DOI: 10.1111/risa.13871

ORIGINAL ARTICLE

A suggestion for the quantification of precise and bounded probability to quantify epistemic uncertainty in scientific assessments

Ivette Raices Cruz[1,2] | **Matthias C. M. Troffaes**[3] | **Ullrika Sahlin**[1]

[1] Centre for Environmental and Climate Science, Lund University, Lund, Sweden

[2] Department of Biology, Lund University, Lund, Sweden

[3] Department of Mathematical Sciences, Durham University, Durham, UK

Correspondence
Ivette Raices Cruz, Centre for Environmental and Climate Science, Lund University, Sölvegatan 37, Lund, Sweden.
Email: ivette.raices_cruz@cec.lu.se

Funding information
Swedish Research Council FORMAS through the project "Scaling up uncertain environmental evidence", Grant/Award Number: (219-2013-1271)

Abstract

An honest communication of uncertainty about quantities of interest enhances transparency in scientific assessments. To support this communication, risk assessors should choose appropriate ways to evaluate and characterize epistemic uncertainty. A full treatment of uncertainty requires methods that distinguish aleatory from epistemic uncertainty. Quantitative expressions for epistemic uncertainty are advantageous in scientific assessments because they are nonambiguous and enable individual uncertainties to be characterized and combined in a systematic way. Since 2019, the European Food Safety Authority (EFSA) recommends assessors to express epistemic uncertainty in conclusions of scientific assessments quantitatively by subjective probability. A subjective probability can be used to represent an expert judgment, which may or may not be updated using Bayes's rule to integrate evidence available for the assessment and could be either precise or approximate. Approximate (or bounded) probabilities may be enough for decision making and allow experts to reach agreement on certainty when they struggle to specify precise subjective probabilities. The difference between the lower and upper bound on a subjective probability can also be used to reflect someone's strength of knowledge. In this article, we demonstrate how to quantify uncertainty by bounded probability, and explicitly distinguish between epistemic and aleatory uncertainty, by means of robust Bayesian analysis, including standard Bayesian analysis through precise probability as a special case. For illustration, the two analyses are applied to an intake assessment.

KEYWORDS

Bayesian analysis, scientific assessment, subjective probability, uncertainty analysis, uncertainty communication

1 | INTRODUCTION

1.1 | Background and Aim

Uncertainty arising from limitations in knowledge is epistemic. An honest communication of uncertainty enhances transparency in scientific advice (Fischhoff & Davis, 2014; SAPEA, 2019). There are different ways to communicate uncertainty (van der Bles et al., 2019). How and what to communicate depends on how uncertainty is characterized in scientific assessment, i.e., the process of using scientific evidence and reasoning to answer a question or estimate a quantity in a specific decision-making context (EFSA et al., 2018). Organizations responsible for scientific assessment have developed guidance and recommendations about characterizing epistemic uncertainty in scientific assessment and communicating uncertainty to decisionmakers and the public (EFSA et al., 2018, 2019; FAO & WHO, 2021; Institute of Medicine, 2013).

The European Food Safety Authority (EFSA)'s guidance on uncertainty analysis (EFSA et al., 2018) emphasizes the importance of well-defined assessment questions, a systematic identification and characterization of individual sources of uncertainty, and the ultimate goal to characterize overall

uncertainty in the answer to the assessment question. This can be done by dividing the assessment into parts, identifying and characterizing uncertainty in each part, making a distinction between aleatory (i.e., inherent randomness, heterogeneity, or variability) and epistemic uncertainty, and evaluating the combined uncertainty on the quantities of interest chosen to support the conclusion of the assessment.

EFSA's guidance includes a description of alternative ways to characterize uncertainty, including qualitative and quantitative methods (EFSA et al., 2018). Since 2019, EFSA recommends that assessors should try to express epistemic uncertainty about any quantity of interest in scientific assessments (also known as overall uncertainty) in a quantitative way using subjective probability (either precise or approximate).

To meet EFSA's recommendation, we revisit a possible framework to quantify these two expressions (precise and bounded probability) for epistemic uncertainty that also comply with the requirements of scientific assessments to

1. distinguish between aleatory and epistemic uncertainty, and
2. integrate available evidence in a systematic way.

Epistemic uncertainty should be distinguished from aleatory uncertainty to allow quantification of uncertainty about a quantity of interest, even when it is based on a summary of aleatory uncertainty.

There is an ongoing debate in the risk analysis community about the use of probability versus other expressions of epistemic uncertainty, and if probability should be restricted to aleatory uncertainty (Aven, 2010, 2020; Ferson & Ginzburg, 1996; Helton & Oberkampf, 2004). There are limitations with any expression of uncertainty, which should be acknowledged in the assessment. The benefits of probability (precise or bounded) are, besides being a quantitative expression, which they comply with the two requirements listed above.

It is possible to maintain a distinction between aleatory and epistemic uncertainty and quantify epistemic uncertainty by a precise probability (see, e.g., Helton, 1997; Kelly & Smith, 2011; Nauta, 2000; O'Hagan, 2013; US EPA, 2011). In some part of an assessment, experts may provide precise and/or imprecise judgments which support the specification of a precise and/or approximate probability for quantifying epistemic uncertainty. When epistemic uncertainty is quantified by precise probability, Bayesian inference allows for a systematic integration of evidence into the assessment (O'Hagan, 2013). There is a need for guidance on how to integrate evidence when epistemic uncertainty is represented by an approximate (bounded) probability. The EFSA guidance for uncertainty analysis describes probability bound analysis as a way to propagate epistemic uncertainty through an assessment model, separating aleatory from epistemic uncertainty (EFSA et al., 2018, section B.13). But, in the current version of the guidance there is no recommendation on how to integrate evidence in assessment when uncertainty is quantified by approximate probability.

The aim of this article is to demonstrate a possible way for EFSA to quantify uncertainty in a quantity of interest (that is an output from an assessment model) by bounded probability, and explicitly distinguish between epistemic and aleatory uncertainty. This way, here referred to as robust Bayesian analysis, includes standard Bayesian analysis where uncertainty is quantified by precise probability as a special case.

1.2 | Bayesian Analysis to Quantify Uncertainty in Scientific Assessments

In scientific assessments, it is useful to make a clear distinction between assessment variables and assessment parameters (EFSA et al., 2018). Parameters are theoretical constructs within a scientific model, with values that are fixed (true) but we are uncertain about (van der Bles et al., 2019). As an example, variability in body weights of individuals of a certain age and sex in a population can be modeled to follow a normal distribution with unknown mean μ and (for simplicity) known variance σ^2. Here, μ and σ^2 are assessment parameters, and "body weight" is an assessment variable modeled as $N(\mu, \sigma^2)$. Uncertainty about the parameter μ can be quantified by subjective probability, e.g., by specifying how probable different values u of the parameter are, i.e., $P(\mu < u)$. A probability distribution consists of probabilities for all possible values on u.

A subjective probability represents someone's degree of belief that a statement is true now (or will be true at a specified time in the future) conditional on the knowledge that person has (Hampton et al., 1973; Lindley, 2006; O'Hagan, 2013). Different interpretations of subjective probability exist. For example, in the betting interpretation, the probability $P(A)$ indicates how much a person is willing to pay (sell) for a bet if he/she would receive a unit reward if the event A occurs and nothing otherwise (de Finetti, 1937; Hájek, 2019). This interpretation has the advantage of giving an operational method for measuring someone's degree of belief in terms of their behavior. However, it suffers from the usual issues associated with operational interpretations of probability. For instance, in some situations, the value of the rewards involved in the bet may depend on the outcome of the bet itself, or on whether the bet is placed at all (Hájek, 2019). This can be particularly problematic in the context of risk analysis, where for instance the life or quality of life of the bettor (decision maker) itself might be at stake (Aven, 2021; Aven and Reniers, 2013; Lindley, 2006).

To circumvent these well-known issues that may arise with the betting interpretation of subjective probability, a nonoperational interpretation of probability may be more suitable. For instance, an alternative interpretation is proposed by Lindley (2006), which links subjective probability to an uncertainty standard such as an urn. Lindley suggested that the probability $P(A)$ is the number such that uncertainty about the occurrence of the event A is considered equivalent to uncertainty about the occurrence of some standard event, e.g., if a person assigns a probability of 0.2 for an event A, his/her degree of belief in A occurring is equivalent to drawing a red ball at random from an urn containing 10 balls where two balls are red. The interpretation given by Lindley (2006) has been

suggested by Aven and Reniers (2013), Aven (2021) as more appropriate in the context of risk analysis because it does not conflate uncertainty judgments with value judgments. We refer the reader to Hájek (2019) for a much more extensive discussion around the operational issues with the betting interpretation and how some of these issues can be addressed. In addition, Aven (2020) argues that in an assessment context, a value-neutral interpretation of subjective probability as a representation of the assessors' uncertainty is preferable. An expert can use a subjective probability to represent her certainty about a statement given her current knowledge (Singpurwalla & Wilson, 2008), which may or may not be followed with Bayesian updating integrating evidence available for the assessment.

In Bayesian analysis, a prior is a subjective probability distribution quantified by expert judgment that represents uncertainty about the parameters prior to considering information in data (Gelman et al., 2013; Kelly & Smith, 2009; McElreath, 2016; O'Hagan, 2013). Data are integrated by specifying a probabilistic model over the possible observations under different values on the parameters, which can be used to get the likelihood. Given prior information and the likelihood for data, uncertainty about parameters is updated using Bayes rule.

