TRANSLATIONAL PERSPECTIVE

Limitations of Animal Studies for Predicting Toxicity in Clinical Trials

Is it Time to Rethink Our Current Approach?

Gail A. Van Norman, MD

SUMMARY

Animal testing is used in pharmaceutical and industrial research to predict human toxicity, and yet analysis suggests that animal models are poor predictors of drug safety in humans. The cost of animal research is high—in dollars, delays in drug approval, and in the loss of potentially beneficial drugs for human use. Human subjects have been harmed in the clinical testing of drugs that were deemed safe by animal studies. Increasingly, investigators are questioning the scientific merit of animal research. This review discusses issues in using animals to predict human toxicity in pharmaceutical development. Part 1 focuses on scientific concerns over the validity of animal research. Part 2 will discuss alternatives to animal research and their validation and use in production of human pharmaceuticals. (J Am Coll Cardiol Basic Trans Science 2019;4:845-54) © 2019 The Author. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

There is no doubt that the use of animals in science and medicine has significantly benefitted human beings (**Table 1**). However, many investigators are increasingly concerned that animal experimentation may be based on a scientifically flawed premise and that it retains its acceptability only because clear alternatives have not been identified. Dramatically rising costs and extremely high failure rates in drug development have led many to re-evaluate the value of animal studies. This review focuses on questions regarding the scientific validity of nonhuman animal models (hereafter referred to simply as "animal research") in predicting human toxicity in preclinical pharmaceutical testing.

HISTORICAL PERSPECTIVE

In the United States, the use of animals to test human pharmaceuticals dates to 1937, when a liquid formulation of a sulfa antibiotic dissolved in ethylene glycol resulted in the deaths of 107 adults and children. The incident resulted in passage of the 1938 U.S. Federal Food, Drug, and Cosmetic Act, mandating animal toxicity testing [1,2]. In 1946, language was incorporated into the Nuremberg code [3] and later the Helsinki Declaration [4] requiring human experiments to be "*designed and based on the results of animal experimentation* [author's italics] and a knowledge of the natural history of the disease." The statement was written by Andrew Ivy, a strong proponent of animal research, but was not based on scientific evidence that such a requirement would improve safety or efficacy of human drug development [3].

Today, the U.S. Food and Drug Administration (FDA) generally requires preclinical testing of any new drug or biological therapeutic "for pharmacologic activity and acute toxicity in animals" prior to

entering human clinical trials (5). In certain cases, such as emergency treatment for hazardous exposure, the FDA may even approve in-human use based solely on animal testing under "The Animal Efficacy Rule" (6).

Despite the deeply rooted assumption that animal models accurately predict human toxicity (7-9), even cursory examination of the concordance of animal and human trials raises concerns. A 2006 review of 76 animal studies, for example, found that approximately 20% were contradicted in humans and only 37% were ever replicated in humans (10). A review of 221 animal experiments found agreement in human studies just 50% of the time—essentially randomly (11). Review of 37 chemicals studied in the U.S. National Toxicology Program concluded that toxicities other than carcinogenesis were not reproducible between rats and mice, between sexes, or compared with historic control animals. Average positive predictive value (PPV) from mouse to rat was 55.3% and 44.8% for long-term and short-term studies, respectively. Combining organ, length of exposure, and sex, PPV between mice and rats hovered around 50%, which is no greater than random chance (12). An analysis of 2,366 drugs concluded that "results from tests on animals (specifically rat, mouse and rabbit models) are highly inconsistent predictors of toxic responses in humans, and are little better than what would result merely by chance—or tossing a coin—in providing a basis to decide whether a compound should proceed to testing in humans" (13). Similar results were found for nonhuman primates and dogs (14). Indeed, we need go no farther than the failure rates in drug development to have serious questions about whether animal testing accurately predicts toxicity in human trials.

About 12% of pharmaceuticals pass preclinical testing to enter clinical trials (15). Of those, only 60% successfully complete phase I trials (16). Overall, approximately 89% of novel drugs fail human clinical trials, with approximately one-half of those failures due to unanticipated human toxicity (**Figure 1**) (17). If animal tests accurately predict human toxicity, then why are toxicity-related failure rates in human clinical trials so high?

THE PRICE OF WRONG DECISIONS

Two critical "wrong" decisions regarding animal tests of human pharmaceuticals are 1) falsely identifying a toxic drug as "safe" and 2) falsely labeling a potentially useful therapeutic agent as toxic.

When a human-toxic drug is identified as "safe" by animal testing, the most likely outcome by far is that the drug will fail in clinical testing, often due to unacceptable adverse human effects, and sometimes significantly harming volunteer research subjects in the process. Drugs that survive clinical trials and attain market approval may still be recalled later due to toxicity identified only after months or years of in-human use. Vioxx (Merck, Kenilworth, New Jersey) was found after release to significantly increase the risk of cardiovascular morbidity and mortality, costing Merck more than $8.5 billion in legal settlements alone (18). An estimated 88,000 people suffered heart attacks after taking Vioxx and 38,000 died (19).

Of 578 discontinued and withdrawn drugs in Europe and the United States, almost one-half were withdrawn or discontinued in post-approval actions due toxicity (20). Van Meer et al. (21) found that of 93 post-marketing serious adverse outcomes, only 19% were identified in preclinical animal studies. In the first decade of the 21st century, approximately one-third of FDA-approved drugs were subsequently cited for safety or toxicity issues. or a combination of both, including human cardiovascular toxicity and brain damage, after remaining on the market for a median of 4.2 years (22,23). The most common toxicity types associated with drug withdrawals in the United States and Europe are hepatic (21%), cardiovascular (16%), hematological (11%), neurological (9%), and carcinogenicity (8%) (**Figure 2**) (20).

Protein-based biologics (e.g., monoclonal antibodies), fusion proteins, and recombinant proteins now account for most development stage and marketed biopharmaceuticals (15). These present a particular challenge in predicting human toxicity, due to their propensity to provoke production of antidrug antibodies. Safety concerns include cross reactivity, potentially exaggerated pharmacology, and slow recovery from toxicity, among others (15,24)—and immunogenic responses in animals do not predict immunogenicity in humans (15,25-27).

