

Artificial intelligence methods for a Bayesian epistemology-powered evidence evaluation

Francesco De Pretis PhD^{1,2}  | Jürgen Landes PhD³  | William Peden PhD^{4,5} 

¹Department of Biomedical Sciences and Public Health, School of Medicine and Surgery, Marche Polytechnic University, Ancona, Italy

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

³Munich Center for Mathematical Philosophy, Faculty of Philosophy, Philosophy of Science and Study of Religion, Ludwig-Maximilians-Universität München, Munich, Germany

⁴Erasmus Institute for Philosophy and Economics, Erasmus School of Philosophy, Erasmus University Rotterdam, Rotterdam, The Netherlands

⁵Department of Philosophy, Durham University, Durham, UK

Correspondence

William Peden, PhD, Department of Philosophy, Durham University, 50 Old Elvet, DH1 3HN Durham, UK.
Email: w.j.peden@durham.ac.uk

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 405961989, 432308570; Horizon 2020 Framework Programme, Grant/Award Number: 639276

Abstract

Rationale, aims and objectives: The diversity of types of evidence (eg, case reports, animal studies and observational studies) makes the assessment of a drug's safety profile into a formidable challenge. While frequentist uncertain inference struggles 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.

Methods: *E-Synthesis* is a Bayesian framework for drug safety assessments built 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. AI systems are being developed for evidence aggregation in medicine, which increasingly are automated.

Results: We find that AI can help *E-Synthesis* 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*. Vice versa, *E-Synthesis* offers a solid methodological basis for (semi-)automated evidence aggregation with AI systems.

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 practice.

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* 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.[†] AI can also make a contribution to diagnostic procedures by doctors² (see Amato et al³ for a general

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Evaluation in Clinical Practice* published by John Wiley & Sons Ltd

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. 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, as well as AI methodology for evidence aggregation in medicine. In the Results section, we show how *E-Synthesis* and AI can be intertwined to the benefit of both. 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: E-SYNTHESIS AND AI

The synthesis of evidence from multiple sources 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 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 (RCT). 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 aggregation. 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 aggregation in pharmacosurveillance to support timely decision making based on all the available ‘safety signals’.⁸⁻¹² 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 of evidence qualities,¹¹ investigated the aggregation of knowledge concerning biological mechanisms and dose-response,^{9,10} and made strides towards applying *E-*

Synthesis in personalized medicine.¹² In this subsection, we give a brief overview of *E-Synthesis*.

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 synthesis, 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 heterogeneous 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 network¹³ incorporating indicators of causality derived from the Bradford-Hill guidelines¹⁴ 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* utilizes 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 formalizes degrees of beliefs as probabilities; it thereby inherits the formal constraints of the probability calculus. Thus, $P(H)$ represents a researcher’s degree of belief in a hypothesis, while $P(H|\mathcal{E})$ 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:

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

where the hypotheses H_i and $\odot = H_1$ constitute a mutually inconsistent and exhaustive partition.[‡]

With this mathematical formula, the posterior probability of the hypothesis given the evidence, $P(\odot|\mathcal{E})$, only depends on prior probabilities $P(H_i)$, and likelihoods $P(\mathcal{E}|H_i)$.[§] Bayesian epistemology focuses on updating (or “conditionalizing”) for propositions or events in general, whereas Bayesian statistics focuses on testing statistical models using conditional probabilities.

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.^{13,16} 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 (which means that the edges are directed such that the graph does not contain a directed cycle, that is, it has no path of directed edges which leads back to its starting point). Secondly, a probability distribution specifying the conditional probabilities of all variables given their parent variables (all other variables which directly point to this variable). See Figure 1 for an example graph.

Technically, this works as follows. Denoting the parents of a variable Y by X_1, \dots, X_n one specifies $P(Y = y | X_1 = x_1, \dots, X_n = x_n) \in [0, 1]$ for all possible values y, x_1, \dots, x_n under the condition that $\sum_{y \in \mathcal{Y}} P(Y = y | X_1 = x_1, \dots, X_n = x_n) = 1$. 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’.

