2019

The Scientific Problems with Using Non-Human Animals to Predict Human Response to Drugs and Disease

Ray Greek
Americans For Medical Advancement

Lisa A. Kramer
University of Toronto

Recommended Citation

Greek, R., & Kramer, L. A. (2019). The scientific problems with using non-human animals to predict human response to drugs and disease. In Animal Experimentation: Working Towards a Paradigm Change (pp. 391-416). Brill.

CHAPTER 17

The Scientific Problems with Using Non-Human Animals to Predict Human Response to Drugs and Disease

Ray Greek
President, Americans for Medical Advancement, California, United States

Lisa A. Kramer
Professor of Finance, University of Toronto, Ontario, Canada

1 Introduction

Every year, and in countries around the world, significant time and resources are devoted to the noble cause of developing drugs to treat and cure human disease. With rare exception, drug interventions cannot reach commercialization without safety and efficacy having first been demonstrated in animal models. The intention of regulations, which require the use of animal models in such contexts, is to ensure that only safe and effective drugs end up being used by patients. Similarly, it is standard practice for researchers to employ animal models in their attempts to understand the way diseases present and progress in humans. Unfortunately, there exist serious theoretical and empirical concerns regarding the standard practice of using non-human animals to model human response to perturbations, such as drugs and disease. These concerns are important because conducting disease research and drug development in a manner that is not supported by science will have suboptimal implications for the humans who rely on that research, which encompass the entire population. Based on complexity science, modern evolutionary biology, and empirical evidence, we demonstrate that animal models have failed as predictors of human response. That is, animal models do not and cannot have acceptably high predictive value for human response to drugs and disease. By this we mean that animal modeling, as a methodology, is for all practical purposes not predictive of human response to drugs and disease; and hence it should be abandoned in favor of human-based research and testing, such as personalized medicine, a new field that takes into account the unique genetic make-up of each individual patient.

People are accustomed to hearing about the ethical issues arising from the use of non-human animals in biomedical research, testing, and science in general. But there are scientific issues with the practice as well. Researchers who employ animal modeling often attempt to justify the practice based on claims of accurately predicting human response to drugs and disease. For example, Giles (2006, p. 981) states: "In the contentious world of animal research, one question surfaces time and again: how useful are animal experiments as a way to prepare for trials of medical treatments in humans? The issue is crucial, as public opinion is behind animal research, only if it helps develop better drugs. Consequently, scientists defending animal experiments insist they are essential for safe clinical trials, whereas animal-rights activists vehemently maintain that they are useless".

One need not search hard to find examples claiming non-human animals play an essential role in the quest to treat and cure human disease. For example, the American Physiological Society (APA) (2017) states on its website: "Animals are used in research to develop drugs and medical procedures to treat diseases." Andrew B. Rudczynski, Yale University's associate vice president for research administration, stated in a letter to the editor (2011): "[T]he basic research model used by Yale University and its peer institutions is scientifically valid and predictive of human disease". Michael F. Jacobson, executive director of the Center for Science in the Public Interest (2008) stated: "We must test animals to determine whether a substance causes cancer". Huff, Jacobsen, and Davis (2008, p. 1439) stated: "Chemical carcinogenesis bioassays in animals have long been recognized and accepted as valid predictors of potential cancer hazards to humans." Lin (1995, p. 1008) stated: "Although the validity of animal testing to predict efficacy and or safety in humans has been questioned, it is generally believed that data from animal studies can be reasonably extrapolated to humans with the application of appropriate pharmacokinetic principles [....] From an evolutionary point of view, all mammals are similar, because they originate from a common ancestor, yet they differentiate because of their dissimilar environmental adaptations".

While it can be argued that there may be scientifically justified grounds for the use of non-human animals in some contexts, other than those that involve predicting human responses, it is most common to see attempts to justify the use of non-human animals for applications to human health (see Kramer and Greek (2018), for additional discussion of this point). Therefore, it is appropriate to carefully examine the claimed validity of the animal model for predicting human outcomes.

To that end, consider *Trans-Species Modeling Theory* (TSMT), a concept that was formalized by Greek and Hansen (2013), based on a combination of

extensive previous research on complex systems science and evolutionary biology, as summarized by authors, including Greek and Rice (2012), LaFollette and Shanks (1996), LaFollette and Shanks (1998), Shanks and Greek (2008), and Shanks and Greek (2009). TSMT states: "While trans-species extrapolation is possible when perturbations concern lower levels of organization or when studying morphology and function on the gross level, one evolved complex system will not be of predictive value for another when the perturbation affects higher levels of organization" (Greek and Hansen, 2013, p. 254). That is, according to science, the observation of a drug response in one species is uninformative about the drug response in another species. This theory is based on complexity science, evolutionary biology, and empirical evidence. In the remainder of this article, we explain why the fields of complexity science and evolutionary biology are relevant to understanding animal modeling and evaluating the inability of animal models to predict human response to drugs and disease.

LaFollette and Shanks (1996) and the Medical Research Modernization Committee (2006) were among the first to document systematically the methodological failure of using one evolved complex system to model another, in terms of predicting outcomes. Subsequent work by Greek and Hansen (2013), Greek and Rice (2012), Shanks and Pyles (2007), and Shanks and Greek (2009) then led to the development of TSMT, which is the only theory (we intentionally use the word *theory* as opposed to *hypothesis*; see National Academies of Science Engineering Medicine, 2016) that accounts for both past and present successes and failures of animal modeling. It is also the only theory that explains why animal models will never offer practical predictive value for disease and drug research. To be clear, the aforementioned authors did not discover evolution, complexity science, or any aspect of probability. Rather, they relied on what had been previously published in those disciplines and combined various insights to formalize the case against the use of animal models to predict outcomes in other species.

TSMT was a paradigm shift in animal modeling analysis. Moreover, TSMT was inclusive of valid past criticisms, while simultaneously explaining and taking those criticisms further. For example, TSMT obviated the need to point out that small differences in environments among lab animals influenced results, as many anti-vivisectionists did and continue to do, because even under perfect environmental conditions, one evolved complex system would not be expected to have predictive value for another. Likewise, there is little to no value in analyzing why one species has historically been inadequate for predicting human response, because according to TSMT, no species, regardless of genetic similarity, will ever be similar enough to another to serve as a valid predictive model. TSMT is also more precise and has more explanatory

power than general criticisms, such as *species differ in their metabolisms*. Furthermore, TSMT explains why increasing scientific rigor, the current mantra for justifying the use of animal models, will have no effect on predictive value.

We now turn to examining the three pillars underlying TSMT, comprising complex systems science, evolutionary biology, and empirical evidence.

2 Complex Systems

Advances in the field of complex systems have highlighted the poor predictive value of animal modeling. The study of complex systems and chaotic systems, currently usually classified under the general heading of complex systems, dates back to the 1950s and began a revolution in physics, similar to that of the early 1900s involving relativity and quantum mechanics (Gell-Mann, 1994; Gleick, 2008; Goodwin, 2001).

The following are characteristics of simple systems:

- They are nothing more than the sum of their parts.
- They have predictable behaviors. (There are no unanticipated or unexpected behaviors.)
- They are usually composed of just a few components.
- They can be intuitively understood.
- They are in equilibrium. (They are non-dynamic.)
- There are few interactions and feedback loops. (For example, compare a primitive barter system in contrast to our modern market-based economy).

Rosen (1999, p. 392) states: "A system is simple if all its models are simulable. A system that is not simple, and that accordingly must have a nonsimulable model, is complex". This should give us pause: A complex system is nonsimulable. Note that *simulable* may mean different things to different people. When scientists state that biological complex systems are nonsimulable, they mean nonsimulable *at the complex level*. The aim of researchers who use animal models is not to gain insight into the simple systems that are basic building blocks of the complex system. For example, at the simple level, we can rely on knowledge about simple systems to extrapolate that the final outcome for two different species will be the same when, for example, they are permanently deprived of water or they are thrown out of an airplane at 30,000-feet elevation. Researchers attempt to use non-human animals to model humans at higher, *complex* levels of organization, because this is the level at which disease and drug effects occur. So, when an animal modeler claims that their model simulates a human, unless they are speaking of low levels of organization (much

simpler than the levels at which drug and disease responses occur), this is not possible.

