Pharmacovigilance as Personalized Evidence

Francesco De Pretis[1] *William Peden*[2], *Jürgen Landes*[3] *and Barbara Osimani*[4]

Abstract

Personalized medicine relies on two points: 1) causal knowledge about the possible effects of X in a given statistical population; 2) assignment of the given individual to a suitable reference class. Regarding point 1, standard approaches to causal inference are generally considered to be characterized by a trade-off between how confidently one can establish causality in any given study (internal validity) and extrapolating such knowledge to specific target groups (external validity). Regarding point 2, it is uncertain which reference class leads to the most reliable inferences.

Instead, pharmacovigilance focuses on both elements of the individual prediction at the same time, that is, the establishment of the possible causal link between a given drug and an observed adverse event, and the identification of possible subgroups, where such links may arise. We develop an epistemic framework that exploits the joint contribution of different dimensions of evidence and allows one to deal with the reference class problem not only by relying on statistical data about covariances, but also by drawing on causal knowledge. That is, the probability that a given individual will face a given side effect, will probabilistically depend on his characteristics and the plausible causal models in which such features become relevant. The evaluation of the causal models is grounded on the available evidence and theory.

Keywords: Drug safety; Personalized Medicine; Pharmacosurveillance; Pharmacovigilance; Precision Medicine.

1 Introduction

Pharmacovigilance (or pharmacosurveillance) has gradually emerged during the last century, in the aftermath of notorious disasters related to pharmaceutical products. The most notorious was the Thalidomide case, but others include the DES case, Croniassal case, and more recently the Vioxx case[5]. The establishment of the institute of pharmacosurveillance emerged in the course of time out of the increased awareness that pharmaceutical products (like other chemicals) are associated with unknown risks, which may go unnoticed during the studies for approval, because they might be related to long-term use, drug-drug interaction, or particular physiological co-factors. Drug monitoring thus registers suspected case of adverse events, collects them, analyses them and constantly evaluates their import in relation to the drug risk-benefit profile.

Like personalized medicine, pharmacovigilance is a fast-growing part of modern medicine. Both fields face many of the same methodological challenges, and developments in each one of them will impact the other; however, their interrelations and potential for integration are under-investigated, especially by philosophers. Both personalized medicine and pharmacovigilance aim to accurately predict what would happen to a given individual, were he to receive a given treatment X. This sort of promise relies on two ingredients: 1) causal knowledge about the possible effects of X *in a given statistical population*; 2) assignment of the given individual to a suitable reference class. These two inferential problems have been extensively analysed in the philosophical and methodological literature (Reichenbach

[1]Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60126 Ancona, Italy and Department of Communication and Economics, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy

[2]Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60126 Ancona, Italy

[3]Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, 80539 München, Germany

[4]Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60126 Ancona, Italy and Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, 80539 München, Germany. Corresponing author b.osimani@univpm.it

[5]For an historical overview of the role and drug agencies in the postmarketing phase and the related regulation see: (Osimani 2008); for an overview of the complex ecosystem of epistemic, regulators, ethical and financial constraints informing the process of evidence collection, evaluation and use for decisions in the medical and pharmaceutical domain see (Osimani 2020).

1951; Pollock 1990; Hájek 2007; Pearl and Bareinboim 2014; Luo et al. 2019; Stoffi and Gnecco 2018; Xie 2013; Nilsson et al. 2019; Pearl 2012; Osimani 2020). We discuss here how they are related and present a methodological tool, *E-Synthesis*, that is able to exploit the evidence for either inferential targets, based both on statistical knowledge about relevant subgroups of users and causal knowledge about the drug-effect relationship. This tool was originally developed for assessing causality in pharmacovigilance; we examine here the prospects for applying it to questions in personalized medicine.

E-Synthesis is a Bayesian approach to aggregating heterogenous evidence for the purpose of assessing causal relationship between pharmaceuticals and adverse reactions in a probabilistic terms. The Bayesian element in *E-Synthesis* consists in its use of Bayesian learning methods, which are explained below. Bayesian reasoning is very popular in contemporary philosophy of science, but the application of Bayesian ideas to complex methodological issues in sciences like medicine is still a work in progress. *E-Synthesis* contributes to the cutting-edge research in applied Bayesian philosophy of science. (Landes et al. 2018; De Pretis and Osimani 2019; De Pretis et al. 2019; Abdin et al. 2019) illustrate this new framework for evidence synthesis focusing on various components of the system. We continue here to develop this methodology to questions concerning the estimation of single patient's adverse drug reactions (ADRs).

Before going on, we warn the reader about our goals and scope. We do not intend here to offer an instrument for causal search (testing hypotheses about causal models against data) or causal inference from data. Neither do we intend to provide instructions on choosing the right kinds of statistical models. Instead, we provide an instrument that uses probabilistic knowledge about such models in order to output a probabilistic estimate that a given individual might be affected by a given ADR. In order to do this, we adopt a Bayesian model averaging approach, that allows us to contemplate diverse causal models as possible generators of the observed data and to compute the individual estimate, on the basis of their plausibility and their relevance to the specific features of the considered individual. We also trace and explicate the epistemic steps that may guide the inference from available causal models plausibly linking the drug to the ADR, to the prediction that an individual may be affected by that ADR upon taking the drug.

We proceed as follows: in Section 2 we discuss the similarities and differences between pharmacovigilance and personalized medicine at a broad level. In Section 3, we introduce *E-Synthesis*. In Section 4, we provide an example application of this approach to determining the probability of an ADR in a single patient. The idealizations and limiting assumptions that we make are indicative of future directions for research and development for attaining this goal in real-world medical decisions. Finally, in Section 5, we close with a general discussion.

2 Personalized Medicine and Pharmacovigilance

In this chapter, we understand the term "personalized medicine" as referring to the targeting of interventions, drugs, and other parts of medicine to particular patients, based upon their membership of multiple populations for which we have at least some relevant statistical data. In an ideal world for medical research, we would know exactly what effect a given treatment would cause in every given patient. In a less ideal world, we would have some probabilistic knowledge about that patient's probability of suffering an adverse effect, given that they have taken a drug. Our world is even less ideal: we typically lack even such probabilistic information, especially regarding ADRs. Personalized medicine uses statistics from various covariates (genotype, clinical biomarkers, age etc.), identified by stratifying patients according to physiological and other characteristics, in order to identify the most adequate reference class for a given individual with respect to a specific effect [6]. Improvements in personalized medicine hold the promise of the systematic and precise application of medicines to patients, rather than a "casual hit and miss" approach (Wertheimer 2016).

Since every patient is a member of many reference classes and we often have conflicting data for these various

[6] A reference class is a group of objects or events that we describe using a statement of probability. For example, the reference class in "The probability of $1 coins landing heads is approximately 50%" is the class of $1 coins.

categories, personalized medicine must deal with the problem of combining and interpreting the diverse data. The Problem of the Reference Class is one of the greatest problems in the philosophy of statistics; it is thereby relevant to all sciences (Venn 2013, p. 194) and (Hempel 1962, p. 374)[7]. In essence, this problem is simple: everything we consider in science (patients, chemical samples, thermometer readings etc.) is a member of many reference classes, and we will often have statistical data for these reference classes that point in different directions with respect to the hypothesis that we are considering. For example, imagine that we know the frequency of a drug's effectiveness among women and patients over 65, but not the intersection of these reference classes (i.e., women aged over 65). Suppose that the data suggests a high degree of effectiveness in women, but a very low effectiveness in patients over 65. Given a female patient, aged over 65, which reference class should we use for estimating whether the drug will be effective in treating her? There is a massive literature on this topic in general philosophy of science and philosophy of medicine, including personalized medicine (Kent et al. 2018) and in philosophies of other disciplines (Franklin 2011; Strevens 2016).

The reference class problem affects the estimation of beneficial and harmful effects of treatments, which is often the cornerstone for risk policy. For example, suppose that policymakers discover that a drug D has a favourable safety profile for its average user, but it has an unfavourable safety profile for a particular subgroup S. The policymakers might decide that D should not be withdrawn from the market, but instead to circumscribe its use to those groups of consumers for which its benefit-risk profile is favourable. This decision-making might seem trivial, but it requires identifying the relevant subgroups in a rational way. Furthermore, both before and after drug approval, evidence of harms is sparse and noisy. Because of the awareness of such latent risk enduring also after approval, and of the epistemic and methodological difficulties surrounding risk assessment, drug monitoring has been institutionalised (both through the development of a sophisticated set of norms in soft and hard law, as well as through the establishment of national and supranational agencies for drug approval and monitoring), with the purpose of updating the drug safety profile in an ongoing manner and make timely decisions on this basis [8].

