CrossMark

Introduction: Foundations of Clinical Reasoning—An Epistemological Stance

Mattia Andreoletti[1] · Paola Berchialla[2] · Giovanni Boniolo[3] · Daniele Chiffi[4,5,6]

The present collection of essays entitled *Foundations of Clinical Reasoning: An Epistemological Stance* aims at shedding light on the epistemological approaches to *the foundations of reasoning in clinical research and practice*, and other related topics. Specifically, this issue of *Topoi* is intended to promote a new epistemological reflection on some foundational issues of clinical reasoning in research and practice. We focused on: (i) methodological validity and limitations of Randomized Controlled Trials (RCTs) and Evidence Based Medicine (EBM), (ii) varieties of clinical reasoning and values, and (iii) the Bayesian approach in clinical research and practice. While the collection by no means aims at covering all the main topics that have been discussed in the contemporary debate on the foundations of clinical reasoning, we hope nevertheless to introduce the reader to some of the most interesting and recently more discussed topics in the field.

✉ Daniele Chiffi
chiffidaniele@gmail.com

Mattia Andreoletti
mattia.andreoletti@gmail.com

Paola Berchialla
paola.berchialla@unito.it

Giovanni Boniolo
giovanni.boniolo@unife.it

[1] Department of Experimental Oncology, European Institute of Oncology, Milan, Italy

[2] Department of Clinical and Biological Sciences, University of Torino, Turin, Italy

[3] Dipartimento di Scienze Biomediche e Chirurgico Specialistiche, Università di Ferrara, Ferrara, Italy

[4] Ragnar Nurske Department of Innovation and Governance, Tallinn University of Technology, Tallinn, Estonia

[5] IFILNOVA Institute of Philosophy, Universidade Nova, Lisbon, Portugal

[6] Dastu, Politecnico di Milano, Milan, Italy

1 Validity and Limitations of RCTs

Over the last few decades of twentieth century, the complexity and the rigor of medical methods have increased exponentially. Over the years, medical doctors have faced the typical issues of philosophy of science, such as the relation between theories and experiments, the causal inference or the nature of explanation. This has unfolded an interdisciplinary dialogue between philosophers and physicians. Among the most discussed epistemological issues, the problem of the external validity might be considered as one of the most urgent as regards the clinical reasoning. This problem arises from the fact that the results of the Randomized Controlled Trials (RCTs) are not always applied to the whole reference population, which is identified with people needing a medical treatment. Following Campbell and Stanley (1966), the expression *external validity* "asks the question of generalizability: To what populations, settings, treatment variables and measurement variables can this effect be generalized?". However, the term 'generalization' referring to controlled clinical experiments might not be sufficiently precise to grasp the problem, because it is not only a matter of justifying an inference from an experiment conducted on a narrow number of patients, but also justifying the inference towards a single patient (i.e. particularization) (Rothwell 2005). Once verified that the result of an RCT is valid, we still have to explain how to apply this result to patients who did not take part in the experiment. As a matter of fact, several individuals who present features are excluded from the groups of patients selected for the experiment. This is because of different reasons, mostly ethical: patients who suffer from serious illnesses other than those targeted in the treatment (i.e. diabetes), the terminally ills and the elderlies (for whom the administration of an experimental treatment is considered too risky) are excluded. Things get even worse for when investigating of rare or chronic diseases. In the first case, it is very difficult to find a sufficient number of patients to enroll, in the second case, since the experiments only last

for a short time, RCTs cannot provide enough information about the effects of an extended treatment. Therefore, how can we justify the belief that a certain treatment has the same effect when applied to a different setting? Without a reasonable answer to this latter question, in theory RCTs would prove useless, as they would only show the results related to a situation, without any guarantee that the same results could apply to other contexts.

It is not just a philosophical matter, or a contemporary version of Hume's problem of induction, as the philosopher of science Worrall (2009) pointed out. In the case of RCTs we know that the context in which the treatment is applied will differ from that of the experiment. Then we should clarify the epistemic relevance of "context", taking in account which background conditions defining the context are relevant for the generalization of results (Cartwright 2011). In conclusion, an analysis of epistemological issues underlying the clinical experimentation (such as the external validity) looks fundamental to better understand the process of clinical reasoning.

2 Varieties of Clinical Reasoning and Values

In order to choose a suitable treatment for a patient, diagnostic and prognostic judgements are essential. An epistemological analysis of these classical themes of clinical reasoning is given by considering the recent logical and epistemological discussions on the role of hypothetical reasoning and on the connection between population-based knowledge and specific clinical cases.

