Regulatory Toxicology and Pharmacology 80 (2016) 9–24

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Bayesian methods for uncertainty factor application for derivation of reference values

Ted W. Simon [a, *], Yiliang Zhu [b], Michael L. Dourson [c], Nancy B. Beck [d]

[a] Ted Simon LLC, 4184 Johnston Road, Winston, GA 30187, United States
[b] Department of Epidemiology and Biostatistics, University of South Florida, 13201 Bruce B. Downs, MDC56, Tampa, FL 33612, United States
[c] Toxicology Excellence for Risk Assessment (TERA) Center, University of Cincinnati, College of Medicine, 160 Panzeca Way, Cincinnati, OH 45267-0056, United States
[d] Regulatory Science Policy, Regulatory and Technical Affairs, American Chemistry Council, 700 2nd Street NE, Washington, DC 20002, United States

ARTICLE INFO

Article history:
Received 10 November 2015
Received in revised form
3 March 2016
Accepted 16 May 2016
Available online 19 May 2016

Keywords:
Uncertainty factor
Bayesian methods
Reference dose
Reference concentration
Risk assessment
Hazard assessment

ABSTRACT

In 2014, the National Research Council (NRC) published *Review of EPA's Integrated Risk Information System (IRIS) Process* that considers methods EPA uses for developing toxicity criteria for non-carcinogens. These criteria are the Reference Dose (RfD) for oral exposure and Reference Concentration (RfC) for inhalation exposure. The NRC Review suggested using Bayesian methods for application of uncertainty factors (UFs) to adjust the point of departure dose or concentration to a level considered to be without adverse effects for the human population. The NRC foresaw Bayesian methods would be potentially useful for combining toxicity data from disparate sources—high throughput assays, animal testing, and observational epidemiology. UFs represent five distinct areas for which both adjustment and consideration of uncertainty may be needed. NRC suggested UFs could be represented as Bayesian prior distributions, illustrated the use of a log-normal distribution to represent the composite UF, and combined this distribution with a log-normal distribution representing uncertainty in the point of departure (POD) to reflect the overall uncertainty. Here, we explore these suggestions and present a refinement of the methodology suggested by NRC that considers each individual UF as a distribution. From an examination of 24 evaluations from EPA's IRIS program, when individual UFs were represented using this approach, the geometric mean fold change in the value of the RfD or RfC increased from 3 to over 30, depending on the number of individual UFs used and the sophistication of the assessment. We present example calculations and recommendations for implementing the refined NRC methodology.

1. Introduction

Uncertainty factors (UFs) were developed in the 1980s by U.S. Environmental Protection Agency (USEPA) scientists based on margins of safety for determining acceptable daily intakes (ADIs) (Dourson and Stara, 1983; Barnes and Dourson, 1988; Dourson, 1993, 1996; Dourson et al., 1996). The development and use of uncertainty factors comprise an attempt to address the lack of specificity in margins of safety and are designed to address specific areas of uncertainty, thus enabling the development of data-derived values to replace default values of generally 10-fold (Dourson et al., 1996). The goal for any toxicity guidance value such as the Reference Dose (RfD), Reference Concentration (RfC) or Tolerable/Acceptable Daily Intake (TDI/ADI) is not only protection of human health consistent with the societal consensus for such protection but also avoidance of an overprotective level that could conceivably lead to excessive regulation (Simon, 2011). This balance notwithstanding, the needs of regulation are immediate and these exigencies are the basis for the continued regulatory embrace of default values for UFs and their use in the derivation of reference values (RfVs).

The individual UFs used in EPA toxicity assessments address five distinct areas of uncertainty. Historical publications by EPA staff in the 1980s provide much of the basis for four of the UFs, excluding UF-D, applied for database deficiencies (Dourson and Stara, 1983; Barnes and Dourson, 1988; Dourson and DeRosa, 1991). Subsequent publications introduced the basis for this latter database factor, generally the absence of evidence regarding developmental

* Corresponding author.
E-mail address: ted@tedsimon-toxicology.com (T.W. Simon).

http://dx.doi.org/10.1016/j.yrtph.2016.05.018

Abbreviations

ADI	Acceptable daily intake
BMD	Benchmark dose
DDEF	Data-derive extrapolation factor
NRC	National Research Council
POD	Point of departure
RfD	Reference dose
RfC	Reference concentration
RfV	Reference value
PBPK	physiologically-based pharmacokinetic
TDI/ADI	Tolerable/Acceptable daily intake
UFs	Uncertainty factors
UF/EF	uncertainty/extrapolation factor
LOAEL:	Lowest observed no adverse effect level
NOAEL	No observed adverse effect level
UF-L:	LOAEL-to-NOAEL uncertainty/extrapolation factor
UF-S	Subchronic-to-chronic uncertainty/extrapolation factor
UF-A	interspecies uncertainty/extrapolation factor
UF-A-TD	toxicodynamic component of the interspecies uncertainty/extrapolation factor
UF-A-TK	toxicokinetic component of the interspecies uncertainty/extrapolation factor
UF-H	intraspecies uncertainty/extrapolation factor
UF-H-TD	toxicodynamic component of the intraspecies uncertainty/extrapolation factor
UF-H-TK	toxicokinetic component of the intraspecies uncertainty/extrapolation factor
UF-D	uncertainty/extrapolation factor for database deficiencies

and reproductive toxicity (DART) (Dourson et al., 1992, 1996; Dourson, 1993). All five areas of uncertainty are discussed in USEPA's *Review of the Reference Dose and Reference Concentration Processes* (USEPA, 2002b). The purposes of the individual UFs were to address these five areas of uncertainty and, according to this document, were:

> *… (1) the variation in sensitivity among the members of the human population (i.e., inter-individual variability); (2) the uncertainty in extrapolating animal data to humans (i.e., interspecies uncertainty); (3) the uncertainty in extrapolating from data obtained in a study with less-than-lifetime exposure to lifetime exposure (i.e., extrapolating from subchronic to chronic exposure); (4) the uncertainty in extrapolating from a LOAEL rather than from a NOAEL; and (5) the uncertainty associated with extrapolation when the database is incomplete*. (USEPA, 2002b)

Before the publication of EPA's 2002 document, any uncertainties not explicitly addressed by the five different UFs had been addressed by the use of a modifying factor (MF). However, the EPA 2002 document recommended discontinuation of the use of MFs (USEPA, 2002b). In IRIS assessments developed today, all five individual UFs are multiplied together and the composite UF applied to the point of departure (POD) by arithmetic division, generally as the final step in the RfV development process.

Over time, the understanding of UFs as individual factors rather than their combination has continued to grow. Each individual UF consists of an adjustment and the uncertainty associated with the adjustment; here, we identify the central value of a UF distribution as a measure of adjustment and the variance as a measure of uncertainty. In many of the IRIS derivations considered here, EPA has chosen the UFs for specific reasons. Even Lehman and Fitzhugh (1954) recognized their composite 100-fold UF was intended to deal with several distinct areas of uncertainty (Dourson and Stara, 1983). However, the encoding of distinct areas of uncertainty as individual factors rather than as an overall or composite factor is an important distinction and is not fully recognized in NRC (2014).

The use of an overall composite UF or "safety factor" masks the compounding conservatism inherent in the use of several UFs set at default values and each intended to provide a highly protective toxicity value (Burmaster and Harris, 1993; Burmaster and Anderson, 1994; Cullen, 1994; Simon, 2011; Tatum et al., 2015). The compounded conservatism in the use of many high-end values will yield an overestimate of risk and the actual risk is likely to be much lower or even non-existent. Indeed, a highly conservative policy-based assessment seems at odds with principles of transparency and the use of science as a basis for societal decision-making (Dourson and Stara, 1983; Lewis et al., 1990).

Although we use the familiar abbreviation UF in this paper, these factors are also called extrapolation or adjustment factors and, ideally, their values, whether chemical-specific or default, will be based upon actual data (e.g., WHO-IPCS, 2005, 2014; Chiu and Slob, 2015; USEPA, 2014).

The conceptual basis of the application of UFs using the standard deviations of Bayesian prior distributions is described in the recent *Review of EPA's Integrated Risk Information System (IRIS) Process* from the National Research Council (NRC, 2014). Here we consider the NRC methodology in terms of both the mean and variance of these distributions, provide several illustrations of this application, and explore ways that these methods could be applied currently to the development of RfVs within the IRIS program or in other similar hazard assessment programs in public and private sectors. The World Health Organization International Programme on Chemical Safety (WHO-IPCS) recently released the *Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization* that also endorsed probabilistic approaches; the methods and practices described here are consistent with both the NRC report and the WHO guidance (WHO-IPCS, 2014; Chiu and Slob, 2015). We also provide a brief narrative on the considerations and best practices for the use of Bayesian methods for development of quantitative uncertainty estimates in RfVs that could be put into practice immediately.

1.1. Chemical-specific adjustment factors and data-derived uncertainty factors

In 2005, the WHO-IPCS released *Chemical-Specific Adjustment Factors for Interspecies Differences and Human Variability*: *Guidance Document for Use of Data in Dose/Concentration Assessment* (WHO-IPCS, 2005). In this guidance, UFs are called chemical-specific adjustment factors (CSAFs). Over a decade in development, this document was introduced to provide methods for the incorporation of quantitative data on toxicokinetics or toxicodynamics into the development of RfV/TDI values by modifying the default value of 10 for each CSAF. Since 1994, Health Canada has been using a data-derived procedure based on the developing WHO-IPCS guidelines (Meek et al., 1994). In 2014, the U.S. Environmental Protection Agency issued a similar document, *Guidance for Applying*

Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation (USEPA, 2014).

1.2. Bayesian and probabilistic approaches to UFs

In *Review of EPA's Integrated Risk Information System (IRIS) Process* (Chapter 7), the NRC outlines a Bayes approach to integrating UFs in deriving toxicity values as well as integrating several studies or data from diverse sources (NRC, 2014). The NRC's approach assumes that the uncertainty in the estimated point of departure (e.g., benchmark dose) as well as that encoded by the UFs can be characterized using appropriately specified distributions and uses the log-normal distribution for illustration. NRC (2014) was explicit in promoting systematic approaches to incorporating quantitative uncertainties into toxicity value development and foresaw that toxicity value development would likely rely on a range of disparate data sources, including high throughput *in vitro* assays, animal testing, and/or epidemiological studies.

The general approach as described in the NRC report differs from the classic UF approach in two aspects: first, uncertainty from a specific source can be characterized with greater flexibility and representativeness as a distribution or range of values rather than as a point estimate; and, second, the use of a distribution provides a system that can accommodate a variety of types of data to refine existing toxicity factors. In the past, a number of authors have characterized UFs using the log-normal distribution, likely because of its mathematical tractability (Price et al., 1997; Swartout et al., 1998; Gaylor and Kodell, 2000).

Perhaps more important, any distribution can be characterized by parameters that specify the location and scale—for the log-normal distribution, these would be the natural logarithms of the geometric mean and geometric standard deviation, here symbolized as μ and σ respectively. Because UFs are intended to address both a location adjustment and the attendant uncertainty, both the location and scale of the UF need to be specified. For those UFs not needing an adjustment, the location will be zero (i.e. geometric mean $= 1$ or $\mu = 0$ for a log-normal distribution) indicating the factor addresses uncertainty only. For other UFs, such as UF-L and UF-S, the extrapolation from a LOAEL to a NOAEL will require some degree of numerical adjustment as well as an expression of uncertainty and the same idea will hold true when extrapolating from a sub-chronic to chronic duration from animal bioassay data. This idea of adjustment and uncertainty will be discussed throughout the rest of this paper.

1.3. Historical context for the NRC IRIS review

Formaldehyde became a high priority public health issue after its discovery in FEMA trailers sent to the Gulf Coast as shelters for Hurricane Katrina victims (Jacobs, 2011). Hence, EPA requested that the NRC review the draft IRIS assessment for formaldehyde; the resulting review was highly critical of the process used, pointing out the assessment both lacked transparency and consistency and did not provide sufficient documentation of methods or the characterization of uncertainty and variability (NRC, 2011, p. 26). The final chapter of the NRC review (Chapter 7) provided a general critique of the IRIS process and suggestions for revising that process (NRC, 2011).

Following the NRC formaldehyde review, the U.S. Congress held several hearings to examine EPA's IRIS program and to address concerns raised regarding EPA's overall IRIS process. The result was that the U.S. Congress requested that NRC assess the existing and planned changes in IRIS, and in 2014, NRC released their review.

NRC's forward-looking approach sought to provide a framework for the inclusion of data from multiple sources such as high throughput *in vitro* assays, animal testing and epidemiological observations in humans. Such an approach would easily accommodate newer types of data to refine existing RfVs. However, what is missing from the NRC report is a discussion of ways to implement these new methods into current IRIS assessments—hence, a primary purpose of this paper is to provide such a discussion with examples as a proof of concept and suggestions for best practices when implementing these methods.

