ANNUAL REVIEWS

Annual Review of Biomedical Engineering

Physiological Modeling and Simulation—Validation, Credibility, and Application

W. Andrew Pruett,[1] John S. Clemmer,[1] and Robert L. Hester[1,2]

[1]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA; email: wpruett@umc.edu, jclemmer@umc.edu, rhester@umc.edu

[2]John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA

Annu. Rev. Biomed. Eng. 2020. 22:185–206

The *Annual Review of Biomedical Engineering* is online at bioeng.annualreviews.org

https://doi.org/10.1146/annurev-bioeng-082219-051740

Keywords

simulation, physiological modeling, validation, credibility

Abstract

In this review, we discuss the science of model validation as it applies to physiological modeling. There is widespread disagreement and ambiguity about what constitutes model validity. In areas in which models affect real-world decision-making, including within the clinic, in regulatory science, or in the design and engineering of novel therapeutics, this question is of critical importance. Without an answer, it impairs the usefulness of models and casts a shadow over model credibility in all domains. To address this question, we examine the use of nonmathematical models in physiological research, in medical practice, and in engineering to see how models in other domains are used and accepted. We reflect on historic physiological models and how they have been presented to the scientific community. Finally, we look at various validation frameworks that have been proposed as potential solutions during the past decade.

Contents

INTRODUCTION

In a search of PubMed in July 2019 using the terms "mathematical model" and "human physiology validation," 367 models dating from 1974 were obtained. Of these, 197 had fewer than 3 citations, with 79 of these having no citations. This search yielded an overall 20% noncitation rate versus the estimated 10% for all science papers (1). While these search terms are not exhaustive, they are representative. Mathematical modeling papers in physiology are rare and are rarely used by other scientists. The papers from this search that were most cited fell into two general groups: clinical papers with a modeling component, such as a nomogram or algorithmic control of a device, and papers with predictions of population observations, such as growth models. Assuming that modelers of biological systems cover topics that are similarly as interesting as experimental and clinical studies, this suggests that a lack of model credibility is to blame for the overall disuse of mathematical modeling in physiological research.

In this review, we survey the various notions of model validity in the physiology literature. Validity is not a straightforward concept, and, currently, there is no simple rubric to follow to ensure that a model is a valid representation of a biological system. Although several physiological models have been validated on an ad hoc, empirical basis by their authors, there is little information in the literature about how a model should be validated. This deficiency is likely due to the origins of modeling itself, which arises as a discipline at the interface of the classical subject of physiology, biochemistry, physics, and the practice of medicine itself. Empirical agreement—that is, the agreement of the model with a collection of data—is an important part of model validation, but it is not sufficient to ensure that a model is an accurate or trustworthy representation of a system (2).

Without a formal definition or even common agreement about what constitutes model validation beyond the empirical, there is a necessity for a deeper understanding of what models are and how they can complement basic and clinical research and generate scientific value. Given a better appreciation for how they are used, we can find better ways of describing whether they are good tools for these purposes. To do this, we consider what a scientific model is and compare the development and defense of mathematical models with animal and cellular models. Currently, cellular and animal models drive basic research, accounting for most of the evidence used to advance physiological knowledge, and they provide the major translational link to clinical applications and novel therapies in humans. In 2010 alone, there were approximately 60,000 publications in PubMed that

used mouse models (3). Investigation of validation modalities in this domain can provide an understanding of what the larger community recognizes as valid and aid in identifying gaps in the ways mathematical models are described, developed, and tested. After gaining a more solid appreciation for what constitutes a good animal model, we can begin to address the problem of validation in mathematical models and increase the use of credible models in basic and clinical research.

MODELS

Kaizer et al. (4, p. 211) define a model to be "a representation of a system, entity, phenomenon, or process." This broad definition can be separated into three essential subclasses, two of which are widely used by most biological scientists and one that is less often used. These are (*a*) conceptual models, (*b*) physical models, and (*c*) mathematical models (5).

A conceptual model is a heuristic representation of a system that allows one to draw qualitative conclusions. A good conceptual model (*a*) is consistent with both itself and a set of observations and (*b*) allows one to make testable hypotheses that extend beyond those observations. An example of a conceptual model is Harvey's 1628 description of the circulation as a loop in which blood moves from the lungs to the heart to peripheral arteries and through pores to the peripheral veins before returning to the right heart and lungs (6). This model obviously simplifies reality. It does not include capillaries (which had not been observed and would not be until the eighteenth century) or basic aspects of circulatory physiology, such as local control of blood flow and the mechanics of the vasculature and heart. However, it gives a coherent description of how blood might transport substances around the body and allows for the development of a causal understanding of how the organs integrate their activities into a whole, all of which went against the well-known physiological notions in the seventeenth century. This heuristic propelled physiology forward by providing a meaningful, self-consistent framework on which to build new hypotheses. Conceptual models can vary in scope and substance, from Harvey's model of circulation to Freudian psychology or an internist's clinical model of disease, which is founded on years of experience and training but not subject to mathematical precision. Conceptual models are common in biological science. They appear as working hypotheses—that is, frameworks for understanding the interactions in complex systems—informing a belief in the response to a stimulus. They are generally not assessed for validity, or they are used as springboards to a more qualitative assessment of systems.

A physical model is a physical representation of a system that is used to investigate the properties of the system under question. The representation is not the whole system or process, despite Rosenblueth & Wiener's (7, p. 320) proclamation that "the only material model of a cat is another, or preferably the same, cat." Examples of physical models range from prototypes of mechanical components used to test the robustness of a load to three-dimensional (3-D) models of uniquely shaped hearts or vasculature used by surgeons to plan a surgery (8, 9) to a model of the Mississippi River used to assess the impact on flooding of particular placements of dikes and dams (10). Physical models trade realism for representation: A printed 3-D heart, for instance, lacks a realistic substance when compared with a real heart, and the model resolution is limited by the printer that creates it. Nevertheless, such models have been shown to be a useful technology, altering the course of surgical plans in half of all cases in a multicenter study (8). In the context of biology, the most important physical models are the cell lines and animal models that comprise the vast majority of basic research. Animal models are discussed in more detail later in this review.

Mathematical models have been more variously defined. At its essence, a mathematical model uses abstract logical systems, rather than physical manifestations, to represent a system. In engineering, a mathematical model can be used to increase the ability to understand, predict, or control a system's behavior (11). The US Department of Defense removes the necessity of clear purpose,

but demands that a model be a "mathematical, or otherwise logical representation," of the system at hand (12, p. 10). Examples of mathematical models in the biomedical domain include systems of ordinary differential equations as 0-dimensional (lumped parameter) models, 1- to 3-D models of blood flow through vessels of regular or irregular shape, statistical models of disease based on demographic or genetic data, models of disease propagation, and stress–strain models of mechanical anatomical components. These models have been associated with different purposes for different stakeholders, complicating the answer to the question, What makes a model good? This review attempts to show how these various purposes are just facets of a single purpose and that the framework of verification, validation, and uncertainty quantification utilized in regulatory science can serve as a scaffolding on which to build a unified theory of mathematical model validation for physiology, bioengineering, pharmacology, and other biological sciences.

REDUCTIONISM—WHY DO WE NEED MODELS?

Reductionism is the idea that complex concepts can be divided into understandable units and that by understanding the units, we understand the whole. In many ways, this has been the philosophy of scientific enquiry since the Renaissance. The very idea of a controlled experiment comes from the reductionist hypothesis. Mathematical modeling is way of applying reductionist thinking: By reducing complexity to a series of equations that each agree with observations individually, and by discarding the parts of the system that do not directly affect the outcome under consideration, biological mechanisms may be identified.

Physics has advanced in no small part due to the successful marriage between mathematics and experiment. Since Isaac Newton, physics has been framed as a mathematical science, in which theory is ensconced in equations and experiment seeks to parameterize those equations in simple cases or to test complex assemblies of hypotheses in more general cases. Because physics-based systems can be specified and controlled, a high degree of precision is attainable. Discrepancies between theory and mathematics and experiment are easily quantified and communicated. Theory itself becomes a sort of experiment, in which hypothetical interactions can be proposed, framed as mathematics, and tested before physical experiments are planned or performed. In this way, the research process itself is streamlined and focused on only the most credible ideas. For instance, the history of astrophysics can be summarized as a progression of models of the universe, each explaining new phenomena that were unexplainable under previous theories, that remain consistent with existing data. Physics, however, can be thought of as the study of forces, each acting in a predictable way. The interactions of these forces may be complex, but the forces themselves obey well-defined laws.