If we know the exact causal relations for a particular patient, then we evade the Problem of the Reference Class. For instance, if matters were as simple as "A patient with Gene X will have adverse reaction Y" and we can easily test each patient for Gene X, then we do not have to consider the patient's class memberships. However, in practice, things are rarely so simple. The relevant causal mechanisms in a patient can be impossible to identify. The causal mechanism might not be a closed system. Gene X might be part of a wider causal mechanism, including the patient's social environment, lifestyle choices, other medications, and so on. (For discussions and examples, see the contributions by Xavier Guchet, Maël Montévil, and Anya Plutynski in this volume.) Thus, in practice, personalized medicine must make use of statistical data about the reference classes to which a patient belongs.

Pharmacovigilance also faces challenges of synthesising and interpreting heterogeneous data. The term 'pharmacovigilance' covers both the empirical study of harms caused by drugs (i.e., ADRs) and the regulation of drugs given these harms. ADRs are a broad category covering a large spectrum of effects. ADRs happen within the so called "therapeutic range". According to the traditional WHO (World Health Organization) definition, ADRs are: "A response which is noxious and unintended, and which occurs at doses normally used in humans for the prophylaxis, diagnosis, or therapy of disease, or for the modification of physiological function" (WHO, 1972) The classical Rawlins and Thompsons categorization of ADRs subdivides them in: 1) Type A: dose-dependent effects, predictable from the drug pharmacology; 2) Type B: effects independent from dosage, unavoidable reactions with no predictable connection with the drug known mechanism (off-target effects); these are generally characterized by high severity and irreversible damage (e.g.,. anaphylactic shock); 3) Type C: chronic reactions; 4) Type D: delayed reactions (e.g., adenocarcinoma in women exposed in utero to diethylstilbestrol); 5) Type E: withdrawal reactions (as a consequence of addiction: e.g.,. benzodiazepines); 6) Type F: failure of therapy (lack of efficacy). Determinants of individual susceptibility

[7] In this chapter, we consider only the epistemic version of the Problem of the Reference Class; there is also a metaphysical version concerning the "true" physical probability of some particular event (Hájek 2007, pp. 565-566).

[8] For philosophical discussions on the underpinnings, rationales and issues concerning pharmaceutical regulation see (Teira and Reiss 2013; Teira 2013; Andreoletti and Teira 2019; Osimani 2013, 2007).

to this range of reactions can be practically anything, including age, sex, genetic make-up, physiological changes, exogenous factors (drug-drug or food-drug interference), and disease-drug interaction; additionally, ethnicity is a carrier of factors (environment, genetics, lifestyle) determining higher or lower susceptibility to drug efficacy and ADRs (Aronson and Ferner 2003).

Mechanisms leading to ADRs may be the same that produce the intended therapeutic effect (e.g., non-steroidal anti-inflammatory drugs' inhibition of prostaglandins production causing both anti-inflammatory effects and gastritis), or may be related to the drug binding to the intended receptors (for the intended pharmacological reaction) at different sites other than the target organ, or else it can be the result of an integrated response of separate levels in the organ system (interaction of different organ levels). ADRs may be collateral effects, that is, effects which are merely distinct from the intended therapeutic goal. However, they can also be paradoxical, i.e., consisting of exactly the opposite effect of the one which the drug was intended to produce, or bidirectional effects in which the same drug produces opposite effects in the same individual at the same time (Smith et al. 2012).

Unlike personalized medicine, the focus in pharmacovigilance is to identify possible causal relations between drugs and adverse effects in the first place, rather than merely identifying statistical relations between drugs and memberships of particular reference classes. The two inferential targets, however, may work together, as we will see in the following.

Interpreting and combining heterogeneous evidence of different sorts of ADRs is fraught with methodological challenges, included the Problem of the Reference Class: drug testing provides data in favour of many statistical hypotheses about many reference classes, our target population will generally be a member of multiple reference classes, and often the statistical evidence will point in different directions regarding likely ADRs in the target population. More often, evidence will be very "local", e.g., individual case reports, which point to very specific subsets of features, any of which may have been relevant for the occurrence of the ADR.

Fortunately, shared problems often have shared solutions. Firstly, advances in statistical methods, medical epistemology, and improved experimental practices in pharmacovigilance can assist with personalized medicine, and vice versa. For example, techniques for synthesising heterogeneous evidence in pharmacovigilance might also be adaptable to applications in personalized medicine. Similarly, answering the Reference Class Problem in personalized medicine, at least partially, would also help address it in pharmacovigilance. Secondly, solving problems in one field might *ipso facto* alleviate problems in the other, at least in some cases. For instance, we sometimes have excellent data of the effects of particular medicines on specific genotypes, and sometimes even causal knowledge of various kinds, such as knowing the interactions that occur with the drug and specific enzymes expressed by people who possess a particular genotype (Wertheimer 2016). Genotyping may also assist the pursuit of external validity in pharmacovigilance, because it would help us identify target groups that may suffer from specific ADRs in view of their genetic make up. Group-specific effects also generate practical issues in drug labelling, which are shared by personalized medicine as well (Fang et al. 2016).

There has not yet been much research on the relationship between personalized medicine and pharmacovigilance. However, the personalized medicine emphasis on targeting treatments to patients with a high probability of a positive response has inspired some research in pharmacovigilance, in proposals for targeted research by pharmaceutical companies (Enjo B 2018) [9].

From the other side, researchers in personalized medicine have noted the importance of better pharmacovigilance data for personalized medicine (Reynolds 2015; Ennezat et al. 2017; Masimirembwa et al. 2016). At the applied level, pharmacogeneticists are seeking methods that will concurrently improve the two research areas (La Russa et al. 2017).

[9]Some philosophers have emphasized the privileged role of "dispositionalist" views on causation as more adequate approaches to causal inference in pharmacovigilance (Anjum and Rocca 2019; Rocca et al. 2020, 2019). In this paradigm, causation cannot be established without a thorough knowledge of the context in which the particular causal event happens: understanding the dispositions at place in one single patient (especially in cases of unexpected effects of a drug) can potentially help improve the knowledge about the intrinsic properties of the drugs, and the way such properties might interact. While the dispositionalist approach to causation promises to be very fruitful in the biological sciences and pharmacology particularly, we emphasise here that *E-Synthesis* does not commit to any specific ontological view on causation and is intentionally flexible as to the metaphysical stance that one assumes with regard to causality.

Given the great importance of pharmacovigilance and personalized medicine for the health and wealth of nations, the integration of insights and methods across these areas is a burning issue for the philosophy of medicine.

We present here an approach to the reference class problem which draws on knowledge about the causal structure about the data generating process. We thereby emphasize a division of labour between the phase of hypothesis generation regarding such causal structures, and that of hypothesis confirmation, through probabilistic evidence. Furthermore, since in the field of pharmacovigilance we are mainly dealing with "little" rather than "big data", and we may lack the required resources to test causal models with several variables involved, we propose here a method for hypothesis confirmation that makes the most out of all the available evidence, of whatever quality and relevance it is. This holds a promise for personalized medicine in that, especially in the early phase of risk detection, the availability of big data is not the common scenario, while tools for the optimization of the available evidence can serve the purpose of approaching an accurate individual estimate of the individual risk associated with drug intake, based on the current state of the art.

3 *E-Synthesis*

E-Synthesis has been developed to aggregate evidence in the process of pharmacosurveillance and make timely decisions in view of incoming "safety signals" (Landes et al. 2018; Abdin et al. 2019; De Pretis and Osimani 2019; De Pretis et al. 2019). In the preceding papers, we have presented its theoretical foundations (Landes et al. 2018), how information about evidence quality may be incorporated in the risk estimation (De Pretis et al. 2019), and how evidence of biological mechanisms or dose-response contributes to the confirmation of causal hypotheses (De Pretis and Osimani 2019; Abdin et al. 2019). In the present paper, we illustrate how issues related to the specification of the causal claims between drug and ADRs with respect to the subgroups to which they relate may be efficiently addressed by our framework. This will help the prediction of drug effects on individuals, based on available knowledge, both derived from established theory and from learning techniques (e.g., machine learning, statistics, artificial intelligence).

3.1 Motivation

From its conception, throughout the development phase, and continuing in the post-marketing phase, each drug risk-benefit profile is assessed and continuously updated. If the utility from the expected benefits outweighs the expected disutilities from harms (and possibly monetary costs), then the assessment is favourable and development and/or circulation of the drug in the market are advisable. Instead, if the negatives outweigh the positives, then there might be a decrease in research effort or a (partial) withdrawal from the market, depending on the development stage. Hypotheses of causal associations between drugs and harms consolidate only gradually on the basis of heterogeneous evidence. However, these causal associations should be anticipated as much as possible in order to minimize harm to exposed subjects. Hence, decisions concerning a drug (e.g., approval, suspension, withdrawal, administration) should be based on all the available evidence at any point in time. This can imply that one may need to base such decisions on evidence that is normally perceived as weak, based on current evidence standards, such as basic science studies.