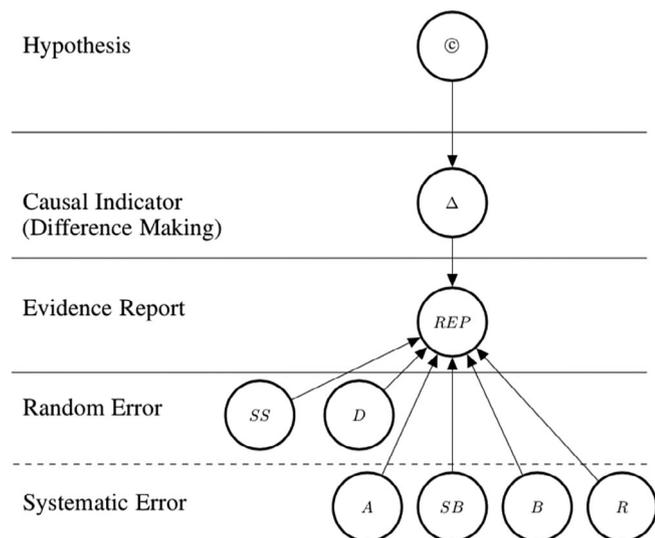


FIGURE 1 Graph structure of the Bayesian network for one randomized controlled trial (RCT) which informs us about difference making (Δ) which in turn informs us about the causal hypothesis. The information provided by the reported study is modulated by how well the particular RCT guards against random 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; and R , Randomization¹¹

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: (a) difference making, (b) probabilistic dependence, (c) dose-response relationship, (d) rate of growth, (e) temporal precedence, and (f) 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 a drug and an adverse effect, such that higher dosages lead to a more and/or stronger adverse effect. However, note that a causal relationship might lack a dose-response relationship (anaphylaxis) and a dose-response relationship might exist without a causal relationship, due to confounding. The indicators are probabilistic consequences in the sense that their truth is more likely, if the hypothesis is also true, than if the latter is false, that is, $P(\text{Ind}|\odot) > P(\text{Ind}) > P(\text{Ind}|\bar{\odot})$. In turn, $P(\odot|\text{Ind}) > P(\odot) > P(\odot|\bar{\text{Ind}})$.

Therefore, 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.^{8,11,17} *E-Synthesis* thus analyses the inferential process from the raw data to the hypothesis that a causal link holds between a drug and an ADR into two steps: (a) from data (study reports) to causal indicators and (b) 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, randomization, sample size, study duration, stratification), see Figure 1.

2.1.4 | Evidential modulators

One key feature of *E-Synthesis* is the possibility of assessing 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, Rep , for every item of evidence and then creating for every such variable a set of pertinent modulator variables Q_1, \dots, Q_k , for example, 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, $P(Rep = rep | Ind = ind, Q_1 = q_1, \dots, Q_k = q_k)$.[¶]

An application of Bayes' Theorem enables one then to calculate the posterior probability of causal indicators. In turn, this posterior probability 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 and evidence aggregation in medicine

As outlined in the Introduction, AI already has a growing impact on healthcare. However, its potential for evidence synthesis is still underdeveloped.^{**} This is despite cautious interest within parts of the healthcare industry.^{††} Greater use of AI in evidence synthesis could have many benefits, which we detail in Section 4.2.

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 are 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

The semi-automation research program has already produced some results. 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, specific 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 labelled, 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

3 | RESULTS

In this section, we explore in detail what may come out from the interactions of *E-Synthesis* and AI. We investigate both directions, that is, what *E-Synthesis* can provide for a better working of AI and how AI itself can improve *E-Synthesis*.

