

WellBeing International  
**WBI Studies Repository**

2019


## Replacing Animal Tests to Improve Safety for Humans

Kathy Archibald  
*Safer Medicines Trust*

Robert Coleman  
*Consultant*

Tamara Drake  
*Center for Responsible Science*

Follow this and additional works at: <https://www.wellbeingintlstudiesrepository.org/appamet>

 Part of the [Bioethics and Medical Ethics Commons](#), [Design of Experiments and Sample Surveys Commons](#), and the [Laboratory and Basic Science Research Commons](#)

---

### Recommended Citation

Archibald, K., Coleman, R., & Drake, T. (2019). Replacing animal tests to improve safety for humans. In *Animal Experimentation: Working Towards a Paradigm Change* (pp. 417-442). Brill.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact [wbisr-info@wellbeingintl.org](mailto:wbisr-info@wellbeingintl.org).



# Replacing Animal Tests to Improve Safety for Humans

*Kathy Archibald*

Director, Safer Medicines Trust

*kathy@safermedicines.org*

*Robert Coleman*

Consultant to Safer Medicines Trust and Center for Responsible Science

*robt.coleman@btinternet.com*

*Tamara Drake*

Director of Research and Regulatory Policy, Center for Responsible Science

*tami.drake@crs501.org*

## 1 Introduction

Animal safety testing for new medicines is arguably the most difficult use of non-human animals (hereinafter referred to as animals) to challenge, for two reasons: first, it is required by governments (regulatory testing); second, protecting patients is a vital goal, and it seems intuitively obvious that animal tests *must* protect patients. Animal testing became institutionalized in the mid twentieth century (Parke, 1994) in response to early drug disasters, with the aim of preventing further tragedies. However, even the laudable aim of protecting patients cannot justify animal testing, unless it is the most effective means to ensure the safety of medicines. European Union (EU) law (European Parliament, 2010, Directive 2010/63/EU) states that animals must not be used if a non-animal method could achieve the same purpose. So, it is crucial to know how well animal tests predict the safety of medicines, and whether any other methods are equally or more predictive. In addition to the question of predictive value, other important issues must also be taken into consideration, including the efficiency of different methods in terms of time and costs; and the ethical acceptability of using animals, if their use is deemed to be of irreplaceable value.

The issue of whether animals should be used as human surrogates for safety testing is highly contentious; individual views range from no use of animals

is acceptable, to any use is justified if it is ultimately for the benefit of human health. In a 2016 Ipsos Mori poll for the United Kingdom government, 35% of respondents felt that “animals should not be used in any scientific research because of the importance I place on animal welfare”; and 51% felt that it is unacceptable to use animals to test chemicals that could harm plants or the environment (Clemence and Leaman, 2016). In the United States (US), a 2017 Gallup poll found that 44% of adults considered medical testing on animals to be morally wrong (Jones and Saad, 2017). A 2015 survey by the US Pew Centre found that 50% of citizens “oppose the use of animals in scientific research” (Pew Research Center, 2015). Thus, it is questionable whether the continued use of animals in research or testing has a mandate from society. What is clear is that their use should be replaced as soon as scientifically possible (European Parliament, 2010, Recital 10), which brings us back to the question of their effectiveness relative to other methods.

Before any new methods can be approved for use in regulatory safety testing, they must be shown to be at least as effective as the methods they are designed to replace, a logic that cannot be faulted. However, herein lie a number of problems. First, we do not know how valuable existing animal-based methods actually are, as none have ever been formally validated in the manner required for potential replacements. One reason for this is that the formal process of test-method validation is so slow, expensive, and demanding, in its current format, that it represents an effective block to testing existing accepted methods and a significant barrier to testing new methods. The situation is further complicated by the fact that the “gold standard” with which new data must be compared, is usually animal data that is of unknown value. We strongly believe that the most relevant gold standard should be human data.

In this chapter, we propose a new, pragmatic approach that could accelerate the replacement of most, if not all, regulatory animal tests with superior tests based on human biology. We also propose that changes to the requirements for safety testing, issued by the US Food and Drug Administration (FDA), must be made in order to enable the use of superior new tests, which are currently disadvantaged by the outdated language of the regulations. But first, it is imperative to establish some level of understanding of the efficacy of existing animal-based methods in order to know whether any possible replacement is better or worse.

## 2 Learning from Clinical Experience

In order to quantify, as best as we can, the effectiveness of animal tests for predicting the safety of medicines, we can begin by assessing about half a century

of clinical experience. We have a significant amount of information gleaned from the use of approved medicines in human subjects and can identify, for such medicines, where animal-based testing failed to predict human safety issues; and whether non-animal methods now exist that would be able to identify the toxic effects that were missed by animal tests.

Many medicines that have been judged safe enough for testing in humans, following all the required safety tests *in vitro*, and in at least two species of animals, have gone on to cause serious adverse reactions in the first volunteers to try them: participants in clinical trials. The most infamous examples include the trials of the candidate medicines TGN1412 in the UK, BIA 10-2474 in France, and fialuridine in the US. TGN1412 is a monoclonal antibody that was intended to treat B cell chronic lymphocytic leukemia and rheumatoid arthritis. The clinical trial, in London in 2006, hit headlines when all six young men in the Phase I (safety assessment) trial were rushed to intensive care with multiple organ failure. Miraculously, they all survived; but they were told that they face “a lifetime of contracting cancers and all the various autoimmune diseases from lupus to MS, from rheumatoid arthritis to ME” (Leppard, 2006). TGN1412 was shown to be safe in monkeys at doses 500 times higher than those that nearly proved fatal to the volunteers (St. Clair, 2008).

In January 2016, a Phase I study of the drug BIA 10-2474 conducted in Rennes, France, left one initially healthy volunteer dead, and four volunteers with serious neurological damage (Sharav, 2016). The drug was intended to target a wide range of conditions including pain, hypertension, multiple sclerosis, obesity, and cancer. Experts convened by the French National Agency for Medicines and Health Products concluded that the compound being tested had caused an “astonishing and unprecedented” reaction in the brain. Why this was not clear in early trials on animals is “inexplicable,” according to the expert panel’s report (Bisserbe, 2016). The drug had been tested in mice, rats, dogs, and monkeys, with few ill effects, despite doses up to 650 times stronger than those given to the volunteers (Temporary Specialist Scientific Committee, TSSC, 2016). A subsequent study indicates that an off-target effect, which can be species dependent, may explain why animal tests in multiple species did not identify the deadly neurological effects (van Esbroeck et al., 2017). The off-target effect could only be found using human cells *in vitro* and in humans.

In 1993, a combined Phase I/Phase II clinical trial (to test both safety and effectiveness) of fialuridine, a potential hepatitis B treatment, conducted by the National Institutes of Health (NIH) in the US, caused unexpected and devastating reactions, such as jaundice, liver failure, and multiple organ failure. Five of the 15 participants died. Emergency liver transplants saved two others. Previous toxicity tests in animals, including a six-month trial in dogs, had given the drug the green light for testing in humans (Thompson, 1994).

Many more medicines have passed both preclinical (mainly animal-based) safety tests and human clinical trials and still gone on to cause serious adverse reactions in patients. This illustrates how difficult it is to predict safety for humans, in general, and even more so for particular members of the human population. There is enormous genetic variability between people, and individual reactions will vary with age, sex, ethnicity, health, diet, environment, and unique genetic characteristics. Adverse drug reactions (ADRs) are now a leading cause of death, killing 197,000 people in the EU each year (European Commission, 2008), and over 125,000 in the US (Light, 2015). In addition to this devastating human cost, the financial cost of ADRs is astronomical, calculated at €79 billion per annum in the EU (European Commission, 2008). A study of new drugs approved by the US FDA between 2001 and 2010 found that 32% were affected by a post market safety event (Downing et al., 2017). Another study of all 454 drugs approved in the US and Canada from 1992 to 2011, found that 52% (236 drugs) were either withdrawn from the market or restricted by a serious safety (black box) warning within the 20-year period (Rawson, 2013). Black box warnings are reserved for ADRs that may lead to death or serious injury. Half of them are detected and documented within seven years after drug approval, during which time their market uptake and sales volume may be explosive. There is a compelling argument that “when safe and effective therapies already exist, any new drug should be considered a black box” (Lasser et al., 2002). When the costs of withdrawn and restricted drugs, as well as failures during development, are factored into the total cost of developing a successful new drug, this results in an estimated average of US\$4 billion and could reach as high as US\$12 billion (Herper, 2012).

