ORIGINAL ARTICLE

Journal Section

Artificial intelligence methods for a Bayesian epistemology-powered evidence evaluation

Francesco De Pretis PhD[1 2] | **Jürgen Landes PhD**[3] | **William Peden PhD**[1 4]

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

[2]Department of Communication and Economics, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy

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

[4]Department of Philosophy, Durham University, DH1 3LE Durham, United Kingdom

Correspondence
William Peden PhD
Department of Biomedical Sciences and Public Health, Marche Polytechnic University, Ancona AN, 60126, Italy
Email: w.j.peden@durham.ac.uk

Rationale, aims and objectives: The diversity of safety signals (e.g., case reports, animal studies and observational studies) makes the assessment of the (un-)safety of a drug a formidable challenge. While frequentist uncertain inference struggle in aggregating these signals, the more flexible Bayesian approaches seem better suited for this quest. Artificial Intelligence (AI) offers great promise to these approaches for information retrieval, decision support and learning probabilities from data.

Method: *E-Synthesis* is a Bayesian framework for drug safety assessments build on philosophical principles and considerations. It aims to aggregate all the available information, in order to provide a Bayesian probability of a drug causing an adverse reaction. We delineate and assess ways in which AI can support *E-Synthesis*.

Results: We find that AI can help with information retrieval, usability (graphical decision making aids), learning Bayes factors from historical data, assessing quality of information and determining conditional probabilities for the so-called "indicators" of causation for *E-Synthesis* .

Conclusions: Properly applied, AI can help the transition of philosophical principles and considerations concerning evidence aggregation for drug safety to a tool that can be used in practise.

KEYWORDS
Artificial intelligence, Drug safety, Evidence evaluation, *E-Synthesis*, Pharmacosurveillance, Pharmacovigilance

1 | INTRODUCTION

Every day, doctors, hospitals, pharmaceutical companies, and others in healthcare face the complexities of the human body and the healthcare environment. There are huge masses of diverse possibly relevant data which, if harnessed properly, can improve the quality of treatment, and if used poorly, can lead to disasters like thalidomide and Lyodura.

Given the challenge of interpreting such varieties of data, it is clear that AI has an important role to play in healthcare.

In fact, it has already had a major impact. Telehealth agencies such as the NHS 24 Self-Help guide[1] use automated reasoning to help patients self-diagnose. An AI system powered by Google LLC predicted hospital inpatient death risks with 95% accuracy[1]. In January 2020, the first AI-developed drug, DSP-1181 (a treatment for obsessive compulsive disorder) entered clinical trials[2]. AI can also made a contribution to diagnostic procedures by doctors[2]. (See[3] for a general overview of AI for medical diagnoses.) The idea of a "smart" hospital, with programs and devices coordinated by AI, is no longer just science fiction[4]. AI also has roles to play in identifying drug interactions interpreting possibly minute details in images, logging and processing health records, and more. Still, rigorous research into the performance of AI in many of these areas is still in its infancy[5;6]. AI's use for public health more widely is at more of a prospective stage, but its potential is obvious[7].

In this article, we focus on pharmacosurveillance and we explore how AI can contribute to the continuous assessment of putative Adverse Drug Reactions (ADRs). This manuscript is organized as follows: in the Methods section, we briefly present *E-Synthesis*, a framework for combining different types of evidence in pharmacovigilance, based on Bayesian epistemology. Then, we describe AI methods actually employed in the realm of evidence synthesis. In the Results section, we assess their potential impacts on the usability and applicability of *E-Synthesis* . Finally, in the Discussion section, we offer some concluding remarks and provide an outlook on a possible research agenda in drug safety assessment.

2 | METHODS

The synthesis of evidence from multiple sources of providing different kinds of information (randomized studies, observational studies, case reports, in vitro evidence), with the aim of evaluating hypotheses and making decisions, plays a fundamental role in in many areas of medicine. In pharmacosurveillance, for instance, relevant evidence only becomes available in an unsystematic and motley way, so that evaluating hypotheses is far from the textbook ideal of interpreting a neat result from a randomized controlled trial. Thus, there is a need for methods of synthesis that assess the significance of heterogeneous evidence in a systematic, well-grounded, and manageable way. Since traditional frequentist statistical methods struggle with aggregating different kinds of information, a more flexible approach is required here. We next present a Bayesian approach to drug safety assessment, and then we outline how AI methods can serve evidence synthesis. The interaction between AI and this Bayesian approach will be explored in the Results section.

