Article

New Insights in Computational Methods for Pharmacovigilance: *E-Synthesis*, a Bayesian Framework for Causal Assessment

Francesco De Pretis [1,2,*,†] and **Barbara Osimani** [1,3,‡]

1   Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60126 Ancona, Italy;
     b.osimani@univpm.it
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
*   Correspondence: f.depretis@univpm.it
†   Current address: Via Tronto 10/A, 60126 Ancona, Italy.
‡   These authors contributed equally to this work.

check for updates

Abstract: Today's surge of big data coming from multiple sources is raising the stakes that pharmacovigilance has to win, making evidence synthesis a more and more robust approach in the field. In this scenario, many scholars believe that new computational methods derived from data mining will effectively enhance the detection of early warning signals for adverse drug reactions, solving the gauntlets that post-marketing surveillance requires. This article highlights the need for a philosophical approach in order to fully realize a pharmacovigilance 2.0 revolution. A state of the art on evidence synthesis is presented, followed by the illustration of *E-Synthesis*, a Bayesian framework for causal assessment. Computational results regarding dose-response evidence are shown at the end of this article.

Keywords: advisory committees; Bayesian epistemology; data fusion; data mining; dose-responsiveness; drug approval process; *E-Synthesis*; evidence synthesis; pharmacovigilance

1. Introduction

The problem of collecting, analyzing and evaluating evidence on pharmaceutical safety is a central problem of health-care practice. Statistics on drug-induced hospitalizations range between 5% and 10% of total hospitalizations in Europe as well as in all Western countries [1–6]. According to the European Medicines Agency some 197,000 people in the European Union die each year as a result of adverse drug reactions (ADRs) [7]. Drug-induced toxicities in the US, which rank among the top 10 causes of death, result in health care costs of over US$30 billion annually [8,9]. The authors of [10] estimated approximately £500 million in costs per annum for the UK as a result of adverse drug events. The annual direct costs for Germany have been estimated to amount to €400 million [11]. ADRs constitute a concern for the industry as well in that the attrition rate (the proportion of would-be drugs whose development is interrupted before reaching the market over the total research and development portfolio) is continuously increasing [12]. This concern is also one of the reasons for the development of the Innovative Medicines Initiative (IMI and IMI2) within the EU 7th and H2020 Framework Programmes [13,14]. ADRs are thus responsible for a heavy economic and social burden [15]; in addition, they constitute an extremely vulnerable point for the health system and a key ethical problem for decisions concerning pharmaceutical products.

In this landscape, evidence synthesis is now rising as a robust approach, especially in the field of post-marketing drug safety. This trend stems mainly from a public interest towards the integration of information coming from different sources when evaluating safety signals (European Parliament and the European Council: Directive 2010/84/EU; Regulation (EU) No 1235/2010; see also the 21st Century Cures Act, recently enacted in the US). An analogous wish has also been expressed by researchers in the field [16,17], and, more generally, for a broader notion and use of evidence in the medical science [18]. Despite these pressures, standard practices of evidence assessment are still based on statistical standards which are at odds with the integration of heterogeneous data [19]. Thus, there is an increasing awareness for the need of tools which support the evaluation of putative causal links between drugs and adverse reactions grounded on such heterogeneous evidence.

This article presents a new theoretical framework to support decisions in pharmacovigilance, defined "*E-Synthesis*", a tool encoded through a Bayesian network able to robustly screen and amalgamate evidence of different types (such as epidemiological studies, molecular biology research findings, case reports) that are combined together in order to form the "critical mass" needed to make decisions about the approval/withdrawal of a particular drug for/from the market [20,21].

The paper is structured as follows: In the first part, we present an overview of the main risk detection methods currently available to decision makers in pharmacosurveillance. In the second part, we present the theoretical underpinnings and the epistemic working of *E-Synthesis*: An instrument that uses Bayesian nets in order to exploit the joint evidential force coming from different lines of evidence. In the third part, we focus on one such line of evidence, namely the support provided by dose-responsiveness (*DR*). We draw on computational methods currently employed to process dose-response evidence in pre-approval trials [22–27] and embed them in our theoretical framework; a final discussion follows.

2. State of the Art

A series of directions have been explored in the field of collecting, evaluating and summarizing evidence of harm in pharmacosurveillance. Tools are subdivided according to: (i) The kind of evidence they allow to be taken into account in risk detection/assessment; (ii) and types of methods used to assess and measure it.

2.1. Aggregation of Spontaneous Reports

Spontaneous reports are a fundamental source for post-marketing drug safety, a set of information that is particularly valuable today when coupled with the current computational methods employed for big data analysis. FAERS (Food and Drug Administration (FDA) Adverse Event Reporting System) (https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/) implemented in the US, and EUDRAVIGILANCE (European Union Drug Regulating Authorities Pharmacovigilance) in the EU largely embodies this vision. (https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance. Differently from FAERS, EUDRAVIGILANCE tackles not only postmarketing safety surveillance but is aimed at reporting and evaluation of suspected adverse reactions since the very development of new drugs).

At a global scale, a well-known reservoir of data in this field is VigiBase (https://www.who-umc.org/vigibase/vigibase/), a unique WHO international database of individual case safety reports—the largest of its kind in the world—with over 19 million reports of suspected adverse effects of medicines, submitted since 1968 by member countries of the World Health Organization (WHO) Programme for International Drug Monitoring. VigiBase is continuously updated with incoming reports. A special focus has been put on spontaneous reports processing and a series of analysis algorithms aimed at the discovery of potential pharmacological harms has been created [28]. The most significant ones are: vigiRank, vigiGrade, and vigiMatch.

vigiRank [29] is a predictive model implemented in signal detection that ranks pharmacovigilance safety signals according to multiple aspects of strength of evidence. Parameters that make up the score are: Disproportionate reporting, recent reporting, geographic spread, informative reports and narratives included in case reports.

vigiGrade [30] is a multidimensional measure of the amount of information on reports that highlights quality issues in collection of individual case safety reports and gives a score for how complete a case report is. The main use of vigiGrade is as part of communication between countries belonging to the WHO on data quality, but it has also proven to be an indicator of a true signal and is therefore one of the parameters included in the vigiRank method used by the WHO in signal detection. vigiMatch [31] is a probabilistic record matching method, a likelihood-based approach to identify unexpectedly similar record pairs in large databases. vigiMatch acts as a data cleaner and computes a match score for each pair of records, where matching information is rewarded and mismatching information penalised. This score reflects the probability that the two records relate to the same underlying entity or that they are duplicates. Record pairs with match scores that exceed a certain threshold are flagged as suspected duplicates. The threshold is derived from a comparison between the match scores of confirmed duplicates and of random record pairs in the database of interest.

2.2. Aggregation of Human and Animal Data

The extrapolation of animal data to estimate effects on humans has in general proved to be very fruitful, especially in fields where hazard measurement definitely results to be complex and awkward. However, animal-human extrapolation is notoriously fraught with difficulties [32]. Integration of animal-human data is a viable solution to some of the problems raised by extrapolation. Regarding chemicals, guidance in this respect has been available for a long time [33]. In particular, ECETOC ([33], chapter 7) has identified the need to review and evaluate the different types of human data that are available, and to provide guidance on how such data could be used best in the risk assessment process. The proposed framework considers when and where human data could be used to support risk assessment and risk management decisions and how human and animal findings could be integrated and used in tandem. Its related model examines the quality of human and animal data and evaluates the risk, by merging all the information through an integrating matrix.

2.3. Bayesian Aggregation of Safety Trial Data

Bayesian methods have been proposed to substitute a frequentist approach in evaluating and aggregating evidence, especially in designing, monitoring and reporting randomized controlled trials (RCTs). Researchers like [34] claim that a Bayesian theory of statistical inference may provide a better match to the dynamic nature of clinical investigation, a context in which knowledge accrues across multiple studies and observations and in which synthesis of evidence often provides a more definitive answer to a particular question of interest. Through a Bayesian approach, accumulating data are used to formally update prior distributions via Bayes rule, yielding posterior distributions that in turn become prior distributions awaiting further data. An implementation of Bayesian methods for designing RCTs is described by [35] where the authors propose a pragmatic approach to Bayesian analysis of randomized trials that avoids the complexity but retains the core features of more elaborate Bayesian approaches. On the same perspective, a similar outlook has been taken for the combination of treatments in clinical trials by [36]. Moreover, a nice overview of the use of Bayesian methods to aggregate evidence in designing RCTs is provided by [37]; additional excellent references include [38–40].