It is argued that most ADRs that were not detected in clinical trials are very rare and/or idiosyncratic, i.e. unique to the individuals who suffered them and, therefore, impossible to identify until large numbers of people are exposed to the drug, once it is on the market. The implication of this position, accepted by our governments, is that we are powerless to prevent rare or idiosyncratic ADRs and must simply accept them as an unavoidable risk of medicine. The problem is that even if an adverse reaction is rare, when millions of people are taking a drug, large numbers will be affected. Not only are hundreds of thousands of people killed, it is estimated that a total of over 80 million ADRs result in 2.7 million hospitalizations each year; in addition, pain, discomfort and dysfunction affect physical or cognitive function and can lead to falls and cause potentially fatal vehicle accidents (Light, 2015). While it can be argued that responsibility for failing to protect participants in clinical trials from dangerous drug candidates lies mainly with animal testing, neither animal tests nor human trials have been able to prevent the large numbers of ADRs

documented above. Clearly, it is imperative to examine all aspects of safety testing, to find every weak point, and to take action to address them all.

### 3 Clinical Trial Flaws

Many problems with clinical trials have been identified, and are being addressed to varying degrees (Evans, Thornton and Chalmers, 2006; Goldacre, 2012). For example, most volunteers in Phase I trials are young men, who are not representative of the often elderly and/or female patients who will be taking the medicines (Abadie, 2010; Johnson et al., 2014). The conduct and reporting of trials are beset by a host of biases, such as selective reporting of results, to emphasize benefits and disguise risks; and non-publication of trials where the desired outcomes were not achieved (Goldacre, 2012; Harris, 2017). In biomedical research as a whole, 235 types of bias have been documented (Chavalarias and Ioannidis, 2010). Many doctors have been campaigning for many years to tackle these biases, which make a mockery of the evidence base for medical treatments. Doctors and patients are unable to choose the best treatments without full, unbiased disclosure of the magnitude of their benefits as well as their risks. With UN endorsement, the AllTrials campaign (2016) has published a roadmap towards ensuring that all clinical trials are properly reported to improve the evidence base for medicine, which is currently badly incorrect and incomplete.

### 4 Preclinical Animal Tests

To assess the performance of preclinical animal tests, the most direct comparison is between data obtained during preclinical (animal) and clinical (human) trials. We have already mentioned three extreme examples of disastrous clinical trials, where animal tests failed to predict toxicity with devastating consequences. But are these isolated examples, and do animal tests usually predict serious toxicities before they manifest in people? This is difficult to answer quantitatively because compounds that are shown to be toxic in animal tests do not usually progress to clinical trials. However, we do know that 95% of potential new drugs fail during clinical trials (Arrowsmith, 2012), either because of toxicities that were not predicted, or because they lack the therapeutic efficacy that was predicted. Data obtained by Freedom of Information legislation shows that from 2010–2014, 7,187 people in the UK suffered serious unexpected ADRs during clinical trials and 761 died, although none of the

deaths could be proven to have been “directly caused” by the test drug (Bagot, 2015). More than 2,600 patients participating in clinical trials in India died between 2005 and 2012, and nearly 12,000 suffered serious adverse effects. Of these, 80 deaths and more than 500 serious adverse effects were directly attributed to the drug being trialed (Nair, 2015). Clearly the record of animal tests in predicting safety is poor.

Another example that illustrates the dangers of both misleading preclinical animal studies and non-publication of clinical trials is lorcinide, which is estimated to have killed over 100,000 people in the US alone over the course of the 1980s (Bruckner and Ellis, 2017). Lorcinide and other anti-arrhythmic drugs (most of which have since been withdrawn) were prescribed routinely to patients recovering from heart attacks, on an assumption, bolstered by the strength of their effectiveness against experimentally induced arrhythmias in animals, that they would help to prevent early deaths. A clinical trial in 1980 indicated that, in fact, they *caused* more deaths; but the trial was not published until 13 years later, to the great regret of the authors, who realize that they could have helped avert tens of thousands of unnecessarily early deaths (Hampton, 2015).

An important point that must be made is the difference between predicting the presence or the absence of toxicity. It seems intuitively obvious that if a compound is overtly toxic for an animal, it is not unreasonable to suspect that it will also be toxic in humans. In a series of studies, Bailey, Thew and Balls (2013, 2014, 2015) examined the likelihood that such suspicions would be correct. They analyzed a data set of 2,366 drugs, for which both animal and human data are available, in the most comprehensive analysis of publicly available animal toxicity data ever compiled. Crucially, they used the appropriate statistical metrics of likelihood ratios, for the first time, to question critically the value of the use of the main preclinical animal species (i.e., rats, mice, rabbits, dogs, and monkeys) in the testing of new human pharmaceuticals. They found that the *presence* of toxicity in animal tests indeed shares some degree of correlation (above random chance) with the presence of toxicity in humans, although such correlation is too variable to be regarded as predictive, as has been demonstrated by many previous studies (Fourches et al., 2010; Geerts, 2009; Green, 2015; Hackam and Redelmeier, 2006; Heywood, 1990; Igarashi, 1994; Ioannidis, 2012; Knight et al., 2006; Matthews, 2008; Pound et al., 2004; Pound and Bracken, 2014; Perel et al., 2007; Salsburg, 1983; Seouk et al., 2013; Spriet-Pourra and Auriche, 1994; Wall and Shani, 2008, van Meer et al., 2012). More importantly, they found that animal tests have essentially no ability to predict the *absence* of toxicity, the very reason for their use in preclinical testing: candidate drugs proceed to testing in humans when no toxicity shows up in tests on animals.

So, while animal tests undoubtedly prevent some toxic compounds from reaching humans, they cannot predict safety for humans. Thus, as we have seen, their use creates a false sense of security. A study published in 2012 found that animal tests missed 81% of the serious side effects of 43 drugs that went on to harm patients (van Meer, 2012). This is disastrous not only for patients but also for the pharmaceutical industry, which is in crisis and urgently needs to stem the unsustainable rate of late-stage attrition of new medicines.

In addition to letting dangerous medicines slip through the net (through *false negative* results), promising medicines may be wrongly discarded due to animal toxicities that do not affect humans (*false positives*). Clear examples of this are few, as any compound causing ADRs in animals is extremely unlikely to progress to the clinic; therefore, its safety profile in humans remains unestablished. However, there are examples. Glivec, an effective cancer treatment, was almost abandoned during development, as it caused liver damage in dogs. Fortunately, its remarkable success in human cells *in vitro* and in early trials in leukemia patients enabled its continued development (Capdeville, 2002). Similarly, tamoxifen was almost lost as a cancer treatment because it causes liver tumors in rats (Carthew, 1995). Evidence for this may also be gleaned from drugs introduced before rigorous safety testing became mandatory. For example, aspirin, introduced over a hundred years ago, has proved useful for pain treatment ever since, but it is highly doubtful it would ever have appeared had it been subjected to modern animal-based safety testing (Hartung, 2009). Other such examples include, benzodiazepines, methylxanthines, such as caffeine, and beta-blockers. It is a similar story with many foodstuffs, such as chocolate and garlic, which are well tolerated by humans but prove toxic to dogs and cats (Cortinovis and Caloni, 2016).

Furthermore, not all failures in animal studies involve adverse events. Many reflect a lack of apparent efficacy in the chosen animal species, a finding that usually consigns a prospective candidate to the waste bin. However, on occasion, a “failed” compound has a champion, sufficiently dogged to proceed despite such a setback. A particularly good example of this are statins (Endo, 2010), the best selling drugs in history, which nearly never emerged from preclinical testing. Based on the belief that elevated levels of cholesterol in the body are, in some way, responsible for coronary heart disease, many approaches to reducing circulating cholesterol have been explored; one of these was through inhibition of HMG-CoA reductase, a key enzyme in cholesterol biosynthesis. In 1976, a report of the first statin, *compactin*, was published (Endo et al., 1976), describing how it inhibited this key enzyme and reduced cholesterol synthesis in isolated mammalian cells. Unfortunately, when tested in rats, this compound proved to be without effect on serum cholesterol levels



(Endo, 2010); and if not for the persistence of the scientists working on the project, this would have been the end of the story. However, despite its lack of effect in rats, compactin was, by chance, found to lower circulating levels of cholesterol in chickens and, subsequently, in other animal species; and the race to develop the ultimate statin was on. Although statins' effectiveness in saving lives is now controversial (de Lorgeril and Rabaeus, 2015), there is no doubt of their effectiveness in lowering cholesterol in humans.

