the rodent model of diabetes is wrong. A substantial amount of human suffering could have been prevented if instead we had focused on studying HIV and diabetes solely through human-based tests.

Treatments that fail to work or are harmful in animals may be effective and safe in people. Robert Wall and Moshe Shani from the USDA wrote that

> it is interesting to speculate that animal models may be just as likely to exhibit false positive results (compound or devise would be OK in humans but show adverse effects in animal studies) as they do false negatives results (OK in animal studies but have adverse outcomes in human trials).[139]

Animal experimental results may have caused us to abandon countless therapies, which could have worked in humans and alleviated untold suffering. Of every 100,000 chemicals tested in the lab, only about 50 pass on to phase 1 clinical trials. Most don't show enough benefit, aren't easily absorbed in the body or are harmful to animals.[140] But many of these agents may have worked spectacularly in humans.

Aspirin is considered one of the best drugs we have today, despite the fact that its discovery took place over 100 years ago. A recent report examined the safety profile of aspirin in experimental animals.[141] The results showed that in different animal species, aspirin is a cancer promoter, 'harmful if swallowed', a 'respiratory irritant' and causes other serious adverse effects. The report concluded that we are extremely fortunate that we did not rely on animal experiments in 1899 to decide whether to approve aspirin for use in humans by saying 'it is not very likely that any substance with such a profile would make it to clinical trials or to the market today'. This holds true for many well-known drugs, including acetaminophen. Experiments on animals delayed the acceptance of cyclosporine, and Fk-506 (tacrolimus) was almost shelved because of high toxicity in animal experiments.[142] Both drugs are widely and successfully used to treat autoimmune disorders and prevent organ transplant rejection in people. Experiments on mice provided no evidence whatsoever of the efficacy of beta-agonist bronchodilators in the treatment of asthma and suggested that thiazolidinedione anti-diabetes drugs would actually make diabetes worse, in contrast to human studies.[143] A report in *Slate* magazine rightly noted that 'an equal source of human suffering may be the dozens of promising drugs that get shelved when they cause problems in animals that may not be relevant for humans'.[144]

The costs to animals

In addition to causing direct and indirect human suffering, reliance on animal experimentation causes a vastly underappreciated amount of pain and suffering in animals. Annually, more than 115 million animals—including mice, rats, frogs, dogs, cats, rabbits, hamsters, guinea pigs, monkeys and birds—are used in experimentation or bred to supply the research industry worldwide, many of whom endure intense suffering. Approximately 42 percent of NIH-funded research involves experimentation on animals.[145] That translates to more than $12 billion spent on animal experimentation in 2009 alone in the USA, not including the substantial amount coming from the pharmaceutical sector and other governmental and private entities.[146] In the USA in 2009, more than 76,000 animals were subjected to pain without being provided with pain relief.[147] This number does not include the majority of animals used in experimentation (rats and mice), birds, reptiles, amphibians and most animals used in agricultural experiments, all of whom are excluded because they are not considered animals under the Animal Welfare Act (AWA).[148] There are no federal requirements to report the number of these animals used in experimentation or the types of procedures conducted on them. Thus potentially hundreds of thousands of animals may be subjected to painful experiments annually without being provided with any pain relief at all.

In Canada, more than 3 million animals were used in research, teaching, testing and the production of biological products in 2009, an increase from prior years.[149] More than 145,000 were subjected to 'severe pain near, at, or above the pain tolerance threshold of unanaesthetized conscious animals'. The number of animals subjected to this severe pain increased from 55,000 in 1998. At least 11 million animals are used each year in experiments in the European Union.[150]