2.4. Data Mining and Fusion Methods

Since their emergence, data mining methods have been considered a natural realm for evidence synthesis, especially in a pharmacosurveillance context. Over the years, signal detection within data of pharmacological interest has been carried along with the progresses of computer science and the availability of increasing computational power. In such contexts, data mining techniques have been

deployed as fast and efficient ways to detect possible ADR signals coming from different sources of evidence. Several data mining algorithms have been described in literature [41], mainly based on nonparametrical statistics theory, including the Reporting Odds Ratio (ROR), Multi-item Gamma Poisson Shrinker (MGPS), the Proportional Reporting Ratio (PRR), and the Information Component (IC) [42] to name a few examples. An overview of statistical signal detection methods has been outlined in [43], whereas the first seminal examination of data mining techniques was given by Hauben over several articles [44–46]. Though successful and broadly implemented also by the pharma-industry sector [47], the theoretical basis and limitations of these data mining methods should be understood by researchers and drug safety professionals, highlighting that such automated methods should not be mechanically accepted [48]. Furthermore, the published evaluations of these techniques are limited mainly to large regulatory studies and performance characteristics may differ from case to case, principally in the smaller security databases of drug developers; therefore, it is worthwhile to remark that conclusive comparisons of the main techniques have not yet been established and such methods still have to be considered an experimental field of research. Eventually, the mathematical basis of such techniques shall not obscure the numerous data confusion and distortion factors [44].

In partial response to such criticism, data fusion methods have been introduced. Data fusion has been defined as the process of integrating multiple data sources to produce more consistent, accurate, and useful information than that provided by any individual data source [49]. From this perspective, many research efforts in pharmacosurveillance can be seen as data fusion driven. For example, ref. [50] estimates the increased risk of drug-induced hyponatremia during high temperatures based on various kind of data such as administered drugs, onset dates, given causality assessments, sodium levels and geographical origin of the ADRs reports.

Given their versatility in processing information from multiple sources, data fusion methods are considered a promising technique for evidence synthesis. In fact, as stated by [51], classical data mining methods for drug safety signal detection have mostly been limited to the examination of one adverse event data stream at a time, whereas data fusion methods for signal detection are explicitly designed to involve multiple data streams, with the goal of detecting safety signals more quickly and reliably using multiple data streams relevant to drug safety. However, shifting to this new paradigm requires drug safety researchers to design and implement novel data fusion algorithms, mainly needed for: (i) Automated data cleansing; (ii) vocabulary mapping and control across cross-functional databases; (iii) statistical detection of safety signals using multiple data streams; (iv) incorporating clinical trial data into the signal detection problem; (v) incorporating known clinical knowledge, or context, into all aspects of the problems of pharmacosurveillance. An overview of possible solutions related to this problem is offered in [51,52].

2.5. Semantic (Web) Methods

Semantic techniques are emerging as another trend in evidence synthesis. Researchers in [53] state that semantic technologies may provide the means to address some of the challenges in processing evidence and synthesizing it within a pharmacological context. Particular emphasis is on: (i) Annotating data sources and analysis methods with quality attributes to facilitate their selection given the analysis scope; (ii) consistently defining study parameters such as health outcomes and drugs of interest, and providing guidance for study setup; (iii) expressing analysis outcomes in a common format enabling data sharing and systematic comparisons; (iv) assessing/supporting the novelty of the aggregated outcomes through access to reference knowledge sources related to drug safety. According to [53], a semantically-enriched framework can facilitate seamless access and use of different data sources and computational methods in an integrated fashion, bringing a new perspective for large-scale, knowledge-intensive signal detection. Furthermore, the implementation of semantic algorithms finds a natural landing on web-based data. Authors in [54] explicitly tackle this perspective, offering a pharmacovigilance model based on a semantic web-based platform for continuous and integrated monitoring of ADRs in open data sources and social media.

3. *E-Synthesis*: A Bayesian Epistemology-Driven Framework for Pharmacovigilance

The methods presented in the previous sections are well developed and established; however they are mainly focused on specific kinds of evidence and methods. We present here a broader, holistic approach, which intends to exploit the *joint contribution* of different kinds of evidence to the (dis)confirmation of hypotheses of causal assessment: *E-Synthesis*.

E-Synthesis emerges as a theoretical framework for causal assessment based on [20,21], aimed to support estimates of causal associations between drugs and adverse events. *E-Synthesis* practically consists in a "scaffold" where all the available evidence may probabilistically increase or decrease causal hypotheses. (Note that the framework does not aim to provide utilities of harms, nor probabilities of expected benefits, nor utilities of benefits. When utilities of harms and benefits, as well as estimation of benefits are provided by further means, a drug regulatory agency can perform an expected utility calculation to determine whether the expected advantages of a drug exceed the expected dis-advantages. The agency will withdraw the drug (or not approve it), if the expected disadvantages outweigh the expected benefits, see ([20], Section 2.2).)

The **hypothesis of interest** is: 'Drug D causes harm E in population U and model $\mathcal{M}$'. To facilitate the inference from all the available evidence, **indicators of causality** are used. These indicators are based on Hill's nine viewpoints for causal assessment [55]. Below, we detail the main components that build up *E-Synthesis*.

3.1. Bayesian Network Model

In *E-Synthesis*, the probability of the causal hypothesis ©️ is modelled via a Bayesian network linking a finite number of propositional variables, see ([20], Section 5). The "root" variable represents the hypothesis of causal association ('Drug D causes harm E in population U and model $\mathcal{M}$'), its children represent abstract "causal indicators", such as statistical association ("Probabilstic Dependence", PD), dose-response (DR), rate of growth (RoG), etc. Every concrete study result probabilistically confirms one or more of such indicators with higher or lower strength, depending on its reliability (internal and external validity). The latter is gauged by "evidential modulators", see Figure 1.

The conditional probabilistic (in-)dependencies can be gleaned from the graph in terms of [56]'s d-separation criterion. While the Bayesian network is used for causal assessment, the arrows in the network are *not* causal arrows in [56]'s sense—here they represent epistemic probabilistic (in-)dependencies only. Figure 2 is an example graph with only one report. The causal hypothesis represented by (©️) is a root node (The abbreviations and symbols used here are listed at the end of the manuscript). The causal indicators are parents of report nodes which mediate causal inference from the concrete data (the reports) towards the causal hypothesis. The parents at the two bottom levels are modulators of the evidential strength of the data. These incorporate considerations about the reliability of the evidence into the assessment of the hypothesis. In particular, they take into account the possibility of random error (as a function of sample size (SS) and study duration (D)), and systematic error, attenuated by adjustment or stratification (A), randomisation (R) and blinding (B), and favoured by sponsorship bias (SB).

This framework also allows for the incorporation of evidential modulators related to external validity (called "relevance" in [20]), however we will not treat them here for ease of exposition, since they are extensively addressed in [57].

A probability function consistent with the conditional independencies of the Bayesian network is selected which expresses our uncertainties in the tradition of Bayesian epistemology [58–60]. In case one does conditionalise on a particular model, conditional probabilities are (virtually) always set to the probabilities of the statistical model. However, unlike in "pure" Bayesian statistics, where one conditionalises on statistical models and hence obtains conditional probabilities mandated by the particular model (parameter), in Bayesian epistemology one may conditionalise on any proposition

Int. J. Environ. Res. Public Health **2019**, *16*, 2221

6 of 19

(or event), since probabilities are interpreted more widely as one's uncertainties about general propositions; the Bayesian statistician Lindley is sympathetic to this approach [61].
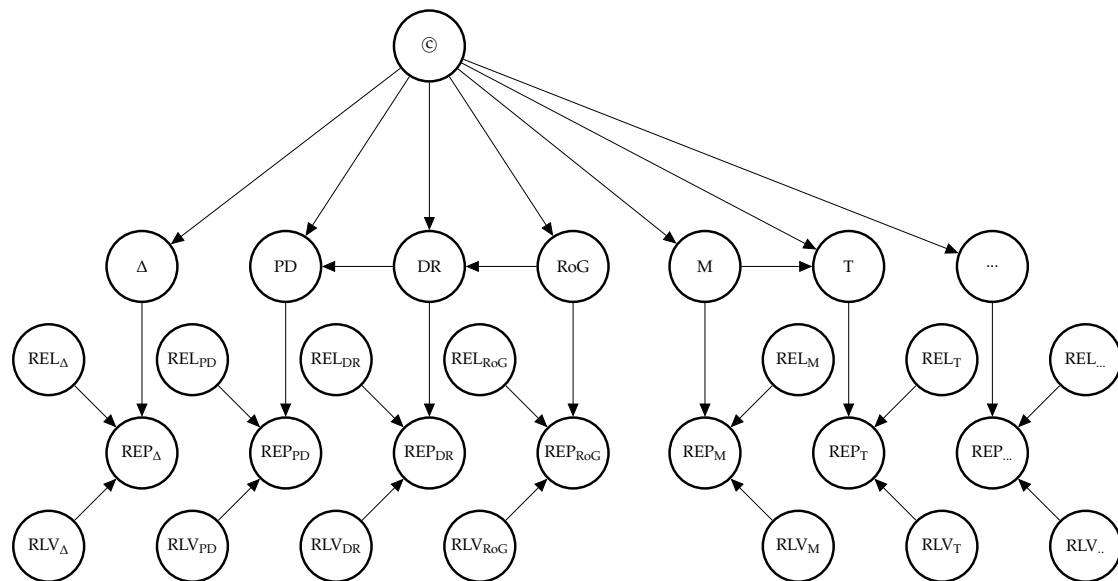