In contrast to simple systems, complex systems are characterized by the following (see Figure 17.1 for a diagrammatic representation of a complex system):

- Complex systems are composed of many parts that themselves have hierarchal levels of organization.
- Complex systems have feedback loops.
- Complex systems exhibit self-organization.
- Complex systems respond to perturbations in a *nonlinear* fashion. Because small changes in a complex system can result in outcomes that are not proportional to the input, one biological complex system can die because of what, at first, appears to be a minor change or difference between it and another almost identical complex system (Morange, 2001; Pearson, 2002). For example, Northrop (2011, p. xiv) states: "Early bioengineers, biophysicists, and systems physiologists tried to characterize certain physiological regulators as linear and stationary. Initially, linear systems analysis was inappropriately applied to certain complex, physiological regulators and

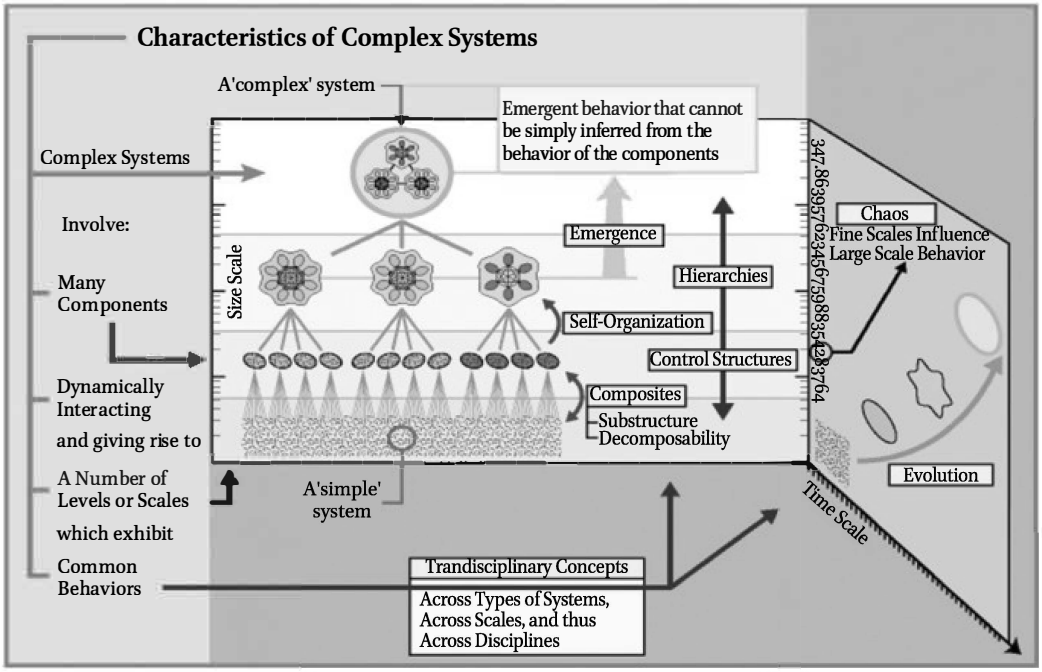

FIGURE 17.1 The characteristics of complex systems.
SOURCE: MARSHALL CLEMENS/IDIAGRAM (HTTPS://WWW.IDIAGRAM.COM)
Note: A complex system is built out of simple systems. As more and more of these simple systems combine and form a complex system, the level of organization increases and simulability decreases.

control systems (e.g., pupil regulation and eye movement control), which resulted in *black-box*, closed-loop models in which linear transfer function modules were connected to a nonlinear module in a single feedback loop. These were phenomenological input/output models that gave little insight into the physiology and complexity of the systems".

- Complex systems demonstrate redundancy and robustness. Complex systems have redundant parts and, therefore, losing a part may not affect function. Adding to this is robustness, which means that perturbations may not result in dysfunction. Complex systems have emergent properties that Aziz-Alaoui and Bertelle (2009, preface) define as follows: "Emergence and complexity refer to the appearance of higher-level properties and behaviors of a system that obviously comes from the collective dynamics of that system's components. These properties are not directly deductable from the lower-level motion of that system. Emergent properties are properties of the "whole" that are not possessed by any of the individual parts making up that whole. Such phenomena exist in various domains and can be described, using complexity concepts and thematic knowledges."
- Examples of emergent properties include the following from Van Regenmortel (2002):
  - The three physical states of water and phase transitions, such as boiling point.
  - The viscosity of water (individual water molecules have no viscosity).
  - The color of a chemical.
  - A melody arising from notes.
  - The saltiness of sodium chloride.
  - The specificity of an antibody.
  - The immunogenicity of an antigen.
- The components of complex systems can be grouped as modules, and the modules communicate with each other. Nevertheless, failure in one module does not necessarily spread to the system as a whole because of redundancy and robustness.
- Complex systems are dynamic. They communicate with, and change in response to, their environment.
- The whole of a complex system is greater than the sum of its parts, and hence complex systems have properties that cannot be determined even with total knowledge of the components of the system. This limits the validity of reductionism when studying complex systems.
- Importantly for our discussion, complex systems are also *very dependent on initial conditions*; for example, genetic make-up in the context of individuals or species. This means that a very small change in the initial conditions of

two otherwise identical complex systems (e.g., monozygotic twin humans), may result in sickness for one but not the other. In strains of mice, knocking out one gene has been shown to result in death for one strain, while the other thrives (Belmaker et al., 2012; Bell and Spector, 2011; Bruder et al., 2008; Castillo-Fernandez et al., 2014; Chapman and Hill, 2012; Czyz et al., 2012; Dempster et al., 2011; LeCouter et al., 1998; Raineri et al., 2001; Pearson, 2002).

The sensitivity of complex systems, also known as nonlinear dynamic systems, to initial conditions, in general, was demonstrated in principle in the 1960s by Massachusetts Institute of Technology mathematician, Edward Lorenz, while he was studying a weather model using a computer. Lorenz found significant differences in outcomes using his model, when the initial conditions were changed by a very small amount:

> On a particular day in the winter of 1961, Lorenz wanted to re-examine a sequence of data coming from his model. Instead of restarting the entire run, he decided to save time and restart the run from somewhere in the middle. Using data printouts, he entered the conditions at some point near the middle of the previous run and re-started the model calculation. What he found was very unusual and unexpected. The data from the second run should have exactly matched the data from the first run. While they matched at first, the runs eventually began to diverge dramatically — the second run losing all resemblance to the first within a few "model" months.
>
> BRADLEY, 2010

FIGURE 17.2 Plots of the data from two simulations of weather response over time.
SOURCE: BRADLEY (2010)

Plots of the time-series data from two of Lorenz's weather simulations appear in Figure 17.2.

Lorenz rounded off a variable to three digits after the decimal instead of six, and this resulted in the different values shown in Figure 17.2. While no one knows which specific weather condition Lorenz recorded on the Y axis (it is commonly assumed that time is shown on the X axis), we do know the fluctuations shown on the right-most portion of the Y axis are between extreme values, and thus we see that a tiny perturbation in starting values (measured in units smaller than three decimal places), eventually yielded opposite predictions in the simulated weather. This experiment is the origin of expressions, such as, "a butterfly flaps its wings in Brazil, and it rains in America." Very small changes in initial conditions can result in dramatically different outcomes in complex systems. In fact, this behavior is a defining characteristic of a complex or chaotic system (Gleick, 2008). Obviously, Lorenz's computer program was intended to simulate weather, but because it lacked sufficiently detailed inputs, the model yielded dramatically different outputs depending on very small changes in the inputs — the initial conditions. This example demonstrates how a particular model, in this case a computer program, can be inadequate for simulating a complex system. Likewise, animal models are inadequate for predicting human response to drugs and disease.

