

## RESEARCH ARTICLE

EA<sup>3</sup>: A softmax algorithm for evidence appraisal aggregationFrancesco De Pretis<sup>1,2</sup>\*, Jürgen Landes<sup>3</sup>

**1** Department of Biomedical Sciences and Public Health, School of Medicine and Surgery, Marche Polytechnic University, Ancona, Italy, **2** Department of Communication and Economics, University of Modena and Reggio Emilia, Reggio Emilia, Italy, **3** Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München, München, Germany

\* These authors contributed equally to this work.

\* [francesco.depretis@unimore.it](mailto:francesco.depretis@unimore.it)

## Abstract

Real World Evidence (RWE) and its uses are playing a growing role in medical research and inference. Prominently, the 21st Century Cures Act—approved in 2016 by the US Congress—permits the introduction of RWE for the purpose of risk-benefit assessments of medical interventions. However, appraising the quality of RWE and determining its inferential strength are, more often than not, thorny problems, because evidence production methodologies may suffer from multiple imperfections. The problem arises to aggregate multiple appraised imperfections and perform inference with RWE. In this article, we thus develop an evidence appraisal aggregation algorithm called EA<sup>3</sup>. Our algorithm employs the softmax function—a generalisation of the logistic function to multiple dimensions—which is popular in several fields: statistics, mathematical physics and artificial intelligence. We prove that EA<sup>3</sup> has a number of desirable properties for appraising RWE and we show how the aggregated evidence appraisals computed by EA<sup>3</sup> can support causal inferences based on RWE within a Bayesian decision making framework. We also discuss features and limitations of our approach and how to overcome some shortcomings. We conclude with a look ahead at the use of RWE.

## OPEN ACCESS

**Citation:** De Pretis F, Landes J (2021) EA<sup>3</sup>: A softmax algorithm for evidence appraisal aggregation. PLoS ONE 16(6): e0253057. <https://doi.org/10.1371/journal.pone.0253057>

**Editor:** Chi-Hua Chen, Fuzhou University, CHINA

**Received:** January 19, 2021

**Accepted:** May 27, 2021

**Published:** June 17, 2021

**Copyright:** © 2021 De Pretis, Landes. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data that support the findings of this study are available via open access in *Frontiers in Pharmacology* at doi: [10.3389/fphar.2019.01317](https://doi.org/10.3389/fphar.2019.01317), reference number 10:1317.

**Funding:** FDP acknowledges funding from the European Research Council (PhilPharm - GA n. 639276) through the Marche Polytechnic University (Ancona, Italy). JL gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 432308570 and 405961989. The funders had no role in study design, data collection

## Introduction

Real World Evidence (RWE) [1] is one of the new frontiers of medical research and inference and attracts growing interests in academic and industrial research. RWE comprises observational data obtained outside the context of Randomised Controlled Trials (RCTs) which are produced during routine clinical practice. According to a broader understanding, it may be possible to point at any source of information, that is related to medications and not directly retrievable from RCTs, as a potential generator of RWE, e.g. social networks [2].

Despite being known for a long time and in some cases applied as an informative support in the drug approval process [3] (e.g. the anticoagulant Rivaroxaban [4]), RWE has recently been brought to the fore by the US Congress with the Pub.L. 114—255 (21st Century Cures Act) which modified in 2016 the Food and Drug Administration (FDA) procedures for

and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

medications licensing. The act allows, under certain conditions, pharmaceutical companies to provide “data summaries” and RWE such as observational studies, insurance claims data, patient input, and anecdotal data rather than RCTs data for drug approval purposes. After the turn to RCTs as gold-standard in the drug approval process, this is the first act allowing for uses of RWE in the drug approval process in an industrialised country. This move sparked interest also of the European Medical Agency (EMA) and the Japanese Pharmaceuticals and Medical Devices Agency (PMDA) [5, 6].

The use and standards for proper use of RWE have ignited a serious debate in the scientific community [7–11]; for a special issue see [12]. Proponents of the use of RWE point to the fact that RWE can be produced much faster than conducting and analysing a clinical study [13, 14]. This allows pharmaceutical companies to obtain approval for new products or new indications (off-label use) quicker, which can benefit companies as well as patients [15]. Faster and safe drug approval procedures are particularly relevant during the current Covid-19 pandemic [16, 17]. However, many researchers have expressed concerns related to data quality, validity, reliability and sensitivity to capture the exposure, adverse effects and outcomes of interest when using RWE [18–22]. Using RWE for medical inference presents methodological challenges [23], though some efforts have been carried out to efficiently merge evidence coming from RCTs and observational studies [24–26], also for causal inference purposes [27, 28]. Attempts to provide a framework for appraising the quality of evidence for medical inference have been going on since long before the current debate on uses of RWE began, e.g. GRADE [29, 30]. However, these frameworks do not provide a clear way to quantitatively solve this problem nor do they lend themselves to an integration into a standard decision making framework [31–34].

The US National Research Council has issued following call: “*The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams*” [35]. This point appears crucial to us for appraising RWE. There is however no commonly accepted methodology for carrying out RWE appraisals. A possible solution to this problem is to split the appraisal of RWE into multiple more manageable appraisals along different dimensions and then to aggregate these appraisals. However, how can we aggregate these multiple appraisals? Subsequently, how can we use this aggregate for decision making?

We here address these two questions by proposing an algorithm based on (1) the softmax function—a generalisation of the logistic function to multiple dimensions—as an instrumental tool for aggregation within (2) a Bayesian decision making framework. While the softmax function was initially introduced in statistical mechanics, it has now found wide-spread applications in machine learning and artificial intelligence methods at large [36–38]. On the other hand, Bayesian approaches are increasing in popularity in part due to their intuitive incorporation of information and updating procedures.

Drawing on these traditions, we present an Evidence Appraisal Aggregation Algorithm, EA<sup>3</sup> (suggested pronunciation: “EA-cube”) compressing a generic vector of evidence appraisals along multiple dimensions into a scalar. Roughly, input data (evidence appraisals) are first processed through the softmax function and next aggregated by the application of a geometric mean. EA<sup>3</sup> is then shown to have some desirable properties. It offers the possibility of emphasizing or the de-emphasizing the maximum values associated to each evidence appraisal via a cautiousness parameter (the thermodynamic  $\beta$  of softmax). Furthermore, EA<sup>3</sup> allows one to incorporate the importance of the dimensions of appraisals. Eventually, we show how EA<sup>3</sup> can be used to support assessments of causal hypotheses within a Bayesian decision making approach.

To the best of our knowledge, EA<sup>3</sup> represents one of the first attempts to solve the problem of evidence appraisal through an easy-to-exploit numerical measure [39, 40]. In line with the previously mentioned US Environmental Protection Agency (EPA) recommendations [35], our appraisals can be understood as risk-of-bias assessments—but also of other possible methodological flaws. We offer a formalisation of such assessments and facilitate a tracking of these assessments through evidence aggregation to the calculation of probabilities of hypotheses of interest. Our proposal commits to be thus “transparent, reproducible and scientifically defensible” as suggested by the EPA [35, p. 79].

The rest of this article is organised as follows: in **Materials and Methods**, we introduce the softmax function as well as a motivating example and then present our softmax algorithm in some detail and discuss its properties. The **Results** section puts forward a method to apply EA<sup>3</sup> in Bayesian decision making problems. A final **Discussion** outlines advantages and limitations of our approach and points to important future work.

### Materials and methods

In this section, we first introduce the softmax function, then we present the EA<sup>3</sup> algorithm and discuss its properties.

#### Softmax

The softmax function (more correctly softargmax, also known as normalised exponential function) is a function from  $\mathbb{R}^k$  to  $\mathbb{R}^k$  ( $k \in \mathbb{N}$ ) mapping a vector  $\vec{A} = \langle a_1 \dots a_1 \dots a_k \rangle \in \mathbb{R}^k$  ( $k \geq 2$ ) to a vector  $\sigma(\vec{A})$  as follows:

$$\sigma(\vec{A})_i = \frac{\exp(\beta a_i)}{\sum_{i=1}^k \exp(\beta a_i)}, \tag{1}$$

where  $\beta$  is a real number different from zero, see Table 1 for an overview of key notation. We now briefly discuss some of the properties of the softmax function (henceforth softmax) and recall some of its applications to mathematical physics, probability theory, statistics, machine learning and artificial intelligence.

**Table 1. Key notation.**

$k$	number of dimensions of appraisals
$\vec{A}$	tuple of $k$ -appraisals
$\vec{R}$	ranking of dimensions of appraisal
$\vec{A} \times \vec{R}$	weighted mean of $\vec{A}$ and $\vec{R}$
$\beta$	cautiousness parameter
$v_f$	appraisal aggregate output of EA <sup>3</sup>
$EA^3(\vec{A}, \vec{R})$	appraisal aggregate output of EA <sup>3</sup> applied to appraisals $\vec{A}$ and ranking $\vec{R}$
$c@k$	$k$ -tuple with all entries equal to $c$
$\Omega$	finite set of possible worlds
$\mathcal{E}$	body of evidence
$P$	probability function
$Q$	probability function when RWE is not taken at face value
$\odot$	drug $D$ causes adverse reaction $E$
Ind	indicator of causation
RoG	Rate of Growth indicator of causation

<https://doi.org/10.1371/journal.pone.0253057.t001>

**Normalisation.** While the input vector may contain any real number, the output of softmax is normalised in the sense that all components of the output vector are in the unit interval and sum to one. The output vector can hence be understood as a probability distribution over  $k$  elementary events where the probabilities are proportional to the exponential of the input vector.

**Translational invariance.** Softmax is invariant under translations: let  $\vec{A}'$  be obtained from a vector  $\vec{A}$  by adding a constant  $c \in \mathbb{R}$  to every component of  $A$  then

$$\begin{aligned}\sigma(\vec{A}')_l &= \frac{\exp(\beta(a_l + c))}{\sum_{i=1}^k \exp(\beta(a_i + c))} = \frac{\exp(\beta a_l) \cdot \exp(\beta c)}{\sum_{i=1}^k \exp(\beta a_i) \cdot \exp(\beta c)} \\ &= \sigma(\vec{A})_l.\end{aligned}$$

So, if  $\vec{A}'$  is obtained from  $\vec{A}$  via translation, then  $\sigma(\vec{A}') = \sigma(\vec{A})$ .

Softmax is **not scale invariant**. It is easy to prove that multiplying every component of an input vector  $\vec{A}$  by some constant  $c$  does, in general, not return the same output vector.

**The  $\beta$  parameter** allows one to change the base of the exponential function. This choice permits one to emphasise or de-emphasise the maximum value belonging to the input vector, the greater  $\beta$  the greater the maximal component of the output vector. For  $\beta = +\infty$  the output vector vanishes everywhere except those components at which the input vector is the greatest (in this case, softmax becomes an *argmax*). Conversely, for  $\beta = -\infty$  the output vector vanishes everywhere except those components at which the input vector is the smallest (*argmin*). In the limit case  $\beta = 0$  the output vector is the uniform probability distribution resulting in a loss of all the information contained in the input.