Figure 1. Graph of the Bayesian network with one report for every causal indicator variable. The dots indicate that there might be further indicators of causality not considered here. As explained in text, we here take it that *M* entails *T* and hence introduces an arrow from *M* to *T* which is not in [20]. *REL* and *RLV* act as evidential modulators of data (*REP* nodes).
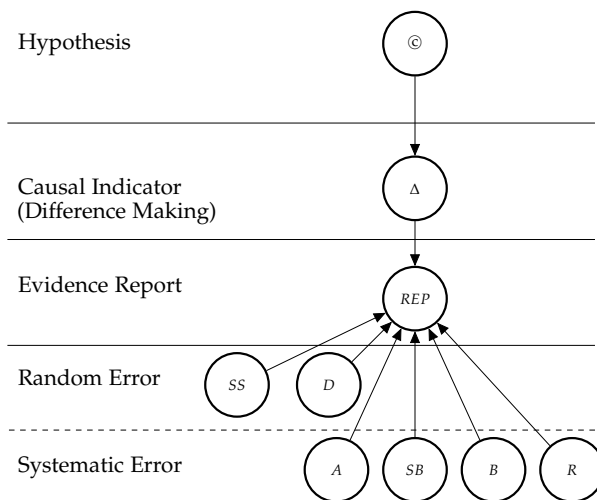

Figure 2. Graph structure of the Bayesian network for one randomised controlled trial (RCT) which informs us about difference making (Δ) which in turn informs us about the causal hypothesis. The information provided by the reported study is modulated by how well the particular RCT guards against random and systematic error.

3.2. Theoretical Entities

Concepts of interest fall into two classes: (i) A class of causal concepts comprised of the hypothesis of interest and the indicators of causation and (ii) a class of evidential concepts including evidential modulators and reports (data).

3.2.1. The Causal Hypothesis (ⓒ)

We are interested in determining the probability of the causal hypothesis that a drug *D* causes a particular adverse effect *E* in a population *U*—given the available evidence. While the hypothesis space in principle could be subdivided into three hypotheses: (1) *D* causes *E*, (2) *D* hinders *E*, and (3) *D* does not cause *E*, we divide it here for simplicity's sake into two alternative hypotheses: (1) *D* causes *E* and (2) *D* does not cause *E*, which consists of the disjunct of 2 and 3 above. To shorten notation, we use the symbol ⓒ in order to denote causation, such as in *D*ⓒ*E*, or simply ⓒ. (We do not commit here to any specific view or definition of causation (e.g., dispositional, probabilistic, counterfactual, manipulationist, etc., see also [20]. That is a question on the ontology of causation that we leave open for the moment. Our causal hypothesis allows for the term "causes" to cover any of the current definitions of causality to the extent that the evidence used for causal inference may be made relevant to them).

3.2.2. Indicators of Causation

Causal inference is mediated in the framework by "indicators of causation" in line with the Bradford Hill Guidelines for causation. As Hill puts it [55]:

> "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us make up our minds in the fundamental question—is there any other way of explaining the set of facts before us, is there any other equally, or more, likely than cause and effect?" (Bradford Hill both refers to explanatory power and likelihood as reliable grounds to justify causal judgements, and presents the respective criteria as opposed to tests of significance: "No formal tests of significance can answer those questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us on the likely magnitude of those effects. Beyond that, they contribute nothing to the proof of our hypothesis." [55])

In epistemic terms, causal indicators can be considered as observable and testable consequences of causal hypotheses. They are however indeterministic consequences of such hypotheses; that is they are more likely to be observed than not in the presence of a causal relationship, but they are not entailed by it. However, the first indicator ("difference-making", Δ) is a perfect one, in that it entails causation (although it is not entailed by it).

Difference-making (Δ): If *D* and *E* stand in a difference-making relationship, then changes in *D* make a difference to *E* (while the reverse might not hold). In contrast with mere statistical measures of association, the difference-making relationship is an asymmetric one. Probabilistic dependence can go in both ways (e.g., if *Y* is probabilistically dependent on *X*, then also *X* is probabilistically dependent on *Y*); the same does not hold for difference making, which provides information about its direction. This explains why experimental evidence is considered particularly informative with respect to causation; the reason is exactly that in experiments, putative causes are intervened upon, in view of establishing whether they make a difference to the effect. (In philosophical terms, difference-making is understood as *ideal controlled variance* along the concept of *intervention* in manipulationist theories of causation (see [20] for a detailed treatment, see also [62] and [56]): *X* is called a cause of *Y* if *Y*'s value can be varied by varying *X* (possibly upon controlling for additional variables in the given situation)). Hence, randomised controlled trials (RCTs) are the privileged source of evidence for Difference-making (Δ): Reports from RCTs contribute to the (dis)confirmation of causal hypotheses via the Δ node.

Probabilistic dependence (*PD*): *PD* encodes whether *D* and *E* are probabilistically dependent or not—such dependence naturally increases our belief in some underlying causal connection (as an indicator of causation; see, e.g., [63]). Probabilistic dependence is an imperfect indicator of causation because neither the former entails the latter nor the reverse. There are cases in which probabilistic dependence is created by confounding factors, and cases where two opposite effects of

a single cause cancel each other out and produce a zero net effect. (A well-known example of this type of cancellation is Hesslow's birth control pills case (see, e.g., [64]): The contraceptive (directly) causes thrombosis but simultaneously (indirectly) prevents thrombosis by preventing pregnancy which is a cause of thrombosis. Cartwright ([64]) discusses this case as one of the pitfalls of reducing causal analysis to probabilistic methodology alone. Of course, if cancellation is suspected, one might disable certain preventative causal routes to check whether the causal relationship actually shows once disabling conditions are held fixed. Cartwright however discusses cases in which this strategy might not even be viable, owing to the complexity of the causal web.).

Dose-response relationship (DR**)**: Dose-response relationships are taken as strong indicators of causation. DR is a stronger indicator than probabilistic dependency alone, because it requires the presence of a clear pattern of ≥ 3 data-points relating input and output. Indeed DR implies PD. Dose-response relationships can be inferred both at the population and at the individual level, and both in observational and experimental studies. DR abstracts away from these specifications and means that for dosages $D > 0$ in the therapeutic range, the adverse effect E shows (approximate) monotonic growth for a significant portion of the range (see below, Figure 3, for an illustration of important types of dose-response curves).

Rate of growth (RoG**)**: This indicator signals that the dose-response relationship is a steep one. If we have evidence of DR and evidence of $\overline{RoG}$ means either that the rate of growth is low, or highly non-linear.

The indicators of causality RoG, DR and PD are independent of the causal structure, in the sense that they could be equally observed either in cases where D causes E, or in cases where E causes D, or when D and E have a common cause. All that matters is whether there is a (certain) systematic relationship between D and E. RoG, DR, PD are semantically and epistemically related and we refer to them as "statistical black-box indicators", denoted by Σ.

Mechanistic Knowledge (M**)**: M represents the proposition: "there is a mechanism", meaning that there is a physiological pathway from drug use to the effect. In the biological realm, a causal relationship obviously entails the presence of a biological pathway connecting the cause to the effect. Therefore, ⓒ $\Rightarrow M$. However, this pathway may not be causally responsible for bringing out the effect due to possible inhibitors, back-up mechanisms, feedback loops, etc. $M \Rightarrow$ ⓒ does hence not necessarily hold.

Time course (T**)**: T encodes whether D and E stand in the right temporal relationship (time course), which can refer to temporal order, distance, or duration. If D causes E, T must hold (as a necessary condition): ⓒ $\Rightarrow T$. T remains an imperfect indicator, nevertheless, because temporal precedence is also compatible with $\neg(D$ causing $E)$ while D and E are connected by a common cause or through reversed causation. Hence $T \Rightarrow$ ⓒ does not necessarily hold.