Examples of complex systems include cells, humans, non-human animals, ecosystems, economies, ant colonies, social interaction, and the United States electrical grid. For more on biological complex systems, see Ahn et al. (2006), Gell-Mann (1994), Goodwin (2001), Greek (2013c), Greek and Rice (2012), Kitano (2002); Morowitz (2002), Solé and Goodwin (2002), Van Regenmortel (2004a, b), Van Regenmortel and Hull (2002), Vojinovic (2015a, b).

It is not easy to understand complex systems. Consider the following summary of the necessary background for understanding complex systems:

> This introductory textbook is intended for use in a one-semester course to acquaint biomedical engineers, biophysicists, systems physiologists, ecologists, biologists, and other scientists, in general, with complexity and complex systems. I have focused on biochemical, genomic, and physiological complex systems, and I have also introduced the reader to the inherent complexity in economic systems [....] Reader background: Readers should have had college courses in algebra, calculus, ordinary differential equations, and linear algebra, and, hopefully, engineering systems analysis. They should also have had basic college courses in chemistry, biochemistry, cell biology, and ideally even in human physiology and anatomy. This is the broad background that is required in the

> interdisciplinary fields of biomedical engineering, biophysics, systems physiology, and economics.
>
> NORTHROP, 2011, pp. xiii–xvii

Northrop (2011, p. xiii) also notes: "Broadly stated, we consider that complexity is a subjective measure of *the difficulty in describing and modeling a system* (thing or process), and thus being able to predict its behavior" (emphasis added). Again we note the fact that complex systems are difficult to model in terms of being able to predict outcomes to perturbations.

Vicsek (2002, p. 131) states:

> In the past, mankind has learned to understand reality through simplification and analysis. Some important simple systems are successful idealizations or primitive models of particular real situations — for example, a perfect sphere rolling down an absolutely smooth slope in a vacuum. This is the world of Newtonian mechanics, and it ignores a huge number of other, simultaneously acting factors. Although it might sometimes not matter that details such as the motions of the billions of atoms dancing inside the sphere's material are ignored, in other cases reductionism may lead to incorrect conclusions. *In complex systems, we accept that processes that occur simultaneously on different scales or levels are important, and the intricate behaviour of the whole system depends on its units in a nontrivial way. Here, the description of the entire system's behaviour requires a qualitatively new theory, because the laws that describe its behaviour are qualitatively different from those that govern its individual units."* (Emphasis added)

Animal modeling seeks to use one complex system, be it a mouse or a monkey, to predict responses to perturbations that occur at higher levels of organization, of another complex system — a human. To do so ignores the most basic fundamental features of complex systems, discussed above. Given those features, it is outside the realm of science to use one complex system in expectation of its having predictive value for another, when the perturbation affects higher levels of organization.

3 Evolutionary Biology

Informally, evolution can be thought of as small changes in genes (i.e., initial conditions) that occur over long periods of time, resulting in new species with traits different from those of the ancestor organism. In other words, chimpanzees

and humans are both different from the primate that we descended from, and we are different from each other. But the notion that differences among genes can result in new species is separate from the fact that very small differences in genes can also lead to members of the same species reacting quite differently to drugs and disease. Humans and non-human animals are examples of complex systems that have evolved over time – their initial conditions changed in the form of genetic make-up, and these changes affected the organism in a nonlinear fashion, just as we saw in Lorenz's computer model of weather.

Even for two individuals within the same species, small differences in DNA can mean the difference between life and death. A tiny difference of one amino acid within the human chromosome is all that separates a patient with life-threatening sickle cell anemia from those of us who can live free of that condition. Dramatic differences can exist across species without changes in amino acid sequences. Genes are regulated, turned on and off, by other genes. For example, mice and humans share the gene that allows mice to grow a tail (Graham, 2002). The reasons humans do not normally grow a tail during development is that the gene is never turned on (or expressed). Differences in gene regulation and expression vary within and between species and account for differences in response to drugs and disease (Kasowski et al., 2010; Marchetto et al., 2013; Morley et al., 2004; Pritchard et al., 2006; Rifkin, Kim and White, 2003; Rosenberg et al., 2002; Sandberg et al., 2000; Seok et al., 2013; Storey et al., 2007; Suzuki and Nakayama, 2003; Warren et al., 2014; Zhang et al., 2008). So, while it is a fact that humans share a large percentage of their genes with other mammals, this fact is largely immaterial in terms of predicting how humans will respond to perturbations, such as drugs and disease. For example, the progression of HIV to AIDS, which is common in humans, has been very rarely observed in great apes. On the matter of non-human primates, Varki and Altheide (2005, p. 1746) write "[I]t is a striking paradox that chimpanzees are in fact not good models for many major human diseases/conditions".

Based on facts from the theory of evolution and complexity science, there are robust theoretical reasons to conclude that, for all practical purposes, one species will have no predictive value for the response to perturbations that occur at higher levels of organization; and drugs and disease affect higher levels of organization. Note that we are not saying humans and non-human animals cannot ever respond similarly to the same drug or disease. They do in some instances. However, in order for there to be scientific merit in using non-human animals as predictive models for humans, the models would have to have a *high predictive value* as calculated using concepts we discuss in the following section. Consistent with theory, extensive empirical evidence shows that animal models *do not* have high predictive value for human response to drugs and disease, rendering their use in that context unscientific.

4 Empirical Evidence: The Failure of the Animal Model in Terms of Predictive Value for Humans

We now delve into empirical evidence regarding the inability of the animal model to predict human response to drugs and disease. By comparing how well an animal-based test or research method corresponds to human results, we can determine how much predictive value the modality has. Predictive value is measured in science by using the calculations summarized in Table 17.1. In the discussion that follows, we refer to quantities from this table, such as gold standard, false positive, and false negative. Any given test or system can generally be compared to a gold standard, which is the most accurate one available under reasonable conditions.

For example, the gold standard for determining whether a patient has a collapsed lung is a computerized axial tomography (CT) scan of the chest. Even clinically insignificant cases of a collapsed lung can be detected with a CT scan and clinically significant collapses are detected essentially 100% of the time. In reality, patients are assessed with a chest x-ray instead of a CT scan because an x-ray is quicker, easier, and less expensive than a CT scan, and clinically significant collapses are detected by x-ray a very, very high percentage of the time. To determine the predictive value of the chest x-ray, one would perform both diagnostic tests on a group of patients and the calculations in Table 17.1. A positive chest x-ray (an x-ray that revealed a collapsed lung) in light of a positive CT scan would be counted as a true positive (TP) and listed under gold standard positive; while a negative chest x-ray (no collapsed lung) in light of a negative CT scan would be listed as true negative (TN) and listed under gold standard negative. Similarly, a negative x-ray in light of a positive CT scan would be labeled a false negative (FN); and a positive chest x-ray in conjunction with a negative CT scan would be a false positive (FP) (see Nagarsheth and Kurek, 2011, for an example of this).

In the case of evaluating animal models, outcomes in humans would be the gold standard. These same calculations can be performed for any test or modality where a gold standard can be known in contexts within and outside of biomedical science, for example to determine whether a patient has cancer, to determine whether a computer model can predict an outcome in engineering or business, or to determine the predictive value of drug sniffing dogs in airports. For more details see Greek (2014b).