1.4. The critical effect—the basis of the reference value

To understand how NRC's suggestions differ from EPA's current practice, this section provides a brief discussion of the concept of the critical effect as the most appropriate biological basis of the RfV. EPA defines the critical effect as "the first adverse effect or its known precursor, that occurs in the most sensitive species as the dose rate of an agent increases" (USEPA, 2011). The critical effect is one effect in a spectrum of biological changes from adaptive responses to frank effects.

At low doses, biological responses occur in a continuum and are most often adaptive responses. These adaptive responses enhance the organism's ability to withstand a challenge. As dose increases, compensatory effects occur, enabling the organism to maintain overall function with neither further enhancement of these effects nor significant harm. As dose increases further, the critical effect and/or downstream adverse effects occur as functional impairments or pathologic lesions that incur significant cost or harm. At some point these adverse effects become irreversible and overtly manifest as disease (Patlewicz et al., 2013).

Generally, the current practice in EPA's IRIS program is to select a single critical study and single critical effect upon which to base the point of departure (POD) for an RfV value. In contrast, NRC (2014) suggested that a meta-analytic approach be used to combine the results of multiple studies. NRC (2014) also provided a few examples of combining studies using such meta-analytic approaches and the application of UFs using Bayesian methods. Categorical regression can also be used to combine results of multiple studies, including studies of effects of different severities with methods developed by EPA (Hertzberg and Dourson, 1993; Dourson et al., 1997; Teuschler et al., 1999).

2. Uncertainty factors as distributions

Historically, the general assumption within the risk assessment community has been that UFs could be represented by a range of generally up to 10-fold (Felter and Dourson, 1998) Although the understanding of the biology underlying the choices of UF values has grown, some imprecision remains (e.g., WHO-IPCS, 2014; Chiu and Slob, 2015; Bokkers and Slob, 2007). Most ranges can be characterized as probability distributions and an uncertainty factor can be assumed to follow a distribution with a specific range of variation. Here, as does NRC (2014), we also characterize the distribution of an uncertainty factor as log-normal with the underlying normal distribution on a natural log scale with a mean denoted by μ and a standard deviation denoted by σ.

For example, Swartout et al. (1998) suggested a 3-parameter log-normal as a "reference" UF distribution with a 50^{th} percentile of 3.16, a 95th percentile of 10 and a lower limit or offset parameter, $\tau = 1$. For such a distribution, all adjustments will be greater than one. These authors give the base ten logarithms of the offset-adjusted median, i.e., $\log_{10}$ (median $-$ offset), with a value of 0.335 and the standard deviation with a value of 0.377. Using natural logarithms, the corresponding values would be $\mu = 0.770$ and $\sigma = 0.867$ and following a selection of a value using these parameters, one would also need to exponentiate the selected value and

add the value of the offset parameter, $\tau = 1$.

Attempts have been made to determine the range of human variability and thus "groundtruth" the intraspecies UF (Dourson et al., 1996; Renwick and Lazarus, 1998). Both EPA's recent guidance on data-derived extrapolation factors (DDEFs) and the WHO-IPCS guidance on CSAFs are silent on the nature of the distribution, although the WHO document mentions log-normal distributions (WHO-IPCS, 2005; USEPA, 2014).

Swartout et al. (1998) likely chose the log-normal distribution more for computational tractability and flexibility rather than for the fidelity with which this distribution represents quantitative adjustment and uncertainty. This point is examined further elsewhere in this paper.

When NOAELs or LOAELs are used as PODs, their value is unavoidably skewed by study design and we recognize that the BMD as a point estimate of the POD is also imprecise. Toxicological responses follow the laws of thermodynamics and logarithms are the most appropriate expression of dose (Clark, 1933; Bruhn et al., 2003; Waddell, 2008, 2010; Hughes and Aronson, 2010). When the sample size of the underlying study is sufficiently large, by the central limit theorem, the maximum likelihood estimator of the POD will approximate a normal distribution; this normal approximation is also the case for this estimator of the POD on a logarithmic scale. Further, the use of lognormal distributions would be more robust against variation skewed toward the heavy right tail when the sample size is small to moderate. Zhu et al. (2007) conducted an extensive bootstrap simulation of four developmental studies that represent the typical dose-response shapes observed in generation II developmental studies conducted by NTP. The distributions of the BMD distributions at BMR levels of both 1% and 5% BMR level are negatively skewed to a considerable extent (Table IV in Zhu et al., 2007). Hence, using normal distributions of the BMD is not advisable in practice because substantial bias would occur in estimating BMDLs (Zhu et al., 2007; USEPA, 2012a).

These explanations notwithstanding, the true form of the distribution of the BMD remains unknown and is likely case-specific. Hence, similar to the NRC and others in the history of RfV development, we acknowledge this imprecision and also chose to represent both the POD and the UFs with log-normal distributions for illustration.

2.1. Adjustment vs. accounting for uncertainty

UFs, as traditionally applied, seek to provide a "reasonable maximum adjustment" both for the necessary extrapolations and for developing a protective value in the face of uncertainty (Dourson and Stara, 1983; Barnes and Dourson, 1988). Hence, any choice of the form of the distribution of uncertainty needs to consider both adjustment and uncertainty. The central value of the distribution is a reasonable choice for numerical adjustment whereas the variance is likely a representation of uncertainty. The central values will be different for different UFs and will affect the choice of a value for the variance to provide adequate coverage.

For example, even if the value of a NOAEL cannot be established from a study, by definition, the value will be less than the measured LOAEL. Hence, the central value of any distribution representing the LOAEL-to-NOAEL UF (UF-L) will be greater than unity and will constitute an adjustment and the variance will be an estimate of the attendant uncertainty.

As noted, a chronic point of departure will be lower than a subchronic POD, and similar to the case of UF-L, the central value of UF-S represents the central value of this adjustment for dosing duration and the variance represents the surrounding uncertainty.

The factors for animal-to-human extrapolation (UF-A) and human variability (UF-H), even when split into toxicokinetic and toxicodynamic components, also contain both elements of adjustment and uncertainty (e.g., Table 2).

Even the factor for database deficiencies assumes that missing toxicity data will necessarily represent an endpoint that, if tested, would result in a lower RfV. The database for a given chemical may not include testing for developmental and reproductive toxicity (DART) and adjustment is deemed necessary because of concern for this type of endpoint (Dourson et al., 1992, 1996; USEPA, 2002b; Blackburn et al., 2015).

2.2. Sources of uncertainty represented by the five uncertainty factors

As previously mentioned, five distinct areas of uncertainty are addressed by the UFs used in EPA assessments. Separating these five types of uncertainty was not addressed in NRC (2014). The use of UFs for human variability and interspecies extrapolation was developed from work in the 1950s of Lehmann and Fitzhugh of the Food and Drug Administration as well as others (Lehman and Fitzhugh, 1954). At that time, a 100-fold UF was proposed to account for interspecies differences and human variability with the latter including allowance for sensitive subpopulations and, to a lesser extent, possible synergistic effects with other chemicals. Data regarding inter- and intra-species differences were considered and the likely intent was that eventually a data-derived probabilistic method would be used in lieu of the 100-fold default. In 1980, EPA recommended an additional UF between 1 and 10 in cases when an ADI value was estimated from a LOAEL rather than a NOAEL (Dourson and Stara, 1983).

Renwick and Lazarus (1998) provide historical details of the separation of the inter- and intraspecies UFs into toxicokinetic and toxicodynamic components. WHO-IPCS (2005) indicates the default interspecies UF should be split into factors of 2.5 and 4 for toxicodynamics and toxicokinetics respectively whereas the intraspecies UF should be split evenly into default components of $\sqrt{10}$ (i.e., 3.16). USEPA (2014) suggests that both default inter- and intraspecies factors can be split evenly as $\sqrt{10}$. Generally, this value is rounded to 3. In addition, when supporting data exist, EPA (2014) indicates that both the toxicokinetic and toxicodynamic components of the interspecies UF can have a value of less than one (USEPA, 2014).

Using the traditional method for UF application by dividing into the POD, when uncertainties are identified from distinct sources, individual UFs should be specified and generally applied at the appropriate stage of RfV development to ensure the transparency and specificity of the process (e.g., Simon et al., 2008, 2009). Using only a composite UF applied at the end of the process diminishes this transparency. Generally, this same principle should be applied when using Bayesian methods to individual UFs singly and at different stages within the RfV process, based on the nature of the adjustment as shown in Fig. 1. The issue of when to apply particular individual UFs will be examined further in the DISCUSSION.

The choice of numerical values for μ and σ for log-normal UF distributions allows one to decide both the degree of adjustment and uncertainty and whether to remain consistent with default uncertainty values or to choose a different value—by specifying the percentile corresponding to the degree of uncertainty. Table 1 provides a range of UF values for a generic log-normal UF distribution; the value of μ is zero and the table shows the values of the Bayesian prior standard deviations on a natural log scale for a range of percentiles corresponding to one-sided confidence intervals. The table is similar to Table 7-2 in NRC (2014). The highest UF value in the NRC table is 1000 whereas the highest UF value shown in Table 1 is 10—the reason is that we believe these methods should be applied to individual UFs/EFs rather than to the composite UF,

Table 1
Standard deviations of the log-normal distributions used for application of individual uncertainty/adjustment/extrapolation factors with geometric means of one, UF-A and UF-H. The value of μ is set equal to zero, under the assumption that no adjustments are needed. Hence, these UF values will be applicable to UF-A and UF-H.

Value of UF	20th percentile or 80th percentile	10th percentile or 90th percentile	5th percentile or 95th percentile	2.5th percentile or 97.5th percentile	0.5th percentile or 99.5th percentile
Z-score	0.842	1.282	1.645	1.960	2.576
	Standard deviation values for UFs > 1 below are added when combined				
2	0.823	0.541	0.421	0.354	0.269
2.5	1.089	0.715	0.557	0.468	0.356
3	1.305	0.857	0.668	0.561	0.427
3.16	1.367	0.898	0.699	0.587	0.447
10	2.736	1.797	1.400	1.175	0.894

Percentile of the left tail (e.g. 5-percentile) applies to UF < 1; percentile of the right applies to UF > 1.
Confidence level is equal to (100 − left) percentile for one-sided values and (100−2 × left percentile) for two-sided values.
The actual UF value corresponds to exp(Z-score * Table Value); hence, for a UF of 10 corresponding to the 95th percentile, UF = exp(0 + 1.645*1.4) = 10.

Table 2
Means and Standard Deviations of log-normal distributions used for UF-A when humans are known to be less sensitive than animals. To use these UFs, the value of μ would be added to the POD and the value of σ combined with the other UFs as shown in Eq. (1b). A value of μ = ln (0.3) assumes that humans are known to be about 1/3 as sensitive as animals. The value of the standard deviation are used to obtain a UF value equivalent to that shown in the leftmost column. Hence, the data might show humans are 1/3 as sensitive but to account for uncertainty, humans and animals are considered equally sensitive. In the bottom two rows, humans are known to be 1/10 as sensitive as animals and to account for uncertainty, humans are considered either 1/3 as sensitive or equally sensitive (Value of UF = 0.3 or 1 respectively).

Value of UF	20th percentile or 80th percentile	10th percentile or 90th percentile	5th percentile or 95th percentile	2.5th percentile or 97.5th percentile	0.5th percentile or 99.5th percentile
Z-score	0.842	1.282	1.645	1.960	2.576
	Standard deviation values for UFs < 1 are subtracted when combined, assuming μ = ln (0.3)				
1	1.431	0.939	0.732	0.614	0.467
	Standard deviation values for UFs < 1 are subtracted when combined, assuming μ = ln (0.1)				
0.3	1.306	0.858	0.669	0.561	0.428
1	2.736	1.797	1.400	1.175	0.894

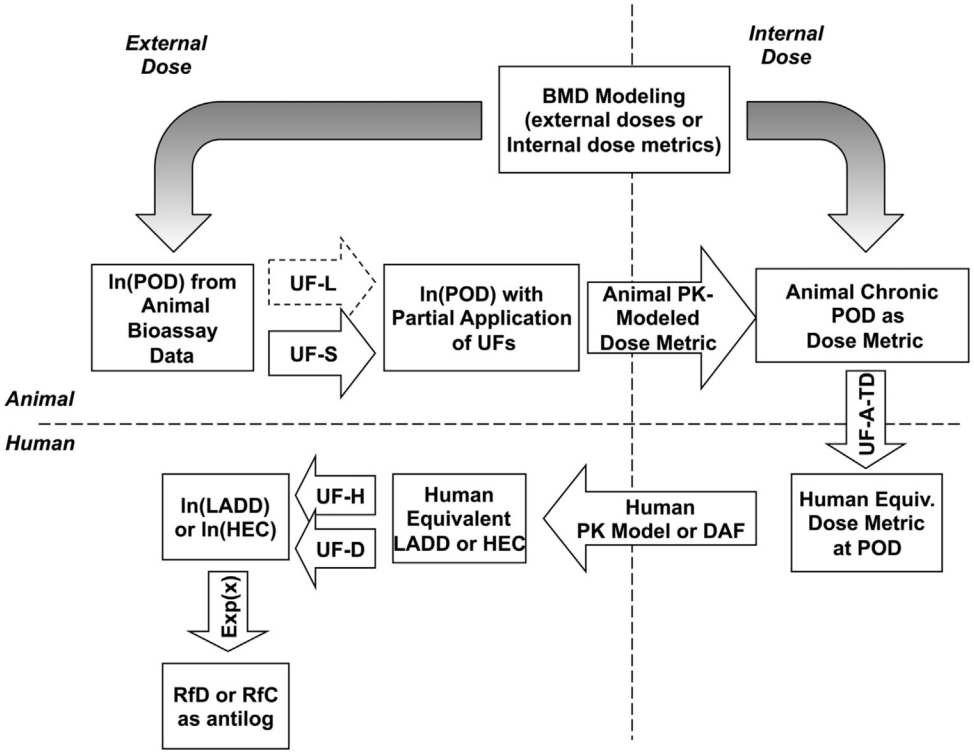

Fig. 1. Scheme for application of UFs at the appropriate stage of the RfD process. Following BMD modeling in animals, the internal dose metric may be adjusted for with UF-L and UF-S. The animal dose metric at the POD is adjusted with the UF-A-TD to obtain the human internal dose metric at the POD. This human POD is adjusted with the UF-H-TK to obtain the POD in sensitive humans. The human PBPK model is used to obtain external doses and the UF-D is applied if necessary.

the product of individual UFs. In addition, the use of a distribution for UFs necessitates the choice of both μ and σ. Both parameters affect coverage, an important consideration for the protectiveness of the UF value chosen.