It is hard to imagine a world without antibiotics, the most life-saving class of drugs ever discovered. Yet, the world's first antibiotic, penicillin, was almost lost to humanity because Alexander Fleming concluded that its rapid clearance from the bloodstream in a rabbit would prevent it from being systemically effective (Hare, 1982). For twelve years following his discovery of "mould juice", Fleming pursued its use merely as a topical antiseptic, until Florey and Chain resurrected interest in its greater potential. Fleming later commented to his student, Dennis Parke, who became an extremely influential pioneering toxicologist: "How fortunate we didn't have these animal tests in the 1940s, for penicillin would probably never have been granted a license, and possibly the whole field of antibiotics might never have been realised" (Parke, 1994).

We have discussed, above, examples of *false negatives* and *false positives* for safety, as well as false negatives for efficacy. Finally, there are many examples of false positives for efficacy, i.e. drugs that were effective in animal tests but turned out to be ineffective in humans. They include the vast majority of new cancer treatments, which have one of the highest failure rates (96%) in clinical trials (Hutchinson and Kirk, 2011); all putative disease-modifying treatments (more than 300) for Alzheimer's disease to date (Langley, 2014; Lowe, 2017); more than 100 candidate AIDS vaccines, all of which were effective in non-human primates, as well as other animal models (Sheets et al., 2016); more than 100 drugs for stroke (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies, CAMARADES, 2017); and 150 drugs for sepsis, the leading cause of death in intensive care units (Seok et al., 2013).

The CAMARADES group was founded to study the translatability of stroke studies from animals to humans, and later expanded to include a number of other diseases that share a high rate of translational failure. They have found that the poor quality of animal studies confounds research in all areas they have studied, so far (CAMARADES, 2017). These failed treatments have been tested on patients in clinical trials. When the director of the US NIH, Dr. Francis Collins, learned of the poor quality of the animal studies that led to clinical trials of treatments for amyotrophic lateral sclerosis (ALS, also known as motor neuron disease), he said: "Humans were being put at risk based on that kind of data, and that took my breath away" (Harris, 2017). This *reproducibility crisis* is now receiving much attention, and many initiatives have begun to attempt

to improve standards and the quality of animal research and reporting (Harris, 2017). However, a review of developments in the field of stroke found that, despite researchers' adherence to recommendations intended to improve the quality of preclinical stroke studies for over 10 years, there is no evidence of an increased rate of successful translation (Sutherland et al., 2012). It is possible that many years and many precious resources will be invested in attempts to surmount a problem that may ultimately be insurmountable, due to inherent interspecies differences in crucial genetic and biochemical systems. Dr. Fiona Godlee, Editor of British Medical Journal, posed the pivotal question: "Where would you place the balance of effort: investment in better animal research or a shift in funding to more clinical research?" (Godlee, 2014).

## 5 Other Preclinical Tests in Current Use

Preclinical testing also includes a number of *in vitro* and *in silico* (computer modeling) methods, whose record of predicting safety must also be acknowledged as lamentable. Indeed, the UK government always uses this argument in defense of animal testing, stating that "prior to testing in animals, new drugs are tested in batteries of *in silico* and *in vitro* tests, including, where available and validated, tests using human tissue samples" (UK Department of Health, 2012). However, many of these tests are based on animal cells and tissues; and even the human-based ones generally do not represent the latest state of the art models, which have long ago moved on from 2D to 3D models and recognized the importance of incorporating more realistic physiological features, such as multiple interconnected organs, metabolic activity, and fluid circulation, among others. Technologies are now becoming available that can identify toxic liabilities more accurately than animal tests and furthermore, some of them are able to identify subtle signals of toxicities that only manifest in rare individuals (Xu et al., 2008). This could enable the detection of potential rare ADRs that are currently unpredictable (Kenna, 2017). Thus, these human biology-based technologies should be recognized as a truly disruptive (i.e. revolutionary) technology, with the potential to transform toxicology from an imprecise science based on inter-species extrapolations to a predictive science based on a deep understanding of human pathways of toxicity. A particularly powerful approach has recently been described by Theil et al. (2017), in which they use a system to "contextualize *in vitro*" data to reflect an *in vivo* situation in patients through computer modeling, using data derived from both human cells and clinical experience. A system such as this allows the identification of potential biomarkers of toxicity, and the use of these biomarkers in an *in vitro* setting to predict potential toxicity in clinical use.

## 6 Non-animal Technologies

Remarkable scientific advances have created a new generation of more relevant and predictive toxicological tools. They include human tissue created by reprogramming cells from people with the relevant disease (dubbed *patient in a dish*); *organ on a chip* devices, where living human tissue samples on a silicon chip are linked by a circulating blood substitute; a variety of computer modeling approaches, such as virtual organs, patients, and clinical trials; and microdosing studies, where tiny doses of drugs given to volunteers allow scientists to study their metabolism in humans, safely and with unsurpassed accuracy. There are also humbler, but no less valuable, studies in ethically donated “waste” tissue. Together, these innovations provide invaluable insight into the functioning of the integrated human system. Such tests are frequently able to detect side effects that were missed by preclinical animal tests. For example:

- A micro-liver (called HepatoPac) comprising human liver cells is able to predict liver damage from fialuridine, the potential hepatitis B treatment that killed five patients in the devastating 1993 clinical trial (Baker, 2011). Furthermore, the same technology is able to identify many other liver-toxic drugs that were missed by animal testing (Xu et al., 2008).
- Following the trial of TGN1412, a method using human cells was rapidly developed to model the *cytokine storm* experienced by the volunteers (Stebbins et al., 2007).
- The US government’s initiative, *Toxicology in the 21st Century* (Tox21), has tested 10,000 chemicals using a panel of human cell-based assays (National Center for Advancing Translational Sciences, 2016). These are automated high-throughput screening assays that expose cells to chemicals and then screen them for changes that could suggest toxic effects. The use of this panel of assays enabled the identification of important safety aspects of drugs and chemicals “markedly better” than toxicity tests in animals (Huang et al., 2016). The human *in vitro* data were mainly assessed against rodent data, as human *in vivo* data are sparse. As expected, the Tox21 data better predicted human toxicity endpoints than rodent data.

Non-animal tests are often faster and cheaper, as well as more accurate and reliable (Balijepalli and Sivaramakrishan, 2017; Bracken, 2009; Garner et al., 2017; Krul, 2014; NIH, 2008). Some of the more valuable technologies are expensive, but worth it—there is nothing more expensive than getting the wrong answer. Human tissue company, Biopta (2017), estimates an average saving of US\$7 for every US\$1 invested in predictive human assays. Director of the US NIH, Dr. Francis Collins, recently predicted before US Congress that within 10 years, human biochips “will mostly replace animal testing for drug toxicity and

environmental sensing, giving results that are more accurate, at lower cost, and with higher throughput” (US Senate Committee on Appropriations, 2016). However, new methods will not automatically implement themselves. Pharmaceutical companies would make much greater use of them if governments encouraged it, but inflexible requirements for animal tests are a major deterrent. Reliance on animals is so entrenched and institutionalized, that the system is “locked-in” (Frank, 2004). Intervention is necessary to overcome the many factors contributing to entrenchment against change.

## 7 Validation

Quite correctly, new technologies must be shown to be robust, reliable, and fit for purpose before they can be recommended for use in any regulatory safety-testing regime. The current validation process involves testing by several different laboratories and is tremendously demanding, taking an average of 10 years and costing up to US\$1 million (Hartung, 2013). This approach protects the status quo by making the bar for acceptance so high and so unaffordable for small technology providers. Moreover, in this fast-moving field, by the time a new technology has finally been validated, it will already have been superseded. Most ironically, new technologies are assessed on how well they can predict the “gold standard” animal data; thus ensuring that they *cannot* succeed, if the drug affects animals differently from humans, which we now know is very often the case (Hartung, 2007, 2010; Leist et al., 2012). The very concept of the use of animal data as a useful standard is fundamentally flawed, as no species is truly representative of any other (Hartung, 2009; Wang & Gray, 2015; Perlman, 2016). Indeed the ability of rats to predict for carcinogenicity in mice has been shown to be useful in less than 60% of cases (Gray et al., 1995).

