A tentative taxonomy for predictive models in relation to their falsifiability

BY MARCO VICECONTI*

Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Via di Barbiano 1/10, 40136 Bologna, Italy

The growing importance of predictive models in biomedical research raises some concerns on the correct methodological approach to the falsification of such models, as they are developed in interdisciplinary research contexts between physics, biology and medicine. In each of these research sectors, there are established methods to develop cause–effect explanations for observed phenomena, which can be used to predict: epidemiological models, biochemical models, biophysical models, Bayesian models, neural networks, etc. Each research sector has accepted processes to verify how correct these models are (falsification). But interdisciplinary research imposes a broader perspective, which encompasses all possible models in a general methodological framework of falsification. The present paper proposes a general definition of ‘scientific model’ that makes it possible to categorize predictive models into broad categories. For each of these categories, generic falsification strategies are proposed, except for the so-called ‘abductive’ models. For this category, which includes artificial neural networks, Bayesian models and integrative models, the definition of a generic falsification strategy requires further investigation by researchers and philosophers of science.

Keywords: biomedicine; predictive models; falsification; integrative models

1. Introduction

A model is an invention, not a discovery [1, p. 277].

As biomedical researchers realize that a purely reductionist approach is not sufficient to investigate some of the most challenging processes of human physiopathology [2–7], the importance of predictive models within biomedical research grows exponentially (see the articles included in Phil. Trans. R. Soc. A [8–13]). This is mostly due to the fact that any integrative approach requires a way to capture the fragments of reductionist knowledge (i.e. knowledge on a phenomenon limited to a single characteristic scale, or to a subset of interactions) into artefacts that make it practically possible to compose these fragments of knowledge into a systemic understanding of physiopathology processes, and predictive models are the most effective way to capture knowledge in a reusable and composable way [5,14,15].

*viceconti@tecnio.ior.it

One contribution of 11 to a Theme Issue ‘Towards the virtual physiological human: mathematical and computational case studies’.
But also this poses a number of new challenges. One relevant to this paper is that these predictive models must be developed and investigated in terms of their correctness in a territory of science that is at the crossroads between biological, medical, physical and engineering sciences. Each of these domains is mature enough to have its own well-established epistemology, in particular with respect to the formulation of theories and the assessment of their truth content; however, these epistemologies are significantly different from each other. Without entering into a difficult and potentially controversial discussion, we can recognize that medicine and engineering are applied sciences, whose scope is to solve problems, whereas biology and physics are fundamental sciences, whose scope is to establish new scientific knowledge about the world. In addition, engineering and physics tend to embrace the scientific method in full, aiming towards a mechanistic explanation of a phenomenon whenever possible, whereas biology and medicine tend to amplify the empirical/observational component of research.

Developing predictive models in this interdisciplinary region of science is dangerous. Is it possible to develop physics-based predictive models when the observations we start from in order to develop the model are subject to severe uncertainty? If, following the proposal by Popper [16], theories can only be falsified but never confirmed, what does it mean to validate a predictive model? Is there an inherent difference in the potential truth content of a mechanistic model with respect to a phenomenological model based exclusively on observations?

In the following pages, I shall try to answer these questions, by first providing a definition of ‘scientific model’ that, while general in nature, provides those prescriptive elements to make them potentially useful for scientists.

2. What is a model?

Before I try to provide a tentative answer to this apparently simple question, it is important to point out that with the term ‘scientific model’ I intend to indicate any artefact that we use to idealize a portion of reality for the purpose of scientific investigation. In this sense what follows is applicable not only to mathematical/computer models, but also to any other ‘idealization vehicle’ that we use in the practice of science. For example, when we investigate an animal under the assumption that it will provide information on certain aspects of human physiopathology, that animal is a model. When we use some tissue-engineering product to investigate properties of real human tissues, that tissue-engineering product is a model. Thus, what is discussed here is potentially of much broader interest.