An assessment model may use probability distributions to model aleatory uncertainty (Cox, 2006; Kelly & Smith, 2011). Bayesian analysis uses subjective probability to quantify uncertainty in parameters. For example, the marginal distribution over parameter μ is interpreted as a characterization of epistemic uncertainty about the population mean (Fig. 1B), while the distribution of body weights given a specific value on the mean $\mu = u$ is interpreted as aleatory uncertainty (one of the cumulative density functions (cdf) in Fig. 1A).

In general, the ability to distinguish between aleatory and epistemic uncertainty relies on the extent to which aleatory uncertainty is explicitly modeled (i.e., if the statistical models consider relevant sources of variability in the system and/or measurement errors), since epistemic uncertainty is attributed to the parameters of the specified model. A separation is possible by modeling the full probability distribution for both parameters and variables (i.e., epistemic and aleatory uncertainty), and then appropriately marginalizing the distribution to characterize epistemic uncertainty about a quantity of interest. The quantity of interest is related to the assessment question.

1.3 | Beyond Bayesian Analysis

A common criticism of Bayesian analysis relying on subjective probability is that it needs subjective judgments to specify the prior (Aven, 2020; Gelman, 2008; Hampton et al., 1973). This issue can be approached by carefully specifying the priors using structured methods for expert knowledge elicitation (EFSA, 2014; O'Hagan, 2019; O'Hagan et al., 2006). On the other hand, there is no assessment free from subjectivity. For example, the choice of statistical model (to derive the likelihood) is always subjective for any type of inference. A scientific approach for treating uncertainty in assessments requires an acknowledgment that uncertainty is subjective and conditional on the current knowledge (EFSA et al., 2018; Goldstein, 2006; O'Hagan, 2019). Using a statistical principle that can work with uncertainty conditional on knowledge (e.g., Bayesian analysis) is useful in scientific assessments. In particular, this holds when the aim is to quantify and revise uncertainty in a transparent and reproducible way.

In scientific assessments, it is common to use precise prior information, instead of flat priors elicited from expert knowledge. However, in some situations scientific experts might struggle to agree on a precise subjective probability (O'Hagan et al., 2006). This difficulty could raise a concern when experts are directly quantifying uncertainty in a quantity of interest, or for Bayesian updating based on information in data that is sparse, weak or conflicting (Walley, 1991; Walter & Coolen, 2016). To account for the issue of specifying a precise probability and because approximate quantitative expressions of uncertainty may be enough to inform decision making, EFSA et al. (2018) allow for epistemic uncertainty in overall uncertainty to be quantified by bounded probability.

A bounded probability can partially reflect someone's strength of knowledge, where a wider difference between the lower and upper bounds indicates less strength. The lower and upper bounds can also be interpreted as a range for the probability expressing epistemic uncertainty without indicating strength of knowledge. This could occur when experts provide judgments with bounds, for example, to possibly simplify the expert elicitation process. The bounds can also be given a betting interpretation (Walley, 1991). In this interpretation, increasing the lower bound strengthens our belief in favor of a statement because we are increasing the amount we are willing to pay for a bet that gives us a unit reward if the statement proves to be true. Similarly, decreasing the upper bound strengthens our belief against a statement, because we are increasing the amount we are willing to pay for a bet that gives us a unit reward if the statement proves to be false. In that sense, when lower and upper bounds coincide, we have expressed the strongest possible belief in a statement (Troffaes & Cooman, 2014; Walley, 1991). In all assessments, there will always be limitations in knowledge that is not considered in the assessment model, which should be acknowledged in the conclusion (EFSA et al., 2018). Therefore, approaches for uncertainty have emerged that communicate judgments on the strength of knowledge supporting the assessment side by side with expressions of uncertainty about the quantity of interest (Aven, 2020; van der Bles et al., 2019).

Probability bounds analysis (PBA) is a method to calculate bounds on the probability of a composite event based on expressions of bounded probability of the other events (EFSA et al., 2018). PBA can propagate aleatory and epistemic uncertainty through a model, by representing uncertainty associated to each assessment variable by a probability box (p-box) (Ferson et al., 2003; Flage et al., 2018). A "p-box for an assessment variable" is a set of

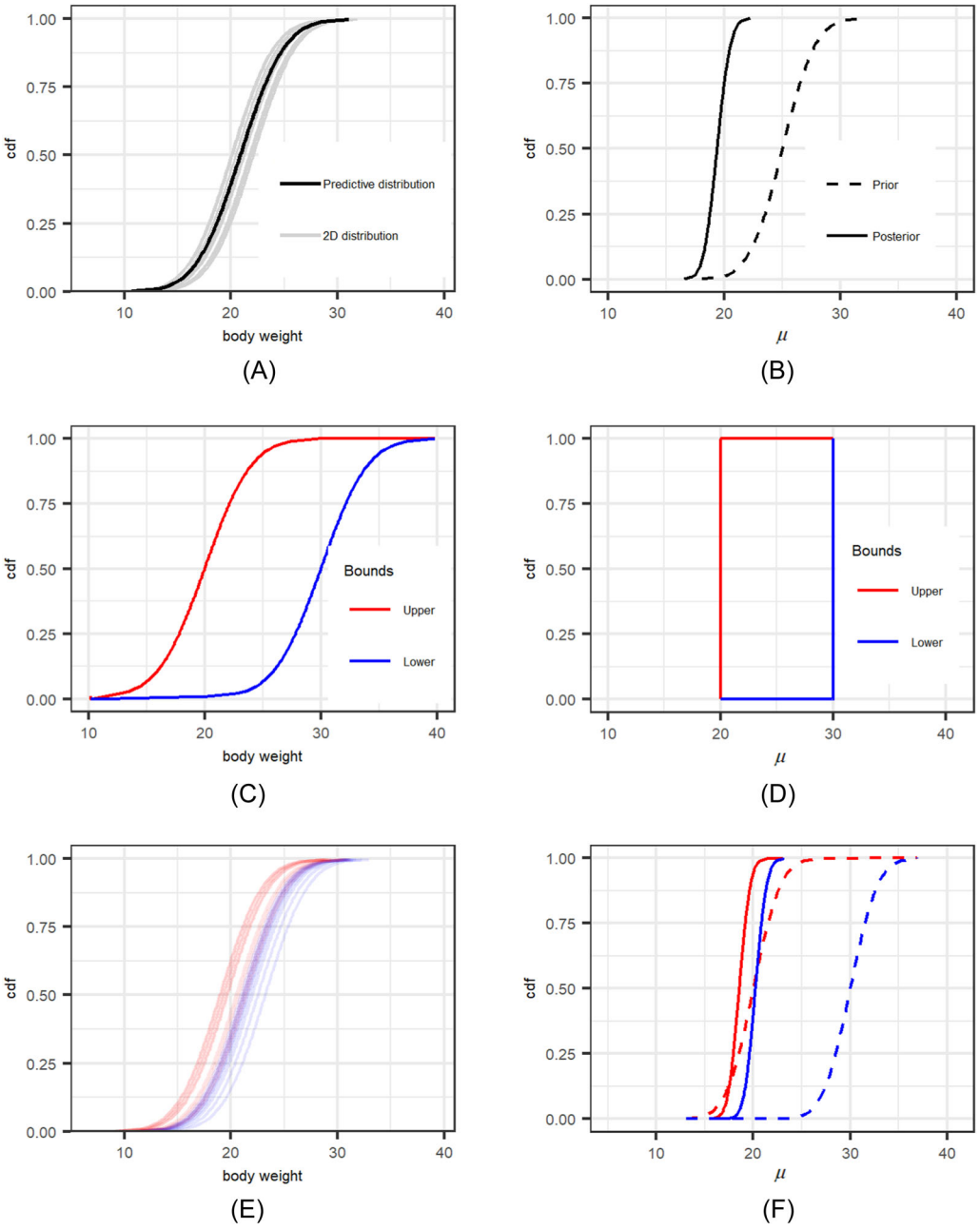

FIGURE 1 Uncertainty in assessment variable ("body weight") and parameter average body weight (μ) quantified using Bayesian analysis (A, B), probability bounds analysis (C, D), and robust Bayesian analysis (E, F). Under these three approaches, uncertainty in the parameter μ is quantified by (B) a precise probability distribution (first a prior, then a posterior), (D) an interval, and (F) a set of probability distributions (first a set of priors (dashed lines), then a set of posteriors (solid lines)). Uncertainty in the variable "body weight" is characterized by (A) a 2D-distribution, (C) a p-box, and (E) 2D-distributions for every distribution in the set of probability distributions for the parameter (light blue and light red lines are 2D-distributions associated to the lower bound and the upper bound of the posterior, respectively). A posterior predictive distribution (black line in (A)) is a marginalization over the 2D-distribution

probability distributions for aleatory uncertainty bounded by a lower and an upper cumulative distribution function. A p-box may arise from a parameterized distribution, where uncertainty in parameters is expressed by intervals with no further probabilistic model on those intervals. That is, no uniform or other distribution on parameters is assumed, and all possible parameter values, within set bounds, are propagated. Going back to our previous example, a p-box of body weight (see, for instance, Fig. 1C) could have been derived by expressing epistemic uncertainty in the expected body weight μ by an interval (for example, Fig. 1D) (instead of a subjective probability distribution).

