OXFORD
SOT | Society of Toxicology
www.toxsci.oxfordjournals.org
ToxSci 20 Years

TOXICOLOGICAL SCIENCES, 165(1), 2018, 74–89

doi: 10.1093/toxsci/kfy124
Advance Access Publication Date: June 12, 2018
Research Article

Dose-Related Severity Sequence, and Risk-Based Integration, of Chemically Induced Health Effects

Salomon Sand,*,[1] Roland Lindqvist,* Dietrich von Rosen,†,‡ and Nils-Gunnar Ilbäck*

*Department of Risk-Benefit Assessment, Swedish National Food Agency, SE-75126 Uppsala, Sweden; †Department of Energy and Technology, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden; and ‡Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden

[1]To whom correspondence should be addressed at Department of Risk-Benefit Assessment, Swedish National Food Agency, PO Box 622, SE-75126 Uppsala, Sweden. Fax: +46 (0)18 10 58 48; E-mail: salomon.sand@slv.se.

ABSTRACT

Risk assessment of chemical hazards is typically based on single critical health effects. This work aims to expand the current approach by characterizing the dose-related sequence of the development of multiple (lower- to higher-order) toxicological health effects caused by a chemical. To this end a "reference point profile" is defined as the relation between benchmark doses for considered health effects, and a standardized severity score determined for these effects. For a given dose of a chemical or mixture the probability for exceeding the reference point profile, thereby provoking lower- to higher-order effects, can be assessed. The overall impact at the same dose can also be derived by integrating contributions across all health effects following severity-weighting. In its generalized form the new impact metric relates to the probability of response for the most severe health effects. Reference points (points of departure) corresponding to defined levels of response can also be estimated. The proposed concept, which is evaluated for dioxin-like chemicals, provides an alternative for characterizing the low-dose region below the reference point for a severe effect like cancer. The shape and variability of the reference point profile add new dimensions to risk assessment, which for example extends the characterization of chemical potency, and the concept of acceptable effect sizes for individual health effects. Based on the present data the method shows high stability at low doses/responses, and is also robust to differences in severity categorization of effects. In conclusion, the novel method proposed enables risk-based integration of multiple dose-related health effects. It provides a first step towards a more comprehensive characterization of chemical toxicity, and suggests a potential for improved low-dose risk assessment.

Key words: risk assessment; BMD; severity; risk-based; integrated; multiple effects.

The term "health risk assessment" carries the notion that it involves determination of the probability of occurrence, and the severity, of an adverse health outcome. In chemical risk assessments, however, margin of exposure related concepts are generally applied. Estimates of the human exposure to a chemical are then compared, in one way or another, to a reference level: ie, a health-based guidance value, or a reference point (RP) (a term by the European Food Safety Authority [EFSA], 2009) also called point of departure (POD). The RP is based on the critical health effect observed in the pivotal study, eg, derived by the benchmark dose (BMD) approach (Crump, 1984; Dourson *et al.*, 1985; EFSA, 2005, 2009, 2017; U.S. Environmental Protection Agency [EPA] 2005, 2012).

For nongenotoxic effects there may be a health concern if the estimated human exposure to a chemical exceeds the health-based guidance value, which is established by the application of adjustment factors to the RP. A default overall adjustment factor of 100 is for example used to account for inter- and intraspecies differences in susceptibility, which was already introduced in the 1950s (Lehman and Fitzhugh, 1954) indicating a

long-lasting practice. Risk assessment is more stringent in the case of genotoxic effects since the absence of chemical-specific thresholds is then assumed. For example, EFSA considered that a margin of exposure of 10 000 or higher would be of "low concern" for compounds that are both genotoxic and carcinogenic (EFSA, 2005).

The critical effect used for development of a health-based guidance value is defined as the first adverse effect or its known precursor (EPA, 2002). Intuitively, this would protect from effects occurring at higher doses that may be more serious. As a rough example, according to EPA's integrated risk information system the search term "mortality" or "survival" appear in 20 text strings describing the critical effect considering 661 reference doses/concentrations; "lesion", "degeneration", "pathology"/"pathologic", or "function" appear in 129 of the remaining (641) text strings, and may represent various severities; and "weight" (eg, as in body/organ weight) appears in 212 of the remaining (512) text strings (EPA, 2018). Selection of critical effects is limited by the available data. The nature and severity of the chosen critical effect may therefore differ between assessments. EPA generally addresses this by an uncertainty factor for incompleteness of the database if required information on eg, different life stages and organ systems are considered missing (EPA, 2002). EFSA (2012) recommends that the possibility/feasibility to derive additional data is first considered before case by case justification of an uncertainty factor.

