Real and Virtual Clinical Trials: a Formal Analysis

Osimani, B., Bertolaso, M., Poellinger, R. Frontoni E. Real and Virtual Clinical Trials: A Formal Analysis. Topoi 38, 411–422 (2019).

https://doi.org/10.1007/s11245-018-9563-3

Keywords: Randomised Clinical Trials**,** In Silico Clinical Trials**,** Computational modeling and simulation**,** External validity**,** Extrapolation**,** Interpolation.

Abstract

If well-designed, the results of a Randomised Clinical Trial (RCT) can justify a causal claim between treatment and effect *in the study population*; however, additional information might be needed to carry over this result to another population. RCTs have been criticized exactly on grounds of failing to provide this sort of information (Cartwright & Stegenga 2011), as well as to black-box important details regarding the mechanisms underpinning the causal law instantiated by the RCT result. On the other side, so-called In-Silico Clinical Trials (ISCTs) face the same criticisms addressed against standard modelling and simulation techniques, and cannot be equated to experiments (see, e.g., Boem & Ratti, 2017, Parker, 2009, Parke, 2014, Diez Roux, 2015 and related discussions in Frigg & Reiss, 2009, Winsberg, 2009, and Beisbart & Norton, 2012).

We undertake a formal analysis of both methods in order to identify their distinct contribution to causal inference in the clinical setting. Britton et al.'s study (Britton et al., 2013) on the impact of ion current variability on cardiac electrophysiology is used for illustrative purposes. We deduce that, by predicting variability through interpolation, ISCTs aid with problems regarding extrapolation of RCTs results, and therefore in assessing their external validity. Furthermore, ISCTs can be said to encode "thick" causal knowledge (knowledge about the biological mechanisms underpinning the causal effects at the clinical level) – as opposed to "thin" difference-making information inferred from RCTs. Hence, ISCTs and RCTs cannot replace one another but rather, they are complementary in that the former provide information about the determinants of variability of causal effects, while the latter can, under certain conditions, establish causality in the first place.

1 Introduction

Systems biology is an approach to the understanding and conceptualization of the biological realm that emphasizes systemic and holistic aspects rather than reductionist and mereological features (Bertolaso & Ratti, forthcoming). Still, crucial to systems biology seems to be the possibility to single out specific causal links in the working of the

cell or of the genome and of putting it at work in isolation, within a tractable and controlled system that allows manipulability and repeatability, configuring multimodal research strategies (Macleod & Nersessian 2013). While 'wet' experiments to this aim are often possible, the use of computer simulations and of quantitative modelling methods borrowed from physics, offer epistemic advantages in terms of controllability, reproducibility, and analyticity (Morrison, 2015).[1]
However, the status of systems biology within the epistemic community of the biological sciences is a long-standing issue. One important aspect is the very possibility of building abstract mechanistic models of biological phenomena (Bertolaso & Campaner, forthcoming). As Keller (2003) famously articulated, models in the biological sciences – such as "animal models" – have traditionally been constructed to be *homologous* to the system they aim to capture, model organisms being a case in point. Mathematical and computational models of biological systems, on the other hand, bear a different kind of relation to the modelled object, namely one of *analogy*: such models aim at describing the behaviour of the system by capturing the underlying mechanistic principles and the abstract mathematical laws of its functioning.[2]

It has been proposed that computer simulation and computer-aided modelling techniques could be employed in the setting of clinical testing, as it is already happening in several other mission-critical domains (surgical intervention, adjustment of prosthetics, etc.), in order to support the planning of clinical trials, refine their conduct and reduce the possibility of their failure. What is envisaged is "the use of individualized computer simulation in the development or regulatory evaluation of a medicinal product, medical device, or medical intervention" (Viceconti et al., 2016, p. 8). This definition has been coined by the *Avicenna alliance*, a European focus group born with the purpose of fostering the adoption of computational modelling in the context of clinical testing. Patient-specific computer models should be used to generate simulated populations, on which new biomedical products can be safely tested. The Alliance refers to this methodology as *In Silico Clinical Trial* (ISCT), a pillar of the more general movement of in silico medicine (Bertolaso & Macleod 2016). In their recently published *Roadmap* (Viceconti et al., 2016), the Avicenna alliance produced an in-depth examination of the scientific, technological and societal challenges that have to be overcome in order to establish a role for the ISCT in medical research.

[1] For the sake of clarity, we will distinguish the two terms as follows: by "computer model" we mean the algorithm or computer program that is built in order to capture the mechanism of the phenomenon under study; by "simulation" instead we mean the actual run of the program.

[2] Keller's stance has been questioned by Rowbottom (2009) – by pointing out the importance of analogy also in extrapolation of results from animal models to humans (or from species to species, more generally). However this misses the point of distinguishing the two kinds of inference in that, even if analogy is also used in inference from animal models, it aids the extrapolation in a different way. Analogical inferences from organ system to organ system rely on ontological assumptions regarding the affinity of various biological species. Instead, computational models are intended to reproduce the target system by modelling its hypothesised underpinning structure; therefore analogical inference is based here on isomorphism (structural similarity).

[2] Keller's stance has been questioned by Rowbottom (2009) – by pointing out the importance of analogy also in extrapolation of results from animal models to humans (or from species to species, more generally). However this misses the point of distinguishing the two kinds of inference in that, even if analogy is also used in inference from animal models, it aids the extrapolation in a different way. Analogical inferences from organ system to organ system rely on ontological assumptions regarding the affinity of various biological species. Instead, computational models are intended to reproduce the target system by modelling its hypothesised underpinning structure; therefore analogical inference is based here on isomorphism (structural similarity).

With the present paper we intend to provide an analysis of their epistemological status, in particular with respect to the gold standard instrument of clinical investigation: Randomized Controlled Trials (henceforth RCT). To illustrate the methodology, we give a detailed account of the study done by (Britton et al., 2013) below (see also Carusi, 2014) for a discussion of this study).

The paper is structured as follows: In the next section we present the rationales, and the epistemic value of randomised controlled trials (RCTs) and the criticisms addressed against them: 1) black-boxing heterogeneity and variability; 2) providing a "thin" account of causation, which neglects the mechanisms underpinning the causal associations they aim to undercover. Next we present Nancy Cartwright's account of the epistemic contribution of RCTs. This provides the basis for a formal analysis of "in silico clinical trials" (ISCTs), which we undertake in the next sections. After illustrating a case study in cardiac electrophysiology, we identify the specific epistemic contribution of RCTs and ISCTs and conclude that they have complementary functions. Whereas RCTs establish causality, in that they come with a warrant of "causal sufficiency" (having taken into account any possible latent cause, or neutralised it), ISCTs cannot provide such a warrant in that they cannot guarantee that all possible relevant causal factors have been taken into account in the model, however, they can provide information about of the variability of the causal effect established through an RCT in different contexts. In the concluding section we elaborate on this distinction and present two main caveats to ISCTs methodology: its heterogeneous robustness in different contexts, and relatedly, its scope of implementation.