Physics-based systems allow reductionism to be applied with great predictability. Because the underlying rules can be easily described, one can calculate the defect that arises between applying a simplified model of a theory and applying a more complex one. For instance, given a question involving gravity, one can calculate the difference between using Newtonian theory and Einsteinian theory. The user can determine whether the simpler theory provides enough accuracy to perform the task at hand and can also determine the effect of model choice on predictions. Hence, reductionism is a choice, not a necessity.

Systems biology and physiology study the mechanisms by which hundreds of agents and actions coordinate to produce biological results. Most often associated with molecular biology, systems biology was defined by Breitling (13, p. 4). as "the comprehensive study of molecular diversity. . . , the identification of simplifying principles and patterns. . .in living and engineered systems, and the integration of this biological knowledge into complex models (of the regulatory networks) that characterize life." According to Noble, systems biology is the consequence of the realization that

biological function arises from networks of interactions between cellular regulatory networks, the environment, and the phenotype of the individual (14). The same realization has been applied more coarsely to physiological systems for decades, recognizing biological function as an emergent property of the communication between organs and tissues. Models offer a mechanism for assembling complex systems from well-understood components in a way that allows assumptions underlying the assembly itself to be tested. In this way, only models can test hypotheses about the properties that arise from interactions between system components.

INFERENCE FROM MODELS—WHAT MAKES A MODEL USEFUL?

In the context of biology, the use of a physical model (a cellular or animal model of disease or pathology) is common, and its use merits exploration as we discuss factors that make a model useful. Biologists have developed cellular models of disease processes by adapting existing cell lines such as HeLa (15), creating appropriate cells from stem cell lines (16, 17), developing animal lines with known susceptibility to diseases such as lupus (18) or obesity and diabetes (19), or by creating an environment through which an animal assumes a pathological phenotype, exemplified by diet (20) or the nonhuman primate addiction models (21–23). In all of these cases, the goal of the model is to replicate some aspect of pathology, behavior, or systemic response that can then be tested to make inferences about physiology or pathophysiology. The "some aspect" part of this is key; the similarity and applicability of the model to a pathology or behavior may not be of the utmost importance.

For example, the obese Zucker rat is a model of obesity and hypertension due to a missense mutation in the long-chain leptin receptor (19). Obese Zucker rats have been used as models to understand the mechanisms underlying the link between hypertension and obesity in several ways: in the consumption of high- and low-fat diets (24, 25), by focusing on renal sympathetic nerve activity that drives hypertension (26, 27), and through many other mechanistic interactions with these symptoms. Overall, as of July 2019, 2,098 papers had appeared in PubMed with "obese Zucker" in the title or abstract. Yet the obese Zucker rat model is not a perfect model of human obesity. Leptin receptor mutations are rare, affecting only 2% to 3% of individuals with severe early onset obesity (28–30). Yet the model is still regarded as a good model of metabolic syndrome (31).

This illustrates one of the primary drivers of interest for modeling: Models are tools to gain an understanding of the mechanisms underlying a disease or therapy in complex, redundant, nonlinear systems, which, hopefully, can give insight into a clinical condition. A mechanistic explanation is one in which the behavior of the whole can be attributed to the actions and interactions of its parts (32). In physiology, this is a common refrain echoing through the work of Guyton, Coleman, and their colleagues (33–36), as well as being voiced by their chief critics (37, 38). In fact, much of the history of physiological modeling, both animal and mathematical, is consumed with trying to use models to provide evidence supporting a mechanism's importance in some observation. Mechanistic explanation is the goal because this level of understanding allows potential interventions to be specified.

Biologists have begun to appreciate that reductionism will not answer all questions of mechanistic explanation in complex systems (39). While it is useful in sciences with lesser complexity, such as physics, biological systems have five fundamental properties that weaken the power of reductionism.

1. Biological variation is widespread and persistent: "In any evolutionary system. . .diversity and complexity will increase on average" (40, p. 1102; see also 41, p. 3). Variation is of

profound clinical relevance. Even among people of the same sex, age, race, and renal function, blood pressure responses to a single antihypertensive drug can be highly variable (42).

2. Biological systems are relentlessly nonlinear. While nonlinear interactions in biology are widely understood and explainable, nonlinear dynamics as a discipline is ill-equipped to analyze systems as complex and large as cellular-level biology, much less organismal- or human-level biology.
3. Biological systems are redundant. Each species is a competition to find robust ways to maintain life in the face of multiple challenges. Many physiological or pathophysiological states are the result of multiple mechanisms supporting an observable response. Redundancy makes it difficult for a researcher or clinician to identify important causal mechanisms—that is, in physiology, if one system fails, another system can compensate and mask any change in the phenotype.
4. Biology consists of multiple systems interacting across different time and spatial scales. The importance of an observation often depends on its timing. For instance, an individual's blood pressure response to a stimuli can depend on a combination of fast-acting neural mechanisms, slow-acting hormonal responses, and long-term effects of the regulation of body fluid volume. The importance that one scientist or another places on these mechanisms depends on when they conduct their measurements.
5. Biological properties are emergent—that is, they come from the complex interactions of subsystems—and, therefore, cannot be simply described as the sum of their inputs from each single component (43).

Knowledge of complex biological systems is gained by studying the intact system and by inferring knowledge about the whole after studying the system's components separately. The factors above, especially the principles of redundancy and nonlinearity, suggest that reducing biological systems to single components, while an important building block in the study of complex systems, is not enough to guarantee an understanding of that system. Some method of integrating the components into an edifice that represents the real system is required to bridge the gap between agents and actions and the functions they elicit (44).

Biological mathematical models allow knowledge and hypotheses to be integrated into a whole, with the goal of testing the consistency of the hypotheses against data. This extends beyond the importance of a single mechanism, and it allows modelers to construct, analyze, and understand the emergent properties of systems. Models are often used to create data in an area where data are sparse or unmeasurable, subject to a collection of hypotheses that can be supported through experimental evidence (45, 46). The mathematical framework is a natural vehicle for modeling redundancy and nonlinearity and for incorporating natural variation. With modern computational methods, multiscale models are within reach (47–50), allowing complex responses to develop over broad time domains. Hence, mathematical models can address all five basic challenges inherent in biological systems. While molecular biology, systems biology, and functional genetics have not yielded significant innovations in clinical diagnosis, therapeutics, or even basic science (51–53), modeling, with appropriate methods for validation, may be the tool that will allow translational knowledge to be synthesized from basic science.

PAST AND PRESENT MATHEMATICAL MODELING EFFORTS—EXAMPLES FROM PHYSIOLOGY

Mathematical modeling has been used to study and describe physiological responses since Harvey estimated cardiac output in *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* in 1628

(6) as a means of defending his general circulatory model. Other early examples of physiological modeling include Krogh's (54) model of oxygen flux in skeletal muscle in 1919 and Hodgkin & Huxley's (55) computational analysis of neuronal membrane potential in 1952. Glomerular function was first quantitatively analyzed by Starling (56) in 1899, and with the advancement of new micropuncture techniques, this led to the first attempts at modeling renal and glomerular function (57, 58).

Mathematical modeling has given insights into complex interactive mechanisms in whole organisms as well, for example the *lac* operon in *Escherichia coli*. This suite of genetic cues has been deeply studied from a variety of perspectives, all giving insight into how this single system processes information from the bacterium's environment and converts it into a metabolic path forward (59–62). This gives insight into how systems-based modeling returns value that cannot be retrieved from reductionism. In these models, the interactions between coefficients and a system's behavior are studied intently, with the variance in a system's behavior attributable not to an interaction but to all of the interactions operating as a system. This behavior cannot be studied by investigating component parts; instead, it must be studied in the system as a whole.

In 1972, Guyton et al. (35) published a model of integrative human physiology that challenged the existing views of blood pressure control. When published, it was widely believed that total peripheral resistance controlled blood pressure while the heart controlled cardiac output. This concept was revolutionized by the concepts of pressure natriuresis and blood flow autoregulation. In 1983, with the rapid advancement of the computer, Coleman & Randall (63) developed a model of human physiology called HUMAN, an extension of the 1972 Guyton et al. model. In the 2000s, this was expanded into a Windows software package called Quantitative Circulatory Physiology (**http://hummod.org**). This model uses several hundred mathematical functions to describe cardiovascular, renal, neural, respiratory, endocrine, and metabolic relationships across multiple organ systems. It is freely downloadable (64; see also **http://hummod.org/hummod-baro.zip**).