A sophisticated set of tools for meta-analyses and systematic reviews has been developed for the purpose of evaluating the intended therapeutic effects of interventions, but the adaptation of these instruments with the aim of evaluating the safety of health technologies encounters various problems. These are mainly due to the sparsity, heterogeneity, and "fragility" of data concerning unintended effects of medical treatments: 1) evidence about unknown risks emerges gradually from spontaneous reports or other kinds of sporadic data: especially in the earlier phases of risk detection, these data can be at the same time very noisy and rare; 2) furthermore, the data may come from heterogeneous sources (such as clinical case series, animal studies, or molecular studies, etc.); and 3) it can be unreliable, because of noise and/or confounding. As a matter of fact, evidence for harm emerges unsystematically and unpredictably. The evidence demands evaluation (and decision) even when it cannot deliver perfect information about the state of nature.

3.2 Aims and Scope

The need to provide an instrument for synthesizing the evidential support of heterogeneous sources for assessing hypotheses of causal association between drugs and harms at any point in time and on the basis of any kind of available evidence led us to develop *E-Synthesis* (Landes et al. 2018; Abdin et al. 2019; De Pretis and Osimani 2019; De Pretis et al. 2019). This framework provides a theoretical basis for justifying the probabilistic confirmation of causal hypotheses, on the basis of all the available evidence. It puts forward a Bayesian epistemic network incorporating indicators of causality derived from Bradford-Hill "guidelines" for inferring causation (Hill 1965) and indicators of evidence quality. The result of the assessment is a probabilistic estimation of the causal hypotheses of interest, which reflects the degree of support provided by the available evidence (Hawthorne 2005). We formalise our hypothesis of causation as follows: "Cause D causes effect E in a given population p and a given causal model M" [10].

However, we have not yet illustrated how updating the causal hypothesis itself may work by taking into account incoming information about the causal structure generating the data and/or evidence about subgroup effects. That is, we aim to determine probabilities over the causal models in which the drug does cause the particular ADR we consider. The causal models allow for predictions which take into account characteristic properties of individual patients and may thus in turn better predict the outcome of individual patients than the generic causal hypothesis ⓒ (D causes E *in a population*). With these probabilities over the causal models in hand, we hence aim to defeasible determine the probability that a particular patient will (not) suffer the ADR after the administration of the drug D.

3.3 Background

3.3.1 A Very Short Introduction to Bayesian Epistemology

E-Synthesis puts forward an epistemic network drawing on Bayesian updating as the main computational tool. For those unfamiliar with the topic, we briefly introduce it here and give the rationales of this inferential tool.

Bayesian epistemology is a philosophical account of rational beliefs in hypotheses that comes in degrees [11]. Bayesianism represents uncertainties by modelling them as probability functions [12]. To determine the conditional posterior probability P of the hypothesis being true - in our case this hypothesis is that of the drug causing an ADR (denoted by ⓒ) - given the available evidence, $\mathcal{E}$, one applies Bayes' Theorem:

$$P(ⓒ|\mathcal{E}) = \frac{P(ⓒ) \cdot P(\mathcal{E}|ⓒ)}{P(ⓒ) \cdot P(\mathcal{E}|ⓒ) + \sum_{i=2}^{N} P(H_i) \cdot P(\mathcal{E}|H_i)} \quad ,$$

where the hypotheses H_i together with $ⓒ = H_1$ form a mutually inconsistent and exhaustive partition [13].

With this mathematical formula, the posterior probability of the hypothesis given the evidence, $P(ⓒ|\mathcal{E})$, only depends on prior probabilities $P(H_i)$, and likelihoods $P(\mathcal{E}|H_i)$. Bayesian epistemology allows one to conditionalise on any proposition (or event) whereas in Bayesian statistics one conditionalises on statistical models [14]. The probabilities in Bayesian epistemology are interpreted more widely as one's uncertainties about general propositions or as degrees of support for the hypotheses of interest.

[10] By "causal model" we mean an ordered set of variables and a description of their interrelationships in terms of varying strengths and functional forms (Cartwright and Stegenga 2011), as exemplified by Structural Equation Models and Directed Acyclic Graphs (Cowell et al. 2006; Dawid 2010; Pearl 2009). See also subsequent sections.

[11] In the philosophy of science, there are controversies regarding the extent to which Bayesianism and whether its subjectivity is a problem (Gelman and Hennig 2017; Sprenger 2018). We assume here that the possible drawbacks related to such issues are compensated by its flexibility in interpolating and extrapolating data when these are fragmentary and heterogeneous. In pharmacovigilance, our data often has these characteristics.

[12] Philosophy of science also draws on the formal tools developed within Bayesian epistemology in order to investigate scientific inference (Bovens et al. 2003; Talbott 2011; Howson and Urbach 2006).

[13] We abuse notation in the usual way by using the same symbol denoting a variable and the variable being true.

[14] These likelihoods are relatively easy to determine, since they provide probabilities on the assumption that a certain hypothesis is true and thereby considerably shrink the set of possible worlds. For example, determining the likelihood for a statistical hypothesis H_i is merely an exercise of probability calculations. It would be much harder to specify probabilities on all states generated by the variables of interest. We demonstrate below that although the task to determine the likelihoods is relatively easy, it is by no means a trivial task.

3.3.2 Bayesian Networks

Bayesian networks are a convenient tool to graphically display and reason with probability functions. They allow us to specify and read-off conditional independencies from a graph. Formally, a Bayesian network is built up on a number of pairwise different propositional variables which form the nodes of a graph. The graph topology on these nodes forms a Directed Acyclic Graph (DAG). This means that edges are directed and that there is no directed cycle in the graph, i.e., there is no path of directed edges which leads back to its starting point.

Finally, one needs to specify the (conditional) probabilities of all variables. For a variable Y one first determines the set of parent variables: a variable X is a parent of Y if and only if a directed arrow originates at X and ends at Y. Denoting the parents of Y by $X_1, \ldots, X_n$ one specifies

$$P(Y = y | X_1 = x_1, \ldots, X_n = x_n) \in [0, 1]$$

for all possible values $y, x_1, \ldots, x_n$ under the condition that

$$\sum_x P(Y = y | X_1 = x_1, \ldots, X_n = x_n) = 1 \ .$$

This condition ensures that one defines a probability function. Conditional and unconditional probabilities of arbitrary events are calculated by (repeated) application of the so-called "chain rule". For more background on Bayesian networks, see (Darwiche 2009; Neapolitan 2004)[15,16].

3.4 General Approach

While Bayes' theorem is essential in Bayesian epistemology, it is by no means clear how to determine the likelihoods $P(\mathcal{E}|H_i)$ in our concrete application where $\mathcal{E}$ is all the available information that might (dis)confirm that a given drug D causes adverse effect E.

To facilitate this task, we introduce abstract indicators of causality. We distilled six indicators by applying a philosophical analysis of current theories of causation to the Bradford Hill Guidelines (Landes et al. 2018). Our indicators of causality are: 1) difference making, 2) probabilistic dependence, 3) dose-response relationship, 4) rate of growth, 5) temporal precedence and 6) mechanistic knowledge [17]. Conceptually, our indicators are testable (probabilistic) consequences of the hypothesis of interest. As such, experiments and observations can, by providing evidence against or in favour of the indicators, thereby probabilistically (dis)confirm the causal hypothesis of interest. The common definition of an "indicator" implies that it is more likely to be true, if the inferential target is also true, than if the latter is false [18]. Hence, in our case:

$$P(Ind|\copyright) > P(Ind|\overline{\copyright})$$

Each experimental study, observational study, case series, case report or basic science finding is then associated with a set of causal indicators which it is informative about (Landes et al. 2018; De Pretis et al. 2019). This procedure routes the inference from medical data to a theoretical entity (causation) via abstract intermediaries (causal indicators). Such

[15] Every Bayesian network uniquely specifies a probability function. However, typically a probability function is representable by multiple Bayesian networks.

[16] Conditional independencies are given by the d-separation criterion (Geiger et al. 1990).

[17] By proposing his guidelines, Bradford Hill meant to offer an alternative method for causal inference in the face of hazard, in a context dominated by hypothesis testing. While he obviously did not exclude experimental sources of evidence as a basis for causal inference, he emphasised that non-experimental evidence may also contribute to assessing the causal import of a specific chemical/molecules. Some paradigmatic examples are when many items replicate the same results, converge towards the same hypothesis, or cohere with available theoretical and/or empirical knowledge. *E-Synthesis* pays homage to such "higher order" considerations too, in that it allows diverse items of evidence to jointly contribute to the probabilistic (dis)confirmation of the hypothesis under investigation, and therefore, to exploit their (lack of) coherence.