3.1 | E-synthesis for AI

As outlined in the previous section, *E-Synthesis* offers a methodologically sound approach for evidence aggregation tasks in general. The methodological choice of a Bayesian network lends itself to further applications in AI, since Bayesian network algorithms are designed to be easily implemented within AI systems. Moreover, we deem that *E-Synthesis* can contribute to strengthen AI use in digital health applications at least in two ways:

1. It is a formal evidence synthesis procedure. Hence, the procedure should ultimately be amenable to semi-automation.
2. Its Bayesian basis aids transparency. The Bayesian methodology requires that we define the prior probabilities of possible events and their interrelations within our model, as a precondition of making inferences using *E-Synthesis*. The elements of this process are standard Bayesian tools, adapted to the particular case of pharmacological evidence synthesis. Therefore, prior to applying *E-Synthesis*, we must articulate our assumptions in a way that those familiar with Bayesian modelling can understand them.

We shall expand on the second point. A significant concern in contemporary AI design is transparency, especially for AI involved in decisions that affects people's lives. Where possible, it is ethical that decision-making processes are understandable for the people affected by them, so that these people can enter into the relevant deliberations, articulate their own viewpoints in an informed manner, and otherwise hold AI designers to account. For instance, if an AI algorithm has features that systematically bias decisions against a particular race or gender, we want this bias to be open to challenges by the groups who are negatively affected or experts working on their behalf. The extent to which AI is understandable for users and stakeholders will vary among contexts, but even when users lack the expertise required to understand some AI's reasoning, the comprehensibility of that reasoning is often possible for experts who are accountable to those users. Transparency of an AI's decision-making process widens the scope of users and stakeholders who can understand (directly or indirectly) the system and the depth of their understanding.

However, some types of AI have limited transparency even for those with relevant expertise. For example, recently, concerns have been raised with respect to medical decision algorithms.²⁹⁻³² Many early applications of computer reasoning in AI used relatively simple if-then reasoning procedures, where the link between the inputs and the decisions was clear.^{33,p. 3032} Yet machine learning functions that use neural networks are distributed over all the neurons, with no unique functional form. The neural network approach offers great gains in the accuracy of inferences made using the AIs, but at the cost of relatively low transparency.

By contrast, the decisions about drug safety that are made by *E-Synthesis* will ultimately be formalizable in algorithms. It is true that there could be some exogenous elements. One example is that, at the input level, the selection of data for the evidential modulators could be decided by non-transparent neural network machine learning. At



the output level, we are not proposing the complete automation of drug safety decisions, but instead just semi-automation, and therefore there will still be human judgements that could be opaque, depending on how the regulators make their choices.

However, *E-Synthesis* shares a common Bayesian advantages that it forces us to make our probabilistic assumptions explicit, and thus open to criticism.^{34, Chapter 11} Therefore, in comparison to some types of AI, using *E-Synthesis* would improve transparency. Note that this superior transparency holds even if we think that the priors are ultimately ‘subjective’ in an epistemological sense: users can still raise challenges on criteria such as alignment of the prior probabilities with well-tested physical probabilities, the liability of priors to help us avoid catastrophic choices,³⁵ and other desiderata that users might have for priors.

For the capacity of *E-Synthesis* to improve pharmacological predictions, we can point to some promising precedents in which AI has been used to improve predictive power.^{36,37} AI is especially promising for orphan drugs³⁸ where the quantity and quality of data cannot compare with largely used medications. We think that *E-Synthesis* may contribute in improving these AI methods with a more sophisticated evidence aggregation and evaluation, favouring a better understanding of causal underpinnings in drug safety management.

3.2 | AI for evidence synthesis

As we have seen, AI methods are already employed in the realm of evidence aggregation 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.³⁹ 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, we pin down 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, (cf. Section 4.2).

3.2.1 | Machine learning

Machine learning can greatly strengthen *E-Synthesis*, 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. 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*: (a) estimation of conditional probabilities of causal indicators and learning the weighting schemes of the evidential modulators from data and (b) modelling the ‘linkage between a direct molecular initiating event [...] and an adverse outcome at a biological level of organization relevant to risk assessment’.^{40, p. 731} The latter occurs through an adverse outcome pathway (AOP), that is, a conceptual construct—expressed in terms of flow-charts—that portrays existing knowledge concerning the linkage between that initiating event at a molecular level and the adverse outcome that can be macroscopically observed. Such ‘mechanisms’ play an important inferential role.⁴¹

3.2.2 | 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 determining how likely it is that a study (observational or an RCT) correctly identifies the absence or presence of a causal relationship between a drug and an ADR given the characteristics of the study, for example, 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.