The traditional method for application of UFs involves division of the POD by the UFs and division is equivalent to subtraction of logarithms. The use of UFs less than one will mean that these values will be added to the logarithm of the POD rather than subtracted and thus will increase the value of the RfV.

When values of UF-A less than unity are used, this choice indicates that humans are measurably less sensitive than the animal test species (e.g., Simon et al., 2008, 2009). USEPA (2014) indicates that the extrapolation factor may be less than one for both the toxicokinetic and toxicodynamic components of the interspecies extrapolation factor when supporting data exist. In such a case, one would base the value of such a data-derived UF on measured differences between humans and the test species. The central value, i.e., the mean or median, of these measured differences would inform the adjustment value or μ and the spread of the differences would inform σ, the measure of uncertainty. Use of an interspecies UF with a value less than unity would, of course, depend on the data in humans and the test species; here, for illustration, Table 2 shows values of μ less than one that represent a reduction of the margin of safety as an adjustment to be added directly to the natural logarithm of the POD and the value of σ would be combined using Eq. (1b) below to increase the margin of safety, thus accounting for uncertainty.

In general, an UF/EF from a given source is considered to be both a numerical adjustment with a conservative upper bound of uncertainty. The Bayes approach described here accounts for both the adjustment and the uncertainty. A point estimate may be selected from this distribution or the entire distribution may be used. We emphasize that the mean value is the central value of the adjustment and standard deviation characterizes the range of uncertainty inherent in this adjustment. Hence, risk analysts will need the flexibility to choose different means and standard deviations and different percentiles and, thus, different confidence levels for a given UF, and will also need to provide a rationale for their choices. The increased flexibility and the need for a rationale will increase the transparency and the likely utility of any RfV.

2.3. Use of UF distributions to represent a percentile of the target population included in an RfV value

Science and Decisions: *Advancing Risk Assessment*, from the NRC in 2009 distinguished between adverse effects occurring on the individual level characterized by severity or magnitude, and adverse effects of a particular magnitude occurring on a population level characterized by their incidence or frequency within the population (NRC, 2009).

This idea of protecting the population at a given risk or incidence level for a particular effect magnitude was first advanced as part of a “Straw Man” proposal to define the RfV in a more quantitative fashion (Hertzberg and Dourson, 1993; Dourson et al., 1997; Teuschler et al., 1999; Hattis et al., 2002).

This paper is a proof-of-concept and a first attempt to learn what aspects of Bayesian methods for UF application can be used either immediately or in the near term. Hence, consistent with current EPA practice, distinguishing magnitude and incidence was not implemented here.

3. Methods

The derivation of RfD and RfC values followed generally the suggestions of NRC (2014). We considered each individual UF, rather than the composite UF, to be represented by a Bayesian prior distribution characterized by its mean and variance. These individual distributions can be combined to obtain a measure of the overall adjustment and attendant uncertainty. Hence, the method shown here represents a methodological refinement of the one used in the phenol example on page 121 of Chapter 7 of NRC (2014).

Practically, we believe that individual UFs represent areas of uncertainty too dissimilar to support the use of a single distribution for the composite UF obtained as the product of individual UFs in a Bayesian framework. The use of a composite UF would ignore the knowledge of uncertainty from specific sources, consistent with the original rationale for the five UFs. For example, the extrapolation of experimental animal toxicokinetics to human toxicokinetics (i.e., the UF-A-TK) is not the same as the extrapolation of a NOAEL or BMDL from a subchronic experimental animal study to its chronic equivalent (i.e., the UF-S); specific knowledge about the uncertainty in interspecies differences in toxicokinetics and chronic vs. subchronic exposure will likely warrant choosing different adjustments and different ranges of uncertainty. Hence, we believe the best practice is to assume that each factor is independent and then to choose distributions that represent each individual UF and apply these at the appropriate stage during the development of an RfD or RfC, as in Fig. 1.

Eq. (1a) below shows the combination of a single distribution for the composite uncertainty factor with the uncertainty in the POD, as in the phenol example in NRC (2014). Eq. (1b) shows the refinement of this method recommended here for combining the distributions representing both adjustment and uncertainty for each individual UF with the uncertainty in the POD.

$$\ln(POD) - Z_{\alpha}\sqrt{\sigma^2_{POD} + \sigma^2_{UF_{comp}}} \tag{1a}$$

$$\ln(POD) - \sum \mu_{UF} - Z_{\alpha}\sqrt{\sigma^2_{POD} + \sigma^2_{UF-S} + \sigma^2_{UF-L} + \sigma^2_{UF-A} + \sigma^2_{UF-H} + \sigma^2_{UF-D}} \tag{1b}$$

Eq. (1b) implies that for each UF both parameters μ and σ are known: 1) the central value characterized by the mean and 2) the range characterized by the variance. In contrast, Eq. (1a) implies that only the range of total uncertainty from all sources aside from that in the POD is known and that the adjustment or mean is assumed to be zero. For any of the UF distributions, the value of μ may be zero, indicating no adjustment, or a positive or negative value, and in Eq. (1b), the sum of the adjustments is shown as Σμ. In contrast, in Eq. (1a), whether an adjustment is associated with this total uncertainty and just how much of this total variation is attributable to each source of uncertainty remains unspecified. Thus, the relative magnitude of any area of uncertainty will also be unspecified and any dominant uncertainty cannot be identified. As noted in the legend to Table 2 and above, the values of μ for UF-A may be less than zero; in such cases, the absolute values of the logarithms of their values will be added to and not subtracted from the POD.

The five examples below serve to suggest initial ways to use the methods described in NRC (2014) in the immediate term. Save for example 2 below, the value of μ is assumed to be zero for all UFs, consistent with NRC (2014). This assumption was maintained in the calculation of RfVs for 24 recent IRIS assessments presented below in order to facilitate comparison with the IRIS values. As discussed, this assumption about the value of μ for all UFs is clearly incorrect, but with this assumption, the results can be meaningfully compared with the method from NRC (2014) as shown in Eq. (1a) and with those in EPA’s IRIS database. These examples are by no means definitive; however they do provide a first look at replacing default single point estimates for UFs with distributions, and, thus, provide an improvement that has been recommended in multiple NRC reports (NRC, 2006, 2007b, 2009, 2014).

3.1. Example 1—combining individual UFs with the POD

In this example, the general method is illustrated. The BMDL represents the lower 95% one-sided confidence limit on the BMD, consistent with USEPA (2012a). Note that in practice, a different level of confidence may be chosen and one can specify this choice in EPA's Benchmark Dose Software. From the BMD modeling conducted by EPA for phenol, the values of the BMD and BMDL were 157 mg/kg/d and 93 mg/kg/d respectively (USEPA, 2002b). The absolute value of the Z-score associated the BMDL is 1.645, indicating the BMDL occurs at the lower 95% one-sided confidence limit of the BMD or the 5th percentile of the distribution of uncertainty in the BMD. Assuming ln (BMD) follows a normal distribution with the mean = ln (157) and the lower 95% one-sided confidence level or 5th percentile = ln (93), the standard deviation, σ, of the distribution of uncertainty in the POD is given by:

$$\frac{\ln(157)-\ln(93)}{1.645}=0.318 \tag{2}$$

where the symbol "ln(x)" indicates the natural logarithm of the value of "x".

In deriving the RfD for phenol, EPA used three individual UFs, (BMDL/(UF-A × UF-H × UF-D)); these were interspecies UF-A with a default value of 10, intraspecies UF-H with a default value of 10 and a database uncertainty factor, UF-D, with a value of 3. USEPA (2002a) provides a rationale for the choice of the default values for UF-A and UF-H and indicates the reduced value of UF-D is based on the relatively complete database; however, the value of UF-D was chosen to be greater than unity because only a single short-term study indicating the potential for hematological effects was available.

Here we assume each UF can be characterized by a log-normal distribution with mean μ representing any needed adjustment and a standard deviation σ representing the attendant uncertainty. When μ = 0, σ is calculated as ln (UF)/Z-score (for the 95th percentile, this would be ln (UF)/1.645). The overall uncertainty in the RfD also follows a log-normal distribution that combines into its overall standard deviation the variation/uncertainty in the POD with the adjustments all equal to zero (not shown explicitly in Eq. (3a)) and uncertainty at the 95th percentile of the log-normal distributions of the inter- and intraspecies UFs having each a value of 10 and the database uncertainty factor having a value of 3. Table 1 provides the values of σ for the UFs for a range of percentiles/confidence limits (Eq. (3a)). The RfD on the log-scale is taken as the 5th percentile of this distribution, i.e. 1.579.

$$\ln(157)-1.645*\sqrt{0.318^2+1.4^2+1.4^2+0.668^2}=1.579 \tag{3a}$$

$$RfD=e^{1.579}=4.85\simeq 5\ \text{mg/kg/d} \tag{3b}$$

An RfV is expressed with one or at most two decimals (Eq. (3b)) (USEPA, 1989). For comparison, the RfD value for phenol from EPA's IRIS database is 93 mg/kg/d divided by 300 or 0.3 mg/kg/d. The RfD value for phenol presented in NRC (2014) was derived using Eq. (1a) above and assuming the BMDL was the one-sided lower 97.5th percentile of the uncertainty in the BMD and that the composite UF represented the 97.5th percentile of the total uncertainty; this value was 0.5 mg/kg/d.

The reduction of total value of uncertainty that occurs by dividing the total uncertainty into components is a general phenomenon, not specific to RfV development; the standard deviation for the composite UF would be the sum of the standard deviations as follows:

$$\sigma_{composite}=\frac{\ln(UF_1\,x\,UF_2\,x\,UF_3\,x\ldots x\,UF_k)-(\mu_1+\mu_2+\mu_3+\ldots+\mu_k)}{Z_\alpha}=\sigma_1+\sigma_2+\sigma_3+\ldots+\sigma_k \tag{4}$$

The variance for the composite UF would be:

$$\sigma^2_{composite}=(\sigma_1+\sigma_2+\sigma_3+\ \ldots\ +\sigma_k)^2 \tag{5}$$

In contrast, when individual UFs are combined within the Bayesian framework described in NRC (2014), the overall variance would be sum of the individual variances as follows:

$$\sigma^2=\sigma_1^2+\sigma_2^2+\sigma_3^2+\ \ldots\ +\sigma_k^2 \tag{6}$$

Thus the variance for the composite UF will be far greater than the variance when an independent distribution is assumed for each component UF.

To demonstrate this point, we will consider the combination of three UFs, each with a default value of 10 representing 95% coverage. On a natural log scale, the distributions would have μ = 0 and σ = 1.4 (Table 1). The composite UF would be 1000, $\sigma_{composite}$ on a natural log scale would be 4.2, and the variance would be 17.6. When the UFs are combined as in Eq. (6), the overall standard deviation would be 2.4 and the variance would be 5.9. Intuitively, the value of the composite UF has a 5% chance to be beyond the overall 95th percentile whereas the chance of being above 95th percentile for all individual UFs simultaneously is $(0.05)^k$, much smaller than 0.05. Hence, this use of a composite UF is another example of compounding conservatism (Burmaster and Lehr, 1991; Burmaster and Harris, 1993; Von Stackelberg and Burmaster, 1993; Cullen, 1994; Simon, 2011; Tatum et al., 2015).