2.1 | *E-Synthesis*: Bayesian epistemology for evidence aggregation in pharmacovigilance

E-Synthesis is a Bayesian framework for evidence evidence aggregation in pharmacosurveillance to support timely decision making based on all the available 'safety signals'[8;9;10;11;12]. The framework rests on Bayesian epistemology, which unlike Bayesian statistics enables representation of and reasoning with uncertainties attaching to arbitrary propositions. In previous papers, we have presented its philosophical foundations[8], studied the incorporation evidence qualities[11], investigated the aggregation of knowledge concerning biological mechanisms and dose-response[10;9], and made strides towards applying *E-Synthesis* in personalized medicine[12]. Next, we give a brief overview of *E-Synthesis*.

[1] https://www.nhs24.scot

[2] https://www.exscientia.ai/news-insights/sumitomo-dainippon-pharma-and-exscientia-joint-development

2.1.1 | Motivation and goal

The risk-benefit profile of a drug is assessed and updated throughout the development process: after its formula is proposed, during its synthesization, and in the post-marketing period. There is no point at which its safety is definitively established: its developers and drug regulators must make multiple judgements at different phases of development, using heterogenous evidence, such as whether to withdraw the drug. Currently, these decisions are made using systematic reviews that combine the wide variety of available evidence (pre-clinical studies, clinical trials, spontaneous reports, basic research etc.) to justify or undermine hypotheses about the presence or absence of causal relations between the drug and harms. However, it is difficult to combine heterogeneous data with various sources, modalities (observational vs. experimental) and different degrees of external and internal validity. The ultimate objective of *E-Synthesis* is to surmount this difficulty, by providing a systematic, epistemologically principled, and usable method for combining evidence.

This framework rests on the paradigmatic philosophical account of uncertain inference (Bayesian epistemology) in order to provide a theoretically justified probability of a drug causing a harm on the basis of all the available evidence. It employs a Bayesian network[13] incorporating indicators of causality derived from Bradford-Hill "guidelines"[14] as well as evidence qualities and uncertainties attaching to these evidence qualities. Unlike the GRADE approach which is not straight-forwardly applicable to decision problems[15], the probability produced by *E-Synthesis* has been designed to be used for making decisions via the maximization of expected utilities.

2.1.2 | Bayesian networks

In order to have an inferential mechanism that can handle heterogeneous types of evidence, *E-Synthesis* utilises the tools of Bayesian networks and Bayesian epistemology. We provide a brief introduction to these ideas and the rationale of their implementation in *E-Synthesis*.

Bayesian epistemology is a philosophical theory about (a) what sort of beliefs and strength ("degree") of beliefs can be rational in a particular context and (b) how those beliefs should be revised upon learning new evidence. Bayesianism formalises degrees of beliefs as probabilities; it thus inherits the formal constraints of the probability calculus. Thus, $P(H)$ represents a researcher's degree of belief in a hypothesis H, while $P(H \mid \mathcal{E})$ represents their degree of belief in H conditional on acquiring evidence $\mathcal{E}$. In the case where our hypothesis is that of the drug causing an ADR (denoted by ©), this conditional probability can be determined using Bayes' Theorem:

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

where the hypotheses H_i and $© = H_1$ constitute a mutually inconsistent and exhaustive partition.[3]

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 $P(\mathcal{E}|H_i)$[4]. Bayesian epistemology focuses on updating (or "conditionalising") for proposition or events in general, whereas in Bayesian statistics focuses on testing statistical models using conditional probabilities.

[3] For convenience, we use the same symbol denoting a variable and the variable being true.

[4] The likelihoods are often (but not always) easy to determine, because the content of the hypothesis will often determine a probability for the evidence due to logical or mathematical reasons. For example, if a hypothesis (with a non-zero prior probability) implies the evidence, then the likelihood must be 1. Meanwhile, determining the likelihood of the evidence given a statistical hypothesis H_i often just requires using purely mathematical reasoning, e.g. calculating the probability of a particular series of independent and identically distributed binomial trials given the hypothesis of a population frequency.

It is generally very difficult to calculate conditional probabilities directly or to make a long and complex series of inferences using them. Bayesian networks offer a convenient means for graphically displaying and reasoning with probability functions [16;13]. We can use them to specify and read-off conditional independencies from a graph. Technically, a Bayesian network is defined on a set of pairwise different variables by a directed acyclic graph (edges are directed such that the graph does not contain a directed cycle, i.e., it has no path of directed edges which leads back to its starting point) and a probability distribution specifying the conditional probabilities of all variables given their parent variables (all other variables which directly point to this variable), see Fig.1 for an example graph.