With rare exceptions, scientific interest always trumps the welfare of the animals. The US *Guide for the Care and Use of Laboratory Animals* stipulates there should be 'proper use of animals, including the avoidance or minimization of discomfort, distress, and pain', but, and this is the important point, '*when consistent with sound scientific practices*' (emphasis added).[151] Thus, the scientific endeavor overrides animal welfare concerns, even for those animals covered by the AWA. All experimentation, no matter the level of pain and suffering, is potentially justifiable by these guidelines. As one bioethicist notes, 'Of particular importance, the appeal to animal welfare in the regulatory guidelines avoids any commitment to limits on what can be done to animals for the sake of

human interests.'[152] Other regulatory guidelines in the USA and abroad are severely deficient in protecting animals from harm.[153] As an example, the AWA does not set forth any standards by which animals are to be kept but leaves that to the USDA.[154] Marian Sullivan, Deputy Chief Court Attorney at the New York State Supreme Court, explains that the AWA requires the USDA to set forth humane care standards. Essentially, however, 'the standards set forth by the USDA . . . require little more than that animals be fed, watered, vetted, and kept in reasonably clean and safe enclosures that allow them to make species-appropriate postural adjustments'.[155] In other words, the AWA is basically a husbandry law that stipulates that animals be fed and be allowed to move about somewhat in their cages.

Ultimately, anything can, and arguably has, be done to animals in the laboratory setting. Every year we poison, bludgeon, shoot, crush, gas, infect, drown, blind, dismember, burn and electrically shock animals in the name of research—often without any pain relief. A survey was recently conducted by one of the top authorities of analgesic use in animals, Paul Flecknell. He found that of the published papers that reported the use of mice or rats in extremely painful, invasive procedures such as burn experiments, spinal cord injury experiments and skull surgeries, post-procedural pain relief was provided to the animals only 20 percent of the time.[156] Moreover, an estimated 50–60 percent of mice and rats receive no pain relief whatsoever both during and after the painful procedures. Signs of psychological distress, including stereotypic or repetitive movements, self-injurious behaviors, near catatonia, vocalizations, inappropriate aggression, fear or withdrawal are all commonly seen in animals in the laboratory.[157] About half of all mice used in experiments are estimated to be afflicted with behavioral stereoptypies.[158]

Earlier in this chapter, studies were presented which demonstrated that animals respond to routine laboratory procedures, such as handling and blood collection, with rapid, pronounced and statistically significant elevations in stress-related markers. Common responses by NHPs to routine procedures include fear grinning, vocalizations, diarrhea and physical resistance (such as struggling or refusing to enter a cage).[159] The simple act of catching an animal and removing him from his cage can cause significant elevation of his plasma cortisone levels.[160] Several studies in monkeys, mice and rats suggest that witnessing other individuals being subjected to unpleasant laboratory procedures is stressful.[161] Animals watching their cagemates being captured for a procedure are affected by 'contagious anxiety'.[162] Rats show significant elevations in

heart rate and blood pressure when present during decapitation of other rats.[163] Cortisol levels shoot up in monkeys able to see other monkeys being restrained and sedated for blood collection.[164]

These findings suggest that the responses in animals to the laboratory procedures are more than mere arousal responses and are indicative of stress and distress.[165] One study found that when an individual in a laboratory coat with a catching net entered the room where monkeys were housed, the monkeys displayed substantial expressions of negative emotion and changes in body temperature indicative of distress.[166] What this and other such studies demonstrate is that animals do not readily habituate to the laboratory environment or procedures; they just don't get used to it. Fear and anxiety are daily phenomena of their lives. Even if we try to make life a little easier for these animals, by housing 'enrichment' and by more routine use of pain medications, for example, ultimately we just cannot get around the fact that the laboratory settings, daily procedures and experiments themselves cause tremendous suffering.

Despite the meager regulations covering only a minority of animals used in experiments, enforcement of even these is pitiful.[167] In the USA, the USDA's Animal and Plant Health Inspection Service (APHIS) is charged with overseeing the AWA. In 2005 the Office of the Inspector General published a scathing report of the USDA's failure to enforce the AWA.[168] It cited APHIS for not pursuing enforcement actions against violators, including repeat offenders, failing to effectively monitor research facilities, and charging minimal fees to violators. The report further concluded that the fines against violators were so minimal that 'violators now consider the monetary stipulation as a normal cost of conducting business rather than as a deterrent for violating the AWA'.