The first use of softmax goes back to 1868 when Ludwig Boltzmann introduced the function for modelling ideal gases. Today, softmax is known as the Boltzmann-Gibbs distribution in statistical mechanics, where the index set  $\{1, \dots, l, \dots, k\}$  represents the microstates of a classical thermodynamic system in thermal equilibrium and  $a_l$  is the energy of that state  $l$  and  $\beta$  the inverse temperature (thermodynamic  $\beta$ ) [41, 42]. Beyond the representation of physical systems, the distribution and this modeling have paved the way to some noteworthy algorithms based on the same statistical mechanics assumptions, e.g. Gibbs sampling [43].

The normalisation property has led to applications of softmax in probability theory to represent a categorical distribution [44] and in statistics to define a classification method through the so-called softmax regression, an equivalent to multinomial logistic regression [45, 46]. This property has been widely used also in medical statistics [47–49].

In recent years, two fields have been seeing a raising interest towards softmax: machine learning and artificial intelligence [50, 51]. The term softmax itself has been first introduced by Bridle in neural networks, where it is usually employed as an activation function to normalise data [52]. In computer science, applications of softmax are varied: classification methods (again, softmax regression) for supervised and unsupervised learning [53–55], computer vision [56–58], reinforcement learning [59–61] and hardware design [62], just to name some current areas of application. Additionally, a considerable number of conference papers is witnessing the popularity of softmax and its proposed variants [63–67].

## Motivating example

Consider the hypothesis that paracetamol use causes asthma in children [68]. Only relatively few RCTs have been conducted that could help us determine the truth of this hypothesis [69]. RWE will thus have to (!) play an important role in treatment and prescription decisions that

have to be made now, that is before (meta-analyses of) RCTs can deliver conclusive evidence [70].

RWE for and against this causal hypothesis is, for example, obtained from relatively large surveys [71–77]. Such evidence is clearly less confirmatory than well-run RCTs and we hence need to find a way to appraise this evidence. De Pretis et al. (2019) [78] suggested that such surveys can be appraised along three independent and relevant dimensions: duration of the surveyed time period, the sample size and the methodology for adjustment and stratification. Appraisals are represented by numbers in the unit interval where 1 represents a perfect appraisal (e.g. perfect methodology for adjustment and stratification) and 0 represents the worst possible score (e.g. tiny sample size). These three appraisals are then aggregated by taking their arithmetic mean.

Simply taking the arithmetic mean is problematic for a number of reasons. Firstly, the dimensions of appraisal are all given the same weight. This problem can be easily addressed by moving to a weighted mean where the weights represent the importance of the dimensions of appraisal. Secondly, every weighted mean of three equal numbers  $c$  is equal to  $c$ . That is, multiple imperfections of RWE of equal degree  $c$  lead to an overall appraisal equal to  $c$ . We think, the overall appraisal ought to be less than  $c$ , multiple imperfections are worse than just one imperfection. Thirdly, a decision maker has no flexibility in the aggregation of appraisal to represent his/her attitude towards the question “how much worse are multiple imperfections than a single imperfection”. We hence think that a suitable aggregation is not idempotent.

We next present and explain the EA<sup>3</sup> algorithm to aggregate evidence appraisals, which addresses these points.

### The evidence appraisal aggregation algorithm EA<sup>3</sup>

We assume that evidence is appraised in  $k$  relevant and pairwise different and mutually independent dimensions represented by a normalised appraisal vector

$\vec{A} = \langle a_1 \dots a_l \dots a_k \rangle \in [0, 1]^k$ , see the *E-Synthesis* subsection for a suggested set of dimensions for appraisal. We do not commit to a fixed number of evidence appraisals (in agreement with multi criteria decision making in medicine [33] and risk prediction for multiple outcomes [79]).

We also make use of a given ranking of the importance of the different dimensions of appraisal. We represent this ranking by a vector  $\vec{R} = \langle r_1 \dots r_l \dots r_k \rangle \in (0, 1)^k$  such that  $\sum_{i=1}^k r_i = 1$ . The more important the appraisal  $a_i$ , the greater the value  $r_i$ .

EA<sup>3</sup> proceeds in 5 steps listed in Table 2 and explained below:

1. Appraisals weighted by ranking:

$$a_l^{[1]} := r_l \cdot a_l \text{ for all } 1 \leq l \leq k.$$

**Description.** Step 1 weighs every appraisal by its importance.

**Table 2. EA<sup>3</sup> algorithm structure with objectives described for each step.**

Step	Objective
1	Appraisals weighted by ranking
2	Softmax with a positive thermodynamic $\beta$
3	Rescaling
4	Geometric averaging
5	Normalisation to unit interval

<https://doi.org/10.1371/journal.pone.0253057.t002>

2. Softmax with a positive thermodynamic  $\beta$

$$a_i^{[2]} := \frac{\exp(\beta a_i^{[1]})}{\sum_{i=1}^k \exp(\beta a_i^{[1]})} = \frac{\exp(\beta \cdot r_i \cdot a_i)}{\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)}$$

**Description.** Step 2 applies, as advertised above, softmax with a parameter  $\beta$  representing cautiousness, cf. the discussion following Proposition 1.

3. Rescaling

$$a_i^{[3]} := a_i^{[2]} \cdot \sum_{i=1}^k r_i \cdot a_i = \frac{\exp(\beta \cdot r_i \cdot a_i)}{\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)} \cdot (\vec{A} \times \vec{R}),$$

where  $\times$  denotes the scalar product between two vectors of the same length  $k$ .

**Description.** Step 3 rescales the softmax of Step 2 by aggregated ranked appraisals. Softmax has the well-known property that it is invariant under uniform pointwise translations,  $\sigma(\langle a_1, \dots, a_k \rangle) = \sigma(\langle a_1 + c, \dots, a_k + c \rangle)$ . This property means for our application that applying softmax to a study  $S_1$  and to a study  $S_2$  which is appraised to be better according to every dimension by the same amount ( $c$ ) it holds that  $\sigma(S_1) = \sigma(S_2)$ . This is clearly undesirable as a uniformly better study should score better than a uniformly worse study. Multiplying by  $\vec{A} \times \vec{R}$  is a simple and intuitive way of ensuring that EA<sup>3</sup> is not invariant under uniform pointwise translations. Not only is our algorithm sensitive to pointwise translations, it is even the case that every improvement of an appraisal leads to a greater number  $v_f$  (see Proposition 2).

4. Geometric averaging:

$$v := \frac{\prod_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)}{\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)} \cdot (\vec{A} \times \vec{R}) = \frac{\exp(\beta \cdot (\vec{A} \times \vec{R}))}{\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)} \cdot (\vec{A} \times \vec{R})$$

**Description.** Step 4 compresses the vector to a scalar. To achieve this task, we apply a geometric mean, as it is routinely performed in machine learning for comparing items with a different number of properties and numerical ranges [80–82].

5. Normalisation to unit interval:

$$v_f := v \cdot \frac{\sum_{i=1}^k \exp(\beta \cdot r_i)}{\exp(\beta)} = \frac{\exp(\beta \cdot (\vec{A} \times \vec{R}))}{\exp(\beta)} \cdot \frac{\sum_{i=1}^k \exp(\beta \cdot r_i)}{\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)} \cdot (\vec{A} \times \vec{R})$$

**Description.** Step 5 ensures that the final output is in the unit interval. We find this normalisation convenient for our application and point out that this step might not be necessary for other applications.

To summarize, given two  $k$ -tuples  $\vec{A}, \vec{R} \in [0, 1]^k \times (0, 1)^k$  as input the algorithm returns a single number in the unit interval as output. We can understand EA<sup>3</sup> as a map and thus write  $EA^3(\vec{A}, \vec{R}) \in [0, 1]$  (see Corollary 1 for a proof that EA<sup>3</sup> maps into the unit interval).

**Properties of EA<sup>3</sup>**

Denoting by  $c@k$  a vector of length  $k$  with all components equal to  $c$ , we find that:

**Proposition 1.** EA<sup>3</sup> is not idempotent, i.e. for all  $c \in [0, 1]$  and all  $\beta > 0$  it holds that

$$EA^3(c@k, \frac{1}{k}@k) = c \cdot \exp\left(\beta \cdot (c - 1) \cdot \left(1 - \frac{1}{k}\right)\right). \tag{2}$$

*Proof.* The computation is straightforward:

$$\begin{aligned} v_f &:= \frac{\exp(\beta \cdot (c@k \times \frac{1}{k}@k))}{\exp(\beta)} \cdot \frac{\sum_{i=1}^k \exp(\beta \cdot \frac{1}{k})}{\sum_{i=1}^k \exp(\beta \cdot \frac{c}{k})} \cdot (c@k \times \frac{1}{k}@k) \\ &= \frac{\exp(\beta \cdot c)}{\exp(\beta)} \cdot \frac{\exp(\beta \cdot \frac{1}{k})}{\exp(\beta \cdot \frac{c}{k})} \cdot c \\ &= \exp(\beta \cdot (c - 1)) \cdot \frac{1}{\exp((c - 1) \cdot \frac{\beta}{k})} \cdot c \\ &= c \cdot \exp\left(\beta \cdot (c - 1) \cdot \left(1 - \frac{1}{k}\right)\right). \end{aligned}$$

This observation demonstrates the role of  $\beta$  and how the simplest ranking scheme (all dimensions are ranked equally) acts in the simple case in which all appraisals are equal to  $c$ , see Fig 1 for an illustration. The greater  $\beta$ , the smaller  $v_f$ , the further away the curves plotted in Fig 1 are away from the identity map. This means that a study with all appraisals equal to  $c$  will have an aggregate,  $v_f$ , equal to less than  $c$ . In other words, RWE that is less than perfect in more than one respect has an even lower aggregated appraisal. This seems right, studies which might produce poor evidence for multiple reasons are considered to produce very poor evidence. It is for this reason that we require that  $\beta > 0$ .

$\beta = +\infty$  represents maximal cautiousness, if the study is not perfect in all respects ( $c < 1$ ), then  $EA^3(c@k, \frac{1}{k}@k) = 0$ .  $\beta = 0$  represents maximal optimism (and in our eyes overly strong optimism) in that  $EA^3(c@k, \frac{1}{k}@k) = c$ , a study with a number of imperfections ( $c < 1$ ) is overall as good as just a single imperfection.

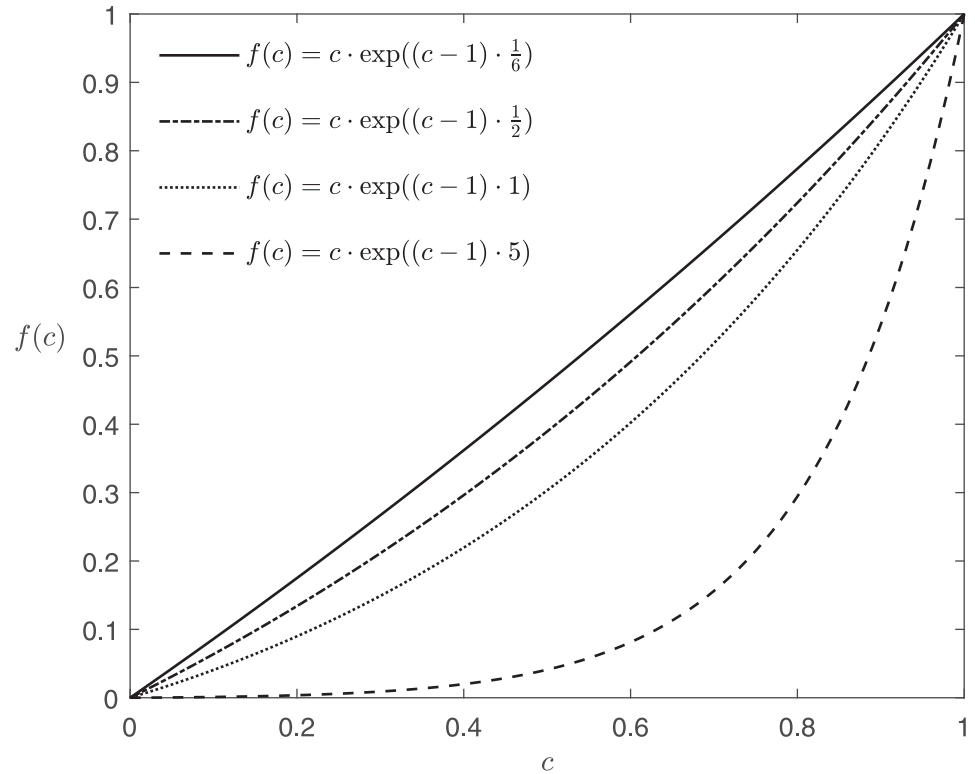
Furthermore, note that if  $\beta \ll 0$ , then Eq (2) may exceed 1. So, in such a case our Step 5 would fail normalise  $v_f$  to the unit interval and a different normalisation step would be required.