[18] With this, *E-Synthesis* relaxes standard necessary and sufficient conditions for causation developed in the philosophical literature on causation (Landes et al. 2018).

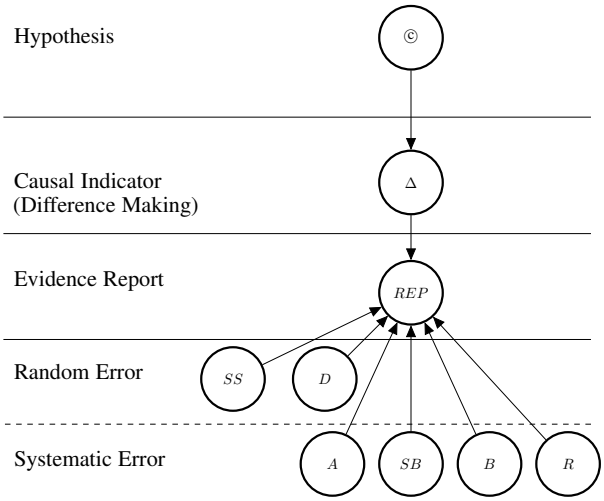

Figure 3.1 Graph structure of the Bayesian network for one randomized controlled trial (RCT) which informs us about difference making (Δ) which in turn informs us about the causal hypothesis. The information provided by the reported study is modulated by how well the particular RCT guards against random and systematic error. The evidential modulators for an evidence report are SS = Sample Size; D = Study Duration; A = Adjustment for covariates or subgroup analyses and the like; SB = Sponsorship Bias; B= Blinding; R = Randomisation. (De Pretis et al. 2019).

an approach recalls Bogen and Woodward's distinction between data and phenomena (Bogen and Woodward 1988): *E-Synthesis* breaks down the inferential process from the raw data to the hypothesis that a causal link holds between Drug and ADR into two steps: 1) from data (study reports) to causal indicators; 2) from causal indicators to causality.

Items of evidence (dis)confirm indicators of causation to different degrees. The degree to which indicators of causation are (dis)confirmed is made partly explicit by spelling out evidential modulators, which signals the quality of evidence as a function of various choices in study design and data analysis (blinding, randomisation, sample size, study duration, stratification), see Fig. 3.1.

4 Pharmacovigilance as Personalized Evidence

We now discuss how to expand *E-Synthesis* for personalized medicine. We investigate how a framework developed for inferring causation in a population can be adapted to inferring the probability of an ADR occurring in a particular patient, and which philosophical problems require to be addressed during this development. As a preliminary step to such a form of "extrapolation", one would normally need to identify the closest "reference class" instantiated by the individual at stake on the basis of available knowledge. We adopt a "structural" approach to this problem, in that we derive the probability P of a particular drug consumer being affected by a given side effect not only by assigning him to a given reference class, but also by drawing our inference on a weighted estimate of different causal models being at play, based on the available evidence and theoretical knowledge. We will not be able to determine for certain which causal model is correct, if any. Instead, we can use the models as tools for assessing the likelihood of a causal connection in the case of a particular patient, by considering their predictions, weighted according to each model's probability given our evidence. As a preliminary step to this we need to update probabilities over plausible causal models, conditional on incoming evidence of various kinds.

We show how this can be done in principle by *E-Synthesis*, by drawing on the distinctive components of *E-Synthesis's* inferential structure. A hypothesis about a causal relationship between the drug and the side effect is represented by the variable ©. On its turn, this is a disjunction of mutually exclusive and jointly exhaustive causal models: $\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_n\}$ partitioning ©. Every model $\mathcal{M}_i$ consists of a set of functions mapping a set of variables

(relevant causal factors for developing a given side effect) to the set of possible values measuring the strength of the effect: thus a model $\mathcal{M}_i$ consists of a set of equations of the form $Y^k = F^k(X)$, where the F^k are functions relating random variable(s) X to random variable(s) Y, representing causal relationships.

The causal models relate the drug dosage D and the ADR E by embedding them in a causal structure. Each model delivers an expectation about the intensity of a given side effect E, given a certain dosage, and other constraints (co-factors contributing to the side effect, such as mediators, moderators, interactive causes etc.). For instance, a model $\mathcal{M}_i$ could read as follows:

$$E = \alpha + \beta_1 D + \beta_2 D^2 + \epsilon_1$$
$$D = \gamma + \beta_3 C + \epsilon_2$$
$$C = \lambda + \beta_4 F^2 + \epsilon_3$$

where, for example, α is the base rate of the adverse effect in the population and the epsilons are error/noise terms. Structural coefficients, such as β_1, represent the strength of the causal relationship between independent and dependent variable. This may due to the intrinsic intensity of such relationship, but also other, possibly unknown, moderating factors (such as age, sex, co-morbidities, etc.). We assume that at each stage, the available domain knowledge (knowledge about biological processes, physiology, the pharmacodynamics and pharmacokinetics of the drug, epidemiological and clinical data, etc.) delivers an exhaustive set of mutually exclusive causal models with respect to the current state of the art. Although, the availability of an *exhaustive* set of mutually exclusive causal models underpinning the data is not a realistic assumption, we make it here for the sake of simplifying the representation of the essential components of our inferential machinery, cf. 6 in the discussions.

A probability distribution P over the set of causal models $\mathcal{M}$ measures the support these enjoy from the theory and the data, and gets updated upon new evidence. The probability that a *specific individual* may get a side effect is calculated on the basis of P over $\mathcal{M}$. It is then updated on the basis of new evidence, as we explain in Section 4.3.

The absence of any causal connection, e.g., between E and D may be formalised by having coefficients for D set to zero. We subsume such models under $\textcircled{c}$ which refers to the state of affairs where there is no causal connection between D and E. For obvious reasons, even in the absence any causal link between D and E, there might still be a non-zero base-rate of people who suffer the adverse event E for causes other than D [19].

4.1 Generating Causal Models/Hypotheses Using *E-Synthesis*

4.1.1 Data & Theory

The system starts with model hypotheses of causation, which does not specify its functional form, nor its parameters; for instance $\mathcal{M}_i : E = f_i(D)$, where D stands for the drug (dosage) and E stands for the effect, the strength of the adverse effect. $\mathcal{M}_i$ represents the assumption that there might be a causal effect from the drug to E without specifying its functional form, possible mediators, interactive factors and the strength of such relationship. Incoming evidence from, say, observational studies, might reveal subgroup effects for specific users or specific dose-response curves. In the following, we will outline an example to illustrate, step-by-step, the inferential procedures we are considering.

Example: First we will give a simple example of a causal model which relate the effect E (a change in body temperature in degrees Celsius) and the dosage of the drug D. Every model depends on the characteristics of a patient. This set of characteristics can be formalised by some tuple $\vec{t}$, say $\vec{t} = \langle$age, genetic make-up, diet$\rangle$. Every model then maps every such tuple $\vec{t}$ to a real number. Let us call this mapping β. For instance, it may turn out that women taking a drug are more intensively affected by the side effect than men; also having a specific genetic make up and age may interact with the effect of the drug, while smoking habits may constitute a reason for suffering from the side effect

[19] In a real-world structural equation model, we would more richly describe the interrelations between D and other variables in the model, e.g., our variables for age, sex, and blood cholesterol. In our examples, we will only consider simple single equation models.