At the present time we do not have a precise definition of what a model is in science, shared and accepted by most scientists. In such cases, we need to resort to those who look to science from outside, the philosophers of science. Most of the philosophical literature debates the role of scientific models with respect to theories. In their relevant extensive review, Frigg & Hartmann [17] argued that authors advocating the so-called syntactic approach to theories recognize a limited role for models, while the advocates of the semantic approach to theories, which reject the concept of calculus as a unique representation of theories, see theories as families of models. While this net interpretation of previous works is being questioned [18], the point remains that most authors until recently have considered models only in the context of scientific theorization.
More recently alternative perspectives have been proposed, wherein models are recognized as autonomous agents possibly independent from theories (i.e. [19]). In these papers, models are recognized as complements, replacements or preliminaries of theories, but much of the recent philosophical debate focuses on the representational purpose of models and whether this has unique characteristics, or if it can be conducted to the more general philosophical discussion on the representation of reality [20–24].

While this debate is of great interest and relevance, it lacks, so far, those prescriptive elements that scientists require in order to improve their work. In addition, in the current philosophical debate, there is an implicit assumption that models are a product of science. However, if we look at the recent neuroscience literature, and in particular at how long-term memory works, we are forced to recognize that idealizing reality is a fundamental cognitive process [25–28]. From this neuroscience perspective, we can define models, in a general sense, as:

finalized cognitive constructs of finite complexity that idealize an infinitely complex portion of reality.

Here I pose, as an a priori concept, the assumption that every portion of reality, no matter how small or limited it is, is always infinitely complex to ‘understand’.¹ On the contrary, any idealization of that portion of reality is of finite complexity; this is made possible by finalizing the cognitive elaboration of the perceptual information. So to remember Mary’s face, we idealize certain elements of our perceptions, and associate them with certain memories, whereas to recognize that Mary does not feel well, we idealize other elements of our perceptions, and associate them with other memories.

A careful reader can spot an apparent contradiction here: before I stressed how scientific models are also physical things, such as tissue-engineering products or research animals. But now I define models as products of our minds. But the contradiction is only apparent. In reality, the animal is not the model; the model is the idealization we apply in interpreting the observations that we make on that animal as representative of human physiology. Knuuttila [32] defines a model as ‘a two-fold phenomenon, comprised of both a material sign-vehicle and an intentional relation of representation that connects the sign-vehicle to whatever it is that is being represented’. I prefer to use the term model to indicate the idealization, and the term analogue to indicate the ‘device that can be used for the purpose of making a model’ [1].

3. What is a scientific model?

If we accept that modelling is an essential process of the human mind, then what is special about scientific models? I propose that scientific models are discernible from the other idealizations of the human mind because of the finality that informs the choice of idealization.

¹The infinite complexity of reality descends, for example, from the so-called ‘many worlds interpretation’ of quantum theory [29]; it is also advocated by various sociologists, such as Durkheim [30] or Weber [31]. If we limit this concept by considering only the disparity between the complexity of reality and the capacity of the human brain, simple considerations across dimensional scales down to sub-atomic particles would support this statement.
But ‘science’ is a very broad term. It is used to indicate formal sciences, such as mathematics, and empirical sciences, such as physics or sociology. In formal sciences, the role of models is not necessarily linked to the concept of reality, and thus our definition appears inappropriate.

Among empirical sciences, we must distinguish between natural sciences and social and behavioural sciences. Hereinafter I shall refer with the term ‘science’ to all areas that use the scientific method. If we define science as ‘a systematic enterprise of gathering knowledge based on the scientific method’ and the scientific method as ‘the typical modality by which science proceeds to reach a knowledge of reality that is objective, reliable, verifiable and shareable’, then we can define scientific models as:

finalized cognitive constructs of finite complexity that idealize an infinitely complex portion of reality through idealizations that contribute to the achievement of knowledge on that portion of reality that is objective, shareable, reliable and verifiable.