Technically, this works as follows. Denoting the parents of a variable 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 $\sum_{y \in Y} P(Y = y | X_1 = x_1, \ldots, X_n = x_n) = 1$. This condition ensures that we have defined a probability function that satisfies the standard probability calculus. To calculate conditional and unconditional probabilities of interest, one may use the so-called "chain rule".

2.1.3 | Indicators of causation

Bayes' theorem is essential in Bayesian epistemology, but it is by no means clear how to determine the likelihoods $P(\mathcal{E}|H_i)$ in pharmacovigilance. To facilitate this task, we employ abstract indicators of causality that are derived from Bradford Hill Guidelines: 1) difference making, 2) probabilistic dependence, 3) dose-response relationship, 4) rate of growth, 5) temporal precedence and 6) mechanistic knowledge. Conceptually, indicators of causality are testable (probabilistic) consequences of the causal hypothesis. For example, we can test whether there is a dose-response relationship between drug and adverse effect (higher dosages lead to more and/or stronger adverse effects). However, note that neither does the presence of a causal relationship entail the presence of a dose-response relationship (anaphylaxis) nor does the presence of a dose-response relationship entail a causal relationship, due to confounding. The indicators are probabilistic consequences in these sense that their truth is more likely, if the hypothesis is also true, than if the latter is false. $P(Ind|\copyright) > p(Ind) > P(Ind|\overline{\copyright})$. In turn $P(\copyright|Ind) > P(\copyright) > P(\copyright|\overline{Ind})$.

Thus, there is an association between each relevant experimental study, observational study, case series, case report or basic science finding with a set of causal indicators which it is informative about [17]. *E-Synthesis* thus analyses 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.

A core idea of Bayesian epistemology is that the confirmatory value of evidence with respect to hypotheses is degree-valued. The same holds here with respect to evidence for or against our causal indicators. We use evidential modulators to make this fine-grained and incremental element in Bayesian reasoning explicit, by determining 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. 1.

2.1.4 | Evidential modulators

One key feature of *E-Synthesis* is the possibility to use assessments of the quality of items of evidence. The assessed quality of evidence then modulates the degree to which the item of evidence (dis-)confirms indicators of causation. This is achieved by first creating a "report" variable for every item of evidence and then creating for every such variable a set of pertinent modulator variables $Q_1, \ldots, Q_k$, e.g., duration of a study, sample size and blinding. In the Bayesian network, these modulator variables are, together with a set of indicator variables, the parents of the report variable. According to the Bayesian approach one then needs to set the conditional probabilities of observing the evidence

given their qualities and given the values of the indicator variables, $P(Rep = rep|Ind = ind, Q_1 = q_1, \ldots, Q_k = q_k)$.[5]

An application of Bayes' Theorem enables one then to calculate the posterior probability of causal indicators which in turn can be used to calculate the posterior probability of the causal hypothesis that the drug causes an ADR in the population of interest.

2.2 | AI for evidence synthesis

As outlined in the Introduction, AI already has a growing impact on healthcare. However, its potential for evidence synthesis is still undeveloped[6]. This is despite cautious interest within parts of the healthcare industry [7]. Greater use of AI in evidence synthesis could have many benefits. One is transparency: an algorithm for evidence synthesis can be made public in ways that can be difficult or impossible for human judgements. Another is computational scale: the accelerating increase in medical data means that the application of AI in evidence synthesis is a growing issue. Insofar as evidence syntheses depend on a lot of human input, it will be difficult to keep track of the ever-greater flow of information e.g. from case reports and clinical trials. Automation can help alleviate some of the strains in the evidence synthesis process.

We stress that the automation of the *entire* evidence synthesis process is not a currently realistic goal. Instead, a plausible ambition is what has been called "semi-automated evidence synthesis" [20] in which parts (perhaps even a majority) of the evidence synthesis process is automated using AI software. This would make evidence synthesis more manageable and transparent, while preserving vital roles for human judgement in many parts of the process. Some researchers are already pursuing such goals on a grand scale [21].

AI can contribute in many ways to the semi-automation research programme. For instance, inference of causality from heterogeneous data have been explored [22], so as semi-automatic transferring of knowledge from one field to another by analogy [23;24]. Moreover, particular efforts have been deployed on machine learning. Machine learning focuses on computer algorithms such that the computers can perform tasks without being expressly compiled to do as such. This AI field utilizes different methodologies. There is a particular interest in two perspectives: supervised and unsupervised learning [25]. Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs. Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs. An algorithm that improves the precision of its outputs after some time is said to have learned how to play out that task. In contrast, unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labeled, classified or categorized. Unsupervised learning is usually considered the most advanced edge of research in this field. For example, machine learning methods like text mining can help to screen studies for relevance [26]. There is also research on automating the extraction of relevant data from particular studies [27]. It might even be possible to create what has recently been dubbed "living systematic reviews": once an evidence synthesis has been completed, there will be automated identification of relevant subsequent research and extraction of the data that directly addresses the subject of the evidence synthesis. Human input would only be required to check the results of this process (which will be imperfect) once it has been completed [28].