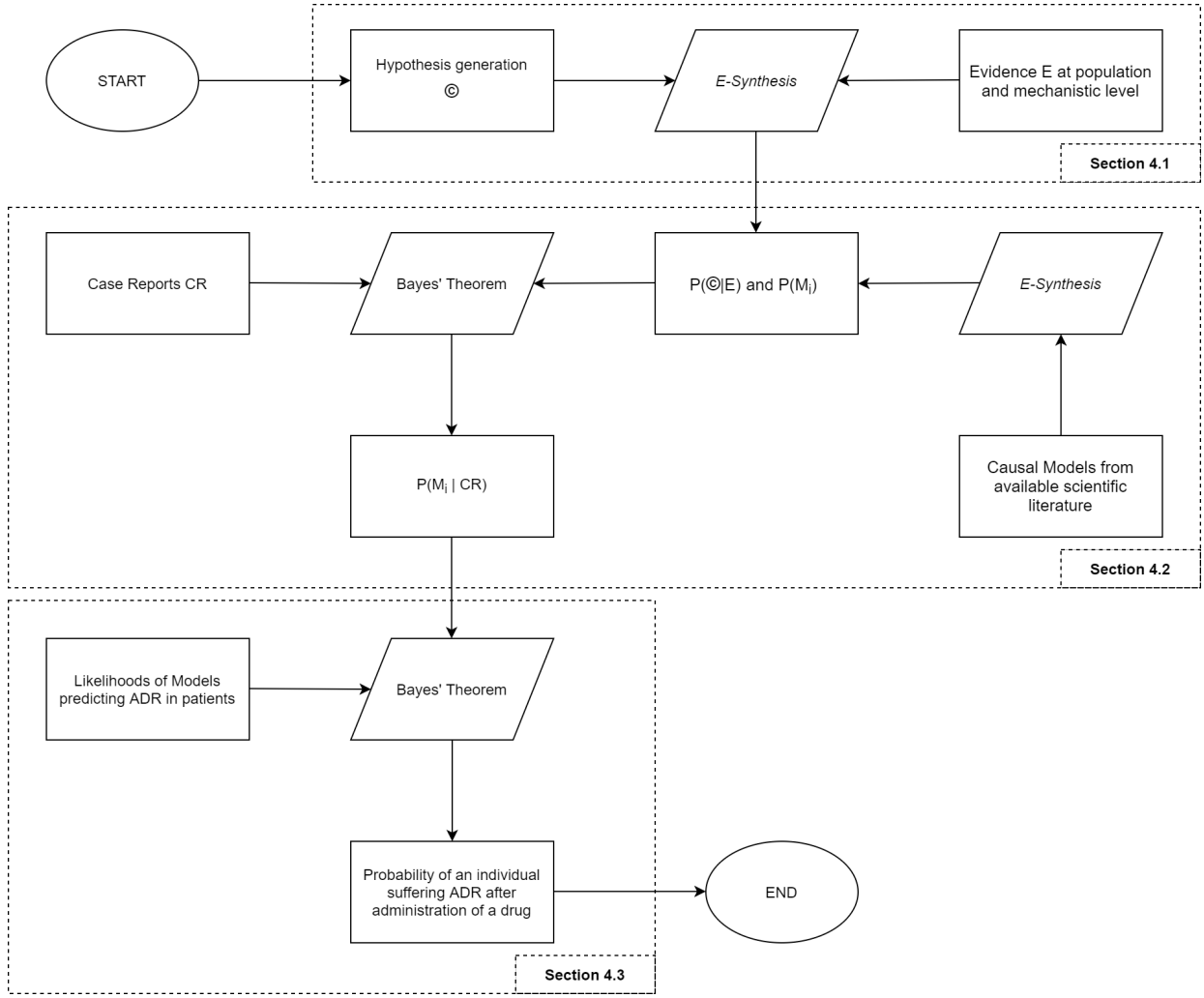

Figure 4.1 Flowchart representing the structure of our reasoning process. Parallelograms represent processes; rectangles represent inputs or outputs of such processes. Section 4.1: The overall process starts with the generation of a hypothesis of adverse drug reaction (ADR) on the basis of spontaneous reports or other signals from nature: ©; this is the first input to *E-Synthesis*, which then incorporates any other incoming evidence (at population, clinical and mechanistic level), by letting the data feed into the Bayesian network. Section 4.2: update of the causal hypothesis is mediated by averaging over available causal models, which the scientific literature puts at disposal. In our example, we then show how the observation of a series of clinical cases (Case Reports, CR) updates the probability distribution over the available causal models. Section 4.3: Once data about the individual patient are taken into account, his individual probability of being affected by the ADR can be then calculated through Bayes Theorem. Further details are explained in the text.

independently of drug consumption. We would put the interacting factors in the coefficient β, smoking in the intercept α, and leave an error term ϵ for other yet unidentified "disturbances". So now our causal model would look like:

$$E = \alpha + \beta D + \epsilon \ .$$

We may also consider non-linear models. For example, a set of causal models could just be:

$$\begin{aligned}\mathcal{M}_1: &\quad E = \alpha + \beta_1 D + \epsilon_1\\ \mathcal{M}_2: &\quad E = \alpha + \beta_2 D^2 + \epsilon_2\\ \mathcal{M}_3: &\quad E = \alpha + \beta_3 D + \beta_4 D^2 + \epsilon_3\ .\end{aligned}$$

The base rate α of of the effect in the population (distribution of fevers in the population) is for ease of exposition assumed to be independent of a particular model and thus it is the same in all models. The different models have, in general, different functional forms, different error terms and a different number of contributing terms.

4.2 Testing Causal Models/Hypotheses

4.2.1 Confirming Models

In order to test the models, we will use the information we have to discern among the models in the available case reports to learn which model is most likely.

At this stage of the inquiry, we have applied *E-Synthesis* to determine a probability of the drug causing an adverse reaction *in the general population* (©) and a "prior" probability distribution over the possible models $\mathcal{M}_i$, $P(\mathcal{M}_i)$. In order to update the probability distribution over the $\mathcal{M}_i$ on the available case reports we apply Bayes' Theorem.

Denoting the set of available case reports by CR, applying it here this theorem tells us that:

$$\begin{aligned}P(\mathcal{M}_i|CR \wedge ©) =& \frac{P(\mathcal{M}_i \wedge CR \wedge ©)}{P(CR \wedge ©)} = \frac{P(\mathcal{M}_i \wedge CR)}{P(CR \wedge ©)}\\ =& P(\mathcal{M}_i) \cdot \frac{P(CR|\mathcal{M}_i)}{P(CR \wedge ©)} = P(\mathcal{M}_i) \cdot \frac{P(CR|\mathcal{M}_i)}{\sum_l P(CR \wedge \mathcal{M}_l)} = P(\mathcal{M}_i) \cdot \frac{P(CR|\mathcal{M}_i)}{\sum_l P(\mathcal{M}_l)P(CR|\mathcal{M}_l)}\\ =& P(\mathcal{M}_i|©)P(©) \cdot \frac{P(CR|\mathcal{M}_i)}{\sum_l P(\mathcal{M}_l|©)P(©)P(CR|\mathcal{M}_l)} = \frac{P(\mathcal{M}_i|©) \cdot P(CR|\mathcal{M}_i)}{\sum_l P(\mathcal{M}_l|©)P(CR|\mathcal{M}_l)}. \qquad (1)\end{aligned}$$

The updated probability of a model can thus be calculated from prior probabilities of models ($P(\mathcal{M}_l|©)$) and likelihoods of observations ($P(CR|\mathcal{M}_l)$). The likelihoods are specified by the probabilistic models (see Section 4.2.2).

How have the posterior probabilities changed? The greater the likelihood of the observations given a model, the more of boost the model will obtain from the observations, *ceteris paribus*. Therefore, while there are many factors involved in confirming models in Bayesianism, the key issue for a model's evidential support by the observations is simply the likelihood of those observations given that model.

Example: Suppose we have $P(©) = 20\%$. The probability of the different models (being mutually exclusive and exhaustive) have to add up to 20%. So, let us suppose that the conditional probabilities of the models – given that © is true – are $25\%, 25\%, 50\%$ for $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ respectively. Now, suppose the likelihoods of the case reports for the three models are $P(CR|\mathcal{M}_1) = 0.05\%, P(CR|\mathcal{M}_2) = 0.1\%$ and $P(CR|\mathcal{M}_3) = 0.025\%$. We obtain a posterior probability of the models:

$$\begin{aligned}P(\mathcal{M}_1|CR \wedge ©) &= \frac{25\% \cdot 0.1\%}{25\% \cdot 0.05\% + 25\% \cdot 0.1\% + 50\% \cdot 0.025\%} = \frac{0.00025}{0.001} = 25\%\\ P(\mathcal{M}_2|CR \wedge ©) &= \frac{25\% \cdot 0.2\%}{25\% \cdot 0.05\% + 25\% \cdot 0.1\% + 50\% \cdot 0.025\%} = \frac{0.0005}{0.001} = 50\%\\ P(\mathcal{M}_3|CR \wedge ©) &= \frac{50\% \cdot 0.05\%}{25\% \cdot 0.05\% + 25\% \cdot 0.1\% + 50\% \cdot 0.025\%} = \frac{0.00025}{0.001} = 25\%\end{aligned}$$

.

We can now see that given CR, the probability of $\mathcal{M}_2$ conditional on © being true has doubled and it is now the most probable model, whereas the previously most probable model $\mathcal{M}_3$ is now equiprobable with the least likely model $\mathcal{M}_1$. This is intuitively plausible, e.g., CR is most probable if $\mathcal{M}_2$ is true, so learning CR gives the probability of $\mathcal{M}_2$ a

relatively large boost. Therefore, while the relative probabilities of the models are initially determined by their priors, updating with some evidence such as case reports can change the ordering of their probability, depending on their respective likelihoods for that evidence.