2.1 Diagnosis and Clinical Reasoning

Diagnoses almost always occur under a veil of uncertainty, so that those who make diagnoses must develop advanced probabilistic reasoning skills, being a well-known fact that intuitive probabilistic arguments are very likely to be biased (Tversky and Kahneman 1974). Moreover, diagnoses are frequently based on diagnostic tests, which require statistical (and clinical) decision making. Assuming that the *null* hypothesis H_0 (the no-effect hypothesis to be tested) can be rejected, we can accept (in a binary context) an alternative hypothesis H_1, when the *p*-value (i.e., the probability of obtaining a result equal or more extreme than the observed one assuming H_0 is true) is equal to or less than a specific value of statistical significance α (generally identified with a probability of 0.01 or 0.05). A type I error occurs when H_0 is rejected although it is true. A type II error occurs when failing to reject H_0 that it is actually false. The probability of a type II error is denoted by *β*. In the diagnostic framework, if H_0 is the absence of disease and H_1 the presence of disease, then type I error is a false positive diagnosis, while the type

II error is a false negative diagnosis. Notice that it is important for an accurate clinical judgement to balance the thresholds for type I (false-positive) and type II (false-negative) statistical errors, i.e. the two main ingredients of what it has been called "inductive risk" (Hempel 1965; Parascandola 2010). A proper equilibrium should be achieved between statistical errors by also taking into account the sample and the effect sizes that one wants to detect, e.g., specific values for accepting or rejecting hypotheses (Cranor 1990). Probabilistic reasoning is often involved in diagnostic judgment. Also in virtue of this fact, a diagnosis is not normatively neutral, since it encompasses both value judgments (also with a non-epistemic nature) and critical assessment of evidence and of different types of probability. It is, in fact, usually assumed that a gap exists between the population-based evidence and the specificities of a particular clinical context. This poses the problem of interpreting different forms of probability and understanding which model of knowledge translation could take (quantitative) population-based evidence into account, and still be suitable for contingent clinical situations. In addition, a diagnosis may be incomplete in many ways–for instance, if the evidence is insufficient or problematic, if the evaluation is not correct according to some decision rules, etc.

Although many different types of clinical diagnosis can be established, diagnoses can be classified in two broad categories: *nosographic* and *pathophysiological* (Federspil 2004). Nosographic diagnoses aim to assign a patient's pathological condition to a specific category with a nosographic taxonomy, taking into account only population-based information. Recognizing the causes, explanations and mechanisms behind the individual's pathological state is not tantamount to formulating a nosographic diagnosis. However, pathophysiological diagnoses focus specifically on the past to shed light on how a patient's disease developed in a specific individual.

2.2 Prognosis and Clinical Reasoning

Prognostication is an action that has been somewhat left on the sidelines of current medical practice and it is sometimes erroneously conflated with the diagnostic process (Rich 2002). A prognosis is the end-result of a fundamental clinical assessment that focuses mainly on the future, and that is why it is considered more difficult than establishing diagnoses or making treatment decisions (Christakis and Sachs 1996). Even when a diagnosis has been soundly formulated, the prognosis may be uncertain because our understanding of pathological conditions is always incomplete, and diseases may develop in unpredictable ways. The aim of prognostics is to judge: (i) the duration of a specific future stage of a disease in a given patient (*prognosi quoad tempus*); or (ii) the probability of a patient recovering from

a disease (*prognosi quoad veletudinem*); or (iii) the chances of a patient's survival (*prognosi quoad vitam*) (Rizzi 1993). All these types of prognosis are made with a margin of uncertainty, and a prognosis is usually formulated as the likelihood of an outcome and/or of a patient's health taking a certain course (in terms of a particular disease) within a specified time interval (Djulbegovic et al. 2011). The matter is of the utmost importance to patients, and to physicians having to make appropriate, tailored treatment decisions.

Like diagnoses, prognoses may be classified within two main categories: *nosographic* and *pathophysiological* (Federspil 2004). A *nosographic prognosis* is based on a nosographic diagnosis and aims to predict when a given patient will reach a certain future stage of their disease. This prediction is based on statistical frequencies taken from population-based data. A *pathophysiological prognosis* is based on a pathophysiological diagnosis. The prediction of a future stage of disease in a given patient is grounded on. Moreover, there is an interplay between different types of probability (population-based frequencies or probability eliciting clinician knowledge), that may be connected, for instance, in a Bayesian framework (see the next Section).