We deem that *E-Synthesis* can contribute to this research programme in two ways:

[5] Uncertainty about study qualities are represented by probabilities in the fashion usual in Bayesian statistics, e.g., $P(Q_i = q_i)$.

[6] The first automated evidence synthesis system was only published in 2019 [18]. See [19] for a recent overview of evidence synthesis automation.

[7] https://blog.evidencepartners.com/past-present-and-future-automation-in-systematic-review-software

1. It is a formal evidence synthesis procedure. Hence, it should ultimately be amenable to semi-automation.
2. It has a solid methodological basis in Bayesian epistemology and medical practice. *E-Synthesis* fits well with the growing Bayesian paradigm in both statistical practice and the philosophy of science. It also incorporating scientists' own successful patterns of reasoning in elements like indicators of causality, as we describe below.

From the other direction, AI itself can improve *E-Synthesis*. In the next section, we explore in detail what may come out from the interactions of *E-Synthesis* and AI.

3 | RESULTS

As we have seen, AI methods are already employed in the realm of evidence amalgamation and may effectively contribute to a better functioning of *E-Synthesis* (Section 2.2). That framework puts forward a decision making model to support drug safety assessments, which are usually performed in a collective way by advisory committees, panels of experts consulting Drug Agencies [29]. However, significant parts of *E-Synthesis* are still left to experts and are not automated. For instance, the strengths of how strongly different evidential modulators (Section 2.1.4) influence confirmation is still input *manually* by the introduction of an *ad hoc* weighting scheme. The application of machine learning and other AI techniques could lead to remarkable improvements of the quality of decisions.

In the following subsections, we identify three main areas of interaction between *E-Synthesis* and AI: machine learning, information retrieval and graphical decision aids. We conclude that evidence synthesis for pharmacosurveillance can be enhanced by AI.

3.1 | Machine Learning

With the aim of creating automated systems that make better use of the vast amount of accumulating publications and promoting the uptake of that evidence into a wide range of contexts, machine learning may represent the backbone of *E-Synthesis*. We deem that, using machine learning, *E-Synthesis* will be enhanced in identifying, extracting, synthesizing and interpreting relevant information, converting this into knowledge that can answer complex questions over causal associations. We identify two main applications of machine learning for improving *E-Synthesis*: 1) estimation of (conditional) probabilities of causal indicators and learning the weighting schemes of the evidential modulators from data and 2) modelling the "linkage between a direct molecular initiating event [..] and an adverse outcome at a biological level of organization relevant to risk assessment" [30] P. 731. Such "mechanisms" play an important inferential role [31].

3.1.1 | Assessing probabilities and predictive powers

As shown above, *E-Synthesis* delivers a probability of causal association between a drug and an ADR, based on a Bayesian updating of evidence that accrues through causal indicators. Machine learning could help *E-Synthesis* in:

- **Learning the weighting scheme of the evidential modulators**. The task is to determine how likely a study (observational or an RCT) is to correctly identify the absence or presence of a causal relationship between drug and ADR given the characteristics of the study (e.g., duration and sample size). Machine learning can be used to estimate frequencies from past studies, since we know whether the causal link was present and the values of

the modulator variables. When selecting the set of studies from which to infer these frequencies we face the reference class problem[32]: Which studies should we learn these frequencies from? Do we include all studies of the same/similar drug, similar/same adverse event (reaction), same type of sponsor of study (commercial or institutional)[33], beneficial and/or adverse effects? There does not seem to be an obvious answer here. What is obvious is that considering only studies which are similar to the study under consideration leads to a small set of specific studies (little but specific data) while considering many, some of which less similar, studies leads to a large set of studies (much but unspecific data). Ample data is the tool of choice to decrease statistical noise while specific data helps ensuring that the actual phenomenon of interest is studied. In our world of limited specific data, it is impossible to say how to optimally strike a balance between the value of these tools in general.