According to our view, study design determines the kind of information that the evidence is able to provide us, hence in the following we define the evidence as a function of the kind of information it delivers: That is, the causal indicator it "speaks" to.

We associate study designs with causal indicators as follows: RCTs provide information about difference making, time course, possibly also dose-response relationship and rate of growth. Cohort studies provide evidence of time course and statistical association (Σ). Case-control studies provide information about Σ only. Individual case reports cannot provide information about statistical association, but they provide very detailed information about time course and possibly difference-making whenever this can be established with confidence (see for instance the Karch-Lasagna or Naranjo algorithm [65–67]). However, such information is "local", that is about an individual subject, and therefore do not license inferences about the general population. A case series can then possibly help delineate a reference class, where the putative causal link holds. Basic science studies (in vitro, or in silico), and in vivo studies, are generally the main source for evidence on the mechanisms underpinning the putative causal link.
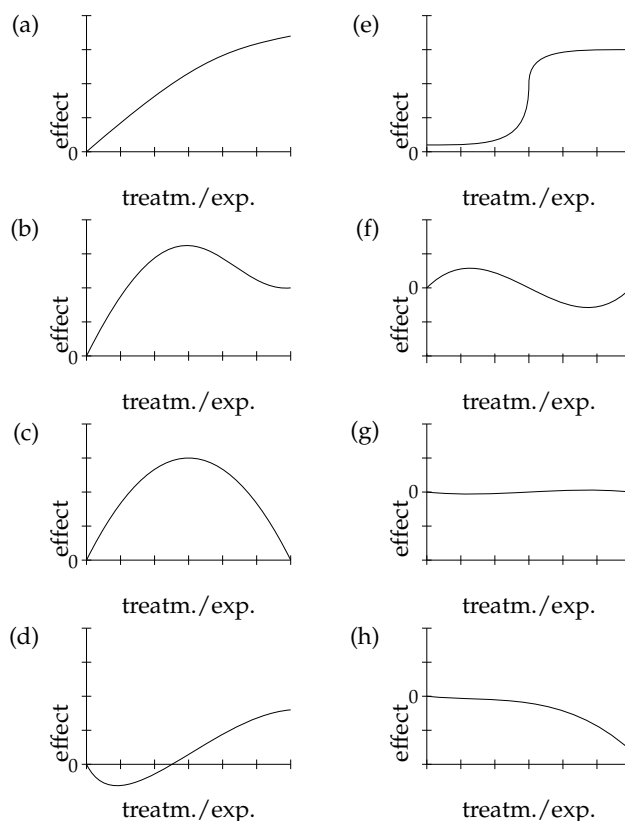
Figure 3. Possible functional forms of the relationship between *D* and *E*. Adapted from [22] (Figure 1).

4. Zooming in *E-Synthesis*: Processing Evidence for Dose-Response

As outlined above, in *E-Synthesis* many indicators of causality contribute to the assessment of pharmacological risk. Among them, we decide in this part to focus on Σ, and, in particular, on the *DR* arm. This choice is motivated by the fact that dose-responsiveness has been investigated for a long time as a manifest signal of pharmaceutical harm in pharmacosurveillance [68] (non-dose-related ADRs have been known and studied in the field too [69]).

In [21] we presented the epistemic dynamics of *E-Synthesis* by applying it to a case study: The debated causal association between paracetamol and asthma. We focus here on the *DR* arm. [70] is a study supporting the claim of paracetamol and asthma association, based on evidence pertaining to both *DR* and *T*. The study reports a prospective cohort study for women in which, between 1990 and 1996, 73,321 subjects were included in the analysis. Proportional hazard models included age, race, socioeconomic status, body mass index, smoking, other analgesic use, and postmenopausal hormone use. During 297,282 person-years of follow-up, 299 participants reported a new physician diagnosis of asthma meeting diagnostic criteria (see Table 1). For this study, in [21] we derived:

$$P(ES = 1|\vec{x}, DR, T) = 1 \tag{1}$$

and:

$$P(ES = 1|\vec{x}, \overline{DR}, \overline{T}) = 0 \tag{2}$$

where *ES* stands for the observed effect size (with significant = 1, not significant = 0) and $\vec{x}$ is a vector containing all the modulators introduced in the previous part. Evaluating such research study allows one to quantify the probability of witnessing a proxy of hypothesis of causation (i.e., the effect size, in this case) conditioned on the presence of dose-responsiveness and temporal

precedence. This example highlights well that *E-Synthesis* is able to process the given evidence, but this process occurs through the mediating "lenses" of an evaluation of research studies, meaning that all the available evidence has always been analyzed ab initio by other researchers.

Table 1. Raw data derived from the asthma association study [70] (Table 2). The study lasted from 1990 to 1996 and accounted for 297,282 person-years. In this table d_i represents the dose-level, n_i the number of subjects in each dose-group, y_i the number of subjects with effect in the corresponding dose-group and r_i is the dose-response computed as the ratio y_i / n_i.

d_i	n_i	y_i	r_i
0	137,568	108	7.85×10^{-4}
10	99,922	112	1.12×10^{-3}
31.67	32,077	41	1.28×10^{-3}
60	10,656	16	1.50×10^{-3}
86.67	17,059	22	1.29×10^{-3}

Nevertheless, from a more general point of view, a preferred approach would be to compute the above probabilities directly, by treating the evidence at glance. If we focus on the *DR* arm, what we would like to compute is thus:

$$P(DR|\mathcal{E}_{DR}, \vec{x}) \tag{3}$$

that is the probability of dose-responsiveness, given the available dose-response evidence and its related modulators (From this point onward, our problem will be approximated by the study of $P(DR|\mathcal{E}_{DR})$. The introduction of modulators into this framework will be addressed in a forthcoming article.)

Hence, a key question here is which may be the most suitable computational methods to be implemented to process *DR* evidence with the final goal of making a sound estimate of such probability. Regarding this problem, in recent years an extensive literature has been developed not directly in the realm of pharmacosurveillance but more specifically for dose-finding assessment, a task that is usually targeted in Phase II of clinical research. In [22], an algorithm aimed at this task named Multiple Comparison Procedures and Modeling (MCP-Mod) is presented. This algorithm is composed of two main steps. The "MCP" step which includes: (i) The assessment of a dose-response signal using a suitable test for trend; (ii) the model selection (or model averaging) out of the set of statistically significant dose-response models. The "Mod" step focuses on dose-response and target dose-estimation based on the previous selected model.

Given our interest in pharmacovigilance, we will focus only on the "MCP" step. The mentioned test for trend is a data analysis procedure able to demonstrate a dose-response association between the risk factor (drug prescription) and the outcome (adverse event), even if the association is not statistically significant for any particular level of exposure. A clear description of this test (also known as p-trend test or just p-trend) is presented in [71], whereas a good mathematical exposition can be found in [72]. A discussion regarding the use of confidence intervals instead of *p*-values in such test is analyzed in [73].

The model selection represents the second part of the "MCP" algorithm. Here, given some kind of association having been established in the test for trend analysis, the focus is on the detection of the best dose-response pattern. In other words, if we look back at Figure 3, our goal here is to select which kind of panel best fits the results obtained by the p-trend test. For instance, if we consider the (g) panel, the final outcome is non dose-responsiveness, whereas the (e) panel presents a sigmoid *DR* pattern. As shown in [22] (Table 5), every model we may select has an associated probability computed on the given evidence. However, our main interest is limited to discriminating between a horizontal pattern (i.e., signalling $\overline{DR}$) and a non horizontal one, whatever functional form it may exhibit. This may offer a welcome shortcut to the computation of $P(DR|\mathcal{E}_{DR})$.

Int. J. Environ. Res. Public Health **2019**, *16*, 2221

11 of 19

Since its introduction in [22], the MCP-Mod algorithm has been subject of subsequent research, with various enhancements presented in [23,24], with highlights also on subgroups analysis [25] and extensions to a continuous set of possible dose-response patterns [26]. A Bayesian approach for dose-finding has been also proposed in several papers [74–77]. In particular, in [27], we find an explicit formula for calculating $P(data|\theta)$ where θ represents a set of parameters embodied in a statistical model $f(d_i|\theta)$ that defines the possible functional patterns of dose-responsiveness exhibited in Figure 3.

$$\ln[P(data|\theta)] = \sum_{i=1}^{G} \left\{ \ln \binom{n_i}{y_i} + y_i \ln\left[f(d_i|\theta)\right] + (n_i - y_i) \ln\left[1 - f(d_i|\theta)\right] \right\} \tag{4}$$

In this formula, d_i represents the dose-level, n_i the number of subjects in each dose-group, y_i the number of subjects with effect in the corresponding dose-group and r_i is the dose-response computed as the ratio y_i/n_i. The index i is a natural number less or equal to G, the number of dose-groups.