There are many notable examples of cases in which animal trials did not predict severe human toxicity. Isuprel for treatment of asthma caused over 3,500 deaths in Great Britain alone, despite safety in rats, guinea pigs, dog, and monkeys, all of which had received doses far exceeding those administered in humans (2,28). Thalidomide caused devastating phocomelia in an estimated 20,000 to 30,000 infants before it was withdrawn. However, animal tests failed to reveal significant teratogenicity in 10 strains of rats; 11 breeds of rabbit; 2 breeds of dog; 3 strains of

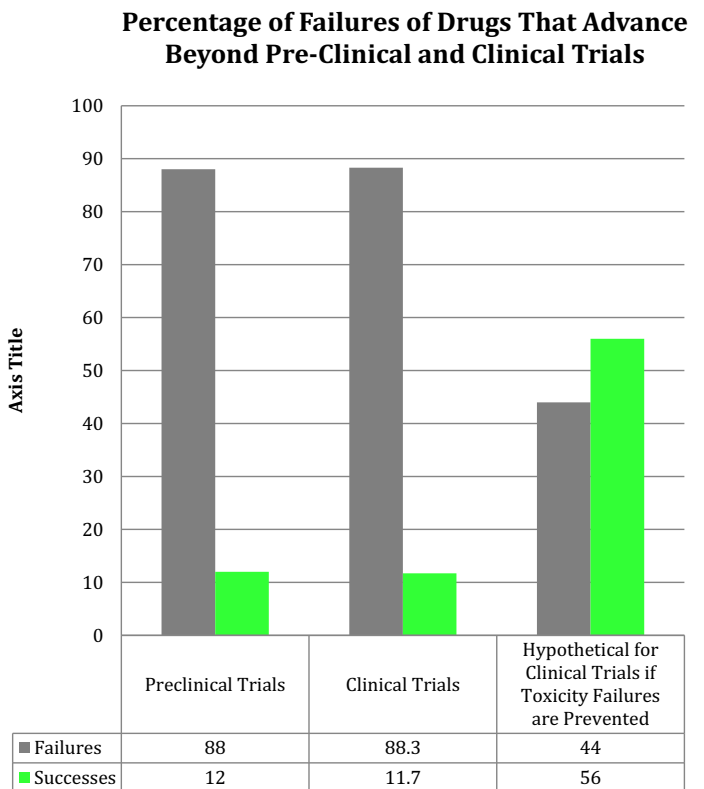| TABLE 1   Animal Uses in Science, Medicine, and Research |
|---|
| Predictive models for human diseases and their processes |
| Predictive models for testing drugs and other chemicals for human toxicity and efficacy |
| "Spare parts"— e.g., pig-derived aortic valve prostheses |
| Bioreactors or factories— e.g., production of monoclonal antibodies |
| Sources of tissue to study physiological principles |
| Educational "material" to educate and train biology and medical students and others |
| Subjects in research to benefit other animals |
| Subjects of research to gain basic knowledge for its own sake |

FIGURE 1   Failures in Translational Research: Preclinical and Clinical Trials

Percentage of Failures of Drugs That Advance Beyond Pre-Clinical and Clinical Trials

	Preclinical Trials	Clinical Trials	Hypothetical for Clinical Trials if Toxicity Failures are Prevented
Failures	88	88.3	44
Successes	12	11.7	56

Percentages of drugs that fail in preclinical trials (due to drug toxicity or failure of efficacy in animal testing) and in clinical trials (due drug toxicity or failure of efficacy in human testing) are shown in columns 1 and 2. The third column demonstrates what would happen if animal and human toxicity were closely correlated and therefore drugs with human toxicity were eliminated at the preclinical testing stage by animal toxicity testing (one-half of all drug failures in clinical trials are due to toxicity issues despite safety in animals). Success rates of clinical trials increase from 11.7% overall to approximately 56%.

hamsters; 8 species of primates; and various cats, armadillos, guinea pigs, swine, and ferrets (29). An antibody to treat human autoimmune disease, TGN1412, was given at 1/500th the dose found safe in animal testing to 6 human volunteers in a phase I trial (30,31), rendering them all critically ill within minutes and leaving them all with long-term complications (32-34). BIA-102474-101, a drug developed for a range of disorders from anxiety to Parkinsonism, caused deep brain hemorrhage and necrosis in all 5 human volunteers during a phase I clinical trial after it was administered in doses that were 1/500th of the safe dose for dogs. One volunteer died (35). Fialuridine, for treatment of hepatitis B, caused the deaths of 5 volunteers during phase II clinical trials despite being safe in mice, rats, dogs, monkeys, and woodchucks in doses that were hundreds of times higher. Two other volunteers only survived after receiving liver transplants (32).

When animal tests falsely identify a safe chemical as "toxic," the almost certain outcome is abandonment of further development. Undoubtedly many potentially beneficial drugs have failed animal testing and been lost to patients, even though they would have been both safe and effective (36,37). Because a drug that shows toxicity in animal models is unlikely to ever undergo human testing, the magnitude of this type of "error" is unknown. However, many highly beneficial drugs would have failed animal testing and would never have been brought to market, except that they were developed before animal testing was required (38). Examples include penicillin (fatal to guinea pigs) (39), paracetamol (toxic in dogs and cats) (40), and aspirin (embryo toxicity in rats and rhesus monkeys) (41).

Lack of animal tests has also caused deleterious delays in critical drug approvals. Compassionate human use of ganciclovir demonstrated efficacy and safety in treating acquired immunodeficiency syndrome-related cytomegalovirus retinitis in more

human patients than would generally be required for a phase I clinical trial, but the FDA refused to license it due to lack of animal studies. Ganciclovir had also been used safely in over 300 patients under compassionate use to treat cytomegalovirus colitis—more than would generally be required in a phase II clinical trial—but the FDA delayed clinical trials for more than a year due to lack of animal studies. The drug was finally approved after a 4-year delay (42).