- **Learning the conditional probabilities of indicators of causation**. The goal is to estimate the conditional probability of an indicator variable given © or its negation (and its other parent variables, if there are any). The predictive power of the causal indicators may be inferred from past drugs with a suspected ADR, such that (1) we now know whether each of those drugs causes the ADR and (2) which of the indicators they had. Concrete learning applications again face a reference class problem. The set of causal indicators was distilled from Hill's Guidelines and the set of modulators was determined from a study of current medical methodology literature. *E-Synthesis* has always been developed with future possible modifications of these sets in mind. Unsupervised machine learning algorithms may discover further predictors (for instance, the number of authors of published study and/or affiliation of study authors), which could give rise to new indicators and/or evidential modulators.

3.1.2 | Modelling mechanisms

Machine learning could play a fundamental role also in modelling mechanisms within *E-Synthesis*. There is already an abundant literature on its use in pharmacokinetics and pharmacodynamics[34;35] to figure out possible and impossible biochemical mechanisms, bypassing *in vitro* and *in vivo* checks by fast and efficient deployment of *in silico* analyses. Likewise, a better understanding of absorption, distribution, metabolisation mechanisms – which prove critical for dose-response and drug concentration estimation in drug delivery processes – has been highly accelerated by computer simulations[36] and machine learning[37;38]. Some steps towards such a direction have been already taken in[9;10], where – in the latter – dose-response algorithms, usually employed in clinical phase II, have been translated to pharmacovigilance.

3.2 | Information retrieval

Given larger and larger amount of publications available, the need for advanced information retrieval (IR) systems increases. AI may also help here. At present, most of present IR systems, such as general search engines (e.g. Google and Yahoo) and scientific literature search engines (e.g. PubMed and ACM Digital Library) use keywords to query and index documents. However, this traditional keyword-based IR model provides little semantic context for the understanding of user information needs. For example, a keyword usually has several senses and its meaning is ambiguous without context. In addition, one meaning can be expressed by many keywords[39]. There is a long-running research programme of trying to addressing these problems[40;41]. The push towards integration of semantic context according to the user's information need and the user's understanding of documents in the collection into IR systems is one of the main topics of current IR research[39]. On the medical side, knowledge extraction may prove fundamental for accelerating the bench to bed passage in pharmacological research[42]. With respect to *E-Synthesis*, we think that evidence retrieval may boost its performances, by querying databases for all known names for a drug (alike what is done

in databases like VigiBase[8]), for similar drugs (similarity in terms of active ingredient, drug carrier, chemical structure) and similar reactions, for disentangling mechanisms of putative causal connections with respect to different drugs causing the same ADR.[9]

3.3 | AI-powered graphical decision aids

Facing an increasing amount of information puts pressure not only on the way such data must be analyzed[44], but also on the way those data have to be presented for an effective decision making. In fact, researchers with limited information processing capability are usually unable to cope with an exponentially increasing amount of information, leading to a phenomenon called "information overload". This phenomenon has widely been recognized to have adverse effects on decision quality[45]. The use of graphs as decision aids to reduce the adverse effects of information overload on decision quality has been positively investigated both in management[46] and communicating risks between patients and physicians[47]. AI could aid these goals by making it easier to visualise the confirmatory impact of (hypothetical) evidence and the confirmatory impact of indicators. An interactive graphical representation of strengths of associations may lead to better decisions based on *E-Synthesis*.

4 | DISCUSSION

We have shown how AI may contribute to pharmacovigilance by improving a Bayesian framework for evidence synthesis. We think that such applications will also benefit other approaches to evidence synthesis. The prospects for AI supported inference in medicine seem bright, yet AI will not cure all ills.

4.1 | Limitations: AI is not a panacea

AI can reduce some of the limitations of *E-Synthesis*, yet some will remain. For instance, while machine learning can help in making the weighting scheme of evidential modulators as well as the probabilities of the causal indicators more objective, it is still a human who chooses the algorithm for these machine learning operations. There will hence continue to be room for subjective choice and disagreement about these choices. Furthermore, graphical decision aids can make improve the usability and explainability of decision processes, good decision making under uncertainty is a complicated task at which we routinely fail to be optimal[48].

One current limitation of *E-Synthesis* is its concept of causation. Consider the (simplified) case of taking a drug *D* and an adverse drug reaction *A*. Currently, *E-Synthesis* treats causation as categorical and binary: either *D* causes *A* or it does not. This reflects the traditional approach to causation in philosophy[49;50;51;52;53;54]. For some decisions, this might be sufficient, e.g. if we regard a causal relation from *D* to *A* is sufficient for rejecting the use of *D* in medicine, then all we need to determine is the presence or absence of that causal relation. However, policymakers, doctors, patients and scientists are often interested in the question of the *strength* of a causal relation. *E-Synthesis* does not commit us to any particular account of causal strength. There are many options in the literature which may be explored[55;56;57;58;59].