4.2.2 Likelihoods of Case Series

In order to determine the value of probability $P(\mathcal{M}_i|CR \wedge \textcircled{c})$, it insufficient to specify the conditional probability of the models given observations and ©. We also require the likelihood of the case series. The likelihood of the case series, provided that the observations are independent, is simply the product of the likelihoods of the individual patients suffering the adverse event within the case series, assuming that all sub-groups of patients have an equal probability of having their adverse effect reported. In case this last assumption is not appropriate, it is possible to use weighting factors to balance over/under-reporting of subgroups.

Example: Returning to our example, suppose that the case reports CR consist of only five reports for five patients: Rep_1, Rep_2, Rep_3, Rep_4, and Rep_5 . Given our assumptions so far, their likelihoods given a model and © may take a number of values; for illustrative purposes, we assume:

$$
\begin{array}{lll}
P(Rep_1|\mathcal{M}_1) = 5\% & P(Rep_1|\mathcal{M}_2) = 75\% & P(Rep_1|\mathcal{M}_3) = 50\% \\
P(Rep_2|\mathcal{M}_1) = 50\% & P(Rep_2|\mathcal{M}_2) = 20\% & P(Rep_2|\mathcal{M}_3) = 25\% \\
P(Rep_3|\mathcal{M}_1) = 20\% & P(Rep_3|\mathcal{M}_2) = 25\% & P(Rep_3|\mathcal{M}_3) = 8\% \\
P(Rep_4|\mathcal{M}_1) = 40\% & P(Rep_4|\mathcal{M}_2) = 33.\dot{3}\% & P(Rep_4|\mathcal{M}_3) = 10\% \\
\underbrace{P(Rep_5|\mathcal{M}_1) = 25\%}_{P(CR|\mathcal{M}_1)=0.05\%} & \underbrace{P(Rep_5|\mathcal{M}_2) = 80\%}_{P(CR|\mathcal{M}_2)=0.1\%} & \underbrace{P(Rep_5|\mathcal{M}_3) = 25\%}_{P(CR|\mathcal{M}_3)=0.025\%} .
\end{array}
$$

We show in the next section how to calculate the likelihood of a particular patient suffering an ADR according to some causal model $\mathcal{M}_i$ given the patient characteristics.

4.3 From Models to Individual Patients

We have performed the necessary steps to finally give a probability that the patient in front of us taking a certain dosage of the drug will suffer an adverse drug effect. As a simplifying measure, we focus on the case of uniform probability distributions for the conditional probabilities that a particular patient has an ADR given a set various characteristics.

If one of our causal models is indeed the one true model, then the model will predict how strong the side effect is going to be for this particular patient. The prediction consists of a probability density function assigning probabilities to effects given the dosage. The probability of an ADR, that is, an adverse event that is attributed to the drug's causal effects, is the probability of an adverse event occurring after taking the drug (whatever its cause), minus the base rate (the probability of such event occurring independently of drug intake).

Of course, we are unsure about which of our models represents the actual world, the predictions made by the models are thus weighted by their respective probabilities.

As for the problem of the reference class, our approach inherits the strengths and weaknesses of the standard Bayesian approach to this issue (Hájek 2007). On the one hand, given a suitably rich probability distribution, determining whether a patient known to possess various characteristics (a particular age group, having/not having a relevant gene etc.) will experience an ADR is just a matter of mathematical calculation. On the other hand, the question of how these priors might be determined rationally (if at all) is a heated question in the philosophy of statistics (the infamous "problem of the priors").

Example: Denoting by P' the probability of taking all but the case series into account, we now calculate the probability that an individual patient k will suffer the adverse event of strength E

$$P'(E_k|D) = P'(\overline{\copyright})\underbrace{P'(E_k|\overline{\copyright})}_{\text{base rate}} + P'(\copyright) \cdot \sum_i P'(\mathcal{M}_i|\copyright, CR)P'(E_k|D, \mathcal{M}_i)$$
$$= P'(\overline{\copyright})\alpha + P'(\copyright) \cdot \sum_i P'(\mathcal{M}_i|\copyright) \cdot [\alpha + \beta_i D^{\delta_i} + \epsilon_i] \ ,$$

where δ_i is the exponent of D in model $\mathcal{M}_i$, and *individual* is the hypothesis that the given patient suffers the side effect. The probability of suffering the adverse event *caused* by the administration of the drug is obtained by subtracting the base rate.

According to Bayes' Theorem (1) only prior probabilities and likelihoods are required to compute posterior probabilities. In particular, there is no need to determine which model/s is/are most relevant to a particular patient.

To determine the probability of a particular patient aged 65 years, who possesses a particular gene, has a daily salt intake of 5 grams per day and takes a daily dosage of 100 mg of drug D for some model $\mathcal{M}_j$ with $\alpha = 1\%$[20], a single coefficient on D , $\beta_j = 0.1(\text{age} - 60) + \text{gene} + 0.4 \cdot (\text{salt intake} - 3)$, $\delta_j = 1$ (linearity) and a normally distributed error term ϵ_j with mean $\mu_j = 0$. So, this patient has a β_j value of $0.5 + 1 + 0.4 \cdot 2 = 2.3$. The mean expected increase of temperature of the adverse effect under model $\mathcal{M}_j$ is thus 2.3 degrees Celsius, the probability density function of the strength of the adverse effect is a normal distribution with mean 2.3 and a variance given by the error term ϵ. To determine the probability of this particular patient suffering an adverse effect, one needs to determine such likelihoods for all models, weigh them according to their probabilities and the probability of $\copyright$, and add the probability of suffering the adverse effect without the drug D causing it (which is just the base rate α).

5 Discussion

Personalized medicine promises to accurately predict what would happen to a given individual, instantiating a given set of relevant characteristics, were he to receive a given treatment X. This promise relies on two points: 1) causal knowledge about the possible effects of X *in a given statistical population*; 2) assignment of the given individual to a suitable reference class. Regarding point 1, standard approaches to causal inference are generally considered to be characterized by a trade-off between how confidently one can establish causality in any given study (internal validity), and extrapolating such knowledge to specific target groups (external validity). Regarding point 2, it is irreducibly uncertain which reference class leads to the most reliable inferences (Reichenbach 1951; Salmon 1977). Furthermore, the reference class problem threatens all approaches to causal inference alike. Even if methods for dealing with population heterogeneity have been developed to different degree of sophistication (Reichenbach 1951; Pollock 1990; Hájek 2007; Pearl and Bareinboim 2014; Dahabreh et al. 2016; Yeh et al. 2018), the problem of identifying proper subgroups explaining variance in the causal effect is of a different nature to that of assigning a given individual to such subgroups.

By its very nature, pharmacovigilance is focused on both elements of the individual prediction at the same time, that is, the establishment of the possible causal link between a given drug and an observed adverse event, and the identification of possible subgroups, where such links may arise. This happens for several reasons: first of all, ADRs are by definition unintended effects that should occur in a minority of users, and it is important that such users are characterized as detailed as possible, in order to minimize the risk of future events in the population of consumers; secondly, risk signals from spontaneous reports, which are the first source of evidence for ADRs in the postmarketing phase, are primarily based on establishing whether the observed adverse event may be causally attributed to the drug or not (Karch and Lasagna 1977; Naranjo et al. 1981); thirdly, in the postmarketing phase it is of utmost importance to carefully weigh the drug benefits against its risks in specific subgroups of patients, so as to exclude from consumption only those groups for which the benefit-risk balance is unfavourable (Osimani and Mignini 2015).

[20] Strictly speaking, α should be a probability density function specifying how likely a fever of a particular temperature is.

We developed an epistemic framework that exploits the joint contribution of different dimensions of evidence, and, specifically for the present discussion, evidence about the causal link between drug and adverse event on one side, and evidence about contextual information related to the determinants of the effect variance in specific subpopulations on the other side. At the same time, this framework keeps the individual inputs of such sources of evidence conceptually and computationally distinct: this allows us to deal with the reference class problem not only by relying on statistical data about covariances, but also by drawing on causal knowledge. That is, the probability that a given individual will face a given side effect, will probabilistically depend on his features (e.g., age, genetic make-up, smoking, blood cholesterol etc.) *and* the plausible causal models in which such features become relevant. The evaluation of the causal models is grounded on the available evidence and theory.

We now briefly enumerate our major limiting assumptions and discuss their impact.

1. There are the limitations our analysis inherited from *E-Synthesis* such as the choice of indicators of causation, the Bayesian tenets and limitations of the Bayesian machinery. These limitations have been discussed in previous work (Abdin et al. 2019; Landes et al. 2018; De Pretis et al. 2019; De Pretis and Osimani 2019).

2. We only use case series to inform the updating of the beliefs in the models in Section 4.2. It is desirable – even normatively required (Carnap 1947, § 3) – to use all the available evidence to inform rational beliefs and make decisions. We must leave it to future work to also use multimodal evidence.