**Definition 1** (Monotonicity) We call a function  $f : [0, 1]^k \rightarrow \mathbb{R}$  monotone, if and only if the restriction of  $f$  to all coordinates is a strictly monotonously increasing function.

**Proposition 2.** For every given fixed ranking scheme  $\vec{R}$ , the function  $EA^3(\cdot, \vec{R})$  is monotone.

This proposition is key for our purposes as it states that every improved appraisal entails a better aggregate. In other words, better methodologies have a greater  $v_f$  which in turn have greater (dis-)confirmatory weight (see the Results section).

*Proof.* It suffices to verify that all the partial derivatives of  $EA^3(\cdot, \vec{R})$  with respect to the  $a_i$  are strictly positive for all  $a_i \in [0, 1]$ . Since the normalisation step is a multiplication by a scalar which does not depend on  $\vec{A}$ , it suffices to verify that all the partial derivatives of  $v$  with respect to the  $a_i$  terms are strictly positive for all  $a_i \in [0, 1]$ .



**Fig 1. Behaviour of EA<sup>3</sup>(c@k, 1/k@k) for varying β, where β · (1 - 1/k) is the second factor within the scope of the exponential function.** The smaller parameter β and the greater the number of appraisals (the greater k), the closer EA<sup>3</sup>(c@k, 1/k@k) gets to the identity map. This graph clearly displays the monotonicity of these functions.

<https://doi.org/10.1371/journal.pone.0253057.g001>

We now compute that this is indeed the case:

$$\begin{aligned} \frac{\partial}{\partial a_1} v &= \exp(\beta \cdot (\vec{A} \times \vec{R})) \cdot \frac{(a_1 + \beta \cdot r_1) \cdot (\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)) - (\beta \cdot r_1) \cdot (\vec{A} \times \vec{R})}{(\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i))^2} \\ &= \exp(\beta \cdot (\vec{A} \times \vec{R})) \cdot \frac{(a_1 + \beta \cdot r_1) \cdot (\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i)) - (\beta \cdot r_1) \cdot (\sum_{i=1}^k a_i \cdot r_i)}{(\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i))^2} \\ &> \exp(\beta \cdot (\vec{A} \times \vec{R})) \cdot \frac{(a_1 + \beta \cdot r_1) \cdot (\sum_{i=1}^k r_i \cdot a_i) - (\beta \cdot r_1) \cdot (\sum_{i=1}^k a_i \cdot r_i)}{(\sum_{i=1}^k \exp(\beta \cdot r_i \cdot a_i))^2} \\ &\geq 0 . \end{aligned}$$

The sharp inequality follow from the fact that  $\exp(\beta \cdot x) > 1 \geq x$  for all  $x \in [0, 1]$  and all  $\beta > 0$ .

**Corollary 1.** For every given fixed ranking scheme  $\vec{R}$ , the function EA<sup>3</sup>(·,  $\vec{R}$ ) maps into the unit interval, [0, 1]. Furthermore, we note that  $v_f(\vec{A}, \vec{R}) = 0 \Leftrightarrow \vec{A} = 0@k$  and  $v_f(\vec{A}, \vec{R}) = 1 \Leftrightarrow \vec{A} = 1@k$ .

*Proof.* Applying Proposition 2 it suffices to show that EA<sup>3</sup>(0@k,  $\vec{R}$ ) = 0 and EA<sup>3</sup>(1@k,  $\vec{R}$ ) = 1. The first condition follows from 0@k ×  $\vec{R}$  = 0 and the second from 1@k ×  $\vec{R}$  = 1.

Also note that if  $\vec{A} = 0@k$ , then  $\vec{A} \times \vec{R} = 0$  and thus  $v_f = 0$ . If  $\vec{A} \neq 0@k$ , then  $\vec{A} \times \vec{R} > 0$  and thus  $v_f > 0$ .



**Table 3. Evidence Appraisal Aggregation according to De Pretis et al. (2019) [78] and EA<sup>3</sup> with equally important appraisal dimensions ( $\vec{R} = \frac{1}{3} @ 3$ ) where SS represents the appraised sample size, D the appraised duration and A represents the appraised adjustment and stratification.**

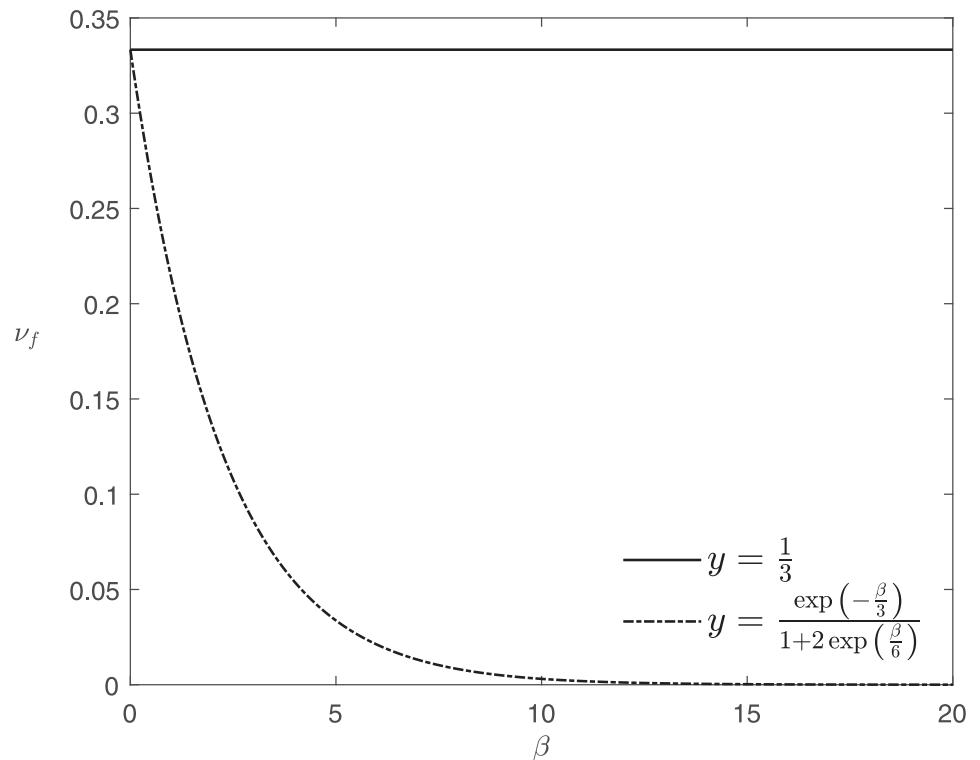
Survey	SS	D	A	De Pretis et al. (2019)	EA <sup>3</sup>
<i>Lesko and Mitchell (1999) [71]</i>	1	0	0.5	$\frac{1}{2}$	$\frac{3}{2[\exp(\frac{\beta}{6}) + \exp(\frac{\beta}{6}) + \exp(\frac{\beta}{6})]}$
<i>Newson et al. (2000) [72]</i>	1	0.5	1	$\frac{5}{6}$	$\frac{5}{2[1 + 2\exp(\frac{\beta}{6})]}$
<i>Lesko et al. (2002) [73]</i>	1	0	0.5	$\frac{1}{2}$	$\frac{3}{2[\exp(\frac{\beta}{6}) + \exp(\frac{\beta}{6}) + \exp(\frac{\beta}{6})]}$
<i>Shaheen et al. (2002) [74]</i>	1	1	1	1	1
<i>Karimi et al. (2006) [75]</i>	0.5	0	0.5	$\frac{1}{3}$	$\frac{\exp(-\frac{\beta}{3})}{1 + 2\exp(\frac{\beta}{6})}$
<i>Amberbir et al. (2011) [76]</i>	0.5	1	1	$\frac{5}{6}$	$\frac{5}{2[1 + 2\exp(\frac{\beta}{6})]}$
<i>Beasley et al. (2011) [77]</i>	1	0.5	1	$\frac{5}{6}$	$\frac{5}{2[1 + 2\exp(\frac{\beta}{6})]}$

<https://doi.org/10.1371/journal.pone.0253057.t003>

Similarly, if  $\vec{A} = 1 @ k$ , then  $\vec{A} \times \vec{R} = 1$  and thus  $v_f = 1$ . If  $\vec{A} \neq 1 @ k$ , then  $\vec{A} \times \vec{R} < 1$  and thus  $v_f < 1$ .