TIME AND DOLLARS. Rodent testing in cancer therapeutics adds an estimated 4 to 5 years to drug development and costs $2 to $4 million. For industrial toxicity testing, it takes about 10 years and $3 million

FIGURE 2   Toxicity Failures in Pharmaceutical Development
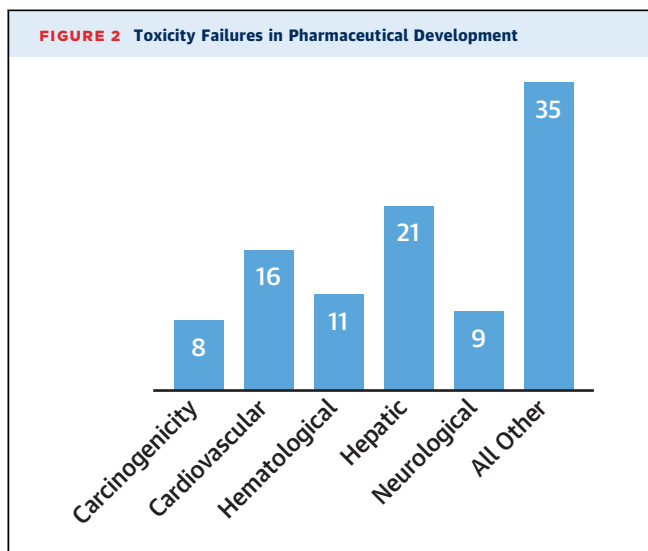
to complete all required animal studies to register a single pesticide (43). Compared with the costs of in vitro testing, animal tests range from 1.5× to >30× as expensive (44,45).

No comprehensive reviews of the total overall cost of animal testing in pharmaceutical development appear to exist. In part, this may be because even the total number of animals or of such studies is unknown. The 2002 amendments to the Animal Welfare Act exempted mice, rats, fish, and birds used in animal research from required reporting to the U.S. Department of Agriculture (46). These are the 4 most common types of animals used, and they account for >90% of all U.S. animal subjects and 81% of European animal subjects (45,47).

Costs of animal toxicity tests can be estimated from other industries, however, and are eye-opening. According to the Organization for Economic Development, which determines animal testing guidelines and methodology for government, industry, and independent laboratories in its several dozen member countries, the average cost of a single, 2-generation reproductive animal toxicity study worldwide is €318,295 and for Europe alone is €285,842 (45), or roughly $349,890 and $314,215, respectively.

Contract research organizations account for most of the animal testing done in the United States and Europe. Statista, a global data portal for market and economic sector statistics, estimates the global markets for animal testing in 2018 at $7.4 billion for drug discovery, $11.2 billion for preclinical development and safety, $58.5 billion for clinical development, and $2.3 billion for central laboratory testing

(48). Keen (49) estimates that annual U.S. biomedical and agricultural research and development investments involving animal research exceed $26 billion.

REPRODUCIBILITY AND INTERSPECIES RELIABILITY OF ANIMAL TESTS

Reproducibility of animal studies within species, even when carried out under rigorous protocols, is questionable. Using a database of more than 800,000 animal toxicity studies performed for 350 chemicals under rigorous guidelines, a reviewer found toxicity was repeatable just 70% of the time in the same species (45). Another reviewer found that results for a single chemical often differed with animal model, strain, dose, and delivery route. About 26% of chemicals demonstrated contradictory results on repeat testing in the same species. Furthermore, discordant results sometimes ranged over 3 orders of magnitude within the same species (50).

PPV, NPV, AND LR. Sensitivity reflects how likely a positive test is to detect all subjects with a condition, and specificity reflects how likely a negative test is to exclude all subjects without the condition (**Figure 3**). PPV reflects how often a positive test actually identifies a subject with the condition, and the power of the negative predictive value (NPV) of a test reflects the proportion of subjects with negative tests that actually do not have the condition. Whereas sensitivity, specificity, PPV, and NPV are often used to describe the accuracy of tests, they are not sufficient to inform us how much "value" to attribute to any given test. For example, suppose a positive toxicity test in mice for a group of drugs always predicts human toxicity (sensitivity = 100%), but it also indicates human toxicity when it is not present—in fact the test results always indicate that the drug is toxic. Such a test would have virtually no use in determining human toxicity despite being 100% sensitive. A useful toxicity test is 1 that also indicates accurately when toxicity in animals is not present in humans or has high specificity. Furthermore, we want to know how often the test accurately indicates human toxicity, compared with how accurately it indicates human nontoxicity.
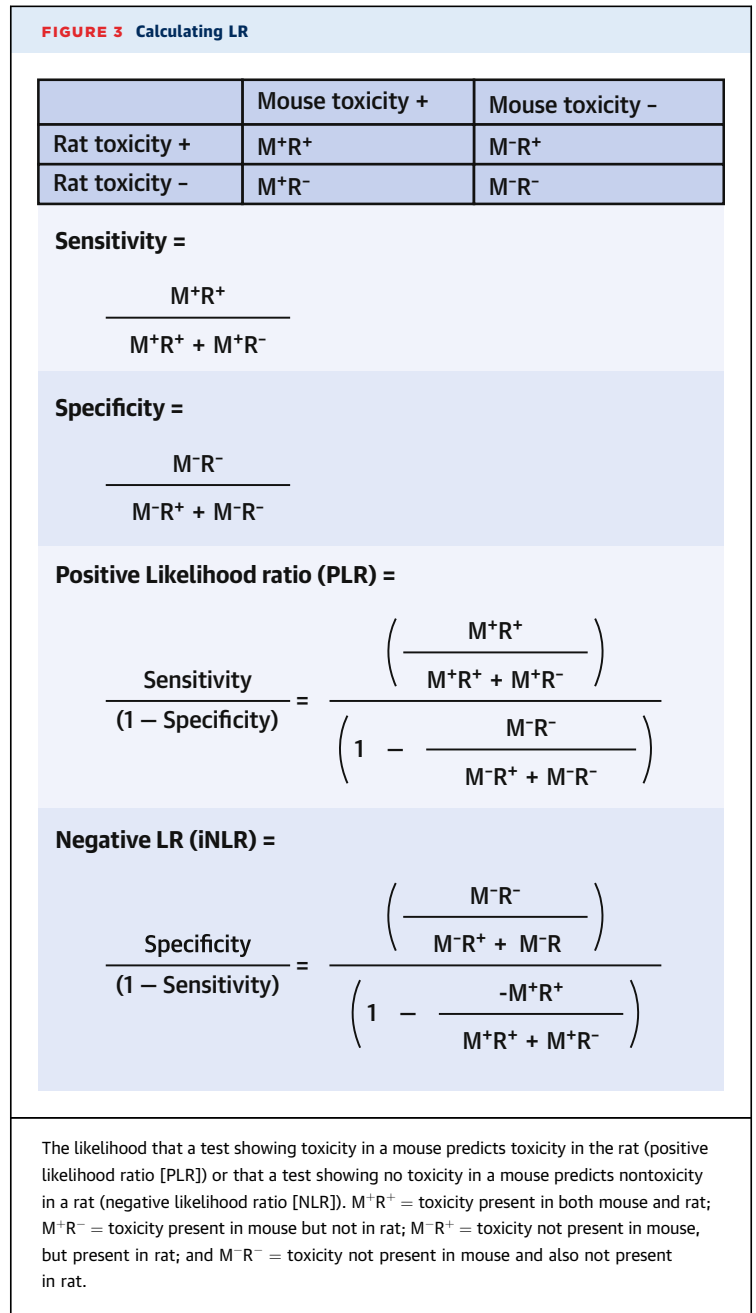
Sensitivity, specificity, PPV, and NPV are all strongly affected by the prevalence of the condition they test for and are therefore of limited value in assessing the reliability of a test when the prevalence of the condition is unknown. Lower prevalence increases the likelihood of false positive results, and higher prevalence increases the