This simple example of the refined method (Eq. (1b)) demonstrates not only the combination and application of individual UFs using Bayesian methods but also the flexibility and potential of the use of informed choices for the values of the mean adjustment, the standard deviation and percentile that each individual UF represents. In addition, the Bayesian framework provides the ability to develop sequential updates to an RfD/RfC as new knowledge about these uncertainties becomes available.

3.2. Example 2: adjusting the POD and accounting for uncertainty

We wished to provide an example of an adjustment to the POD for UF-L that implies a non-zero adjustment and we chose the RfC for 1,4-dioxane. In the IRIS derivation, the composite UF was 1000, consisting of UF-A = 3, UF-H = 10, UF-L = 10 and UF-D = 3. The POD was a NOAEL based on the critical effect of sclerosis of the lamina propria, and a LOAEL for the critical effects of atrophy and respiratory metaplasia of the olfactory epithelium. The value of the POD was 32.2 mg/m^3 and was considered a LOAEL (USEPA, 2013b).

The values of μ for UF-A, UF-H and UF-D were assumed to be zero. The values for μ and σ for UF-L were obtained from Table 3 of Pieters et al. (1998). The geometric mean and geometric standard deviation of the LOAEL/NOAEL ratio values from 175 chronic studies are reported as 4.5 and 1.7 respectively. These values are equivalent to μ = 1.504 and σ = 0.531.

Hence, the value of the RfC is calculated as:

$$\ln(32.2)-1.504-1.645*\sqrt{0.668^2+1.400^2+0.531^2+0.668^2}=-0.940 \tag{7a}$$

$$RfC=e^{-0.940}=0.39\cong 0.4\ \text{mg/m}^3 \tag{7b}$$

For comparison, the RfC is calculated using the assumption of no adjustment and the value from Table 1 for UF-L = 10, i.e. $\mu = 0$ for all UFs, including UF-L.

$$\ln(32.2) - 1.645*\sqrt{0.668^2 + 1.400^2 + 1.400^2 + 0.668^2} = -0.137 \quad (7c)$$

$$RfC = e^{-0.137} = 0.87 \cong 0.9\ \mathrm{mg/m^3} \quad (7d)$$

In this instance, the adjustment with a value of ln (4.5) for the use of a LOAEL along with the measured variance in a dataset of 175 LOAEL/NOAEL ratios led to a more protective toxicity factor than using $\mu = 0$ (Table 3 and Eq. (7)).

A full exploration of the issues of adjustment and coverage is beyond the scope of this paper; nonetheless, we wanted to provide an example calculation as a gateway to future work. The point of this example is that both the choices of both the mean and standard deviation of the log-normal distribution chosen to represent a particular UF affect the amount of adjustment and the degree of coverage.

3.3. Example 3: methanol RfC—incorporation of physiologically-based pharmacokinetic (PBPK) modeling for interspecies extrapolation

In this example, we illustrate the stages at which the various UFs are applied, as shown in the scheme in Fig. 1. This example will be revisited in the DISCUSSION. EPA selected decreased brain weight in rats as the critical effect and used a PBPK model to obtain the POD as an area-under-the-curve (AUC) in rat dams in units of mg-hr/L (USEPA, 2013a). A decrement of one standard deviation from the control mean was used as the benchmark response (BMR). The Hill model was chosen and a BMD_{-1SD} of 1730 mg-hr/L and a $BMDL_{-1SD}$ of 858 mg-hr/L were obtained. Assuming the $BMDL_{-1SD}$ is the 95% lower one-sided confidence level or 5th percentile, the standard deviation of the distribution of uncertainty in the animal POD is

$$\frac{\ln(1730) - \ln(858)}{1.645} = 0.426 \quad (8)$$

The value of the interspecies factor or UF-A used by EPA was 3, reduced from 10 because a PBPK model was used to extrapolate the toxicokinetic component. This choice represents the 95% one-side confidence limit of uncertainty associated with the toxicodynamic component of interspecies extrapolation; further, this choice assumes that humans could be at most slightly more than three-fold more sensitive than the test species in terms of toxicodynamics and that the value of 3 as a high percentile represents a conservative choice. An estimate of the actual value representing interspecies toxicodynamic differences is not available; if so, the value of this estimate would be used as the value of μ of UF-A-TD instead of assuming $\mu = 0$. The human POD as the AUC derived from the animal AUC BMD_{-1SD} would be

$$\ln(1730) - 1.645*\sqrt{0.426^2 + 0.668^2} = 6.152$$
$$\exp(6.152) = 470\ \mathrm{mg-hr/L} \quad (9)$$

To obtain the human-equivalent concentration (HEC) in ppm, EPA (2013) provided a single equation that represented the human PBPK model and is shown below for completeness:

$$\mathrm{HEC}(ppm) = 0.02308*AUC + \frac{1734*AUC}{1098 + AUC} \quad (10)$$

Using Eq. (10), the HEC corresponding to 470 mg-hr/L is 531 ppm. USEPA indicates that the PBPK model captured in Eq. (10) is valid only for HEC values below 500 ppm. Nonetheless, the model was used at this stage to remain faithful to the scheme shown in Fig. 1. Additional information on the derivation of the Methanol RfC indicates that the most accurate range of Eq. (10) is between 500 and 1000 ppm; this information is provided in Supplementary Content.

EPA chose a value of 10 for the combination of the toxicodynamic and toxicokinetic components of the intraspecies UF-H accounting for lack of knowledge of true human variability and assuming that the value of 10 represents the "best guess" of the 95% one-sided confidence limit of human variability (WHO-IPCS, 2014). UF-H is best applied to the human-equivalent concentration obtained from Eq. (10) representing the result of applying human PBPK model. Because of the lack of data on reproductive, developmental and chronic effects in non-human primates, EPA chose to use a database deficiency UF-D with a value of 3. The choice of this value indicates that should a full database become available, one could be 95% certain that the value of the RfC would be reduced by no more than three-fold. The values of both UF-H and UF-D represented the 95% one-sided confidence limits (Table 1) were applied to the HEC of 531 ppm as follows:

$$\ln(531) - 1.645*\sqrt{1.4^2 + 0.668^2} = 3.721 \quad (11a)$$

$$\mathrm{RfC} = e^{3.721} = 41.3\ \mathrm{ppm} = 54\ \mathrm{mg}/\mathrm{m^3} \cong 50\ \mathrm{mg}/\mathrm{m^3} \quad (11b)$$

For comparison, this value is 2.5 fold greater than the RfC for methanol in IRIS of 20 mg/m^3.

3.4. Example 4—combining PODs from multiple studies-PCE

In this example, Bayesian updating is shown for the accommodation of different data sets. The advantage of this method is a reduction in uncertainty in the POD. For perchlorethylene, EPA considered two separate studies as sources for POD values; both studies measured mean exposure of dry cleaning workers (Cavalleri et al., 1994; Echeverria et al., 1995). These two studies observed changes in vision in dry cleaning workers and the exposure concentrations are thus considered LOAELs (USEPA, 2012b).

Cavalleri et al. (1994) provide the arithmetic mean and arithmetic standard deviation of 6.23 ± 6.66 ppm as the concentrations to which 35 dry cleaning workers were exposed as a time-weighted average, equivalent to 42 ± 45 mg/m^3. This exposure was duration-adjusted to represent constant exposure as follows:

$$(42\pm45)\,mg/m^3 \times \frac{5d}{7d} \times \frac{10\,m^3/d}{20\,m^3/d} = (15\pm16)\,mg/m^3 \quad (12)$$

The air concentrations of PCE were measured with personal passive samplers. The range of values is from 0.38 to 31.19 ppm. The original data are not available, but the coefficient of variation of these air concentrations is 6.66/6.23, over 100%, and the concentration range is about 100 fold. These wide variations are suggestive of a heavy right-tailed distribution consistent with the characteristic shape of a log-normal distribution. The geometric mean of 10.3 mg/m^3 and geometric standard deviation of 2.391 were calculated from the arithmetic mean and arithmetic standard deviation (Limpert et al., 2001).

With such a wide range of exposure, considerable uncertainty exists in the interpretation of the mean value as an effect level. Within the approximately 100-fold range of exposure, some fraction of those exposed were sufficiently sensitive to experience visual changes; however, the unknown variations in both exposure

and sensitivity both contribute to the occurrence of the critical effect, and the contributions of these two factors cannot be separated. This compounded uncertainty of exposure and sensitivity notwithstanding, the standard error of the duration-adjusted arithmetic mean concentration of 15 mg/m^3 will be used here as an estimate of the uncertainty in this LOAEL with the understanding that this estimate is likely inaccurate. The purpose here is to illustrate the Bayesian combination of PODs. The value of the SEM is 2.68 mg/m^3 and was estimated with a simulation, similar to a parametric bootstrap.[1] Hence, the uncertainty in this duration-adjusted LOAEL of 15 ± 2.68 mg/m^3 can be represented by a log-normal distribution with parameters $\mu = 2.692$ and $\sigma = 0.177$ (Limpert et al., 2001).

Echeverria et al. (1995) provide the arithmetic mean and arithmetic standard deviation equal to 23.2 ± 17.7 ppm in air as the concentration to which 18 workers were exposed in dry cleaning shops using a wet-transfer process. They collected PCE measurements in breath and air in the breathing zone of 17 dry cleaners in Detroit. These concentrations are equivalent to 156 ± 120 mg/m^3. The duration-adjusted values are 56 ± 43 mg/m^3. This group is the moderate exposure group in Table 3 of Echeverria et al. (1995) and the one chosen by EPA in the PCE Toxicological Review in Table 5-1 as representing a LOAEL (USEPA, 2012b). Using the same simulation method, the standard error of the mean (SEM) was estimated to be 10.36 mg/m^3. Hence, the mean exposure is 56 mg/m^3 and the uncertainty as the SEM is 10.36 mg/m^3. Based on these values, the uncertainty in the LOAEL from Echeverria et al. (1995) can be represented as a log-normal distribution with $\mu = 4.009$ and $\sigma = 0.183$, with the same caveats about the inherent uncertainty from unknown ranges of exposure and sensitivity.

NRC (2014) suggests variance weighting to combine PODs from different sources and gives several examples. In the following example, the distribution from Cavalleri et al. (1994) is considered the prior distribution and that from Echeverria et al. (1995) is considered as new data to update the prior. This example calculation is very similar to that on page 123 of NRC (2014). The variance-weighted POD would be calculated as:

$$\frac{\sigma_{prior}^2}{\sigma_{prior}^2+\sigma_{new}^2}\mu_{new}+\frac{\sigma_{new}^2}{\sigma_{prior}^2+\sigma_{new}^2}\mu_{prior}=$$
$$\frac{0.1773^2}{0.1773^2+0.1834^2}\cdot 4.009+\frac{0.1834^2}{0.1773^2+0.1834^2}\cdot 2.692=3.328 \quad (13a)$$

Combining the values from the two studies would give a POD of 28 mg/m^3. The standard deviation in natural log space of the log-normal distribution representing uncertainty in the combined POD would be:

$$\sqrt{\left(\frac{1}{\sigma_{prior}^2}+\frac{1}{\sigma_{new}^2}\right)^{-1}}=\sqrt{\left(\frac{1}{0.1773^2}+\frac{1}{0.1834^2}\right)^{-1}}=0.1274 \quad (13b)$$

[1] The simulation was implemented by selecting 35 values at random from a log-normal distribution with an arithmetic mean of 15 mg/m^3 and an arithmetic standard deviation of 16 mg/m^3 and calculating the arithmetic mean of these selected values. This process was repeated 10,000 times and the standard deviation of the collection of 10,000 simulated means was assumed to represent the standard error of the mean.

3.5. Example 5—determining the percentile of the distribution of overall uncertainty for RfD development

In the four examples above, we have chosen the lower 95th percentile (i.e. the 5th percentile) of the log-normal distribution representing the individual UFs or the POD. Here, examples of different choices are explored for both the value of the UF and the percentile this value represents reflect the state of knowledge or confidence regarding the various sources of uncertainty. The choices of UF value and the percentile dictate the value of the standard deviation (Table 1). We expect guidance to emerge regarding these choices as more experience with Bayesian application of UFs is gained. Here, we provide three illustrations of altering the value of UFs using the example of phenol.

The first method involves choosing a different value for an uncertainty factor. The 95th percentile would be maintained and these two choices would dictate the value of σ. Hence, for the phenol example, the value of 2 rather than 3 would be chosen as UF-D indicating a lower level of concern for the completeness of the toxicity database; EPA indicated that this database UF was used because immunologic and hematologic endpoints were not examined in the rodent developmental studies, and thus, any values derived from consideration of developmental and reproductive studies may not be applicable because of this missing endpoint. From Table 1, the value of the standard deviation is 0.421 corresponding to a UF value of 2 at the 95th percentile is used, and the RfD for phenol would be calculated as:

$$\ln(157)-1.645*\sqrt{0.318^2+1.4^2+1.4^2+0.421^2}=1.686 \quad (14a)$$

$$\mathrm{RfD}=e^{1.686}=5.4\cong 5\ \mathrm{mg/kg/d} \quad (14b)$$

Because of the need to round to one significant figure (USEPA, 1989), the choice of a UF of 2 versus 3 makes no difference in the final value of the RfD.