## 8 A Way Forward: Pragmatic Evaluation

The need for better ways to protect the public from the ever-increasing epidemic of ADRs is so urgent that a new approach to implementing more predictive methods is critical. This is now widely recognized and much attention is being devoted to making validation more flexible. The FDA is considering accepting methods that have been through a process of “qualification”, rather than traditional validation (Food and Drug Administration, FDA, 2017). Others have suggested streamlining validation, through greater use of reference chemicals and performance standards and the development of an objective, transparent, online peer review process (Judson et al., 2013).

We believe most strongly that any superior system must be based on *human* biology, and if that aim is compromised, predictive value is bound to fall. Advocates of animal testing say that this is unrealistic, and that it is not possible to gain sufficient understanding of the intact human system from isolated cells and tissues. However, if we look at other fields of technology, such as computing, automotive manufacture, telephonic communication, or space exploration, we see that yesterday's impossibility becomes today's challenge and tomorrow's commonplace. There is no reason why this should not equally apply to safety testing. In all other areas, technological advances are made in a step-wise fashion, seldom, if ever, in a single leap. We argue that the only practical way forward is a process of pragmatic evaluation of new technologies, whereby those that demonstrate success in predicting safety issues for humans, where the current system failed (as well as where it succeeded), should be accepted for use in appropriate circumstances and with sufficient justification. This approach will be iterative, and as shortcomings of the new tests are identified, further tests may be developed to overcome these problems. The truth is that we may never identify tests that will allow prediction of all safety issues, but by tackling these in a manageable fashion, we will get much closer than we can currently manage using animal-based approaches.

Of course, we cannot test potential new medicines on humans prospectively, using new methods in place of old ones, in case they perform less well. Therefore, new methods must be evaluated using historical "legacy" data. By studying the safety profiles of drugs that have been extensively used in human subjects, which will have necessarily passed the mandatory animal-based safety tests, we can identify where those tests failed to detect safety issues in human subjects. A selection of drugs whose toxicities were missed by animal tests can then form the basis of a test panel, to be submitted to a range of non-animal tests. In this way, the predictive performance of the new tests can be compared to that of the animal-based methods. To increase the scientific rigor of such studies, pairs of closely-related compounds should be used, where one has a particular toxicity that the other does not share. This will identify tests that are capable of differentiating between toxic and non-toxic compounds, the key attribute of any desirable test. Rather than assessing each new test in isolation, different types of tests will be combined in testing batteries, designed to complement each other in their ability to detect a variety of toxicities. Different batteries will be appropriate for different types of compounds. We need to forget the beguilingly simplistic approach of attempting to model humans with one system, even when that system is an integrated whole animal. No single test, however integrated, will ever be an adequate model for the breadth of human genetic variability. Combinations of tests at the molecular, cellular, organ, and system levels will need to be performed to generate sufficient

confidence to advance into cautious first-in-human testing, using safe approaches, such as microdosing, before proceeding to test therapeutic doses in patients in carefully designed adaptive clinical trials.

The Evidence-Based Toxicology Collaboration (EBTC) at the Johns Hopkins Bloomberg School of Public Health, is currently undertaking an evidence-based evaluation study, as described above. Using systematic reviews, they are comparing drug-induced toxicity in humans to preclinical animal data and to *in vitro* data from the Toxicity Forecaster (ToxCast) program of the US Environmental Protection Agency. The results will provide an objective comparison of the relative predictive abilities of animal versus non-animal methods. If successful, this study will demonstrate that a limited compound set can be used, if proper negative and positive controls are present, to compare the performance of a battery of tests relative to the current system. A clear demonstration of multiple successes, especially where the current regime has failed, would create a powerful impetus for governments and pharmaceutical companies to allocate more resources to tackling this problem more urgently. Substantial funding is required, as is greatly increased access to data.

Pharmaceutical companies are sitting on a treasure trove of preclinical and clinical data, which could yield immensely valuable information if made available for analysis. Former FDA Commissioner, Robert Califf called for a preclinical database to be established (Scott, 2016). This initiative must be seized; it has the potential to save time, money, and animals by avoiding futile repetitive testing; and, more importantly, the potential to revolutionize the evaluation of both old and new technologies, through statistical comparisons with a gold mine of millions of data points.

## 9 Regulatory Change

Former NIH Director, Elias Zerhouni, and former FDA Commissioner, Margaret Hamburg, state that the “regulation of drugs can either grease the wheels of progress or throw a wrench in the works” (Zerhouni and Hamburg, 2016). Calling for global harmonization of regulatory requirements, they note that differences between regulations in different countries create unnecessary barriers to the efficient delivery of safe, innovative, and effective treatments to patients. They acknowledge that regulatory authorities are struggling to keep up with rapid advances in science and technology and advocate high-level cooperation to ensure progress is not delayed by bureaucratic stagnation that promotes the status quo. Change needs to be driven by a *top-down* strategy to drive harmonization forward, urgently (Zerhouni and Hamburg, 2016).

Decades-old regulations have not been updated to reflect rapid advances in science and technology. It is acknowledged that regulations requiring the use of animal tests are a major barrier to adoption and use of more predictive human-relevant test methods (Malloy, 2016). Without regulatory updates reflecting the acceptability of the most predictive test methods available, the scientific advancements of the past decade will not be utilized.

International guidelines for preclinical testing remain focused on the use of traditional animal tests and merely mention the availability of more predictive human-relevant test methods. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), *Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals*, states: "The development of a pharmaceutical is a stepwise process involving an evaluation of both animal and human efficacy and safety information" (ICH, 2009). While the ICH guideline states that consideration should be given to the 3Rs, specifically reduction of the use of animals, and suggests consideration of the use of *in vitro* methods that could possibly replace animals, it does not discuss specifics of acceptable non-animal methods. This focus on reduction of animal use addresses only the ethics of animal testing, not the safety of human patients. From a public health perspective, the focus must be shifted to the replacement of animal tests with human-relevant test methods to provide safer, more effective medicines.

One has only to look at the FDA regulations on investigational new drugs and devices to understand the regulatory barrier to acceptance and adoption of modern test methods. FDA claims that the regulations give them the flexibility to accept modern, non-animal test methods (NATMs), such as *in vitro* studies, or prior experience with the drug or biological product in humans (Dorsey, 2010); however, current FDA regulations explicitly require animal testing. This requirement discourages the use of NATMs, which may be more predictive of human response. Twenty-nine FDA regulations clearly require animal testing and promote the status quo, creating an unreceptive environment that fails to encourage innovation and development of more predictive test methods (Center for Responsible Science, 2015). Modification of regulatory language would promote use of existing modern test methods and encourage further development to advance modernization of preclinical testing. Regulations must be changed to state clearly that the test most predictive of human response should (or even must) be used. In 2015, a coalition of non-profits, technology developers, and patient advocacy groups petitioned the FDA to make modest, non-controversial regulation amendments that would be an important first step in advancing the use of NATMs (Center for Responsible Science, 2015). These minor amendments to outdated existing regulations would have great impact

on the use and development of better tools for drug and device development. For example, when a current regulation states, “...*extensive animal and clinical tests are required as a condition of approval*,” the petition proposes a change of wording to state, “...*extensive preclinical and clinical tests are required as a condition of approval*.” Adoption of these test-neutral, conservative regulatory amendments would be an important first-step in moving forward. The FDA has yet to provide a substantive response to the petition.

While the US is a world leader in biomedical research and technology development, it lagged behind the EU in developing a strategy and *roadmap* for the advancement and use of new technology, until very recently. In December 2017, the FDA’s *Predictive Toxicology Roadmap* was issued to advance predictive toxicology in regulatory risk assessments (FDA, 2017). In January 2018, after considering input from 16 federal agencies, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) issued its *Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States* (ICCVAM, 2018). Both *roadmaps* outline a way forward to successful implementation of new technology. Crucially, they have been issued by government agencies, which should ensure that real progress is achieved; indeed, many activities towards implementation are already underway. The European roadmap calls for many things, including a joint taskforce to gather all current data on a wide variety of compounds into a toxicity database; abolition of useless tests; and, crucially, reasonable investment (Basketter et al., 2012). However, without an effective *top-down* (i.e., government-led) implementation strategy, advances in science and technology will languish and the EU will lag behind.