P-Boxes on assessment variables may be hard to justify in scientific assessment. First, the intervals are treated

as 100% probability intervals, which means that there is a zero probability of a parameter taking a value outside the interval. Second, some information from the underlying evidence is lost when epistemic uncertainty is summarized into intervals, instead of probability distributions (where different values are weighted according to how probable they are). Therefore, we argue that p-boxes on variables are appropriate when epistemic uncertainty is directly quantified as intervals on parameters instead of probability distributions, and there has been no formal updating using Bayes rule. But this is often not the case in scientific assessments, where available evidence (including data) are to be taken into account. This leads to the question, what to do instead in order to quantify epistemic uncertainty by bounded probability?

1.4 | Bayesian and Robust Bayesian Analysis as a Possible Way Forward

Robust Bayesian analysis has been used to evaluate the sensitivity to the choice of priors (e.g., the choice of family or hyperparameters of the prior distribution) (Berger, 1990; Roos et al., 2015). An alternative use of "robust" Bayesian analysis is to apply Bayesian inference over a set of priors for quantifying uncertainty about parameters, resulting in uncertainty quantified by bounded (imprecise) probability (Walley, 1991; Walter & Coolen, 2016). Used in this way, robust Bayesian analysis is based on statistical principles for inference within the theory of imprecise probability (Walley, 1991). For example, Sahlin et al. (2021) use bounds on probability for expressing severe uncertainty about the state of the system. Because robust Bayesian analysis allows integration of evidence, it is a useful approach to quantify epistemic uncertainty by bounded probability in scientific assessments.

A set of priors represents prior beliefs measured through bounds. This results in a set of posterior distributions (see, for example, Fig. 1F) and bounds are derived for relevant summaries from probability distributions such as expectations, percentiles, or probability intervals (for instance, see Fig. 1E). Imprecision can also emerge from the likelihood of a given parametric model, as a consequence of doubt about the specification of the likelihood and/or imprecise data, e.g., interval data or categorical data with unclear class membership. For more details, see Benavoli and Ristic (2011) and Cattaneo and Wiencierz (2012).

In this article, we demonstrate a way to quantify epistemic uncertainty by precise and bounded probability derived by Bayesian analysis and robust Bayesian analysis. To illustrate the approach, we specify an assessment model and quantify epistemic uncertainty by precise and bounded probability for an assessment of the weekly intake of aluminium via consumption of cocoa and chocolate products (Schendel et al., 2018). This example is chosen since this is a common type of scientific assessment aiming to answer if a continuous assessment variable (here exposure) exceeds a safety threshold.

2 | A MATHEMATICAL FRAMEWORK FOR THE QUANTIFICATION OF EPISTEMIC UNCERTAINTY USING BAYESIAN AND ROBUST BAYESIAN ANALYSIS

2.1 | The Separation of Epistemic and Aleatory Uncertainty in Bayesian Analysis

Let us start with a simple model with one variable X and one parameter θ. In Bayesian analysis, uncertainty about parameter θ is quantified in the form of a subjective probability distribution. Uncertainty about the parameter can be updated using Bayes rule with evidence, e.g., observations of the variable X. This is done by combining the model for the uncertainty in the parameter before the variable X is observed (the prior), and a model for how likely it is to observe the data $\mathbf{x} := (x_1, \ldots, x_n)$ given specific values on the parameter θ (the likelihood). In what follows, we assume that data are conditionally independent and identically distributed given the parameters in the model (i.e., exchangeable). Uncertainty about the parameter after updating (the posterior) is

$$P(\theta \mid \mathbf{x}) \propto P(\theta)P(\mathbf{x} \mid \theta), \tag{1}$$

where $P(\theta)$ is the prior probability and $P(\mathbf{x} \mid \theta)$ is the likelihood.

The prior can, but does not have to, be expressed with a parametric probability distribution, for which the probability density function (pdf) has a mathematical form with a fixed number of parameters (Cox, 2006). To avoid double use of the term parameter, parameters of a parametric prior distribution are in Bayesian terminology referred to as hyperparameters. Using a parametric prior, the posterior distribution is conditional on data and the hyperparameters:

$$f(\theta \mid \mathbf{x}, t_0) \propto f(\theta \mid t_0) \prod_{i=1}^{n} f(x_i \mid \theta), \tag{2}$$

where f denotes a pdf, and t_0 denotes the hyperparameters.

A Bayesian model is a joint probability distribution of parameters and variables, which are both mathematically treated as random variables. Although both parameters and variables are modeled as random variables, probability distributions for parameters should always represent epistemic uncertainty interpreted as subjective probabilities. Note that a parameter in a statistical model may not be a parameter in the strict sense as defined in the context of an assessment (i.e., a fixed, but uncertain number). An example of this is a random effect in a multilevel or hierarchical model, which is a quantity with variability (this issue is raised in Kadane, 2011, Chapter 9). When marginalizing the probability distribution of a parameter, the resulting distribution represents only epistemic uncertainty. The probability distributions for assessment variables represent aleatory uncertainty with a frequency interpretation. This is in line with the probability

on frequency approach (Apostolakis, 1990; Kaplan & Garrick, 1981).

Since the distribution of X is dependent on the parameter θ, it is not possible to summarize aleatory uncertainty without epistemic uncertainty. In Bayesian analysis, the predictive distribution for a future value of a variable, expresses a mixture of aleatory and epistemic uncertainty which is represented by a single probability distribution (Fig. 1A) (Gelman et al., 2013; McElreath, 2016). To illustrate the contribution of epistemic uncertainty in relation to aleatory uncertainty, an assessment variable is specified or visualized as a two-dimensional probability distribution (2D-distribution) (Aldenberg & Jaworska, 2000; EFSA et al., 2018). A 2D-distribution consists of pdfs or cdfs for the variable, that are sampled from probability distributions for the parameters within the model of the variable (Fig. 1B) (EFSA et al., 2018). Thus, a 2D-distribution (Fig. 1A) represents aleatory uncertainty at the variable level by taking into account epistemic uncertainty in parameters.

2.2 | Quantifying Uncertainty in the Quantity of Interest By Precise Probability

In order to characterize the magnitude of epistemic uncertainty in the quantities of interest, there is a need to evaluate uncertainty. For illustration, the quantity of interest is here (see next section) the frequency of a target population exceeding a critical threshold. More specifically, the quantity of interest is the frequency of the event "a function $g(\cdot)$ of a future value of the assessment variable X_i, where $i > n$, exceeds the critical value g^*." We use a function $g(\cdot)$ to allow for possible transformations of the assessment variable and later on to generalize it into a function of several assessment variables.

In standard Bayesian analysis, the posterior probability for this event can be written as the integral of the quantity of interest for a given value of the parameter θ, times the posterior distribution of θ:

$$P(g(X_{n+1}) > g^* \mid \mathbf{x}, t_0) = \int P(g(X_{n+1}) \geq g^* \mid \theta) f(\theta \mid \mathbf{x}, t_0) d\theta. \tag{3}$$

If we define the quantity of interest as a function of assessment parameters θ:

$$h(\theta) := P(g(X_{n+1}) \geq g^* \mid \theta), \tag{4}$$

and Θ the random variable corresponding to uncertainty in the parameter θ, Equation (3) can be written as:

$$\int h(\theta) f(\theta \mid \mathbf{x}, t_0) d\theta = \mathrm{E}(h(\Theta) \mid \mathbf{x}, t_0). \tag{5}$$

When presented in this way, the quantity of interest $h(\theta)$ is a characterization of aleatory uncertainty (without necessarily being directly observed). Uncertainty in the quantity of interest, $h(\Theta)$ in Equation (5), is a random variable in the form of a subjective probability distribution. This result holds when generalizing the quantity of interest to be an event over several assessment variables and when $h(\cdot)$ is a function of several assessment parameters.

2.3 | Bayesian Inference on the Quantity of Interest

Monte Carlo simulation is a useful method for estimation when it is difficult to obtain an analytical solution (Vose, 1996). To estimate the expected value of the quantity of interest, let $\theta_1, \ldots, \theta_M$ be random samples drawn from the posterior distribution $f(\theta \mid \mathbf{x}, t_0)$. For each $j \in \{1, \ldots, M\}$, let $x^j_{n+1}, \ldots, x^j_{n+N}$ be random samples drawn from the assessment variable conditional on θ_j. Then, $\widehat{h}(\theta_j)$ is a Monte Carlo estimate for $h(\theta_j)$:

$$h(\theta_j) \approx \widehat{h}(\theta_j) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\left\{ g\left(x^j_{n+i}\right) \geq g^* \right\}}. \tag{6}$$

Thus, sample j of the quantity of interest $\widehat{h}(\theta_j)$ is not directly computed from θ_j, but from the samples $x^j_{n+1}, \ldots, x^j_{n+N}$.

The posterior expected value of the frequency of exceeding a threshold can be estimated by

$$\mathrm{E}(h(\Theta) \mid \mathbf{x}, t_0) \approx \frac{1}{M} \sum_{j=1}^{M} \widehat{h}(\theta_j). \tag{7}$$

In this approach, the quantity of interest is estimated in two steps. First, we randomly generate parameters, θ_j, from the posterior distribution which characterizes epistemic uncertainty. Second, we generate values on the assessment variable x^j_{n+i} conditional on the generated parameters θ_j, by sampling from the model expressing aleatory uncertainty. This two-step procedure is called two-dimensional Monte Carlo (2D-MC) simulation and is used to separate epistemic from aleatory uncertainty (EFSA et al., 2018; Vose, 1996).