2. Randomised Controlled Trials: rationales, criticisms and epistemic value

Pharmacology is an intrinsic interdisciplinary science: it does not investigate a specific level of reality such as physics, or biology, but rather works across levels. Whereas the direct domain of action of drug molecules is limited to protein receptors, the desired end-effects are clinically observable results. However, because the proteins with which the drug molecules interact are embedded in a web of - possibly nonlinearly interacting - biological pathways (metabolic, genetic, signal transduction), most end-effects are unpredictable.

Knowledge of these various interactions and the *biological laws* governing them, as well as the contingent *initial conditions* holding in any specific context is far from being exhaustive to allow reliable prediction or causal inference. Hence, until recently, drug approval has mainly relied on a black-box methodology, grounded on a hypothetico-deductive method: null-hypothesis testing (see Landes, Osimani, Poellinger, 2017 for a detailed account of causal inference in pharmacology).

However, philosophers have raised objections against the privileged role accorded to RCTs as an instrument for causal inference in medicine, mainly on two grounds: 1) the neglect or downplay of *context-sensitivity* of causation by way of black-boxing the above mentioned interacting factors, and related issues with external validity (that is, the extent to which conclusions drawn on the basis of the study sample can be extended to the target

population/individual); 2) black-boxing of the *mechanisms* underpinning the cause-effect relationship and thereby again failing to take into account relevant information for the purpose of extrapolation and interpretation of results.

In the following we present such criticisms in more detail, illustrate Cartwright's analysis of RCTs as a basis for a formal analysis of their epistemic value, and then explain the virtues of randomization with respect to the problem of "causal sufficiency".

2.1 Randomisation and variability

The rationale underpinning RCT methodology is the maximisation of internal validity (i.e. avoidance of systematic error produced by confounding and biases), (see Cartwright 2011, Osimani 2014), often at the expense of external validity, in that studies conducted under stricter conditions are generally less representative of real-life conditions. More generally, RCTs are considered to black-box variability (or to abstract from heterogeneity); that is, the existence of statistically and clinically relevant differences among individuals or groups which interact with the treatment effect tends to be pushed to the background.[3] Indeed, random allocation[4] of the treatment to one of the two experimental groups serves the purpose of obtaining two (probabilistically) balanced groups as to all possible *known and unknown* "confounders" or "moderators", that is, to isolate the putative causal chain from drug to end-effect from other possible causes of the same effect, and, more generally, from the causal web in which it may be embedded.[5]

[3] Heterogeneity is explicitly taken into account in the evaluation of meta-analyses, where it is used to up/downgrade study quality, but it is rarely considered in its own right. However, heterogeneity is one of the main issues when predicting the effects of drugs in specific individuals or target groups, given that they are the result of causal interaction between the drug and various combinations of triggering factors, which may not be equally represented in study and target populations or the individual user.

[4] Random allocation of the treatment to the experimental group should be not confused with random sampling. This refers to the sampling procedure and is aimed at guaranteeing a representative sample with respect to the population from which the study sample is drawn. Hence, whereas the purpose of random sampling is to have a study population as close as possible to the sampled population, the goal of random allocation of treatment to the experimental group is to obtain two groups ("treatment" and "control" group) as close as possible to each other, except for the treatment itself. This should guarantee that the possibly observed difference in the outcome is due to the treatment and only to it.

[5] Random allocation of the treatment to the experimental group has putatively two main roles: 1) in the long run, it should allow the investigator to approach the true mean difference between treatment and control group; however it is unclear what this true underlying population probability denotes when we are dealing not with population of molecules for instance, but with population of patients undergoing medical interventions, where heterogeneity among individuals can at most allow for an aggregate average measure. Furthermore, it is obviously unethical and unfeasible to re-sample the same subjects of an experiment again and again, and even if this were possible, the subjects who were administered the drug in the first round would undergo physiological change; consequently, the successive trial population would no longer be the "same" (Worrall 2007); 2) random allocation (together with intervention and blinding) should guarantee the internal validity of the study by severing any common cause, or common effect, between the investigated treatment and its putative effects (i.e., avoidance of confounders and (self-)selection bias). This property is supposed to justify the primary role assigned to randomised evidence by so called evidence hierarchies (see Osimani, 2014).

This provides a (probabilistic)[6] warrant of "causal sufficiency" that is, that no latent variables are confounding the observed effect, and therefore that the detected difference-making relationship is a real one.

However, black-boxing heterogeneity has considerable consequences as to the validity of results in relation to predicting the *variability* of effects in the population of users, in that the effect (size) observed in the study may not be the same as the one that will be produced in specific individuals, or subgroups of the population.

Whereas classical experiments in physics allow the scientist to observe the behaviour of the investigated phenomenon in an array of different "possible worlds" (in different *scenarios* or under different *initial or boundary conditions*), and compare systematic differences among such situations, randomisation is blind to such specific settings. Its outcome is rather to neutralise their effects on the final result, by creating two populations, one for the treatment and one for the control group, where the same "worlds" should be represented in the same proportion. This should guarantee that the different results possibly observed at the end of the trial are due to the treatment and only to it (see also Landes et al., 2017, section 3.2). As we will see below, the purpose of ISCTs is exactly to reproduce the kind of setting where phenomena are systematically explored under different initial conditions and states of perturbation.

Random allocation of the treatment, as a means to isolate the putative cause from other possible confounding factors, loses much information with regard to the specification of possibly relevant mediating and interacting causes. As a consequence, the *exclusive* reliance on RCTs for the justification of causal claims is considered to be wrongheaded both because of the kind of information which they provide, and because of the kind of information which they are not able to incorporate or account for.

2.2 RCTs as instrumental to "thin" accounts of causation

Another reason why exclusive reliance on RCTs for causal inference has been criticized is that they produce only difference-making knowledge about causation (whether and to what extent the treatment produces the claimed effect), but no information about the underlying physiological mechanisms leading to such outcome. RCTs are thereby associated with a "thin" notion of causation, relying on mere formal relationships between cause and effect (as, e.g., captured in interventionist Bayes net accounts of causation – see Pearl, 2000 and Spirtes Glymour, Scheines, 2000), to be contrasted to a "thick" one, focusing on substantive knowledge about the web of entities and processes involved in bringing about the effects.

Indeed, knowledge of the mechanisms underpinning the effects observed in black-box studies has been invoked by philosophers has having various roles in assessment and prediction of treatment effects. The related debate has ontological, epistemological and

[6] That is, modulo the absence of random and systematic error.

methodological branches.

The ontological side refers precisely to the opposition between "thin" and "thick" concepts of causality. To the former belong definitions of causality based on difference making and counterfactual or regularity notions of causation (e.g., Lewis 1973a, b, 2000); to the latter belong instead process accounts of causality, i.e. accounts which appeal to *how* causes bring about their effects (e.g., Dowe 1992, Mumford 2009, Anjum & Mumford 2012). As Dowe puts it (Dowe 2000), this debate is also related to a distinction between causation as the analysis of truth conditions for causal claims vs. an objective feature of the world.[7]

From an epistemological point of view, Salmon (1997, 1984), following Reichenbach's (1956) ontological account of probability and probabilistic dependence, emphasizes that "mechanisms" provide the ontological explanation for observed regularities.