This definition applies to all natural sciences. Whether it might also be useful in social sciences such as economics is less clear. The adoption of the scientific method in social sciences is debated, in particular with reference to the requirement of objectivity. But in social sciences such as economics, where modelling is a standard practice, I believe the definition above can be relevant and prescriptive.

4. Attributes of scientific models

As I declared in §1, I do not pretend that this definition has any particular philosophical value; however, as I shall show in the following pages, it has the considerable advantage of yielding a number of prescriptive elements that are potentially useful in dealing with the epistemological confusion that the use of predictive models in biomedicine is posing.

A first set of prescriptive observations emerge directly from the definition, by noticing that every scientific model should present the four fundamental attributes the definition provides. Alternatively, following the approach of Popper, we should consider ranking scientific models according to the degree to which they are capable of achieving these fundamental attributes. Now, let us take a close look to these four attributes.

— Objective is the opposite of subjective: agreeable by many peers, mostly unbiased by personal views. In empirical sciences, no model will ever be entirely objective; this is an abstract concept, unachievable in full, but to which nevertheless we all should aim.\(^2\) If we recognize that the search for objectivity does not imply the existence of a unique perspective. To cite one of the reviewers of this paper: ‘Instead of searching for the meta-viewpoint, i.e. the point from which the condition of objectivity as correspondence (with the structure of reality) is achieved, multiple viewpoints could be employed in describing or explaining a given phenomenon. These might complement each other and allow one to acquire a more complete understanding of it’. See also the concept of complementarity applied to biology [33].
scientific truths\textsuperscript{3} is a collective endeavour, then the search for objectivity is strongly linked with the second attribute, shareability.

— **Shareable** means that it can be shared with our peers. Again, from a prescriptive perspective, objectivity and shareability are strongly related to the need for **reproducibility**. This opens up a whole new argument, which I cannot discuss in full here. However, it should be pointed out that while, in principle, reproducibility of experiments is possible through a textual description (materials and methods) as the artefacts they involve are generally available, the reproducibility of numerical studies too frequently relies on digital artefacts that are not readily available (i.e. in-house-developed software), or the fine details about parameters and settings that are necessary to reproduce the study are not provided. This is why many peers are now advocating the need for mechanisms for model sharing attached to peer-review journals that host numerical studies of this kind.\textsuperscript{4}

— **Reliable** means ‘that can be trusted’. In this context, I define this property as the ability of a model to provide comparable quality of performance within a given range of conditions. The quality of performance depends on the finality of the model, which thus must always be stated explicitly. We also recognize that almost no model remains reliable for every possible range of conditions (or more correctly to every possible state the portion of reality under examination can assume). This range is the defined **limit of validity** of the model, and it also must be stated explicitly every time we develop a scientific model.

— **Verifiable** means ‘conceived and presented so that the quality of its performance can be assessed’. This brings us directly to one of the problems I listed at the beginning of this paper: if, according to Popper, no theory can be confirmed, what is the sense of validating a predictive model? This apparent contradiction derives from the fact that the definition of science I provided above is somehow incomplete. In fact, science pursues two distinct objectives: problem-solving and knowledge production.

Traditionally, each scientific domain is aimed at one or the other of these objectives: typically medicine and engineering aim to solve problems, whereas physics and biology aim to understand why some problems can be solved in a certain way, i.e. to form new knowledge about how nature works. However, in integrative biomedical research, this separation is much more blurred, and the risk of mixing up these two scopes is greater. This point is relevant here because, when the scope of a model is to solve a given problem, that model can be validated. Indeed, we can imagine one or more experiments that confirm that the model is capable of predicting the phenomenon of interest with a level of

\textsuperscript{3}This is the only point where I use the word *truth*, whereas in the rest of the paper I use *truth content*: the ‘truth content’ of a theory is the class of true propositions that may be derived from it, and the ‘falsity content’ of a theory is the class of the theory’s false consequences [34]. Here, the accent is on the search for truth, rather than on the existence of an absolute scientific truth, which in agreement with Popper I do not believe will ever exist. In the rest of the paper, the word *true* is used only in the context of logic.