The second method would be the selection of a different percentile of UF-D to be represented by the numerical value of 3. This choice also dictates the value of σ. One can assume this study deficiency was of less concern than the complete lack of any developmental study and thus warranted the assumption that the UF value of 3 actually occurred at the 99.5% one-sided confidence limit and would dictate σ equal to 0.427. Hence, the RfD for phenol would be calculated as:

$$\ln(157)-1.645*\sqrt{0.318^2+1.4^2+1.4^2+0.427^2}=1.667 \quad (15a)$$

$$\mathrm{RfD}=e^{1.667}=5.3\cong 5\ \mathrm{mg/kg/d} \quad (15b)$$

Although numerically slightly higher than the unrounded value from example 1, rounding to one significant figure would give the same final value of the RfD. If UF-D were of greater concern, one could also set this value at a lower not a higher percentile. Hence, if one chose UF-D = 3 to occur at the 90th percentile, this choice would dictate a value of σ equal to 0.857 and the resulting unrounded RfD would be 4.3 mg/kg/d, which would round to 4 mg/kg/d.

The third method uses Monte Carlo simulation and generally requires specialized software. Here, we used 100,000 Monte Carlo iterations to obtain percentile values of the overall standard deviation corresponding to the overall uncertainty in the phenol RfD. Here, we use the notation Φ as representing the underlying normal distribution of the logarithmic values representing the various

uncertainties and use these to calculate a distribution for the natural log of the overall standard deviation.

$$\ln(UF_{overall}) = \Phi_{POD} + \Phi_{UF-A} + \Phi_{UF-H} + \Phi_{UF-D}$$
$$\text{where} \quad \begin{aligned} \Phi_{POD} &= N(0.0,\ 0.318); \\ \Phi_{UFA} &= \Phi_{UFH} = N(0.0,\ 1.4); \\ \Phi_{UFD} &= N(0.0,\ 0.668); \end{aligned} \tag{16a}$$

$$RfD = \exp(\ln(157) + p05(\ln(UF_{overall}))) \tag{16b}$$

The value of the 5th percentile of the ln ($UF_{overall}$) used in Eq. (16b) would correspond to the 95th percentile of uncertainty. Values below the 50th percentile of this combined distribution would be negative because no adjustments were made (i.e. $\mu = 0$ for all UFs) and the value of this chosen percentile would be added to the POD (or the absolute value subtracted from the POD). From the Monte Carlo results, the values of the RfD corresponding to the 1st, 2.5th, 5th and 10th percentiles were 1.1, 2.5, 4.9 and 10.4 mg/kg/d respectively. At the 5th percentile, the resulting rounded RfD of 5 mg/kg/d is the same as that obtained with Eq. (3) and would not be expected to be different from the original phenol estimate except by sampling error. Hence, considering the traditional sources of uncertainty in the RfD, one could be 95% confident that an RfD value below 5 mg/kg/d would be protective of all humans, including potentially sensitive subpopulations; also, one could be 90% confident that an RfD value below 10 mg/kg/d would also be protective.

In the first example in this section, the choice of different UF value, which prescribed the value of σ at a given percentile, was illustrated. In the second example, the chosen UF value was placed at a higher percentile, reflecting a lower contribution of UF-D to the overall uncertainty. This choice also prescribed the value of σ for the distribution of UF-D. The third example used a relatively simple Monte Carlo simulation to obtain a range of UF values and corresponding RfDs. Several NRC reviews and EPA guidance documents point out that providing policy makers with risk estimates spanning a range of overall uncertainty may serve as a more useful decision tool than a single point estimate of risk (USEPA, 1992; NRC, 1994; USEPA, 2001; NRC, 2009). As experience is gained with the Bayesian methods for UFs, some wisdom and resulting guidelines for the choices of percentile values or ranges of uncertainty will likely emerge.

4. Results

Twenty-four chemicals were examined in this exercise. The results and detailed calculations are provided in the Supplementary Content. Table 3 provides an overview of the results; in this table, three values of each RfD or RfC from each of three approaches are shown—the IRIS value, the value obtained by using a single log-normal distribution for the composite UF (Eq. (1a)), and the value obtained when one treats the individual UFs as separate log-normal distributions (Eq. (1b)). Because each assessment has its own characteristics, there are differences in the application of the three approaches. For example, in the cases of perchlorethylene and chloroprene, PODs were combined by variance weighting as illustrated earlier and described on page 123 of NRC (2014).

All RfD/RfC values listed under the heading "Individual. UFs" in Table 3, save that for ammonia, increased when compared with the value found on IRIS. The reason is that the refined Bayesian method (Eq. (1b)) assumes the uncertainty is along a distribution, which is a convolution of individual distributions representing the uncertainty of the POD and each individual UF, respectively. In comparison, the Composite UF approach (Eq. (1a) and with the heading "Composite UF" in Table 3) assumes the total uncertainty is characterized by a single distribution of which the variance is determined by the composite UF, with generally a default-based value of up to 3000, in combination with the uncertainty in the POD; hence, in the Composite UF approach, only two sources of uncertainty are considered within the Bayesian framework—that in the POD and that in the composite UF. Considering the value of the combined variance as a measure of uncertainty, the range of uncertainty under the Composite UF approach is greater than that under Individual UF approach as discussed earlier.

This point about multiple sources of uncertainty is well illustrated by the RfC for ammonia—this value was based on a NOAEL to which a single UF accounting for human variability (UF-H) was applied. Because the POD was a NOAEL and not a BMD, the uncertainty in the POD could not be determined and only a single component of variance was present. In all other assessments considered here, two or more variance components were present (Table 3; Fig. 2A).

In comparison to the IRIS derivations, the RfV values developed with the Individual UF method increased with both the value of the composite UF and the number of UFs used. When 2, 3 and 4 UFs were used, the geometric means of the fold change increase over the IRIS value were approximately 2.6, 8.6 and 26 respectively among the chemicals considered in this exercise. Fig. 2A shows graphically how the fold change values increase with the value of composite UF. For composite UF values of 30, 100, 300, 1000 and 3000, the geometric mean fold increases were 3.2, 3.7, 13, 20 and 34 respectively. In contrast, the geometric mean fold change when using the composite UF did not show an increase, or any pattern, with the value of the composite UF.

Fig. 2B shows that when individual UFs are used in a Bayesian framework, the fold change increase of this refinement of the NRC methodology over the IRIS value gets smaller with increased sophistication of the assessment; assessments that used PBPK models had generally the smallest fold change values, and assessments that relied solely on UFs had the greatest fold change values. With this method refinement, when UFs only are used, the geometric mean fold change is 22; when dosimetric adjustment factors are used, the geometric mean fold change is 8.6; and, when PBPK models are used, the geometric mean fold change is 3.2. When the composite UF was used in lieu of individual UFs, the fold change over the IRIS did not show a decrease with the level of sophistication in the assessment.

5. Discussion

Science and Decisions: *Advancing Risk Assessment* suggests that the choice of percentiles of uncertainty/adjustment/extrapolation factors is the purview of the risk practitioner (NRC, 2009). For human variability, this choice represents the population level of protection from the adverse effect of a particular magnitude. However, with default UFs applied by dividing them into the POD, the percentile of protection is completely obscured. NRC (2009) devoted an entire chapter to default values and recommended that EPA enhance the science basis of default values. Such enhancement is seen in the Bayesian approach suggested in NRC (2014). Although the approach is based on default values for UFs, additional flexibility and transparency is inherent in the choice of the percentile represented by the default value. In time, default distributions for uncertainty/adjustment/extrapolation factors based on data can be developed and used (e.g., WHO-IPCS, 2014).

5.1. POD and UFs as log-normal distributions

Toxicity results from the interaction of a xenobiotic chemical

Table 3
Assessment Details and Results of the Use of Bayesian Methods Compared to the IRIS Approaches (page 1 of 2). The column "Composite UFs" shows the result of using Eq. (1a) for the overall σ and the column "Individual UFs" shows the result of using Eq. (1b) for the overall σ. For all UFs, the value of μ was assumed to be zero. The RfD or RfC is in the same units as the POD.

Substance and date assessed	Type of toxicity value	POD	Status	Uncertainty factors	Dosimetric adjustment called "method of analysis" in IRIS	Resulting RfD/RfC (mg/kg/d or mg/m^3)			Fold change individual UFs vs. IRIS method
						IRIS method	Composite UFs (Eq. (1a))	Individual UFs (Eq. (1b))	
Ammonia[a] (Aug. 2013)	RfC	NOAEL = 3.1 mg/m^3	Draft	UFH = 10	None	0.3	NA	NA	1
N-butanol (Sept. 2011)	RfD	BMD_{10} = 56.5 mg/kg/d $BMDL_{10}$ = 26.1 mg/kg/d	Draft	UFA = 10; UFH = 10; UFD = 3;	None	0.09	0.2	2	22
	RfC	NOAEL-HEC = 59 mg/m^3	Draft	UFA = 3; UFH = 10; UFD = 3; UFS = 10;	PBPK Model	0.06	0.06	0.2	3.3
Vanadium Pentoxide (Sept. 2011)	RfD	NOAEL = 10.5 mg/kg/d	Draft	UFA = 3; UFH = 10; UFS = 10; UFD = 10;	DAF = 0.24	9.E-04	9E-04	8.E-03	8.9
	RfC	BMC_{10} = 0.0045 mg/m^3 $BMCL_{10}$ = 0.0031 mg/m^3	Draft	UFA = 3; UFH = 10; UFD = 10;	DAF = 0.26	1.E-05	1E-05	1.E-04	10
Methanol (Sept. 2013)	RfC	BMD_{1SD} = 1730 mg-hr/L $BMDL_{1SD}$ = 858 mg-hr/L	Final	UFA = 3; UFH = 10; UFD = 3;	PBPK	20	30	50	2.5
1,4-Dioxane (Sept. 2013)	RfD	NOAEL - 9.6 mg/kg-d	Final	UFA = 10; UFH = 10; UFD = 3	None	0.03	0.03	0.3	10
	RfC	NOAEL = 32.2 mg/m^3	Final	UFA = 3; UFH = 10; UFL = 10; UFD = 3	Allometric DAF = 1	0.03	0.03	1	33
Biphenyl (Aug. 2013)	RfD	BMD_{10} = 92 mg/kg/d $BMDL_{10}$ = 58 mg/kg/d	Final	UFA = 10; UFH = 10	Allometric DAF = 0.24	0.5	0.7	2	4
TCDD (Feb. 2012)	RfD	LOAEL = 0.02 ng/kg/d	Final	UFH = 3; UFL = 10	PBPK model	7.E-10	7E-10	2.E-09	2.9
Perchlorethylene[b] (Feb. 2012)	RfC	LOAELs = 56 & 15 mg/m^3	Final	UFH = 10; UFL = 10; UFD = 10	None	0.04	0.03	0.5	12.5
Phenol (Sept. 2002)	RfD	BMD_{1SD} = 157 mg/kg/d $BMDL_{1SD}$ = 93 mg/kg/d	Final	UFA = 10; UFH = 10; UFD = 3	None	0.3	0.5	5	17
Dichloromethane (Nov. 2011)	RfD	1st %ile internal HED BMD_{10} = 18.4 mg/L liver/d $BMDL_{10}$ = 13.0 mg/L liver/d	Final	UFA = 3; UFH = 3; UFD = 3	PBPK + allometric scaling	0.006	0.02	0.05	8.3
	RfC	1st %ile internal HED BMD_{10} = 532 mg/L liver/d $BMDL_{10}$ = 130 mg/L liver/d	Final	UFA = 3; UFH = 3; UFD = 3	PBPK; allometric scaling	0.6	0.7	1.5	2.5