In the methodological debate around evidence standards, knowledge about mechanisms has been considered to complement knowledge about regularities: for instance Clarke et al. (2014) considers that evidence about difference making helps in de-masking causes which might be cancelled out by back-up/compensatory mechanisms in the organ system, whereas evidence about mechanisms is needed in order to design and interpret statistical studies. Hence, such different kinds of evidence reciprocally support each other and jointly (dis-)confirm the causal claim under investigation.

However, it is Nancy Cartwright who most closely addresses the status of mechanism knowledge with respect to RCTs. In her account, knowledge about mechanisms constitute the basis for extrapolation and the assessment of external validity, in that they guarantee that the same causal laws are working in the study and in the target population.

2.3 Epistemic analysis of RCTs according to the causal principle reading

According to the standard methodology of RCTs, the "treatment effect" is measured by the difference of the average outcome of interest in the treatment group minus the average obtained in the control group. There are two interpretations underpinning this approach, one is the counterfactual reading (also known as "potential outcome approach"; Rubin 2005, Holland 1986) and the other is the "causal principle" reading. According to the former, the treatment effect is defined as the effect that it would obtain in the subject if and only if she were administered the treatment. Since this is a counterfactual object of enquiry, and therefore cannot possibly be measured, a substitute measure for it is the difference between the two groups averages, under the assumption that no other differences affect the two groups apart from the treatment administration. Hence, the mathematical definition of treatment effect reads as follows:

$$T =_{def} < Y(u) \mid X(u)=x > - < Y(u) \mid X(u)=x' >$$

[7] See also Poellinger (forthcoming) for a discussion of the ramifications of theory choice in causal assessment.

where T is the treatment effect, x and x' denote the two possible values taken by the treatment variable X, $Y(u)$ denotes the value taken by the outcome variable Y for unit u, $Y(u)/X(u)=x$ denotes the value taken by the outcome variable Y for unit u given that the treatment variable X has value x, and the angled brackets mark that the subtraction is between the estimates (expectations) of the sample means in the two groups (treatment and control).

However, if one considers the causal laws underpinning both such counterfactuals and the statistical results of RCTs, one could interpret the observed treatment effect as following from the interaction of X with other contributing factors. The law underpinning the observed effect in relation to treatment X can be formalized as follows:

$$L: \quad Y =_c \quad a + \beta X + W$$ [8]

Then, the average value of Y in the treatment group would be measured by:

$$< Y(u) \mid X(u)=x > = < a(u) \mid X(u)=x > + < \beta(u) \mid X(u)=x > + < W(u) \mid X(u)=x >,$$

and, consequently, the treatment effect would be measured by the following equation:

$$T =_{def} < a(u) \mid X(u)=x > - < a(u) \mid X(u)=x' >$$

$$+ < \beta(u) \mid X(u)=x > x - < \beta(u)/X(u)=x' > x'$$

$$+ < W(u) \mid X(u)=x > - < W(u) \mid X(u)=x' >.$$

Since random allocation of X is meant to warrant that the expectation of a(u), β(u) and W(u) are the same *whatever value X assumes* (that is, that X is probabilistically independent from a, β and W), the first and last two terms in the equation cancel out, and the measure of the treatment effect given by an RCT results from the following formula:

$$T =_{def} < \beta(u) > (x-x').$$

The critical issue raised by Cartwright is that β represents all the combinations of factors that determine not only whether X contributes to Y (if β is 0, then this contribution is null); but also, how (positively or negatively) and to what extent. By drawing on the notion of causes as INUS condition (Mackie 1980), Cartwright also considers the possibility that β is a disjunction of sets of interacting factors:

$$\beta = f_1 (z_{11}, \ldots, z_{1n}) + \ldots + f_k (z_{k1}, \ldots, z_{km}).$$

So β can be any combination of factors (from macro-aspects such as age and health

[8] In analogy for instance to similar laws in macroeconomics, such as the *expectations-augmented Phillips curve* used to predict the rate of inflation at time t, given that a particular level of unemployment persisted for some time:

$$Y_t = b_0 + b_1 X_{2t} + b_2 X_{3t} + e_t$$

Where Y_t is the actual rate of inflation, X_{2t} the unemployment rate, and X_{3t} the expected inflation rate, at time t (see Hoover, 2008).

history, down to genetic make-up), and this means that the average measure provided by the result of RCTs can be considered as a good measure of the treatment effect in the target population (or individual), only to the extent that β takes the same value there, i.e., the target population of users is characterized by the same combination of factors. More fundamentally, being the effect size an average measure, the fact that this is co-determined by a proportion of co-factors black-boxed in the coefficient β, there is no way to estimate the effect size for any other population, or individual, (where another causal law may hold), from the study itself alone. Hence, not only RCTs are blind as to the relevant combinations of co-factors, which allow X to produce the intended effect, but also as to the causal principles (and related co-factors) involved in producing effects.

This is also critical for the assessment of unintended harmful effects. In fact, the treatment X may produce, in combination with specific subsets of factors other effects as a result of other biological laws in which X also plays a role; e.g. effect Q resulting from another causal principle:

$$L\text{:}\ Q_c = c + \beta' X + U$$

where Q may also be an undesired side-effect of the treatment X, X is again the treatment variable; β' is the set of combinations of factors that contribute to bring about the side-effect Q by interacting with X – these obviously need not be necessarily the same as those represented by β, whence the denotation of β'. U and c are the error term variable and a constant respectively.

Possible ways to partly remedy this problem and identify moderators of causal effects are stratification (subgroup analysis) and adjustment. However, several problems constrain the adoption of such measures *de dicto* and *de facto*. *De dicto*, because the possible combinations of co-factors are not known in principle, *de facto*, even if they were known, very large samples would be needed to have sufficiently powered subgroups, for the analysis of the moderating effects of such combinations of co-factors.

ISCTs may be considered an answer to this problem. Computational modelling has been recently used to develop virtual populations in order to explore interacting effects of drug in combination with various mediating factors. That is, to look into the black-box and explore the space of possibilities for our β term. ISCTs in particular are considered to supplement RCTs with such information. Are such claims warranted? What is the epistemic status of ISCTs (in comparison to RCTs)?

In the following, we use Cartwright's analysis of RCTs as a benchmark to evaluate the specific epistemic contribution of ISCTs and to determine both what kind of information they provide, and what kind of knowledge input they need in order to deliver their results.

3 In Silico Clinical Trials and their epistemic status

At the roots of the ISCT enterprise is the Virtual Physiological Human (VPH) project.

The VPH is an effort to replicate and model relevant bits of the human physiology *in silico*, by using quantitative data gathered through advanced imaging and sensing techniques to build computer models targeting the context of a specific disease (quite in line with the aims of systems biology, cf. Wang et al., 2015).[9] The model must contain, quoting the Roadmap, "mechanistic representation of the pathophysiological processes at play" (Viceconti et al., 2016, 81). The roadmap describes how such models are built: the process of "Knowledge-based disease model design [...] starts with an extensive review of the scientific literature to identify relevant pieces of knowledge describing the various mechanisms thought to play a part in the pathophysiology (e.g., inflammation, cell adhesion, apoptosis, etc.). Hence, the model is a product of "knowledge integration" with the purpose of exploring the behaviour of organ systems and their reactions to perturbations. The idea behind *in silico* clinical trials is that of exploiting a VPH knowledge base for the construction of a population of virtual models representing individual organisms reacting idiosyncratically to the drug or other interventions.