However, it may be the case (in some difficult circumstances) that prognostic judgements face 'fundamental uncertainty' (also called as 'Keynesian uncertainty' (Keynes 2008)), since probability measures may be hardly assign to specific future events (Chiffi and Zanotti 2017). This is due to the fact that relevant prognostic information cannot be available when formulating a prognosis and some future scenarios associated with a patient's disease course might be previously inconceivable from a cognitive and methodological perspective. Despite this remarkable fact, prognostics is still classically based on the concept of probabilistic risk, without recognizing that it is almost often connected with events showing fundamental uncertainty. Some argumentative and abductive strategies have been explored in some of the papers of the present issue in order to make sense of those clinical judgements that may face fundamental uncertainty.

2.3 Bayesianism in Clinical Research and Practice

Statistical thinking has a central role in raising the scientific standards of clinical research and clinical trial has been a major field of innovation of statistical methodology.

Traditional frequentist statistics has been the dominant, and often exclusive, approach. The classical approach, as exemplified by Fisher's p-values, Neyman–Pearson hypothesis tests, and Neyman's confidence intervals, is a procedure for concluding the inference about the population parameters.

In contrast, the Bayesian approach is based on the learning from the experimental data and it is generally view as an inductive inferential approach of learning about the general from particulars (Savage 1954; Bernardo and Smith 1994). Rather than employing tests and attempting falsification, the available evidence is summarized by a posterior distribution, and the central goal of Bayesian inference is computing the posterior probabilities of hypotheses avoiding the null hypothesis paradigm and the logic of p-values.

Three points are important in Bayesian approach: (i) the prior distribution used to summarize the information known *a priori;* (ii) the likelihood to generate the posterior and (iii) distribution of the posterior. The relation between posterior, prior and likelihood is given by posterior proportional to likelihood × prior, according to the Bayes's rule.

Suppose the aim of a study is to determine the treatment effect to reduce the tumor growth in oncological patients. The distribution about the effectiveness of both the experimental and the treatment are required. This kind of information is considered as prior information and can be available from preliminary data or earlier studies.

The priors can be distinguished into informative or non-informative. The non-informative prior gives equal chances of all possible values to occur in the experiment. The informative prior is selective to be included in the sample and there are different types of priors, which can be used to carry out the analysis: conjugate prior distribution (family of distribution that yield to the get the posterior distribution which is still of the same family of the prior distribution); locally uniform prior (prior that changes over the region in which the likelihood is appreciable, and does not assume large values outside that range); location-invariant prior and scale-invariant prior (prior that is invariant to the scale of measurement).

The second point consists in formulating the experimental data as likelihood of the experimental treatment over the standard one. In this step the prior information is multiplied by the likelihood and the result is a distribution that can be used to get directly any information about the experimental treatment (the result observed through non-informative prior is image of the data alone).

In contrast, the reasoning behind the classical approach is not similarly straightforward. For example, when performing a statistical test, the aim is to know whether the observed difference can be explained by simple chance variation. If the test is statistically significant, the probability of getting a result equal or more extreme to the one obtained in the study (the classical p-value) under the null hypothesis of absence of any difference (i.e. by chance) is lower than the threshold of statistical significance. This probability is considered sufficiently small to conclude that the observed difference does not arise from chance. However, the p-values do not directly express the probability of the truth of the null hypothesis, i.e. that the observed difference can be explained by chance variation. Being a probability conditional on the null hypothesis,

they are a measure of the fit of the null hypothesis to the data, which is not always relevant to the aim of the research (Greenland et al. 2016). The same considerations apply to any equivalent procedure, like summarizing results by whether the confidence interval includes the value of the null hypothesis.

The conceptual model of the repeated sampling behind the traditional approach entails the issue of statistical multiplicity. While performing many statistical tests (i.e. when more than endpoint is going to be analyzed) the probability of making at least one type I error increases with the number of tests performed. These issues are not encountered in a Bayesian framework of data analysis since it works in the context of the data gathered from the actual experiment and not in the perspective of an infinite repetition of the same experiment. Despite these pitfalls of the traditional approach in addressing directly the research questions, the Bayesian approach is less common in clinical research since it does not easily lend itself to an oversimplified decision-making system, being less suited to a final judgment of true/false type. Regulatory agencies, especially for phase III pivotal trials, necessarily need to draw true/false conclusions about the superiority of an experimental therapy.