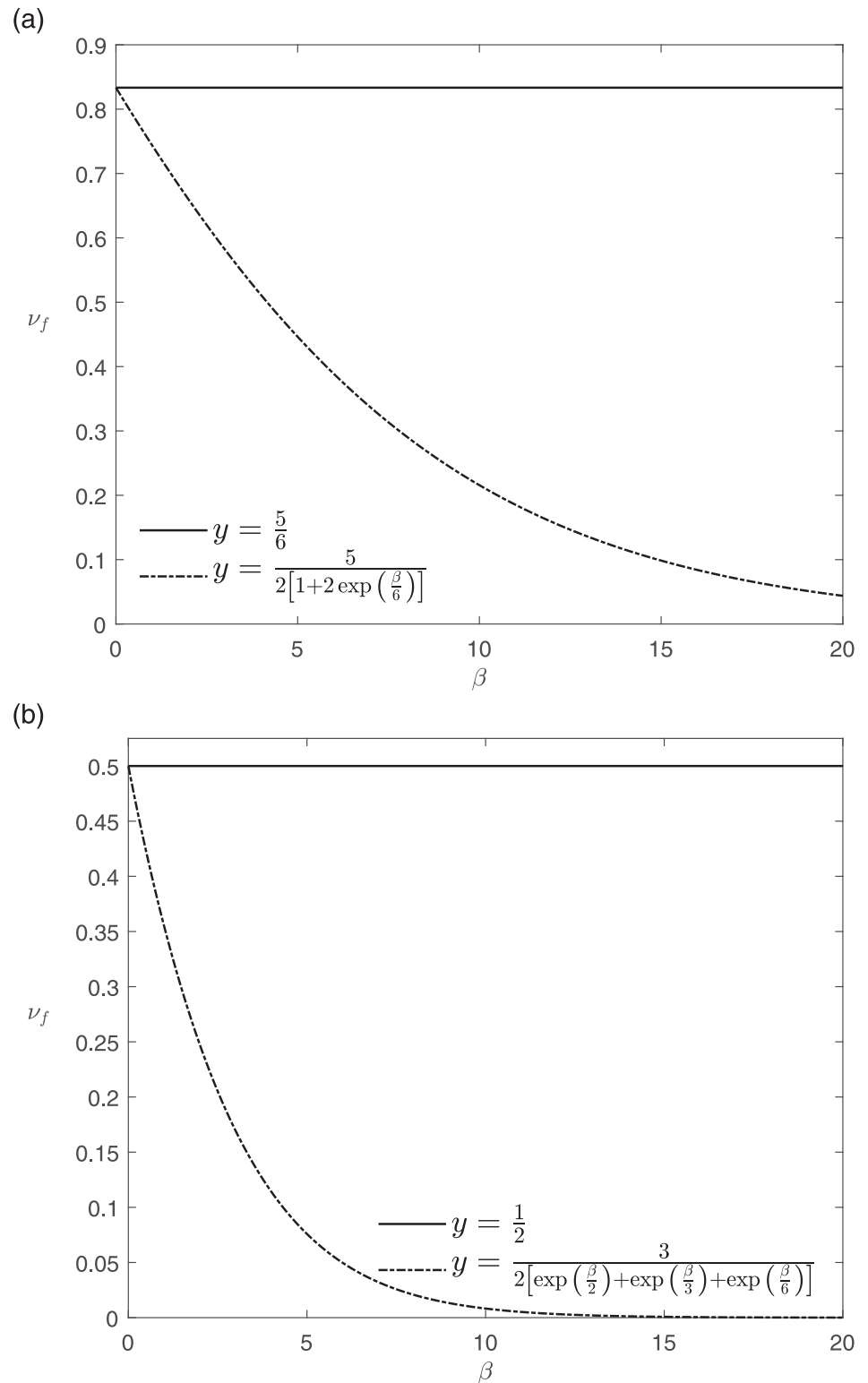
### The motivating example—reconsidered

Returning to the suspected causal link between paracetamol use and asthma, we now compare the aggregated appraisals of several RWE-providing surveys involving children, previously considered in [78], according to De Pretis et al. (2019) [78] and according to EA<sup>3</sup>. See Table 3 for the formulae and Figs 2 and 3 for a graphical comparison under the assumption of equally



**Fig 2. Aggregated appraisals of Karimi et al. (2006) [75] according to De Pretis et al. (2019) [78] (solid line) and EA<sup>3</sup> (dash-dot line).** The latter, lower, curve displays the behaviour with respect to the cautiousness parameter  $\beta$ . Both curves agree for  $\beta = 0$  where EA<sup>3</sup> equals the weighted mean.

<https://doi.org/10.1371/journal.pone.0253057.g002>



**Fig 3.** Similarly to Fig 2, the upper panel shows the aggregated appraisals of Newson et al. (2000) [72], Amberbir et al. (2011) [76] and Beasley et al. (2011) [77] according to De Pretis et al. (2019) [78] (solid line) and EA<sup>3</sup> (dash-dot line). The lower panel depicts the aggregated appraisals for Lesko and Mitchell (1999) [71] and Lesko et al. (2002) [73].

<https://doi.org/10.1371/journal.pone.0253057.g003>

important appraisal dimensions,  $\vec{R} = \frac{1}{3} @ 3$ . We note that for  $\beta = 0$  both approaches agree and that the aggregate appraisal computed with EA<sup>3</sup> decreases with increasing cautiousness parameter  $\beta$ .

We are not aware of other approaches of qualitative aggregations of multiple evidence appraisals for medical inference. We hence lack a standard against which to benchmark our proposal. However, there are substantive bodies of literature on aggregating numerically represented judgements and preferences, which, at times, tackle a formally equivalent aggregation problem. A related proposal for medical inference is the GRADE methodology, which puts forward a way to obtain a qualitative confidence rating in hypotheses. The suggestion is to use the lowest confidence ranking for critical outcomes as the aggregate confidence [83]. By contrast, our approach is quantitative and all appraisals contribute to the aggregate.

Another field relevant our work is the current research on Bayesian hierarchical models for aggregation. In the already mentioned [24, 25] such models are employed to combine different study types in meta-analysis and account for bias, with the objective of its correction. Whereas in this article we consider one study and multiple appraisals of bias, the inverse may be considered true in [24]. There, the author employs a bias-correcting Bayesian hierarchical model [84] to combine different study types in meta-analysis. That model is based on a mixture of two random effects distributions, where the first component corresponds to the model of interest and the second component to the hidden bias structure. The resulting model is thus adjusted by the internal validity bias of the studies included in a systematic review.

## Results. Application of EA<sup>3</sup> to Bayesian decision making problems

### The bayesian framework

We now illustrate how EA<sup>3</sup> can be incorporated into the Bayesian decision making framework [85], in which decisions are based on all the available evidence [86]. In this framework, a decision maker is facing a decision problem in which a number of possible acts are at his/her disposal. However, the decision maker is unsure about the state of the world and thus adopts a prior probability function defined over a finite set of possible worlds,  $\Omega$ .

All the available evidence is then used to determine a posterior probability function by conditionalising the prior probability function. In order to represent the decision maker's preferences all pairs of acts and worlds, the possible outcomes, are assigned a utility value in the real numbers. Normatively correct decisions are those which maximise the decision maker's expected utilities, where expectations are calculated with respect to the updated probability function [87–89].

One immediate issue in this framework is that it is hard to calculate a posterior probability function. This issue is normally solved by applying Bayes' Theorem (see the following subsection). Bayes' Theorem is ubiquitous in Bayesian analyses and it is straight-forwardly applied, if the evidence can be taken at face value. In medical inference, where evidence cannot be taken at face value, numerous methodological design features and choices (conscious and subconscious) bear on the information a study provides.

### Bayes' Theorem

Consider a set of exhaustive and mutually exclusively statistical hypotheses  $H_1, \dots, H_m$ , i.e. the states of the world. Let us denote the available evidence by  $\mathcal{E}$ . Bayes' Theorem then allows us to

compute the posterior probability of the hypothesis  $H_h$

$$\begin{aligned}
 \underbrace{P(H_h|\mathcal{E})}_{\text{Posterior}} &= \frac{P(\mathcal{E} \wedge H_h)}{P(\mathcal{E})} = P(H_h) \cdot \frac{P(\mathcal{E} \wedge H_h)}{P(H_h) \cdot P(\mathcal{E})} \\
 &= P(H_h) \cdot \frac{P(\mathcal{E}|H_h)}{P(\mathcal{E} \wedge H_1) + P(\mathcal{E} \wedge H_2) + \dots + P(\mathcal{E} \wedge H_n)} \\
 &= \underbrace{P(H_h)}_{\text{Prior}} \cdot \frac{P(\mathcal{E}|H_h)}{P(\mathcal{E}|H_1)P(H_1) + P(\mathcal{E}|H_2)P(H_2) + \dots + P(\mathcal{E}|H_n)P(H_n)} .
 \end{aligned}$$

So, the posterior probability can be computed from prior probabilities over hypotheses and conditional probabilities. The prior probabilities are provided by the decision maker’s prior beliefs about the state of the world. The conditional probabilities are likelihoods specified by the statistical hypotheses. Hence, computing the posterior probability is a simple exercise in the probability calculus—*under the assumption* that the conditional probabilities are likelihoods specified by statistical models.

In medical inference problems with RWE, the calculations of Bayes’ Theorem remain valid, the statistical models however *do not* specify the relevant likelihoods for RWE. The challenge hence arises to specify these conditional probabilities. We next show how this can be done via an application of EA<sup>3</sup>.

### EA<sup>3</sup> and posterior probabilities of hypotheses based on a single RWE study

How should the posterior probabilities  $Q(\mathcal{E}_1|H)$  look like, given a single study  $\mathcal{E}_1$ ? For starters, the evidence can be taken at face value,  $\vec{A} = 1@k$ , then  $Q(\mathcal{E}_1|H)$  should just be  $P(\mathcal{E}_1|H)$ . If the evidence contains no information whatsoever,  $\vec{A} = 0@k$  and  $v_f = 0$ , then the posterior  $Q(\mathcal{E}_1|H)$  should just equal the prior probability  $P(\mathcal{E}_1)$ , so  $Q(\mathcal{E}_1|H) = P(\mathcal{E}_1)$ . That is, whether  $H$  is true or not, this does not change the probability of obtaining  $\mathcal{E}_1$ . In all other cases, the posterior probability  $Q(\mathcal{E}_1|H)$  should be somewhere between the posterior  $P(\mathcal{E}_1|H)$  and the prior probability  $P(\mathcal{E}_1)$ .

These considerations suggest that  $Q(\mathcal{E}_1|H)$  may be computed as a weighted mean of the posterior and the prior probability:

$$Q(\mathcal{E}_1|H) = v_f \cdot P(\mathcal{E}_1|H) + (1 - v_f)P(\mathcal{E}_1) . \tag{3}$$

Applying Corollary 1 we see that  $Q(\mathcal{E}_1|H)$  is different from the prior, if the posterior and the prior are different and  $v_f > 0$ .

From a theoretical point of view, one may interpret the convex combination in Eq (3) as a Jeffrey update [90]. Under this interpretation,  $v_f$  is interpreted as the probability that the evidence can be taken at face value and  $1 - v_f$  can be interpreted as the probability that the evidence is completely uninformative.

The modified posterior probability of a hypothesis given one available RWE study is

$$\begin{aligned}
 Q(H_h|\mathcal{E}_1) &= \frac{Q(H_h) \cdot Q(\mathcal{E}_1|H_h)}{Q(\mathcal{E}_1|H_1)Q(H_1) + Q(\mathcal{E}_1|H_2)Q(H_2) + \dots + Q(\mathcal{E}_1|H_n)Q(H_n)} \\
 &= \frac{P(H_h) \cdot Q(\mathcal{E}_1|H_h)}{Q(\mathcal{E}_1|H_1)P(H_1) + Q(\mathcal{E}_1|H_2)P(H_2) + \dots + Q(\mathcal{E}_1|H_n)P(H_n)} \tag{4} \\
 &= \frac{P(H_h) \cdot (v_f \cdot P(\mathcal{E}_1|H_h) + (1 - v_f)P(\mathcal{E}_1))}{\sum_{g=1}^n P(H_g) \cdot (v_f \cdot P(\mathcal{E}_1|H_g) + (1 - v_f)P(\mathcal{E}_1))} .
 \end{aligned}$$

### EA<sup>3</sup> and posterior probabilities of hypotheses based on multiple RWE studies

The assumption of a single available RWE study is, of course, rather unrealistic. We now show how to deal with multiple available RWE studies,  $\mathcal{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_s\}$ . We begin by applying EA<sup>3</sup> to all every study individually, thus obtaining  $s$ -many outputs  $v_f^1, \dots, v_f^s$ .