We exploit Equation (4) to compute $P(data|\overline{DR})$. In fact, beyond many statistical models representing dose-responsiveness, a non dose-response pattern is expressed only by a family of constant functions, that is $f(d_i|\theta) = k$ where k is a real number subjected to the following bound: $0 \leq k \leq 1$. Among all the possible values of k, it is possible to calculate the optimal one $(\widetilde{k})$ through a least squares regression analysis based on the given dose-response (r) data. That can be achieved by solving this problem:

$$\widetilde{k} = \arg\min_{k} \sum_{i=1}^{G} (r_i - k)^2 = \langle r \rangle \tag{5}$$

The bracket-operator $\langle \cdot \rangle$ expressed in Equation (5) coincides with the arithmetic mean over the set $\mathcal{R}$ containing the given dose-response (r) data. (Instead of using a common overline notation to express the arithmetic mean (i.e., $\bar{r}$), we prefer to introduce here the bracket-operator $\langle \cdot \rangle$, since the overline already appears for the nonoccurrence of events in the probabilistic formulas.).

Given this information, we can now compute the likelihood of the data, given the absence of any dose-response pattern $\overline{DR}$:

$$\ln[P(data|\overline{DR})] = \sum_{i=1}^{G} \left\{ \ln \binom{n_i}{y_i} + y_i \ln\langle r \rangle + (n_i - y_i) \ln(1 - \langle r \rangle) \right\} \tag{6}$$

Equation (6) turns out to be very useful to compute $P(DR|\mathcal{E}_{DR}, \vec{x})$. In fact, through the Bayes theorem is easy to prove that:

$$P(DR|data) = 1 - P(\overline{DR}|data) = 1 - \frac{P(data|\overline{DR})P(\overline{DR})}{P(data)} = 1 - \frac{P(data|\overline{DR})[1 - P(DR)]}{P(data)} \tag{7}$$

where $P(data)$ can be written as:

$$P(data) = P(data|DR)P(DR) + P(data|\overline{DR})P(\overline{DR}) \tag{8}$$

In order to compute such expression, it appears crucial to determine which statistical model $f(d_i|\theta)$, i.e., which functional pattern of dose-responsiveness has to be used in Equation (4). Here, we do not commit to any particular kind of likelihood but we point at the fact that only data-analysis can provide a correct estimation for Equation (8). Moreover, Equation (7) requires the estimation of $P(DR)$, that is the a priori probability of dose-responsiveness. This probability can be estimated through clinical considerations regarding the expectation of dose-responsiveness for a certain drug, or, more generally, in a specific field of scientific inquiry. The degree of specificity of such a prior depends on the available knowledge.

We now present a numerical example of this new methodology to assess $P(DR|\mathcal{E}_{DR}, \vec{x})$. We again consider the study [70]. By analyzing the data shown in Table 1 (In order to make a more unbiased analysis, data have been divided into two separate sets, one used for model estimation and the other one for the computation of probabilities.), we compute different likelihood and evaluate their performance according to several established criteria for curve fitting, such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), R^2 and its adjusted version. Such analysis is outlined in Table 2: The exponential model emerges as the best according to all criteria (AIC, BIC, R^2 and adjusted R^2), whereas the linear and the cubic models perform worst.

Table 2. Model evaluation through several criteria: Akaike information criterion (AIC), Bayesian information criterion (BIC), R^2 and adjusted R^2. The best statistical model corresponds to an exponential function where $\theta = (\alpha, \beta) = (1.01266 \times 10^{-3}; 4.17805 \times 10^{-3})$.

Model	θ	Specification of $f(d_i|\theta)$	AIC	BIC	R^2	Adjusted R^2
exponential	(α, β)	$\alpha \exp(\beta d_i)$	-66.29	-67.46	0.98	0.97
linear	(α, β)	$\beta d_i + \alpha$	-66.79	-67.96	0.56	0.41
quadratic	(α, β, γ)	$\gamma d_i^2 + \beta d_i + \alpha$	-74.57	-76.13	0.95	0.90
cubic	$(\alpha, \beta, \gamma, \delta)$	$\delta d_i^3 + \gamma d_i^2 + \beta d_i + \alpha$	-70.10	-72.06	0.95	0.80

From the data, it is possible to estimate the likelihood for non dose-responsiveness:

$$\ln[P(data|\overline{DR})] \approx -25.903 \tag{9}$$

$$P(data|\overline{DR}) \approx 5.63 \times 10^{-12} \tag{10}$$

and, for instance, the related likelihood for DR according to an exponential model: Table 3 presents the predicted values that are neeeded for such computation.

$$\ln[P(data|DR_{exp})] \approx -18.6327 \tag{11}$$

$$P(data|DR_{exp}) \approx 8.09 \times 10^{-9} \tag{12}$$

If we again consider Equation (7), we conclude that:

$$P(DR|data) \approx 1 \tag{13}$$

This results are consistent with what was previously done in [21] and shown in Equation (1). In Figure 4, we present the curves associated to various $P(DR|data)$ for the best and worst tested statistical models at varying levels of $P(DR)$.

Table 3. Predicted values through exponential model.

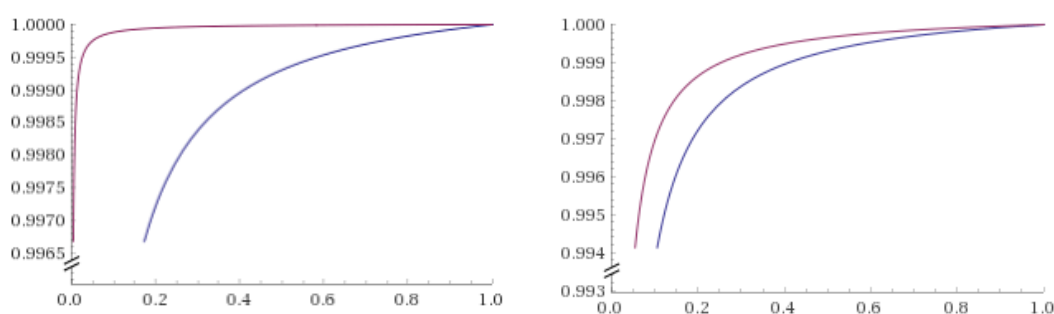| $\hat{r}_i$ |
|---|
| 1.01×10^{-3} |
| 1.06×10^{-3} |
| 1.16×10^{-3} |
| 1.30×10^{-3} |
| 1.46×10^{-3} |

Figure 4. This figure shows the conditional probability $P(DR|data)$ computed for the data presented in Table 1. X-axis represents $P(DR)$ and Y-axis $P(DR|data)$. On the left panel, $P(DR_{exp}|data)$ (blue line), embodying an exponential likelihood for DR, is pictured against $P(DR_{cub}|data)$ (violet line), calculated by using a cubic likelihood. On the right panel, $P(DR_{exp}|data)$ (blue line) is again pictured against $P(DR_{lin}|data)$ (violet line), computed with a linear likelihood.

5. Discussion

Adverse drug reactions present a significant problem to drug manufacturers, are a serious risk to patients, and constitute a major ethical problem in licensing decisions of pharmaceutical products. Current standards for drug evaluation are based on a methodology developed for the assessment of benefits that focuses on internal validity and privileges randomized controlled trials. However, evidence for harm often emerges unsystematically and unpredictably in form of anecdotal reports, case series and survey data. Hence, this methodology faces overwhelming challenges in integrating such diverse evidence.

Whereas a sophisticated set of tools for meta-analysis and systematic reviews has been developed for the purpose of evaluating intended effects of interventions, the adaptation of these instruments with the aim of evaluating safety of health technologies encounters various problems, mainly due to the (i) relative scarcity, (ii) heterogeneity, and (iii) "fragility" of data concerning unintended effects of medical treatments: (i) Evidence about unknown risks emerges gradually from spontaneous reports or other kinds of sporadic data: Especially in the earlier phases of risk detection, these data can be at the same time very noisy and rare; (ii) furthermore, they may come from heterogeneous sources (such as clinical case series, animal studies, or molecular studies, etc.); (iii) and it can be unreliable, because of noise and bias. Risk minimization requires however, that timely decisions are made on the basis of the available, albeit inconclusive and possibly "low-quality" evidence.