3. The current state of the art of *E-Synthesis* still uses a binary variable © to reason about causation in the general population. Incorporating strength of causation would be desirable, in particular in light of the result of our approach here which produces a probability density function capturing the intensity of causation (β).

4. We employ some notion of causation which is expressible in terms of (structural) equations. While this view is virtually unchallenged in the social sciences, it is not without critics in philosophy. For example, causal pluralists (Reiss 2009) might deny that all of the concepts of causation in medicine can be formalised by structural equations. Furthermore, we can't know a priori which models – or which parameters – have to be considered into our ensemble of structural equations. However, we resort here to a Bayesian approach, allowing a certain grade of uncertainty attaching to choices of parameters/models.

5. We limited ourselves to causal models $\mathcal{M}_i$ which are mutually exclusive to make use of the standard Bayesian machinery. In practice, one may want to consider models which are not mutually exclusive, because different models specify the causal connection in different levels of detail, e.g., model $\mathcal{M}_1$ might have a wider error distribution rather than $\mathcal{M}_2$ due to the latter being able to explain more variance. For non-exclusive models one cannot apply Bayesian updating. In that case, one can use comparative measures of evidence, such as Bayes factors, to inform beliefs about which of the models (best) describes the actual world.

6. We assumed that the different causal models are jointly exhaustive, which was not realistic. The problem of Bayesian learning in the absence of a fully-specified partition of possible alternative hypotheses is a major issue in Bayesian philosophy of science. One option is to define a "catch-all" hypothesis, to assign it non-zero probability, and to allocate probability from the "catch-all" to alternative hypotheses as they are conceived (Shimony 1970; Wenmackers and Romeijn 2016). There are alternative approaches to this issue in Bayesianism (Salmon 1990).

There are a number of possible directions for future research. One is relating our discussion more closely to evidence assessment in medical practice and the many disagreements among practitioners (Stegenga 2014b). We shall now discuss the conception and formulation of these new causal models. In pharmacovigilance, hypotheses about side-effects may be (jointly) generated by various sources: asymmetries in databases of spontaneous case reports, clinical case studies, basic science, animal studies etc. (De Pretis and Osimani 2019). At the time of its generation the hypothesis about drug-induced side-effects is relatively unspecific, while it progressively becomes more

articulated as new evidence of various sources comes in. The incoming evidence may be informative about particular subgroups showing specific patterns of reactions to a range of drug dosage. These diverse types of information allows one to progressively refine the entertained causal models by pointing to relevant factors, their interactions and their importance (Stegenga 2014a; Reiss 2015). For instance, observations relating drug dosage and strength of effect made in populations or basic science can be informative about the functional form of the relationship between drug and effect, regardless of whether the relationship is causal or not. Furthermore, the formulation of more specific models makes variables salient, which were previously incorporated in the error term. By having such variables being represented explicitly in the model equations, the random error/noise in the model is reduced.

Acknowledgements

Francesco De Pretis, Barbara Osimani and William Peden acknowledge funding from the European Research Council (PhilPharm - GA n. 639276) and the Marche Polytechnic University (Ancona, Italy). Jürgen Landes gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 432308570 and 405961989.

ORCID

Francesco De Pretis: orcid.org/0000-0001-8395-7833
Jürgen Landes: orcid.org/0000-0003-3105-6624
Barbara Osimani: orcid.org/0000-0001-5212-9525
William Peden: orcid.org/0000-0002-3474-7861

References

Abdin, Ahmad Y, Daniel Auker-Howlett, Jürgen Landes, Glorjen Mulla, Claus Jacob, and Barbara Osimani. 2019. Reviewing the mechanistic evidence assessors e-synthesis and ebm+: A case study of amoxicillin and drug reaction with eosinophilia and systemic symptoms (dress). *Current Pharmaceutical Design* 25 (16): 1866–1880. doi:10.2174/1381612825666190628160603.

Andreoletti, Mattia, and David Teira. 2019. Rules versus standards: what are the costs of epistemic norms in drug regulation? *Science, Technology, & Human Values* 44 (6): 1093–1115.

Anjum, Rani Lill, and Elena Rocca. 2019. From ideal to real risk: philosophy of causation meets risk analysis. *Risk Analysis* 39 (3): 729–740.

Aronson, Jeffrey K, and Robin E Ferner. 2003. Joining the dots: new approach to classifying adverse drug reactions. *Bmj* 327 (7425): 1222–1225.

Bogen, James, and James Woodward. 1988. Saving the phenomena. *The Philosophical Review* 97 (3): 303–352. https://doi.org/10.2307/2185445.

Bovens, Luc, Stephan Hartmann, et al.. 2003. *Bayesian epistemology*. Oxford: Oxford University Press.

Carnap, Rudolf. 1947. On the Application of Inductive Logic. *Philosophy and Phenomenological Research* 8 (1): 133–148. https://doi.org/10.2307/2102920.

Cartwright, Nancy, and Jacob Stegenga. 2011. A Theory of Evidence for Evidence-Based Policy. In *Evidence, Inference and Enquiry*, eds. Phil Dawid and Mimi Twinning William Vasilaki, 291–322. OUP. Chap. 11.

Cowell, Robert G, Philip Dawid, Steffen L Lauritzen, and David J Spiegelhalter. 2006. *Probabilistic networks and expert systems: Exact computational methods for bayesian networks*. Springer.

Dahabreh, Issa J, Rodney Hayward, and David M Kent. 2016. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *International Journal of Epidemiology* 45 (6): 2184–2193.

Darwiche, Adnan. 2009. *Modeling and reasoning with bayesian networks*. Cambridge: Cambridge University Press.

Dawid, A. Philip. 2010. Beware of the dag! In *Proceedings of workshop on causality: Objectives and assessment at nips 2008*, eds. Isabelle Guyon, Dominik Janzing, and Bernhard Schölkopf. Vol. 6 of *Proceedings of machine learning research*, 59–86. Whistler, Canada: PMLR. http://proceedings.mlr.press/v6/dawid10a.html.

De Pretis, Francesco, and Barbara Osimani. 2019. New insights in computational methods for pharmacovigilance: E-synthesis, a bayesian framework for causal assessment. *International Journal of Environmental Research and Public Health* 16 (12). doi:10.3390/ijerph16122221. https://www.mdpi.com/1660-4601/16/12/2221.

De Pretis, Francesco, Jürgen Landes, and Barbara Osimani. 2019. E-synthesis: A bayesian framework for causal assessment in pharmacosurveillance. *Frontiers in Pharmacology* 10: 1317. doi:10.3389/fphar.2019.01317. https://www.frontiersin.org/article/10.3389/fphar.2019.01317.

Enjo B, Aisabokhae E. 2018. Pharmacovigilance literature review in the age of precision medicine, Technical report, Arthur D. Little Services SAS. https://www.adlittle.de/sites/default/files/viewpoints/adl_a_new_approach_to_pv_literature_0.pdf.

Ennezat, Pierre-Vladimir, Shona Cosgrove, Helene Bouvaist, Sylvestre Marechaux, Raphaelle-Ashley Guerbaai, Thierry Le Jemtel, Michel Andrejak, and Denis Vital-Durand. 2017. From evidence-based medicine to personalized medicine, with particular emphasis on drug-safety monitoring. *Archives of cardiovascular diseases* 110 (6-7): 413–419.

Fang, Hong, Stephen C Harris, Zhichao Liu, Guangxu Zhou, Guoping Zhang, Joshua Xu, Lilliam Rosario, Paul C Howard, and Weida Tong. 2016. Fda drug labeling: rich resources to facilitate precision medicine, drug safety, and regulatory science. *Drug discovery today* 21 (10): 1566–1570.

Franklin, James. 2011. The objective bayesian conceptualisation of proof and reference class problems. *Sydney L. Rev.* 33: 545.

Geiger, Dan, Thomas Verma, and Judea Pearl. 1990. d-separation: From theorems to algorithms. In *Uncertainty in artificial intelligence*, 139–148. Elsevier. doi:10.1016/b978-0-444-88738-2.50018-x. https://doi.org/10.1016/b978-0-444-88738-2.50018-x.

Gelman, Andrew, and Christian Hennig. 2017. Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180 (4): 967–1033. https://doi.org/10.1111/rssa.12276.

Guchet, Xavier. In this volume. Exposomics in the era of personalized medicine: A critical analysis. In *Personalized medicine in the making. Philosophical perspectives from biology to healthcare*, eds. Chiara Beneduce, and Marta Bertolaso. Springer.

Hájek, Alan. 2007. The reference class problem is your problem too. *Synthese* 156 (3): 563–585.

Hawthorne, James. 2005. Degree-of-belief and degree-of-support: Why bayesians need both notions. *Mind* 114 (454): 277–320.