For communication of uncertainty in assessments, it is recommended to also summarize uncertainty in the quantity of interest by one or several percentiles (EFSA et al., 2019). Substituting the arithmetic mean by the posterior percentile function of $h(\Theta)$ in Equation (7), we get $h_p(t_0)$ with

$$P(h(\Theta) \leq h_p(t_0) \mid \mathbf{x}, t_0) = p \tag{8}$$

for a given $p \in [0, 1]$. The percentile is approximated by the pth percentile in the sample of quantities of interest $(\widehat{h}(\theta_1), \ldots, \widehat{h}(\theta_M))$ as follows:

$$\widehat{h}_p(t_0) := \widehat{h}^{(r)}, \tag{9}$$

where $r = M\frac{p}{100}$ (smallest integer) and $\widehat{h}^{(r)}$ is the rth order statistics (the rth smallest value in the sample $\widehat{h}(\theta_1), \ldots, \widehat{h}(\theta_M)$). The percentile function can be used to estimate the cdf for uncertainty in the quantity of interest.

2.4 | Quantifying Uncertainty in the Quantity of Interest by Bounded Probability

Robust Bayesian analysis specifies a set of prior distributions to quantify uncertainty in parameter θ. A set of prior distributions represents uncertainty, arising from ambiguity or imprecision, in the specification of a unique (precise) prior distribution. A bound on a probability is obtained by optimizing a summary from a distribution, e.g., an expectation or percentile, over the set of priors. To simplify the analysis, one can assume that all prior distributions belong to the same family of probability distribution. Thus, the set of prior distributions is specified by a set of hyperparameter values T_0. For the assessment, we want to estimate a bound on a probability or on the expectation of the quantity of interest over a set of hyperparameters T_0.

Let us first look how to estimate the upper bound on the expected frequency of exceeding a threshold over our set of prior distributions:

$$\overline{\mathrm{E}}(h(\Theta) \mid \mathbf{x}, t_0) := \max_{t_0 \in \mathrm{T}_0} \int h(\theta) f(\theta \mid \mathbf{x}, t_0) d\theta, \tag{10}$$

which can be approximated by

$$\overline{\mathrm{E}}(h(\Theta) \mid \mathbf{x}, t_0) \approx \max_{t_0 \in \mathrm{T}_0} \frac{1}{M} \sum_{j=1}^{M} \widehat{h}(\theta_j(t_0)). \tag{11}$$

Here, we explicitly write $\theta_j(t_0)$ to emphasize its dependency on t_0.

This estimator gives a conservative bound on the upper expectation, as it is a biased estimator (i.e., the expected value of the estimator given by Equation (11) is higher than the upper expectation given in Equation (10)); for details, see Troffaes (2018). The bias can be kept small by expressing the samples as smooth functions of the hyperparameters and of the sources of randomness that produce the sample, and thus by not resampling within the optimization process. In our analysis, we do not quantify the bias explicitly because the estimated bounds are already reasonable. However, if needed, the bias can be quantified with additional computational effort, for instance by rerunning the sampling and studying the variation in the approximation of the bounds as described in Troffaes (2018).

The upper bound on the pth percentile for the quantity of interest in Equation (8) can be approximated by

$$\overline{h}_p(t_0) := \max_{t_0 \in \mathrm{T}_0} h_p(t_0) \approx \max_{t_0 \in \mathrm{T}_0} \widehat{h}_p(t_0). \tag{12}$$

Lower bounds can be computed using the minimum instead.

It is possible to visualize uncertainty in the quantity of interest by combining estimated lower and upper bounds of percentiles into a "p-box for uncertainty in a parameter." This is then a p-box for a parameter conditional on data and the set of priors T_0 (see, for example, Fig. 1F). Hence, robust Bayesian analysis makes it possible to quantity bounds on uncertainty quantified by probability and integrate data and expert knowledge in a coherent way. In robust Bayesian analysis, there is no ideal way to visualize uncertainty at the variable level (see, for example, Fig. 1E) in a way that clearly separate aleatory and epistemic uncertainty similar to the 2D-distribution in the precise probability case (see, for instance, Fig. 1B).

3 | A CASE STUDY

3.1 | Introduction to the Case Study

To illustrate both frameworks, Bayesian analysis and robust Bayesian analysis are applied on an assessment model that contains several continuous assessment variables which are combined to estimate a quantity of interest. This is a common situation in scientific assessments, including both aleatory and epistemic uncertainty.

To evaluate the main principles of EFSA's new guidance for uncertainty analysis (EFSA et al., 2018), the German Federal Institute for Risk Assessment (BfR) applied different methods for uncertainty analysis on an assessment of the long-term aluminum intake (chronic toxicity) by consumption of chocolate and cocoa products for children (Schendel et al., 2018). The population group consisted of children in Germany that are infants from age 0.5 years to less than five years (which are not breastfed). The weekly intake of aluminum via consumption of cocoa and chocolate products is considered to be safe if it does not exceed the tolerable weekly intake of aluminium (1 mg/(kg bw)/week). The examples in the BfR report did not indicate any concern for safety.

In this article, we apply both frameworks on the safety assessment in the report done by BfR. The intake of aluminum via consumption of cocoa and chocolate products is modeled by the intake equation:

$$Y = C \cdot A, \tag{13}$$

where C is consumption per body weight per week and A is aluminum concentration in the consumed products. We define as our quantity of interest, the limit of the relative frequency of children in the target population exceeding the tolerable weekly intake of aluminum (referred to as the frequency of exceeding TWI).

The BfR assessment (Schendel et al., 2018) used consumption data collected in Germany during 2001/2002 on chocolate consumption per body weight of children. In particular, the consumption of the following products

was taken into consideration: sugar panned chocolate, milk chocolate/baking chocolate, chocolate icing/chocolate sprinkles/chocolate coating, chocolate with fillings, dark chocolate, cocoa powder and beverages containing cocoa powder, chocolate-, nougat-, and cocoa-cream and mixed-milk beverages containing cocoa. Because the consumption study was judged out of date, they performed expert knowledge elicitation to assess the rate of change of cocoa powder consumption from 2001/2002 to 2017. Measurements of aluminum concentrations in these chocolate products were taken from a separate study. They used the data to inform the assessment variables and weekly intake was derived by the intake equation (Equation (13)).

3.2 | The Assessment Model

To assess the quantity of interest, we specify an assessment model that allows us to (1) consider relevant variability, and (2) integrate information in data and expert judgment using Bayesian principles for inference. The model will consider variability in consumption between individuals in the target population and variability in aluminum concentrations between samples of chocolate products. Here, we consider seven of the nine products in the BfR report (Schendel et al., 2018), which were products with data on both consumption and aluminum concentration. It is possible to consider all products, but this decision was taken to simplify the description of the assessment model. As a consequence, the weekly intake will be underestimated.

The weekly intake Y_i for the ith child is modeled as a sum of the intake from each of the seven products

$$Y_i := \sum_{k=1}^{7} C_{ik} A_{ik}, \tag{14}$$

where C_{ik} is the weekly consumption of product k by child i, and A_{ik} is the aluminum concentration in product k for child i during a given week. There is variability in both aluminum concentration A_{ik} and consumption C_{ik} between consumption events. In this assessment, we assume that there is one potential consumption event per week per product.

Variability in aluminum concentration across consumption events is modeled by a lognormal probability distribution with the parameters mean μ_k^A and precision τ_k^A.

$$\log(A_{ik}) \mid \mu_k^A, \tau_k^A \sim \mathcal{N}\left(\mu_k^A, \tau_k^A\right). \tag{15}$$

Precision is the inverse of the variance, and is an alternative specification of the second parameter for a normal distribution[1] (Gelman et al., 2013).

[1] The pdf for a normal distribution with precision as the spread parameter is $f(x) = \sqrt{\frac{\tau}{2\pi}} \exp(-\frac{\tau(x-\mu)^2}{2})$, where x is a sample from X. If Y is lognormally distributed, then $\log(Y)$ is normally distributed with the same parameters.

In order to model consumption events among children, we first acknowledge that not all children consume all these products every week. Let B_{ik} be a variable that indicates whether child i consumes chocolate product k taking the value 1 if consumption occurs and 0 otherwise. Let π_k^B be the frequency a child consumes chocolate product k during a given week. Then B_{ik} is a Bernoulli variable

$$B_{ik} \mid \pi_k^B \sim \text{Bernoulli}\left(\pi_k^B\right). \tag{16}$$

An advantage of modeling consumption as a mixture of consumption events and consumption at a consumption event is that it is easier to find suitable probability model for the amount consumed at consumption events. Also, a high proportion of zeros in consumption data might give rise to lower reliability in the statistical estimates if no mixture is considered.

Consumption by child i of product k during a given week is

$$\begin{cases} C_{ik} & \text{if } B_{ik} = 1, \\ 0 & \text{if } B_{ik} = 0, \end{cases} \tag{17}$$

where $C_{ik} = C'_{ik}(1 + \frac{\psi}{100})$ is defined as the product of consumption from 2001/2002, C'_{ik} and the relative change in consumption in 2017 compared to 2001/2002, ψ, elicited by experts. The weekly consumption from 2001/2002 of a child who consumes chocolate product k a given week is modeled as a lognormal distribution with the parameters μ_k^C and τ_k^C (mean and precision),

$$\log(C'_{ik}) \mid \mu_k^C, \tau_k^C \sim \mathcal{N}\left(\mu_k^C, \tau_k^C\right). \tag{18}$$

The assessment model contains several variables and parameters (1) and a structure for their dependencies is shown in Fig. 2. The next step is to quantify uncertainty in these parameters using Bayesian inference and combine them to evaluate uncertainty in the frequency of exceeding the threshold (i.e., the quantity of interest, later introduced in Equation (26)).