One major obstacle for the optimization of the kind of diverse evidence that we encounter in pharmacosurveillance is inherited from the epistemology underpinning the standard methodological paradigm entrenched in the medical community, i.e., frequentist hypothesis testing. This only works with predefined research protocols and can pool only homogeneous data. Thus, current approaches lack the justificatory underpinning for anecdotal reports, case series, or survey data, which are the main source of evidence for drug safety assessment. Such a perspective, together with the implementation of new computational methods [78], challenges the structure of decision making processes within the pharma-industry and the regulatory bodies.

Starting from these premises, in Section 2 we have shown ongoing trends in evidence synthesis methods. There, we have presented a state of the art of current methods employed to assess and measure risk in pharmacovigilance, highlighting in particular aggregation of spontaneous reports and data mining and fusion techniques.

Following this path, in Section 3 we have outlined *E-Synthesis*, a Bayesian framework for causal assessment that represents a new approach to evidence synthesis grounded on theoretical bases [20,21]. *E-Synthesis* can be pictured as a Bayesian network able to synthesize evidence of different types (such as epidemiological studies, molecular biology research findings, case reports) that are combined together in order to form the "critical mass" needed to make decisions about the approval/withdrawal

of a particular drug on the market. Through several indicators of causality (difference-making, probabilistic dependence, dose-response relationship, rate of growth, mechanistic knowledge and time course), *E-Synthesis* acts as a theoretical *ordering scaffold* on the available evidence and arrange it in order to assess the causal relationship between a certain drug and harm in a given population.

Among the indicators of causality, we have focused here on dose-responsiveness (Section 4). In particular, we have drawn on current methods applied in clinical research for dose-finding and tried to embed them in *E-Synthesis*. By combining different approaches and frameworks (MCP-Mod [22], Bayesian benchmark dose-estimation [27]), we have provided a new model to estimate the probability that a drug and a suspected harm are truly related by a non-spurious dose-response curve and offered also a numerical example, based on an association study between paracetamol and asthma. This numerical example agrees with the results already presented in [21].

However, there are some restrictions regarding the presented results that have to be highlighted. First of all, the numerical example tests a limited number of statistical models (see Table 2) and certainly an expansion of this number (like in [22], for instance) would be a welcome extension. Secondly, the example draws on a cohort study, whereas some changes to our model will be required to embed in a full way other kinds of evidence such as those arising from case reports. Moreover, as outlined in Equation (3), in [20,21] the analysis conducted for the probability estimates also included the implementation of modulators, which—for reasons of brevity—have been omitted here. We reserve to rejoinder all these issues in a forthcoming article.

A more general objection regards the use of Bayesian networks per se, in that the underpinning methodology is suspected to rely on "subjective probabilities". Indeed, unlike in "pure" Bayesian statistics, where one conditionalises on statistical models and hence obtains conditional probabilities mandated by the particular model (parameter), in Bayesian networks one may conditionalise on any proposition (or event), since probabilities are interpreted more widely as one's uncertainties about general propositions. While a certain degree of subjectivity is undeniable, there is a good argument to be made that some subjectivity is unavoidable in any approach to statistical/uncertain inference [79]. Furthermore a more articulated sense of the Bayesian way to interpret probabilities associated to hypotheses is that of "degree of support" (from the evidence), as opposed to "degree of belief" (see [80]) (What probabilities at a population level may mean to an individual has recently been explored in [81].) In the former sense, the probability does not represent the strength with which an agent may believe in a proposition, but rather, the amount of support provided by a body of evidence with respect to a hypothesis (and its alternatives). Finally, having (un)conditional probabilities being explicitly stated in the network makes the overall inference process more transparent, traceable and ultimately conducive to objectivity, both in the sense expressed by [82], in response to the debate inaugurated among others by Longino's analysis of the scientific enterprise [83], and in the more traditional sense expressed by Bayesians to this kind of objections, see for instance [84].

6. Conclusions

ADRs are responsible for a heavy economic and social burden [15]; additionally, they constitute an extremely vulnerable point for the health system and a key ethical problem for decisions concerning pharmaceutical products. In view of this, the European Parliament and the European Council have changed the regulation of pharmacovigilance practice (Directive 2010/84/EU; Regulation (EU) No 1235/2010, entered into force in July 2012) putting a special emphasis on joint efforts for what can be considered an information-based (rather than power-based) approach to pharmaceutical risk assessment. The related guidelines encourage the integration of information coming from different sources of safety signals (spontaneous case reports, literature, data-mining, pharmacoepidemiological studies, post-marketing trials, drug utilization studies, non-clinical studies, late-breaking information, see also [17]). However, the methodological bases for implementing such a policy are shaky in that causal assessment of ADRs is still parasitic on the (statistical) methods developed to test drug

efficacy [85], which are more focused on hypothesis testing, rather than detection and therefore fall short of efficiently using sporadic, fragile and heterogeneous evidence.

Within such perspective, methods for synthesizing heterogeneous evidence are rising as a more and more robust approach to drug safety. While developing its framework on a sound foundational basis, *E-Synthesis* aims to respond to the challenge of merging information from diverse sources by providing a structured way to exploit their joint (dis)confirmatory strength.

The theoretical system developed in [21] is being now translated into a full-fledged computational tool; here we have presented one of its modules: The dose-response line of evidence. Further steps in our research agenda include the development of computational algorithms related to each of the other causality indicators and related evidential modulators. A successful implementation of *E-Synthesis* to better predict ADRs and improve the risk management process (by minimizing both false positives and false negatives) will benefit the pharmaceutical industry, patients and drug agencies.

Abbreviations

The following abbreviations are used in this manuscript:

A	Adjustment for Confounders
ADR	Adverse drug reaction
AIC	Akaike information criterion
B	Blinding
BIC	Bayesian information criterion
D	Duration
DR	Dose-response Relationship
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ES	Effect Size
EUDRAVIGILANCE	European Union Drug Regulating Authorities Pharmacovigilance
FAERS	FDA Adverse Event Reporting System
FDA	Food and Drug Administration (US)
IC	Information Component
M	Mechanistic Knowledge
MCP-Mod	Multiple Comparison Procedures and Modeling algorithm
MGPS	Multi-item Gamma Poisson Shrinker
PD	Probabilistic Dependence
PRR	Proportional Reporting Ratio
R	Randomisation
RCT	Randomized controlled trial
REP	Report Variable
RoG	Rate of Growth
ROR	Reporting Odds Ratio
SB	Sponsorship Bias
SS	Sample Size

Int. J. Environ. Res. Public Health **2019**, *16*, 2221

16 of 19

T	Temporal Precedence
WHO	World Health Organization
©	Hypothesis of Causation
Δ	Difference Making
$\mathcal{E}$	Available Evidence
Σ	Set of statistical Indicators: *PD*, *DR* and *RoG*

References

1. Stausberg, J.; Hasford, J. Drug-related admissions and hospital-acquired adverse drug events in Germany: A longitudinal analysis from 2003 to 2007 of ICD-10-coded routine data. *BMC Health Serv. Res.* **2011**, *11*, 134. [CrossRef] [PubMed]

2. Wu, T.Y.; Jen, M.H.; Bottle, A.; Molokhia, M.; Aylin, P.; Bell, D.; Majeed, A. Ten-year trends in hospital admissions for adverse drug reactions in England 1999–2009. *J. R. Soc. Med.* **2010**, *103*, 239–250. [CrossRef] [PubMed]

3. Brvar, M.; Fokter, N.; Bunc, M.; Mozina, M. The frequency of adverse drug reaction related admissions according to method of detection, admission urgency and medical department specialty. *BMC Clin. Pharmacol.* **2009**, *9*, 8. [CrossRef] [PubMed]

4. van der Hooft, C.S.; Dieleman, J.P.; Siemes, C.; Aarnoudse, A.J.L.; Verhamme, K.M.; Stricker, B.H.; Sturkenboom, M.C. Adverse drug reaction-related hospitalisations: A population-based cohort study. *Pharmacoepidemiol. Drug Saf.* **2008**, *17*, 365–371. [CrossRef] [PubMed]

5. Kongkaew, C.; Noyce, P.R.; Ashcroft, D.M. Hospital admissions associated with adverse drug reactions: A systematic review of prospective observational studies. *Ann. Pharmacother.* **2008**, *42*, 1017–1025. [CrossRef] [PubMed]

6. Aktionbündnis Patientensicherheit. Agenda Patientensicherheit 2007. Available online: http://www.aps-ev.de/wp-content/uploads/2016/10/Agenda_2007_mit_Titelblatt.pdf (accessed on 2 April 2019).