likelihood of false negative results. Once a drug tests positive for toxicity in animals, it is rarely tested against humans, and the prevalence of the real condition the test is being used to "detect"—human toxicity—remains unknown.

However, the "value" of using a given test to improve the post-test probability of ruling in or ruling out a condition can be calculated using likelihood ratios even if the prevalence of the condition is unknown, so long as the sensitivity and specificity of the test are known. LR are indicators of whether the results of a given test will "add weight" over the pre-test probabilities (i.e., prevalence rate) of the condition in deciding what the probability is that a condition is actually present or absent.

There are 2 types of LR: the positive likelihood ratio (PLR) indicates how much more likely it is that a condition exists after a positive test result, when compared with its pre-test probability. The negative likelihood ratio (NLR) indicates how much the probability that a condition exists decreases compared with its pre-test probability, given a negative test result. The change in post-test probability from pre-test probability is calculated by multiplying the pre-test probability (prevalence) by the PLR or NLR. If the change in post-test probability from the pre-test probability is small (i.e., LR, the multiplier, is small), then the test is unlikely to help determine the presence or absence of a condition over simply knowing its prevalence. LR of <1.0 actually indicate a negative shift in post-test probabilities. In other words, if a PLR is <1.0, then for any subject that has a positive test result, the probability that they have the condition decreases compared with the pre-test probability. For an NLR of <1.0, for any subject with a negative test, the probability that they do not have the condition also decreases compared with the pre-test probability. For an LR of 1.0, there is no change from pre-test probabilities (pre-test probabilities are simply multiplied by 1), and the test also was not useful. For LR >1.0, the probability of the condition being present increases in the face of a positive test, and the probability of the condition being absent increases in the presence of a negative test. For LR from 1.0 to 10, these changes are relatively small (meaning the test will not add much), but for LR >10, the changes increase exponentially and are considered significant (51-54).

Using LR to calculate the probability that a test will improve detection of a condition or ruling it out is complex; it requires knowing the sensitivity and specificity of a test and pre-test probabilities, conversion of probabilities to odds and back again, and then using a log table (i.e., a Fagan's nomogram) or log calculator to determine how much a test is likely

FIGURE 3 Calculating LR

	Mouse toxicity +	Mouse toxicity –
Rat toxicity +	M⁺R⁺	M⁻R⁺
Rat toxicity –	M⁺R⁻	M⁻R⁻

Sensitivity =

$$\frac{M^+R^+}{M^+R^+ + M^+R^-}$$

Specificity =

$$\frac{M^-R^-}{M^-R^+ + M^-R^-}$$

Positive Likelihood ratio (PLR) =

$$\frac{\text{Sensitivity}}{(1 - \text{Specificity})} = \frac{\left(\dfrac{M^+R^+}{M^+R^+ + M^+R^-}\right)}{\left(1 - \dfrac{M^-R^-}{M^-R^+ + M^-R^-}\right)}$$

Negative LR (iNLR) =

$$\frac{\text{Specificity}}{(1 - \text{Sensitivity})} = \frac{\left(\dfrac{M^-R^-}{M^-R^+ + M^-R}\right)}{\left(1 - \dfrac{-M^+R^+}{M^+R^+ + M^+R^-}\right)}$$

The likelihood that a test showing toxicity in a mouse predicts toxicity in the rat (positive likelihood ratio [PLR]) or that a test showing no toxicity in a mouse predicts nontoxicity in a rat (negative likelihood ratio [NLR]). M⁺R⁺ = toxicity present in both mouse and rat; M⁺R⁻ = toxicity present in mouse but not in rat; M⁻R⁺ = toxicity not present in mouse, but present in rat; and M⁻R⁻ = toxicity not present in mouse and also not present in rat.

to improve (or decrease) the chances of detecting the condition (53).

LR are increasingly being used to express translatability of animal toxicity testing (52-55). Bailey et al. (14) found that the presence of toxicity in a species sometimes added evidentiary weight to the risk of toxicity in another, but the reverse was not true: negative toxicity tests in animals did not significantly increase the probability that a toxic test would also be negative in humans, and a lack of toxicity in any species would not reliably indicate a

probable lack of toxicity in any other species, including comparisons of primate to human toxicity tests (14). Furthermore, even in the presence of animal toxicity, LRs were extremely inconsistent and varied considerably for different classes of drugs (13,52). Similar findings have been reported in multiple analyses and reviews in other studies (52,54-56).