The new gold standard: Human-based tests

The last critical point against the argument that gleaning some information from animal experiments is better than none at all is that this argument assumes there is no alternative means of gaining medical knowledge. In addition to this being a false assumption, there is an array of proven alternative methods of testing that are in wide use today that reveal that we can gain *better* knowledge by not using animals. Sophisticated in vitro tests, human skin models for corrosion tests, genetic techniques, population studies, modeling methods, virtual whole-human modeling, virtual clinical trials, three-dimensional cell and tissue cultures, organs on a chip and imaging studies (using magnetic resonance imaging (MRI), functional MRI and positron emission

tomography scans) are just a few examples of human-based testing methods currently available. Microdosing provides information on how an experimental drug is metabolized and its bioavailability throughout the human body. By administering an extremely small dose (i.e. well below the threshold necessary for any potential pharmacologic, and thus harmful, effect to take place), microdosing can be used safely in human volunteers.

Currently, many of these testing methods are being used in conjunction with animal experiments prior to the conduction of clinical trials. The problem with using both human-based and animal experiments, however, is that the latter may contradict findings from the former. When this occurs, as is often the case, the animal experimental results may be incorrectly favored (leading researchers down the wrong path of investigation) because they represent 'whole animal system' results. However, the animal tests provide the wrong whole systems. For genetic and physiologic reasons that are immutable, animal experiments are less trustworthy than even incomplete systems of the human body.

Some have argued that in vitro or other similar testing methods are simplistic and cannot accurately mimic the complexities of the human body, hence the need for animal experiments. In vitro tests certainly are prone to some of the same problems as animal experiments in that they can be relatively simplistic models of disease or physiologic mechanisms and are not always accurate. But are the animal experiments necessarily *more* accurate or predictive? A multicenter team of researchers evaluated 68 different methods to predict the toxicity of 50 different chemicals.[169] The animal tests were only 59 percent accurate, but a combined human cell in vitro test was 83 percent accurate in predicting actual human toxicity. Human skin cultured cells outperformed live rabbit tests in detecting chemical skin irritants. Tests in rabbits misclassified 10 out of 25 chemical irritants, while the cultured cells classified all irritants correctly.[170] Researchers compared in vitro human tumor cell lines with mouse cancer models for their reliability in predicting clinical phase 2 trial results of 31 potential cancer drugs. The study found that the in vitro tests were reliable in predicting the clinical utility of these drugs for all four cancer types tested, whereas the mouse allograft cancer model (in which cancerous tissue from one mouse is transplanted into another) was not predictive.[171] The human xenograft mouse model (in which cancerous tissue from a human is transplanted into a mouse) was predictive for only two of the four cancer types studied. The study authors concluded that cancer drug development emphasis should be placed on in vitro cell lines.

An in vitro test developed by UK researchers could have predicted TGN 1412's serious adverse effects before it was ever tested on humans.[172] In all of these examples, 'test tube' experiments were far more accurate than whole animal model systems. Asterand has confirmed that studies on human bronchial smooth muscle and pancreatic Islets tissues are far better at predicting human responses to asthmatic and diabetic drugs than animal experiments.[173] One of the best features of in vitro methods is that we have better control and understanding of the testing parameters. With animal experiments, especially because of the many inter-species differences in physiology of which we are not aware, our control and understanding of the testing influences are greatly limited in comparison. Regardless of the preference to study whole biological systems, non-human animals are not the correct systems. An understanding of human physiology is critical. And it cannot be overstated that in order to be the most accurate and predictive as possible, in vitro tests must use *human* cells and tissues, not cells from other species, otherwise interspecies differences still come into play. While there is no perfect predictive approach to human medicine, a combination of human-based testing methods, including in vitro tests, will likely get us closer to the true answers than animal experiments, which are inherently flawed. Human-based in vitro tests may not always be accurate predictors of human responses, but they have great potential to become more accurate, particularly as new methods are developed that are closer to depicting whole human systems. At a fundamental level, non-human models, on the other hand, cannot be accurate, and cannot be made to be accurate, because of distinctions in genetic make-up and expression, and evolutionary issues such as causal disanalogy.