Substance and date assessed	Type of toxicity factor	POD	Status	Uncertainty factors	Dosimetric adjustment called "Method of Analysis" in IRIS	Resulting RfD/RfC (mg/kg/d or mg/m^3)			Fold change individual UFs vs. IRIS method
						IRIS method	Composite UFs (Eq. (1a))	Individual UFs (Eq. (1b))	
Trichloroacetic Acid (Sept. 2011)	RfD	BMD_{10} = 40.7 mg/kg/d $BMDL_{10}$ = 17.9 mg/kg/d	Final	UFA = 10; UFH = 10; UFD = 10	None	0.02	0.04	0.7	35
Hexachloroethane (Sept. 2011)	RfD	BMD_{10} = 1.34 mg/kg/d $BMDL_{10}$ = 0.728 mg/kg/d	Final	UFA = 10; UFH = 10; UFS = 3; UFD = 3	None	7.E-04	0.001	0.03	43
	RfC	NOAEL = 83 mg/m^3	Final	UFA = 3; UFH = 10; UFS = 10; UFD = 10	None	0.03	0.03	0.3	10
Acrylamide (March 2010)	RfD	BMD_{10} = 0.58 mg/kg/d $BMDL_{05}$ = 0.27 mg/kg/d	Final	UFA-TD = 3; UFH = 10	Hb adduct equivalence and PK conversion	0.002	0.004	0.003	1.5
Chloroprene[b] (Sept. 2010)	RfC	BMDLs = 2.3, 2.1, 2.1 mg/m^3	Final	UFA = 3; UFH = 10; UFD = 3	DAF = RDDR = 1	0.02	0.03	0.08	4
cis-1,2-Dichloroethylene (Sept. 2010)	RfD	BMD_{10} = 19.8 mg/kg/d $BMDL_{10}$ = 5.1 mg/kg/d	Final	UFA = 10; UFH = 10; UFS = 10; UFD = 3	None	0.002	0.006	0.3	150
trans-1,2-Dichloroethylene (Sept. 2010)	RfD	BMD_{1SD} = 126 mg/kg/d $BMDL_{1SD}$ = 65 mg/kg/d	Final	UFA = 10; UFH = 10; UFS = 10; UFD = 3	None	0.02	0.04	2	100
Pentachlorophenol (Sept. 2010)	RfD	LOAEL = 1.5 mg/kg/d	Final	UFA = 10; UFD = 10; UFL = 3	None	0.005	0.005	0.05	10
1,1,2,2-Tetrachloroethane (Sept. 2010)	RfD	BMD_{1SD} = 22 mg/kg/d $BMDL_{1SD}$ = 15 mg/kg/d	Final	UFA = 10; UFH = 10; UFS = 3; UFD = 3	None	0.02	0.02	0.6	30
Carbon tetrachloride (March 2010)	RfC	BMD_{10} HEC = 18.1 mg/m^3 $BMDL_{10}$ HEC = 14.3 mg/m^3	Final	UFA = 3; UFH = 10; UFD = 3	PBPK model	0.1	0.2	0.5	5

[a] Bayesian methods were not used in the case of the ammonia RfC because a single source of uncertainty—human variability—was considered in the development of the ammonia RfC. The value using either Bayesian method is the same as that in IRIS.
[b] Variance-weighted POD was used for NRC and refined methods.

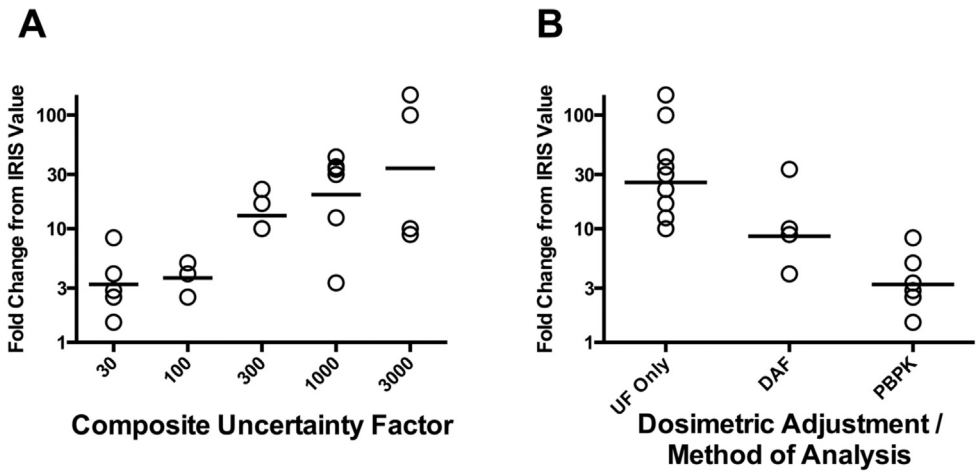

Fig. 2. Separation by Method of Analysis and Value of the Composite UF. A. Dot plot of fold change increases of the refined method over the IRIS value of RfD/RfCs by the value of the composite UF. B. Dot plot of fold change increases over the IRiS value by dosimetric adjustment/method of analysis. The horizontal lines show the geometric mean.

with one or more biological molecules. As earlier noted, based on the principles of thermodynamics, dose is most appropriately expressed as a logarithm (Waddell, 2008). The history of toxicology is often divided into three eras: Pre-classical from the time of Paracelsus to the start of the 20th century, Classical from 1900 until 1965, and Current, from 1965 until the present (Waddell, 2010). In *The Mode of Action of Drugs on Cells*, Clark (1933) points out that the "intensity of maximum action and the duration will vary as the logarithm of dosage, whilst the amount of action will vary as some multiple of the logarithm of the dosage." In the Current era, the use of the logarithm of dosage has been recommended to achieve homogeneity of variance (Igarashi and Sekido, 1996; Igarashi et al., 1996), for calculation of the plasma concentration of propofol when used as a human anesthetic (Bruhn et al., 2003) and for examination of the effects of antinociceptive drugs in mice (Luszczki and Czuczwar, 2008; Luszczki, 2010). Rozman et al. (1996) have even suggested a dose scale using the number of molecules of an agent, which necessitates the use of a logarithmic scale because of the very large numbers, around 10^{19}.

In the Current Era of toxicology, the linearized multistage model emerged as a purely statistical approach to dose-response and risk assessment (Waddell, 2010). Regulatory agencies have accepted the linear no-threshold hypothesis as the basis of the biological effects of ionizing radiation and chemical exposure (Calabrese, 2009, 2015); this acceptance has spurred considerable and continuing diversity of opinion among toxicologists regarding the best quantitative representation of dose (Crump and Clewell, 2003; Waddell, 2003; Bogen, 2014; Simon, 2014).

NRC (2014) takes a welcome long view regarding the current state of risk assessment and acknowledges the value of different types of data, such as animal testing, human epidemiologic data, and high throughput *in vitro* test results in developing toxicity values as well as suggesting that Bayesian methods provide a more systematic and flexible approach for quantitative consideration of uncertainties. The consideration of these various data types will necessitate the use of biological knowledge and mode of action in risk assessment. The increased use of adverse outcome pathways (AOPs) to understand and contextualize data from various sources reflects this growing necessity (Villeneuve et al., 2014a,b; Becker et al., 2015; Patlewicz et al., 2015). Perhaps even more importantly, Bayesian methods place value on the knowledge and beliefs of the risk assessor.

However, despite the historical use of the log-normal distribution, what remains unknown is whether uncertainty or variation in the estimator of a POD at a designated effect level is also log-normally distributed. Log-normal distributions have the advantage of reproducibility, i.e. the ratio involving multiple factors remains a log-normal distribution, enabling simple computation as illustrated in this paper. The extent to which a lognormal distribution is satisfactory is case-specific. Our choice of using log-normal distribution is to further illustrate the approach of the NRC (2014). The true form of the distribution of the BMD remains unknown, and further research in this area is needed. Hence, similar to the NRC and others in the history of RfV development, we acknowledge this imprecision and also chose to represent both the POD and the UFs with log-normal distributions.

In the case when the distribution of the uncertainties or variation of POD deviates from a log-normal distribution, modern statistical tools such as the Markov chain Monte Carlo methods support the integrative process of combining uncertainties across multiple areas (Jones et al., 2009; Hennessey et al., 2010; Shao and Gift, 2013; Wilson et al., 2014). To this end, more experience needs to be gained and access to appropriate software may be also be needed. However, using distributions to representing uncertainty will provide policy makers additional information upon which to base their decisions and also provide greater transparency regarding the various sources of uncertainty (NRC, 2007a).

The long history and considerable experimental and conceptual support for the use of logarithms as the most appropriate expression of quantitative dose also suggests that the log-normal distribution is an appropriate way to express uncertainty in dose. Recently, the relationship between the logarithms of both dose and response is useful in understanding the shape and steepness of the dose-response curve (Slob and Setzer, 2014). The mathematical tractability of this distribution is a highly useful consequence. Other distributions may be used to express the various types of uncertainty accounted for in RfD/RfC derivation and Bayesian Monte Carlo simulation is one means of using other distributions. The use of Bayesian Markov Chain Monte Carlo methods for PBPK modeling in risk assessment is a well-studied and mature technique (Bois et al., 2010). In addition, such a quantitative framework may help integrate data for use by risk managers (Hill, 1996; Linkov et al., 2009).

5.2. The use of distributions to represent UFs

In the early days of using safety factors, a number of individual components of uncertainty were combined into an overall factor without quantifying these individual components. When

individual UFs representing the five areas of uncertainty were developed, their quantification was rudimentary—based only on the number of human digits, the same as our base ten numbering system. Considerable effort has been put forth to provide a basis for quantification of UFs (Pieters et al., 1998; Renwick and Lazarus, 1998; Burin and Saunders, 1999; Bruckner, 2000; Haber et al., 2002; Meek et al., 2002; Pelekis et al., 2003; Pelekis and Krishnan, 2004; Bokkers and Slob, 2005, 2007; Dorne, 2005; Price et al., 2008; Hasegawa et al., 2010; Blackburn et al., 2015). Despite this effort, almost all IRIS assessments use a value of 10 or 3 ($\sqrt{10}$) for all UFs (Stedeford et al., 2007).

The understanding of traditional UFs is grounded in the individual factors and not their combination. Assigning distributions to the UFs is fraught with difficulties in addition to the choice of distributional parameters. How much should UFs be subdivided? Both WHO-IPCS (2005) and EPA (2014) split UF-A and UF-H into toxicokinetic and toxicodynamic components. Thus far, the value of these individual components is almost always three (3). We would thus interpret this choice as a log-normal distribution with $\mu = 0$ and $\sigma = 0.668$ (Table 1). But this question remains unanswered: does this the value of three as the 95th percentile or the use of this log-normal distribution provide sufficient coverage?

In keeping with the concept of UF distributions that provide both an adjustment and a degree of uncertainty, an early paper by Baird et al. (1996) was prescient regarding the current WHO-IPCS uncertainty guidance. These authors presented distributions representing UF-A, UF-H, UF-S and UF-L based on the available data at that time, assumptions about species heterogeneity in humans and test species, and general statistical considerations.

Chiu and Slob (2015) also separate adjustment from uncertainty and also add variability as a category of uncertainty for UF-H. The distinction here is between aleatory uncertainty or variability referring to known population heterogeneity combined with uncertainty in selection an individual from this population and epistemic uncertainty referring to lack of knowledge about phenomena (Helton and Burmaster, 1996). Both Chiu and Slob (2015) and WHO-IPCS (2014) add a dosimetric adjustment based body size expressed as the ratio of a fractional powers of the body weights of humans and the test species in addition to UF-A based on toxicokinetic or toxicodynamic differences. Such dosimetric adjustments have been used in RfC development for many years (USEPA, 1994).

WHO-IPCS (2014) also chose to represent UF distributions as log-normal. This document also presents values for the parameters of missing distributions based on previously compiled data. Supplemental Table 1 provides values for the means and standard deviations of these distributions from WHO-IPCS (2014) and other sources. Included are distributions for uncertainty in NOAEL values, UF-S and subacute-to-chronic extrapolation, the portion of UF-A remaining after body size adjustment, UF-H-TK, and UF-H-TD. The WHO-IPCS guidance also provides a discussion of the inappropriateness of using UF-L to estimate a NOAEL and that the distribution of the LOAEL-to-NOAEL ratio reflects predominantly the dose spacing used in the toxicological studies. WHO-IPCS (2014) also does not mention the use of UF-D for database deficiencies.

5.3. Are the UF distributions independent?

The other issue is that of dependencies between the distributions representing the various UFs. As discussed, generic distributions have been developed to represent the UFs (Price et al., 1997; Swartout et al., 1998). Others have used data on a relatively large number of chemicals to determine distributions for UF-S, UF-L, UF-H (Renwick and Lazarus, 1998; Burin and Saunders, 1999; Bokkers and Slob, 2005, 2007; Dorne, 2005; Dourson et al., 2013; WHO-IPCS, 2014). WHO-IPCS (2014) suggests that the distributions used for UF-S, UF-A and UF-H are considered independent; while independence of these distributions seems likely, this statement is without empirical support at present.

In the future, the combination of extant *in vivo* data along with *in vitro* data from human cell lines considered in a systems biology perspective may be the most fruitful approach for understanding potential dependencies between intra- and interspecies toxicokinetic and toxicodynamic variability—although this remains to be seen (Zeise et al., 2013; Abdo et al., 2015a, 2015b). The number of dependencies in this area would be limited by the complexity of the biology and understanding of mode of action.

Considering UF-S and UF-L, one can envision potential dependencies that may arise from budget constraints on experimental design—too few doses to find a NOAEL and a LOAEL or insufficient resources to perform a chronic study—or other non-scientific factors.

Until such time that an array of chemical dose-response data of sufficient quantity is assembled to be able to assess any dependencies between the various UFs, the assumption of independence made in WHO-IPCS (2014) seems appropriate.