Note that, while machine learning can help us to obtain values for the evidential modulators, we still face ‘The Problem of the Reference Class’: the challenge of selecting the set of studies from which to infer these frequencies.⁴² 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),⁴³ beneficial and/or adverse effects? There does not seem to be an obvious answer. 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. However, a Bayesian framework like *E-Synthesis* helps us make our answers to the methodological questions (in the form of our Bayesian probabilities for particular events) more rigorously formulated and open to scrutiny than if choice among reference classes is left implicit.

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, which could give rise to new indicators and/or evidential modulators. Possible new predictors might include the number of authors of published study and/or affiliation of the study's authors.

3.2.3 | 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^{44,45} 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, metabolization mechanisms—which prove critical for dose-response and drug concentration estimation in drug delivery processes—has been highly accelerated by computer simulations⁴⁶ and machine learning.^{47,48} Some steps towards such a direction have been already taken in Abdin et al⁹ and De Pretis and Osimani,¹⁰ where—in the latter—dose-response algorithms, usually employed in clinical phase II, have been translated to pharmacovigilance.

3.2.4 | 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 IR systems, such as general search engines (eg, Google and Yahoo) and scientific literature search engines (eg, 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.⁴⁹ There is a long-running research program of trying to addressing these problems.^{50,51} 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.⁴⁹ On the medical side, knowledge extraction may prove fundamental for accelerating the bench to bedside passage in pharmacological research.⁵² With respect to *E-Synthesis*, evidence retrieval may boost its performances, by querying databases for all known names for a drug (alike what is done in databases like VigiBase^{‡‡}), for similar drugs (similarity in terms of active ingredient, drug carrier, chemical structure) and similar reactions, as well as disentangling mechanisms of putative causal connections with respect to different drugs causing the same ADR.^{§§}

3.2.5 | AI-powered graphical decision aids

Facing an increasing amount of information puts pressure not only on the way such data must be analysed,⁵⁴ 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.⁵⁵ The use of graphs as decision aids to reduce the adverse effects of information overload on decision quality has been positively investigated both in management⁵⁶ and communicating risks between patients and physicians.⁵⁷ AI could aid these goals by making it easier to visualize 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 we stress that 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, while graphical decision aids can 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.⁵⁸

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.⁵⁹⁻⁶⁴ For some decisions, binary causation might be sufficient: for example, if we regard a causal relation from *D* to *A* as 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 that might be explored.⁶⁵⁻⁶⁹



4.2 | AI and human judgements

We have assumed that that AI can improve human decision making. While we do not think that AI always improves our decision making, there are good reasons and evidence that, in medicine, AI is already improving decision qualities and that AI support will lead to even better decisions in the future.

Firstly, AI can perform tasks on large data sets we are simply not able to do, for example, searching, summarizing and revising probability distributions. AI thus expands the computational capacity of evidence evaluators. The accelerating increase in medical data means that the application of AI in evidence synthesis is increasingly difficult. Insofar as evidence syntheses depend on a lot of human input, it will be hard to keep track of the ever-greater flow of evidence such as case reports and clinical trials. Automation via AI can help alleviate some of these information processing strains in the evidence synthesis process.

Secondly, AI can make the decision-making procedure more transparent. Such systems can offer graphical decision aids which can be used by evidence evaluators when explaining their decisions to patients, policymakers, and other stakeholders. Additionally, all outputs of AI systems depend ultimately in a formal and (in principle) traceable manner on the input. In some cases, AI reasoning can be summarized in terms of algorithms that are accessible for many users and groups affected by the reasoning. Human decision-making procedures are by contrast most often not open to inspection.