Biotechnology company Selventa (formerly called Genstruct) has compared animal models of human disease with the actual human diseases.[174] For instance, it has studied mouse and rat models of type 2 diabetes. In many cases, it found that other than having aberrations in insulin signaling and glucose levels, there was no similarity between the animal model and the human disease condition. 'If you're developing a drug in that animal model, it's clearly not going to work in humans because they have a different disease,' says Keith O. Elliston, Selventa's president and chief executive officer. Selventa focuses on human-based tests. These are far from simplistic and deal with the complexity of human biological systems. It has developed in vitro models that include all the genes, proteins and metabolites present in human cells. The company then applies artificial intelligence tools to work through all the

predicted and observed relationships among these components and put them in context within the complex system.

Many other forward-thinking companies are exploring modern alternatives. Pharmagene Laboratories, based in Royston, UK, is the first company to use only human tissues and sophisticated computer technology in the process of drug development and testing; it does not conduct any animal experiments.[175] With tools from molecular biology, biochemistry and analytical pharmacology, Pharmagene conducts extensive studies of human genes and how drugs affect those genes or the proteins they make. One of the co-founders asked, 'If you have information on human genes, what's the point of going back to animals?'[176]

Neurologists and other neuroscientists collaborating on the Miami Project to Cure Paralysis are using cutting-edge science to model human spinal cord injury.[177] Studying spinal cord-injured patients, researchers are gleaning a more complete understanding of human spinal cord injury. For example, they are comparing postmortem spinal cord tissue with MRIs of living patients to determine what changes in cells and tissues are detrimental. The project correlates neurological function, neurophysiology and findings from imaging studies and tissue pathology to design targeted therapies to improve the quality of life of injured patients and prevent further damage after acute injury. After only a few years, this project has made several notable discoveries about human spinal cord injury that were not made through animal experiments. It is the first project to provide evidence that humans possess specialized nerve circuitry that influences walking and could possibly be enhanced by rehabilitation training. It is also the first to show conclusive evidence of a critical neurological feature, chronic demyelination (disruption of the nerve coating necessary for proper nerve signaling) after spinal cord injury in humans, and to conceive and develop a novel intra-operative monitoring technique that makes spinal surgery safer.

In response to the limitations of animal immunology experiments for human health research, scientists at Stanford University are working on a 'Human Immunology Project'.[178] The investigators are using high-throughput screens to catalog a host of cellular parameters.[179] They are using a systems biology approach to understand the many facets of the human immune system and how the whole system fits together. Researchers have now created a virtual model of all the biochemical reactions that occur in human cells.[180] A major report released in 2007 by the National Academies' National Research Council (NRC) called for a transformation in toxicology testing—one that largely shifts away from

animal experiments.[181] The NRC recommended the development and use of in vitro methods using human cells, in combination with computer modeling and other testing techniques, to evaluate changes in biologic processes and markers that would indicate toxicological effects in the human body. It concluded that not only would these new testing methods be more evidenced-based but they would also save significant resources and time in comparison with animal toxicity experiments.

These few examples of human-based testing methods are just a tiny sample of the sophisticated non-animal approaches currently available. Human-based methods must be validated and the ones that have undergone validation thus far are largely proving to be better than animal experiments in predicting human responses. While gaining momentum, human-based tests are still in their infancy and there are many areas in medicine where these methods need further development. This fact has been used to argue for the continued use of animal experimentation. But not having a viable alternative is not sufficient justification for continuing a misguided research paradigm. Instead, this line of thinking prevents us from any true commitment to finding new or improving existing alternative testing methods. It will cause us to continue to waste years and precious research dollars on sub-par methods, place humans at risk, cause suffering in animals, with perhaps the greatest tragedy of all being that we would likely abandon therapies that would have been effective.

Financial investments in the study of alternative testing methods pale in comparison with investments in animal experimentation.[182] For example, the US Government's agencies have spent less than $10 million over a ten-year period on validating alternatives for regulatory use, and validating alternative methods is rarely a priority for government funding.[183] The development of human-based alternatives to animal research is an underdeveloped field largely because so few resources are devoted to its development as a result of our commitment to animal-based methods.[184] Another major hurdle to the development and use of non-animal testing methods is that government regulations tend to require far more validation than was ever required, if at all, for the animal experimental methods they are intended to replace. Ironically, these new methods are often required to be validated against existing animal experimental methods, most of which have never been validated themselves.[185] This creates a double standard that allows the acceptance of most animal experimental methods as the 'gold standards' (based on tradition, rather than proven efficacy), providing a disincentive to the development of alternative methods.