[8] https://www.who-umc.org/vigibase/vigibase/

[9] There are known examples of linking different drugs to the same ADR[43]. Such evidence can help to exonerate a drug under consideration by putting the blame on a different drug causing the ADR. However, such evidence may also incriminate the drug under consideration by elucidating the mechanism between the drug under consideration and the ADR.

4.2 | Future Work

While we can understand causal relations between binary variables by how much (in some sense) the presence of the cause variable causes the probability of the effect variable to increase, there is also a pertinent graded sense of causation between many valued variables: how *strong* an ADR does a particular *dosage* cause? AI holds great promise to squeeze such more fine-grained information from evidence, which will require continued interaction between stakeholders and scientists from numerous areas. We echo the call for an increase of such interactions to improve pharmacovigilance for the good of us all [60;61;9].

references

[1] Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine* 2018; 1(1). doi: 10.1038/s41746-018-0029-1

[2] Paydar S, Pourahmad S, Azad M, et al. The Evolution of a Malignancy Risk Prediction Model for Thyroid Nodules Using the Artificial Neural Network. *Middle East Journal of Cancer* 2016.

[3] Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine* 2013; 11(2): 47–58. doi: 10.2478/v10136-012-0031-x

[4] Mokhtar AM. The Future Hospital: A Business Architecture View. *Malaysian Journal of Medical Sciences* 2017; 24(5): 1–6. doi: 10.21315/mjms2017.24.5.1

[5] Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health* 2019; 1(6): e271–e297. doi: 10.1016/s2589-7500(19)30123-2

[6] Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020: m689. doi: 10.1136/bmj.m689

[7] Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. *The Lancet Digital Health* 2019; 1(1): e13–e14. doi: 10.1016/s2589-7500(19)30002-0

[8] Landes J, Osimani B, Poellinger R. Epistemology of causal inference in pharmacology. *European Journal for Philosophy of Science* 2018; 8(1): 3–49.

[9] Abdin AY, Auker-Howlett D, Landes J, Mulla G, Jacob C, Osimani B. 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* 2019; 25(16): 1866-1880.

[10] De Pretis F, Osimani B. New Insights in Computational Methods for Pharmacovigilance: E-Synthesis, a Bayesian Framework for Causal Assessment. *International Journal of Environmental Research and Public Health* 2019; 16(12). doi: 10.3390/ijerph16122221

[11] De Pretis F, Landes J, Osimani B. E-Synthesis: A Bayesian Framework for Causal Assessment in Pharmacosurveillance. *Frontiers in Pharmacology* 2019; 10: 1317. doi: 10.3389/fphar.2019.01317

[12] De Pretis F, Peden W, Landes J, Osimani B. Pharmacovigilance as Personalized Evidence. In: Springer International Publishing. 2020 (pp. 1–21)

[13] Neapolitan RE. *Learning Bayesian Networks*. Upper Saddle River, NJ: Pearson Prentice Hall . 2004.

[14] Hill AB. The environment and disease: association or causation?. *Journal of the Royal Society of Medicine* 2015; 108(1): 32–37. This article was first published by JRSM in Volume 58 issue 5, May 1965.doi: 10.1177/0141076814562718

[15] Mercuri M, Baigrie B, Upshur RE. Going from evidence to recommendations: Can GRADE get us there?. *Journal of Evaluation in Clinical Practice* 2018; 24(5): 1232-1239.

[16] Darwiche A. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press . 2009

[17] Bogen J, Woodward J. Saving the phenomena. *The Philosophical Review* 1988; 97(3): 303-352.

[18] Brassey J, Price C, Edwards J, Zlabinger M, Bampoulidis A, Hanbury A. Developing a fully automated evidence synthesis tool for identifying, assessing and collating the evidence. *BMJ Evidence-Based Medicine* 2019: bmjebm–2018–111126. doi: 10.1136/bmjebm-2018-111126

[19] O'Connor AM, Tsafnat G, Gilbert SB, et al. Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). *Systematic Reviews* 2019; 8(1). doi: 10.1186/s13643-019-0975-y

[20] Marshall IJ, Johnson BT, Wang Z, Rajasekaran S, Wallace BC. Semi-Automated evidence synthesis in health psychology: current methods and future prospects. *Health Psychology Review* 2020; 14(1): 145–158. doi: 10.1080/17437199.2020.1716198

[21] Michie S, Thomas J, Johnston M, et al. The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science* 2017; 12(1). doi: 10.1186/s13012-017-0641-5

[22] Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* 2016; 113(27): 7345–7352. doi: 10.1073/pnas.1510507113

[23] Bareinboim E, Pearl J. A General Algorithm for Deciding Transportability of Experimental Results. *Journal of Causal Inference* 2013; 1(1). doi: 10.1515/jci-2012-0004

[24] Bareinboim E, Pearl J. Transportability from Multiple Environments with Limited Experiments: Completeness Results. In: NIPS'14. MIT Press; 2014; Cambridge, MA, USA: 280–288.