Other models may be suitable for these variables. The fit of each model could have been verified if data at individual level had been available. Given the models specified for the variables above, sufficient statistics derived from data on aluminum concentration and consumption from 2001/2002 in the assessment by Schendel et al. (2018) were used to support inference. The result from the expert knowledge elicitation is taken from the published report.

3.3 | Bayesian Analysis

3.3.1 | Uncertainty in Parameters

In this example, parameters associated with aluminum concentration (A), chocolate consumption (C'), and consumption

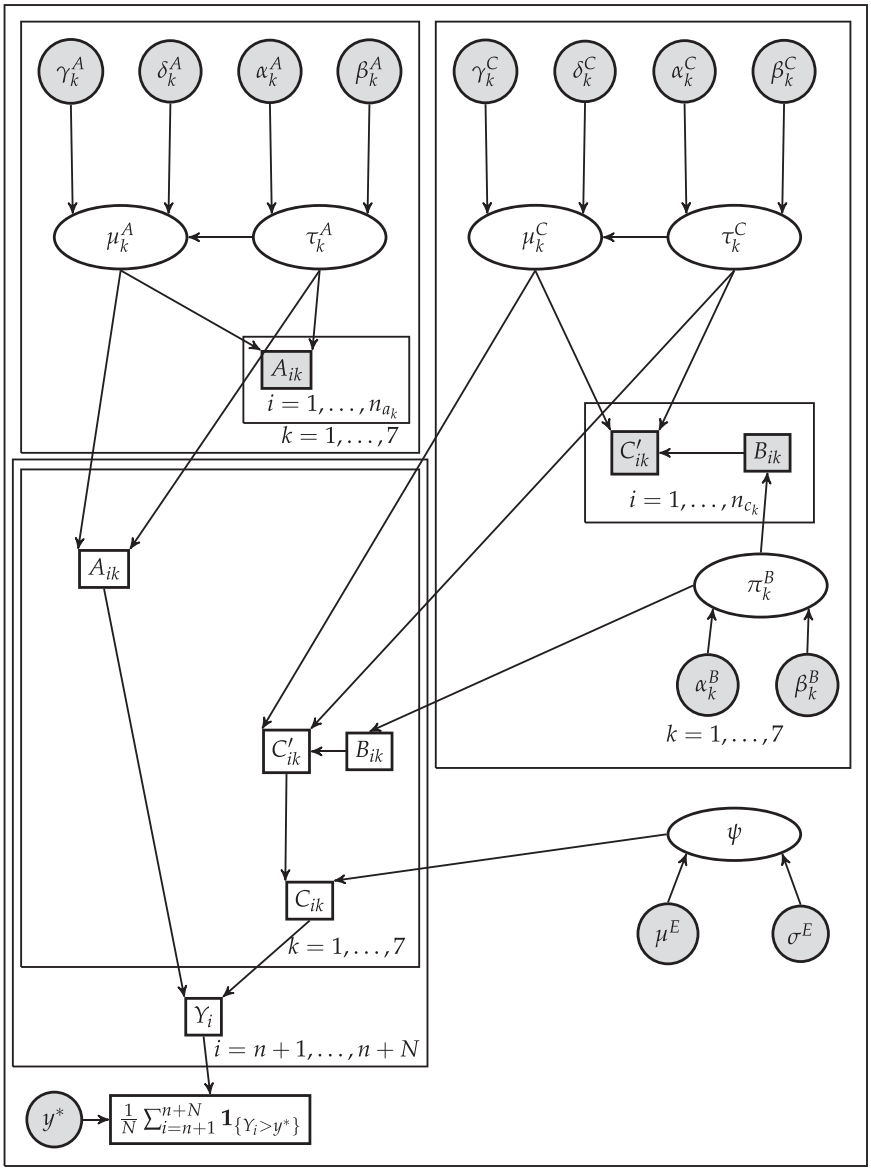

FIGURE 2 The assessment model represented as a directed acyclic graph. Parameters are represented by ellipses, hyperparameters (or other fixed quantities) by gray circles, observed variables (i.e., data) by gray squares. Predicted variables (white squares) are used to approximate the frequency of exceeding the threshold y^* using Monte Carlo simulation, where n is greater than n_{a_k} and n_{c_k}. Plates indicate repeated cases.

events B are to be informed with data. The parameter relative change ψ in consumption is on the other hand informed by expert judgment only. We use parametric distributions for all priors.

Updating uncertainty about parameters in Bayesian analysis can be done by analytical or sampling-based approaches (McElreath, 2016). Bayesian updating with a conjugate model is an example of an analytical approach, which has the convenient property that the distribution of the parameters belongs to the same family before and after Bayesian updating. Hence, updating is done by using a suitable parametric distribution for the prior and changing the hyperparameters based on information in the data (Gelman et al., 2013; McElreath, 2016). Let $f(\theta \mid t_0)$ be the pdf of a conjugate prior for the parameter θ with hyperparameters t_0. Then, there are hyperparameters t_1 such that

$$f(\theta \mid \mathbf{x}, t_0) = f(\theta \mid t_1). \tag{19}$$

For a conjugate model, the hyperparameters for the posterior t_1 are a function of t_0 and data $\mathbf{x}$.

Here, we choose suitable conjugate priors for the parameters that are updated with data. The selected priors are common for these types of parameters, i.e., mean, precision, and frequency of binary events. In practice, the choice of prior or sets of priors should be done with care and based on expert knowledge elicitation at the variable or parameter level (for details, see Daimon, 2008; Gosling, 2018; Oakley & O'Hagan, 2020; Schad et al., 2019). Moreover, the choice of prior should not be limited to conjugate priors.

The normal-gamma distribution is chosen for the parameters within the model for aluminum concentration in product k:

$$\tau_k^A \sim \text{Gamma}\left(\alpha_0^A, \beta_0^A\right), \tag{20}$$

$$\mu_k^A \mid \tau_k^A \sim \mathcal{N}\left(\gamma_0^A, \delta_0^A \tau_k^A\right), \tag{21}$$

TABLE 1 The Assessment Variables, Parameters, and Hyperparameters Included in the Assessment of Weekly Daily Intake (Equation (14))

Name	Variable	Distribution	Parameters	Distribution	Hyperparameters
Aluminum concentration	$(A_{ik})_{k=1}^{7}$	lognormal	$(\mu_k^A, \tau_k^A)_{k=1}^{7}$	normal-gamma	$(\alpha_k^A, \beta_k^A, \gamma_k^A, \delta_k^A)_{k=1}^{7}$
Consumption	$(C'_{ik} \mid B_{ik} = 1)_{k=1}^{7}$	lognormal	$(\mu_k^C, \tau_k^C)_{k=1}^{7}$	normal-gamma	$(\alpha_k^C, \beta_k^C, \gamma_k^C, \delta_k^C)_{k=1}^{7}$
Consumption event	$(B_{ik})_{k=1}^{7}$	Bernoulli	$(\pi_k^B)_{k=1}^{7}$	beta	$(\alpha_k^B, \beta_k^B)_{k=1}^{7}$
Change in consumption			ψ	normal	(μ^E, σ^E)

Note: The change in consumption 2001/2002–17 is denoted by the parameter ψ.

TABLE 2 Results of Expert Knowledge Elicitation for the Relative Change in Cocoa Powder Consumption from 2001/2002–17 Given as Precise (Values Taken from Table 27 in Schendel et al., 2018) and Bounded Probability (Values Assigned by the Authors)

	Percentile	1%	25%	50%	75%	99%
Change of cocoa consumption in %	Precise probability	−30	−15	−5	7.5	20
	Bounded probability	–	[−20, −10]	–	7.5	–

where $\alpha_0^A, \beta_0^A, \gamma_0^A$, and δ_0^A are hyperparameters for the normal-gamma prior distribution (1). The same prior distribution is used for all products k.

Due to conjugacy, the posterior distribution for these parameters is also a normal-gamma distribution with hyperparameters $\alpha_k^A, \beta_k^A, \gamma_k^A$, and δ_k^A:

$$\tau_k^A \mid \mathbf{a}_k \sim \text{Gamma}\left(\alpha_k^A, \beta_k^A\right), \tag{22}$$

$$\mu_k^A \mid \tau_k^A, \mathbf{a}_k \sim \mathcal{N}\left(\gamma_k^A, \delta_k^A \tau_k^A\right), \tag{23}$$

where $\mathbf{a}_k := (a_{k1}, \ldots, a_{kn_k})$ denotes the observed aluminum log concentrations for product k and $\alpha_k^A, \beta_k^A, \gamma_k^A$, and δ_k^A are calculated by Equations (A5) to (A8) (Appendix A).

Epistemic uncertainty in consumption parameters (mean μ_k^C and precision τ_k^C) is characterized by a normal-gamma distribution (Equations (20) and (21)), but with different hyperparameters α_0^C, β_0^C, γ_0^C, and δ_0^C (Table 1).

Uncertainty in the parameter relative change in consumption ψ, is quantified based on expert knowledge elicitation on percentiles (Table 2). Since relative change is continuous and able to take both positive and negative values, a normal distribution is fitted by minimizing the sum of the squared distances between elicited percentiles and the probability distribution function, using the *fitdist* function from the *SHELF* package in R (Oakley, 2020). As a result, uncertainty in the relative change in consumption is characterized by a normal distribution with parameters mean, $\mu^E = -4.414$ and standard deviation, $\sigma^E = 14.854$.