Not all tests or methods need to have a high predictive value to be useful. For example, if you devised a method of winning at the blackjack table more than 50% of the time and bet appropriately each time and played long enough, probabilistically you would beat the house. But in medical science, we need much higher predictive values than 0.5. Even a probability of 0.999 can be

TABLE 17.1 Binary classification test and formulas for determining how well a test or practice compares with the most accurate test available under reasonable conditions.

		Gold Standard:	
		GS+	GS-
Test:	T+	TP	FP
	T-	FN	TN

Calculations:
Sensitivity = TP/(TP+FN)
Specificity = TN/(FP+TN)
Positive Predictive Value (PPV) = TP/(TP+FP) = % of all positives that are true positives
Negative Predictive Value (NPV) = TN/(FN+TN) = % of all negatives that are true negatives

Abbreviations:
T- = Test negative
T+ = Test positive
FP = False positive
TP = True positive
FN = False negative
TN = True negative
GS- = Gold standard negative
GS+ = Gold standard positive

inadequate. Drugs that harm even a very small percentage of patients, even one out of 1,000, have been pulled off the market because of life-threating side effects, such as total liver failure, heart attack, or stroke. Examples of widely-marketed drugs that have been withdrawn due to unanticipated fatalities include Vioxx (rofecoxib), Propulsid (cisapride), and Rezulin (troglitazone). See Graham et al. (2005) and Attarwala (2010) for details on such instances.

So what is an acceptable level of predictive value to expect from animal modeling? To answer this question, first we need to emphasize that acceptable predictive value, like many things in life, varies depending on the context, as the blackjack example illustrates. Consider the case of deeming whether a species exhibits the trait of sentience, which is highly valued in the animal

protection movement as a feature to take into account when considering the ethics of animal modeling. Sentience can be assessed using criteria for which we could attempt to measure predictive value; but, nevertheless, a large gray zone emerges. Chimpanzees are clearly sentient, as are mammals in general. But when we consider invertebrates, the situation becomes less clear. Octopi appear to be both sentient and sapient, but what about sponges, worms, jellyfish, and the common fruit fly? To date we do not have strong evidence that these entities exhibit sentience, but we may simply lack the power to detect sentience in all cases where it exists. Yet, our inability to conclude with certainty that sponges are sentient does not mean we can ignore the fact that chimpanzees do demonstrably exhibit sentience. The precautionary principle should be employed in cases where great suffering is at stake, meaning that our ability to deem a particular species as sentient should not be predicated on the requirement that there exists an assessment method with a predictive value as high as 0.99.

Turning back to the matter at hand, predictive values for responses to drugs in development typically cluster around or below 0.5, which makes them no more useful for prediction than flipping a coin. Predictive values this low are of no use in medical science. When values in the 0.7 to 0.9 range are seen, physicians and medical scientists cannot rely on the results, test, or modality alone, without verifying the item in question with other tests or modalities. To do so would be unethical; the patient deserves greater certainty before proceeding. Science in general relies on consilience, and medical research is not an exception. In this case, when deciding which modality to use, one must consider the mathematics of complex systems and the initial conditions in the form of evolutionary biology. Because animal models are used to make the life-altering decision of whether to take a drug to human trials or to abandon it, even values greater than 0.9 can be deemed inadequate and unacceptably costly in terms of the likelihood of adverse human consequences.

The way around this problem of identifying the *right* predictive value is addressed by Greek and Greek (2004), Greek, Menache and Rice (2012), and Shanks and Greek (2009), and is summarized by Kramer and Greek (2018). The solution involves the use of human-based research and testing through personalized medicine; that is, matching gene(s) to drugs and disease in each patient. Based on the science of complex systems and evolutionary biology, we know categorically that using non-human animal models has unacceptably low predictive value for human responses to drugs and disease. Thus, on balance, the use of animal models in drug development and disease research should be abandoned immediately for the same reasons that society has abandoned wrong or harmful medical practices such as phrenology, bloodletting, and trephination; they were simply ineffective.

We now turn to specific examples of the poor predictive value of animal models, starting with early empirical evidence dating back as early as the 1990s and ending with recent sets of evidence from 2016 that summarize decades of findings.

Data from Suter (1990) and the 11th edition of the *Catalog of Teratogenic Agents* (Shepard and Lemire, 2004) demonstrate the importance of using predictive values. Suter reported on the development of six drugs where humans and non-human animals shared 22 side effects. Suter's data revealed that animal models had a positive predictive value of 0.31. That is, if a side effect was seen in the animal models it had only a 31% chance of being seen in humans for these six drugs. This prediction rate, which is below that expected from a coin toss (heads we abandon the drug because of danger, and tails we continue to develop the drug), illustrates the failure of these animal models as predictors for human response. A naive but common retort to this fact is that if animal models derailed *any* drug that would have harmed humans, it is worth using animal models. The fallacy of this view becomes evident when considering the following assessment of empirical evidence on using animal models to predict human birth defects.

The *Catalog of Teratogenic Agents* lists more than 3,100 agents, of which about 1,500 can produce congenital anomalies (birth defects) in experimental animals but not in humans. These are known as false positives. Furthermore, only about 40 cause birth defects in both humans and non-human animals. These are known as true positives. Based on these numbers and the formulas in Table 17.1, one can calculate a value of 3% for the positive predictive value. A positive predictive value of 3% tells us that for any given birth defect noted in non-human animals, there is only a 3% chance that it will also be seen in humans. A predictive value of 3% is obviously extremely poor but is consistent with the general lack of predictive value in using animal models to determine whether compounds are harmful to developing fetuses (see Greek, Shanks and Rice, 2011, for more on teratogenicity and animal models). This means that for any drug that tests positive for birth defects, when tested for teratogenicity in animal models, there is about a 3% chance that it will harm human babies in utero. Predictive value does not mean that 3% of drugs that would have caused birth defects will be abandoned in development. Instead it means that of 100 drugs tested and shown to harm animal fetuses, about three may harm the human fetus. Unfortunately, we do not know which three. So, abandoning a drug in development based on a test that has a low predictive value does not save babies. Moreover, when human health is involved, low predictive value means anything below 90%–95%; and, often times, even a probability of 99% is inadequate to base treatment on. The predictive value of animal modeling

falls far below 99%; for example, 3% in the above teratogenicity example. For more on this point, see Greek (2013a, b, 2014b), Greek and Greek (2010), and Shanks, Greek, and Greek (2009).

Values this low mean animal modeling per se has, for all practical purposes, *no predictive value* for human response to drugs and disease. Some researchers argue that any predictive value greater than zero means animal models have some predictive value. However, given the scope for serious adverse consequences, including death, the threshold number required in medical science has to be much higher than the typically observed 3% to 55% range of values seen when calculating the predictive value of animal modeling (see previous references); hence the paradigm of animal modeling cannot be justified scientifically in this context. Medical science requires higher predictive values than one needs for winning at the blackjack table.

In our discussion of the predictive value of animal models, we have focused so far on the context of response to drugs. It is also illuminating to consider predictive value in the context of disease research. Scientists are now matching gene response to disease, and great variation is being observed across species. For instance, Seok et al. (2013) studied inflammatory processes, such as sepsis, in mice and humans and found no correlation between what the genes and responses did in mice versus what they did in humans. The following statement, by science journalist Dolgin (2013, p. 118), puts Seok's and colleagues' findings in context: "Yet, despite the fact that some compounds have repeatedly reversed the symptoms of sepsis in animal tests, not a single drug has proven effective in human clinical trials, even though more than 30,000 people have been included in randomized controlled studies, involving candidate antisepsis agents over the past 25 years".

Thus, in searching for a treatment for sepsis, tens of thousands of people were exposed to the risks of a new drug, and billions of dollars were wasted based on animal studies, the results of which proved unrelated to human outcomes. Even more patients were unable to access a potentially effective drug that might have been identified had the resources been dedicated instead to human-based research.