It is true that the superior transparency of AI is not guaranteed. We noted above that machine learning systems are often incomprehensible, in some sense, even for experts. Yet, even in these cases, it is not clear that AI is any *less* transparent than human reasoning, since the latter might involve intuitive judgements that are also impossible to articulate formally.^{70,p. 7} Furthermore, while a neural network's learning algorithm might have no explicit representation, the network's overall dynamics can be articulated and scrutinized—something far beyond what we can currently do with the human mind.

Thirdly, AI offers us the possibility to better understand our judgements by performing hypothetical analyses of how different judgements influence decision-making procedures. Let us recall the Reference Class Problem (Section 3.2.1). Applying AI systems to different reference classes allows us to perform sensitivity analyses, thereby shedding light on how our judgements of relevance influence decision making. Depending on our answer to the Reference Class Problem, such analysis might even help in finding an appropriate reference class.

4.3 | 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.^{9,71,72}

ACKNOWLEDGEMENTS

Francesco De Pretis and William Peden acknowledge funding from the European Research Council (PhilPharm—GA n. 639276) through the Marche Polytechnic University (Ancona, Italy). The Authors are grateful to Durham University for providing open access for this article. 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.

AUTHOR CONTRIBUTIONS

All authors contributed to the paper's conception, drafting, and revisions. All the authors gave their approval to the final version and take accountability for the research.

ENDNOTES

* <https://www.nhs24.scot>.

† <https://www.exscientia.ai/news-insights/sumitomo-dainippon-pharma-and-exscientia-joint-development>.

‡ For convenience, we use the same symbol denoting a variable and the variable being true.

§ In basic statistical applications of Bayesianism, 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, for example, calculating the probability of a particular series of independent and identically distributed binomial trials given the hypothesis of a population frequency. However, in more complex applications, determining the likelihoods can be very difficult, as we discuss later.

¶ Uncertainty about study qualities is represented by probabilities in the fashion usual in Bayesian statistics, for example, $P(Q_i = q_j)$.

** The first automated evidence synthesis system was only published in 2019^[18]. See O'Connor et al^[19] for a recent overview of evidence synthesis automation.

†† <https://blog.evidencepartners.com/past-present-and-future-automation-in-systematic-review-software>.

‡‡ <https://www.who-umc.org/vigibase/vigibase/>.

§§ There are known examples of linking different drugs to the same ADR^[53]. 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.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Francesco De Pretis  <https://orcid.org/0000-0001-8395-7833>