A number of studies have reviewed LR of specific drug toxicity tests for which both animal and human data are available. In a review of 2,366 drugs, including data from 3 of the most common animal research species—rat, mouse, and rabbit—PLRs were generally high (i.e., there is a likelihood that positive toxicity tests in animals would show toxicity in humans). But median NLRs were very low—1.12 (rabbit), 1.39 (mouse), and 1.82 (rat); in other words, they were of little or no value in excluding human toxicity (13). The investigators also examined canine models and found that PPV and PLR for human toxicity were not correlated with 1 another: NLR were low, indicating that the dog provided little evidentiary weight to ruling out toxicity in humans (52). Later analysis of 3,000 drugs found that tests inferring no toxicity in any 1 species, including nonhuman primates, have no evidentiary weight with regard to toxicity in any other species (14). In a comparison study reported by pharmaceutical companies of 150 drugs associated with adverse events or toxicity in humans (55), LR could not be determined due to a lack of specificity reporting on the tests. Paglialunga et al. (56) examined translatability of respiratory safety pharmacology studies from animal models to humans and found that PPV and PLR were so low that animal tests provided little value in predicting human toxicity.

GROWING SCIENTIFIC CRITICISM

As early as 1962, scientists questioned the assumption that animal models reliably predicted human responses. Lichtfield (57) examined 6 drugs studied in animal models and found that rats and dogs demonstrated PPVs (for human response) of 0.49 and 0.55, respectively, essentially random chance. He opined that the differences between species in specific drug responses were so striking that one could actually use the results of drug toxicity tests alone to identify whether an entity was a rat, rather than a dog or a man, and concluded there was no basis for predicting adverse human effects for the 6 drugs from animal studies. A 1990 analysis of the toxicities of 24 drugs abandoned during human clinical trials demonstrated that 16 had no animal model toxicity correlation (58).

In 1981, the Council on Scientific Affairs of the American Medical Association stated, "The Council's consultants agree that to identify carcinogenicity in animal tests does not *per se* predict either risk or outcome in human experience. . . . the Council is concerned about the hundred[s] of millions of dollars that are spent each year (both in the public and private sectors) for the carcinogenicity testing of chemical substances. The concern is particularly grave in view of the questionable scientific value of the tests when used to predict the human experience" (59).

The 2019 West Coast Regional Safety Pharmacology Society Meeting discussed concerns about the lack of concordance between animal and human safety studies, including lack of canine and human concordance for proarrhythmia risks of new cardiovascular drugs and the failure of animal research to predict drug-related risks in the human central nervous and respiratory systems (60,61).

Regulatory and research leaders are increasingly taking notice of the issue. In 2006, Michael Levitt, then U.S. Secretary of Health and Human Services, stated, "nine out of ten experimental drugs fail in clinical studies because we cannot accurately predict how they will behave in people based on laboratory and animal studies" (62,63). A landmark review and report by the Institute of Medicine in 2011 concluded that the use of chimpanzees in biomedical research is unnecessary (64). Although the reasons for it are complex, in 2015, the National Institutes of Health announced they would be ending all chimpanzee research (65). Andrew Wheeler, administrator of the U.S. Environmental Protection Agency, pledged in September 2019 to phase out all toxicity testing in mammals over the next 16 years (66).

IS THE SCIENTIFIC PREMISE BEHIND ANIMAL MODELS VALID?

Many concerns regarding reliability of animal models in predicting human toxicity are not based on the scientific underpinnings of interspecies translation, but rather call out collateral, potentially correctable issues, such as technical competence in executing animal research, the soundness of animal research study design, and publication bias (**Table 2**) (67-71). Indeed, Knight (72) could find no review of animal research studies that rated a majority of the experiments as having "good" methodological quality. An obvious solution would be correction of these problems to improve translation rates of animal research. Despite widespread efforts to improve the quality of methodology in animal studies, however, studies examining whether such measures consistently improve the reliability of animal models in predicting human toxicity have yet to be published, although a

TABLE 2	Commonly Used Arguments Against Animal Research	
Argument	**Critique**
Methodological: Animal models should be abandoned because the scientific methodology of the experiment was poor.	The quality of methodology in an individual experiment cannot be extrapolated to the question of whether animal experimentation as a whole is invalid, merely to whether the individual experiment is yielding true results.
Historical: Historically, medical dependence on animal modeling is much less robust than we are led to believe.	Historical use of animal modeling is a poor measure of the validity of current experimentation and methods. To determine whether animal modeling is reliable in current science, we need to use modern scientific knowledge and examine modern methodology to determine whether animal modeling is predictive of human outcomes today. This takes into account information and methods that may or may not have been historically available.
Reviews: Review articles have determined that certain animal species have not been critical in various medical developments, and therefore animal experimentation should be abolished.	The invalidity of using certain specific animals does not necessarily rule out animal models as a whole.
Alternatives: The existence of alternative models requires us to abandon animal research.	Whereas alternatives to animal research exist or are developing in many areas of medical research, in many instances such alternatives do not exist. This argument does not address whether continued use of animal models is scientifically valid, regardless of alternative methods, and it does not attempt to define whether certain animal models are predictably successful and others are predictably unsuccessful.

number of studies do demonstrate continued problems with predicting human efficacy (73).

Instead of methodologies and publication bias, an increasing number of investigators propose that the problem may lie with the basic premise of animal testing itself (69,74). The biological sciences have increasingly embraced theories regarding complex systems (e.g., chaos theory and complexity theory) to explain mechanisms in evolution, the biology of cancer, the divergent properties of animal species, as well as the failures of translation of drug therapeutics from animal species to humans (75,76). Because animals and humans are classic examples of incompletely understood complex systems, some investigators propose that it may simply be scientifically invalid to assume that toxicity of a substance in any one species can reliably predict toxicity in any other, no matter how stringent animal testing standards are made (69).

ALTERNATIVES TO ANIMAL TESTING

Alternatives to animal testing will be discussed in more detail in part 2 of this review; they include in vitro tests using cell lines, tissue samples, use of alternative organisms such as bacteria, 3-dimensional modeling and bioprinting, in silico tests, organ-on-chip technologies such as 3-dimensional organoids, computer modeling, and phase 0 in-human microdosing trials (77-82). A comprehensive study of the accuracy, LR, and costs of alternative testing methods compared with animal toxicity testing has not been published; however, there is data suggesting that in vitro testing and other methods are significantly faster and less expensive than animal models (42,44).