A side effect of the frequentist rigour and narrowness of focus to the experiment at hand is the inflexibility, which can limit innovation especially in the design and analysis of clinical trials. Therefore, clinical trials tend to be overly large, increasing the cost of developing new therapy and unnecessarily exposing some patients to inferior experimental treatments.

The Bayesian updating process has profound implications for trial design. Perhaps its most useful consequence is the ability to quantify (probabilistically) what is going to happen in a trial from any point on, given the currently available evidence. More generally, the posterior probability can be used as predictive probability, i.e. the posterior distribution becomes the new prior distribution to use along with the likelihood to assess the prediction of hypothetical new results, without the need to implement complex group sequential boundaries to the type I error rate inflation arising from testing repeatedly interim data (DeMets and Lan 1994).

The Bayesian process of continuous learning makes it possible to modify a trial in midcourse. This kind of modifications include stopping the trial, adaptively assigning patients to treatment that are performing better, adding and dropping treatment arms, and extending accrual.

Finally, in analyzing data, the Bayesian aptitude is to bring all the available information, which is typically overlooked because traditional approach struggles to dealing with it. For example, when survival is the primary endpoint, patient-specific outcomes that might be correlated with survival are usually not considered in the classical approach, whereas they can be easily considered in the hierarchical modeling strategy allowed by the Bayesian framework (Bhattacharjee 2013).

3 The Papers of this Issue

The papers of our special issue are grouped in the aforementioned three sections:

3.1 I. Validity and Limitations of RCTs and EBM

(1) *Lalumera & Fanti* raise the problem of applicability of RCTs to validate nuclear diagnostic imaging tests. In spite of the wide application of PET and other similar techniques that use radiopharmaceuticals for diagnostic purposes, RCT evidence is sparse. The authors argue that this occurs due to the specific characteristics, that they extensively discuss, of nuclear diagnostic imaging and of radiopharmaceuticals. The paper is meant to contribute both to the philosophical discussion on the EBM hierarchy of evidence, and on the specific debate on radiopharmaceuticals in nuclear medicine.

(2) *Hovda* points out that EBM may be a controversial model when it comes to research into the effectiveness of psychotherapeutic treatments. This is partially due to the so-called 'Dodo Bird verdict', a thesis claiming that that all psychotherapies are equally effective, and that their effectiveness is largely due to the placebo effect. In response to this controversy, Hovda argues that EBM can nevertheless be made to fit research into the effectiveness of psychotherapy, once a piecemeal approach to conducting RCTs is considered. Such an approach requires studying the contributions made by the individual components of a treatment.

(3) *Osimani and colleagues* focus on the so-called In-Silico Clinical Trials (ISCTs), undertaking a formal analysis of them, as opposed to RCTs, in order to identify their distinct contribution to causal inference in the clinical setting. The authors claim that ISCTs can improve the 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. With regard to causality, Osimani and her co-authors conclude that ISCTs and RCTs cannot replace one another, but rather they are complementary.

3.2 II. Varieties of Clinical Reasoning and Values

(4) *Boniolo & Campaner* try to spur the debate on causation in the philosophy of biomedicine by highlight-

ing some serious clinical aspects not yet satisfactorily tackled. The authors point out that some pathological situations addressed by molecular medicine actually prove resistant to (at least) some of our major epistemological accounts of causal explanation. They show the relevance of this problem not only for the philosophy of causation in the field of medicine, but for the real ethical issues involved and concerning the existential decisions that patients with severe molecular diseases have to take.

(5) *Stanley* describes the patterns of reasoning that experienced clinicians use under different diagnostic circumstances and shows how inductive, deductive and abductive reasoning are meshed together. Through a realistic and detailed clinical case of a 54-year-old man with a sudden acute chest pain, it is shown how plausible diagnoses are generated, and selected on the basis of a combination of strategies. These strategies make use of probabilistic reasoning, which is inherently Bayesian, and they mix deduction on predicted signs and symptoms, given a conjectured disease, to new facts unfolding from continuing clinical examination.

(6) *Chiffi & Pietarinen* point out that clinical equipoise (CE) has been proposed as an ethical principle relating uncertainty and moral leeway in clinical research. CE is usually indicated as a necessary condition for a morally justified introduction of a new RCT. However, questions related to the interpretation of this principle remain woefully open. Recent proposals to rehabilitate CE have divided the bioethical community on its ethical merits. In the light of this, Chiffi & Pietarinen present a new argument, based on the methodology of hypothetical retrospection, that brings out the epistemological difficulties we encounter in justifying CE as a principle connecting uncertainty and moral leeway in the ethics of clinical research.