Application of extra adjustment factors, or similar, specifically for the nature and/or severity of the critical health effect is with a few exceptions, including genotoxicity, not applied even though guidelines note this possibility (World Health Organization, International Programme on Chemical Safety [WHO/IPCS], 2009; European Chemicals Agency [ECHA], 2012; EFSA, 2012), and "severity" is part of the Codex definition of risk characterization (Food and Agriculture Organization of the United Nations, World Health Organization [FAO/WHO], 2008). To compensate for differences in the severity of the critical effect the Swedish National Food Agency introduced the "Risk Thermometer" (NFA, 2015)—a tool based on a severity-adjusted margin of exposure approach involving systematic use of a management-based (extra) adjustment factor. Inconsistencies that might result if severity is not accounted for has been pointed out previously (eg, Bogdanffy *et al.*, 2001), and the possibility of using adjustment factors to account for dose-related differences between effects that differ in severity has been mentioned (Renwick *et al.*, 2004). This work aims to advance the consideration of adjustment factors by quantifying dose-severity relations in the process of deriving a toxicological impact measure that accounts for multiple health effects spanning different severities.

To support risk assessment the adverse outcome pathway (AOP) concept has for example been introduced as a framework within which data and knowledge collected at many levels of biological organization can be synthesized (Ankley *et al.*, 2010; Organization for economic cooperation and development [OECD], 2013). An AOP begins at the molecular level where a chemical interacts with a biological target, leading to a sequential series of higher-order events to produce an adverse health effect (Ankley *et al.*, 2010). As a more descriptive approach, it is here hypothesized that the total amount of different health effects (occurring) in a target organ increases with dose, and that the severity of these different events may also gradually increase and interact across the dose continuum. This concept introduced here is generally termed "sequential severity of toxicity" (SST).

As a starting point the SST concept is implemented as a generalization of the current approach for hazard and risk characterization. Instead of using a specific RP (for a single critical health effect) as a basis for development of human exposure guidelines, the sequence of RPs (or PODs) associated with lower- to higher-order health effects, representing low to high severities, is attempted to be characterized. This sequence of RPs is called reference point profile (RPP). Although not mechanistically describing an AOP, or similar, the RPP would satisfy evolved Bradford Hill considerations in terms of dose-response concordance between key events, which is essential but not necessarily sufficient for a hypothesized mode of action (Meek *et al.*, 2014).

Quantitative modeling of data across different severities has been performed by using regression methods designed to handle categories (discrete data). Hertzberg and Miller (1985) and Hertzberg (1989) introduced such approaches to address species extrapolation. These methods have been discussed as superior compared with only considering the NOAEL for the critical effect since they enable incorporation of more information, and assessment of risks above the reference dose (Hertzberg and Dourson, 1993). The latter issue has been addressed in more detail (Dourson *et al.*, 1997; Teuschler *et al.* 1999), and analysis stratified over groups (eg, sex, species) has also been discussed (Guth *et al.*, 1997). Recently, a severity scoring system for manganese toxicity was developed by an international panel (Mattison *et al.*, 2017), which was used for characterizing the U-shaped dose-response (Milton *et al.*, 2017). The level of toxicity (associated with excess) was described by a 9-graded severity scale. In this work the severity domain of health effects is also defined by a 9-graded scale according to which chemically induced health effects more generally can be ranked.

The regression methods discussed earlier have been used to calculate the probability for a given severity category. As one part, the new methodology proposed in this article also provides this type of output but simultaneously across all categories. Specifically, the probability for exceeding the RPP (the sequence of RPs associated with lower- to higher-order health effects) across the severity domain can be assessed. In addition, results (probabilities) are integrated over the entire severity domain to produce an overall impact measure, RTR. This additional step requires weighting of severity categories, and the developed system with 9 categories is therefore mapped to a quantitative severity scale, $S = 0$ to 1. As shown, this enables that quantitative severity-weighting of (dose-related) health effects can be performed using a (modifiable) continuous function, with few parameters, rather than having to weight the severity of various effects, in relation to each other, one at a time. The overall impact measure, RTR (probability $\times$ severity integrated across S), can be derived for given exposure conditions, and stratified over different groups (eg, different chemicals). Conversely, an RP (or POD) associated with a specified level of the new response metric can be estimated, eg, representing a starting point for exposure guideline development that accounts for multiple effects. The proposed method is illustrated and evaluated using dioxin-like chemicals as a case study, and a generalization of the method to the level of the overall dose-response is also presented.