Several groups around the world are creating environments for developing integrative models of human physiology. The International Union of Physiological Sciences' (IUPS) Physiome Project is a worldwide effort to develop databases and models to aid in understanding physiological responses. This group states that its mission is "the databasing of physiological, pharmacological, and pathological information on humans and other organisms and integration through computational modeling" (**https://www.physiome.org/About/index.html**). It comprises databases and software for developing computational models of cellular functions (**http://models.cellml.org/**). Some of these models are based on the 1972 Guyton et al. (35) model of integrative physiology. The major limitation of the IUPS Physiome Project is that there is no integration of the submodels into a whole-body or integrative model of human physiology. The National Simulation Resource Physiome Project at the University of Washington is affiliated with the IUPS Physiome Project. Their submodels of physiology are written using JSim, a Java-based simulation system. There are more than 250 separate models, but similar to the IUPS project, there is no integration. SimTK is a free project-hosting platform that focuses on models of cellular dynamics and protein folding, musculoskeletal and neuromuscular dynamics, drug–target dynamics, and neural–prosthetic interactions (**https://simtk.org/**). This group also has cardiovascular projects, but that work focuses mainly on finite element meshing and computational fluid dynamics.

Quantitative Circulatory Physiology has been expanded into HumMod by the Center for Computational Medicine and the Department of Physiology at the University of Mississippi Medical Center. HumMod is a large, multiscale model of human physiology that integrates multiple physiological systems, including the renal, autonomic, endocrine, and cardiovascular systems. With considerable content added to the backbone of the model developed by Guyton et al. (35), published more than 45 years ago, HumMod is composed of more than 8,000 independent

variables and ~2,000 parameters and equations (**http://hummod.org/hummod-baro.zip**). The development of HumMod has led to accurate predictions in numerous areas of physiology, including glucose homeostasis (65), insulin resistance (66), salt-induced hypertension (67), fluid homeostasis (68), the effects of spaceflight on the cardiovascular system (69), cardiovascular responses to hemorrhage (70), and the cardiovascular, neural, and hormonal responses to baroreflex stimulation (71).

VALIDATION—INTUITIVE BUT ILL-DEFINED

In this section, we summarize the idea of model validation, which turns a model's value from potential to actual. The ultimate goal of a model is prediction. The word predict has a specific usage here, and it is integrally tied to the word hypothesis (72). Scientific hypotheses are tested by comparing what is expected with what actually happens. If expectations and actuality agree, we presume that some level of understanding has been reached, and a model's behavior serves as a prediction of the real system. Considering a model as a component of a scientific hypothesis of human disease, actuality is what happens in humans, while the model conveys the expectations. An often-missed part of model development is that the model itself is the hypothetical framework on which prediction rests, and so model construction is part of the process to determine whether a model is predictive. A successful predictive model can then be a tool for explanation. The model's components can be tested for their contributions to an effect, and the powers of redundancy and emergence can be measured. In order to extract this value from models, we need a concrete understanding of what makes a model predictive.

Ideally, validation is the process by which the actions of a model are proved to accurately represent the system under consideration or predict its response to perturbation. This is a misleading definition, as it hides many pitfalls that can confuse modelers and end users. Chief among these pitfalls is the question, What does it mean to accurately represent a system? It is not clear whether a model should simulate all aspects of the system, including all known components and their interactions, or whether coherence on 1, 2, or 100 of the variables is enough. It is unclear as well whether a model should duplicate the mechanisms comprising the system or whether simple agreement with an end point through one or more system perturbations is sufficient for validation. This allows different users to interpret the concept of validation differently, which reduces the overall effectiveness of modeling as a scientific enterprise.

In physiological models (mathematical or otherwise), validation is typically empirical. In empirical validation, one selects a collection of experimental end points and compares them with the model's outputs. Comparison can be qualitative or quantitative, steady-state or transient (43). For example, systemic lupus erythematosus (SLE) is an autoimmune disease believed to be caused by a combination of environmental factors and genetic variation in multiple alleles, each with minor functional contribution, resulting into a multifactorial disease (18, 73, 74). Mouse models of SLE were originally confirmed as models of the disease by noting a shortened life span, renal disease, increased prevalence in females, and peripheral vascular disease, which together resemble the phenotype of SLE in humans (75, 76). Some, but not all, of the genetic variation seen in humans with SLE has been observed in the mouse model, further confirming it as a model of the disease, but the original assertion of validity rested on a few empirical agreements in the simple end points mentioned previously.

If animal models of diabetes exhibit the blood glucose characteristics of human diabetes, including changes in fasting glucose levels and perturbations in the oral glucose tolerance test, is this enough to confirm that the model is a valid representation of human diabetes? We consider the systemic response to glucose infusion as the primary intervention. If the model does not exhibit

the comorbidity of diabetic nephritis that appears in ~40% of human diabetics, is the model still valid? Conversely, if diabetic nephritis appears 100% of the time in a certain animal model, can we say the model is valid? In animal research, there is no objective, quantitative discrimination between these situations, and models that display one or more levels of similarity with the human disease might be accorded more or less validity by different scientists in this domain, depending on their particular area of interest in a disease. If a drug is shown to meet a clinical end point, such as improving renal function, in an animal model that always develops diabetic nephritis, that does not mean it will meet that same end point in an animal model that does not. This suggests that models need to be attuned to a purpose in order for the question of validity to even make sense. This restriction is in line with the assumption that a predictive model remains true in similar situations (72).

In recent years, different schemes for assessing the validity of animal models have been proposed. There is basic recognition that a disease model should share some mechanisms with the human version of the disease. For instance, one definition states that an animal model is considered valid if it "resembles the human condition in etiology, pathophysiology, symptomology, and response to therapeutic interventions" (77, p. 956). This definition borrows from the philosophy-of-science literature, in which the standard of animal model validity is suggested to rest on three factors: face validity (i.e., manifestation of symptoms similar to clinical presentation), construct validity (i.e., similar underlying biological mechanism), and predictive validity (i.e., similar response to intervention) (78).

This intuitive framework again masks several underlying problems. For diseases without a clear etiology, construct validity is an impossible standard to meet. Even in cases in which the natural history of the disease process is clear, replicating it in a laboratory may not be feasible for economic or time reasons. Face validity has been called into question as a reproducible and well-defined construct (79). Predictive validity, the raison d'être of using animal models as tools for the preclinical evaluation of an intervention, is rarely satisfied also. For example, this is evident in the poor translation rate of new chemical entities in clinical trials for several diseases, including stroke and cancer (80–82). Numerous suggestions have been made about how to increase the predictive validity of animal models. One suggestion was to extend the modeling of pharmacokinetic and pharmacodynamic relationships because of their potentially profound effects on dosing assumptions in humans (83). Another was to develop more and better biomarkers of disease (84–86). However, these proposals are aimed at specific marginal problems in translation as opposed to being a more general appraisal of whether current validation schemes ensure that so-called valid models are actually predictive.

FAILURE OF VALIDITY—CLINICAL TRANSLATION

Animal models are one of the major sources of knowledge leading to the development of new pharmacological or device-based therapies and are key for regulatory evaluations (87, 88). The scientific use of animal models dates back to a period when they were the only complex system available for studying integrative biology. Unfortunately, during the past decades, numerous potential drug targets have been discovered in animal models that have failed to materialize into effective human therapies (89, 90). These failures, along with the enormous costs of developing novel therapies into marketable products, led to an investigation into this process by the US Food and Drug Administration (FDA) (91). While multiple systemic features play a role in the failure rate and high cost of innovation, one of the chief points of failure is the clinical translation from animal models (91). The translation step has been deeply studied during the past 10 years, as it exists as a scientific, as opposed to a policy or regulatory, impediment to progress (92–97). These

findings are especially relevant to mathematical models because they expose potential failures in translating these models to the clinical domain as well. For a model to be credible, it must address potential failures in modes of translation and quantify or qualify the risk associated with each.