Evidence shows that animal methods are often still used, both in the US and the EU, even when superior validated methods are available. This is likely due to existing regulations that explicitly require animal tests. Applicants worry that forgoing the inclusion of animal data in product submissions risks rejection by regulatory reviewers, which would be costly in time and expense for drug sponsors. For example, since 2005, the FDA has informally stated that Draize data are not required for primary skin and eye irritation testing; yet, drug sponsors continue to submit Draize data. A review of the 137 New Molecular Entities approved by the FDA between 2011–2014 showed that the Draize test was used in 94% of all skin irritation and 60% of all eye irritation tests, despite the availability of validated methods that are more predictive of human response (Archibald, Drake and Coleman, 2015).

Regulatory submission reviewers require continuing education to be up to date on available new technologies. Without reviewer education and uniform acceptance criteria, variability between reviewers’ acceptance of new



technologies will discourage their use and cause confusion for sponsors on their acceptability. It is essential that regulators become knowledgeable about available NATMS to facilitate early communication with sponsors on their acceptability.

## 10 Conclusions

There is a clear ethical imperative to replace unreliable animal-based safety tests, not just for the animals but to protect human safety. Remarkable knowledge and tools are emerging from projects, such as ToxCast; Tox21; Innovative Medicines Initiative; Safety Evaluation Ultimately Replacing Animal Testing (SEURAT); Integrated European “Flagship” Program Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21<sup>st</sup> Century (EU-ToxRisk); and the Precision Medicines Initiative. These initiatives have the potential to revolutionize our ability to advance and protect human health, but only if they are implemented. A 2018 report by the UK BioIndustry Association and the Medicines Discovery Catapult emphasizes that *humanizing* the process of drug discovery and testing is the most important way to ease the *productivity crisis* in pharmaceutical research.

We must acknowledge that predicting the safety of medicines is an enormous challenge, and that a major obstacle to paradigm change is lack of confidence in the new methods. To tackle this, we suggest that a new, pragmatic approach to demonstrating that novel methods are more fit for purpose than existing methods could help to accelerate the replacement of most, if not all, animal toxicity tests with superior tests based on human biology. We believe that only through utilizing human-based systems to evaluate new medicines can we truly gain confidence in their clinical safety. In a 2014 debate on the proposal that “Animal experimentation in toxicology can be phased out in five-years’ time,” there was unanimous agreement that disruptive technologies must be properly funded and that more systematic, comparative data is needed (van der Meer, 2014).

In 2007, the US National Research Council called for a “paradigm shift from the use of experimental animals [...] toward the use of more efficient *in vitro* tests and computational techniques” in their landmark report, *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The authors expected the paradigm shift to encounter resistance, as toxicological testing practices are “deeply ingrained.” They envisioned that “toxicity testing will be radically overhauled over the next 10 years, with the animal testing component virtually, if not actually, eliminated within the next 20 years” (National Research Council, 2007).

The science of toxicity testing has indeed been transformed over the past 10 years; but in the absence of any regulatory pressure, practical change has been occurring at a glacial pace, while revolution rather than evolution is required (Hartung, 2017). Deadlines create tremendous impetus for change, as can be seen with the EU Cosmetic and Registration, Evaluation, Authorisation and Restriction of CHemicals (REACH) regulations. If we are serious about reducing the ever-increasing burden of death and disability caused by ADRs, we must agree on a deadline for the adoption of more human-relevant methods, and the phasing out of methods whose predictive ability has not been proven. As with the replacement of horses by cars, there will need to be a brief period of *sharing the road*, while confidence in the new methods grows. The Netherlands now leads the world with its announcement that it intends to phase out all legally prescribed, animal-based safety testing by 2025 (Netherlands National Committee for the protection of animals used for scientific purposes, NCad, 2016). The Committee recognizes that the transition will not happen of its own accord and will require clear strategic direction to change attitudes and practices.

Crucially, the regulations that govern how drugs are tested must be updated to encourage the adoption of the best new approaches. Current regulations are stifling innovation by failing to keep pace with scientific progress. We argue that several aspects of current practice can no longer be justified:

1. The continued use of testing methods that have never been validated, while novel methods must demonstrate a level of performance that current methods not only have never been asked to perform but are clearly unable to perform.
2. Resistance to the adoption of non-animal methods that, although not formally validated, show greater predictive performance than animal tests.
3. The continued *blind eye* turned to the use of animal-based tests, where viable non-animal methods exist, on the pretext that they may be required by regulators at home or abroad.
4. The exposure of human patients and volunteers to potentially unsafe substances on the basis of demonstrably unreliable animal data.
5. The risk of the loss of potentially life-saving/modifying treatments on the basis of demonstrably unreliable animal data.

In March 2016, Safer Medicines Trust commissioned a survey of 2,500 UK healthcare professionals. 79% agreed that pharmaceutical companies should be legally obliged to test new medicines using methods demonstrated to be the most predictive of safety for humans (Dods Information, 2016). Governments must act to protect the public by updating regulations, whose *raison d'être* is patient safety, that now prevent their own aim from being realized.

## References

- Abadie, R. (2010). *The professional guinea pig: Big pharma and the risky world of human subjects*. Durham, NC: Duke University Press.
- AllTrials (2016). *A Roadmap to Clinical Trial Transparency*. [online] Available at: <http://www.alltrials.net/news/roadmap/> [Accessed 13 October 2016].
- Archibald, K., T. Drake and R. Coleman (2015). Barriers to the Uptake of Human-based Test Methods, and How to Overcome Them. *Alternatives to Laboratory Animals*, 43, pp. 301–308.
- Arrowsmith, J. (2012). A Decade of Change. *Nature Reviews Drug Discovery*, 11, pp. 17–18, Figure 2: Probability of success to market from key milestones.
- Bagot, M. (2015). Trial and Error: Thousands Left Seriously ill or Disabled by Clinical Tests of New Drugs. *The Mirror* [online] Available at: <http://www.mirror.co.uk/news/uk-news/trial-error-thousands-left-seriously-5867052> [Accessed 13 October 2016].
- Bailey, J., M. Thew and M. Balls (2013). An Analysis of the Use of Dogs in Predicting Human Toxicology and Drug Safety. *Alternatives to Laboratory Animals*, 41(5), pp. 335–350.
- Bailey, J., M. Thew and M. Balls (2014). An Analysis of the Use of Animal Models in Predicting Human Toxicology and Drug Safety. *Alternatives to Laboratory Animals*, 42(3), pp. 181–199.
- Bailey, J., M. Thew and M. Balls (2015). Predicting Human Drug Toxicity and Safety Via Animal Tests: Can Any One Species Predict Drug Toxicity in Any Other, and Do Monkeys Help?. *Alternatives to Laboratory Animals*, 43 (6), pp. 393–403.
- Baker, M. (2011). Tissue Models: A Living System on a Chip. *Nature*, 471, pp. 661–665.
- Balijepalli, A. and V. Sivaramakrishnan (2017). Organs on Chips: Research and Commercial Perspectives. *Drug Discovery Today*, 22(2), pp. 397–403.
- Basketter, D.A., H. Clewell, I. Kimber, A. Rossi, B.J. Blaauboer, R. Burrier, M. Daneshian, C. Eskes, A. Goldberg and N. Hasiwa (2012). t4 Report: A roadmap for the Development of Alternative (Non-animal) Methods for Systemic Toxicity Testing. *Alternatives to Animal Experimentation*, 29(1), pp. 3–91.
- BioIndustry Association and Medicines Discovery Catapult (2018). *State of the Discovery Nation 2018 and the Role of the Medicines Discovery Catapult* [online] Available at: <https://md.catapult.org.uk/report-state-of-the-discovery-nation-2018/> [Accessed 15 February 2018].
- Biopta (2017). *How We Help*. [online] Available at: <http://www.biopta.com/How%20we%20help.cfm> [Accessed 17 May 2017].
- Bisserbe, N. (2016). Drug's Toxicity Caused Clinical-trial Death, Panel Says. *The Wall Street Journal*. [online] Available at: <http://www.wsj.com/articles/drugs-toxicity-caused-clinical-trial-death-panel-says-1461092409> [Accessed 13 October 2016].