However, it is clear that, in order for the concept of the VPH (and the ISCT) to make sense at all, one must think that the relevant physiological processes and phenomena can be extracted from the organism and that they can be abstracted in terms of a computer algorithm. Hence, the question arises as to their epistemic contribution with respect to traditional experimental methods and their methodological validity.

We address in the present paper the former question and undertake a formal analysis of the epistemic status of ISCTs, with a special focus on their complementary output with respect to RCTs. The main output of this analysis is that ISCTs provide information exactly about, at least some components of the vector of co-factors (our β term), which jointly moderate the causal effect of the treatment under investigation, and which differs from population to population, or better, from individual to individual.

To this end we draw on Britton et al.'s study (Britton et al. 2013) on the impact of ion current variability on cardiac electrophysiology, and in particular on curvature and duration of the Action Potential (AP).

3.1 Case study: variability in cardiac cellular electrophysiology

Britton et al. (2013) set out from a computational model of the cellular ion channels in a particular tissue of the heart (the Purkinje fibre) and of their conductivity properties (which is a combination of the CGR model (Corrias, Giles, Rodriguez 2011). The model

[9] Clearly, the simulation of a complete human physiology (the 'virtual patient') remains a far-fetched objective of the VPH project. What is possible however is the simulation of bits of human physiology, aimed at reproducing the possible response of the targeted organ or system to a new intervention. As an example, an early effort of this kind of complex computational model of physiology was a model of cardiac physiology aimed at simulating the electrical activity of the heart (Winslow et al., 2002). Winslow and colleagues used different kinds of data –imaging techniques and measurements of electrical activity– in order to construct a full three-dimensional model of the cardiac ventricle. See also the case study illustrated below.

of rabbit Purkinje electrophysiology was developed by integrating experimental data (voltage clamp experiments and AP recordings, see Corrias et al. 2011) in a system of equations.[10]

Figure 1 shows a schema of the rabbit Purkinje cell model, its mechanistic ingredients, and a diagram of their interaction in the calcium handling subsystem. The mechanistic knowledge underlying this abstract visualization is encoded in a computational model (written in C++ code and based on Chaste - Cancer, Heart and Soft Tissue Environment), that is a simulation package aimed at multi-scale, computationally demanding problems arising in biology and physiology.[11]

The investigated hypothesis is that variability in APs is caused by quantitative differences in the properties of ionic currents (rather than by qualitative differences in the biophysical processes underlying the currents). As a further result of such exploration, the study identifies the main co-determinants of variability in the prolongation of AP duration (APD) under the action of the potassium channel blocker[12] Dofetilide, at four different concentrations.

The research built on previous studies, by the same and other authors, that used mathematical methods in order to carry out sensitivity analyses on the ionic determinants of inter-subject variability in AP morphology and duration, on the basis of established cardiac models (in particular the rabbit Purkinje electrophysiology) (2-4: Romero et al. 2009, Sarkar et al. 2012, Davies et al. 2012).

[10] Hence, the proposed model is based on a deterministic approach with no probabilistic or machine learning methods.

[11] Current functionality includes tissue and cell level electrophysiology, discrete tissue modelling, and soft tissue modelling. The package is being developed by a team mainly based in the Computational Biology Group at the Department of Computer Science, University of Oxford, and the development draws on expertise from software engineering, high performance computing, mathematical modelling and scientific computing.

[12] To be precise, Dofetilide is a blocker of the apid component of the delayed rectifier potassium current I_{kr}. The authors oft he study explicitly chose this kind of intervention because I_{kr} block is the main assay required in safety pharmacology assessment, due to ist importance in long QT-related arrythmias (Britton et al. 2013, E 2099).

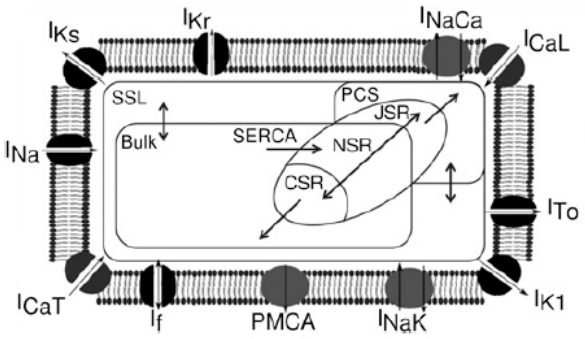

Fig. 1: Schema of Rabbit Purkinje cell model with twelve ionic currents (labels around the cell) and the calcium handling subsystem (depicted inside the cell). [Britton et al., 2013, Fig. 9]

The study consisted in computationally generating individual AP models out of the general one (fig. 1), by means of varying the values of relevant parameters of ionic currents in order to represent natural variation in the virtual population. The "individual models" here represent instantiations of the general model on which the simulation is based. Other than in the domain of computational modelling, where simulation is used to test the empirical adequacy of the source model with respect to the target system, within the bound of some natural variability, here the aim of the simulation in ISCTs is rather to explore such variability, i.e., to identify the determinants of the different behaviours the system instantiates depending on the different combinations of input parameter values (whose interaction is encoded in the mechanistic model).

All 10,000 models in Britton et al. study are generated from the same system of equations (i.e., the same ionic physiological processes), derived from a refined version of the rabbit Purkinje AP-Corrias-Geles-Rodriguez (CGR) model, each with different parameter values for the ionic properties, randomly selected within a wide range.

Each model returns a specific AP curve, typified on the basis of biomarkers, derived from electrophysiologic and fundamental biochemical knowledge, including natural constants (such as the Faraday constant and the Gas constant), constants taken from the rabbit Purkinje cell model (calcium velocities etc.), stimulus-related constants (such as amplitude and duration), and the system of equations describing the mechanistic interaction of all these ingredients with the input parameters (the set of ionic properties).[13]

[13] See, e.g., Carusi et al. (2012) for a systematic overview over the processes and constituents involved in the construction of multiscale models of cardiac electrophysiology.

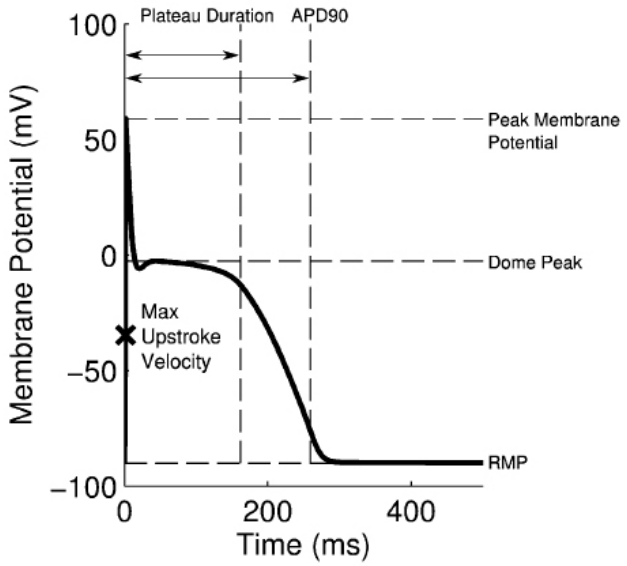

Fig. 2: AP curve with the six chosen biomarkers: dV/dt Max (max upstroke velocity), Vm Peak (Peak membrane potential), APD_{90} (Action Potential Duration), RMP (Resting Membrane Potential), Plateau Duration and Dome Peak. [Britton et al., 2013, Fig. 10]

The generated population of 10,000 AP curves (individual models) was then pruned, by comparing simulation outputs with experimental data of cellular biomarkers extracted from AP recordings at three pacing frequencies (0.2, 1 and 2 Hz), and by keeping only those models (213 in number) that were fully consistent with the experimental dataset (model calibration). Full consistency means here that among all models of AP morphology an duration produced by all possible combinations of ionic currents parameters (10,000), only those "models" were retained, whose six AP biomarkers values were *each* within the range determined by traces of AP from experiments on rabbit Purkinje fibres.