The translational problems in animal models consist of failures of either intrinsic or extrinsic validity (98). Intrinsic validity speaks to the clinical trial methodology itself. End points of clinical trials are often derived from previous animal or cellular studies. These include early estimates of effect size, the end points used in the trial, and the inclusion and exclusion criteria for the trial. Together, these choices can bias a cohort that may misinform a statistical model used to explain treatment effects. These methodological problems can be reduced with a better use of modeling. Extrinsic validation problems exist because animal models are, in fact, models of human disease and not the disease itself. For example, animals are induced into a disease state via pharmaceuticals, surgical or genetic modification, or environmental intervention at ages and in ways that differ from the natural history of the disease in humans (85, 98). This is not by chance: If we understood the natural mechanistic history of the pathologies that interest us, there would be no need to study them. But by establishing a pathology outside of its natural setting, there is a risk of creating something with the same phenotype but with a different root cause and different response to intervention than in the human presentation.

Here, animal models vividly illustrate the problem of the model of the possible as opposed to the model of the actual (99). An animal model is built on assumptions (e.g., a genetic background, a diet, an initial intervention that begins the progression of the disease, or normality). The model represents one possible way in which the symptoms of a disease may come into being. It is not surprising to find an animal model that exhibits a specific phenotype and set of responses to a limited group of interventions that match those seen in a cohort of similarly diseased or normal humans. This does not imply that the animal model is a good model of the cohort, neither does it imply the opposite. The model may coincide with the human cohort in some way; for instance, it may have etiological, histological, or other parallels with the human disease. Each such similarity reduces the range of possibilities and adds additional evidence that the animal model is a good representation of the human disease, but more is necessary to be predictive of a specific cohort. Conversely, differences between a model and its disease increase the opportunity for systemic differences to develop and for the model to lose predictive capability. Consider the genetic models of rodent obesity. Broadly used monogenic models of obesity, such as the *ob/ob*, *db/db*, and *s/s* mouse models, and the obese Zucker rat, Koletsky rat, Zucker diabetic fatty rat, and Wistar Kyoto fatty rat models, are acceptable, based on the existence of hyperphagia, decreased energy expenditure, and the potential for hyperglycemia and insulin resistance at some point in their lives, all of which mirror obesity in humans. Each model brings a different facet of obesity into view. The models themselves are based on clear assumptions: a specific genetic defect that echoes throughout the organism to create a phenotype. The single defect provides an underlying logic to the model that allows for explanation of the phenotype, but it does not match the heterogeneous disease seen in most obese humans (29, 30). Because of this, a therapy that reduces obesity in one of these cases would not be guaranteed to be effective in humans. However, the use of one of these models to investigate the effects of obesity-derived inflammation might be completely predictive of human inflammation; this linkage would depend on the similarity between the model's inflammatory profile and that seen in humans. This would require a careful analysis of the effects of the natural history of obesity and the effects on inflammation. The models are useful but not necessarily translatable. The usefulness depends on matching the construction of the model with the question being asked of it. An example of a well-constructed animal model is the animal model of diabetes that results from pharmacologically inducing the loss of insulin secretion by destroying pancreatic β-cells with streptozotocin. This animal model represents human type 1 diabetes by eliminating a

particular cell type, and the animals are used to study the downstream effects of long-term insulin deficiency. This is an excellent model of the disease because it acts along the same mechanism as the human disease (100). A pharmaceutical that affects end points directly related to the disease in the animal model is more likely to perform well in humans because of the shared similarity.

Perhaps the appropriate filter under which to consider the failure of models in translation comes from the Duhem thesis. Duhem (101, p. 203) stated that,

> The [scientist] can never submit an isolated hypothesis to the controls of experiment, but only a whole set of hypotheses; when the experiment disagrees with the [scientist's] prediction, he learns that at least one of the hypotheses that constitute the set is erroneous, but the experiment does not indicate which.

If a model fails to replicate an experimental observation, as models are made of components, then one of the components must be wrong, but we do not know which one. The negative form of the Duhem thesis stands inspection as well: If a model replicates an experimental observation, one does not know which model components are responsible for this replication. In a state of ignorance, a model may replicate all of some set of observations but not generalize in any way. This observation is true regardless of the model context: cellular, animal, mathematical, or even a clinical trial, which is a model of a population. For example, a mathematical model could be generated by interpolating polynomials to fit a data set without forcing them to conform to any logic underlying the system. Every data set underdetermines an infinite number of models that can generate that data set.

To assert predictive capability, more information is necessary. The model, again mathematical or animal, is more than the phenotype: It is must represent the human cohort as a whole (2). The model is a combination of assumptions, each of which must be present and none of which can be separated. In the domain of mathematical models, this is like the process of equation selection and the calibration of model coefficients. The coefficients are a function of the equation forms chosen as much as they are of the data used for calibration.

CREDIBILITY—VALIDATION WITH PURPOSE

In the examples above, the primary concern is whether a model is an accurate representation of the human system and whether it makes high-quality predictions about the system. This is a value statement rather than an objective declaration. Different users can envisage different utilities for a model, and their evaluations of the model as a tool for these utilities may well be different. Recognizing this subjectivity explicitly leads to another conception of model quality and one that is arguably better for all users. This is the concept of model credibility. Credibility has been defined as the willingness of individuals to base decisions on information obtained from models (102, 103). Credibility associates qualities of a model with qualities of the system being considered. It moves the focus from asking whether it is a good representation to asking whether it gives actionable evidence. Credibility is more than empirical validation: it is confidence in model-based decisions (104). For clinical utility, this is a high bar. Various rationales can be used to argue for the credibility of a computational model. These include:

1. Evidence relating to the validity of assumptions underlying the equations governing a model. The governing equations here are considered as mathematical forms combined with coefficients. The equation forms (e.g., polynomial, exponential, decay equation, logistic function) relate to the theory underlying the interaction being modeled (e.g., physics, chemistry), while the coefficients are concerned with particular experiments or clinical data.

This evidence exists before the model is constructed and is obtained ex post facto from the experimental record;

2. Calibration evidence. Calibration is the process of fitting a model to match clinical or experimental data. Model coefficients that cannot be directly measured must be calculated from model states that are measurable. This process of fitting coefficients may yield evidence of a model's credibility—for example, a correlation between coefficients that are believed to be related—or the discovery that a particular response can be obtained from a hypothesized system;
3. Validation evidence. This refers to the comparison of a model's predictions with real world data not used in calibration (105).

Determining credibility requires a complete analysis of the problem the model is asked to solve, the assumptions underlying the model, the methods used to integrate those assumptions, the effect of the modeler's choices on the model's outcomes, the quality of the data used to build and test the model, the tools used to transform the assumptions into simulation form, the means by which comparisons with data are made, and the choice of which data are used as comparators. This holistic view tests not just the end points but also every aspect of the model to determine where it is inadequate and where it is strong.

A more general quantification of model validity has been proposed using a credibility–risk matrix (102). Biology and engineering share significant overlap when man-made products interact with human biology. While engineering has used modeling and simulation as parts of its design process at every step in a product's life cycle (e.g., design, deployment, monitoring, and redesign), a similar process-oriented perspective may benefit biological simulation (106). The most general parts of the schema are an assessment of the risk of making a bad decision because of information obtained through simulation and an assessment of the credibility of the model against a range of properties, including the amount of data available, the epistemic foundations of the model, and the assumptions made by the modeler. Altogether, the credibility of the model increases with the amount of data available to support it (107). The most modern approaches to model validation in biology, biological engineering, and regulatory science take a broad view of validation as it pertains to clinical applications. We discuss the validation frameworks of animal and mathematical models because the two inform and build off of one another.

Animal Model Credibility

One approach for increasing the predictive validity of animal models has been to develop standards, such as the ARRIVE (Animal Research Reporting of In Vivo Experiments) guidelines (108), to improve the reporting quality of preclinical data that are analogous to the CONSORT (Consolidated Standards of Reporting Trials) guidelines for clinical trial data. These guidelines assume that with better data governance, a more comprehensive evaluation of preclinical data can be performed and better predictions obtained (109). The ARRIVE guidelines rest on extensive abstracting of research studies to encapsulate a study's goals, motivations, assumptions, and methodologies. While these standards increase the amount of information available to help understand why a model may or may not be predictively valid, the information they retain is only descriptive. Such standards facilitate meta-analyses and, thereby, increase the evidentiary quality of the data (110). These standards merely facilitate analysis, however, without providing better evidence of links between the model and its human presentation.