- Bracken, M.B. (2009). Why Animal Studies Are often Poor Predictors of Human Reactions to Exposure. *Journal of the Royal Society of Medicine*, 102(3), pp. 120–122.
- Bruckner, T. and B. Ellis (2017). *Clinical trial transparency: A key to better and safer medicines*. Bristol, UK [online] Available at: <https://www.scribd.com/document/347308262/Clinical-Trial-Transparency-A-Key-to-Better-and-Safer-Medicines-Till-Bruckner-and-Beth-Ellis-2017> [Accessed 22 May 2017].
- Capdeville, R., E. Buchdunger, J. Zimmermann and A. Matter (2002). Glivec (STI571, imatinib), A Rationally Developed, Targeted Anticancer Drug. *Nature Reviews Drug Discovery*, 1, pp. 493–502.
- Carthew, P., E.A. Martin, I.N. White, F. De Matteis, R.E. Edwards, B.M. Dorman, R.T. Heydon and L.L. Smith (1995). Tamoxifen Induces Short-term Cumulative DNA Damage and Liver Tumors in Rats: Promotion by Phenobarbital. *Cancer Research*, 55(3), pp. 544–547.
- Center for Responsible Science (2015). *Citizen Petition to Food and Drug Administration, requesting that the FDA modify existing regulations in Title 21 of the Code of Federal Regulations (CFR) that govern requirements for investigational new drug (IND) applications, investigational device exemptions (IDE), and new drug applications (NDAs)*. [online] Docket ID: FDA-2015-P-2820. Available at: <https://www.regulations.gov/docket?D=FDA-2015-P-2820> [Accessed 13 October 2016].
- Chavalarias, D. and J. Ioannidis (2010). Science Mapping Analysis Characterizes 235 Biases in Biomedical Research. *Journal of Cardiovascular Electrophysiology*, 63(11), pp. 1205–1215.
- Clemence, M. and J. Leaman. (2016). *Public Attitudes to Animal Research in 2016*. A report by Ipsos MORI for the Department for Business, Energy & Industrial Strategy, Ipsos MORI Social Research Institute. [online] Available at: <https://www.ipsos.com/sites/default/files/publication/1970-01/sri-public-attitudes-to-animal-research-2016.pdf> [Accessed 18 May 2017].
- Collaborative Approach to Meta-Analysis and Review of Animal Data from Experiment (CAMARADES) (2017). *About CAMARADES*. [online] Available at: <http://www.dcn.ed.ac.uk/camarades/default.htm> [Accessed 24 May 2017].
- Cortinovis, C. and F. Caloni (2016). Household Food Items Toxic to Dogs and Cats. *Frontiers in Veterinary Science*, 22(3), p. 26.
- De Lorgeril, M. and M. Rabaeus (2015). Beyond Confusion and Controversy. Can We Evaluate the Real Efficacy and Safety of Cholesterol-lowering with Statins? *Journal of Controversies in Biomedical Research*, 1(1), pp. 67–92.
- Dods Information (2016). *Health Care Workforce Perceptions of Pharmaceutical Testing Regulations, A Study Conducted on Behalf of Safer Medicines Trust*. Dods Parliamentary Communications, Ltd. [online] Available at: <http://www.safermedicines.org/pdfs/dods-health-care-springs-survey.pdf> [Accessed 13 October 2016].

- Dorsey, D. (2010). *Food and Drug Administration Office of the Commissioner Response Letter to Meyer Glitzenstein and Crystal Physicians Committee for Responsible Medicine PCRM*. Petition Denial, 3–4. [online] Available at: <https://www.regulations.gov/document?D=FDA-2007-P-0109-0012> [Accessed 10 October 2018].
- Downing, N.S., N.D. Shah, J.A. Aminawung, A.M. Pease, J.D. Zeitoun, H.M. Krumholz and J.S. Ross (2017). Post market Safety Events Among Novel Therapeutics Approved by the US Food and Drug Administration Between 2001 and 2010. *Journal of the American Medical Association*, 317(18), pp. 1854–1863.
- Endo, A., M. Kuroda and Y. Tsujita, (1976). ML-236A, ML-236B, and ML-236C, New inhibitors of cholesterol synthesis produced by *Penicillium citrinum*. *The Journal of Antibiotics*, 29(12), pp. 1346–1348. [online] Available at: [https://www.jstage.jst.go.jp/article/antibiotics1968/29/12/29\\_12\\_1346/\\_article](https://www.jstage.jst.go.jp/article/antibiotics1968/29/12/29_12_1346/_article) [Accessed 10 October 2018].
- Endo, A. (2010). A historical perspective on the discovery of statins. *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences*, 86(5), pp. 484–493. [online] Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108295/> [Accessed 10 October 2018].
- European Commission (2008). *Strengthening Pharmacovigilance to Reduce Adverse Effects of Medicines*. [online] Available at: [http://europa.eu/rapid/press-release\\_MEMO-08-782\\_en.htm?locale=en](http://europa.eu/rapid/press-release_MEMO-08-782_en.htm?locale=en) [Accessed 13 October 2016].
- European Parliament (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Communities*, L276, p. 33–79. [online] Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32010L0063> [Accessed 13 October 2016].
- Evans, I., H. Thornton and I. Chalmers (2006). *Testing treatments*. London, UK: The British Library.
- Fourches, D., J.C. Barnes, N.C. Day, P. Bradley, J.Z. Reed and A. Tropsha (2010). Cheminformatics Analysis of Assertions Mined from Literature That Describe Drug-induced Liver Injury in Different Species. *Chemical Research in Toxicology*, 23(1), pp. 171–183.
- Frank, J. (2004). Technological Lock-in, Positive Institutional Feedback, and Research on Laboratory Animals. *Structural Change Economic Dynamics*, 16, pp. 557–575.
- Garner, J.P., B.N. Gaskill, E.M. Weber, J. Ahloy-Dallaire and K.R. Pritchett-Corning, (2017). Introducing Therioepistemology: The Study of How Knowledge Is Gained from Animal Research. *Nature Lab Animal*, 20, pp. 103–113. [online] Available at: <https://www.nature.com/articles/labani.1224> [Accessed 13 August 2018].
- Geerts, H. (2009). Of Mice and Men: Bridging the Translational Disconnect in CNS Drug Discovery. *CNS Drugs*, 23, pp. 915–926.
- Godlee, F. (2014). How Predictive and Productive Is Animal Research?. *British Medical Journal*, 348: g3719.
- Goldacre, B. (2012). *Bad pharma*. London, UK: Fourth Estate, Harper Collins.

- Gray, G.M., P. Li, I. Shlyakhter and R. Wilson, (1995). An Empirical Examination of Factors Influencing Prediction of Carcinogenic Hazard Across Species. *Regulatory Toxicology and Pharmacology*, 22(3), pp. 283–291.
- Green, S. (2015). Can Animal Data Translate to Innovations Necessary for a New Era of Patient-centred and Individualised Healthcare? Bias in Preclinical Animal Research. *BMC Medical Ethics*, 16(1), p. 53.
- Hackam, D. and D. Redelmeier (2006). Translation of Research Evidence from Animals to Humans, *Journal of the American Medical Association*, 296, pp. 1731–1732.
- Hampton, J. (2015). Therapeutic Fashion and Publication Bias: The Case of Anti-arrhythmic Drugs in Heart Attack. *Journal of the Royal Society of Medicine*, 108(10), pp. 418–420.
- Hare, R. (1982). New Light on the History of Penicillin. *Medical History*, 26, pp. 1–24.
- Harris, R. (2017). *Rigor mortis. How sloppy science creates worthless cures, crushes hope, and wastes billions*. New York: Basic.
- Hartung, T. (2007). Food for thought...on validation. *Alternatives to Animal Experimentation*, 24, pp. 67–80.
- Hartung, T. (2009a). Per Aspirin ad Astra... *Alternatives to Laboratory Animals*, 37(2), pp. 45–47.
- Hartung, T. (2009b). Toxicology for the Twenty-first Century. *Nature*, 460, pp. 208–212.
- Hartung, T. (2010). Evidence-based Toxicology – The Toolbox of Validation for the 21st Century?. *Alternatives to Laboratory Animals*, 27, pp. 253–263.
- Hartung, T. (2013). Look Back in Anger – What Clinical Studies Tell Us About Preclinical Work. *Alternatives to Laboratory Animals*, 30(3), pp. 275–291.
- Hartung, T. (2017). Evolution of Toxicological Science: The Need for Change. *Journal of Risk Assessment and Management*, 20(1–3), pp. 21–45.
- Herper, M. (2012). The Truly Staggering Cost of Inventing New Drugs. *Forbes Magazine*. [online] Available at: <https://www.forbes.com/forbes/2012/0312/strategies-pharmaceuticals-lilly-stagger-cost-inventing-new-drugs.html> [Accessed 23 May 2017].
- Heywood, R. (1990). Clinical toxicity –Could it have been predicted? Post-marketing experience. In: C. Lumley and S. Walker, eds., *Animal Toxicity Studies: Their Relevance for Man*. Lancaster: Quay, pp. 57–67.
- Huang, R., M. Xia, S. Sakamuru, J. Zhao, S.A. Shahane, M. Attene-Ramos, T. Zhao, C.P. Austin and A. Simeonov (2016). Modelling the Tox21 10 K Chemical Profiles for *In Vivo* Toxicity Prediction and Mechanism Characterization, *Nature Communications*, 7, pp. 10425.
- Hutchinson, L. and R. Kirk (2011). High Drug Attrition Rates—Where Are We Going Wrong?. *Nature Reviews Clinical Oncology*, 8, pp. 189–190.
- Igarashi, T. (1994). The duration of toxicity studies required to support repeated dosing in clinical investigation—A toxicologist’s opinion. In: C. Parkinson, C. Lumley and