Hempel, Carl G. 1962. Inductive inconsistencies. In *Logic and language*, 128–158. Dordrecht: Springer.

Hill, Austin Bradford. 1965. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* 58 (5): 295–300.

Howson, Colin, and Peter Urbach. 2006. *Scientific reasoning: the bayesian approach*. Chicago and La Salle: Open Court Publishing.

Karch, Fred E, and Louis Lasagna. 1977. Toward the operational identification of adverse drug reactions. *Clinical Pharmacology & Therapeutics* 21 (3): 247–254.

Kent, David M, Ewout Steyerberg, and David van Klaveren. 2018. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *Bmj* 363.

La Russa, Raffaele, Vittorio Finesch, Mariantonia Di Sanzo, Vittorio Gatto, Alessandro Santurro, Gabriella Martini, Matteo Scopetti, and Paola Frati. 2017. Personalized medicine and adverse drug reactions: the experience of an italian teaching hospital. *Current pharmaceutical biotechnology* 18 (3): 274–281.

Landes, Jürgen, Barbara Osimani, and Roland Poellinger. 2018. Epistemology of causal inference in pharmacology. *European Journal for Philosophy of Science* 8 (1): 3–49.

Luo, Wei, Wenbo Wu, and Yeying Zhu. 2019. Learning heterogeneity in causal inference using sufficient dimension reduction. *Journal of Causal Inference* 7 (1).

Masimirembwa, Collen, Collet Dandara, and Peter Derek Christian Leutscher. 2016. Rolling out efavirenz for hiv precision medicine in africa: are we ready for pharmacovigilance and tackling neuropsychiatric adverse effects? *Omics: a journal of integrative biology* 20 (10): 575–580.

Montévil, Maël. In this volume. Conceptual and theoretical specifications for accuracy in medicine. In *Personalized medicine in the making. Philosophical perspectives from biology to healthcare*, eds. Chiara Beneduce, and Marta Bertolaso. Springer.

Naranjo, Cláudio A, Usoa Busto, Edward M Sellers, P Sandor, I Ruiz, EA Roberts, E Janecek, C Domecq, and DJ Greenblatt. 1981. A method for estimating the probability of adverse drug reactions. *Clinical Pharmacology & Therapeutics* 30 (2): 239–245.

Neapolitan, Richard E. 2004. *Learning bayesian networks*. Upper Saddle River, NJ: Pearson Prentice Hall.

Nilsson, Anton, Carl Bonander, Ulf Strömberg, and Jonas Björk. 2019. Assessing heterogeneous effects and their determinants via estimation of potential outcomes. *European journal of epidemiology* 34 (9): 823–835.

Osimani, Barbara. 2007. Probabilistic information and decision making in the health context: the package leaflet as basis for informed consent. PhD diss, Doctoral Thesis, Università della Svizzera italiana.

Osimani, Barbara. 2008. The epistemic nature of package leaflet information. *Medic* 2007/2008: 72–90.

Osimani, Barbara. 2013. The precautionary principle in the pharmaceutical domain: a philosophical enquiry into probabilistic reasoning and risk aversion. *Health, risk & society* 15 (2): 123–143.

Osimani, Barbara. 2020. Epistemic gains and epistemic games: Reliability and higher order evidence in medicine and pharmacology. In *Boston studies in the philosophy and history of science*, 345–372. Springer. doi:10.1007/978-3-030-29179-2_15. https://doi.org/10.1007/978-3-030-29179-2_15.

Osimani, Barbara, and Fiorenzo Mignini. 2015. Causal assessment of pharmaceutical treatments: why standards of evidence should not be the same for benefits and harms? *Drug safety* 38 (1): 1–11.

Pearl, Judea. 2009. *Causality Models, Reasoning and Inference*, 2nd edn. Cambridge: Cambridge University Press.

Pearl, Judea. 2012. The do-calculus revisited. In *Proceedings of the twenty-eighth conference on uncertainty in artificial intelligence*. *Uai'12*, 3–11. Arlington, Virginia, USA: AUAI Press. ISBN 9780974903989.

Pearl, Judea, and Elias Bareinboim. 2014. External validity: From do-calculus to transportability across populations. *Statistical Science*.

Pollock, John L. 1990. *Nomic probability and the foundations of induction*. Oxford University Press on Demand.

Plutynski, Anya. In this volume. Why precision oncology is not very precise (and why this should not surprise us). In *Personalized medicine in the making. Philosophical perspectives from biology to healthcare*, eds. Chiara Beneduce, and Marta Bertolaso. Springer.

Reichenbach, Hans. 1951. *The theory of probability: An inquiry into the logical and mathematical foundations of the calculus of probability*. Los Angeles: University of California Press.

Reiss, Julian. 2009. Causation in the social sciences: Evidence, inference, and purpose. *Philosophy of the Social Sciences* 39 (1): 20–40. doi:10.1177/0048393108328150.

Reiss, Julian. 2015. A pragmatist theory of evidence. *Philosophy of Science* 82 (3): 341–362. doi:10.1086/681643.

Reynolds, Michael S. 2015. Pharmacovigilance and data for precision medicine. *Journal of Precision Medicine*.

Rocca, Elena, Rani Lill Anjum, and Stephen Mumford. 2020. Causal insights from failure: Post-marketing risk assessment of drugs as a way to uncover causal mechanisms. In *Boston studies in the philosophy and history of science*, 39–57. Springer. doi:10.1007/978-3-030-29179-2_2. https://doi.org/10.1007/978-3-030-29179-2_2.

Rocca, Elena, Samantha Copeland, and I Ralph Edwards. 2019. Pharmacovigilance as scientific discovery: An argument for trans-disciplinarity. *Drug safety* 42 (10): 1115–1124.

Salmon, Wesley. 1990. Rationality and objectivity in science or tom kuhn meets tom bayes. *Scientific theories* 14: 175–204.

Salmon, Wesley C. 1977. Objectively homogeneous reference classes. *Synthese*.

Shimony, Abner. 1970. Scientific inference. In *The nature and function of scientific theories*, ed. Robert Colodny. University of Pittsburgh Press.

Smith, Silas W, Manfred Hauben, and Jeffrey K Aronson. 2012. Paradoxical and bidirectional drug effects. *Drug safety* 35 (3): 173–189.

Sprenger, Jan. 2018. The objectivity of subjective bayesianism. *European Journal for Philosophy of Science* 8 (3): 539–558.

Stegenga, Jacob. 2014a. Down with the hierarchies. *Topoi* 33 (2): 313–322. doi:10.1007/s11245-013-9189-4.

Stegenga, Jacob. 2014b. Herding QATs: Quality assessment tools for evidence in medicine. In *Classification, disease and evidence*, 193–211. Springer. doi:10.1007/978-94-017-8887-8_10. https://doi.org/10.1007/978-94-017-8887-8_10.

Stoffi, Falco Joannes Bargagli, and Giorgio Gnecco. 2018. Estimating heterogeneous causal effects in the presence of irregular assignment mechanisms. In *2018 IEEE 5th international conference on data science and advanced analytics (DSAA)*. IEEE. doi:10.1109/dsaa.2018.00010. https://doi.org/10.1109/dsaa.2018.00010.

Strevens, Michael. 2016. The reference class problem in evolutionary biology: distinguishing selection from drift. *Chance in Evolution. University of Chicago Press, Chicago*.

Talbott, William. 2011. Bayesian epistemology. *EN Zalta (Principal Ed.), Stanford encyclopedia of philosophy. Summer 2011 Edition*. https://plato.stanford.edu/entries/epistemology-bayesian/.

Teira, David. 2013. On the impartiality of early british clinical trials. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (3): 412–418.

Teira, David, and Julian Reiss. 2013. Causality, impartiality and evidence-based policy. In *Mechanism and causality in biology and economics*, 207–224. Springer.

Venn, John. 2013. *The logic of chance*. Courier Corporation.

Wenmackers, Sylvia, and Jan-Willem Romeijn. 2016. New theory about old evidence. *Synthese* 193 (4): 1225–1250. https://doi.org/10.1007/s11229-014-0632-x.

Wertheimer, Albert. 2016. The impact of personalized medicine on pharmacovigilance. *Journal of Pharmacovigilance* 3: 1000–133. doi:10.4172/2329-6887.1000e133.

Xie, Yu. 2013. Population heterogeneity and causal inference. *Proceedings of the National Academy of Sciences* 110 (16): 6262–6268.

Yeh, Robert W, Linda R Valsdottir, Michael W Yeh, Changyu Shen, Daniel B Kramer, Jordan B Strom, Eric A Secemsky, Joanne L Healy, Robert M Domeier, Dhruv S Kazi, et al.. 2018. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *bmj* 363: 5094.