The failure of animal models in these cases appears to be due to differences in gene response between humans and mice (Seok et al., 2013; Warren et al., 2014). Considering that humans and non-human animals are evolved complex systems, there is no reason to expect other diseases or conditions would allow animal models to have high predictive value. Indeed, many diseases have been studied and similarity in responses among species found only at very low rates and usually in retrospect (Enna and Williams, 2009; Hau, 2003; Lin, 1995). (Note that *basic* science research is prone to the same critique. Many researchers now

claim that basic research on non-human animals has high predictive value for humans. See, for example, Devoy et al., 2012; Groenink, Folkerts and Schuurman, 2015; Katzner et al., 2009; National Science Foundation, 2011; Rudczynski, 2011; van Meer, Graham and Schuurman, 2015. Such claims invite the same scrutiny as claims about predictive value in drug development and disease research.)

Based on the track record of drugs that have been tested on non-human animals to date, the poor predictive value of animal models used in preclinical research, and the fact that humans and non-human animals are evolved complex systems, there is every reason to believe yet-to-be-developed drugs identified through the use of animal models will similarly exhibit profoundly different responses in non-human animals versus humans. The exceptions to this rule occur when the perturbation affects levels of organization where the system under analysis is simple or where conserved processes are involved. But even when conserved processes are being studied (e.g., the mechanism for cell replication, the cytochrome P450s, and the presence of various receptors), the outcomes to perturbations to these processes vary among species (Greek and Rice, 2012).

Turning to other medical applications, around 100 vaccines have been shown to be effective against HIV-like viruses in animal models, to date. None have been effective in humans (Bailey, 2008; Editorial, 2007; Gamble and Matthews, 2010). More than a thousand drugs have been seen to protect against nervous system damage in animal models of stroke. Again, none have been protective in humans (Dirnagl, 2006; Dirnagl and Macleod, 2009; Macleod, 2004; O'Collins et al., 2011; O'Collins et al., 2006; Sena et al., 2007). Fouad, Hurd and Magnuson (2013) identify over 10,000 publications modeling spinal cord injury in rats and mice. Many treatments identified in those publications have been effective in non-human animals but failed in humans, and spinal cord injury resulting in paralysis remains incurable in humans.

The predictive value of the above-mentioned medical applications would be roughly zero. In order to prove a test or practice has poor predictive value (as opposed to predictive value numerically equal to zero), one only has to show a relatively small number of failures compared to the successes. The above examples are adequate. Conversely, proving a practice has high predictive value requires examples from a large number of studies. To the best of our knowledge, there are no studies of any kind that show high predictive value of animal models for drugs or disease. Drawing on knowledge from complex systems and the theory of evolution, one can easily infer that the above examples are representative of all animal models and are not exceptions to the rule. Moreover, the studies described above are a small sample of the many such instances that have been recorded in the medical literature showing the

animal model's overall lack of predictive value. For more examples see Arrowsmith (2011a, b), Chiou et al. (2000), Ennever et al. (1987), Fletcher (1978), Grass and Sinko (2002), Hughes (2008), Litchfield (1962), Igarashi et al. (1995, 1996), Johnson et al. (2001), Kola and Landis (2004), Kummar et al. (2007), Lesko and Woodcock (2004), Lumley (1990), Mahmood (2000), Smith and Caldwell (1977), Spriet-Pourra and Auriche (1994), van Meer et al. (2012), and Weaver et al. (2003). Despite the above, important international regulatory bodies still require animal-based research and testing. See for example, the Organisation for Economic Co-operation and Development (2018) and the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (2011).

The overall consequence of continued reliance on animal models is evident when considering the costly failures seen in drug development. For the past few decades, arguably the period when our advanced scientific sophistication should have been yielding the greatest progress in drug development, the success rate in human clinical trials of drugs that entered those trials, based on data from animals, was about 10% (see, e.g., BIO, Biomedictracker and Amplion, 2016; Smietana, Siatkowski and Moller, 2016). Safety/toxicity and efficacy are the two characteristics researchers seek to evaluate when using animal models in drug development. But drugs developed using animal models have systematically failed in human clinical trials for both safety/toxicity reasons and efficacy reasons. Moreover, even more drugs have failed when prescribed to large numbers of people, dropping the success rate below 10%. Granted there are many reasons that drugs fail to enter the market, but these are rare in comparison to the frequency with which efficacy and safety issues have failed to be revealed by animal modeling.

Based on our discussion above of evolved complex systems, evolution, and the empirical data, we conclude that animal models, overall, do not and cannot have a numeric predictive value above about 50%; and, hence, we conclude that, for all practical purposes, they have no predictive value. By this we do not mean the predictive value of any given animal model is exactly equal to zero, but rather that the predictive value is so low that it is necessarily below any reasonable threshold to be considered useful in medical science in general.

5 Summary

Drawing on theoretical principles, based on evolutionary biology and complex systems, and based on extensive empirical evidence, the position that animal modeling has predictive value for human response to drugs in general has

been falsified. TSMT is a theory, and, like all scientific theories, it is consistent with this definition from the National Academies of Science Engineering Medicine (2016): "In everyday usage, *theory* often refers to a hunch or a speculation. When people say, 'I have a theory about why that happened,' they are often drawing a conclusion based on fragmentary or inconclusive evidence. The formal scientific definition of theory is quite different from the everyday meaning of the word. It refers to a comprehensive explanation of some aspect of nature that is supported by a vast body of evidence".

Researchers who aim to improve human outcomes cannot continue to treat humans and non-human animals as simple systems and expect results based on non-human animals to translate to human patients. TSMT is the first comprehensive theory that explains the past failures and apparent successes of animal modeling and also explains why animal models will never achieve predictive value and, thus, should be abandoned.

We acknowledge that the scientific community as a whole is not yet familiar with TSMT; but we are confident that, in time, a consensus will be reached. Kramer and Greek (2018) explain the obstacles that must be overcome to ensure that drug development and the study of diseases are based on sound science. This will require changes to the regulations that currently mandate the use of animal models. Furthermore, Kramer and Greek (2018) discuss modern techniques that fall under the heading of personalized medicine, which offer treatments and cures that are customized to a patient's individual genetic make-up and, hence, sidestep the significant risks associated with the continued blind reliance on methods arising from the use of animal models.

Acknowledgements

We thank Marshall Clemens for allowing us to use his complex systems figure.

References

Ahn, A.C., M. Tewari, C.S. Poon and R.S. Phillips (2006). The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative?. *PLoS Medicine*, 3(6), p. e208.

American Physiological Society (APA) (2017). *Why Do Scientists Use Animals in Research?* [online] Available at: http://www.the-aps.org/mm/sciencepolicy/animalresearch/publications/animals/quest1.html [Accessed 8 June 2017].

Arrowsmith, J. (2011a). Trial Watch: Phase II Failures: 2008–2010. *Nature Reviews Drug Discovery*, 10, pp. 328–329.

Arrowsmith, J. (2011b). Trial Watch: Phase III and Submission Failures: 2008–2010. *Nature Reviews Drug Discovery*, 10, p. 87.

Attarwala, H. (2010). TGN1412: From Discovery to Disaster. *Journal of Young Pharmacists*, 2(3), pp. 332–336.

Aziz-Alaoui, M. and C. Bertelle, eds. (2009). *From system complexity to emergent properties.* Berlin: Springer.

Bailey, J. (2008). An Assessment of the Role of Chimpanzees in AIDS Vaccine Research. *Alternatives to Laboratory Animals*, 36, pp. 381–428.

Bell, J.T. and T.D. Spector (2011). A Twin Approach to Unraveling Epigenetics. *Trends in Genetics*, 27(3), pp. 116–125.

Belmaker, R.H., Y. Bersudsky and G. Agam (2012). Individual Differences and Evidence-based Psychopharmacology. *BMC Medicine*, 10:110.