- S. Walker, eds., *CMR Workshop: The Timing of Toxicological Studies to Support Clinical Trials*. Boston: Kluwer, pp. 67–74.
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) (2009). *Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals*. [online] Available at: [https://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Multidisciplinary/M3\\_R2/Step4/M3\\_R2\\_Guideline.pdf](https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2_Guideline.pdf) [Accessed 10 October 2018].
- Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2018). *Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States*. [online] Available at: [https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam\\_strategicroadmap\\_january2018\\_document\\_508.pdf](https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam_strategicroadmap_january2018_document_508.pdf) [Accessed 11 February 2018].
- Ioannidis, J.P.A. (2012). Extrapolating from Animals to Humans. *Science Translational Medicine*, 4, pp. 1–3.
- Johnson, P.A., T. Fitzgerald, A. Salganicoff, S.F. Wood, J.M. Goldstein and Y.L. Colson (2014). *Sex-specific Medical Research – Why Women’s Health Can’t Wait*. [online] A Report of the Mary Horrigan Connors Center for Women’s Health & Gender Biology at Brigham and Women’s Hospital. [online] Available at: <https://www.brighamand-womens.org/assets/BWH/womens-health/pdfs/ConnorsReportFINAL.pdf> [Accessed 10 October 2018].
- Jones, J. and L. Saad (2017). *Gallup Poll Social Series: Values and Beliefs*. Gallup News Service. [online] Available at: [http://www.gallup.com/poll/210542/americans-hold-record-liberal-views-moral-issues.aspx?g\\_source=2017+poll+animals&g\\_medium=search&g\\_campaign=tiles](http://www.gallup.com/poll/210542/americans-hold-record-liberal-views-moral-issues.aspx?g_source=2017+poll+animals&g_medium=search&g_campaign=tiles) [Accessed 29 August 2017].
- Judson, R. R. Judson, R. Kavlock, M. Martin, D. Reif, K. Houck, T. Knudsen, A. Richard, R.R. Tice, M. Whelan, M. Xia and R. Huang (2013). Perspectives on Validation of High-throughput Assays Supporting 21st Century Toxicity Testing. *Alternatives to Laboratory Animals*, 30(1), pp. 51–56.
- Kenna, J.G. (2017). Human Biology-based Drug Safety Evaluation: Scientific Rationale, Current Status and Future Challenges. *Expert Opinion on Drug Metabolism & Toxicology*, 13(5), pp. 567–574.
- Knight, A., J. Bailey and J. Balcombe (2006). Animal Carcinogenicity Studies: 1. Poor Human Predictivity. *Alternatives to Laboratory Animals*, 34, pp. 19–27.
- Krul, C. (2014). *SLiM Project Final Symposium Report*. [online] Available at: <http://www.innovativetesting.nl/slim-project-final-symposium-report> [Accessed 24 May 2017].
- Langley, G. (2014). Considering a New Paradigm for Alzheimer’s Disease Research. *Drug Discovery Today*, 19(8), pp. 1114–1124.
- Lasser, K.E., P.D. Allen, S.J. Woolhandler, D.U. Himmelstein, S.M. Wolfe and D.H. Bor (2002). Timing of New Black Box Warnings and Withdrawals for Prescription Medications. *Journal of the American Medical Association*, 287(17), pp. 2215–2220.

- Leist, M., N. Hasiwa, M. Daneshian and T. Hartung (2012). Validation and Quality Control Of Replacement Alternatives – Current Status and Future Challenges. *Toxicology Research*, 1, pp. 8–22.
- Leppard, D. (30 July 2006). Elephant Man Drug Victims Told to Expect Early Death. *The Sunday Times*. [online] Available at: <https://www.thetimes.co.uk/article/elephant-man-drug-victims-told-to-expect-early-death-cm5287v9bxn> [Accessed 23 May 2017].
- Light, D. (2015). Serious Risks and Few New Benefits from FDA-approved Drugs, *Health Affairs*. [online] Available at: <http://healthaffairs.org/blog/2015/07/06/serious-risks-and-few-new-benefits-from-fda-approved-drugs/> [Accessed 13 October 2016].
- Lowe, D. (2017). *Guess How This Latest Alzheimer's Trial Came Out*. In the Pipeline. [online] Available at: <http://blogs.sciencemag.org/pipeline/archives/2017/02/15/guess-how-this-latest-alzheimers-trial-came-out> [Accessed 24 May 2017].
- Malloy, T. (2016). Implementing the Vision for Toxicity Testing in the 21st Century: If You Build It, Will They Come?. *Scientific Advisory Committee on Alternative Toxicological Methods*. Available at: [http://ntp.niehs.nih.gov/ntp/about\\_ntp/sacatm/2016/september/presentations/malloy\\_visionfortoxtest\\_508.pdf](http://ntp.niehs.nih.gov/ntp/about_ntp/sacatm/2016/september/presentations/malloy_visionfortoxtest_508.pdf) [Accessed 13 October 2016].
- Matthews, R. (2008). Medical Progress Depends on Animal Models—Doesn't It? *Journal of the Royal Society of Medicine*, 101, pp. 95–98.
- Nair, A. (2015). Clinical Research: Regulatory Uncertainty Hits Drug Trials in India. *The Pharmaceutical Journal*. [online] Available at: <http://www.pharmaceutical-journal.com/news-and-analysis/features/clinical-research-regulatory-uncertainty-hits-drug-trials-in-india/20068063.article> [Accessed 13 October 2016].
- National Center for Advancing Translational Sciences (2016). *Toxicology in the 21st Century (Tox21)* [online] Available at: <https://ncats.nih.gov/tox21> [Accessed 24 May 2017].
- National Institutes of Health (NIH) (2008). *NIH Collaborates with EPA to Improve the Safety Testing of Chemicals – New Strategy Aims to Reduce Reliance on Animal Testing*. [online] Available at: <https://www.nih.gov/news-events/news-releases/nih-collaborates-epa-improve-safety-testing-chemicals> [Accessed 23 May 2017].
- National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The National Academies Press [online] Available at: <http://www.thecre.com/forum8/wp-content/uploads/2016/05/EPI-Toxicity-Testing-in-21st-Century.pdf> [Accessed 23 May 2017].
- National Toxicology Project Scientific Advisory Committee on Alternative Toxicological Methods (2016). *A Strategy for Implementing the Vision for Toxicity Testing in the 21st Century*. [online] Available from: [http://ntp.niehs.nih.gov/ntp/about\\_ntp/sacatm/2016/september/vision20160927\\_508.pdf](http://ntp.niehs.nih.gov/ntp/about_ntp/sacatm/2016/september/vision20160927_508.pdf) [Accessed 13 October 2016].
- Netherlands National Committee for the protection of animals used for scientific purposes (NCad) (2016). *Transition to Non-animal Research – About the Possibilities*