An even larger problem with policies requiring the validation of human-based tests against animal experiments is that the latter are unlikely to predict human responses consistently, and may not even be consistent in general. Thus a human-based model might actually be consistent and predict human responses but would fail validation, while it is the animal test that is in fact inferior. Additionally, a final hurdle is that regulatory agencies do not usually mandate the use of alternative testing methods, where they exist and have been proven valid, in place of the traditional animal experiments. Thus, there is little incentive for pharmaceutical companies and others to switch gears and use alternative methods in research and drug development if they are already wedded to an animal model. Arguably, there has been a net loss of ground because alternative human-based methods, which would have likely gotten us further scientifically, have been neglected in favor of animal experimentation. It is time for this to change. It is incumbent upon investigators and research-supporting institutions to prioritize the replacement of animals in experiments. Failing to do so means delaying the development of more effective and accurate research techniques that could save thousands or millions of human and animal lives.

Dubious experiments we can eliminate

In the short term, we can agree that many experiments currently being conducted could be eliminated today. Consensus can be reached that a substantial proportion of animal experiments are highly irrelevant to human health. A quick exploration of some recently funded animal experiments attests to this. A survey of experiments conducted at US universities that were funded by the NIH was conducted in 2008 through the use of two databases: the Computer Retrieval of Information on Scientific Projects (CRISP), maintained by the NIH, and the CRISP*er* database, maintained by the non-profit Sunshine Project.[186] A literature review provided additional information. Examples of experiments funded by public tax dollars include:

> An experiment conducted between 2006 and 2007 by Emory University School of Medicine cost more than $97,000. In this experiment, muscle-recording electrodes were placed in anesthetized cats' hindlimb muscles. The cats were positioned over a treadmill, with their heads fixed in stereotaxic frames. Their brainstems were then cut and all brain matter above the incision removed. Anesthesia

was then eliminated and, as the cats initiated spontaneous stepping movements, the treadmill was turned on and muscle activity was recorded while the cats' heads were positioned in three different ways. The results were compared with results from intact cats (cats with brains intact). The main results suggest that modifying head pitch in a walking decerebrate (cerebral brain removed or disconnected) cat causes significant muscle activity changes that are similar to what occurs in an intact cat.[187]

At the Keck School of Medicine in southern California, an area of the frontal brain necessary for the sense of smell was removed through aspiration in male hamsters. The hamsters were then tested for their sexual attraction to male versus female hamsters. The goal of this experiment was to assess if, and how, testosterone, sexual experience and chemosensory cues play a role in sexual motivation in male hamsters. Between 1997 and 2006, this and similar experiments cost more than $1.8 million.[188]

At the University of Washington, sparrows were caught from the wild and deafened by puncture of their tympanic membranes. Their song production was then measured. The primary goal of this experiment was to assess whether deafening sparrows affected their singing and the seasonal growth of their song nuclei. This and other similar experiments on sparrows cost the public more than $3.4 million between 1997 and 2007.[189]

A series of mating behavior experiments on ferrets at Boston University between 1998 and 2007 cost more than $4 million. One of the major findings suggested that damage by electrical lesions to both sides of a part of the hypothalamus in the brain causes male ferrets to display a preference for sexual and body odors from other males over females.[190]

Experiments conducted at the University of California in which rats received repeated electric shocks revealed that as a defensive mechanism against a perceived threat, rats will hide and freeze in a familiar enclosure. This and similar experiments cost the public more than $8.6 million between 1997 and 2007.[191]

Between 1997 and 2007, the University of Michigan spent more than $21 million on experiments to assess whether alcohol reinforces the use of other drugs in monkeys.[192]