5.4. Additional considerations for Bayesian methods of UF application

5.4.1. Magnitude of effect and population incidence

An exposure to an individual will result in a toxicological effect of a particular magnitude; over a population, the same exposure will result in variable effect magnitudes across individuals. Hence, if one specifies a particular magnitude of effect, a fixed exposure to a population will result in an incidence of the toxicological effect of a given magnitude because individuals differ in their susceptibility (Chiu and Slob, 2015; WHO-IPCS, 2014). If one considers both effect magnitude and population incidence, the application of an extrapolation/adjustment/uncertainty factor becomes more difficult—both the effect size and the population coverage of the factor need to be defined.

In current EPA practice, the selection of the effect magnitude is tacit, implicit in the choice of critical effect. Hence, the RfV, as defined by EPA and considered here, represents a dose of a chemical associated with an effect of an unspecified magnitude for which the incidence in the human population is believed to be zero, i.e. 100% coverage (WHO-IPCS, 2014).

5.4.2. At what stage of the assessment should PBPK models be applied?

If an assessment involves a PBPK model for toxicokinetic species extrapolation, the stage at which the model is applied becomes another choice. Fig. 1 indicates that UF-A-TD is applied to the animal dose metric before using the human PBPK model to convert this dose metric into an external dose or human-equivalent concentration. The choice of the point in the RfV process to use the PBPK model makes a difference, and below, we use the methanol RfC discussed earlier as an example.

Using the human PBPK model for methanol (Eq. (10)) as the first step to obtain the BMD and BMDL as HEC values provides 1101 ppm and 780 ppm. The uncertainty in this POD and its application along with all the UFs would be:

$$\ln(1101) - 1.645*\sqrt{0.209^2 + 0.668^2 + 1.4^2 + 0.668} = 4.20$$
$$RfC = e^{4.20} = 67 \text{ ppm} = 101 \text{ mg/m}^3 \quad (17)$$

In contrast, using Eq. (6) after applying the uncertainty in the POD and all the UFs to the animal dose metric would give:

$$\ln(1730) - 1.645*\sqrt{0.426^2 + 0.668^2 + 1.4^2 + 0.668^2} = 4.59$$
$$\text{Adjusted Human Dose Metric} = e^{4.59} = 98.5 \text{ mg} - \text{hr/L} \quad (18)$$

The human PBPK model equation (Eq. (10)) would then be used to convert this result to a HEC of 145 ppm or 190.7 mg/m^3 or rounded to 200 mg/m^3 that would become the RfC.

Application of the human PBPK model as indicated in Fig. 1 produced a RfC value 2.5-fold higher than the IRIS value. The two values immediately above are five-fold and ten-fold respectively higher than EPA's derivation. For comparison the OSHA PEL for methanol has a value of 260 mg/m^3 and is based on a non-quantitative consideration of acute effects in workers (NIOSH, 1976).

Hence, guidance on the best practices for the use of PBPK models together with Bayes methods for applying uncertainty factors is needed. Each of the five UFs has a different specificity for a stage or aspect of the assessment. The specificity of UF application provides the conceptual basis of Fig. 1. However, this is an area that clearly needs more examples and additional thought.

5.5. Assessment decisions and best practices

At this point in time, we can offer two recommendations to risk assessors.

- Use Bayesian Methods for UF/CSAF/DDEF application as illustrated in this paper consistent with the suggestions of NRC (2014). One major advantage is the ability to update RfD/RfC values rapidly as new information becomes available; and
- Use care in determining the default values or percentiles for the various UFs. As noted, our expectation is that data-derived distributions characterized by both means and standard deviations will emerge to replace the default distributions in Table 1 and in NRC (2014).
- When using Bayesian methods in a RfV derivation that includes a PBPK model, the stage at which the UFs are applied will likely make a potentially large difference in the resulting value of the RfV and more experience in this area is needed.

As time goes by and greater familiarity with these Bayesian methods grows, additional ideas for best practices will very likely emerge. Nonetheless, the methods described here can be implemented immediately in the IRIS program and used for the development of RfD/RfC values henceforth. The imprecision in the extrapolations used to develop RfVs indicate that expressing an RfV as a distribution or as a selected percentile of distribution will improve the clarity and transparency of these necessary and important regulatory criteria.

Funding and acknowledgements

The American Chemistry Council (ACC) provided support to TWS, MLD, and YZ. NBB is an employee of ACC. The authors retain sole responsibility for the content of this paper. The American Chemistry Council had an opportunity to review this text as part of their internal clearance; however, this paper is exclusively the work product of the authors and does not necessarily represent views or policies of the ACC or the authors' employers. We also wish to acknowledge the contribution of one of the peer reviewers whose thoughtful suggestions improved this paper considerably.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2016.05.018.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.yrtph.2016.05.018.

References

Abdo, N., Wetmore, B.A., Chappell, G.A., Shea, D., Wright, F.A., Rusyn, I., 2015a. In vitro screening for population variability in toxicity of pesticide-containing mixtures. Environ. Int. 85, 147–155. http://dx.doi.org/10.1016/j.envint.2015.09.012.

Abdo, N., Xia, M., Brown, C.C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y.H., Jack, J.R., Gallins, P., Xia, K., Li, Y., Chiu, W.A., Motsinger-Reif, A.A., Austin, C.P., Tice, R.R., Rusyn, I., Wright, F.A., 2015b. Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 genomes high-throughput screening study. Environ. Health Perspect. 123, 458–466. http://dx.doi.org/10.1289/ehp.1408775.

Baird, S.J.S., Cohen, J.T., Graham, J.D., Shlyakhter, A.I., Evans, J.S., 1996. Noncancer risk assessment: a probabilistic alternative to current practice. Hum. Ecol. Risk Assess. Int. J. 2, 79–102.

Barnes, D.G., Dourson, M., 1988. Reference dose (RfD): description and use in health risk assessments. Regul. Toxicol. Pharmacol. 8, 471–486.

Becker, R.A., Ankley, G.T., Edwards, S.W., Kennedy, S.W., Linkov, I., Meek, B., Sachana, M., Segner, H., Van Der Burg, B., Villeneuve, D.L., Watanabe, H., Barton-Maclaren, T.S., 2015. Increasing scientific confidence in adverse outcome pathways: application of tailored Bradford-Hill considerations for evaluating weight of evidence. Regul. Toxicol. Pharmacol. 72, 514–537. http://dx.doi.org/10.1016/j.yrtph.2015.04.004.

Blackburn, K., Daston, G., Fisher, J., Lester, C., Naciff, J.M., Rufer, E.S., Stuard, S.B., Woeller, K., 2015. A strategy for safety assessment of chemicals with data gaps for developmental and/or reproductive toxicity. Regul. Toxicol. Pharmacol. 72 (2), 202–215. http://dx.doi.org/10.1016/j.yrtph.2015.04.006.

Bogen, K.T., 2014. Mechanistic models fit to ED001 data on >40,000 trout exposed to dibenzo[A,L]pyrene indicate mutations do not drive increased tumor risk. Dose Response 12, 386–403. http://dx.doi.org/10.2203/dose-response.13-019.Bogen.

Bois, F.Y., Jamei, M., Clewell, H.J., 2010. PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278, 256–267. http://dx.doi.org/10.1016/j.tox.2010.06.007.

Bokkers, B.G., Slob, W., 2005. A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicol. Sci. 85, 1033–1040. http://dx.doi.org/10.1093/toxsci/kfi144.

Bokkers, B.G., Slob, W., 2007. Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Crit. Rev. Toxicol. 37, 355–373. http://dx.doi.org/10.1080/10408440701249224.

Bruckner, J.V., 2000. Differences in sensitivity of children and adults to chemical toxicity: the NAS panel report. Regul. Toxicol. Pharmacol. 31, 280–285. http://dx.doi.org/10.1006/rtph.2000.1393.

Bruhn, J., Bouillon, T.W., Ropcke, H., Hoeft, A., 2003. A manual slide rule for target-controlled infusion of propofol: development and evaluation. Anesth. Analg. 96, 142–147 table of contents.

Burin, G.J., Saunders, D.R., 1999. Addressing human variability in risk assessment–the robustness of the intraspecies uncertainty factor. Regul. Toxicol. Pharmacol. 30, 209–216. http://dx.doi.org/10.1006/rtph.1999.1351.

Burmaster, D.E., Anderson, P.D., 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. Risk Anal. 24, 477–481.

Burmaster, D.E., Harris, R.H., 1993. The magnitude of compounding conservatisms in superfund risk assessments. Risk Anal. 13, 131–134.

Burmaster, D.E., Lehr, J.H., 1991. It's time to make risk assessment a science. Ground water monitoring review. Summer.

Calabrese, E.J., 2009. The road to linearity: why linearity at low doses became the basis for carcinogen risk assessment. Arch. Toxicol. 83, 203–225. http://dx.doi.org/10.1007/s00204-009-0412-4.

Calabrese, E.J., 2015. An abuse of risk assessment: how regulatory agencies improperly adopted LNT for cancer risk assessment. Arch. Toxicol. 89, 647–648. http://dx.doi.org/10.1007/s00204-015-1454-4.

Cavalleri, A., Gobba, F., Paltrinieri, M., Fantuzzi, G., Righi, E., Aggazzotti, G., 1994. Perchloroethylene exposure can induce colour vision loss. Neurosci. Lett. 179, 162–166.

Chiu, W.A., Slob, W., 2015. A unified probabilistic framework for dose-response assessment of human health effects. Environ. Health Perspect. http://dx.doi.org/10.1289/ehp.1409385.

Clark, A.J., 1933. The Mode of Action of Drugs on Cells. The Williams and Wilkins Company, Baltimore, MD.

Crump, K.S., Clewell, H.J., 2003. Evidence of a "clear and consistent threshold" for