Another example, the Framework to Identify Models of Disease rubric, was recently developed to categorize and quantify the validity of animal models by surveying a snapshot of the animal

model literature and analyzing the validation methodologies used therein (111). This framework relies on yes/no answers to questions about model characteristics across eight evenly weighted domains. These domains are (111):

1. Epidemiological validity (i.e., sex differentiation and age differentiation)
2. Symptomology and natural history (i.e., symptom type and duration, time to disease onset, disease progression, and severity)
3. Genetic validity (i.e., similarity between the animal and human genetic bases for disease and mutations)
4. Biochemical validity (i.e., similarity between biomarkers, prognostic and otherwise)
5. Etiological validity
6. Histological validity
7. Pharmacological validity (do the same drugs work in humans and animals?) and
8. Clinical translation (are the end points in animal studies translatable to human end points?).

Plotting these domain scores on a radar plot yields a visual comparison between potential models, and grading the responses allows a reproducible categorization of model validity.

This approach has numerous benefits, but a significant detriment. For benefits, it considers the whole organism, making assumptions clear (e.g., genetic basis, natural history of disease, biomarkers). It forces the user to recognize the dependence of model validity on the disease parameters, especially etiology and pathophysiology. Without a clear understanding of the mechanisms of human disease, it is difficult to claim that the model is valid. There is a dependency as well on the data that are used for the comparison; a pharmaceutical manufacturer may have access to different biomarkers or data than an academic. Hence, models may not have a consistent validation score, increasing the difficulty in communicating results.

The greatest failing in this validation schema is that it does not account for anything outside of the declared study design. For example, adverse effects, which by definition spring from mechanisms secondary to the effect of a therapy, are not included as part of the determination of validity. Predicting adverse effects is as important as predicting the effect of the therapy on the disease. Adverse drug effects are costly, accounting for up to 30% of hospital admissions in the United States and Canada, and significantly prolong hospital stays (112). Understanding adverse effects may lead to better science by suggesting new interactions that may not yet be known.

A third set of examples comes from the US FDA. As a response to the decline in new molecular entities and the general decline in novel submissions and qualifications through the 1990s and early 2000s, the FDA launched the Critical Path Initiative in 2004. In order to provide clear expectations to industry, the FDA developed a collection of qualification processes for models being used in regulatory filings. These include the Drug Development Tool and Animal Model Qualification Program (AMQP) of, respectively, the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research, as well as the Medical Device Development Tools framework of the Center for Devices and Radiological Health (113–116). These three frameworks are all directed at the goal of creating clear industry standards for using advanced technologies in the creation, description, and qualification of novel therapies, but they also serve as models of validation for more widespread use. We consider only the AMQP here.

The AMQP is a voluntary validation framework for animal models used for product approval under the Animal Rule, which covers products for which efficacy trials are unethical or field trials are not feasible. This rule is the most stringent set of guidelines on the use of animal models in the United States, exceeding those used for academic or basic research purposes. All four of the following criteria must be met for animal models to be the chief source of evidence: (*a*) the pathophysiological mechanism of toxicity of the substance is well understood as is the prevention

or reduction of this by the product under consideration; (*b*) the effect has been demonstrated in more than one animal species whose response is predictive for humans; (*c*) the animal study end point is clearly related to the benefit in humans; and (*d*) the kinetic and dynamic data are sufficient to allow for effective dose selection in humans (115).

Models are qualified independently of any particular product. Scientifically, this follows from the construction of the model: The animal model is of a disease or of a toxicity and is a tool to address the hypothesis that a drug or biologic agent effectively reduces the danger of that disease or toxicity. Critically, animal models are qualified only for a specific context of use (COU). The COU is a precise statement that describes the appropriate use of the animal model as a scientific tool. The COU distills the assumptions of a model and the utility of a model envisaged by the designers and defines a scope of validation. In this way, a new user has a summary expectation of the correct usage of a model. A single model may be useful in more than one COU, but each COU requires independent examination and qualification. Once qualified, the model is useful in any regulatory or developmental action within its COU and, by extension, in any scientific utility within that COU.

Mathematical Model Credibility in Health Care

Computational modeling and simulation (CM&S) have been used to support medical device design and development for decades. Since 1997, CM&S has been allowed as indirect evidence in submissions for regulatory qualification. By 2018, CM&S was usable as direct evidence for qualification, in some cases obviating Phase III clinical trials (117, 118). The foundation of the FDA's review of drugs, biologics, and devices is an assessment of the benefit–risk relationship (119). Following the US National Aeronautics and Space Administration's (NASA's) standards (120) and as an extension of ongoing efforts to define industry standards for engineering models, the American Society for Mechanical Engineers (ASME) developed the Verification and Validation (V&V) 40 guideline in partnership with the FDA, industry, and academia as a single framework designed to parameterize the credibility–risk balance detailed above (119–121). While it was intended for the types of modeling and simulation work seen in biomedical engineering (e.g., computational fluid dynamics, solid mechanics, electromagnetics and optics, ultrasound, and heat transfer), the plan is generalizable to any mathematical modeling effort. While other standards are in place in the FDA's Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research, the V&V 40 guideline is most broadly applicable to the biological and physiological sciences. The framework consists of four phases: plan, risk assessment, verification, and validation (**Table 1**). We describe these phases individually.

Plan. In the planning part of V&V 40, the user addresses two concepts: the question being asked and the COU of the model. Mathematical models of biological systems are intended to address a question that takes the form of one or more responses to a stimulus of some kind. In a computational fluid dynamics model, the response of interest might be hemolysis or platelet aggregation (or both) resulting from a certain type of blood flow. In a lumped parameter model, it might be the transient response to and steady-state outcome of a drug or surgical intervention. The validation process is intended to describe when a model is credible—that is, when data obtained from the model can influence a medical decision pertaining to that question. As such, formulating the question that the model addresses is the first essential step in validation because it provides a context in which to interpret all other information. The COU describes the conditions in which the model will be considered. Once use conditions pass out of this context, model credibility suffers. Together with the question, the COU describes the full intent of the modeler.

Table 1 Overview of verification and validation efforts (V&V 40) for evaluating the credibility of mathematical models in health care (121)

Identification number	Stage	Step	Activity
1.1	Verification plan	NA	Description of the question
1.2			Description of context of use
2.1	Model risk assessment	NA	Model influence
2.2			Decision consequence
3.1.1	Verification	Code	Software quality assurance
3.1.2			Numerical code verification
3.2.1		Calculation	Discretization error
3.2.2			Numerical solver error
3.2.3			Use error
4.1.1	Validation	Computational model	Model form
4.1.2			Model input
4.2.1		Comparator	Test samples
4.2.2			Test conditions
4.3.1		Assessment	Equivalence of Input Parameters
4.3.2			Equivalence of Output Parameters
4.4.1		Applicability	Relevance of Quantities of Interest
4.4.2			Relevance of Validation to COU

Abbreviations: COU, context of use; NA, not applicable.

Model risk assessment. Risk assessment is a combination of the weight that is given to simulation evidence in the medical decision process and the risk associated with an incorrect decision. The risk can be to a patient in the form of an adverse reaction, injury due to the intervention, or injury due to the disease process itself that is exacerbated by using the simulation-suggested intervention instead of another one. An example of the latter risk would be the use of an ineffective treatment in place of an effective treatment in a disease that rapidly progresses, such as cancer. The model risk assessment step describes the modeler's assumptions about how simulation evidence will be used.

Verification. The verification process is an engineering assessment of all modeling assumptions and tools, implemented to check that the results of the simulation consistently reflect the model code. It is composed of two types of activities: code verification and calculation verification. Code verification is a quality checking step, consisting of software quality assurance and numerical code checking. Software quality assurance tests the user interface, data pipelines, application protocol interfaces between model components, and all of the computational components that control the simulation, its inputs, and its outputs. Numerical code–checking establishes that the mathematical solver outputs the correct values against benchmark code.

The second verification activity is calculation verification, which tests the assumptions underlying the numerical solver. The first step is calculation of discretization error, which tests the mesh of solutions in the case of finite element methods, step sizes in the case of ordinary differential equation methods, or similar decisions regarding prediction intervals. The expectation is that the assumption of interval size can have a bearing on a model's performance, and this effect must

be quantified for the purpose of decision-making. Other solver parameters may also influence the final predictions for a model, and these parameters are highly dependent on the solution framework itself. Finally, the inputs and outputs of the model must be verified. This includes verifying that model inputs are inserted appropriately and that the correct outputs are recorded for comparison with external data.