The examples presented here are far from isolated cases. Public funds are used to support numerous dubious experiments at medical centers and universities throughout the USA and abroad, regardless of their lack of relevance to human health. There are much better ways to use our tax dollars to improve human health than the examples above. Rather than continuing to pour millions of dollars each year into experiments on drug and alcohol use in animals, we could instead fund treatment centers for drug abusers. Rather than studying the song nuclei in sparrows, a nucleus that humans don't even have, we could instead fund experiments such as functional MRI studies of the changes in various areas of the brain in humans with deafness. Why not divert more funding to studies on human spinal cord injury, such as the Miami Project, rather than remove the brains of cats to monitor their spontaneous stepping activity, especially when humans, unlike cats, have little to no spontaneous stepping activity without input from the higher brain? Support for these experiments will inevitably revolve around suggestions that they will help elucidate underlying physiologic mechanisms that will one day have human health applicability. However, such a connection is extremely doubtful as we have seen how underlying mechanisms can differ so vastly between species. On the other hand, there is no doubt that there are many other ways to use these funds, which will benefit humans and will do so without causing animals harm.

Someone might claim that we don't know what benefit animal experiments, particularly basic research, may provide down the road. But as bioethicist Bernard Rollin pointed out, 'if that were a legitimate point, we could not discriminate between funding research likely to produce benefits and that unlikely to do so; however, we do. If we appeal to unknown but possible benefits, we are literally forced to fund everything, which we do not.'[193] Many researchers and funding institutions are aware of the fact that basic research on animals has come under intense criticism because society has hinted that that there are limits to what it would fund in terms of knowledge for the sake of knowledge. Consequently, much basic research on animals is now conducted under the guise of applied research.[194] However, as demonstrated in this chapter, the usefulness of basic research on animals to produce medical treatments is highly questionable. And we have seen how so few of even the most highly regarded studies in basic research ever translate to human benefit. Given the highly questionable usefulness and the immense suffering animals in laboratories experience, the appeal to serendipity in research is insufficient to justify an animal experiment.

Steps we can take

Regardless of the rationalizations given to support the use of animal experiments, the final test of their success is whether or not they improve our health and lead to new, effective, treatments or preventions. In this, they are largely disappointing. Animal experiments are proving to be extremely unreliable in predicting human outcomes. Is this a risk we want to continue to take? In drug production, the failure rate is at least 92 percent. More of us need to ask why we are failing so often. A 92 percent failure rate of *anything* should be cause for alarm. A 92 percent failure rate in drug development should likewise be unacceptable. Because the practice of animal experimentation is so entrenched in our current research paradigm, scientists who question the foundational relevance of animal experiments are often marginalized within the scientific community. Alternative opinions and studies critically examining the relevance of animal experiments are rarely, with some of the exceptions provided in this chapter, given an opportunity to be published in the biomedical literature. The failure by the scientific community as a whole to tolerate different opinions and publish critical examination of animal experiments is contrary to the very spirit of science and is a major obstacle to the advancement of human health.

Moving away from animal experimentation will no doubt take time. There are steps we can take today, though, to move us in a positive direction: a direction that is immensely beneficial to humans and animals, and that embraces more sophisticated and accurate testing methods. A thorough examination of how we spend our research dollars and the relevance of animal experiments to human health is vital. The public deserves accountability for how we spend their money. These steps require greater transparency in animal experimentation so that the evaluation of the experiments' human relevance and accurate assessment of the costs to both humans and animals can be made. They should include:

1. prioritization of the conduction and publication of critical and systematic studies evaluating the human health relevancy of animal experiments;
2. identification and immediate replacement of animal experiments agreed to be highly irrelevant to human health;
3. provision of transparency and registration of all animal experiments conducted by public and private institutions similar to clinical trials registries (such a registry should include the numbers and types

of animals used, details about the health and welfare of the animals, funding amounts, housing procedures and details on the experimental procedures conducted);
4. demand for a serious and primary dedication to development of non-animal testing methods; and
5. mandate the use of validated non-animal alternatives that currently exist in place of animal experiments.

We owe it to the public to use the best possible research methods. These are human-based tests. Their use also has the added benefit of avoiding the use of animals in harmful experiments. All we need is the willingness to question our own assumptions and the dedication to follow where this leads us. By doing this we will create a new gold standard for medical research—one based on sound science.