Jürgen Landes  <https://orcid.org/0000-0003-3105-6624>

William Peden  <https://orcid.org/0000-0002-3474-7861>

REFERENCES

- Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *npj Digit Med*. 2018;1(1):1–10. <https://doi.org/10.1038/s41746-018-0029-1>.
- 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 J Cancer*. 2016;7(1):47–52.
- Amato F, López A, Peña-Méndez EM, Vañhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed*. 2013;11(2):47–58. <https://doi.org/10.2478/v10136-012-0031-x>.
- Mokhtar AM. The future hospital: a business architecture view. *Malays J Med Sci*. 2017;24(5):1–6. <https://doi.org/10.21315/mjms2017.24.5.1>.
- 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. *Lancet Digital Health*. 2019;1(6):e271–e297. [https://doi.org/10.1016/s2589-7500\(19\)30123-2](https://doi.org/10.1016/s2589-7500(19)30123-2).
- 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;368:m689. <https://doi.org/10.1136/bmj.m689>.
- Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. *Lancet Digital Health*. 2019;1(1):e13–e14. [https://doi.org/10.1016/s2589-7500\(19\)30002-0](https://doi.org/10.1016/s2589-7500(19)30002-0).
- Landes J, Osimani B, Poellinger R. Epistemology of causal inference in pharmacology. *Eur J Philos Sci*. 2018;8(1):3–49. <https://doi.org/10.1007/s13194-017-0169-1>.
- 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). *Curr Pharm Des*. 2019;25(16):1866–1880. <https://doi.org/10.2174/1381612825666190628160603>.
- De Pretis F, Osimani B. New insights in computational methods for pharmacovigilance: E-synthesis, a Bayesian framework for causal assessment. *Int J Environ Res Public Health*. 2019;16(12):1–19. <https://doi.org/10.3390/ijerph16122221>.
- De Pretis F, Landes J, Osimani B. E-synthesis: a Bayesian framework for causal assessment in pharmacosurveillance. *Front Pharmacol*. 2019;10:1–20. <https://doi.org/10.3389/fphar.2019.01317>.
- De Pretis F, Peden W, Landes J, Osimani B. Pharmacovigilance as personalized evidence. In: Beneduce C, Bertolaso M, eds. *Personalized Medicine in the Making. Philosophical Perspectives from Biology to Healthcare*. Cham, Switzerland: Springer; 2021:19 forthcoming.
- Neapolitan RE. *Learning Bayesian Networks. Prentice Hall Series in Artificial Intelligence*. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.
- Hill AB. The environment and disease: association or causation? *J R Soc Med*. 2015;108(1):32–37. This article was first published by JRSM in Volume 58 issue 5, May 1965. <https://doi.org/10.1177/0141076814562718>.
- Mercuri M, Baigrie B, Upshur RE. Going from evidence to recommendations: can GRADE get us there? *J Eval Clin Pract*. 2018;24(5):1232–1239. <https://doi.org/10.1111/jep.12857>.
- Darwiche A. *Modeling and Reasoning with Bayesian Networks*. Cambridge, UK: Cambridge University Press; 2009.
- Bogen J, Woodward J. Saving the phenomena. *Philos Rev*. 1988;97(3):303–352. <https://doi.org/10.2307/2185445>.
- 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 Evid-Based Med*. 2019;26:24–27. <https://doi.org/10.1136/bmjebm-2018-111126>.
- 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). *Syst Rev*. 2019;8(1):1–5. <https://doi.org/10.1186/s13643-019-0975-y>.
- Marshall IJ, Johnson BT, Wang Z, Rajasekaran S, Wallace BC. Semi-automated evidence synthesis in health psychology: current methods and future prospects. *Health Psychol Rev*. 2020;14(1):145–158. <https://doi.org/10.1080/17437199.2020.1716198>.
- 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. *Implement Sci*. 2017;12(1):121. <https://doi.org/10.1186/s13012-017-0641-5>.
- Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci*. 2016;113(27):7345–7352. <https://doi.org/10.1073/pnas.1510507113>.
- Bareinboim E, Pearl JA. General algorithm for deciding transportability of experimental results. *J Causal Inference*. 2013;1(1):107–134. <https://doi.org/10.1515/jci-2012-0004>.
- Bareinboim E, Pearl J. Transportability from multiple environments with limited experiments: completeness results. *NIPS'14. Proceedings of the 27th International Conference on Neural Information Processing Systems—Vol. 1*. Cambridge, MA: MIT Press; 2014:280–288.
- Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. Harlow, UK: Pearson; 2020.
- Ananiadou S, Rea B, Okazaki N, Procter R, Thomas J. Supporting systematic reviews using text mining. *Soc Sci Comput Rev*. 2009;27(4):509–523. <https://doi.org/10.1177/0894439309332293>.
- Jonnalagadda SR, Goyal P, Huuman MD. Automating data extraction in systematic reviews: a systematic review. *Syst Rev*. 2015;4(1):78. <https://doi.org/10.1186/s13643-015-0066-7>.
- Maas A. Living systematic reviews: a novel approach to create a living evidence base. *J Neurotrauma*. 2018;1–4. <https://doi.org/10.1089/neu.2018.6059>.
- Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. *Philos Technol*. 2020;1–23. <https://doi.org/10.1007/s13347-019-00391-6>.
- Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics*. 2019;46(3):205–211. <https://doi.org/10.1136/medethics-2019-105586>.
- Milano S, Taddeo M, Floridi L. Recommender systems and their ethical challenges. *AI Soc*. 2020;35(4):957–967. <https://doi.org/10.1007/s00146-020-00950-y>.
- McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. *J Med Ethics*. 2018;45(3):156–160. <https://doi.org/10.1136/medethics-2018-105118>.
- Wagholikar KB, Sundararajan V, Deshpande AW. Modeling paradigms for medical diagnostic decision support: a survey and future directions. *J Med Syst*. 2011;36(5):3029–3049. <https://doi.org/10.1007/s10916-011-9780-4>.
- Sprenger J, Hartmann S. *Bayesian Philosophy of Science*. Oxford: Oxford University Press; 2019.
- Williamson J. Motivating objective Bayesianism: from empirical constraints to objective probabilities. *Probability and Inference: Essays in Honor of Henry E. Kyburg Jr*. London: College Publications; 2007:155–183.