Validation. The final step in determining a model's credibility is validation, which spans three types of activities: model validation, assessment of the comparators proposed to address the question within the context of use, and model assessment. Model validation consists of listing and testing assumptions about a model's form and also sensitivity analysis. Model assumptions take the form of the relationships that are included (and excluded) and the mathematical representations of those relationships. The sensitivity analysis is intended to test the robustness of the model to small changes in model coefficients. There are multiple methodologies for sensitivity analyses, and the methods chosen must be clear to end users.

The comparators' step in validation details the closeness of the model to the evidence. Credibility assessment of comparators begins with an assessment of the comparator data itself. The data is rated for its quantity and the similarity of the comparator system to the system detailed in the COU. Measurement uncertainty is accounted for in this step as well. The breadth of experimentation done to generate the comparators is also considered; model performance tests that challenge extreme and median values of variables are generally preferable to ones that stay within a narrower range. Finally, model variables are checked for similarity against the comparators. Input parameters (the interventions) must be compared with the experimental interventions. This comparison covers the range of values utilized in the model, as well as the closeness of the model intervention to the experimental one. For instance, a mathematical model of type 2 diabetes that determines insulin resistance by a single, user-defined parameter will be less credible than a model that includes glucose transporter kinetics and the mechanisms that affect those kinetics. Similarly, model outputs must be compared for similarity in both substance and value to the data being used for end point comparisons. This is the empirical validation component that is most commonly used as the sole measure of validation in physiological models.

The final step of validation is an assessment through the COU of the relevance of the model to the question being asked. This activity is concerned with establishing the relationships between the measured variables in the model and those used in the comparator and verification sets. In the context of physiological modeling, two questions need to be asked: What degree of reduction in system complexity is present in the simulation with respect to the quantities of interest? How does such a reduction affect simulation outcomes on the timescale inherent in the COU? The first question addresses the philosophy and intent of the model directly and interprets the effects of epistemic validation on the confidence that the simulation outputs describe the system. The answer to the second question should convey a level of confidence that the model's equations approximate the comparator variables in a reasonable way, taking into account the simulated scale of the computation. The context determines the appropriateness of the measure and its model. This brings the credibility assessment full circle.

CONCLUSIONS

In this review, we have attempted to convey some understanding of the richness of the problem of validating mathematical models. By contrasting simulation with widely used animal models, we have demonstrated that the difficulties encountered in validating mathematical models are intimately related to the same difficulties in animal models, challenges that have only recently begun

to be appreciated. Practical consideration of model credibility has caught up to philosophical concerns, and we are in an exciting period in the history of modeling. Armed with computational tools that grow more powerful each year, along with deep access to data, physiological modeling and simulation can be potent factors accelerating mechanistic understanding and have a real impact on the world through clinical utility, provided modelers can successfully demonstrate their value.

SUMMARY POINTS

1. The purpose of physiological mathematical models is to test hypotheses that cannot be feasibly tested in other ways. These restrictions may be due to ethics, measurement impossibility, the inability to construct an appropriate animal model of a process or disease, or some other factor.
2. In the past, validation of animal or mathematical models has been largely empirical. The failure of novel therapies to translate from animal models has forced a reassessment of validation modalities.
3. Currently, a notion of credibility—the confidence that a model represents reality well enough to justify making a decision with it—is the new standard of validation.
4. Validation for regulatory submission or for clinical adoption should include an assessment of all assumptions underlying a model, all processes for implementing and solving the model, all choices used for comparisons, and methodologies. Every assumption that can be communicated and tested should be.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors sincerely thank Dr. Tom Coleman for his decades of work spent in modeling and understanding how experimental findings impact model development, which, in turn, stimulates new experiments. We also thank the scientists and physicians that labored with Dr. Coleman at the University of Mississippi Medical Center, including Drs. Arthur Guyton, John Hall, Davis Manning, and Richard Summers. These individuals were integral both to the process of developing the technology for physiological modeling and for molding our understanding of the value of simulation and its validation. Finally, we would like to acknowledge the importance of the regulators at the FDA, especially Drs. Donna Lochner and Tina Morrison in the Center for Devices and Radiological Health, for bringing engineering sensibilities to nebulous questions about model credibility and quality.

LITERATURE CITED

1. Van Noorden R. 2017. The science that's never been cited. *Nature* 552:162–64
2. Jebeile J, Barberousse A. 2016. Empirical agreement in model validation. *Stud. Hist. Philos. Sci. A* 56:168–74
3. Ericsson AC, Crim MJ, Franklin CL. 2013. A brief history of animal modeling. *Mo. Med.* 110:201–5

4. Kaizer JS, Heller AK, Oberkampf WL. 2015. Scientific computer simulation review. *Reliab. Eng. Syst. Saf.* 138:210–18
5. Batzel J, Bachar M, Karemaker JM, Kappel F. 2012. Merging mathematical and physiological knowledge: dimensions and challenges. In *Mathematical Modeling and Validation in Physiology: Applications to the Cardiovascular and Respiratory Systems*, ed. J Batzel, M Baschar, F Kappel, pp. 3–20. Berlin: Springer-Verlag
6. Schultz SG. 2002. William Harvey and the circulation of the blood: the birth of a scientific revolution and modern physiology. *News Physiol. Sci.* 17:175–80
7. Rosenblueth A, Wiener N. 1945. The role of models in science. *Philos. Sci.* 12:316–21
8. Valverde I, Gomez-Ciriza G, Hussain T, Suarez-Mejias C, Velasco-Forte MN, et al. 2017. Three-dimensional printed models for surgical planning of complex congenital heart defects: an international multicentre study. *Eur. J. Cardiothorac. Surg.* 52:1139–48
9. Hoashi T, Ichikawa H, Nakata T, Shimada M, Ozawa H, et al. 2018. Utility of a super-flexible three-dimensional printed heart model in congenital heart surgery. *Interact. Cardiovasc. Thorac. Surg.* 27:749–55
10. Fatheree BH. 2004. *The First 75 Years: History of Hydraulics Engineering at the Waterways Experiment Station*. Vicksburg, MS: US Army Corps Eng., Eng. Res. Dev. Cent.
11. Am. Inst. Aeronaut. Astronaut. 2002. *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations* (AIAA G-077–1998(2002)). Am. Inst. Aeronaut. Astronaut.
12. US Dept. Def. 2018. *Department of Defense Instruction: Number 5000.61*. Washington, DC: Dept. Def.
13. Breitling R. 2010. What is systems biology? *Front. Physiol.* 1:159
14. Noble D. 2011. The aims of systems biology: between molecules and organisms. *Pharmacopsychiatry* 44(Suppl. 1):S9–14
15. Masters JR. 2002. HeLa cells 50 years on: the good, the bad and the ugly. *Nat. Rev. Cancer* 2:315–19
16. Starfield B, Hyde J, Gervas J, Heath I. 2008. The concept of prevention: a good idea gone astray? *J. Epidemiol. Community Health* 62:580–83
17. Dhiman N, Kingshott P, Sumer H, Sharma CS, Rath SN. 2019. On-chip anticancer drug screening—recent progress in microfluidic platforms to address challenges in chemotherapy. *Biosens. Bioelectron.* 137:236–54
18. Perry D, Sang A, Yin Y, Zheng YY, Morel L. 2011. Murine models of systemic lupus erythematosus. *J. Biomed. Biotechnol.* 2011:271694
19. Kurtz TW, Morris RC, Pershadsingh HA. 1989. The Zucker fatty rat as a genetic model of obesity and hypertension. *Hypertension* 13:896–901
20. Perez C, Fanizza LJ, Sclafani A. 1999. Flavor preferences conditioned by intragastric nutrient infusions in rats fed chow or a cafeteria diet. *Appetite* 32:155–70
21. Ankeny RA, Leonelli S, Nelson NC, Ramsden E. 2014. Making organisms model human behavior: situated models in North-American alcohol research, since 1950. *Sci. Context* 27:485–509
22. Shelley C. 2010. Why test animals to treat humans? On the validity of animal models. *Stud. Hist. Philos. Biol. Biomed. Sci. A* 41:292–99
23. Ramsden E. 2015. Making animals alcoholic: shifting laboratory models of addiction. *J. Hist. Behav. Sci.* 51:164–94
24. Kim JH, Snider T, Abel M, Zemel MB. 1994. Hypertension in young, healthy Zucker obese rats is not responsive to reduced salt intake. *J. Nutr.* 124:713–16
25. Pamidimukkala J, Jandhyala BS. 2004. Effects of salt rich diet in the obese Zucker rats: studies on renal function during isotonic volume expansion. *Clin. Exp. Hypertens.* 26:55–67
26. Morgan DA, Anderson EA, Mark AL. 1995. Renal sympathetic nerve activity is increased in obese Zucker rats. *Hypertension* 25:834–38
27. Carlson SH, Shelton J, White CR, Wyss JM. 2000. Elevated sympathetic activity contributes to hypertension and salt sensitivity in diabetic obese Zucker rats. *Hypertension* 35:403–8
28. Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, et al. 1998. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* 392:398–401
29. Matsuoka N, Ogawa Y, Hosoda K, Matsuda J, Masuzaki H, et al. 1997. Human leptin receptor gene in obese Japanese subjects: evidence against either obesity-causing mutations or association of sequence variants with obesity. *Diabetologia* 40:1204–10