BIO, Biomedtracker, and Amplion (2016). *Clinical Development Success Rates 2006–2015*. [online] Available at: https://www.bio.org/sites/default/files/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf [Accessed 1 February 2018].

Bradley, L. (2010) *The Butterfly Effect*. [online] Available at: http://www.stsci.edu/~lbradley/seminar/butterfly.html [Accessed 8 June 2017].

Bruder, C.E., A. Piotrowski, A.A. Gijsbers, R. Andersson, S. Erickson, T. Diaz de Ståhl, U. Menzei, J. Sandgren, D. von Tell, A. Poplawski, M. Crowley, C. Crasto, E.C. Partridge, H. Tiwari, D.B. Allison, J. Komorowski, G.J. van Ommen, D.I. Boomsma, N.L. Pedersen, J.T. den Dunnen, K. Wirdefeldt and J.P. Dumanski (2008). Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-number-variation Profiles. *American Journal of Human Genetics*, 82(3), pp. 763–771.

Castillo-Fernandez, J.E., T.D. Spector and J.T. Bell (2014). Epigenetics of Discordant Monozygotic Twins: Implications for Disease. *Genome Medicine*, 6: 60.

Center for Science in the Public Interest (2008). *Longer Tests on Lab Animals Urged for Potential Carcinogens*. [online] Available at: http://www.cspinet.org/new/200811172.html [Accessed 8 June 2017].

Chapman, S.J. and A.V.S. Hill (2012). Human Genetic Susceptibility to Infectious Disease. *Nature Reviews Genetics*, 13, pp. 175–188.

Chiou, W.L., H.Y. Jeong, S.M. Chung and T.C. Wu (2000). Evaluation of Using Dog as an Animal Model to Study the Fraction of Oral Dose Absorbed of 43 Drugs in Humans. *Pharmaceutial Research*, 17(2), pp. 135–140.

Czyz, W., J.M. Morahan, G.C. Ebers and S.V. Ramagopalan (2012). Genetic, Environmental, and Stochastic Factors in Monozygotic Twin Discordance with a Focus on Epigenetic Differences. *BMC Medicine*, 10(93).

Dempster, E.L., R. Pidsley, L.C. Schalkwyk, S. Owens, A. Georgiades, F. Kane, S. Kalidindi, M. Picchioni, E Kravarti. T. Toulopoulou, R.M. Murray and J. Mill (2011). Disease-associated Epigenetic Changes in Monozygotic Twins Discordant for Schizophrenia and Bipolar Disorder. *Human Molecular Genetics*, 20(24), pp. 4786–4796.

Devoy, A., R.K.A. Bunton-Stasyshyn, V.L.J. Tybulewicz, A.J.H. Smith and E.M.C. Fisher (2012). Genomically Humanized Mice: Technologies and Promises. *Nature Reviews Genetics*, 13, pp. 14–20.

Dirnagl, U. (2006). Bench to Bedside: The Quest for Quality in Experimental Stroke Research. *Journal of Cerebral Blood Flow and Metabolism*, 26(12), pp. 1465–1478.

Dirnagl, U. and M.R. Macleod (2009). Stroke Research at a Road Block: The Streets from Adversity Should Be Paved with Meta-analysis and Good Laboratory Practice. *British Journal of Pharmacology*, 157(7), pp. 1154–1156.

Dolgin, E. (2013). Animal Rule for Drug Approval Creates a Jungle of Confusion. *Nature Medicine*, 19(2), pp. 118–119.

Editorial (2007). Cold Shower for AIDS Vaccines. *Nature Medicine*, 13(12), pp. 1389–1390.

Enna, S.J. and M Williams. (2009). Defining the Role of Pharmacology in the Emerging World of Translational Research. *Advances in Pharmacology*, 57, pp. 1–30.

Ennever, F.K., T.J. Noonan and H.S. Rosenkranz (1987). The Predictivity of Animal Bioassays and Short-term Genotoxicity Tests for Carcinogenicity and Non-carcinogenicity to Humans. *Mutagenesis*, 2(2), pp. 73–78.

Fletcher, A.P. (1978). Drug safety tests and subsequent clinical experience. *Journal of the Royal Society of Medicine*, 71(9), pp. 693–696.

Fouad, K., C. Hurd and D.S.K. Magnuson (2013). Functional Testing in Animal Models of Spinal Cord Injury: Not as Straight Forward as One Would Think. *Frontiers in Integrative Neuroscience*, 7, p. 85.

Gamble, L.J. and Q.L. Matthews (2010). Current Progress in the Development of a Prophylactic Vaccine for HIV-1. *Drug Design, Development and Therapy*, 5(9), pp. 9–26.

Gell-Mann, M.. (1994). *The quark and the jaguar: Adventures in the simple and complex*. Boston: Little, Brown and Company.

Giles, J. (2006). Animal Experiments Under Fire for Poor Design. *Nature*, 444(7122), p. 981.

Gleick, J. (2008). *Chaos. Making a new science.* London: Penguin Books.

Goodwin, B. (2001). *How the leopard changed its spots: The evolution of complexity*. Princeton: Princeton University Press.

Graham, D.J., D. Campen, R. Hui, M. Spence, C Cheetham. G. Levy, S. Shoor and W.A. Ray (2005). Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with Cyclo-oxygenase 2 Selective and Non-selective, Non-steroidal Anti-inflammatory Drugs: Nested Case-control Study. *The Lancet*, 365(9458), pp. 475–481.

Graham, S. (2002). Mouse Genome Sequenced. *Scientific American*. [online] Available at: https://www.scientificamerican.com/article/mouse-genome-sequenced/ Accessed 8 June 2017.

Grass, G.M. and P.J. Sinko (2002). Physiologically-based Pharmacokinetic Simulation Modelling. *Advanced Drug Delivery Review*, 54(3), pp. 433–451.

Greek, R. (2013a). Animal models in drug development. In: S. Gowder, ed., *New Insights into Toxicity and Drug Testing.* Manhattan: InTech, pp. 124–152.

Greek, R. (2013b). Animal models of cancer in light of evolutionary biology and complexity science. In: *The Research and Biology of Cancer*. Hong Kong: iConcept Press.

Greek, R. (2013c). Evolved Complex Systems. *Biological Systems: Open Access,* 2, p. e107.

Greek, R. (2014a). A Discussion of the Role of Complex Evolved Systems in the Development of invasive Cardiovascular Interventions as Illustrated by the Blalock-Taussig Shunt and Intra-arterial Stents. *Biological Systems: Open Access*, 3, p. 1.

Greek, R. (2014b). The Ethical Implications for Humans in light of the Poor Predictive Value of Animal Models. *International Journal of Clinical Medicine,* 5(16), pp. 966–1005.

Greek, R. and J. Greek (2004). *What will we do if we don't experiment on animals?*. Victoria: Trafford Publishing.

Greek, R. and J. Greek (2010). Is the Use of Sentient Animals in Basic Research Justifiable?. *Philosophy, Ethics, and Humanities in Medicine*, 5:14.

Greek, R. and L.A. Hansen (2013). Questions Regarding the Predictive Value of One Evolved Complex Adaptive System for a Second: Exemplified by the SOD1 Mouse. *Progress in Biophysics and Molecular Biology*, 113(2), pp. 231–153.

Greek, R., A. Menache and M.J. Rice (2012). Animal Models in an Age of Personalized Medicine. *Personalized Medicine*, 9(1), pp. 47–64.

Greek, R. and M.J. Rice (2012). Animal Models and Conserved Processes. *Theoretical Biology and Medical Modelling*, 9:40. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3533755/pdf/1742-4682-9-40.pdf [Accessed 1 February 2018].

Greek, R., N. Shanks and M.J. Rice (2011). The History and Implications of Testing Thalidomide on Animals. *The Journal of Philosophy, Science, & Law*, 11(3) pp. 1–32.