Under the assumption that the studies have been conducted independently from each other, we can generalise Eq (4) as follows:

$$\begin{aligned}
 Q(H_h|\mathcal{E}) &= \frac{P(H_h) \cdot Q(\mathcal{E}|H_h)}{Q(\mathcal{E}|H_1)P(H_1) + Q(\mathcal{E}|H_2)P(H_2) + \dots + Q(\mathcal{E}|H_n)P(H_n)} \\
 &= \frac{P(H_h) \cdot \prod_{r=1}^s (v_f^r \cdot P(\mathcal{E}_r|H_h) + (1 - v_f^r)P(\mathcal{E}_r))}{\sum_{g=1}^n P(H_g) \cdot \prod_{r=1}^s (v_f^r \cdot P(\mathcal{E}_r|H_g) + (1 - v_f^r)P(\mathcal{E}_r))} .
 \end{aligned}$$

### E-Synthesis

*E-Synthesis* is a Bayesian framework developed for determining probabilities of particular drugs causing a specific adverse reaction [78, 91–95]. In order to facilitate the inference from real world data to a causal hypothesis a layer of so-called “indicators” has been inserted between the hypothesis of interest and the data. The indicators have been derived from Hill’s Guidelines [96] and serve the role as (probabilistic) testable consequences of the causal hypothesis. Learning that an indicator is true raises the probability of the causal hypothesis to a degree. For example, learning that there is correlation between a drug and an adverse effect does *not* entail that the drug causes an adverse reaction. Nevertheless, the presence of a correlation does increase our suspicion that there indeed might be a causal relationship between a drug and an adverse event.

Evidence for adverse reactions often emerges spontaneously in form of case reports and suspected adverse reactions are often confirmed only from observational data [97]. Such RWE is at a high risk of bias and hence the RWE needs to be appraised. *E-Synthesis* has been designed to incorporate such appraisals of RWE, making their role explicit by formalising them as variables (previously, these variables have been termed “evidential modulator” variables). The following dimensions of appraisal have been suggested within the *E-Synthesis* framework: sample size, duration of the study, degree of sponsorship bias, degree of adjustment for covariates and the degree of analogy between the study population and the studied population. Randomised studies can also be appraised for how well blinding, randomization and placebo control were implemented.

*E-Synthesis* was originally intended for philosophical applications, however it has also recently been developed for more practical matters. As yet, no suggestion has been made of how to aggregate evidence appraisals and how to incorporate these appraisals for decision making. We next show how this can be done for a specific indicator of causation applying EA<sup>3</sup>. Denoting by © the causal hypothesis of a drug *D* causing a specific adverse drug reaction (ADR) and by Ind an indicator variable, we have for the posterior probability of © for RWE,  $Q(©|\mathcal{E})$ ,

$$\begin{aligned}
 Q(©|\mathcal{E}) &= \frac{Q(© \wedge \mathcal{E})}{Q(\mathcal{E})} \\
 &= \frac{Q(© \wedge \text{Ind} \wedge \mathcal{E}) + Q(© \wedge \overline{\text{Ind}} \wedge \mathcal{E})}{Q(© \wedge \text{Ind} \wedge \mathcal{E}) + Q(© \wedge \overline{\text{Ind}} \wedge \mathcal{E}) + Q(\overline{©} \wedge \text{Ind} \wedge \mathcal{E}) + Q(\overline{©} \wedge \overline{\text{Ind}} \wedge \mathcal{E})} \\
 &= \left[ 1 + \frac{Q(\overline{©} \wedge \text{Ind} \wedge \mathcal{E}) + Q(\overline{©} \wedge \overline{\text{Ind}} \wedge \mathcal{E})}{Q(© \wedge \text{Ind} \wedge \mathcal{E}) + Q(© \wedge \overline{\text{Ind}} \wedge \mathcal{E})} \right]^{-1} \\
 &= \left[ 1 + \frac{Q(\overline{©}) \cdot Q(\text{Ind}|\overline{©}) \cdot Q(\mathcal{E}|\text{Ind}) + Q(\overline{©}) \cdot Q(\overline{\text{Ind}}|\overline{©}) \cdot Q(\mathcal{E}|\overline{\text{Ind}})}{Q(©) \cdot Q(\text{Ind}|©) \cdot Q(\mathcal{E}|\text{Ind}) + Q(©) \cdot Q(\overline{\text{Ind}}|©) \cdot Q(\mathcal{E}|\overline{\text{Ind}})} \right]^{-1} \\
 &= \left[ 1 + \frac{P(\overline{©}) \cdot P(\text{Ind}|\overline{©}) \cdot Q(\mathcal{E}|\text{Ind}) + P(\overline{©}) \cdot P(\overline{\text{Ind}}|\overline{©}) \cdot Q(\mathcal{E}|\overline{\text{Ind}})}{P(©) \cdot P(\text{Ind}|©) \cdot Q(\mathcal{E}|\text{Ind}) + P(©) \cdot P(\overline{\text{Ind}}|©) \cdot Q(\mathcal{E}|\overline{\text{Ind}})} \right]^{-1}.
 \end{aligned}$$

This calculation uses the fact that the causal indicator variable mediates the inference from data to the causal hypothesis © in the technical sense that conditionalisation on it renders the data and © independent.

**Motivating example—coda**

We now return to the motivating example of determining a probability of the causal hypothesis (©) that paracetamol use causes asthma in children. In the *E-Synthesis* approach, the Beasley et al. (2011) [77] study is informative about the “rate of growth” indicator, so Ind = RoG. The posterior probability of © (given only this study) is thus computed as:

$$Q(©|\mathcal{E}_1) = \left[ 1 + \frac{P(\overline{©}) \cdot P(\text{RoG}|\overline{©}) \cdot Q(\mathcal{E}_1|\text{RoG}) + P(\overline{©}) \cdot P(\overline{\text{RoG}}|\overline{©}) \cdot Q(\mathcal{E}_1|\overline{\text{RoG}})}{P(©) \cdot P(\text{RoG}|©) \cdot Q(\mathcal{E}_1|\text{RoG}) + P(©) \cdot P(\overline{\text{RoG}}|©) \cdot Q(\mathcal{E}_1|\overline{\text{RoG}})} \right]^{-1}.$$

Using Eq (3) and the suggested conditional probabilities of  $P(\text{RoG}|\cdot)$  ( $P(\text{RoG}|©) = 26\frac{2}{7}\% \approx 26.3\% > P(\text{RoG}|\overline{©}) = \frac{3}{7}\% \approx 0.4\%$  [78, p. 3]) this becomes

$$\begin{aligned}
 Q(©|\mathcal{E}_1) &= \\
 &= \left[ 1 + \frac{P(\overline{©}) \cdot \left[ \frac{3}{7}\% \cdot (v_f \cdot P(\mathcal{E}_1|\text{RoG}) + (1 - v_f)P(\mathcal{E}_1)) + 99\frac{4}{7}\% \cdot (v_f \cdot P(\mathcal{E}_1|\overline{\text{RoG}}) + (1 - v_f)P(\mathcal{E}_1)) \right]}{P(©) \cdot \left[ 26\frac{2}{7}\% \cdot (v_f \cdot P(\mathcal{E}_1|\text{RoG}) + (1 - v_f)P(\mathcal{E}_1)) + 73\frac{5}{7}\% \cdot (v_f \cdot P(\mathcal{E}_1|\overline{\text{RoG}}) + (1 - v_f)P(\mathcal{E}_1)) \right]} \right]^{-1}.
 \end{aligned}$$

[78, p. 11] gives  $P(\mathcal{E}_1|\text{RoG}) = 0.825$  and  $P(\mathcal{E}_1|\overline{\text{RoG}}) = 0$  and so

$$\begin{aligned}
 Q(\odot|\mathcal{E}_1) &= \left[ 1 + \frac{P(\overline{\odot}) \cdot \left[ \frac{3}{7}\% \cdot (v_f \cdot 0.825 + (1 - v_f)P(\mathcal{E}_1)) + 99\frac{4}{7}\% \cdot (1 - v_f)P(\mathcal{E}_1) \right]}{P(\odot) \cdot \left[ 26\frac{2}{7}\% \cdot (v_f \cdot 0.825 + (1 - v_f)P(\mathcal{E}_1)) + 73\frac{5}{7}\% \cdot (1 - v_f)P(\mathcal{E}_1) \right]} \right]^{-1} \\
 &= \left[ 1 + \frac{P(\overline{\odot})}{P(\odot)} \cdot \frac{\frac{3}{7}\% \cdot v_f \cdot 0.825 + (1 - v_f)P(\mathcal{E}_1)}{26\frac{2}{7}\% \cdot v_f \cdot 0.825 + (1 - v_f)P(\mathcal{E}_1)} \right]^{-1}.
 \end{aligned}$$

The posterior probability of  $\odot$  given by De Pretis et al. (2019) [78] is instead:

$$\begin{aligned}
 P(\odot|\mathcal{E}_1) &= \left[ 1 + \frac{P(\overline{\odot})}{P(\odot)} \cdot \frac{P(\text{RoG}|\overline{\odot})}{P(\text{RoG}|\odot)} \right]^{-1} = \left[ 1 + \frac{P(\text{RoG} \wedge \overline{\odot})}{P(\text{RoG} \wedge \odot)} \right]^{-1} \\
 &= \frac{P(\text{RoG} \wedge \odot)}{P(\text{RoG} \wedge \odot) + P(\text{RoG} \wedge \overline{\odot})} = P(\odot|\text{RoG}).
 \end{aligned}$$

We note that in the model of De Pretis et al. (2019) [78] this single study is *conclusive evidence* that RoG holds, i.e. there does exist a strongly increasing dose-response relationship between paracetamol use in children and severe onset of asthma. This probability is

$$P(\odot|\mathcal{E}_1) = P(\odot|\text{RoG}) = \left[ 1 + \frac{P(\overline{\odot})}{P(\odot)} \cdot \frac{\frac{3}{7}\%}{26\frac{2}{7}\%} \right]^{-1} = \left[ 1 + \frac{P(\overline{\odot})}{P(\odot)} \cdot \frac{3}{184} \right]^{-1}.$$

See Figs 4 and 5 for comparisons of  $P(\odot|\mathcal{E}_1)$  [78] and  $Q(\odot|\mathcal{E}_1)$  (EA<sup>3</sup>).