Using human cells, tissue, or organ models to form the basis of an in vitro test may improve accuracy in weeding out drugs with significant adverse human effects; however, this assumption, too, will require rigorous study.

Researchers will undoubtedly be challenged sooner rather than later to reduce animal research as the result of public advocacy efforts. A 2019 spending bill passed by the U.S. House of Representatives includes a directive to the National Institutes of Health to accelerate the replacement of nonhuman primates in research with alternative research models (83).

The FDA states that for the purposes of cosmetic testing, they believe "that prior to use of animals, consideration should be given to the use of scientifically valid alternative methods to whole-animal testing" (84). The Interagency Coordinating Committee on Validation of Alternative Methods and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods were established in 1997 to coordinate the development, validation, acceptance, and harmonization of alternative toxicological test methods throughout the U.S. government (85) and have as a part of their mission the explicit mandate to reduce or eliminate whole animal testing. The Biennial Progress Report of the Interagency Coordinating Committee on Validation of Alternative Methods for 2016 to 2017 details actions they have taken, including, among others: 1) publication of guidance documents waiving all acute dermal lethality studies for pesticides and describing a process for evaluating; 2) publication of notices reducing the number of hamsters for potency testing of certain vaccines; and 3) publication of a roadmap for

integrating predictive toxicology methods into safety and risk assessments by the FDA (86). At this time, the FDA generally still requires submission of preclinical animal data in investigational new drug applications (5).

CONCLUSIONS

Although animal toxicity testing has been the stalwart basis of "ensuring" safety of in-human clinical testing and use, examination of the published data raises significant questions about whether it is reliable and should be abandoned or at least significantly curtailed in favor of other potentially more reliable methods. Savings in time and cost for new therapeutics could be substantial, if the safety of nonanimal preclinical testing is proven. Increasingly, scientific organizations and government regulatory agencies are recognizing that alternative methods may replace animal testing and improve the flow and safety of new therapeutics to human use.

ADDRESS FOR CORRESPONDENCE: Dr. Gail A. Van Norman, Department of Anesthesiology and Pain Medicine, University of Washington, 2141 8th Avenue West, Seattle, Washington 98119. E-mail: gvn@uw.edu.

REFERENCES

1. Wax PM. Elixirs, diluents and the passage of the 1938 Federal Food, Drug and Cosmetics Act. Ann Intern Med 1995;122:456–61.

2. Greek R, Pippus A, Hansen LA. The Nuremberg Code subverts human health and safety by requiring animal modeling. BMC Med Ethics 2012; 13:16.

3. National Institutes of Health. Regulations and Ethical Guidelines. Reprint from: Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law no. 10, Vol. 2. 1949 1949. Washington, DC: U.S. Government Printing Office, 1949:181–2. Available at: https://history.nih.gov/research/downloads/nuremberg.pdf. Accessed October 5, 2019.

4. World Medical Assembly. Declaration of Helsinki 1964: Recommendations Guiding Doctors in Clinical Research. World Medical Association. June 1964. Available at: https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/doh-jun1964/. Accessed October 5, 2019.

5. U.S. Food and Drug Administration. Investigational New Drug (IND) Application. U.S. Food and Drug Administration. October 5, 2017. Available at: https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application. Accessed August 20, 2019.

6. U.S. Food and Drug Administration. Product Development Under the Animal Rule: Guidance for Industry. U.S. Food and Drug Administration. October 2015. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/product-development-under-animal-rule. Accessed September 23, 2019.

7. Fomchenko EK, Holland EC. Mouse models of brain tumors and their applications in preclinical trials. Clin Cancer Res 2006;12: 5288–97.

8. Huff J, Jacobson MF, Davis DL. The limits of two-year bioassay exposure regimens for identifying chemical carcinogens. Environ Health Perspect 2008;116:1439–42.

9. Gad S. Preface. In: Gad S, editor. Animal Models in Toxicology. Boca Raton, FL: CRC Press, 2007:1–18.

10. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. JAMA 2006;296:1731–2.

11. Perel P, Roberts I, Sena E, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 2007; 334:197–203.

12. Wang B, Gray G. Concordance of noncarcinogenic endpoints in rodent chemical bioassays. Risk Anal 2015;35:1154–66.

13. Bailey J, Thew M, Balls M. An analysis of the use of animal models in predicting human toxicology and drug safety. Altern Lab Anim 2014;42: 189–99.

14. Bailey J, Thew M, Balls M. Predicting human drug toxicity and safety via animal tests: can any one species predict drug toxicity in any other, and do monkeys help? Altern Lab Anim 2015;43: 393–403.

15. American Pharmaceutical Review. Preclinical Development: The Safety Hurdle Prior to Human Trials. American Pharmaceutical Review. April 30, 2016. Available at: https://www.americanpharmaceuticalreview.com/Featured-Articles/187349-Preclinical-Development-The-Safety-Hurdle-Prior-to-Human-Trials/. Accessed September 24, 2019.

16. Van Norman GA. Drugs, devices and the FDA: part 1: an overview of approval processes for drugs. J Am Coll Cardiol Basic Trans Science 2016; 1:170–9.

17. Van Norman GA. Phase II trials in drug development and adaptive trial design. J Am Coll Cardiol Basic Trans Science 2019;4:428–37.

18. Compton K. Vioxx Lawsuits. Drugwatch. Available at: https://www.drugwatch.com/vioxx/lawsuits/. Accessed September 16, 2019.

19. Juni P, Dip LN, Reichenback S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet 2004;364:2021–9.

20. Siramshetty BV, Nickel J, Omieczynski C, et al. WITHDRAWN—a resource for withdrawn and discontinued drugs. Nucl Acids Res 2016;44: D1080–6.

21. van Meer PJ, Kooijman M, Gispen-de Wied CC, Moors EH, Schellekens H. The ability of animal studies to detect serious post marketing adverse events is limited. Reg Tox Pharm 2012;64:345–9.