Groenink, L. (2015). Introduction: European Journal of Pharmacology, Special Issue on Translational Value of Animal Models. *European Journal of Pharmacology*, 759, pp. 1–2.

Hau, J. (2003). Animal models. In: J. Hau and G.L. van Hoosier, Jr., eds., *Handbook of Laboratory Animal Science,* 2nd ed. Vol. II. Boca Raton: CRC Press, pp. 1–9.

Huff, J., M.F. Jacobson and D.L. Davis (2008). The Limits of Two-year Bioassay Exposure Regimens for identifying Chemical Carcinogens. *Environmental Health Perspectives*, 116, (11), pp. 1439–1442.

Hughes, B. (2008). Industry Concern over EU Hepatotoxicity Guidance. *Nature Reviews Drug Discovery*, 7, p. 719.

Igarashi, T., S. Nakane and T. Kitagawa (1995). Predictability of Clinical Adverse Reactions of Drugs by General Pharmacology Studies. *Journal of Toxicological Science*, 20(2), pp. 77–92.

Igarashi, T., T. Yabe and K. Noda (1996). Study Design and Statistical Analysis of Toxicokinetics: A Report of JPMA Investigation of Case Studies. *Journal of Toxicological Science*, 21(5), pp. 497–504.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) (2011). *ICH Harmonised Tripartite Guideline. Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals S6(R1)*. [online] Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S6_R1/Step4/S6_R1_Guideline.pdf [Accessed 1 February 2018].

Johnson, J.I., S. Decker, D. Zaharevitz, L.V. Rubinstein, J.M. Venditti, S. Schepartz, S. Kalyandrug, M. Christian, S. Arbuck, M. Hollingshead and E.A. Sausville (2001). Relationships Between Drug Activity in NCI Preclinical *In Vitro* and *In Vivo* Models and Early Clinical Trials. *British Journal of Cancer*, 84(10), pp. 1424–1431.

Kasowski, M., F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. Waszak, L. Habegger, J. Rozowsky, M. Shi, A.E. Urban, M.Y. Hong, K.J. Karczewski, W. Huber, S.M. Weissman, M.B. Gerstein, J.O. Korbel and M. Snyder (2010). Variation in Transcription Factor Binding Among Humans. *Science*, 328(5975), pp. 232–235.

Katzner, S., I. Nauhaus, A. Benucci, V. Bonin, D.L. Ringach and M. Carandini (2009). Local Origin of Field Potentials in Visual Cortex. *Neuron*, 61(1), pp. 35–41.

Kitano, H. (2002). Systems Biology: A Brief Overview. *Science*, 295(5560), pp. 1662–1664.

Kola, I. and J. Landis (2004). Can the Pharmaceutical Industry Reduce Attrition Rates? *Nature Reviews Drug Discovery*, 3(8), pp. 711–715.

Kramer, L.A. and R. Greek (2018). Human Stakeholders and the Use of Animals in Drug Development. *Business and Society Review,* , 123(1), pp. 3–58.

Kummar, S., R. Kinders, L. Rubinstein, R.E. Parchment, A.J. Murgo, J. Collins, O. Pickeral and J. Low (2007). Compressing Drug Development Timelines in Oncology Using Phase '0' Trials. *Nature Reviews Cancer*, 7(2), pp. 131–139.

LaFollette, H. and N. Shanks (1996). *Brute science: Dilemmas of animal experimentation*. London: Routledge.

LaFollette, H. and N. Shanks (1998). Claude Bernard: The founder of the paradigm. In: M. Bekoff, ed., *The Encyclopedia of Animal Rights*. Westport, CT: Greenwood Publishing, pp. 91–92.

LeCouter, J.E., B. Kablar, P.F. Whyte, C. Ying, M.A. Rudnicki (1998). Strain-dependent Embryonic Lethality in Mice Lacking the Retinoblastoma-related p130 Gene. *Development*, 125, pp. 4669–4679.

Lesko, L.J. and J. Woodcock (2004). Translation of Pharmacogenomics and Pharmacogenetics: A Regulatory Perspective. *Nature Reviews Drug Discovery*, 3(9), pp. 763–769.

Lin, J.H. (1995). Species Similarities and Differences in Pharmacokinetics. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, 23(10), pp. 1008–1021.

Litchfield, J.T., Jr. (1962). Symposium on clinical drug evaluation and human pharmacology. XVI. Evaluation of the safety of new drugs by means of tests in animals. *Clinical Pharmacology and Therapeutics*, 3, pp. 665–672.

Lumley, C. (1990). Clinical toxicity: could it have been predicted? Premarketing experience. In: C. Lumley and S. Walker, eds., *Animal Toxicity Studies: Their Relevance for Man*. London: Quay, pp. 49–56.

Macleod, M. (2004). Systematic Review and Meta-analysis of Experimental Stroke. *International Journal of Neuroprotection and Neuroregeneration* 1, pp. 9–12.

Mahmood, I. (2000). Can Absolute Oral Bioavailability in Humans Be Predicted from Animals? A Comparison of Allometry and Different Indirect Methods. *Drug Metabolism and Drug Interactions*, 16(2), pp. 143–155.

Marchetto, M.C., I. Narvaiza, A.M. Denli, C. Denner, T.A. Lazzarini, J.L. Nathanson, A.C. Paquola, K.N. Desai, R.H. Herai, M.D. Weitzman, G.W. Yeo, A.R. Muotri and F.H. Gage (2013) Differential L1 Regulation in Pluripotent Stem Cells of Humans and Apes. *Nature*, 503(7477), pp. 525–529.

Medical Research Modernization Committee (2006). *A Critical Look at Animal Experimentation*. [online] Available at: http://mrmcmed.org/Critical_Look.pdf [Accessed 6 February 2018].

Morange, M. (2001). A Successful Form for Reductionism. *The Biochemist* 23(6), pp. 37–39.

Morley, M., C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Wewns, R.S. Spielman and V.G. Cheung (2004). Genetic Analysis of Genome-wide Variation in Human Gene Expression. *Nature*, 430(7001), pp. 743–747.

Morowitz, H.J. (2002). *The emergence of everything: How the world became complex*. Oxford: Oxford University Press.

Nagarsheth, K. and S. Kurek (2011). Ultrasound Detection of Pneumothorax Compared with Chest X-ray and Computed Tomography Scan. *The American surgeon*, 77(4), pp. 480–484.

National Academies of Science Engineering Medicine (2016). *Is Evolution a Theory or a Fact?*. [online] Available at: http://www.nas.edu/evolution/TheoryOrFact.html [Accessed 8 June 2017].

National Science Foundation (2011) *Press Release: Of Mice and Men*. [online] Available at: http://www.nsf.gov/news/news_summ.jsp?cntn_id=121653&org=NSF&from=news [Accessed 8 June 2017].

Northrop, R.B. (2011). *Introduction to complexity and complex systems*. Boca Raton: CRC Press.

O'Collins, V.E., M.R. Macleod, S.F. Cox, L. Van Raay, E. Aleksoska, G.A. Donnan and D.W. Howells (2011). Preclinical Drug Evaluation for Combination Therapy in Acute Stroke Using Systematic Review, Meta-analysis, and Subsequent Experimental Testing. *Journal of Cerebral Blood Flow and Metabolism*, 31(3), pp. 962–975.

O'Collins, V.E., M.R. Macleod, G.A. Donnan, L.L. Horky, B.H. van der Worp and D.W. Howells (2006). 1,026 Experimental Treatments in Acute Stroke. *Annals of Neurology*, 59(3), pp. 467–477.

Organisation for Economic Co-operation and Development (2018). *Animal Welfare*. [online]. Available at: http://www.oecd.org/chemicalsafety/testing/animal-welfare.htm [Accessed 1 February 2018].

Pearson, H. (2002). Surviving a Knockout Blow. *Nature*, 415(6867), pp. 8–9.