Uncertainty in the consumption event parameter (π_k^B) is quantified by the conjugate beta distribution (Appendix B):

$$\pi_k^B \sim \text{Beta}\left(\alpha_0^B, \beta_0^B\right), \tag{24}$$

with hyperparameters α_0^B and β_0^B. Due to conjugacy, the posterior distribution for this parameter is also a beta distribution:

$$\pi_k^B \mid \mathbf{b}_k \sim \text{Beta}\left(\alpha_k^B, \beta_k^B\right), \tag{25}$$

where α_k^B and β_k^B are given by Equations (B3) and (B4).

A graphical representation of the assessment model embedded in a Bayesian framework helps to visualize the dependencies between variables (observed and predicted), parameters and hyperparameters within the assessment model (see, for instance, Fig. 2). Epistemic uncertainty by precise probability is quantified when, given data $\mathbf{x}$ and prior knowledge t_0, we use Bayesian inference to update the hyperparameters into $t_1 := (\alpha_k^A, \beta_k^A, \gamma_k^A, \delta_k^A, \alpha_k^C, \beta_k^C, \gamma_k^C, \delta_k^C, \alpha_k^B, \beta_k^B)_{k=1}^{7}$. The following values on hyperparameters are chosen by the authors for illustrating the Bayesian analysis: $t_0 := (\alpha_0^A = 1, \beta_0^A = 1, \gamma_0^A = 3.5, \delta_0^A = 5, \alpha_0^C = 1, \beta_0^C = 1, \gamma_0^C = -3, \delta_0^C = 5, \alpha_0^B = 1, \beta_0^B = 1)$.

3.3.2 | Uncertainty in a Quantity of Interest

The assessment example in this article is based on a model informed by data and expert judgment from Schendel et al. (2018). There are uncertainties associated with expert judgment, the choice of model, or how data have been been collected, which can be evaluated by critical appraisal tools (e.g.,

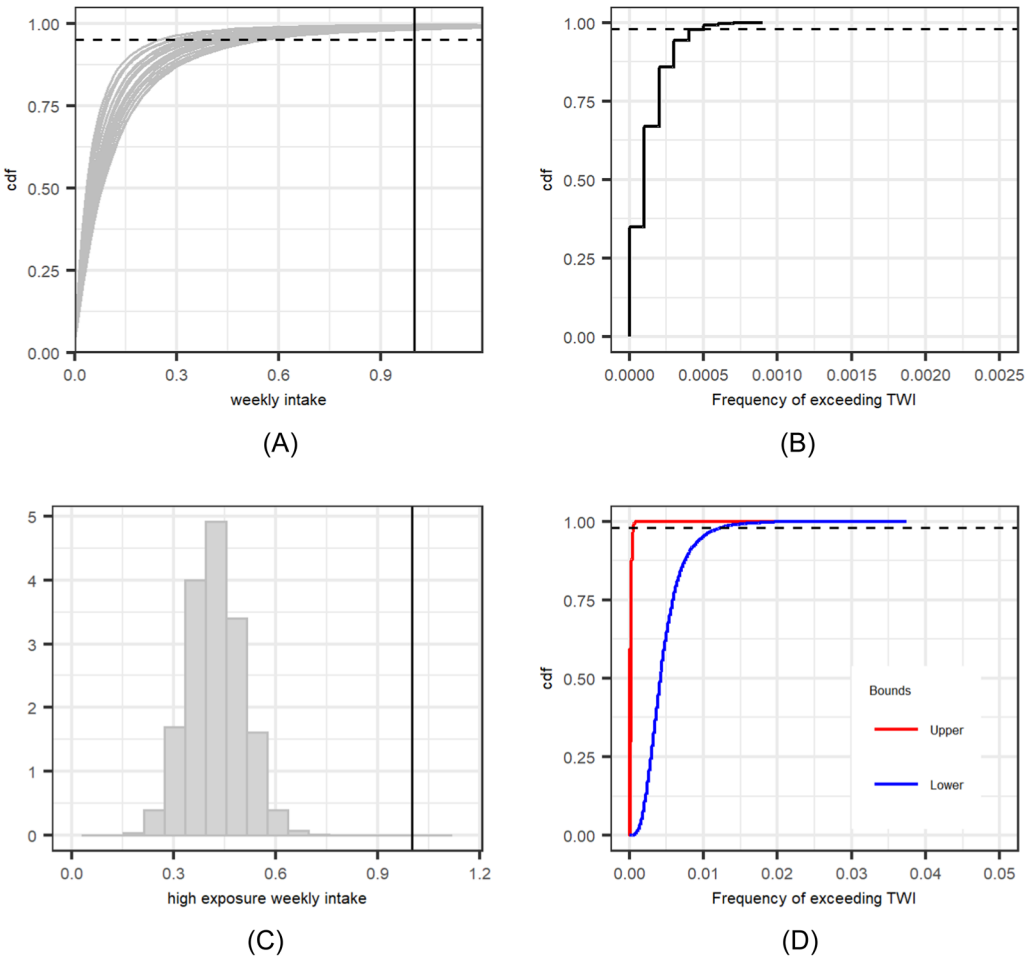

FIGURE 3 Different representations of uncertainty in the assessment output: (A) a 2D-distribution of the weekly intake of aluminum via chocolate consumption with the critical threshold (solid vertical line) and the probability 95% (dashed horizontal line), (B) the frequency of exceeding the tolerable weekly intake (TWI) and the probability 98% (dashed horizontal line), (C) the weekly intake for a high consumer summarized as the 95th percentile with the critical threshold (solid vertical line), and (D) the lower (blue) and upper (red) bounds for the probabilities quantifying uncertainty in the frequency of exceeding and the probability 98% (dashed horizontal line)

Goerlandt & Montewka, 2015). Uncertainty in the weekly intake of aluminum Y as quantified by the assessment model is visualized by a 2D-distribution (Fig. 3A). The figure shows that weekly intake is unlikely to exceed the critical threshold. In this assessment, safety is evaluated based on the frequency of the event that Y exceeds the threshold y^*. Conditional on the parameters θ,

$$h(\theta) := P(Y \geq y^* \mid \theta) \tag{26}$$

represents the frequency of this event, and captures aleatory uncertainty (see also Equation (4)).

Epistemic uncertainty in the parameters θ is expressed through the posterior distribution $f(\theta \mid t_1)$, which then translates into epistemic uncertainty about the quantity of interest $h(\theta)$. The subjective probability representing epistemic uncertainty in the quantity of interest is approximated using 2D-MC simulation using 10,000 iterations for both aleatory and epistemic uncertainty (Fig. 3B).

The analysis shows[2] that for a random child in the target population, the frequency of exceeding the safety threshold is less than 0.05% with a probability of 98%. We find this a useful way to decompose and represent the posterior uncertainty in the event $Y \geq y^*$.

There are other ways to marginalize the 2D-distribution. For example, weekly intake for a child with high exposure can be summarized as the 95th percentile of the output variable Y (see also Equation (8)). The analysis shows that the 95th percentile is unlikely to exceed the safety threshold (Fig. 3C). However, this marginalization does not say anything about individuals with higher exposures and is not describing uncertainty about the frequency of exceeding in the target population. This example illustrates the importance of carefully selecting the quantity of interest, which guides how to do the summary over aleatory uncertainty.

[2] Note that the conclusion of this assessment is not a result from an actual scientific assessment done by EFSA. We have done several modifications from the example used in Schendel et al. (2018).

3.4 | Robust Bayesian Analysis

Robust Bayesian analysis is an extension of the standard Bayesian analysis of the intake assessment. Here, this is done by defining a set of prior distributions for the parameters of the assessment model.[3]

We let all prior distributions associated with the assessment variables A and C belong to the same family of probability distributions, but intervals have been defined on a selection of the hyperparameters for μ_k^A and μ_k^C. The hyperparameters γ_0^A and γ_0^C are expanded by interval around the value chosen for the Bayesian analysis. Instead of 3.5, γ_0^A is set to be between 1 and 6, and instead of -3, γ_0^C is specified to be between -5 and 1. For the assessment, the set of hyperparameters for the prior T_0 is defined by $\alpha_0^A = 1$, $\beta_0^A = 1$, $\gamma_0^A \in [1, 6]$, $\delta_0^A = 5$, $\alpha_0^C = 1$, $\beta_0^C = 1$, $\gamma_0^C \in [-5, 1]$, $\delta_0^C = 5$, $\alpha_0^B = 1$, and $\beta_0^B = 1$ (Appendix C).

For illustration, we quantify uncertainty in parameter for the change in consumption, ψ, by a bounded probability distribution. We set up a hypothetical expert knowledge elicitation where the experts provide percentiles with bounds. In this case, we assign the 25th percentile to be in the interval $[-20, -10]$ and the 75th percentile to be 7.5 (Table 2).

In robust Bayesian analysis, the quantification and propagation of uncertainty are merged in the process of finding the bounds (see Equations (10) and (12)) by optimizing each bound over the set of priors. Due to conjugacy, here the objective function is a summary of the quantity of interest as a function of hyperparameters. This is maximized (or minimized) considering the set of prior distributions defined by γ_0^A, γ_0^C and the elicited intervals on the 25th and 75th percentiles for ψ.

Bounds on subjective probabilities expressing epistemic uncertainty are approximated using the Nelder–Mead algorithm. This is a heuristic algorithm which is available in the *dfoptim* package through the function *nmkb* in R (Varadhan & Borchers, 2018). If sets of nonconjugate priors are used, the optimization is more computational complex since requires optimizing using Markov chain Monte Carlo (MCMC) sampling.