36. Yom-Tov E. Predicting drug recalls from internet search engine queries. *IEEE J Transl Eng Health Med.* 2017;5:1-6. <https://doi.org/10.1109/jtehm.2017.2732945>.
37. Galeano D, Li S, Gerstein M, Pacanaro A. Predicting the frequencies of drug side effects. *Nat Commun.* 2020;11(1):4575. <https://doi.org/10.1038/s41467-020-18305-y>.
38. Price J. What can big data offer the pharmacovigilance of orphan drugs? *Clin Ther.* 2016;38(12):2533-2545. <https://doi.org/10.1016/j.clinthera.2016.11.009>.
39. 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. *Am J Gastroenterol.* 2014;109(10):1508-1512. <https://doi.org/10.1038/ajg.2014.85>.
40. Ankley GT, Bennett RS, Erickson RJ, et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem.* 2010;29(3):730-741. <https://doi.org/10.1002/etc.34>.
41. Rocca E. The judgements that evidence-based medicine adopts. *J Eval Clin Pract.* 2018;24(5):1184-1190. <https://doi.org/10.1111/jep.12994>.
42. Hájek A. The reference class problem is your problem too. *Synthese.* 2007;156(3):563-585. <https://doi.org/10.1007/s11229-006-9138-5>.
43. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract.* 2014;20(6):908-914. <https://doi.org/10.1111/jep.12147>.
44. Poynton M, Choi B, Kim Y, et al. Machine learning methods applied to pharmacokinetic modelling of Remifentanyl in healthy volunteers: a multi-method comparison. *J Int Med Res.* 2009;37(6):1680-1691. <https://doi.org/10.1177/147323000903700603>.
45. Bunte K, Smith DJ, Chappell MJ, et al. Learning pharmacokinetic models for in vivo glucocorticoid activation. *J Theor Biol.* 2018;455:222-231. <https://doi.org/10.1016/j.jtbi.2018.07.025>.
46. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics.* 2005;61(3):738-748. <https://doi.org/10.1111/j.1541-0420.2005.00344.x>.
47. Tang J, Liu R, Zhang YL, et al. Application of machine-learning models to predict Tacrolimus stable dose in renal transplant recipients. *Sci Rep.* 2017;7(1):1-8. <https://doi.org/10.1038/srep42192>.
48. You W, Widmer N, Micheli GD. Personalized modeling for drug concentration prediction using Support Vector Machine. *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*. Piscataway, NJ: IEEE; 2011:1523-1527.
49. Huang X. Machine learning approaches to Information Retrieval and its applications to the web, medical informatics and health care. *2008 IEEE International Conference on Granular Computing*. Piscataway, NJ: IEEE; 2008:1-2.
50. Jones KS. Information retrieval and artificial intelligence. *Artif Intell.* 1999;114(1-2):257-281. [https://doi.org/10.1016/s0004-3702\(99\)00075-2](https://doi.org/10.1016/s0004-3702(99)00075-2).
51. Boughanem M, Akermi I, Pasi G, Abdulahhad K. Information retrieval and artificial intelligence. In: Marquis P, Papini O, Prade H, eds. *A Guided Tour of Artificial Intelligence Research*. Cham, Switzerland: Springer International Publishing; 2020:147-180.
52. 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. <https://doi.org/10.20900/mo20200001>.
53. Singer JB, Lewitzky S, Leroy E, et al. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat Genet.* 2010;42(8):711-714. <https://doi.org/10.1038/ng.632>.