(7) *Fillion* begins his analysis by considering that in specific circumstances biostatisticians think that adaptive design is the only ethically permissible experimental design. However, some proponents of CE argue that adaptive trials are neither ethically required nor permissible. More specifically, they affirm that adaptive trials fail to meet the moral requirement of CE, since they presuppose an epistemic state that is not coherent with a physician's duty of care. Fillion emphasizes that the debate on adaptive design and equipoise conflates two different conceptions of statistical evidence (namely, frequentist and Bayesian). Such distinction helps the author to propose an epistemological framework in which adaptive trials are both consistent with and recommended by the principle of CE.

3.3 III. Bayesian Approach

(8) *Berchialla, Gregori, and Baldi* explain the key role played by the randomization in clinical trials designs. While randomization is primarily intended to prevent the source of bias in treatment allocation, in the frequentist analysis it provides a justification for the validity of statistical testing. Instead, in the Bayesian framework, the role of randomization is more nuanced, and the Bayesian analysis can afford a valid rationale for selective controls. In clinical trials, there have been many heated debates on the ethics of randomization since it has been claimed that it is ethical to randomize patients to different health-related interventions only in a state of true equipoise. In this paper the authors offer a view of randomization from the perspective of both frequentist and Bayesian inference and discuss adaptive randomization as point of encounter for an ethical evaluation of clinical trials.

(9) *Andreoletti & Oldofredi* make explicit an epistemological tension between the conduct of clinical trials and their interpretation: the reasoning underlying the methodological design of clinical trials hinges on a frequentist interpretation of probability, whereas the interpretation of their results often implies a Bayesian reasoning. The authors suggest that acknowledging the potentiality of Bayesianism might contribute to clarify and improve comprehension of medical research. They present a formal method which has been developed in the context of objective Bayesian statistics to define priors which have a minimal or null impact on posterior probabilities. Finally, the authors show how this method does not violate the senses of scientific objectivity which traditionally pose challenges to the Bayesian perspective.

References

Bernardo JM, Smith AFM (1994) Bayesian Theory. Wiley, Chichester

Bhattacharjee A (2013) A Bayesian joint analysis and imputation model for longitudinal data: an application in type 2 diabetes drug effect comparison. Int J Collabor Res Intern Med Public Health 5(2):103–111

Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Rand McNally, Chicago

Cartwright N (2011) A philosopher's view of the long road from RCTs to effectiveness. Lancet 377(9775):1400–1401

Chiffi D, Zanotti R (2017) Fear of knowledge: clinical hypotheses in diagnostic and prognostic reasoning. J Eval Clin Pract 23(5):928–934

Christakis NA, Sachs GA (1996) The role of prognosis in clinical decision-making. J Gen Intern Med 11(7):422–425

Cranor CF (1990) Some moral issues in risk assessment. Ethics 101(1):123–143

DeMets DL, Lan KK (1994) Interim analysis: the alpha spending function approach. Stat Med 13:1341–1352

Djulbegovic B, Hozo I, Greenland S (2011) Uncertainty in clinical medicine. In: Gifford F (ed) Philosophy of medicine. Handbook of the Philosophy of Science, vol 16. Elservier, Amsterdam, pp 299–356

Federspil G (2004) Logica clinica. McGraw-Hill, Milan

Greenland S, Senn SJ, Rothman K, Carlin JB, Poole C, Goodman N, Altman DG (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31:337–350

Hempel CG (1965) Science and human values. In: Hempel CG (ed) Aspects of scientific explanation and other essays in the philosophy of science. The Free Press, New York, pp 81–96

Keynes JM (2008) The general theory of employment, interest, and money. Atlantic Publishers [1936], New Delhi

Parascandola M (2010) Epistemic risk: empirical science and the fear of being wrong. Law Probab Risk 9(3–4):201–214

Rich BA (2002) Prognostication in clinical medicine: phrophecy or professional responsibility? J Leg Med 23(3):297–358

Rizzi DA (1993) Medical prognosis–some fundamentals. Theor Med 14(4):365–375

Rothwell PM (2005) External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 365(9453):82–93

Savage LJ (1954) The foundations of statistics. Wiley, New York

Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131

Worrall J (2010) Evidence: philosophy of science meets medicine. J Eval Clin Pract 16(2):356–362