Robust Bayesian analysis results in lower and upper bounds for the subjective probability distributions expressing uncertainty in the frequency of exceeding the safety threshold (Fig. 3D). The frequency of exceeding the safety threshold is less than 1.2% with a probability of at least 98%. Note that using bounded probability to quantify uncertainty in the quantity of interest allows us to express uncertainty about the conclusion in an approximate way, using "at least" instead of "is equal to." The lower bound is obtained at $\gamma_0^A = 3.02$, $\gamma_0^C = -2.51$ and elicited value on the 25th percentile equal to -13.45 and the upper bound is obtained at $\gamma_0^A = 6$, $\gamma_0^C = 1$ and elicited value on the 25th percentile equal to -10.17.

[3] The analysis that is presented below is made by the authors to illustrate an application of robust Bayesian analysis, and choices of sets are not elicited from experts.

3.5 | Summary of the Case Study

It is worthwhile to carefully specify the assessment model. A probability model for aleatory uncertainty allows for a clear distinction between variables and parameters, which is necessary for quantifying uncertainty in the parameters. With a probabilistic assessment model, it is possible to consider different quantities of interest, e.g., the frequency of exceeding the critical threshold (Fig. 3B) or the weekly intake for a high consumer (Fig. 3C). Uncertainty in the quantity of interest is derived by appropriate marginalization. This was done as the last step of the assessment.

The assessment done here is not fully comparable to the assessment done by BfR (Schendel et al., 2018) because aleatory and epistemic uncertainty are separated, quantified, and reported in different ways. We find that the result of the assessment using precise probability is in the same range as one done by BfR (comparing weekly intake for a child with high exposure assuming no brand loyalty) (Schendel et al., 2018). We did not expect a large underestimation due to the two products left out from the assessment. Major differences between this assessment and the one done by BfR are that we consider aleatory uncertainty by a probability model for variability in aluminum concentration and consumption events and that we quantify epistemic uncertainty by Bayesian analysis. In addition, we use a different probability distribution for uncertainty in parameter change in consumption, ψ.

Both the precise and bounded analysis did not exclude the possibility, although with small probability, of exceeding the weekly intake safety threshold. From the precise Bayesian analysis, the conclusion is that weekly intake is too high in less than 5 out of 10,000 children in the target population with a certainty of 98%. Robust Bayesian analysis shows that weekly intake is too high in less than 120 of 10,000 children in the target population with a certainty of 98% or more. The latter is a more conservative assessment. The result of standard Bayesian analysis is bounded by the result of robust analysis, as long as the prior in the standard Bayesian analysis is within the set of priors in robust Bayesian analysis. Therefore, a robust analysis can be useful to evaluate the sensitivity of the conclusions made by a precise analysis.

4 | DISCUSSION

This work is motivated by the recommendation of EFSA to use quantitative expressions of epistemic uncertainty in scientific assessment. To meet this recommendation, we demonstrate a framework to support assessment where uncertainty is quantified by probability (precise or bounded) and show that it meets two important requirements for scientific assessment, the possibility to distinguish aleatory from epistemic uncertainty, and a systematic principle to integrate evidence to the assessment. The framework, robust Bayesian analysis, allows for epistemic uncertainty to be quantified by bounded probability, with Bayesian analysis and precise probability as a special case. The framework is consistent with

the probability on frequency approach (Kaplan & Garrick, 1981) with 2D probability distributions for all assessment variables.

Robust Bayesian analysis allows for a more conservative quantification of uncertainty, compared to standard Bayesian analysis, by considering sensitivity to the choice of prior or likelihood. It has the potential to reveal data-prior conflict (i.e., mismatch between an informative prior distribution and the observed data). If so, the difference between the bounds is an emerging property from conflicting information between priors and data, which would not be seen in a standard Bayesian analysis (Walter & Coolen, 2016).

The advantages of robust Bayesian analysis compared to standard Bayesian analysis, will vary from case to case, because it depends on whether bounds can be approximated in a reliable way. The optimization step adds complexity because bounds must be approximated using optimization algorithms. For instance, bounds on probability or expectation could be directly estimated in the case of conjugate models as analytical forms are known. Otherwise, it is necessary to apply Monte Carlo or MCMC sampling to estimate bounds which could be slow under sets of priors as well as add complexity to the problem at hand (Troffaes, 2017, 2018).

There are several ways to summarize a precise or bounded probability distribution into quantitative expressions, such as an expected value, a probability interval or a percentile (EFSA et al., 2018; van der Bles et al., 2019). Consider as an example the message: "we expect that the weekly intake is exceeding the critical threshold for 12 out of 100,000 children in the target population." The message expresses an expected value derived from an assessment where uncertainty has been quantified by subjective (precise) probability. Although uncertainty is taken into account, the expected value does not explicitly reveal any uncertainty. Decision makers may be sensitive to the magnitude of uncertainty in the quantity of interest. To consider this, uncertainty in the quantity of interest can be summarized by a probability interval, with one or two sides.

EFSA recommends using combinations of quantitative and verbal expressions of uncertainty (EFSA et al., 2019) to enhance understanding of the uncertainty being communicated. A message communicating uncertainty with quantitative and verbal expressions can be: "we are 95% certain (i.e. with a probability equal to 95%) that the weekly intake is exceeding the safety threshold for 40 out of 100 000 children in the target population." The results of bounded probability can be communicated in a similar way as "we are at least 95% certain (i.e., with a probability greater or equal to 95%) that the weekly intake is exceeding the safety threshold for less than 980 out of 100 000 children in the target population."

Whether to use bounded or precise probability will depend on a variety of concerns, including the nature of the information at hand, the amount of risk involved, the purpose of the analysis, and possibly also computational requirements. Precise probability is a special case of bounded probability, which when the modeling is performed well, always falls between the lower and upper bounds of the probabilities characterizing uncertainty in the quantity of interest. Methods of statistical (Bayesian) inference and expert elicitation resulting in uncertainty quantified by precise probability are well established, and useful for assessment as long as the assessor is confident about her (un)certainty about the conclusion. In practice, it can be easier for an expert to provide, or for experts to agree on, a bounded compared to a precise probability. Bounded probabilities can be used in a coarse uncertainty analysis to evaluate the need for a refined approach. We have shown that, when expert knowledge is provided with imprecision, it is still possible to do statistical (robust Bayesian) inference and propagate uncertainty without being limited to quantifying uncertainty by precise probability.

A subjective probability (precise or approximate) is an unambiguous expression for uncertainty (EFSA et al., 2018; van der Bles et al., 2019). This can be accompanied by judgments about the strength of knowledge supporting the assessment (Aven, 2020; van der Bles et al., 2019) to support the final characterization of overall uncertainty about the conclusion. In any assessment, the usefulness of subjective probability for quantifying epistemic uncertainty in risk assessment depends on having specified clear assessment questions and an assessment model that can consider aleatory uncertainty and integrate evidence from data and expert knowledge. These requirements hold for any quantitative measure of epistemic uncertainty.

SUPPLEMENTARY MATERIAL

The code to produce the assessment and generate graphs in this article is available as an R-package at https://github.com/Iraices/PrecisePvsBoundedP.

ACKNOWLEDGMENTS

This work was supported by information from Thomas Schendel on the assessment. Fundings from the Swedish Research Council FORMAS through the project "Scaling up uncertain environmental evidence" (219-2013-1271) and the strategic research areas BECC (Biodiversity and Ecosystem Services in a Changing Climate) are appreciated. The authors would like to thank Andy Hart and Scott Ferson for helpful comments on early drafts of the manuscript and three anonymous reviewers for their valuable comments and interesting debates. The authors declare that no competing interests exist.

REFERENCES

Aldenberg, T., & Jaworska, J. S. (2000). Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicology and Environmental Safety*, *46*(1), 1–18.

Apostolakis, G. (1990). The concept of probability in safety assessment of technological systems. *Science*, *250*(4986), 1359–1364.

Aven, T. (2010). On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, *30*(3), 354–360.

Aven, T. (2020). Bayesian analysis: Critical issues related to its scope and boundaries in a risk context. *Reliability Engineering & System Safety*, *204*, 107209.

Aven, T. (2021). Further reflections on EFSA's work on uncertainty in scientific assessments. *Journal of Risk Research*, *24*(5), 553–561.

Aven, T., & Reniers, G. (2013). How to define and interpret a probability in a risk and safety setting. *Safety Science*, *51*(1), 223–231.

Benavoli, A., & Ristic, B. (2011). Classification with imprecise likelihoods: A comparison of TBM, random set and imprecise probability approach. In *14th International Conference on Information Fusion, Chicago, IL* (pp. 1–8).

Berger, J. O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, *25*, 303–328.

Cattaneo, M. E., & Wiencierz, A. (2012). Likelihood-based imprecise regression. *International Journal of Approximate Reasoning*, *53*(8), 1137–1154.

Cox, D. (2006). *Principles of statistical inference*. Cambridge: Cambridge University Press.

Daimon, T. (2008). Predictive checking for Bayesian interim analyses in clinical trials. *Contemporary Clinical Trials*, *29*(5), 740–750.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *7*(1), 1–68.

EFSA. (2014). Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA Journal*, *12*(6), 278.

EFSA, Hart, A., Maxim, L., Siegrist, M., Goetz, N. V., da Cruz, C., Merten, C., Mosbach-Schulz, O., Lahaniatis, M., Smith, A., & Hardy, A. (2019). Guidance on communication of uncertainty in scientific assessments. *EFSA Journal*, *17*(1), e05520.