7. EMA. Better Vigilance for Public Health Protection. Overview of the New European Union Pharmacovigilance Legislation. Available online: https://www.ema.europa.eu/en/documents/leaflet/better-vigilance-health-protection-innovation_en.pdf (accessed on 2 April 2019).

8. Mokdad, A.H.; Marks, J.S.; Stroup, D.F.; Gerberding, J.L. Actualcauses of death in the United States, 2000. *JAMA* **2004**, *291*, 1238–1245. [CrossRef] [PubMed]

9. Wysowski, D.K.; Swartz, L. Adverse drug event surveillance and drug withdrawals in the United States, 1969–2002: The importance of reporting suspected reactions. *Arch. Intern. Med.* **2005**, *165*, 1363–1369. [CrossRef] [PubMed]

10. Pirmohamed, M.; James, S.; Meakin, S.; Green, C.; Scott, A.K.; Walley, T.J.; Farrar, K.; Park, B.K.; Breckenridge, A.M. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18,820 patients. *BMJ* **2004**, *329*, 15–19. [CrossRef] [PubMed]

11. Schneeweiss, S.; Hasford, J.; Göttler, M.; Hoffmann, A.; Riethling, A.K.; Avorn, J. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: A longitudinal population-based study. *Eur. J. Clin. Pharmacol.* **2002**, *58*, 285–291. [CrossRef]

12. Hay, M.; Thomas, D.W.; Craighead, J.L.; Economides, C.; Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **2014**, *32*, 40. [CrossRef]

13. Goldman, M. Reflections on the innovative medicines initiative. *Nat. Rev. Drug Discov.* **2011**, *10*, 321. [CrossRef] [PubMed]

14. Goldman, M. The innovative medicines initiative: A European response to the innovation challenge. *Clin. Pharmacol. Ther.* **2012**, *91*, 418–425. [CrossRef] [PubMed]

15. Lundkvist, J.; Jönsson, B. Pharmacoeconomics of adverse drug reactions. *Fundam. Clin. Pharmacol.* **2004**, *18*, 275–280. [CrossRef] [PubMed]

16. Cooper, N.; Coyle, D.; Abrams, K.; Mugford, M.; Sutton, A. Use of evidence in decision models: An appraisal of health technology assessments in the UK since 1997. *J. Health Serv. Res. Policy* **2005**, *10*, 245–250. [CrossRef] [PubMed]

17. Herxheimer, A. Pharmacovigilance on the turn? Adverse reactions methods in 2012. *Br. J. Gen. Pract.* **2012**, *62*, 400–401. [CrossRef] [PubMed]

18. Anjum, R.L.; Copeland, S.; Rocca, E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion "of evidence". *BMJ Evid.-Based Med.* **2018**. [CrossRef] [PubMed]

19. Amrhein, V.; Greenland, S.; McShane, B. Scientists rise up against statistical significance. *Nature* **2019**, *567*, 305–307. [CrossRef]

20. Landes, J.; Osimani, B.; Poellinger, R. Epistemology of causal inference in pharmacology. Towards a framework for the assessment of harms. *Eur. J. Philos. Sci.* **2018**, *8*, 3. [CrossRef]

21. De Pretis, F.; Landes, J.; Osimani, B. E-Synthesis: A Bayesian framework for causal assessment in pharmacosurveillance. *Front. Pharmacol.* **2019**, Submitted.

22. Bretz, F.; Pinheiro, J.C.; Branson, M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **2005**, *61*, 738–748. [CrossRef]

23. Pinheiro, J.; Bornkamp, B.; Glimm, E.; Bretz, F. Model-based dose finding under model uncertainty using general parametric models. *Stat. Med.* **2014**, *33*, 1646–1661. [CrossRef] [PubMed]

24. Schorning, K.; Bornkamp, B.; Bretz, F.; Dette, H. Model selection versus model averaging in dose finding studies. *Stat. Med.* **2016**, *35*, 4021–4040. [CrossRef] [PubMed]

25. Bornkamp, B.; Ohlssen, D.; Magnusson, B.P.; Schmidli, H. Model averaging for treatment effect estimation in subgroups. *Pharm. Stat.* **2017**, *16*, 133–142. [CrossRef] [PubMed]

26. Gutjahr, G.; Bornkamp, B. Likelihood ratio tests for a dose-response effect using multiple nonlinear regression models. *Biometrics* **2017**, *73*, 197–205. [CrossRef] [PubMed]

27. Shao, K.; Shapiro, A.J. A web-based system for Bayesian benchmark dose estimation. *Environ. Health Perspect.* **2018**, *126*, 017002. [CrossRef]

28. Watson, S.; Chandler, R.E.; Taavola, H.; Härmark, L.; Grundmark, B.; Zekarias, A.; Star, K.; van Hunsel, F. Safety concerns reported by patients identified in a collaborative signal detection workshop using vigibase: Results and reflections from Lareb and Uppsala Monitoring Centre. *Drug Saf.* **2018**, *41*, 203–212. [CrossRef] [PubMed]

29. Caster, O.; Sandberg, L.; Bergvall, T.; Watson, S.; Norén, G.N. vigiRank for statistical signal detection in pharmacovigilance: First results from prospective real-world use. *Pharmacoepidemiol. Drug Saf.* **2017**, *26*, 1006–1010. [CrossRef]

30. Bergvall, T.; Norén, G.N.M. vigiGrade: A tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf.* **2014**, *37*, 65–77. [CrossRef]

31. Tregunno, P.M.; Fink, D.B.; Fernandez-Fernandez, C.; Lázaro-Bengoa, E.; Norén, G.N. Performance of probabilistic method to detect duplicate individual case safety reports. *Drug Saf.* **2014**, *37*, 249–258. [CrossRef]

32. Shanks, N.; Greek, R.; Greek, J. Are animal models predictive for humans? *Philos. Ethics Humanit. Med.* **2009**, *4*, 2. [CrossRef]

33. ECETOC. Framework for the Integration of Human and Animal Data in Chemical Risk Assessment. Available online: http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-104.pdf (accessed on 25 March 2019).

34. Price, K.L.; Amy Xia, H.; Lakshminarayanan, M.; Madigan, D.; Manner, D.; Scott, J.; Stamey, J.D.; Thompson, L. Bayesian methods for design and analysis of safety trials. *Pharm. Stat.* **2014**, *13*, 13–24. [CrossRef] [PubMed]

35. Spiegelhalter, D.J.; Freedman, L.S.; Parmar, M.K.B. Bayesian approaches to randomized trials. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1994**, *157*, 357–387. [CrossRef]

36. Thorlund, K.; Golchi, S.; Mills, E. Bayesian adaptive clinical trials of combination treatments. *Contemp. Clin. Trials Commun.* **2017**, *8*, 227–233. [CrossRef] [PubMed]

37. Berry, D.A. Bayesian clinical trials. *Nat. Rev. Drug Discov.* **2006**, *5*, 27; Review Article. [CrossRef]

38. Spiegelhalter, D.J. *Bayesian Approaches to Clinical Trials and Health Care Evaluation*; Spiegelhalter, D.J., Abrams, K.R., Myles, J.P., Eds.; Statistics in Practice; Wiley: Chichester, UK, 2006.

39. Berry, S.M.; Carlin, B.P.; Lee, J.J.; Muller, P. *Bayesian Adaptive Methods for Clinical Trials*; Bioscience, Mathematics & Statistics; Taylor & Francis: Boca Raton, FL, USA, 2010.

40. Emmanuel Lesaffre, A.B.L. *Bayesian Biostatistics*; John Wiley & Sons: Hoboken, NJ, USA, 2012. [CrossRef]

41. Poluzzi, E.; Raschi, E.; Piccinni, C.; De Ponti, F. Data mining techniques in pharmacovigilance: Analysis of the publicly accessible FDA adverse event reporting system (AERS). In *Data Mining Applications in Engineering and Medicine*; Karahoca, A., Ed.; IntechOpen: London, UK, 2012; pp. 201–213. [CrossRef]

42. Subeesh, V.; Maheswari, E.; Saraswathy, G.R.; Swaroop, A.M.; Minnikanti, S.S. A comparative study of data mining algorithms used for signal detection in FDA AERS database. *J. Young Pharm.* **2018**, *10*, 444–449. [CrossRef]

43. Candore, G.; Juhlin, K.; Manlik, K.; Thakrar, B.; Quarcoo, N.; Seabroke, S.; Wisniewski, A.; Slattery, J. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf.* **2015**, *38*, 577–587. [CrossRef] [PubMed]

44. Hauben, M.; Zhou, X. Quantitative methods in pharmacovigilance. Focus on signal detection. *Drug Saf.* **2003**, *26*, 159. [CrossRef] [PubMed]

45. Hauben, M.; Madigan, D.; Gerrits, C.M.; Walsh, L.; Van Puijenbroek, E.P. The role of data mining in pharmacovigilance. *Exp. Opin. Drug Saf.* **2005**, *4*, 929–948. [CrossRef] [PubMed]

46. Hauben, M.; Patadia, V.; Gerrits, C.; Walsh, L.; Reich, L. Data mining in pharmacovigilance. The need for a balanced perspective. *Drug Saf.* **2005**, *28*, 835–842. [CrossRef] [PubMed]

47. Klepper, M.J.; Cobert, B. *Drug Safety Data: How to Analyze, Summarize and Interpret to Determine Risk*; Jones & Bartlett Learning: Burlington, MA, USA, 2011.