\textsuperscript{4}See, for example, the petition for model reproducibility promoted by the Virtual Physiological Human Network of Excellence (http://www.biomedtown.org/VPH/petitions/reproducibility/).

---

*Phil. Trans. R. Soc. A* (2011)
accuracy sufficient to solve the problem. On the contrary, if a predictive model aims to provide some new knowledge on a certain phenomenon, then we can only try to falsify this model, but it is not possible to develop a finite number of experiments that conclusively confirm that model to be valid.

5. Idealization and truth content

In order to be objective and shareable, the idealizations of scientific models should always be formulated through logical reasoning. This consideration makes it possible to elaborate a tentative taxonomy for scientific models that makes possible a general discussion on the truth content and the falsification strategies for each family of models.

There are three recognized kinds of logical reasoning:

— deduction, deriving \( a \) as a consequence of \( b \);
— induction, inferring \( a \) from multiple instantiations of \( b \); and
— abduction, inferring \( a \) as an explanation of \( b \).

While deduction and induction should be familiar to most readers, it might be worth providing a bit more detail on the third type of logical reasoning, abduction. First proposed by Charles S. Peirce [35], this type of reasoning can be summarized as follows:

— the surprising fact \( b \) is observed;
— but if \( a \) was true, \( b \) would be a matter of course; and
— hence, there is a reason to suspect that \( a \) is true.

Every scientific model contains multiple idealizations, which, in general, are formulated according to different types of logical reasoning. However, it is usually possible to identify one particular type of reasoning as predominant in the construction of the model. We can thus taxonomize scientific models according to the predominant type of logical reasoning used to build its idealizations, and use this partitioning to discuss in a general context the potential truth content and the possible falsification strategies that can be used for each category.

(a) Deductive models

To this family belong all physics-based mechanistic models. In these models, we idealize the observed phenomenon according to a given set of physics laws, capable of capturing the aspects of the phenomenon that the model scope regards as relevant. All the rest of the model is deduced from these general laws. Typical examples are the solid and fluid mechanics models used in biomechanics (i.e. [36,37]), or the electro-chemo-mechanical models used in cardiac physiology (see Bassingthwaighte’s review article on the cardiac physiome [38]).

The potential truth content of deduction is totally dependent on the truth content of the initial assumptions: if the precedents are true, deduction ensures that the consequents are also true. If the set of laws we choose to represent the

Phil. Trans. R. Soc. A (2011)
phenomenon of interest (i) have an extensive truth content and (ii) are accurate in capturing the relevant aspects of the phenomenon of interest, everything deduced from them has the same truth content.

Another important property of deductive models is potential generality. If the laws I used as antecedents are sufficiently general, I might find that the same model is effective in predicting a very large set of similar phenomena.

Because in a deductive model, the consequents' truth content derives from that of the antecedents, deductive models are usually challenged by questioning the laws of physics or of biology that we used to build our deductive model. As these laws frequently have already been challenged for a long time, the point is not to prove the laws false in themselves, but rather to challenge the assumption that the phenomenon under observation can be accurately modelled in its aspects relevant to the modelling scope by that choice of laws. So, for example, if we idealize the biomechanical behaviour of the mineralized extracellular matrix of bones with the theory of elasticity, in order to falsify this idealization, we should conduct experiments to see if, within the range of physiological skeletal loading, bone tissue shows a significant deviation from proportionality in the relationship between stress and strain.

The other falsification strategy is to investigate phenomena that, while independent from the one we modelled, logically descend from the idealization we made. So, staying with our bone example, if bone behaviour in physiological conditions is linear elastic, then, in the range of deformation rates that is physiologically possible, bones should exhibit no significant viscous dissipation, which can be verified with isothermal experiments [39].

(b) Inductive models

All models belong to this category, where the prediction is generated by induction, i.e. models based on the interpolation or extrapolation of observational data or measurements. Typical examples are predictive models based on epidemiological studies (e.g. population-based models to predict the risk of bone fractures in osteoporotic patients [40]).