The study delivers three results:

1) A partial correlation analysis reveals which ionic properties determine the shape, amplitude, and rate dependence (0.1 Hz, 1 Hz, 2 Hz) of specific action potentials (this amounts to the identification of the "causal laws" linking ionic current properties and AP properties);

2) Prediction of APD prolongation caused by the I_{kr} channel blocker drug dofetilide (in four concentrations: 0.001, 0.01, 0.05, 0.1 μM (perturbation of the causal laws and prediction of related outcomes);

3) specific components that influences the outcomes of the perturbation (identification co-factors: components of β)

We briefly describe them in the following and then use them as a basis for our analysis of the distinct epistemic contribution of ISCTs with respect to RCTs.

1. Partial correlation analysis reveals which ionic properties determine the shape, amplitude, and rate dependence of specific action potentials. A partial correlation analysis was used as a sort of model fitting procedure, in order to show that different combinations of ionic properties determine different profiles action potentials (in terms of shape, amplitude and rate dependence). From this it follows that the variability of the AP curve can be predicted by the different values of the ionic currents parameters. Although the profile of each AP biomarker is influenced by many ionic currents parameters, still "individual parameter values are important for determining the exact balance of currents and therefore, the specific AP properties of each model (the investigated outcome encodes a vector of biomarkers jointly describing AP morphology and duration: figure 2 illustrates the six biomarkers and how they describe the AP curve).
The procedure is hypothetico-deductive, in that the cell-model arises from the integration of the various pieces of empirical knowledge formalised in mathematical formulas, and then statistical associations are searched between the parameters of individual ionic curves of the model and the biomarkers determining the different AP curves produced in the simulation. However, the partial correlation analysis delivers the net influence of *each* ionic current parameter on *each* AP curve biomarker, rather than delivering the *combined influence of the all* ionic current parameters on the entire curve profile. Hence this procedure is a sort of piecemeal model-fitting, and leaves room for uncertainty regarding the joint effect of such parameters on the AP curve. This question is in part and indirectly answered through "intervention" in the model, that is by simulating the effect of a channel blocking drug on one of the ionic channels, and in observing the resulting AP curves in the "treatment" and "control" groups. Treatment consisted in the administration of the potassium channel (I_{kr}) blocker drug dofetilide.[14]

2. Prediction of APD prolongation caused by the I_{kr} channel blocker drug dofetilide. The drug is *virtually* "administered" in four different dosages in the population of models (0.001, 0.01, 0.05, and 0.1 μM), in order to investigate the interaction of the ionic current variability with the drug dosage. If we let the drug administration correspond to our *X* in the causal principle interpretation of RCTs, and the outcome variable *Y* correspond to the shape of the AP curve, the simulation of "control" AP curves and "treatment" AP curves (that is AP curves resulting from blocking the potassium channel I_{kr}) shows that such intervention induces both APD prolongation, and increased APD variability. This is in agreement with previous experimental studies. This agreement "validates" the model and, consequently the empirical adequacy of the system of equation the constitute it.

[14] The choice of I_{kr} block as an intervention to evaluate the predictive power of the population of models is motivated by the fact that I_{kr} block is the main assay required in safety pharmacology assessment, due to its link to QT-related arrhytmias.

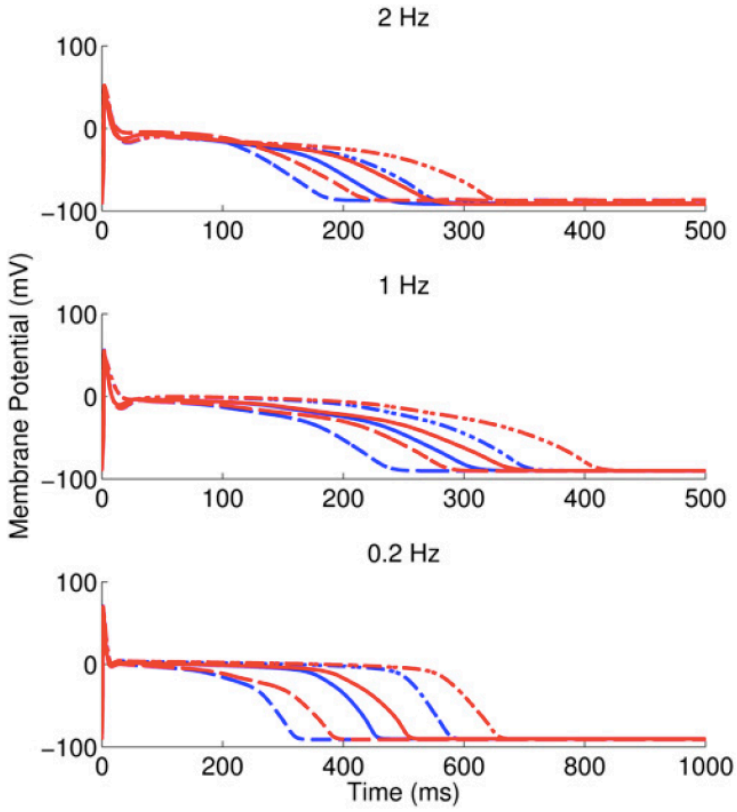

Fig. 2: Simulated AP traces obtained for three representative models accepted in the population in control conditions (blue) and following application of 0.01 μM concentration of dofetilide (red) at 2-, 1-, and 0.2-Hz pacing frequencies. This is the concentration closest to the experimentally determined IC50 (therapeutic dose) for dofetilide (0.0124 μM). Plots extend to 500 ms for 2 and 1 Hz and 1,000 ms for 0.2 Hz. Line style indicates which of the control and dofetilide traces correspond to each model. [Britton et al., 2013, Fig. 6]

3. Specific components that influences the outcomes of the perturbation. Against the background provided by such results, the most interesting finding derives from a further partial correlation analysis: the one examining the ΔAPD variable (that is, the difference in AP duration between "treatment" and "control" in the simulation: "treatment APD" – "control APD") against each of the 12 parameters that were varied to create the population of models (each time controlling for the other 11 parameters). This result is clinically interesting, because it identifies the main moderator(s) of heart arrhythmias when the drug is administered: at all pacing frequencies (0.2 Hz, 1 Hz, and 2 Hz), the scientists registered a strong positive correlation between G_{kr} (the maximal conductance of channel Kr) and the difference between AP duration under "control" vs. after I_{kr} block: "Rabbit Purkinje cells with a larger Gkr are more dependent on the I_{kr} current for repolarization and so will have a greater increase in APD 90 following I_{kr} block than cells with a smaller G_{kr}" (Britton et al. 2013, E2101).