48. Edwards, I.R.; Lindquist, M. *Pharmacovigilance. Critique and Ways Forward*; ADIS: Zurich, Switzerland, 2017.

49. Haghighat, M.; Abdel-Mottaleb, M.; Alhalabi, W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1984–1996. [CrossRef]

50. Jönsson, A.K.; Lövborg, H.; Lohr, W.; Ekman, B.; Rocklöv, J. Increased risk of drug-induced hyponatremia during high temperatures. *Int. J. Environ. Res. Public Health* **2017**, *14*, 827. [CrossRef] [PubMed]

51. Streit, R.; Silver, J. Data fusion aspects of pharmacovigilance. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; pp. 1522–1528.

52. Shibata, A.; Hauben, M. Pharmacovigilance, signal detection and signal intelligence overview. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; pp. 1515–1521.

53. Koutkias, V.G.; Jaulent, M.C. Computational approaches for pharmacovigilance signal detection: Toward integrated and semantically-enriched frameworks. *Drug Saf.* **2015**, *38*, 219. [CrossRef] [PubMed]

54. Piccinni, C.; Poluzzi, E.; Orsini, M.; Bergamaschi, S. PV-OWL—Pharmacovigilance surveillance through semantic web-based platform for continuous and integrated monitoring of drug-related adverse effects in open data sources and social media. In Proceedings of the 2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), Modena, Italy, 11–13 September 2017; pp. 1–5. [CrossRef]

55. Hill, A.B. The environment and disease: Association or causation? *Proc. R. Soc. Med.* **1965**, *58*, 295–300. [CrossRef] [PubMed]

56. Pearl, J. *Causality: Models, Reasoning, and Inference*, 1 ed.; Cambridge University Press: Cambridge, UK, 2000.

57. Poellinger, R. Analogy-based inference patterns in pharmacological research. In *Uncertainty in Pharmacology: Epistemology, Methods, and Decisions*; La Caze, A., Osimani, B., Eds.; Boston Studies in Philosophy of Science; Springer: Berlin, Germany, 2018; forthcoming.

58. Bovens, L.; Hartmann, S. *Bayesian Epistemology*; Oxford University Press: Oxford, UK, 2003.

59. Howson, C.; Urbach, P. *Scientific Reasoning*, 3rd ed.; Open Court Publishing: Chicago, IL, USA, 2006.

60. Talbott, W. Bayesian epistemology. In *Stanford Encyclopedia of Philosophy*, Summer 2011 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2011; Available online: http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian (accessed on 25 March 2019).

61. Lindley, D.V. The philosophy of statistics. *J. R. Stat. Soc. Ser. D Stat.* **2000**, *49*, 293–337. [CrossRef]

62. Woodward, J. *Making Things Happen: A Theory of Causal Explanation*; Oxford Studies in the Philosophy of Science; Oxford University Press: Oxford, UK, 2003.

63. Reichenbach, H. *The Direction of Time*; University of California Press: Berkeley, CA, USA, 1956.

64. Cartwright, N. What is wrong with Bayes nets? *Monist* **2001**, *84*, 242–264. [CrossRef]

65. Naranjo, C.A.; Busto, U.; Sellers, E.M.; Sandor, P.; Ruiz, I.; Roberts, E.; Janecek, E.; Domecq, C.; Greenblatt, D. A method for estimating the probability of adverse drug reactions. *Clin. Pharm. Ther.* **1981**, *30*, 239–245. [CrossRef]

66. Karch, F.E.; Lasagna, L. Toward the operational identification of adverse drug reactions. *Clin. Pharm. Ther.* **1977**, *21*, 247–254. [CrossRef]

67. Varallo, F.R.; Planeta, C.S.; Herdeiro, M.T.; de Carvalho Mastroianni, P. Imputation of adverse drug reactions: Causality assessment in hospitals. *PLoS ONE* **2017**, *12*, e0171470. [CrossRef]

68. Phillips, E.J. Classifying ADRs—Does dose matter? *Br. J. Clin. Pharmacol.* **2016**, *81*, 10–12. [CrossRef]

Int. J. Environ. Res. Public Health **2019**, *16*, 2221

19 of 19

69. Aronson, J.K.; Ferner, R.E. The law of mass action and the pharmacological concentration-effect curve: Resolving the paradox of apparently non-dose-related adverse drug reactions. *Br. J. Clin. Pharmacol.* **2016**, *81*, 56–61. [CrossRef]

70. Barr, R.G.; Wentowski, C.C.; Curhan, G.C.; Somers, S.C.; Stampfer, M.J.; Schwartz, J.; Speizer, F.E.; Carlos, A.; Camargo, J. Prospective study of acetaminophen use and newly diagnosed asthma among women. *Am. J. Respir Crit. Care Med.* **2004**, *169*, 836–841. [CrossRef] [PubMed]

71. Patino, C.M.; Ferreira, J.C. Test for trend: Evaluating dose-response effects in association studies. *J. Bras. Pneumol.* **2016**, *42*, 240. [CrossRef] [PubMed]

72. Vittinghoff, E.; Glidden, D.V.; Shiboski, S.C.; McCulloch, C.E. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*; Springer: New York, NY, USA, 2005; p. 340.

73. Hothorn, L.A. Trend tests in epidemiology: P-values or confidence intervals? *Biometr. J.* **1999**, *41*, 817–825. [CrossRef]

74. Liu, S.; Johnson, V.E. A robust Bayesian dose-finding design for phase I/II clinical trials. *Biostatistics* **2016**, *17*, 249–263. [CrossRef] [PubMed]

75. Takeda, K.; Morita, S. Bayesian dose-finding phase I trial design incorporating historical data from a preceding trial. *Pharm. Stat.* **2018**, *17*, 372–382. [CrossRef]

76. Toumazi, A.; Comets, E.; Alberti, C.; Friede, T.; Lentz, F.; Stallard, N.; Zohar, S.; Ursino, M. dfpk: An R-package for Bayesian dose-finding designs using pharmacokinetics (PK) for phase I clinical trials. *Comput. Methods Programs Biomed.* **2018**, *157*, 163–177. [CrossRef] [PubMed]

77. Mu, R.; Xu, J. A new Bayesian dose-finding design for drug combination trials. *Stat. Biopharm. Res.* **2017**, *9*, 384–389. [CrossRef]

78. De Pretis, F. New mathematical perspectives to understand the Information Society: the statistical mechanics approach to model and analyze big-data. In Proceedings of the International Conference of Young Scientists and Specialists "Information Society as Contemporary System of Defense and Attack", Baku, Azerbaijan, 27–28 November 2014; Mütercim Publishing House: Baku, Azerbaijan, 2016; pp. 3–10.

79. Gelman, A.; Hennig, C. Beyond subjective and objective in statistics. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2017**, *180*, 967–1033. [CrossRef]

80. Hawthorne, J. Degree-of-Belief and Degree-of-Support: Why Bayesians Need Both Notions. *Mind* **2005**, *114*, 277–320. [CrossRef]

81. Dawid, A.P. On individual risk. *Synthese* **2017**, *194*, 3445–3474. [CrossRef]

82. Sprenger, J. The objectivity of Subjective Bayesianism. *Eur. J. Philos. Sci.* **2019**. [CrossRef]

83. Longino, H.E. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*; Princeton University Press: Princeton, NJ, USA, 1990.

84. Edwards, W.; Lindman, H.; Savage, L.J. Bayesian statistical inference for psychological research. *Psychol. Rev.* **1963**, *70*, 193. [CrossRef]

85. Senn, S.S. *Statistical Issues in Drug Development*; John Wiley & Sons: Hoboken, NJ, USA, 2008; Volume 69. [CrossRef]