EFSA Scientific Committee, Benford, D., Halldorsson, T., Jeger, M. J., Knutsen, H. K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J. R., Silano, V., Solecki, R., Turck, D., Younes, M., Craig, P., Hart, A., … Hardy, A. (2018). Scientific opinion on the principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. *EFSA Journal*, *16*(1), e05122.

FAO, & WHO. (2021). *Microbiological risk assessment—Guidance for food*. Microbiological Risk Assessment Series No. 36, Rome.

Ferson, S., & Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, *54*(2), 133–144.

Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. & Sentz, K. (2003). Constructing probability boxes and Dempster-Shafer structures. Technical report, Sandia National Laboratories.

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 13664–13671.

Flage, R., Aven, T., & Berner, C. L. (2018). A comparison between a probability bounds analysis and a subjective probability approach to express epistemic uncertainties in a risk assessment context: – A simple illustrative example. *Reliability Engineering & System Safety*, *169*, 1–10.

Gelman, A. (2008). Objections to Bayesian statistics. *International Society for Bayesian Analysis*, *3*(3), 445–450.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). New York: Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Goerlandt, F., & Montewka, J. (2015). A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a ship-ship collision. *Safety Science*, *76*, 42–66.

Goldstein, M. (2006). Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, *1*(3), 403–420.

Gosling, J. P. (2018). The Sheffield Elicitation framework. Luis C. Dias, Alec Morton & John Quigley, *Elicitation: The Science and Art of Structuring Judgement* (pp. 61–93). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65052-4_4

Hájek, A. (2019). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2019 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2019/entries/probability-interpret

Hampton, J., Moore, P., & Thomas, H. (1973). Subjective probability and its measurement. *Journal of the Royal Statistical Society. Series A (General)*, *136*(1), 21–42.

Helton, J. C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation*, *57*(1-4), 3–76.

Helton, J. C., & Oberkampf, W. L. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety*, *85*(1-3), 1–10.

Institute of Medicine. (2013). *Environmental decisions in the face of uncertainty*. Washington, DC: The National Academies Press.

Kadane, J. B. (2011). *Principles of uncertainty*. Chapman & Hall/CRC Texts in Statistical Science Series: 92. Boca Raton, FL: Chapman & Hall/CRC.

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, *1*(1), 11–27.

Kelly, D., & Smith, C. (2011). *Bayesian inference for probabilistic risk assessment: A Practitioner's Guidebook*. London: Springer-Verlag.

Kelly, D. L., & Smith, C. L. (2009). Bayesian inference in probabilistic risk assessment: The current state of the art. *Reliability Engineering & System Safety*, *94*(2), 628–643.

Lindley, D. V. (2006). *Understanding uncertainty*. Hoboken, NJ: Wiley.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press/Taylor & Francis Group.

Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology*, *57*(1), 9–18.

Oakley, J. (2020). *SHELF: Tools to support the Sheffield elicitation framework*. R Package Version 1.7.0.

Oakley, J., & O'Hagan, T. (2020). *SHELF: The Sheffield elicitation framework (version 4.0)*. School of Mathematics and Statistics, University of Sheffield, UK.

O'Hagan, A. (2013). *Probability: Methods and measurement*. Dordrecht: Springer Science & Business Media.

O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *American Statistician*, *73*(sup1), 69–81.

O'Hagan, A., Buck, C., Daneshkhah, A., & Eiser, J. (2006). *Uncertain judgments: Eliciting experts' probabilities*. Chichester: Wiley.

Roos, M., Martins, T. G., Held, L., & Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis*, *10*(2), 321–349.

Sahlin, U., Troffaes, M. C. M., & Edsman, L. (2021). Robust decision analysis under severe uncertainty and ambiguous tradeoffs: An invasive species case study. *Risk Analysis*, *41*(11), 2140–2153.

SAPEA (2019). Making sense of science for policy under conditions of complexity and uncertainty. Berlin: SAPEA. https://doi.org/10.26356/MASOS

Schad, D. J., Betancourt, M., & Vasishth, S. (2019). Toward a principled Bayesian workflow in cognitive science. arXiv:1904.12765 [stat.ME].

Schendel, T., Jung, C., Lindtner, O., & Greiner, M. (2018). Guidelines for uncertainty analysis: Application of the respective documents of EFSA and BfR for exposure assessments. *EFSA Journal*, *15*(7), 1–113.

Singpurwalla, N. D., & Wilson, A. G. (2008). Probability, chance and the probability of chance. *IIE Transactions*, *41*(1), 12–22.

Troffaes, M. C., & Cooman, G. D. (2014). *Lower previsions*. Wiley series in probability and statistics. Chichester: Wiley.

Troffaes, M. C. M. (2017). A note on imprecise Monte Carlo over credal sets via importance sampling. In *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*.

Troffaes, M. C. M. (2018). Imprecise Monte Carlo simulation and iterative importance sampling for the estimation of lower previsions. *International Journal of Approximate Reasoning*, *101*, 31–48.

US EPA (2011). Exposure Factors Handbook (Final Report): Chapter 2. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-09/052F, 2011.

van derBles, A. M., van der Linden, S., Freeman, A. L., Mitchell, J., Ana, B., Galvao, L. Z., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, *6*, 181870.

Varadhan, R., & Borchers, H. W. (2018). *Derivative-free optimization*. R Package Version 2018.2-1.

Vose, D. (1996). *Quantitative risk analysis: A guide to Monte Carlo simulation modelling*. New York: Wiley.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman & Hall.

Walter, G., & Coolen, F. P. (2016). Sets of priors reflecting prior-data conflict and agreement. In *Proceedings of the 16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*.

How to cite this article: Raices Cruz, I., Troffaes, M.C.M., Sahlin, U. (2021). A suggestion for the quantification of precise and bounded probability to quantify epistemic uncertainty in scientific assessments. *Reliability Engineering & System Safety*, 1–15. https://doi.org/10.1111/risa.13871

APPENDIX A: NORMAL-GAMMA MODEL

Let $\mathbf{x} := (x_1, \ldots, x_n) \mid \mu, \tau \sim \mathcal{N}(\mu, \tau)$ with unknown mean (μ) and precision (τ) and the prior distribution for parameters (μ, τ) has a normal-gamma distribution

$$\tau \sim \text{Gamma}(\alpha_0, \beta_0), \tag{A1}$$

$$\mu \mid \tau \sim \mathcal{N}(\gamma_0, \delta_0 \tau). \tag{A2}$$

Then, due to conjugacy, the posterior distribution is also a normal-gamma distribution

$$\tau \mid \mathbf{x} \sim \text{Gamma}(\alpha, \beta), \tag{A3}$$

$$\mu \mid \tau, \mathbf{x} \sim \mathcal{N}(\gamma, \delta\tau), \tag{A4}$$

where

$$\alpha := \alpha_0 + \frac{n}{2}, \tag{A5}$$

$$\beta := \beta_0 + \frac{1}{2}(n-1)S^2 + \frac{n\delta_0}{\delta_0 + n}\frac{(\bar{x} - \gamma_0)^2}{2}, \tag{A6}$$

$$\gamma := \frac{\delta_0 \gamma_0 + n\bar{x}}{\delta_0 + n}, \tag{A7}$$

$$\delta := \delta_0 + n, \tag{A8}$$

where $\bar{x}$ is the sample mean and S^2 is the sample variance of $\mathbf{x}$.

APPENDIX B: BETA MODEL

Let $B \sim \text{Bernoulli}(\pi)$ a binary random variable that indicates the occurrence of an event taking the value 1 and 0 otherwise. The prior distribution for parameter π follows a Beta distribution

$$\pi \sim \text{Beta}(\alpha_0'', \beta_0''), \tag{B1}$$

where α_0'' and β_0'' are hyperparameters for this prior. Due to conjugacy, the posterior distribution for this parameter is also Beta distribution

$$\pi \mid \mathbf{b} \sim \text{Beta}(\alpha'', \beta''), \tag{B2}$$

where $\mathbf{b} := (b_1, \ldots, b_n)$ and

$$\alpha'' := \alpha_0'' + \sum_{i=1}^{n} b_i, \tag{B3}$$

$$\beta'' := \beta_0'' + n - \sum_{i=1}^{n} b_i. \tag{B4}$$

APPENDIX C: SET OF PRIOR DISTRIBUTIONS

The set of prior hyperparameters T_0, is defined as

$$\text{T}_0(\gamma_0^A, \gamma_0^C) = \left\{ \begin{array}{c} t_0(\gamma_0^A, \gamma_0^C) := (\alpha_0^A, \beta_0^A, \gamma_0^A, \delta_0^A, \alpha_0^C, \beta_0^C, \gamma_0^C, \delta_0^C, \alpha_0^B, \beta_0^B) : \\ \underline{\gamma_0^A} \le \gamma_0^A \le \overline{\gamma_0^A}, \\ \underline{\gamma_0^C} \le \gamma_0^C \le \overline{\gamma_0^C}. \end{array} \right\}, \tag{C1}$$

where $\alpha_0^A = 1$, $\beta_0^A = 1$, $\underline{\gamma_0^A} = 1$, $\overline{\gamma_0^A} = 6$, $\delta_0^A = 5$, $\alpha_0^C = 1$, $\beta_0^C = 1$, $\underline{\gamma_0^C} = -5$, $\overline{\gamma_0^C} = 1$, $\delta_0^C = 5$, $\alpha_0^B = 1$, $\beta_0^B = 1$.