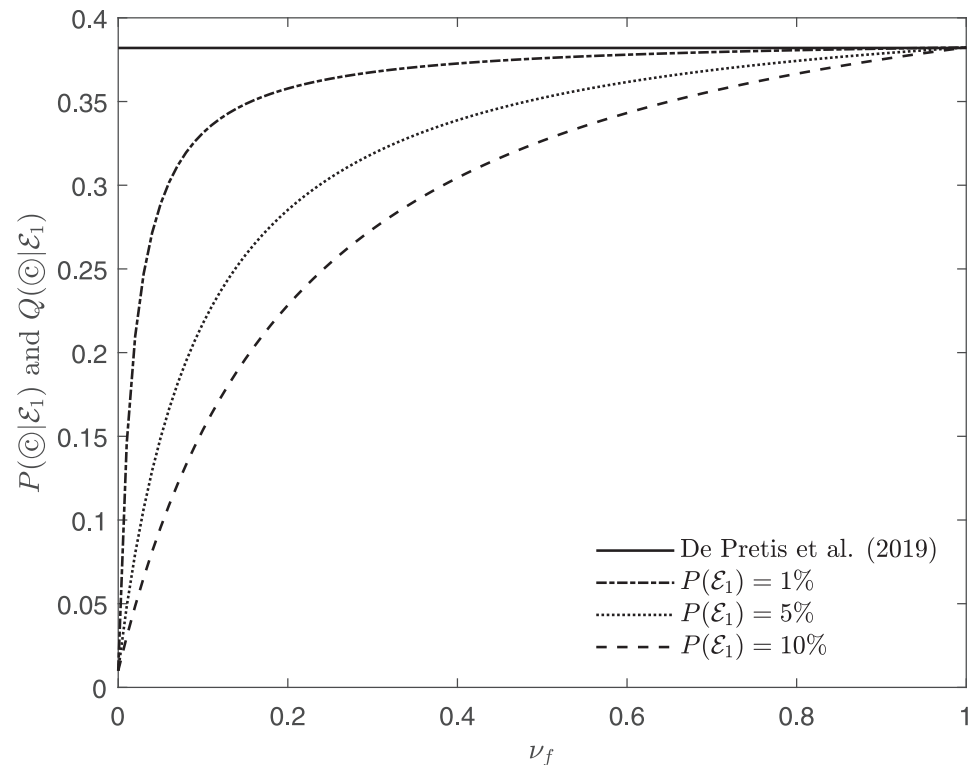
### Discussion

In this article, we presented an algorithm to support the assessment of the inferential strength of RWE in order to make sound decisions. We proceeded by considering different dimensions of appraisal and then moved on to aggregate multiple appraisals according to the different dimensions into an aggregate. Subsequently, we showed how such an aggregate can be used within a Bayesian decision making framework. Our formal approach carries forward evidence appraisals, incorporates them into an overall appraisal of the evidence and integrates it into decision making [35]. It also enables sensitivity analyses of these appraisals via variation of appraisals, variations of  $\vec{A}$ , as well as sensitivity analyses of the ranking, variations of  $\vec{R}$ , and the cautiousness parameter  $\beta$ . Furthermore, our approach is transparent, reproducible and scientifically defensible, thus satisfying the desiderata suggested by the US Environmental Protection Agency [35, p. 79].

While our formal aggregation approach is motivated by the need to appraise RWE for medical inference, the developed algorithm is, in principle, applicable to other aggregation problems, too. Whether it is suitable to a particular problem depends on particular circumstances.

Our approach is limited by the assumptions we made, e.g. we assumed that the dimensions of appraisal are independent of each other and that rankings and appraisals can be represented numerically. If at least one of our assumptions fails to hold in an application, then the theoretical considerations made here might not apply. These limitations may be overcome by applications of multi-criteria decision making methodology [98].

In future work, we aim to determine empirically supported dimensions for evidence appraisal, calibrate ranking schemes and determine (normatively and/or descriptively) appropriate values of the  $\beta$ -parameter in order to assess the validity and reliability of EA<sup>3</sup> based on



**Fig 4. Posterior probability of the causal hypothesis ©, considering Beasley et al. (2011) [77] as evidence  $\mathcal{E}_1$ , and computed in agreement with De Pretis et al. (2019) [78] (solid line) and EA<sup>3</sup> (dash-dot, dotted and dashed lines). For EA<sup>3</sup>, different lines represent different priors  $P(\mathcal{E}_1)$ , whereas the prior  $P(\textcircled{c})$  is always set to 1%. All curves agree for  $\nu_f = 1$  where De Pretis et al. (2019) [78] becomes a special case of EA<sup>3</sup>.**

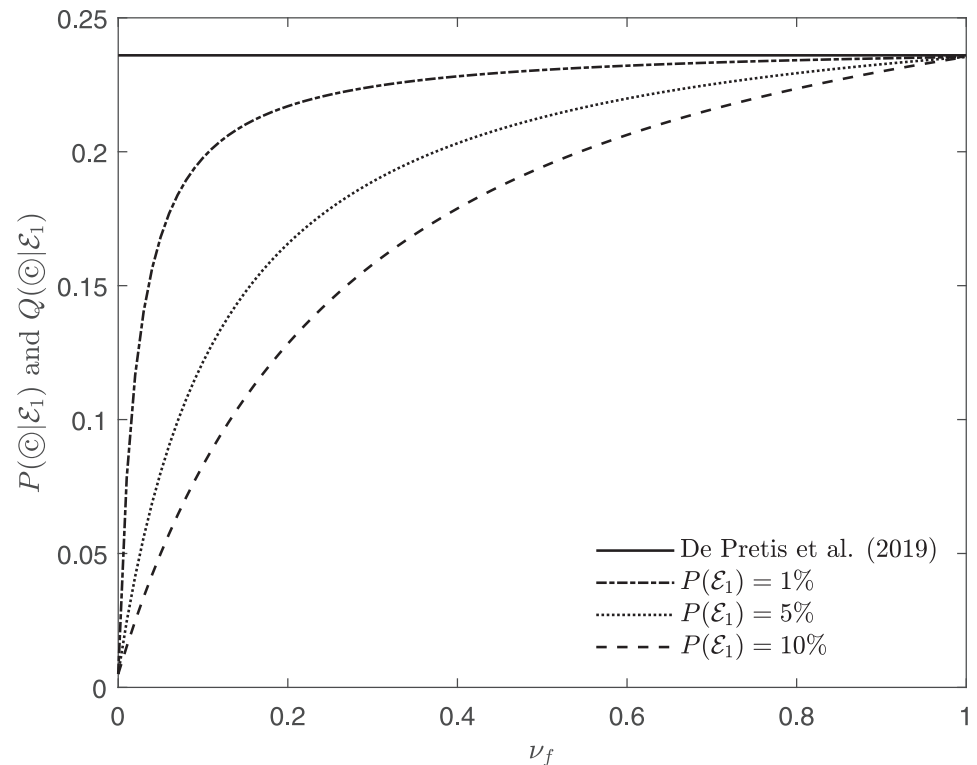
<https://doi.org/10.1371/journal.pone.0253057.g004>

actual data [35]. The  $\beta$ -parameter which represents cautiousness reflects risk attitudes which can differ from user to user and from application to application.

Furthermore, EA<sup>3</sup> reflects the position of a single agent (or of a unanimous committee). In reality, drug approval or withdrawal decisions are a group effort involving experts from different areas (toxicologists, pharmacists, clinicians, statisticians as well as patient representatives [99]), which have different risk attitudes (different  $\beta$ ), different appraisals (different  $\vec{A}$ ) and different rankings (different  $\vec{R}$ ). We thus plan to integrate EA<sup>3</sup> into a multi-agent framework which represents different (risk) attitudes, preferences and areas of expertise of stakeholders in drug (un-)safety assessments.

We expect the assessment and use of RWE for medical inference to continue to grow in coming years, drawing on scientific fields in which there are, by the very nature of the investigation, (next to) no randomised studies. For example, in macroeconomics we cannot simply randomly assign countries into different trial arms to learn about the disputed causal relationships between minimum wages and employment [100] and in nutrition science it is not possible to randomise people into drinkers of red wine and non drinkers for a trial lasting several years to learn about the hypothesised causal influences of red wine on health and well-being [101]. Similarly, in pharmacovigilance ADRs may take too long to manifest (years of treatment with olanzapine cause tardive dyskinesia [102]) or be too rare yet fatal (in some cases, 1 fatality in every 10,000 patients [103]) to be detected by RCTs. We think that the use of RWE for pharmacovigilance and medical inference more widely is an area holding great promise despite





**Fig 5.** Similarly to Fig 4, the graph pictures the posterior probability of the causal hypothesis  $\textcircled{S}$ , considering Beasley et al. (2011) [77] as evidence  $\mathcal{E}_1$ , and computed in agreement with De Pretis et al. (2019) [78] (solid line) and EA<sup>3</sup> (dash-dot, dotted and dashed lines). In this plot, the prior  $P(\textcircled{S})$  is set to 0.5%.

<https://doi.org/10.1371/journal.pone.0253057.g005>

justified worries about biases and confounding. The development and application of RWE appraisal methods hence seems to become even more important in the future.

## Supporting information

**S1 Text.**  
(TXT)

## Acknowledgments

The authors are grateful to Bolin Gao (University of Toronto, Canada) for discussing his recent works on the softmax function. The authors would also like to thank Martin Posch (Medical University of Vienna, Austria) for helpful suggestions to improve the manuscript.

## Author Contributions

**Conceptualization:** Francesco De Pretis, Jürgen Landes.

**Formal analysis:** Francesco De Pretis, Jürgen Landes.

**Funding acquisition:** Jürgen Landes.

**Investigation:** Francesco De Pretis, Jürgen Landes.

**Methodology:** Francesco De Pretis, Jürgen Landes.

**Validation:** Francesco De Pretis, Jürgen Landes.

**Writing – original draft:** Francesco De Pretis, Jürgen Landes.

**Writing – review & editing:** Francesco De Pretis, Jürgen Landes.

## References

1. Sherman RE, Anderson SA, Pan GJD, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence—What Is It and What Can It Tell Us? *New England Journal of Medicine*. 2016; 375(23):2293–2297. <https://doi.org/10.1056/NEJMsb1609216> PMID: 27959688
2. Audeh B, Calvier FE, Bellet F, Beyens MN, Pariente A, Louet ALL, et al. Pharmacology and social media: Potentials and biases of web forums for drug mention analysis—case study of France. *Health Informatics Journal*. 2019; 26(2):1253–1272. <https://doi.org/10.1177/1460458219865128> PMID: 31566468
3. Bolisliis WR, Fay M, Kühler TC. Use of Real-world Data for New Drug Applications and Line Extensions. *Clinical Therapeutics*. 2020; 42(5):926–938. <https://doi.org/10.1016/j.clinthera.2020.03.006> PMID: 32340916
4. Camm A, Coleman C, Tamayo C, Beyer-Westendorf J. Rivaroxaban real-world evidence: Validating safety and effectiveness in clinical practice. *Thrombosis and Haemostasis*. 2016; 116(S 02):S13–S23. <https://doi.org/10.1160/TH16-06-0485> PMID: 27623681
5. Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clinical Pharmacology & Therapeutics*. 2019; 106(1):36–39. <https://doi.org/10.1002/cpt.1426> PMID: 30970161
6. Andre EB, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiology and Drug Safety*. 2019; 29(10):1201–1212. <https://doi.org/10.1002/pds.4932>
7. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making. *JAMA*. 2017; 318(8):703. <https://doi.org/10.1001/jama.2017.9991> PMID: 28715550
8. Miksad RA, Abernethy AP. Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. *Clinical Pharmacology & Therapeutics*. 2017; 103(2):202–205. <https://doi.org/10.1002/cpt.946> PMID: 29214638
9. Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *JAMA*. 2018; 320(9):867. <https://doi.org/10.1001/jama.2018.10136> PMID: 30105359
10. Katkade VB, Sanders KN, Zou KH. Real world data: an opportunity to supplement existing evidence for the use of long-established medicines in health care decision making. *Journal of Multidisciplinary Healthcare*. 2018; Volume 11:295–304. <https://doi.org/10.2147/JMDH.S160029> PMID: 29997436
11. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *eGEMs*. 2016; 4(1):18. <https://doi.org/10.13063/2327-9214.1244> PMID: 27713905
12. Pan J, Lin DY. Guest editors' note on special issue on real-world experience and randomized clinical trials. *Journal of Biopharmaceutical Statistics*. 2019; 29(4):579–579. <https://doi.org/10.1080/10543406.2019.1643360> PMID: 31362612
13. Kim HS, Lee S, Kim JH. Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. *Journal of Korean Medical Science*. 2018; 33(34). <https://doi.org/10.3346/jkms.2018.33.e213> PMID: 30127705
14. Forstag EH, Kahn B, Gee AW, Shore C, editors. *Examining the Impact of Real-World Evidence on Medical Product Development*. Washington DC, USA: The National Academies Press; 2019. Available from: <https://doi.org/10.17226/25352>.
15. Gould AL. Substantial evidence of effect. *Journal of Biopharmaceutical Statistics*. 2002; 12(1):53–77. <https://doi.org/10.1081/BIP-120005740> PMID: 12146720
16. Martini N, Trifirò G, Capuano A, Corrao G, Corrao G, Racagni G, et al. Expert opinion on Real World Evidence RWE in drug development and usage. *Pharmadvances*. 2020; 02(02). <https://doi.org/10.36118/pharmadvances.01.2020.01>
17. Rome BN, Avorn J. Drug Evaluation during the Covid-19 Pandemic. *New England Journal of Medicine*. 2020; 382(24):2282–2284. <https://doi.org/10.1056/NEJMp2009457> PMID: 32289216