30. Rolland V, Clement K, Dugail I, Guy-Grand B, Basdevant A, et al. 1998. Leptin receptor gene in a large cohort of massively obese subjects: no indication of the *fa/fa* rat mutation. Detection of an intronic variant with no association with obesity. *Obes. Res.* 6:122–27
31. Nilsson C, Raun K, Yan FF, Larsen MO, Tang-Christensen M. 2012. Laboratory animals as surrogate models of human obesity. *Acta Pharmacol. Sin.* 33:173–81
32. Bokulich A. 2011. How scientific models can explain. *Synthese* 180:33–45
33. Guyton AC, Coleman TG, Cowley AV Jr., Scheel KW, Manning RD Jr., Norman RA Jr. 1972. Arterial pressure regulation: overriding dominance of the kidneys in long-term regulation and in hypertension. *Am. J. Med.* 52:584–94
34. Guyton AC, Coleman TG, Fourcade JC, Navar LG. 1969. Physiologic control of arterial pressure. *Bull. N.Y. Acad. Med.* 45:811–30
35. Guyton AC, Coleman TG, Granger HJ. 1972. Circulation: overall regulation. *Annu. Rev. Physiol.* 34:13–46
36. Clemmer JS, Pruett WA, Hester RL, Iliescu R, Lohmeier TE. 2018. Role of the heart in blood pressure lowering during chronic baroreflex activation: insight from an in silico analysis. *Am. J. Physiol. Heart Circ. Physiol.* 315:H1368–82
37. Beard DA, Pettersen KH, Carlson BE, Omholt SW, Bugenhagen SM. 2013. A computational analysis of the long-term regulation of arterial pressure. *F1000Research* 2:208
38. Kurtz TW, DiCarlo SE, Pravenec M, Jezek F, Silar J, et al. 2018. Testing computer models predicting human responses to a high-salt diet. *Hypertension* 72:1407–16
39. Strange K. 2005. The end of "naive reductionism": rise of systems biology or renaissance of physiology? *Am. J. Physiol. Cell Physiol.* 288:C968–74
40. McShea DW, Wang SC, Brandon RN. 2019. A quantitative formulation of biology's first law. *Evolution* 73:1101–15
41. McShea DW, Brandon RN. 2010. *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems*. Chicago: Univ. Chicago Press
42. Mokwe E, Ohmit SE, Nasser SA, Shafi T, Saunders E, et al. 2004. Determinants of blood pressure response to quinapril in black and white hypertensive patients: the Quinapril Titration Interval Management Evaluation trial. *Hypertension* 43:1202–7
43. Hester RL, Iliescu R, Summers R, Coleman TG. 2011. Systems biology and integrative physiological modelling. *J. Physiol.* 589:1053–60
44. Ahn AC, Tewari M, Poon CS, Phillips RS. 2006. The limits of reductionism in medicine: Could systems biology offer an alternative? *PLOS Med.* 3:e208
45. Winberg E. 2010. *Science in the Age of Computer Simulation*. Chicago: Univ. Chicago Press
46. Oreskes N, Shrader-Frechette K, Belitz K. 1994. Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* 263:641–46
47. Rudiger D, Kupke SY, Laske T, Zmora P, Reichl U. 2019. Multiscale modeling of influenza A virus replication in cell cultures predicts infection dynamics for highly different infection conditions. *PLOS Comput. Biol.* 15:e1006819
48. Bonnaffoux A, Herbach U, Richard A, Guillemin A, Gonin-Giraud S, et al. 2019. WASABI: a dynamic iterative framework for gene regulatory network inference. *BMC Bioinform.* 20:220
49. Dura-Bernal S, Suter BA, Gleeson P, Cantarelli M, Quintana A, et al. 2019. NetPyNE, a tool for data-driven multiscale modeling of brain circuits. *eLife* 8:e44494
50. Clancy CE, An G, Cannon WR, Liu Y, May EE, et al. 2016. Multiscale modeling in the clinic: drug design and development. *Ann. Biomed. Eng.* 44:2591–610
51. Collins FS. 2011. Reengineering translational science: The time is right. *Sci. Transl. Med.* 3:90cm17
52. Scruggs SB, Watson K, Su AI, Hermjakob H, Yates JR 3rd, et al. 2015. Harnessing the heart of big data. *Circ. Res.* 116:1115–19
53. Longtin R. 2005. An integrated approach: systems biology seeks order in complexity. *J. Natl. Cancer Inst.* 97:476–78
54. Krogh A. 1919. The number and distribution of capillaries in muscles with calculations of the oxygen pressure head necessary for supplying the tissue. *J. Physiol.* 52:409–15

55. Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117:500–44
56. Starling EH. 1899. The glomerular functions of the kidney. *J. Physiol.* 24:317–30
57. Deen WM, Robertson CR, Brenner BM. 1972. A model of glomerular ultrafiltration in the rat. *Am. J. Physiol.* 223:1178–83
58. Chang RS, Robertson CR, Deen WM, Brenner BM. 1975. Permselectivity of the glomerular capillary wall to macromolecules. I. Theoretical considerations. *Biophys. J.* 15:861–86
59. Kim SW, Cho HG, Park CR. 2009. Fabrication of unagglomerated polypyrrole nanospheres with controlled sizes from a surfactant-free emulsion system. *Langmuir* 25:9030–36
60. Yildirim N, Santillan M, Horike D, Mackey MC. 2004. Dynamics and bistability in a reduced model of the *lac* operon. *Chaos* 14:279–92
61. Veliz-Cuba A, Stigler B. 2011. Boolean models can explain bistability in the *lac* operon. *J. Comput. Biol.* 18:783–94
62. Dreisigmeyer DW, Stajic J, Nemenman I, Hlavacek WS, Wall ME. 2008. Determinants of bistability in induction of the *Escherichia coli lac* operon. *IET Syst. Biol.* 2:293–303
63. Coleman TG, Randall JE. 1983. HUMAN: a comprehensive physiological model. *Physiologist* 26:15–21
64. Abram SR, Hodnett BL, Summers RL, Coleman TG, Hester RL. 2007. Quantitative Circulatory Physiology: an integrative mathematical model of human physiology for medical education. *Adv. Physiol. Educ.* 31(2):202–10
65. Summers R, Montani J. 1989. Mathematical model of glucose homeostasis for the study of metabolic states. *J. Miss. Acad. Sci.* 34:25–32
66. Summers RL, Montani JP, Woodward LH, Coleman TG, Hall JE. 1997. Theoretical analysis of the mechanisms of chronic hyperinsulinemia. *Comput. Biol. Med.* 27:249–56
67. Guyton AC, Coleman TG, Manning RD Jr., Hall JE. 1984. Some problems and solutions for modeling overall cardiovascular regulation. *Math. Biosci.* 72:141–55
68. Pruett WA, Clemmer JS, Hester RL. 2016. Validation of an integrative mathematical model of dehydration and rehydration in virtual humans. *Physiol. Rep.* 4:e13015
69. Summers RL, Martin DS, Meck JV, Coleman TG. 2007. Computer systems analysis of spaceflight induced changes in left ventricular mass. *Comput. Biol. Med.* 37:358–63
70. Summers RL, Ward KR, Witten T, Convertino VA, Ryan KL, et al. 2009. Validation of a computational platform for the analysis of the physiologic mechanisms of a human experimental model of hemorrhage. *Resuscitation* 80:1405–10
71. Clemmer JS, Pruett WA, Hester RL, Iliescu R, Lohmeier TE. 2018. Role of the heart in blood pressure lowering during chronic baroreflex activation: insight from an in silico analysis. *Am. J. Physiol. Heart Circ. Physiol.* 315:H1368–82
72. Shanks N, Greek R, Greek J. 2009. Are animal models predictive for humans? *Philos. Ethics Humanit. Med.* 4:2
73. Drake CG, Rozzo SJ, Vyse TJ, Palmer E, Kotzin BL. 1995. Genetic contributions to lupus-like disease in (NZB×NZW)F_1 mice. *Immunol. Rev.* 144:51–74
74. Drake CG, Rozzo SJ, Hirschfeld HF, Smarnworawong NP, Palmer E, Kotzin BL. 1995. Analysis of the New Zealand Black contribution to lupus-like renal disease: multiple genes that operate in a threshold manner. *J. Immunol.* 154:2441–47
75. Helyer BJ, Howie JB. 1963. Renal disease associated with positive lupus erythematosus tests in a cross-bred strain of mice. *Nature* 197:197
76. Burnet FM, Holmes MC. 1965. The natural history of the NZB/NZW F_1 hybrid mouse: a laboratory model of systemic lupus erythematosus. *Australas. Ann. Med.* 14:185–91
77. Van Dam D, De Deyn PP. 2006. Drug discovery in dementia: the role of rodent models. *Nat. Rev. Drug Discov.* 5:956–70
78. Willner P. 1984. The validity of animal models of depression. *Psychopharmacology* 83:1–16
79. Royal K. 2016. "Face validity" is not a legitimate type of validity evidence! *Am. J. Surg.* 212:1026–27
80. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 2006. 1,026 experimental treatments in acute stroke. *Ann. Neurol.* 59:467–77