54. De Pretis F. New mathematical perspectives to understand the Information society: the statistical mechanics approach to model and analyze big-data. *Proceedings of the international conference of young scientists and specialists "Information society as contemporary system of defense and attack" (Baku, Azerbaijan, Nov 27-28 2014)*. Vol 2016. Baku, Azerbaijan: Mütercim Publishing House; 2016:3-10.
55. Chan SY. The use of graphs as decision aids in relation to information overload and managerial decision quality. *J Inf Sci.* 2001;27(6):417-425. <https://doi.org/10.1177/016555150102700607>.
56. Benbasat I, Dexter AS. An experimental evaluation of graphical and color-enhanced Information presentation. *Manag Sci.* 1985;31(11):1348-1364. <https://doi.org/10.1287/mnsc.31.11.1348>.
57. Franklin L, Plaisant C, Shneiderman B. An Information-centric framework for designing patient-centered medical decision aids and risk communication. *AMIA 2013 Annual Symposium Proceedings*. Washington, DC: AMIA; 2013:456-465.
58. Crupi V, Elia F. Understanding and improving decisions in clinical medicine (I): reasoning, heuristics, and error. *Intern Emerg Med.* 2017;12(5):689-691. <https://doi.org/10.1007/s11739-017-1665-1>.
59. Ross W. *Aristotle's Physics, A Revised Text with Introduction and Commentary*. Oxford, UK: The Clarendon Press; 1936.
60. Hume D. *A Treatise on Human Nature*. Oxford, UK: The Clarendon Press; 1978.
61. Mill JSA. *System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London, UK: Longman; 1970.
62. Suppes P. *A Probabilistic Theory of Causality*. Amsterdam, The Netherlands: North-Holland Publishing Company; 1970.
63. Woodward J. *Making Things Happen*. Oxford, UK: Oxford University Press; 2004.
64. 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. London, UK: Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science; 2007.
65. Good IJ. A causal Calculus (I). *Br J Philos Sci.* 1961;XI(44):305-318. <https://doi.org/10.1093/bjps/xi.44.305>.
66. Good IJ. A causal Calculus (II). *Br J Philos Sci.* 1961;XII(45):43-51. <https://doi.org/10.1093/bjps/xii.45.43>.
67. Eells E. *Probabilistic Causality*. Cambridge, UK: Cambridge University Press; 1991.
68. Pearl J. *Causality: Models, Reasoning and Inference*. 2nd ed. Cambridge, UK: Cambridge University Press; 2009.
69. Sprenger J. Foundations of a probabilistic theory of causal strength. *Philos Rev.* 2018;127(3):371-398. <https://doi.org/10.1215/00318108-6718797>.
70. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. *Big Data Soc.* 2016;3(2):1-21. <https://doi.org/10.1177/2053951716679679>.
71. Rocca E. Bridging the boundaries between scientists and clinicians-mechanistic hypotheses and patient stories in risk assessment of drugs. *J Eval Clin Pract.* 2016;23(1):114-120. <https://doi.org/10.1111/jep.12622>.
72. Rocca E, Copeland S, Edwards IR. Pharmacovigilance as scientific discovery: an argument for trans-Disciplinarity. *Drug Saf.* 2019;42(10):1115-1124. <https://doi.org/10.1007/s40264-019-00826-1>.

How to cite this article: De Pretis F, Landes J, Peden W. Artificial intelligence methods for a Bayesian epistemology-powered evidence evaluation. *J Eval Clin Pract.* 2021;1-9. <https://doi.org/10.1111/jep.13542>