Validation of such results is often implicit and tacitly relies on no miracles arguments (see Sprenger, 2016, for a probabilistic defense of such kinds of higher order evidence, and related literature) relating the strong systematic analogy of simulation and experimental results under different initial conditions and states of perturbation. This systematic analogy would be surprising if not explained by the fact that the mathematical laws underpinning the simulation at least partially reliably represent the mechanisms underpinning the experimental phenomena. However, we set this issue aside in the present paper and focus our analysis on the distinct informative contribution of ISCTs with respect to RCTs, drawing on the present case study.

3.2 From Laws to Models: peering into the blackbox

Since Rubin (1974), the standard conceptualisation of causal claims resulting from RCTs (and comparative studies) is counterfactual: the "causal effect" is the difference between what would have happened to the subject, had it been exposed to the treatment and what would have happened to it, had it been exposed to the control. Since the subject cannot undergo the same experimental conditions at the same time, the causal effect is calculated as the average difference of the effects observed in the group of exposed and the group of unexposed subjects.[15] What ISCTs promise to give, is precisely the "counterfactual information" about each possible individual either exposed to the treatment or not. A simplified representation of an individual model could look like this:

$$i = \pi\,(u, x),$$

where an individual i is described by a set of functions π (usually designed and coded in a program, encoding the physiological relations underlying the mechanistic model in a deterministic or probabilistic way) for a certain combination of relevant parameters u (a vector of constitutive parameter values) and a possible intervention x.[16]

Since blindly reshuffling the u vector in a combinatorial way (such that the components are varied modularly) might produce individuals that do not (or cannot) occur in empirical studies, the set of possible combinations of parameters (individuals) is narrowed down through calibration:

$$I_{cal} = \{\, i \mid i \text{ is empirically attainable}\}$$

The calibrated class I_{cal} allows one to infer the admissible initial conditions p (those units that are compatible with empirical data) and to learn about their variability.

If the treatment X is an exogenous variable in the causal model encoded in π, it is now possible – in obvious contrast with RCTs – to compute the effect of this treatment *for each individual*, such that

[15] This is also known as the "potential outcome approach" to causal inference.
[16] Plus possibly the addition of some bounded random error to express finer-grained natural variation in the sense of a Monte-Carlo simulation.

$$T = \Delta\,(i_x\,,\, i_{x'})\,,$$

where x indicates the treatment, x' the default or control, and Δ the quantitative or qualitative difference between individual under control vs. treatment (e.g., shape variations of AP curves).[17] In ISCTs, comparing control and treatment is precisely done for the same co-factor set. I.e., drawing on Cartwright's notation,

$$\Delta\,(i_x\,,\, i_{x'}) = (x - x')$$

without potential β interactions, since the instantiation of co-factors $< \beta(u) >$ is indeed kept fixed for i_x and $i_{x'}$.

Hence the specific contribution of ISCT methods is ultimately to use already established knowledge about the physiological mechanism for the formulation of π (i.e., the mechanistic interaction of p components), together with experimental findings, in order to restrict the parameter space to experimentally attainable value combinations of u (through calibration of the outcome space). Built on such a knowledge base, the computational model is then used to predict variability of the outcome as a function of the interaction of different instantiation of β vectors with the intervention variable X.

Note however, that to be precise the set of factors considered in the model is by no means exhaustive with respect to the phenotypic effect of the drug at the clinical level. Although the cell-model is a good predictor of the drug effect at the cell-level, it obviously fails to provide information for other kinds of causal contributors. In this sense, it provides only one set of components of the β vectors, which possibly comprises also other sets of relevant phenomena (genetic make-up, clinical history, possible co-morbidities, age, etc.).

More importantly, ISCTs cannot establish causation by themselves in that they bring no guarantee that all possible latent causes have been taken into account, or neutralised, not even within the limits of the biological level modelled (here for instance, the cell). The model comes with no guarantee that the system of equation representing the cell is exhaustive and that the functional forms are all correct. Indeed this is exactly the point of running the simulation and comparing the "performance" of the model with empirical data.
Instead, RCTs can, at least in principle, establish whether X causes Y, because they satisfy the requirement of causal sufficiency, that is, precisely the fact all possible latent causes have been taken into account, or neutralised.
However, ISCTs provide information as to variability of the outcomes in a heterogeneous population.

Hence the two methods are complementary in that RCTs establish causation, but black-box the set of co-factors that contribute to variability, whereas ISCTs cannot establish

[17] Averaging would now be possible, but an aggregate effect size is not the goal of the investigation here, since the value of the functional model lies precisely in the fact, that its mechanistic core is able to treat input vectors *individually* and generate predictions that are sensitive to the input details.

causation, since they cannot warrant causal sufficiency, but they can predict variability of causal effects on the basis of known interacting factors.

For this reason, validation of the model with experimental data, provides confirmatory support both to the hypothesized mechanism that informs the simulation, and to the mathematical equations that constitute it: the systematic analogy of results in simulations and experiments provides indirect confirmatory evidence to the hypothesis that the mathematical equations representing the mechanisms underpinning the investigated phenomenon are empirically adequate (no-miracles argument).

The information provided by RCTs and ISCTs is therefore complementary both in the sense that ISCTs provide the information on variability, which is necessarily neglected in RCTs, and in the sense of ensuring that knowledge about mechanisms (thick causation) is incorporated in the prediction of variability

As observed by Carusi, "validation shifts from being about how one model matches up to experiment [...] to how the range of variability in an experimentally calibrated population of models maps onto the range of variability in an experimental data set. " (Carusi, 2014). Population variability becomes part of the model and of the model validation process.

4 Summary and Conclusion

Our epistemic analysis shows that ISCTs and RCTs cannot replace each other, but rather, that they are complementary: while the former provide information about the determinants of variability of causal effects, the latter can, under certain conditions, establish causality in the first place. We elaborate on the implications of our analysis in the following.

4.1 Extrapolation and interpolation

The aim of Randomized Clinical Trial is to uncover difference-making relationships that can possibly be transferred to a population of future drug users. If well-designed, the results of an RCT can justify a causal claim between treatment and effect *in the study population*. Additional information might be needed to carry over this result to another population.[18] This sort of information corresponds to the β factor in Cartwright's formal analysis of RCTs; RCTs have been criticized for failing to provide such information and therefore to lack a warrant for external validity (see also Cartwright & Stegenga 2011), as well as to black-box important information as to the mechanisms underpinning the causal law instantiated by the RCT result.

On the other side, as mere models, ISCTs face the same criticisms addressed against standard modelling and simulation techniques and cannot be equated to experiments (see,

[18] See also Poellinger (forthcoming) for an analysis of analogy-based inference in pharmacology.

e.g., Boem & Ratti, 2017, Parker, 2009, Parke, 2014, Diez Roux, 2015 and related discussions in Frigg & Reiss, 2009, Winsberg, 2009, and Beisbart & Norton, 2012). However, even if computer models do not go beyond the information, which the model was constructed upon (see Bertolaso 2013), once an ISCT model population is validated, one can dispose of knowledge about variability in a population. Such knowledge can be used to extend the boundaries of measurement beyond what would be possible with real populations, via interpolation.