- for Phasing Out Animal Procedures and Stimulating Innovation without Laboratory Animals*. [online] Available at: <https://www.ncadierproevenbeleid.nl/documenten/rapport/2016/12/15/ncad-opinion-transition-to-non-animal-research> [Accessed 17 May 2017].
- Parke, D. (1994). Clinical Pharmacokinetics in Drug Safety Evaluation. *Alternatives to Laboratory Animals*, 22, pp. 207–209.
- Perel, P., I. Roberts, E. Sena, P. Wheble, C. Briscoe, P. Sandercock, M. Macleod, L.E. Mignini, P. Jayaram and K.S. Khan (2007). Comparison of Treatment Effects Between Animal Experiments and Clinical Trials: Systematic Review. *British Medical Journal*, 334(7586), p. 197.
- Perlman, R. (2016). Mouse Models of Human Disease: An Evolutionary Perspective. *Evolution, Medicine, and Public Health*, 1, pp. 170–176.
- Pew Research Center (2015). *Public and Scientists' Views on Science and Society: Use of Animals in Scientific Research*. [online] Available at: [http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/pi\\_2015-01-29\\_science-and-society-03-05/](http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/pi_2015-01-29_science-and-society-03-05/) [Accessed 13 October 2016].
- Pound, P., S. Ebrahim, P. Sandercock, M.B. Bracken and I. Roberts (2004). Where Is the Evidence That Animal Research Benefits Humans?. *British Medical Journal*, 328, pp. 514–517.
- Pound, P. and M. Bracken (2014). Is Animal Research Sufficiently Evidence Based to Be a Cornerstone of Biomedical Research?. *British Medical Journal*, 348:g3387.
- Rawson, N. (2013). New Drug Approval Times and Safety Warnings in the United States and Canada, 1992–2011. *Journal of Population Therapeutics and Clinical Pharmacology*. 20(2):e67–e81.
- Salsburg, D. (1983). The Lifetime Feeding Study in Mice and Rats – An Examination of Its Validity as a Bioassay for Human Carcinogens. *Fundamental and Applied Toxicology*. 3, pp. 63–67.
- Scott, D. (2016). Califf's Big Idea: Build a Database for Research Done Before Clinical Trials, *STAT News*. [online] Available at: <https://www.statnews.com/2016/06/10/califf-database-preclinical-trials/> [Accessed 13 October 2016].
- Seok, J., H.S. Warren, A.G. Cuenca, M.N. Mindrinos, H.V. Baker, W. Xu, D.R. Richards, G.P. McDonald-Smith, H. Gao, L. Hennessy and C.C. Finnerty (2013). Inflammation and Host Response to Injury, Large Scale Collaborative Research Program. Genomic Responses in Mouse Models Poorly Mimic Human Inflammatory Diseases. *Proceedings of the National Academy of Science USA*. 110, pp. 3507–3512.
- Sharav, V. (2016). *A Recent Deadly Drug Experiment – BIA 10-2474 – Rennes, France*. [Blog] Alliance for Human Research Protection. [online] Available at: <http://ahrp.org/a-deadly-drug-experiment-bia-10-2474-rennes-france/> [Accessed 13 October 2016].

- Sheets, R., T. Zhou and I. Knezevic (2016). Review of Efficacy Trials of HIV-1/AIDS Vaccines and Regulatory Lessons Learned: A Review from a Regulatory Perspective. *Biologicals*, 44(2), pp. 73–89.
- Spriet-Pourra, C. and M. Auriche (1994). *Drug withdrawal from sale*. New York: PJP Publications.
- St Clair, E.W. (2008). The Calm After the Cytokine Storm: Lessons from the TGN1412 Trial. *Journal of Clinical Investigation*, 118(4), pp. 1344–1347.
- Stebbing, R., L. Findlay, C. Edwards, D. Eastwood, C. Bird, D. North, Y. Mistry, P. Dilger, E. Liefoghe, I. Cludts and B. Fox (2007). “Cytokine Storm” in the Phase I Trial of Monoclonal Antibody TGN1412: Better Understanding the Causes to Improve Preclinical Testing of Immunotherapeutics. *The Journal of Immunology*, 179(5), pp. 3325–3331.
- Sutherland, B.A., J. Minnerup, J.S. Balami, F. Arba, A.M. Buchan and C. Kleinschnitz (2012). Neuroprotection for Ischaemic Stroke: Translation from the Bench to the Bedside. *International Journal of Stroke*, 7, pp. 407–418.
- Temporary Specialist Scientific Committee (TSSC) (2016). *Minutes of the TSSC meeting on “FAAH (Fatty Acid Amide Hydrolase) inhibitors”*, 15 February 2016. [online] Available at: [http://ansm.sante.fr/content/download/86439/1089765/version/1/file/CR\\_CSST-FAAH\\_15-02-2016\\_Version-Anglaise.pdf](http://ansm.sante.fr/content/download/86439/1089765/version/1/file/CR_CSST-FAAH_15-02-2016_Version-Anglaise.pdf) [Accessed 22 May 2017].
- Theil, C., H. Cordes, L. Fabbri, H.E. Aschmann, V. Baier, I. Smit, F. Atkinson, L.M. Blank and L. Kuepfer (2017). A Comparative Analysis of Drug-induced Hepatotoxicity in Clinically Relevant Situations. *PLoS Computational Biology*, 13(2), p. e1005280. [online] Available at: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005280> [Accessed 10 October 2018].
- Thompson, L. (1994). The Cure That Killed. *Discover Magazine*, 15(3): pp. 56–62.
- UK Department of Health (2012). *Response to Safer Medicines Campaign*. [online] Available at: <http://transparency.dh.gov.uk/2012/03/30/response-to-safer-medicines-campaign/> [Accessed 13 October 2016].
- US Department of Health and Human Services, National Toxicology Program. *Alternative Methods Accepted by US Agencies* [online] Available at: <https://ntp.niehs.nih.gov/pubhealth/evalatm/accept-methods/index.html> [Accessed 10 October 2018].
- US Food and Drug Administration (FDA) (2017). *FDA’s Predictive Toxicology Roadmap*. [online] Available at: <https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RegulatoryScience/UCM587831.pdf> [Accessed 11 February 2018].
- US Senate Committee on Appropriations (2016). *Hearing on FY2017 NIH Budget*. Testimony of Collins, Francis S., M.D., Ph.D., Director, National Institutes of Health. [online] Available at: <http://www.appropriations.senate.gov/hearings/hearing-on-fy2017-national-institutes-of-health-budget-request> [Accessed 13 October 2016].

- van der Meer, A. (2014). *Can We Get Rid of Animal Experimentation in Toxicology within Five Years?* [blog] Available at: <http://www.andriesvandermeer.com/2014/06/can-we-get-rid-of-animal.html> [Accessed 13 October 2016].
- van Esbroeck, A., A.P.A. Janssen, III Cognetta, D. Ogasawara., G. Shpak., M. van der Kroeg., V. Kantae., M.P. Baggelaar., F.M.S de Vrij., H. Deng., M. Allara., F. Fezza., Z. Lin., T. van der Wel., M. Soethoudt., E.D. Mock., H. den Dulk., I.L. Baak., B.I. Florea., G. Hendriks., L. De Petrocellis., H.S. Overkleeft., T. Hankemeier., C.I. De Zeeuw., V. Di Marzo., M. Maccarrone., B.F. Cravatt., S.A. Kushner. (2017). Activity-Based Protein Profiling Reveals Off-target Proteins of the FAAH Inhibitor BIA 10-2474. *Science*, 356, (6342), pp. 1084–1087.
- van Meer, P.J., M. Kooijman, C.C. Gispen-de Wied, E.H. Moors and H. Schellekens (2012). The Ability of Animal Studies to Detect Serious Post Marketing Adverse Events Is Limited. *Regulatory Toxicology and Pharmacology*, 64(3), pp. 345–349.
- Wall, R. and M. Shani (2008). Are Animal Models as GOOD as We Think?. *Theriogenology*, 69(1), pp. 2–9.
- Wang, B. and G. Gray (2015). Concordance of Noncarcinogenic Endpoints in Rodent Chemical Bioassays. *Risk Analysis*, 35, pp. 1154–1166.
- Xu, J., P. Henstock, M. Dunn, A. Smith, J. Chabot and D. de Graaf (2008). Cellular Imaging Predictions of Clinical Drug-induced Liver Injury. *Toxicological Sciences*, 105, (1), pp. 97–105. [online] Available at: <https://academic.oup.com/toxsci/article/105/1/97/1662976> [Accessed 10 February 2018].
- Zerhouni, E. and M. Hamburg (2016). The Need for Global Regulatory Harmonization: A Public Health Imperative. *Science Translational Medicine*, 8(338), p. 338ed6. [online] Available at: <http://stm.sciencemag.org/content/8/338/338ed6.short> [Accessed 10 February 2018].