Pritchard, C., D. Coil, S. Hawley, L Hsu and P.S. Nelson (2006). The Contributions of Normal Variation and Genetic Background to Mammalian Gene Expression. *Genome Biology*, 7:R26.

Raineri, I., E.J. Carlson, R. Gacayan, S. Carra, T.D. Oberley, T.T. Huang and C.J. Epstein (2001). Strain-dependent High-level Expression of a Transgene for Manganese Superoxide Dismutase Is Associated with Growth Retardation and Decreased Fertility. *Free Radical Biology and Medicine*, 31(8), pp. 1018–1030.

Rifkin, S.A., J. Kim and K.P. White (2003). Evolution of Gene Expression in the Drosophila Melanogaster Subgroup. *Nature Genetics*, 33(2), pp. 138–144.

Rosen, R. (1999). *Essays on life itself*. New York: Columbia University Press.

Rosenberg, N.A., J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky and M.W. Feldman (2002). Genetic Structure of Human Populations. *Science*, 298(5602), pp. 2381–2385.

Rudczynski, A.B. (2011). Letter. *New Haven Register*, [online]. Available at: http://www.nhregister.com/article/NH/20110325/NEWS/303259946 [Accessed 8 June 2017].

Sandberg, R., R. Yasuda, D.G. Pankratz, T.A. Carter, J.A. Del Rio, L. Wodicka, M. Mayford, D.J. Lockhart and C. Barlow (2000). Regional and Strain-specific Gene Expression Mapping in the Adult Mouse Brain. *Proceedings of the National Academy of Science of the United States of America,* 97(20), pp. 11038–11043.

Sena, E., H.B. van der Worp, D. Howells and M. Macleod (2007). How Can We Improve the Preclinical Development of Drugs for Stroke? *Trends in Neurosciences*, 30(9), pp. 433–439.

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, C.C. Finnerty, C.M. López, S. Honari, E.E. Moore, J.P. Minei, J. Cuschieri, P.E. Bankey, J.L. Johnson, J. Sperry, A.B. Nathens, T.R. Billiar, M.A. West, M.G. Jeschke, M.B. Klein, R.L. Gamelli, N.S. Gibran, B.H. Brownstein, C. Miller-Graziano, S.E. Calvano, P.H. Mason, J.P. Cobb, L.G. Rahme, S.F. Lowry, R.V. Maier, L.L. Moldawer, D.N. Herndon, R.W. Davis, W. Xiao, R.G. Tompkins and Inflammation and Host Response to Injury, Large Scale Collaborative Research Program (2013). Genomic Responses in Mouse Models Poorly Mimic Human Inflammatory Diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 110(9), pp. 3507–3512.

Shanks, N. and R. Greek (2008). Experimental Use of Non-human Primates Is Not a Simple Problem. *Nature Medicine* 14(10), p. 1012.

Shanks, N. and R. Greek (2009). *Animal models in light of evolution.* Boca Raton: Brown Walker.

Shanks, N., R. Greek and J. Greek (2009). Are Animal Models Predictive for Humans?. *Philosophy, Ethics, and Humanities in Medicine*, 4:2.

Shanks, N. and R.A. Pyles. (2007). Evolution and Medicine: The Long Reach of Dr. Darwin. *Philosophy, Ethics Humanities in Medicine,* 2, p. 4.

Shepard, T. and R. Lemire (2004). *Catalog of teratogenic agents,* 11th ed. Baltimore: Johns Hopkins.

Smietana, K., M. Siatkowski and M. Moller (2016). Trends in Clinical Success Rates. *Nature Reviews Drug Discovery,* 15(6), pp. 379–380.

Smith, R.L. and J. Caldwell (1977). Drug metabolism in non-human primates. In: D.V. Parke and R.L. Smith, eds., *Drug Metabolism. From Microbe to Man.* London: Taylor & Francis, pp. 331–356.

Solé, R. and B. Goodwin (2002). *Signs of life: How complexity pervades biology*. New York: Basic Books.

Spriet-Pourra, C. and M. Auriche (1994). *SCRIP reports,* 2nd ed. New York: PJB Publications.

Storey, J.D., J. Madeoy, J.L. Strout, M. Wurfel, J. Ronald and J.M. Akey (2007). Gene-expression Variation within and among Human Populations. *American Journal of Human Genetics*, 80(3), pp. 502–509.

Suter, K. (1990). What can be learned from case studies? The company approach. In: C. Lumley and S. Walker, eds., *Animal Toxicity Studies: Their Relevance for Man.* Lancaster: Quay, pp. 71–78.

Suzuki, Y. and M. Nakayama (2003). Differential Profiles of Genes Expressed in Neonatal Brain of 129X1/SvJ and C57BL/6J Mice: A Database to Aid in Analyzing DNA Microarrays Using Nonisogenic Gene-targeted Mice. *DNA Research*, 10(6), pp. 263–275.

van Meer, P.J., M.L. Graham and H.J. Schuurman (2015). The Safety, Efficacy, and Regulatory Triangle in Drug Development: Impact for Animal Models and the Use of Animals. *European Journal of Pharmacology*, 759, pp. 3–13.

van Meer, P.J.K., M. Kooijman, C.C. Gispen-de Wied, E.H.M. 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.

Van Regenmortel, M. (2002). Reductionism and the Search for Structure-function Relationships in Antibody Molecules. *Journal of Molecular Recognition*, 15(5), pp. 240–247.

Van Regenmortel, M. (2004a). Biological Complexity Emerges from the Ashes of Genetic Reductionism. *Journal of Molecular Recognition*, 17(3), pp. 145–148.

Van Regenmortel, M. (2004b). Reductionism and Complexity in Molecular Biology. Scientists Now Have the Tools to Unravel Biological Complexity and Overcome the Limitations of Reductionism. *EMBO Reports*, 5(11), pp. 1016–1020.

Van Regenmortel, M.H. and D.L. Hull (2002). *Promises and limits of reductionism in the biomedical sciences* (*Catalysts for fine chemical synthesis*). West Sussex: Wiley.

Varki, A. and T.K. Altheide (2005). Comparing the Human and Chimpanzee Genomes: Searching for Needles in a Haystack. *Genome Research*, 15(12), pp. 1746–1758.

Vicsek, T. (2002). The BIGGER PICTure. *Nature*, 418(6894), p. 131.

Vojinovic, M. (2015a). *Reductionism, Emergence, and Burden of Proof — Part I*. Scientia Salon. [online] Available at: https://scientiasalon.wordpress.com/2015/01/14/reductionism-emergence-and-burden-of-proof-part-i/ [Accessed 8 June 2017].

Vojinovic, M. (2015b). *Reductionism, Emergence, and Burden of Proof — Part II*. Scientia Salon. [online] Available at: https://scientiasalon.wordpress.com/2015/01/16/reductionism-emergence-and-burden-of-proof-part-ii/ [Accessed 8 June 2017].

Warren, H.S., R.G. Tompkins, L.L. Moldawer, J. Seok, W. Xu, M.N. Mindrinos, R.V. Maier, W. Xiao and R.W. Davis (2014). Mice Are Not Men. *Proceedings of the National Academy of Sciences*, 112(4), pp. E347-E348.

Weaver, J.L., D. Staten, J. Swann, G Armstrong, M. Bates and K.L. Hastings (2003). Detection of Systemic Hypersensitivity to Drugs Using Standard Guinea Pig Assays. *Toxicology*, 193(3), pp. 203–217.

Zhang, W., S. Duan, E.O. Kistner, W.K. Bleibel, R.S. Huang, T.A. Clark, T.X. Chen, A.C. Schweitzer, J.E. Blume, N.J. Cox and M.E. Dolan (2008). Evaluation of Genetic Variation Contributing to Differences in Gene Expression Between Populations. *American Journal of Human Genetics*, 82(3), pp. 631–640.