In models where the idealizations are predominantly formulated by induction, within a deterministic framework the potential truth content is null. In fact, if an inductive model accurately predicts a phenomenon \( n \) times, nothing guarantees that for the \( n + 1 \) set of parameters the model will also be accurate. In a probabilistic framework, which postulates that all physical phenomena show a certain degree of regularity, the potential truth content of an inductive model depends on statistical indicators such as level of significance, power of the test, etc.

The first approach to the falsification of inductive models should also be inductive. We frequently forget that, if we want to use a data regression to make predictions, it is not sufficient to verify that regression is accurate ‘on average’. In this sense, attempts to falsify an inductive model by verifying if the predicted values and the measured values simply correlate should be considered largely inadequate. Of course, a predictive model must make predictions that are correlated with the reality they intend to model. But predicting is much more than correlating; it implies an explanatory nature that simple correlation lacks. What we must do is to verify if the predictive error remains for every prediction with the acceptance limits over the entire range of validity of the model.
Another way to see this problem is to recognize that the peak predictive error we observe over a large set of validation experiments should be considered as the interval of uncertainty of our model’s predictions. If we run our model with two different sets of parameters, and the two predictions differ by less than the peak error we observed during the validation experiments, nothing can be concluded, because we cannot positively confirm that the difference really exists and it is not due to the inadequacy of our model.

But this is not sufficient. Inductive models must also be challenged on the statistical side. If the model assumes the data are normally distributed, are they? What is the level of statistical significance for our predictor? What is the statistical power? These tests should be done not only on the data that we use to inform our inductive model, but also on the predictions produced by our model. A typical example is the following. Say that we want to model the relationship between the electric charge on and the concentration of a chemical species in an electrophysiology experiment. The concentration is measured for a wide range of positive and negative charges, and then the entire set of data pairs is fitted with a linear regression that shows a regression coefficient $R^2 > 0.9$. We thus use the regression equation as a predictive model of species concentration as a function of the electric charge. If the process is polarized, i.e. positive charges produce opposite effects to negative charges, such a model might be easily falsified by repeating the regression separately on the positive and the negative charge measurements. We might discover that the two resulting regressions are significantly different in slope and/or intercept from the one obtained by pooling all data together. This is an example of a regression that is statistically significant, but that has a very low statistical power.

\[(c)\text{ Abductive models}\]

This is not only the most difficult but also the most interesting family of models. What are the models that change as we add new observations? So far I have identified three examples:

— **Neural networks** are models that change according to the training set we provide them, i.e. new observations.
— **Bayesian models** are models that change the probability associated with the model as we add new observations.
— **Integrative models** are models made of component idealizations and of relational idealizations defining the relations between components. Integrative models are abductive: by adding a new component idealization into our model, it predicts a surprising fact $b$; if we observe something close to $b$ under the same conditions, the compositional idealization that forms the model is the fact $a$ that by abduction is suspected to be true.

Neural network models are being used in a myriad of biomedical research domains, such as analytical chemistry [41], diagnosis of pancreatic diseases [42], pharmacokinetics [43] and drug discovery [44], to predict renal colic in emergency settings [45], in the recognition of the activity type in accelerometer signals [46], in neuromotor control modelling [47], etc. In fact, ‘artificial neural network’ returns 2736 papers in PubMed.
Similarly, Bayesian models are used in myriad applications, including the interpretation of magnetic resonance imaging of the human brain [48], the study of metabolic systems [49], etc. ‘Bayesian model’ returns 6346 papers in PubMed.

The term ‘integrative model’ is quite recent and not yet universally adopted. It was probably first used extensively in psychology, to indicate approaches that combined psychology and other scientific disciplines such as psychoneuroimmunology [50]. With respect to physiology and pathology modelling, the idea of integrative modelling largely remains a conceptual manifesto, although under the umbrella of the Virtual Physiological Human initiative a considerable amount of research is being conducted that should produce a large number of examples in the near future. Some early examples are the multi-scale study of skeletal strength [14], a multi-scale model of gas exchange in fruit [51], a multi-scale model of gastric electrophysiology [52], a multi-scale model of bone remodelling [53] and a multi-scale model of cardiovascular regulation [54].