81. Kaitin KI, DiMasi JA. 2011. Pharmaceutical innovation in the 21st century: new drug approvals in the first decade, 2000–2009. *Clin. Pharmacol. Ther.* 89:183–88
82. Pevarello P. 2009. Recent drug approvals from the US FDA and EMEA: what the future holds. *Future Med. Chem.* 1:35–48
83. Gabrielsson J, Green AR. 2009. Quantitative pharmacology or pharmacokinetic pharmacodynamic integration should be a vital component in integrative pharmacology. *J. Pharmacol. Exp. Ther.* 331:767–74
84. Wendler A, Wehling M. 2012. Translatability scoring in drug development: eight case studies. *J. Transl. Med.* 10:39
85. Prabhakar S. 2012. Translational research challenges: finding the right animal models. *J. Investig. Med.* 60:1141–46
86. Wendler A, Wehling M. 2017. Translatability score revisited: differentiation for distinct disease areas. *J. Transl. Med.* 15:226
87. US FDA (Food Drug Adm.). 2006. *Guidance for Industry: Nonclinical Safety Evaluation of Drugs or Biologic Combinations*. Rockville, MD: FDA
88. US FDA (Food Drug Adm.). 2010. *Guidance for Industry and FDA staff: General Considerations for Animal Studies for Cardiovascular Devices*. Rockville, MD: FDA
89. Lowenstein PR, Castro MG. 2009. Uncertainty in the translation of preclinical experiments to clinical trials. Why do most phase III clinical trials fail? *Curr. Gene Ther.* 9:368–74
90. Mak IW, Evaniew N, Ghert M. 2014. Lost in translation: animal models and clinical trials in cancer treatment. *Am. J. Transl. Res.* 6:114–18
91. US FDA (Food Drug Adm.). 2004. *Innovation Stagnation: Challenge and Opportunity on the Critical Path to New Medical Technologies*. Rockville, MD: FDA
92. Adams CP, Brantner VV. 2006. Estimating the cost of new drug development: Is it really $802 million? *Health Aff.* 25:420–28
93. DiMasi JA, Grabowski HG, Hansen RW. 2015. The cost of drug development. *N. Engl. J. Med.* 372:1972
94. DiMasi JA, Grabowski HG, Hansen RW. 2016. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* 47:20–33
95. DiMasi JA, Hansen RW, Grabowski HG. 2003. The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22:151–85
96. DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. 1991. Cost of innovation in the pharmaceutical industry. *J. Health Econ.* 10:107–42
97. DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. 1995. Research and development costs for new drugs by therapeutic category: a study of the US pharmaceutical industry. *Pharmacoeconomics* 7:152–69
98. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, et al. 2010. Can animal models of disease reliably inform human studies? *PLOS Med.* 7:e1000245
99. Hughes RIG. 1999. The Ising model, computer simulation, and universal physics. In *Models as Mediators: Perspectives on Natural and Social Science*, ed. MS Morgan, M Morrison, pp. 97–145. Cambridge: Cambridge Univ. Press
100. Wu KK, Huan Y. 2008. Streptozotocin-induced diabetic models in mice and rats. *Curr. Protoc. Pharmacol.* 5:47.1–20
101. Duhem P. 1954. *The Aim and Structure of Physical Theory*. Princeton: Princeton Univ. Press
102. Patterson EA, Whelan MP. 2017. A framework to establish credibility of computational models in biology. *Progress Biophys. Mol. Biol.* 129:13–19
103. Schruben LW. 1980. Establishing the credibility of simulations. *Simulation* 34:101–105
104. Pathmanathan P, Gray RA. 2018. Validation and trustworthiness of multiscale models of cardiac electrophysiology. *Front. Physiol.* 9:106
105. Roache PJ. 2009. Validation—What does it mean? *J. Fluids Eng.* 131:034503
106. Morrison TM, Dreher ML, Nagaraja S, Angelone LM, Kainz W. 2017. The role of computational modeling and simulation in the total product life cycle of peripheral vascular devices. *J. Med. Device* 11:024503
107. Audi R. 2010. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. New York: Routledge

108. Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG, et al. 2010. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br. J. Pharmacol.* 160:1577–79
109. McGonigle P, Ruggeri B. 2014. Animal models of human disease: challenges in enabling translation. *Biochem. Pharmacol.* 87:162–71
110. Masic I, Miokovic M, Muhamedagic B. 2008. Evidence based medicine—new approaches and challenges. *Acta Inform. Med.* 16:219–25
111. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, et al. 2019. A standardised framework to identify optimal animal models for efficacy assessment in drug development. *PLOS ONE* 14:e0218014
112. Sultana J, Cutroneo P, Trifiro G. 2013. Clinical and economic burden of adverse drug reactions. *J. Pharmacol. Pharmacother.* 4(Suppl. 1):S73–77
113. US FDA (Food Drug Adm.). 2017. *Qualification of Medical Device Development Tools: Guidance for Industry, Tool Developers, and Food and Drug Administrative Staff*. Rockville, MD: FDA
114. US FDA (Food Drug Adm.). 2014. *Guidance for Industry and FDA Staff: Qualification Process for Drug Development Tools*. Rockville MD: FDA
115. US FDA (Food Drug Adm.). 2015. *Product Development Under the Animal Rule—Guidance for Industry*. Rockville MD: FDA
116. Woodcock J, Buckman S, Goodsaid F, Walton MK, Zineh I. 2011. Qualifying biomarkers for use in drug development: a US Food and Drug Administration overview. *Expert Opin. Med. Diagn.* 5:369–74
117. Gottlieb S. 2017. How FDA plans to help consumers capitalize on advances in science. *FDA Blog*, July 7. **https://www.fda.gov/news-events/fda-voices-perspectives-fda-leadership-and-experts/how-fda-plans-help-consumers-capitalize-advances-science**
118. Subcomm. Health. 2018. Hearing on "21st century cures implementation: updates from FDA and NIH." In *Committee on Energy and Commerce*. Washington, DC: Subcommittee on Health. **https://www.nih.gov/about-nih/who-we-are/nih-director/testimony-21st-century-cures-implementation-updates-fda-nih**
119. US FDA (Food Drug Adm.). 2018. *Benefit-Risk Assessment in Drug Regulatory Decision-Making: Draft PDUFA VI Implementation Plan (FY 2018–2022)*. Rockville, MD: FDA
120. NASA (Natl. Aeronaut. Space Adm). 2009. *Standard for Models and Simulations*. Washington, DC: NASA
121. ASME (Am. Soc. Mech. Eng.). 2018. *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*. New York: ASME