18. Camm AJ, Fox KAA. Strengths and weaknesses of ‘real-world’ studies involving non-vitamin K antagonist oral anticoagulants. *Open Heart*. 2018; 5(1):e000788. <https://doi.org/10.1136/openhrt-2018-000788> PMID: 29713485
19. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Network Open*. 2019; 2(10):e1912869. <https://doi.org/10.1001/jamanetworkopen.2019.12869> PMID: 31596493
20. Evans K. Real World Evidence: Can We Really Expect It to Have Much Influence? *Drugs—Real World Outcomes*. 2019; 6(2):43–45. <https://doi.org/10.1007/s40801-019-0155-3> PMID: 31016548
21. Raphael MJ, Gyawali B, Booth CM. Real-world evidence and regulatory drug approval. *Nature Reviews Clinical Oncology*. 2020; 17(5):271–272. <https://doi.org/10.1038/s41571-020-0345-7> PMID: 32112057
22. Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. *New England Journal of Medicine*. 2020; 382(7):674–678. <https://doi.org/10.1056/NEJMs1901642> PMID: 32053307
23. Ryan P. Statistical challenges in systematic evidence generation through analysis of observational healthcare data networks. *Statistical Methods in Medical Research*. 2013; 22(1):3–6. <https://doi.org/10.1177/0962280211403601> PMID: 23439684
24. Verde PE. A bias-corrected meta-analysis model for combining, studies of different types and quality. *Biometrical Journal*. 2020; 63(2):406–422. <https://doi.org/10.1002/bimj.201900376> PMID: 32996196
25. Efthimiou O, Mavridis D, Debray TPA, Samara M, Belger M, Siontis GCM, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*. 2017; 36(8):1210–1226. <https://doi.org/10.1002/sim.7223> PMID: 28083901
26. Ferguson J, Alvarez-Iglesias A, Newell J, Hinde J, Donnell MO. Joint incorporation of randomised and observational evidence in estimating treatment effects. *Statistical Methods in Medical Research*. 2017; 28(1):235–247. <https://doi.org/10.1177/0962280217720854> PMID: 28745132
27. Nguyen TL, Collins GS, Pellegrini F, Moons KGM, Debray TPA. On the aggregation of published prognostic scores for causal inference in observational studies. *Statistics in Medicine*. 2020; 39(10):1440–1457. <https://doi.org/10.1002/sim.8489> PMID: 32022311
28. Wang C, Rosner GL. A Bayesian nonparametric causal inference model for synthesizing randomized clinical trial and real-world evidence. *Statistics in Medicine*. 2019; 38(14):2573–2588. <https://doi.org/10.1002/sim.8134> PMID: 30883861
29. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is “quality of evidence” and why is it important to clinicians? *BMJ*. 2008; 336(7651):995–998. <https://doi.org/10.1136/bmj.39490.551019.BE> PMID: 18456631
30. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008; 336(7650):924–926. <https://doi.org/10.1136/bmj.39489.470347.AD> PMID: 18436948
31. Ansari MT, Tsertsvadze A, Moher D. Grading Quality of Evidence and Strength of Recommendations: A Perspective. *PLoS Medicine*. 2009; 6(9):e1000151. <https://doi.org/10.1371/journal.pmed.1000151> PMID: 19753108
32. Stegenga J. Down with the Hierarchies. *Topoi*. 2013; 33(2):313–322. <https://doi.org/10.1007/s11245-013-9189-4>
33. Landes J. An Evidence-Hierarchical Decision Aid for Ranking in Evidence-Based Medicine. In: Osimani B, La Caze A, editors. *Uncertainty in Pharmacology: Epistemology, Methods and Decisions*. vol. 338 of Boston Studies in Philosophy of Science. Cham, Switzerland: Springer; 2020. p. 231–259. Available from: [https://doi.org/10.1007/978-3-030-29179-2\\_11](https://doi.org/10.1007/978-3-030-29179-2_11).
34. Mercuri M, Baigrie B, Upshur REG. Going from evidence to recommendations: Can GRADE get us there? *Journal of Evaluation in Clinical Practice*. 2018; 24(5):1232–1239. <https://doi.org/10.1111/jep.12857> PMID: 29314554
35. National Research Council. Review of EPA’s integrated risk information system (IRIS) process. Washington DC, USA: The National Academies Press; 2014. Available from: <https://doi.org/10.17226/18764>.
36. Saitta L, Giordana A, Cornuejols A. Statistical physics and machine learning. In: *Phase Transitions in Machine Learning*. Cambridge, UK: Cambridge University Press; 2011. p. 140–167. Available from: <https://doi.org/10.1017/CBO9780511975509.009>.
37. Bahri Y, Kadmon J, Pennington J, Schoenholz SS, Sohl-Dickstein J, Ganguli S. Statistical Mechanics of Deep Learning. *Annual Review of Condensed Matter Physics*. 2020; 11(1):501–528. <https://doi.org/10.1146/annurev-conmatphys-031119-050745>

38. Gabrié M. Mean-field inference methods for neural networks. *Journal of Physics A: Mathematical and Theoretical*. 2020; 53(22):223002. <https://doi.org/10.1088/1751-8121/ab7f65>
39. Stewart GB, Higgins JPT, Schönemann H, Meader N. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 1. *PLOS ONE*. 2015; 10(4):e0114497. <https://doi.org/10.1371/journal.pone.0114497> PMID: 25837450
40. Llewellyn A, Whittington C, Stewart G, Higgins JP, Meader N. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLOS ONE*. 2015; 10(12):e0123511. <https://doi.org/10.1371/journal.pone.0123511> PMID: 26716874
41. Balian R. The Boltzmann-Gibbs Distribution. In: *From Microphysics to Macrophysics*. Heidelberg, Germany: Springer; 1991. p. 141–180. Available from: [https://doi.org/10.1007/978-3-540-45475-5\\_5](https://doi.org/10.1007/978-3-540-45475-5_5).
42. Landau LD, Lifshitz EM. The Gibbs Distribution. In: *Statistical Physics*. Oxford, UK: Butterworth-Heinemann; 1980. p. 79–110. Available from: <https://doi.org/10.1016/b978-0-08-057046-4.50010-5>.
43. Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984; PAMI-6(6):721–741. <https://doi.org/10.1109/TPAMI.1984.4767596> PMID: 22499653
44. Ruiz F, Titsias M, Dieng AB, Blei D. Augment and Reduce: Stochastic Inference for Large Categorical Distributions. In: Dy J, Krause A, editors. *Proceedings of the 35th International Conference on Machine Learning*. vol. 80 of *Proceedings of Machine Learning Research*. Stockholm, Sweden: PMLR; 2018. p. 4403–4412. Available from: <http://proceedings.mlr.press/v80/ruiz18a.html>.
45. Siddique AA, Schnitzer ME, Bahamyirou A, Wang G, Holtz TH, Migliori GB, et al. Causal inference with multiple concurrent medications: A comparison of methods and an application in multidrug-resistant tuberculosis. *Statistical Methods in Medical Research*. 2018; 28(12):3534–3549. <https://doi.org/10.1177/0962280218808817> PMID: 30381005
46. Wolfe J, Jin X, Bahr T, Holzer N. Application of softmax regression and its validation for spectral-based land cover mapping. *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2017; XLII-1/W1:455–459. <https://doi.org/10.5194/isprs-archives-XLII-1-W1-455-2017>
47. Magnusson BP, Schmidli H, Rouyrre N, Scharfstein DO. Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by postrandomization event occurrence. *Statistics in Medicine*. 2019; 38(23):4761–4771. <https://doi.org/10.1002/sim.8333> PMID: 31386219
48. McCurdy S, Molinaro A, Pachter L. Factor analysis for survival time prediction with informative censoring and diverse covariates. *Statistics in Medicine*. 2019; 38(20):3719–3732. <https://doi.org/10.1002/sim.8151> PMID: 31162708
49. Tu C, Koh WY. Comparison of balancing scores using the ANCOVA approach for estimating average treatment effect: a simulation study. *Journal of Biopharmaceutical Statistics*. 2018; 29(3):508–515. <https://doi.org/10.1080/10543406.2018.1559855> PMID: 30561245
50. Bishop CM. *Pattern Recognition and Machine Learning*. Information Science and Statistics series. New York NY, USA: Springer; 2006.
51. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Adaptive Computation and Machine Learning series. Cambridge MA, USA: MIT Press; 2016.
52. Bridle JS. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In: Fogelman Soulié F, Héroult J, editors. *Neurocomputing. Algorithms, Architectures and Applications*. Heidelberg, Germany: Springer; 1990. p. 227–236. Available from: [https://doi.org/10.1007/978-3-642-76153-9\\_28](https://doi.org/10.1007/978-3-642-76153-9_28).
53. Li J, Gao M, D'Agostino R. Evaluating classification accuracy for modern learning approaches. *Statistics in Medicine*. 2019; 38(13):2477–2503. <https://doi.org/10.1002/sim.8103> PMID: 30701585
54. Somanchi S, Neill DB, Parwani AV. Discovering anomalous patterns in large digital pathology images. *Statistics in Medicine*. 2018; 37(25):3599–3615. <https://doi.org/10.1002/sim.7828> PMID: 29900578
55. Chopra P, Yadav SK. Restricted Boltzmann machine and softmax regression for fault detection and classification. *Complex & Intelligent Systems*. 2017; 4(1):67–77. <https://doi.org/10.1007/s40747-017-0054-8>
56. Kwon Y, Won JH, Kim BJ, Paik MC. Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis*. 2020; 142:106816. <https://doi.org/10.1016/j.csda.2019.106816>
57. Qi X, Wang T, Liu J. Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision. In: *2017 Second International Conference on Mechanical, Control and Computer Engineering*