22. Lupkin S. One-Third of New Drugs Had Safety Problems After FDA Approval. NPR: Shots: Health News From NPR. May 9, 2017. Available at: https://www.npr.org/sections/health-shots/2017/05/09/527575055/one-third-of-new-drugs-had-safety-problems-after-fda-approval. Accessed October 5, 2019.

23. Downing NS, Shah ND, Aminawung JA, et al. Postmarket safety events among novel therapeutics approved by the US Food and Drug Administration between 2001 and 2010. JAMA 2017;317: 1854–63.

24. Brinks V, Jiskoot W, Schellekns H. Immunogenicity of therapeutic proteins: the use of animal models. Pharm Res 2011;28:2379–85.

25. McDougal AJ. Preclinical development [CDER]: biological therapeutics for cancer treatment. Paper presented to iSBTc Oncology Biologics Development Primer; February 28, 2008; Gaithersburg, MD. Available at: https://sitc.sitcancer.org/meetings/am08/primer08_oncology/presentations/3_mcdougal.pdf. Accessed September 24, 2019.

26. De Groot AS, Moise L. Prediction of immunogenicity for therapeutic proteins: state of the art. Curr Opin Drug Discov Devel 2007;10:332–40.

27. Pedras-Vasconcelos JA. The Immunogenicity of therapeutic Proteins—What You Don't Know Can Hurt YOU and the Patient. U.S. Food and Drug Administration. Fall 2014. Available at: https://www.fda.gov/media/91668/download. Accessed September 24, 2019.

28. Stolley PD. Asthma mortality: why the United States was spared an epidemic of deaths due to asthma. Am Rev Respir Dis 1972;105:883–90.

29. Schardein J. Drugs as Teratogens. Cleveland, OH: CRC Press, 1976:49.

30. Suntharalingam G, Perry MR, Ward S, et al. Cytokine storm in phase 1 trial of the anti-CD28

monoclonal antibody TGN1312. N Engl J Med 2006;355:1018–28.

31. Dayan CM, Wraith DC. Preparing for first-in-man studies: the challenges for translational immunology post TGN1412. Clin Exp Immunol 2008;151:231–4.

32. Attarwala H. TGN1412: from discovery to disaster. J Young Pharm 2010;2:332–6.

33. Allen F. "Like a horror film": what was the "Elephant Man" drug testing trial, what is TGN1412 and what happened to the men involved? The Sun (UK). February 21, 2017. Available at: https://www.thesun.co.uk/news/2917810/elephant-man-drug-testing-trial-tgn1412/. Accessed October 5, 2019.

34. Vial T, Descotes J, Braun F, Behrend M. Chapter 37. Drugs that act on the immune system: 37—-cytokines and monoclonal antibodies. Side Effects Drugs Annu 2007;29:383–423.

35. Eddleston M, Cohen AF, Webb DJ. Implications of the BIA-102474-101 study for review of first-into-human clinical trials. Brit J Clin Pharm 2016; 81:582–6.

36. Sankar U. The delicate toxicity balance in drug discovery. Scientist 2005;19:32.

37. Gura T. Systems for identifying new drugs are often faulty. Science 1997;273:1041–2.

38. Koppanyi T, Avery MA. Species differences and the clinical trial of new drugs: a review. Clin Pharmacol Ther 1966;7:250–70.

39. Cormia FE, Lewis GM, Hopper ME. Toxicity of penicillin for the guinea pig. J Invest Dermatol 1947;9:261–7.

40. Villar D, Buck WB, Gonzalez JM. Ibuprofen, aspirin and acetaminophen toxicosis and treatment in dogs and cats. Vet Human Toxicol 1998; 40:156–62.

41. Wilson JG, Ritter EJ, Scott WJ, Fradkin R. Comparative distribution and embryotoxicity of acetylsalicylic acid in pregnant rats and rhesus monkeys. Toxicol Appl Pharmacol 1977;41:67–78.

42. Buhles WC. Compassionate use: a story of ethics and science in the development of a new drug. In: Monticello M, editor. Perspectives and Biology in Medicine, Vol. 54. Baltimore, MD: Johns Hopkins University Press, 2011:304–15.

43. Humane Society International. Limitations of Animal Tests. Humane Society International. February 24, 2019. Available at: https://www.hsi.org/news-media/limitations-of-animal-tests/. Accessed September 14, 2019.

44. Humane Society International. Costs of Animal and Non-animal Testing. Humane Society International. October 23, 2012. Available at: https://www.hsi.org/news-media/time_and_cost/. Accessed September 14, 2019.

45. Meigs L, Smirnova L, Rovida C, Leist M, Hartung T. Animal testing and its alternatives—the most important omics is economics. ALTEX 2018; 35:275–305.

46. USDA National Agricultural Library, U.S. Department of Agriculture. Animal Welfare Act. USDA National Agricultural Library, U.S. Department of Agriculture. Available at: https://www.nal.usda.gov/awic/animal-welfare-act. Accessed October 5, 2019.

47. Taylor K, Gordon N, Langley G, Higgins W. Estimates for worldwide laboratory animal use in 2005. Altern Lab Anim 2008;36:327–42.

48. Statista Research Department. Total Addressable Contract Research Organization (CRO) Market Worldwide in 2018, by Function (in Million U.S. Dollars). Statista. July 22, 2019. Available at: https://www.statista.com/statistics/814232/total-addressable-cro-market-estimates-worldwide-by-function/. Accessed September 16, 2019.

49. Keen J. Wasted money in United States biomedical and agricultural animal research. In: Hermann K, Jayne K, editors. Animal Experimentation: Working Towards a Paradigm Change, Vol 22. Boston, MA: Brill, 2019:244–72.

50. Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. Screening chemicals for estrogen receptor bioactivity using a computations model. Environ Sci Technol 2015;49: 8804–14.

51. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329:168.

52. Bailey J, Thew M, Balls M. An analysis of the use of dogs in predicting human toxicology and drug safety. Altern Lab Anim 2013:335–50.

53. McGee S. Simplifying likelihood ratios. J Gen Intern Med 2002;17:646–9.

54. Grimes DA, Schultz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005;365: 1500–5.

55. Matthews RA. Medical progress depends on animal models—doesn't it? J R Soc Med 2008;101:95–8.

56. Paglialunga S, Morimoto BH, Clark M, Friedrichs GS. Translatability of the S7A core battery respiratory safety pharmacology studies: preclinical respiratory and related clinical adverse events. J Pharmacol Toxicol Methods 2019 Jun 5 [E-pub ahead of print].