This last category, in particular, is central to our reflection; as I already mentioned in §1, various authors are suggesting that, to overcome the drift in biomedical research to frequently transform the necessary methodological reductionism5 into the very dangerous causal reductionism6 and to properly account for complex systemic emergence [57] and the issue of complementarity [33], a possible approach is to capture reductionist knowledge into component idealizations, and then add them into integrative models together with the appropriate relational idealizations.

However, at the current state of our analysis very little can be said on the potential truth content of abductive models, and on the best strategies to falsify them. The best advice for integrative models is first to attempt the falsification of each component idealization and of each relational idealization separately, and then to repeat it for the integrative model now considered as a whole.

However, this strategy does not apply to Bayesian models or to neural networks; thus a more general falsification framework, valid for all abductive models, needs to be elaborated. A possible direction of research is that of the so-called bootstrap methods [58]: the repetition of the construction of the model through the abduction cycle using multiple sets of independent observations, all relative to the phenomenon to be modelled, should yield the same model.

6. Discussion

The potential truth content of deductive models is very high. Using the Popperian approach, the falsifiability of deductive models is, in general, excellent. By formulating our idealizations in terms of a set of fundamental principles from which we deduce, we provide an excellent target for the falsification of the model.

5Methodological reductionism is defined as ‘an approach to understanding the nature of complex things by reducing them to the interactions of their parts, or to simpler or more fundamental things’ [55].

6Causal reductionism implies that the causes acting on the whole are simply the sum of the effects of the individual causalities of the parts [55,56]. In other words, macro-level causal relations are seen as reducible to micro-level causal relations.
On the contrary, the potential truth content of inductive models is low. With the exception of the most trivial cases, inductive models can be falsified only on a statistical basis, which in the end transfers the problem of model validity to that more general one of the regularity of the predicted phenomenon. In this sense, while we recognize the usefulness of inductive models in a number of cases, everything else being the same (ceteris paribus clause), deductive models should always be preferred to inductive models for their superior potential truth content, i.e. their higher falsifiability.

The discussion for abductive models is more complex; we are tempted to say that their recursive nature perfectly matches the process of scientific research, and because of this they promise to embody the scientific method better than any other modelling approach. However, the current incompleteness of the falsification framework for this family of models casts some concerns on their use. Because of this I strongly recommend that the scientific community, possibly with the help of the philosophy of science community, adopts the problem of the falsification of abductive models as a central goal for future research.

This paper results from the collective reflection on the use of integrative models in biomedicine that is permeating the research community that participates in the Virtual Physiological Human initiative. In particular, I would like to acknowledge the contributions of various individuals, including Denis Noble, Peter Hunter, Ralph Müller and Peter Kohl, who found the time and the patience to discuss with me these fairly convoluted arguments.

This study was supported by the European Commission through the projects ‘The Osteoporotic Virtual Physiological Human—VPHOP’ grant FP7-ICT2008-223865 and ‘Virtual Physiological Human Network of Excellence’ grant FP7-2007-IST-223920. The author declares that he does not have any financial or personal relationships with other people or organizations that could have inappropriately influenced this study.

References

2 Van Regenmortel, M. H. V. 2004 Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological complexity and overcome the limitations of reductionism. EMBO Rep. 5, 1016–1020. (doi:10.1038/sj.embor.7400284)

For more information on the Virtual Physiological Human initiative refer to the VPH Network of Excellence web site: http://www.vph-noe.eu.


23 Giere, R. N. 2004 How models are used to represent reality. *Phil. Sci.* **71**, 742–752. (doi:10.1086/425063)


Phil. Trans. R. Soc. A (2011)