- (ICMCCE), vol. 1. Harbin, China: IEEE; 2017. p. 151–155. Available from: <https://doi.org/10.1109/icmcce.2017.49>.
58. Saito S, Yamashita T, Aoki Y. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. *Electronic Imaging*. 2016; 2016(10):1–9. <https://doi.org/10.2352/ISSN.2470-1173.2016.10.ROBVIS-392>
  59. Gao B, Pavel L. On Passivity, Reinforcement Learning and Higher-Order Learning in Multi-Agent Finite Games. *IEEE Transactions on Automatic Control*. 2020; p. 1–16. <https://doi.org/10.1109/TAC.2020.3045094>
  60. Pan L, Cai Q, Meng Q, Chen W, Huang L. Reinforcement Learning with Dynamic Boltzmann Softmax Updates. In: Bessiere C, editor. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. Yokohama, Japan: International Joint Conferences on Artificial Intelligence Organization; 2020. p. 1992–1998. Available from: <https://doi.org/10.24963/ijcai.2020/276>.
  61. Gao B, Pavel L. On Passivity and Reinforcement Learning in Finite Games. In: 2018 IEEE Conference on Decision and Control (CDC). Miami FL, USA: IEEE; 2018. p. 340–345. Available from: <https://doi.org/10.1109/cdc.2018.8619157>.
  62. Kouretas I, Paliouras V. Hardware Implementation of a Softmax-Like Function for Deep Learning. *Technologies*. 2020; 8(3):46. <https://doi.org/10.3390/technologies8030046>
  63. Kim S, Asadi K, Littman ML, Konidaris G. Removing the Target Network from Deep Q-Networks with the Mellowmax Operator. In: N Agmon EE M E Taylor, Veloso M, editors. *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019)*. Montreal, Canada: International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS); 2019. p. 2060–2062. Available from: <http://www.ifaamas.org/Proceedings/aamas2019/pdfs/p2060.pdf>.
  64. Jain V, Doshi P, Banerjee B. Model-Free IRL Using Maximum Likelihood Estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019;33:3951–3958.
  65. Laha A, Chemmengath SA, Agrawal P, Khapra M, Sankaranarayanan K, Ramaswamy HG. On Controllable Sparse Alternatives to Softmax. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Information Processing Systems 31*. Red Hook NY, USA: Curran Associates, Inc.; 2019. p. 6422–6432. Available from: <http://papers.nips.cc/paper/7878-on-controllable-sparse-alternatives-to-softmax.pdf>.
  66. Asadi K, Littman ML. An Alternative Softmax Operator for Reinforcement Learning. In: Precup D, Teh YW, editors. *Proceedings of the 34th International Conference on Machine Learning*. vol. 70 of *Proceedings of Machine Learning Research*. Sydney, Australia: PMLR; 2017. p. 243–252. Available from: <http://proceedings.mlr.press/v70/asadi17a.html>.
  67. Tokic M, Palm G. Value-Difference Based Exploration: Adaptive Control between Epsilon-Greedy and Softmax. In: Bach J, Edelkamp S, editors. *KI 2011: Advances in Artificial Intelligence*. Heidelberg, Germany: Springer; 2011. p. 335–346. Available from: [https://doi.org/10.1007/978-3-642-24455-1\\_33](https://doi.org/10.1007/978-3-642-24455-1_33).
  68. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? *The Lancet*. 2015; 386(9998):1075–1085.
  69. Sherbath M, Furuya-Kanamori L, Nader JD, Thalib L. Risk of wheezing and asthma exacerbation in children treated with paracetamol versus ibuprofen: a systematic review and meta-analysis of randomised controlled trials. *BMC Pulmonary Medicine*. 2020; 20(1). <https://doi.org/10.1186/s12890-020-1102-5> PMID: 32293369
  70. McBride JT. The Association of Acetaminophen and Asthma Prevalence and Severity. *Pediatrics*. 2011; 128(6):1181–1185. <https://doi.org/10.1542/peds.2011-1106> PMID: 22065272
  71. Lesko SM, Mitchell AA. The Safety of Acetaminophen and Ibuprofen Among Children Younger Than Two Years Old. *Pediatrics*. 1999; 104(4):e39–e39. <https://doi.org/10.1542/peds.104.4.e39> PMID: 10506264
  72. Newson RB, Shaheen SO, Chinn S, Burney PGJ. Paracetamol sales and atopic disease in children and adults: an ecological analysis. *European Respiratory Journal*. 2000; 16(5):817–823. <https://doi.org/10.1183/09031936.00.16581700> PMID: 11153577
  73. Lesko SM, Louik C, Vezina RM, Mitchell AA. Asthma Morbidity After the Short-Term Use of Ibuprofen in Children. *Pediatrics*. 2002; 109(2):e20–e20. <https://doi.org/10.1542/peds.109.2.e20> PMID: 11826230
  74. Shaheen SO, Newson RB, Sherriff A, Henderson AJ, Heron JE, Burney PGJ, et al. Paracetamol use in pregnancy and wheezing in early childhood. *Thorax*. 2002; 57(11):958–963. <https://doi.org/10.1136/thorax.57.11.958> PMID: 12403878
  75. Karimi M, Mirzaei M, Ahmadi MH. Acetaminophen Use and the Symptoms of Asthma, Allergic Rhinitis and Eczema in Children. *Iranian Journal of Allergy, Asthma and Immunology*. 2006; 5(2):63–67. PMID: 17237578

76. Amberbir A, Medhin G, Alem A, Britton J, Davey G, Venn A. The Role of Acetaminophen and Geohelminth Infection on the Incidence of Wheeze and Eczema. *American Journal of Respiratory and Critical Care Medicine*. 2011; 183(2):165–170. <https://doi.org/10.1164/rccm.201006-0989OC> PMID: 20935107
77. Beasley RW, Clayton TO, Crane J, Lai CKW, Montefort SR, von Mutius E, et al. Acetaminophen Use and Risk of Asthma, Rhinoconjunctivitis, and Eczema in Adolescents. *American Journal of Respiratory and Critical Care Medicine*. 2011; 183(2):171–178. <https://doi.org/10.1164/rccm.201005-0757OC> PMID: 20709817
78. De Pretis F, Landes J, Osimani B. E-Synthesis: A Bayesian Framework for Causal Assessment in Pharmacosurveillance. *Frontiers in Pharmacology*. 2019; 10:1317. <https://doi.org/10.3389/fphar.2019.01317> PMID: 31920632
79. Dudbridge F. Criteria for evaluating risk prediction of multiple outcomes. *Statistical Methods in Medical Research*. 2020; 29(12):3492–3510. <https://doi.org/10.1177/0962280220929039> PMID: 32594841
80. Cao X, Deng Y. A New Geometric Mean FMEA Method Based on Information Quality. *IEEE Access*. 2019; 7:95547–95554. <https://doi.org/10.1109/ACCESS.2019.2928581>
81. Kim MJ, Kang DK, Kim HB. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*. 2015; 42(3):1074–1082. <https://doi.org/10.1016/j.eswa.2014.08.025>
82. Baldi P, Sadowski P. The dropout learning algorithm. *Artificial Intelligence*. 2014; 210:78–122. <https://doi.org/10.1016/j.artint.2014.02.004> PMID: 24771879
83. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *Journal of Clinical Epidemiology*. 2013; 66(2):151–157. <https://doi.org/10.1016/j.jclinepi.2012.01.006> PMID: 22542023
84. Verde PE. The hierarchical metaregression approach and learning from clinical evidence. *Biometrical Journal*. 2019; 61(3):535–557. <https://doi.org/10.1002/bimj.201700266> PMID: 30600534
85. Savage LJ. *The Foundations of Statistics*. Chelmsford MA, USA: Courier Corporation; 2012.
86. Carnap R. On the Application of Inductive Logic. *Philosophy and Phenomenological Research*. 1947; 8(1):133–148. <https://doi.org/10.2307/2102920>
87. Howson C, Urbach P. *Scientific Reasoning*. 3rd ed. Chicago IL, USA: Open Court; 2006.
88. Sprenger J, Hartmann S. *Bayesian Philosophy of Science*. Oxford, UK: Oxford University Press; 2019.
89. Williamson J. *In Defence of Objective Bayesianism*. Oxford, UK: Oxford University Press; 2010.
90. Jacobs B. The Mathematics of Changing One's Mind, via Jeffrey's or via Pearl's Update Rule. *Journal of Artificial Intelligence Research*. 2019; 65:783–806. <https://doi.org/10.1613/jair.1.11349>
91. Abdin Y, Auker-Howlett DJ, Landes J, Mulla G, Jacob C, Osimani B. Reviewing the Mechanistic Evidence Assessors E-Synthesis and EBM+: A Case Study of Amoxicillin and Drug Reaction with Eosinophilia and Systemic Symptoms (DRESS). *Current Pharmaceutical Design*. 2019; 25(16):1866–1880. <https://doi.org/10.2174/1381612825666190628160603> PMID: 31264541
92. Landes J, Osimani B, Poellinger R. Epistemology of Causal Inference in Pharmacology. *European Journal for Philosophy of Science*. 2018; 8:3–49. <https://doi.org/10.1007/s13194-017-0169-1>
93. De Pretis F, Osimani B. New Insights in Computational Methods for Pharmacovigilance: E-Synthesis, a Bayesian Framework for Causal Assessment. *International Journal of Environmental Research and Public Health*. 2019; 16(12):2221. <https://doi.org/10.3390/ijerph16122221> PMID: 31238543
94. De Pretis F, Landes J, Peden W, Osimani B. Pharmacovigilance as personalized evidence. In: Beneduce C, Bertolaso M, editors. *Personalized Medicine in the Making. Philosophical Perspectives from Biology to Healthcare*. Cham, Switzerland: Springer; 2021. p. 1–19. Available from: <https://doi.org/10.1007/978-3-030-74804-3>.
95. De Pretis F, Landes J, Peden W. Artificial intelligence methods for a Bayesian epistemology-powered evidence evaluation. *Journal of Evaluation in Clinical Practice*. 2021; 27(3):504–512. <https://doi.org/10.1111/jep.13542> PMID: 33569874
96. Hill AB. The environment and disease: association or causation? *Journal of the Royal Society of Medicine*. 2015; 108(1):32–37. <https://doi.org/10.1177/0141076814562718> PMID: 25572993
97. Onakpoya IJ, Heneghan CJ, Aronson JK. Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. *Critical Reviews in Toxicology*. 2016; 46:477–489. <https://doi.org/10.3109/10408444.2016.1149452> PMID: 26941185
98. Triantaphyllou E. *Multi-criteria decision making methods: A comparative Study*. Dordrecht, The Netherlands: Kluwer; 2000. Available from: <https://doi.org/10.1007/978-1-4757-3157-6>.

99. Ciociola AA, Karlstadt RG, Pambianco DJ, Woods KL, Ehrenpreis ED. The Food and Drug Administration Advisory Committees and Panels: How They Are Applied to the Drug Regulatory Process. *American Journal of Gastroenterology*. 2014; 109(10):1508–1512. <https://doi.org/10.1038/ajg.2014.85> PMID: 25001252
100. Reiss J. *Philosophy of Economics*. New York NY, USA: Routledge; 2013.
101. Jukola S. On the evidentiary standards for nutrition advice. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. 2019; 73:1–9. <https://doi.org/10.1016/j.shpsc.2018.05.007> PMID: 29866402
102. Beasley CM, Dellva MA, Tamura RN, Morgenstern H, Glazer WM, Ferguson K, et al. Randomised double-blind comparison of the incidence of tardive dyskinesia in patients with schizophrenia during long-term treatment with olanzapine or haloperidol. *British Journal of Psychiatry*. 1999; 174(1):23–30. <https://doi.org/10.1192/bjp.174.1.23>
103. Food and Drug Administration. Drug Induced Liver Injury: Premarketing Clinical Evaluation—Guidance for Industry; 2009. Web page. Available from: <http://www.fda.gov/downloads/Drugs/Guidance/UCM174090.pdf>.