Hence ISCTs, cannot establish new causal laws *per se*, but they can use information about already established causal laws in order to inform prediction about variability of outcomes generated by the interaction of such laws with different initial conditions of the relevant interacting factors. By predicting variability through interpolation, ISCTs aid with problems regarding extrapolation of RCTs results to other populations than the one on which the trial was carried out, and therefore in assessing its external validity. The prediction of the determinants of variability through interpolation straightforwardly provides information about what kind of outcomes can be expected in specific populations or individuals as a function of such determinants.

4.2 Thick causal hypotheses, again

In our electrophysiology example above, we emphasized the fact, that mechanistic knowledge represents an important prerequisite for the generation of a virtual model population. The case study used biochemical knowledge drawn from the rabbit Purkinje cell model in order to compute AP curves from a set of mechanistically interacting ionic properties.

In that sense, ISCTs can be said to *encode* thick causal knowledge – as opposed to thin difference-making information *inferred* from RCTs. If the model is validated in comparison with empirical data (to a chosen degree of accuracy), the mechanistic model (i.e., the mechanistic causal hypothesis) underlying the computer program can be said to be confirmatorily supported. If empirical data raise doubts as to the validity of the mechanistic model, the researchers have to decide whether to dismiss the model or refine the mechanistic knowledge base in details. Entrenched parts of the mechanistic model will remain intact (e.g., natural constants and well-tested functional dependencies), others will be corrected or removed (see also Osimani and Poellinger, forthcoming).

It is important to note that, necessarily, the computational model consists of "thin" mathematical equations – it is only that these thin equations jointly refer to the complex interacting system they encode. In that sense, ISCTs can be said to encode a *thicker* understanding of the causal associations, when contrasted with RCTs. *Focusing on one specific portion of Cartwright's β term*, they can be understood as supporting a specific how-possibly explanation for both the individual outcomes *and* their variability.

4.3 Caveats

The presented case study adopted a deductive approach: first construct the cell-model on the basis of available empirical knowledge and then fit the data to the model to test its validity. However, regarding our epistemic analysis, analogous conclusions could be applied if inductive approaches were used, including machine-learning approaches.[19] A fundamental caveat concerns instead the knowledge base on which the model is constructed, and especially the kind of phenomenon that is modelled. Electrophysiology lends itself to be modelled mathematically relatively easily, because it regards reasonably well-known physical laws. However, other biological pathways may be much more challenging because of the complex interaction of chemical reactions, cell signalling and homeostatic mechanisms. Therefore, they are not amenable to mathematical modelling in the same way. Hence, the epistemic warrants of ISCTs may be much more fragile in these cases, and more general, vary substantially from context to context. As we show in this paper, a pre-requisite for an ISCT is a mechanistic understanding of the investigated system. Hence the predictive warrant of ISCTs is critically dependent on the stability of such knowledge. Moreover, since the majority of biological phenomena are of the latter kind, it is still to be judged how wide the scope of ISCT technology can be.

Acknowledgments

Barbara Osimani and Roland Poellinger gratefully acknowledge funding from the European Research Council (Grant Number: GA 639276).

References

Anjum, R.L., Mumford, S. (2012) Causal dispositionalism. In: A. Bird, B. Ellis, H. Sankey (eds.) Properties, Powers and Structure, Routledge, chap. 7, pp. 101–118

Barter, Z. E., Tucker, G. T., and Rowland-Yeo, K. (2013) Differences in cytochrome p450-mediated pharmacokinetics between chinese and caucasian populations predicted by mechanistic physiologically based pharmacokinetic modelling. Clinical pharmacokinetics, 52(12): 1085–1100.

[19] The main difference is that in some cases (i.e. neural networks) the model in not explicitly known by the users but is in the network learning phase. All the conclusions presented here are also valid, with different formulations on models for biological processes based on discrete mathematics (graphs, combinatorics) and probabilistic and optimization methods, such as Markov chains and Markov fields, Monte-Carlo simulation, maximum-likelihood estimation, entropy and information. Applications selected from epidemiology, inheritance and genetic drift, combinatorics and sequence alignment of nucleic acids, energy optimization in protein structure prediction, topology of biological molecules.

Beisbart, C., Norton, J. D. (2012) Why monte carlo simulations are inferences and not experiments, International Studies in the Philosophy of Science 26 (4) 403-422.

Bertolaso M, Giuliani A, Filippi S (2013) The Mesoscopic Level and its Epistemological Relevance in Systems Biology. In: Recent Advances in Systems Biology. Nova Science Publishers, Inc., pp. 19-36.

Bertolaso M (2015) Philosophy within Science, in: Bertolaso M. (Ed.) The Future of Scientific Practice: 'Bio-Techno-Logos'. Pickering & Chatto Publishers.

Bertolaso, M., and Buzzoni, M. (2016) Causality and Levels of Explanation in Biology, in Michele Paolini Paoletti and Francesco Orilia (eds.) Philosophical and Scientific Perspectives on Downward Causation, Part III - Downward Causation and the Sciences, Routledge.

Bertolaso, M., and Campaner, R. (2018) Scientific Practice in Modelling Diseases: Stances from Cancer Research and Neuropsychiatry, the Journal of Medicine & Philosophy (forthcoming).

Bertolaso, M., and Macleod, M., eds. (2016) In Silico Modeling: the Human Factor, Humana.Mente — Journal of Philosophical Studies, Issue 30 (April 2016).

Bertolaso, M., and Sterpetti, F. (2017) Evidence Amalgamation, Plausibility, and Cancer Research, Synthese, S.I. Evidence Amalgamation in the Sciences. https://doi.org/10.1007/s11229-017-1591-9

Boem F, Malagrinò I, Bertolaso M (2018), In Silico Clinical Trials: a Possible Response to Complexity in Pharmacology, in: Adam La Caze, Barbara Osimani (eds.) Uncertainty in Pharmacology: Epistemology, Methods, and Decisions. Boston Studies in the Philosophy and History of the Sciences, Springer.

Boem F., Ratti E., (2017) Toward a notion of intervention in Big-Data biology and Molecular Medicine. In: Boniolo G., Nathan M.J., (eds.) Philosophy of Molecular Medicine: Foundational Issues in Research and Practice, Routledge.

Britton, O. J., Bueno-Orovio, A., Van Ammel, K., Lu, H. R., Towart, R., Gallacher, D. J., and Rodriguez, B. (2013) Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. Proceedings of the National Academy of Sciences, 110(23): E2098–E2105.

Butler, D. (2008) Crossing the valley of death. Nature, 453:840–2.

Carusi, A. (2014) Validation and variability: Dual challenges on the path from systems biology to systems medicine. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 48:28–37.

Carusi A, Burrage K & Rodriguez B (2012) Bridging experiments, models and simulations: an integrative approach to validation in computational cardiac electrophysiology. AJP: Heart and Circulatory Physiology, 303(2), H144-H155.