[25] Russell S, Russell S, Norvig P, Davis E. *Artificial Intelligence: A Modern Approach*. Pearson . 2020.

[26] Ananiadou S, Rea B, Okazaki N, Procter R, Thomas J. Supporting Systematic Reviews Using Text Mining. *Social Science Computer Review* 2009; 27(4): 509–523. doi: 10.1177/0894439309332293

[27] Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. *Systematic Reviews* 2015; 4(1). doi: 10.1186/s13643-015-0066-7

[28] Maas A. Living systematic reviews: A novel approach to create a living evidence base. *Journal of Neurotrauma* 2018. doi: 10.1089/neu.2018.6059

[29] Ciociola AA, Karlstadt RG, Pambianco DJ, Woods KL, Ehrenpreis ED. The Food and Drug Administration Advisory Committees and Panels: How They Are Applied to the Drug Regulatory Process. *American Journal of Gastroenterology* 2014; 109(10): 1508–1512. doi: 10.1038/ajg.2014.85

[30] Ankley GT, Bennett RS, Erickson RJ, et al. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry* 2010; 29(3): 730-741.

[31] Rocca E. The judgements that evidence-based medicine adopts. *Journal of Evaluation in Clinical Practice* 2018; 24(5): 1184–1190. doi: 10.1111/jep.12994

[32] Hájek A. The reference class problem is your problem too. *Synthese* 2007; 156(3): 563–585. doi: 10.1007/s11229-006-9138-5

[33] Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *Journal of Evaluation in Clinical Practice* 2014; 20(6): 908-914.

[34] Poynton M, Choi B, Kim Y, et al. Machine Learning Methods Applied to Pharmacokinetic Modelling of Remifentanil in Healthy Volunteers: A Multi-Method Comparison. *Journal of International Medical Research* 2009; 37(6): 1680–1691. doi: 10.1177/147323000903700603

[35] Bunte K, Smith DJ, Chappell MJ, et al. Learning pharmacokinetic models for in vivo glucocorticoid activation. *Journal of Theoretical Biology* 2018; 455: 222–231. doi: 10.1016/j.jtbi.2018.07.025

[36] Bretz F, Pinheiro JC, Branson M. Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies. *Biometrics* 2005; 61(3): 738–748. doi: 10.1111/j.1541-0420.2005.00344.x

[37] Tang J, Liu R, Zhang YL, et al. Application of Machine-Learning Models to Predict Tacrolimus Stable Dose in Renal Transplant Recipients. *Scientific Reports* 2017; 7(1). doi: 10.1038/srep42192

[38] You W, Widmer N, Micheli GD. Personalized modeling for drug concentration prediction using Support Vector Machine. In: IEEE; 2011: 1523-1527

[39] Huang X. Machine learning approaches to Information Retrieval and its applications to the web, medical informatics and health care. In: IEEE; 2008: 1-2

[40] Jones KS. Information retrieval and artificial intelligence. *Artificial Intelligence* 1999; 114(1-2): 257–281. doi: 10.1016/s0004-3702(99)00075-2

[41] Boughanem M, Akermi I, Pasi G, Abdulahhad K. Information Retrieval and Artificial Intelligence. In: Springer International Publishing. 2020 (pp. 147–180)

[42] Frey LJ, Talbert DA. Artificial Intelligence Pipeline to Bridge the Gap between Bench Researchers and Clinical Researchers in Precision Medicine. *Med One* 2020; 5(1): 1-18. doi: 10.20900/mo20200001

[43] Singer JB, Lewitzky S, Leroy E, et al. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nature Genetics* 2010; 42(8): 711-714.

[44] De Pretis F. New mathematical perspectives to understand the Information Society: the statistical mechanics approach to model and analyze big-data. In: Mütercim Publishing House; 2016: 3-10.

[45] Chan SY. The use of graphs as decision aids in relation to information overload and managerial decision quality. *Journal of Information Science* 2001; 27(6): 417–425. doi: 10.1177/016555150102700607

[46] Benbasat I, Dexter AS. An Experimental Evaluation of Graphical and Color-Enhanced Information Presentation. *Management Science* 1985; 31(11): 1348–1364. doi: 10.1287/mnsc.31.11.1348

[47] Franklin L, Plaisant C, Shneiderman B. An Information-Centric Framework for Designing Patient-Centered Medical Decision Aids and Risk Communication. In: AMIA; 2013; Washington, DC, USA: 456-65.