bladder and liver cancer in the large ED01 carcinogenicity study. Toxicol. Sci. 74 (485) http://dx.doi.org/10.1093/toxsci/kfg117 author reply 485-6.
Cullen, A.C., 1994. Measures of compounding conservatism in probabilistic risk assessment. Risk Anal. 14, 389–393.
Dorne, J.L.C.M., 2005. The refinement of uncertainty/safety factors in risk assessment by the incorporation of data on toxicokinetic variability in humans. Toxicol. Sci. 86, 20–26. http://dx.doi.org/10.1093/toxsci/kfi160.
Dourson, M., 1996. Uncertainty factors in noncancer risk assessment. Regul. Toxicol. Pharmacol. 24 (107) http://dx.doi.org/10.1006/rtph.1996.0115.
Dourson, M., Becker, R.A., Haber, L.T., Pottenger, L.H., Bredfeldt, T., Fenner-Crisp, P.A., 2013. Advancing human health risk assessment: integrating recent advisory committee recommendations. Crit. Rev. Toxicol. 43, 467–492. http://dx.doi.org/10.3109/10408444.2013.807223.
Dourson, M.L., 1993. Reference dose of the United States environmental protection agency. Scand. J. Work Environ. Health 1 (19 Suppl. I), 115–118.
Dourson, M.L., DeRosa, C.T., 1991. The use of uncertainty factors in establishing safe levels of exposure. In: Krewski, D., Franklin, C. (Eds.), Statistics in Toxicology. Gordon and Breach Science Publishers, New York, NY.
Dourson, M.L., Felter, S.P., Robinson, D., 1996. Evolution of science-based uncertainty factors in noncancer risk assessment. Regul. Toxicol. Pharmacol. 24, 108–120. http://dx.doi.org/10.1006/rtph.1996.0116.
Dourson, M.L., Knauf, L.A., Swartout, J.C., 1992. On reference dose (RfD) and its underlying toxicity data base. Toxicol. Ind. Health 8, 171–189.
Dourson, M.L., Stara, J.F., 1983. Regulatory history and experimental support of uncertainty (safety) factors. Regul. Toxicol. Pharmacol. 3, 224–238.
Dourson, M.L., Teuschler, L.K., Durkin, P.R., Stiteler, W.M., 1997. Categorical regression of toxicity data: a case study using aldicarb. Regul. Toxicol. Pharmacol. 25, 121–129. http://dx.doi.org/10.1006/rtph.1996.1079.
Echeverria, D., White, R.F., Sampaio, C., 1995. A behavioral evaluation of PCE exposure in patients and dry cleaners: a possible relationship between clinical and preclinical effects. J. Occup. Environ. Med. 37, 667–680.
Felter, S.P., Dourson, M.L., 1998. The inexact science of risk assessment (and implications for risk management). Hum. Ecol. Risk Assess. . Int. J. 2, 245–251.
Gaylor, D.W., Kodell, R.L., 2000. Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal. 20, 245–250.
Haber, L.T., Maier, A., Gentry, P.R., Clewell, H.J., Dourson, M.L., 2002. Genetic polymorphisms in assessing interindividual variability in delivered dose. Regul. Toxicol. Pharmacol. 35, 177–197. http://dx.doi.org/10.1006/rtph.2001.1517.
Hasegawa, R., Hirata-Koizumi, M., Dourson, M.L., Parker, A., Sweeney, L.M., Nishikawa, A., Yoshida, M., Ono, A., Hirose, A., 2010. Proposal of new uncertainty factor application to derive tolerable daily intake. Regul. Toxicol. Pharmacol. 58, 237–242. http://dx.doi.org/10.1016/j.yrtph.2010.06.006.
Hattis, D., Baird, S., Goble, R., 2002. A straw man proposal for a quantitative definition of the RfD. Drug Chem. Toxicol. 25, 403–436. http://dx.doi.org/10.1081/DCT-120014793.
Helton, J.C., Burmaster, D.E., 1996. Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems. Reliab. Eng. Syst. Saf. 54, 91–94.
Hennessey, V.G., Rosner, G.L., Bast Jr., R.C., Chen, M.Y., 2010. A Bayesian approach to dose-response assessment and synergy and its application to in vitro dose-response studies. Biometrics 66, 1275–1283. http://dx.doi.org/10.1111/j.1541-0420.2010.01403.x.
Hertzberg, R.C., Dourson, M.L., 1993. Using categorical regression instead of a NOAEL to characterize a Toxicologist's judgment in noncancer risk assessment. In: Yang, R.S. (Ed.), H., (Ed.), Toxicology of Chemical Mixtures: Case Studies, Mechanisms and Novel Approaches. Academic Press, San Diego, CA.
Hill, R.A., 1996. From science to decision-making: the applicability of Bayesian methods to risk assessment. Hum. Ecol. Risk Assess. Int. J. 2, 636–642. http://dx.doi.org/10.1080/10807039609383641.
Hughes, D.A., Aronson, J.K., 2010. A systematic review and empirical analysis of the relation between dose and duration of drug action. J. Clin. Pharmacol. 50, 17–26. http://dx.doi.org/10.1177/0091270008329555.
Igarashi, T., Sekido, T., 1996. Case studies for statistical analysis of toxicokinetic data. Regul. Toxicol. Pharmacol. 23, 193–208. http://dx.doi.org/10.1006/rtph.1996.0043.
Igarashi, T., Yabe, T., Noda, K., 1996. Study design and statistical analysis of toxicokinetics: a report of JPMA investigation of case studies. J. Toxicol. Sci. 21, 497–504.
Jacobs, J.P., 2011. NAS Reviewers Slam EPA's Formaldehyde Assessment. New York Times. The New York Times Company, New York.
Jones, D.R., Peters, J.L., Rushton, L., Sutton, A.J., Abrams, K.R., 2009. Interspecies extrapolation in environmental exposure standard setting: a Bayesian synthesis approach. Regul. Toxicol. Pharmacol. 53, 217–225. http://dx.doi.org/10.1016/j.yrtph.2009.01.011.
Lehman, A.J., Fitzhugh, O.G., 1954. 100-Fold margin of safety. Assoc. Food Drug Off. Cl. S. Q. Bull. 18, 33–34.
Lewis, S.C., Lynch, J.R., Nikiforov, A.I., 1990. A new approach to deriving community exposure guidelines from "no-observed-adverse-effect levels". Regul. Toxicol. Pharmacol. 11, 314–330.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. Bioscience 51, 341–352.
Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K., Bridges, T., 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Sci. Total Environ. 407, 5199–5205. http://dx.doi.org/10.1016/j.scitotenv.2009.05.004.
Luszczki, J.J., 2010. Dose-response relationship analysis of pregabalin doses and their antinociceptive effects in hot-plate test in mice. Pharmacol. Rep. 62, 942–948.
Luszczki, J.J., Czuczwar, S.J., 2008. Dose-response relationship analysis of vigabatrin doses and their antinociceptive effects in the hot-plate test in mice. Pharmacol. Rep. 60, 409–414.
Meek, M.E., Newhook, R., Liteplo, R.G., Armstrong, V.C., 1994. Approach to assessment of risk to human health for priority substances under the Canadian Environmental Protection Act. Environ. Carcinog. Ecotoxicol. Rev. C12, 105–134.
Meek, M.E., Renwick, A., Ohanian, E., Dourson, M., Lake, B., Naumann, B.D., Vu, V., International Programme on Chemical, S, 2002. Guidelines for application of chemical-specific adjustment factors in dose/concentration-response assessment. Toxicology 181–182, 115–120.
NIOSH, 1976. Criteria for a Recommended Standard: Occupational Exposure to Methyl Alcohol. US Department of Health, Education and Welfare, Washington, DC.
NRC, 1994. Science and Judgement in Risk Assessment. National Academy Press, Washington, DC.
NRC, 2006. Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. The National Academies Press, Washington, DC.
NRC, 2007a. Models in Environmental Regulatory Decision Making. The National Academies Press, Washington, DC.
NRC, 2007b. Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget. The National Academies Press, Washington, DC.
NRC, 2009. Science and Decisions: Advancing Risk Assessment. National Academies Press, Washington, DC.
NRC, 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. National Academies Press, Washington, DC.
NRC, 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. The National Academies Press, Washington, DC.
Patlewicz, G., Simon, T., Goyak, K., Phillips, R.D., Rowlands, J.C., Seidel, S., Becker, R.A., 2013. Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul. Toxicol. Pharmacol. 65, 259–268. http://dx.doi.org/10.1016/j.yrtph.2012.12.008.
Patlewicz, G., Simon, T.W., Rowlands, J.C., Budinsky, R.A., Becker, R.A., 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul. Toxicol. Pharmacol. 71, 463–477. http://dx.doi.org/10.1016/j.yrtph.2015.02.011.
Pelekis, M., Krishnan, K., 2004. Magnitude and mechanistic determinants of the interspecies toxicokinetic uncertainty factor for organic chemicals. Regul. Toxicol. Pharmacol. 40, 264–271. http://dx.doi.org/10.1016/j.yrtph.2004.07.004.
Pelekis, M., Nicolich, M.J., Gauthier, J.S., 2003. Probabilistic framework for the estimation of the adult and child toxicokinetic intraspecies uncertainty factors. Risk Anal. 23, 1239–1255.
Pieters, M.N., Kramer, H.J., Slob, W., 1998. Evaluation of the uncertainty factor for subchronic-to-chronic extrapolation: statistical analysis of toxicity data. Regul. Toxicol. Pharmacol. 27, 108–111. http://dx.doi.org/10.1006/rtph.1997.1196.
Price, P.S., Keenan, R.E., Swartout, J.C., 2008. Characterizing interspecies uncertainty using data from studies of anti-neoplastic agents in animals and humans. Toxicol. Appl. Pharmacol. 233, 64–70. http://dx.doi.org/10.1016/j.taap.2008.03.026.
Price, P.S., Keenan, R.E., Swartout, J.C., Gillis, C.A., Carlson-Lynch, H., Dourson, M.L., 1997. An approach for modeling noncancer dose responses with an emphasis on uncertainty. Risk Anal. 17, 427–437.
Renwick, A.G., Lazarus, N.R., 1998. Human variability and noncancer risk assessment–an analysis of the default uncertainty factor. Regul. Toxicol. Pharmacol. 27, 3–20.
Rozman, K.K., Kerecsen, L., Viluksela, M.K., Osterle, D., Deml, E., Viluksela, M., Stahl, B.U., Greim, H., Doull, J., 1996. A toxicologist's view of cancer risk assessment. Drug Metab. Rev. 28, 29–52. http://dx.doi.org/10.3109/03602539608993990.
Shao, K., Gift, J.S., 2013. Model uncertainty and bayesian model averaged benchmark dose estimation for continuous data. Risk Anal. http://dx.doi.org/10.1111/risa.12078.
Simon, T., 2011. Just who is at risk? the ethics of environmental regulation. Hum. Exp. Toxicol. 30, 795–819. http://dx.doi.org/10.1177/0960327110379252.
Simon, T., Aylward, L.L., Kirman, C.R., Rowlands, J.C., Budinsky, R.A., 2009. Estimates of cancer potency of 2,3,7,8-tetrachlorodibenzo(p)dioxin using linear and nonlinear dose-response modeling and toxicokinetics. Toxicol. Sci. 112, 490–506. http://dx.doi.org/10.1093/toxsci/kfp232.
Simon, T., Kirman, C.R., Aylward, L.L., Budinsky, R.A., Rowlands, J.C., Long, T.F., 2008. Estimates of cancer potency of 2,3,4,7,8-pentachlorodibenzofuran using both nonlinear and linear approaches. Toxicol. Sci. 106, 519–537.
Simon, T.W., 2014. Environmental Risk Assessment: a Toxicological Approach. CRC Press; Taylor & Francis Group, Boca Raton, FL.
Slob, W., Setzer, R.W., 2014. Shape and steepness of toxicological dose-response relationships of continuous endpoints. Crit. Rev. Toxicol. 44, 270–297. http://dx.doi.org/10.3109/10408444.2013.853726.
Stedeford, T., Zhao, Q.J., Dourson, M.L., Banasik, M., Hsu, C.-H., 2007. The application of non-default uncertainty factors in the U.S. EPA's integrated risk information system (IRIS). Part I: UFL, UFS, and "Other uncertainty factors". J. Environ. Sci. Health, Part C 25, 245–279. http://dx.doi.org/10.1080/10590500701569430.
Swartout, J.C., Price, P.S., Dourson, M.L., Carlson-Lynch, H.L., Keenan, R.E., 1998.

A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. 18, 271–282.
Tatum, V., Wiegand, P., Stratton, S., Louch, J., Ebert, E., Anderson, P., 2015. Derivation of human health-based ambient water quality criteria: a consideration of conservatism and protectiveness goals. Integr. Environ. Assess. Manag. 11, 298–305. http://dx.doi.org/10.1002/ieam.1584.
Teuschler, L.K., Dourson, M.L., Stiteler, W.M., McClure, P., Tully, H., 1999. Health risk above the reference dose for multiple chemicals. Regul. Toxicol. Pharmacol. 30, S19–S26. http://dx.doi.org/10.1006/rtph.1999.1321.
USEPA, 1989. Risk Assessment Guidance for Superfund: Volume 1: Human Health Evaluation Manual (Part A) (RAGS). USEPA, Washington, DC.
USEPA, 1992. Guidance on Risk Characterization for Risk Managers and Risk Assessors.
USEPA, 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry.
USEPA, 2001. Risk Assessment Guidance for Superfund (RAGS) Volume III - Part A: Process for Conducting Probabilistic Risk Assessment. Washington, DC.
USEPA, 2002a. IRIS Toxicological Review of Phenol. USEPA, Washington, DC.
USEPA, 2002b. A Review of the Reference Dose and Reference Concentration Processes. USEPA, Washington, DC.
USEPA, 2011. Integrated Risk Information System (IRIS) Glossary. USEPA, Washington, DC.
USEPA, 2012a. Benchmark Dose Technical Guidance. USEPA.
USEPA, 2012b. IRIS Toxicological Review of Tetrachloroethylene (Perchlorethylene). USEPA.
USEPA, 2013a. IRIS Toxicological Review of Methanol (noncancer). USEPA.
USEPA, Toxicological, 2013b. Review of 1,4-Dioxane (with Inhalation Update) (CAS No. 123-91-1). USEPA, Washington, DC.
USEPA, 2014. Guidance for Applying Quantitative Data to Develop Data-derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. USEPA.
Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M., 2014a. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol. Sci. 142, 312–320. http://dx.doi.org/10.1093/toxsci/kfu199.
Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M., 2014b. Adverse outcome pathway development II: best practices. Toxicol. Sci. 142, 321–330. http://dx.doi.org/10.1093/toxsci/kfu200.
Von Stackelberg, K., Burmaster, D.E., 1993. Compounding Conservatisms–EPA's health risk assessment methods. HazMat World. March.
Waddell, W.J., 2003. Thresholds of carcinogenicity in the ED01 study. Toxicol. Sci. 72, 158–163. http://dx.doi.org/10.1093/toxsci/kfg004.
Waddell, W.J., 2008. Thermodynamic basis for expressing dose logarithmically. Toxicol. Appl. Pharmacol. 228, 156–157. http://dx.doi.org/10.1016/j.taap.2007.12.004.
Waddell, W.J., 2010. History of dose response. J. Toxicol. Sci. 35, 1–8.
WHO-IPCS, 2005. Chemical-specific Adjustment Factors for Interspecies Differences and Human Variability: Guidance Document for Use of Data in Dose/Concentration Assessment. WHO, Geneva, Switzerland.
WHO-IPCS, 2014. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. WHO, Geneva, Switzerland.
Wilson, A., Reif, D.M., Reich, B.J., 2014. Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals. Biometrics 70, 237–246. http://dx.doi.org/10.1111/biom.12114.
Zeise, L., Bois, F.Y., Chiu, W.A., Hattis, D., Rusyn, I., Guyton, K.Z., 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ. Health Perspect. 121, 23–31. http://dx.doi.org/10.1289/ehp.1205687.
Zhu, Y., Wang, T., Jelsovsky, J.Z., 2007. Bootstrap estimation of benchmark doses and confidence limits with clustered quantal data. Risk Anal. 27, 447–465. http://dx.doi.org/10.1111/j.1539-6924.2007.00897.x.