Cartwright, N., & Stegenga, J. (2011) A theory of evidence for Evidence-Based policy. In

Dawid, P., & Twinning William Vasilaki, M. (Eds.) Evidence, Inference and Enquiry, chap. 11, OUP: 291–322.

Diez Roux, A. V. (2015) The virtual epidemiologist - promise and peril, American Journal of Epidemiology 181 (2), 100-102.

Dowe, P. (1992) Wesley salmon's process theory of causality and the conserved quantity theory. Philosophy of Science 59 (2):195-216.

Dowe, P. (2000) Physical Causation. Cambridge University Press.

Frigg, R., Reiss, J. (2009) The philosophy of simulation: hot new issues or same old stew?, Synthese 169 (3) 593-613.

Grüne-Yanoff, T. (2007) Bounded rationality. Philosophy Compass, 2(3):534–563.

Haddad, T., Himes, A., and Campbell, M. (2014) Fracture prediction of cardiac lead medical devices using bayesian networks. Reliability Engineering and System Safety, 123:145–157.

Holland, Paul W. (1986) "Statistics and Causal Inference". J. Amer. Statist. Assoc. 81 (396): 945–960. JSTOR 2289064. doi:10.1080/01621459.1986.10478354.

Hoover K. D. "Phillips Curve." The Concise Encyclopedia of Economics. 2008. Library of Economics and Liberty. 29 August 2017. URL: http://www.econlib.org/library/Enc/PhillipsCurve.html.

Huang, S. (2004) Back to the biology in systems biology: What can we learn from biomolecular networks? Briefings in functional genomics & proteomics, 2(4):279–297.

Hunter, P. J. and Borg, T. K. (2003) Integration from proteins to organs: the physiome project. Nature reviews Molecular cell biology, 4(3):237–243.

Keller, E. F. (2003) Making sense of life: Explaining biological development with models, metaphors, and machines. Harvard University Press.

Landes, J., Osimani B., Poellinger R. (2017) Epistemology of Causal Inference in Pharmacology. Towards a Framework for the Assessment of Harms. European Journal for the Philsophy of Science. DOI 10.1007/s13194-017-0169-1

Lewis D. (1973a) Counterfactuals. Oxford: Blackwell Publishers and Cambridge: Harvard University Press, 1973, Reprinted with revisions, 1986.

Lewis D. (1973b) Causation. The Journal of Philosophy, 70 (17): 556–567.

Lewis D. (2000) Causation as influence. The Journal of Philosophy, 97 (4): 182-197.
Mackie, J. L. (1980) The Cement of the Universe: A study of Causation. Oxford University Press.

MacLeod, M., and Nersessian, N.J. (2013) "Coupling Simulation and Experiment: the Bimodal Strategy in Integrative Systems Biology", Studies in History and Philosophy of Biological and Biomedical Sciences, 44(4, part A): 572–584. doi:10.1016/j.shpsc.2013.07.001

Morrison, M. (2012) Models, measurement and computer simulation: The changing face of experimentation. Philosophical Studies, 143:33–57.

Morrison, M. (2015) Reconstructing Reality: Models, Mathematics, and Simulations. Oxford University Press.

Mumford, S. (2009) Causal Powers and Capacities, chap. 12, pp. 265–278. In: Beebee, H., Hitchcock, C., Menzies, P. (eds.): The Oxford Handbook of Causation.

Noble, D. (1962) The voltage dependence of the cardiac membrane conductance. Biophysical journal, 2(5):381.

Osimani, B. & Poellinger, R. (forthcoming) A Protocol for Model Validation and Causal Inference form Computer Simulation, submitted to Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences

Parke, E. C. (2014) Experiments, simulations, and epistemic privilege. Philosophy of Science 81 (4), 516-536.

Parker, W. S. (2008) Computer simulation through an error-statistical lens. *Synthese*, 163(3):371–384.

Parker, W. S. (2009) Does matter really matter? Computer simulations, experiments, and materiality. Synthese 169 (3) 483-496.

Pearl, J. (2000) Causality: Models, Reasoning, and Inference, 1st Edition, Cambridge University Press.

Petersen, A. C. (2012) Simulating nature: a philosophical study of computer-simulation uncertainties and their role in climate science and policy advice. CRC Press.

Poellinger, R. (forthcoming) Analogy-Based Inference Patterns in Pharmacological Research. In: La Caze, A. & Osimani, B (eds.): Uncertainty in Pharmacology: Epistemology, Methods, and Decisions. Boston Studies in Philosophy of Science. Springer (forthcoming).

Poellinger R. (forthcoming) On the Ramifications of Theory Choice in Causal Assessment: Indicators of Causation and Their Conceptual Relationships. Submitted to Philosophy of Science.

Reichenbach, H. (1956) The Direction of Time. University of Los Angeles Press.

Rowbottom, D. P. (2009) Models in biology and physics: What's the difference? Foundations of science, 14(4):281–294.

Rubin, Donald (2005) "Causal Inference Using Potential Outcomes". J. Amer. Statist. Assoc. 100 (469): 322–331. doi:10.1198/016214504000001880.

Salmon W. (1997) Causality and Explanation: A Reply to Two Critiques. Philosophy of Science, 64: 461-477.

Salmon W. (1984) Scientific Explanation and the Causal structure of the world. Princeton, NJ: Princeton University Press.

Spirtes, P., Glymour, C., Scheines, R. (2000) Causation, Prediction, and Search, Adaptive Computation and Machine Learning. MIT Press, Boston.

Sprenger, J. (2016) Euro Jnl Phil Sci, The probabilistic no miracles argument 6:173–189.

Teira, D. (2011) Frequentist versus Bayesian Clinical Trials. In Gifford, F., editor, Philosophy of Medicine. North Holland.

Viceconti, M. (2011) A tentative taxonomy for predictive models in relation to their falsifiability. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369(1954):4149–4161.

Viceconti, M., Henney, A., and Morley-Fletcher, E. (2016) In silico clinical trials: how computer simulation will transform the biomedical industry. International Journal of Clinical Trials, 3(2):37–46.

Wang, R.-S., Maron, B. A., and Loscalzo, J. (2015) Systems medicine: evolution of systems biology from bench to bedside. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 7(4):141–161.

Winsberg, E. (2009) A tale of two methods, Synthese 169 (3) 575-592.

Winslow, R. L., Helm, P., Baumgartner, W., Peddi, S., Ratnanather, T., McVeigh, E., and Miller, M. I. (2002) Imaging-based integrative models of the heart: Closing the loop between experiment and simulation. In 'In Silico'Simulation of Biological Processes: Novartis Foundation Symposium 247, pages 129–143. Wiley Online Library.

Worrall, J. (2007) Evidence in medicine and evidence-based medicine. Philosophy Compass, 2(6), 981–1022. doi:10.1111/j.1747-9991.2007.00106.x.

Worrall, J. (2010) Do we need some large, simple randomized trials in medicine? In Suárez, M., Dorato, M., & Rédei, M. (Eds.) EPSA Philosophical issues in science: Launch of the European Philosophy of Science Association (pp. 289–301). doi:10.1007/978-90-481-3252-2 27.