[48] Crupi V, Elia F. Understanding and improving decisions in clinical medicine (I): Reasoning, heuristics, and error. *Internal and Emergency Medicine* 2017; 12(5): 689-691.

[49] Ross W. *Aristotle's Physics, A Revised Text with Introduction and Commentary*. Clarendon Press, Oxford. Reprinted from the first edition of . 1936.

[50] Hume D. *A Treatise on Human Nature*. Clarendon Press . 1739.

[51] Mill JS. *A System of Ratiocinative and Inductive Logic: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London, Longman . 1970.

[52] Suppes P. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland . 1970.

[53] Woodward J. *Making Things Happen*. Oxford University Press . 2004

[54] Cartwright N. Causal Powers: What Are They? Why Do We Need Them? What Can be Done with Them and What Cannot?. Tech. Rep. 04/07, 2007.

[55] Good IJ. A Causal Calculus (I). *The British Journal for the Philosophy of Science* 1961; XI(44): 305–318. doi: 10.1093/bjps/xi.44.305

[56] Good IJ. A Causal Calculus (II). *The British Journal for the Philosophy of Science* 1961; XII(45): 43–51. doi: 10.1093/bjps/xii.45.43

[57] Eells E. *Probabilistic Causality*. Cambridge University Press . 1991

[58] Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge University Press. 2 ed. 2009

[59] Sprenger J. Foundations of a Probabilistic Theory of Causal Strength. *The Philosophical Review* 2018; 127(3): 371–398. doi: 10.1215/00318108-6718797

[60] Rocca E. Bridging the boundaries between scientists and clinicians – mechanistic hypotheses and patient stories in risk assessment of drugs. *Journal of Evaluation in Clinical Practice* 2017; 23(1): 114-120.

[61] Rocca E, Copeland S, Edwards IR. Pharmacovigilance as Scientific Discovery: An Argument for Trans-Disciplinarity. *Drug Safety* 2019; 42(10): 1115–1124. doi: 10.1007/s40264-019-00826-1

ACKNOWLEDGEMENTS

Francesco De Pretis 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICAL APPROVAL

All authors have equally contributed to the writing of this manuscript.

ORCID

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

FIGURES

| Figure 1

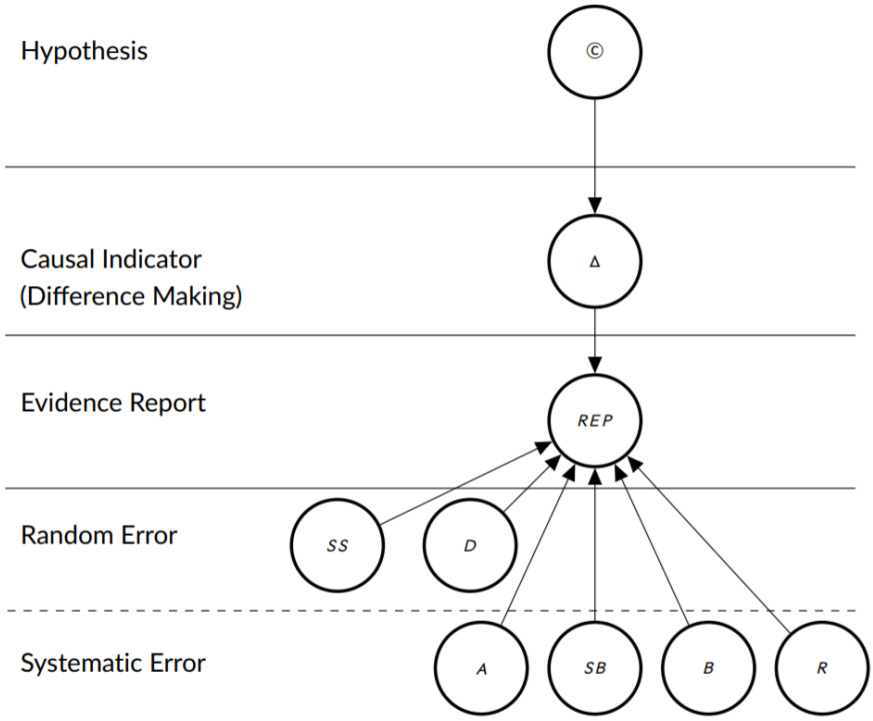

FIGURE LEGENDS

| Legend of Figure 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.[11].