57. Lichtfield JT Jr. Symposium on clinical drug evaluation and human pharmacology. XVI. Evaluation of the safety of new drugs by means of tests in animals. Clin Pharmacol Ther 1962;3:665–72.

58. Lumley C. Clinical toxicity: could it have been predicted? Premarketing experience. In: Lumley C, Walker S, editors. Animal Toxicity Studies: Their Relevance for Man. London, UK: Quay Publishing, 1990:49–56.

59. Council on Scientific Affairs. Carcinogen regulation. JAMA 1981;246:253–6.

60. Abi-Gerges N, McMahon C, Vargas H, et al. The west coast regional safety pharmacology society meeting update: filling in the translational gaps in safety assessment. J Pharmacol Toxicol Methods 2019;98:106582.

61. Ewert L, Aylott M, Deurinck M, et al. The concordance between nonclinical and phase I clinical cardiovascular assessment from a cross-company data sharing initiative. Toxicol Sci 2014; 142:427–35.

62. Shanks N, Greek R, Greek J. Are animal models predictive for humans? Philos Ethics Humanit Med 2009;4:2.

63. U.S. Food and Drug Administration. Press release: FDA issues advice to make earliest stages

of clinical drug development more efficient. January 12. Washington DC: U.S. Food and Drug Administration, 2006.

64. Institute of Medicine Committee on the Use of Chimpanzees in Biomedical and Behavioral Research. Chimpanzees in Biomedical and Behavioral Research: Assessing the Necessity. Washington, DC: National Academic Press, 2011.

65. Milman O. National Institutes of Health announces end to chimpanzee research. The Guardian. November 19, 2015. Available at: https://www.theguardian.com/us-news/2015/nov/19/national-institutes-health-end-chimpanzee-research. Accessed September 26, 2019.

66. Greenfieldboye N. EPA Chief Pledges to Severely Cut Back on Animal Testing of Chemicals. NPR Health News. September 10, 2019. Available at: https://www.npr.org/sections/health-shots/2019/09/10/759435118/epa-chief-pledges-to-severely-cut-back-on-animal-testing-of-chemicals. Accessed September 26, 2019.

67. Bracken MB. Why animal studies are often poor predictors of human reactions to exposure. J R Soc Med 2009;102:120–2.

68. van der Worp HB, Howells DW, Sena ES, et al. Can animal models of disease reliably inform human studies? PLoS Med 2010;7:e10000245.

69. Greek R, Kramer LA. How to evaluate the science of non-human animal use in biomedical research and testing: a proposed format for debate. In: Herrmann K, Jayne K, editors. Animal Experimentation: Working Towards a Paradigm of Change, Vol. 22. Boston, MA: Brill Open, 2019: 65–87.

70. Ioannidis JPA. Extrapolating from animals to humans. Sci Transl Med 2012;4:151.

71. Greek R, Menache A. Systematic reviews of animal models: methodology versus epistemology. Int J Med Sci 2013;10:206–21.

72. Knight A. Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. Altern Lab Anim 2007;35: 641–59.

73. Akhtar A. The flaws and human harms of animal experimentation. Camb Q Healthc Ethics 2015; 24:407–19.

74. Pound P, Ritskes-Hoiting M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. J Transl Med 2018;16:304.

75. Miquel PA. Extended physics as a theoretical framework for systems biology? Prog Biophys Mol Biol 2011;106:348–52.

76. Soto AM, Sonnenschein C, Maini PK, Novle D. Systems biology and cancer. Prog Biophys Mol Biol 2011;106:337–46.

77. Groell F, Jordan O, Borchard G. In vitro models for immunogenicity prediction of therapeutic proteins. Eur J Pharm Biopharm 2018;130:128–42.

78. Mittal R, Woo FW, Castro CS, et al. Organ-on-chip models: implications in drug discovery and clinical applications. J Cell Physiol 2019;234: 8352–80.

79. Benam KH, Gilchrist S, Kleensang A, Satz AB, Willet C, Zhang Q. Exploring new technologies in

biomedical research. Drug Discov Today 2019;24: 1242-7.

80. Doke SK, Dhawale SC. Alternatives to animal testing: a review. Saudi Pharm J 2015;23:223-9.

81. Burt T, Vuong LT, Baker E, et al. Phase 0, including microdosing approaches: applying the three Rs and increasing the efficiency of human drug development. Altern Lab Anim 2018;46. 335-4.

82. Herrmann K, Pistollato F, Stephens ML. Beyond the 3 Rs: expanding the use of human-relevant replacement methods in biomedical research. ALTEX 2019;36:343-52.

83. Reardon S. U.S. Lawmakers Propose Plan to Reduce Primate Research at the National In-stitutes of Health. Nature. May 9, 2019. Available at: https://www.nature.com/articles/d41586-019-01511-0?utm_source=twt_nnc and utm_medium=social and utm_campaign=naturenews and sf212424116=1. Accessed September 26, 2019.

84. U.S. Food and Drug Administration. Animal Testing and Cosmetics. U.S. Food and Drug Adminis-tration. May 31, 1999, updated 2006. Available at: https://www.fda.gov/cosmetics/product-testing-cosmetics/animal-testing-cosmetics. Accessed October 1, 2019.

85. U.S. ICCVAM. Interagency Coordinating Committee on Validation of Alternative Methods. Research Triangle Park, NC: U.S. ICCVAM, 2019. Available at: https://www.niehs.nih. gov/health/materials/interagency_coordinating_committee_on_the_validation_of_alternative_methods_5 08.pdf. Accessed October 12, 2019.

86. U.S. ICCVAM. Biennial Progress Report 2016–2017: Interagency Coordinating Committee on Validation of Alternative Methods. July 2018. Available at: https://ntp.niehs.nih.gov/iccvamreport/2017/bienrpt-txt.pdf. Accessed October 10, 2019.

KEY WORDS animal research, drug development, toxicity, translational research