MATERIALS AND METHODS

Dose-Response Data

The concept of SST is introduced by considering organ-specific toxicity data. Data available from the U.S. National Toxicology Program (NTP) on dioxin-like chemicals, and their mixtures,

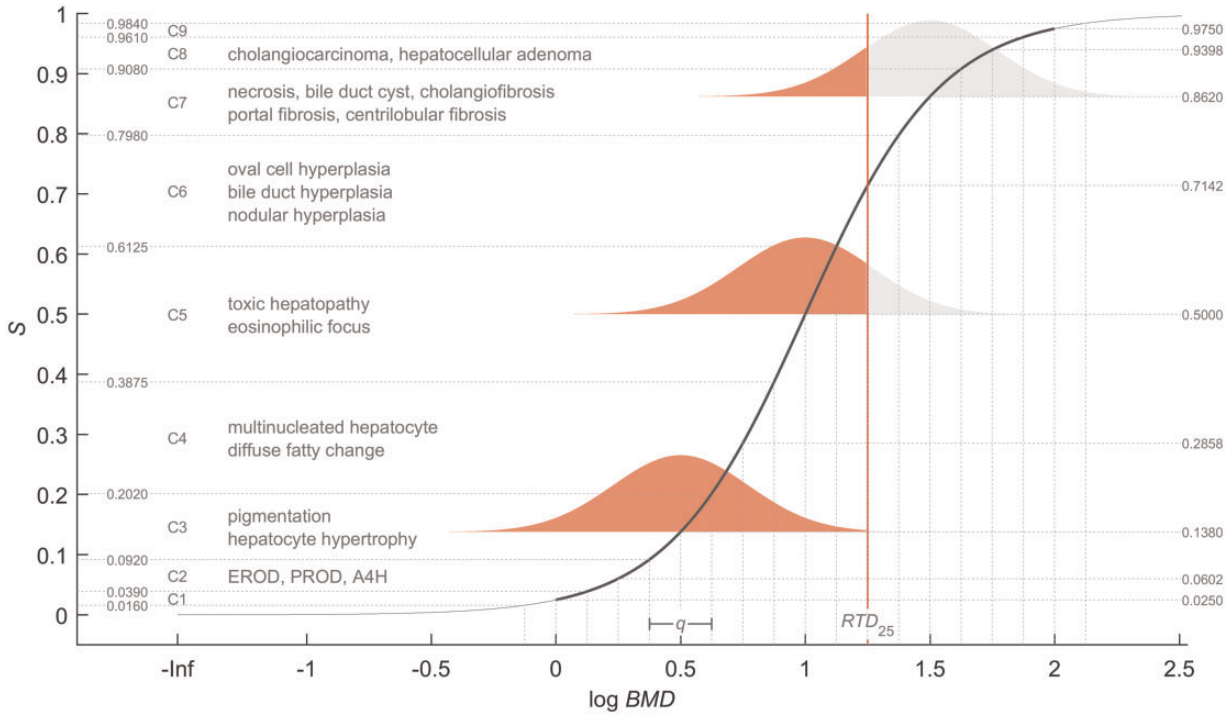

Figure 1. Technical illustration of the RPP, which is a cross-section of the dose-severity-response volume. The RPP (s-shaped curve) describes the relation between the *BMD* for selected health effects, and the severity of toxicity (S) determined for these effects. Health effects are categorized by a 9-graded scale, C1–C9, for increasing severity of toxicity (Supplementary Table 2). Each severity category is mapped to an interval of S-values. For derivation of the default severity-weighting scheme, C1 and C9 are centered at S = 0.025 and S = 0.975, respectively, and S-intervals are then constructed in order to correspond to constant dose intervals, *q*, according to a logistic cumulative distribution function. The variation in the RPP is assumed to be normally distributed on the log-scale with constant variance. In this hypothetical example, the RPP model 1a location (H), shape (λ), and standard deviation (σ) is 10, 1.6, and 0.625, respectively. The RPP characterizes the (low) dose region below the *BMDs* for the most severe health effects. Areas (fractions of the normal distribution) describe probabilities, *p*, for exceeding the RPP, provoking lower- to higher-order toxicological effects, at an exposure level corresponding to the vertical line. In this example the vertical line represents RTD_{25} which is the dose corresponding to a RTR of 0.25 (50%). The RTR is the average severity-weighted probability of exceeding the RPP on the interval S = 0 to S = 1 (equation 3).

were selected since it covers a wide range of long-term health effects in the liver describing lower- to higher-order target organ toxicity. Changes in enzyme activity, nonneoplastic and neoplastic liver lesions were considered, comprising a total of 117 dose-response datasets across 6 NTP studies (NTP, 2006a–e, 2010) (2, 3, 7, 8-tetrachlorodibenzo-*p*-dioxin [TCDD], 2, 3, 4, 7, 8-pentachlorodibenzofuran [PeCDF], 3, 3′, 4, 4′, 5-pentachlorobiphenyl [PCB_{126}], 2, 3′, 4, 4′, 5-pentachlorobiphenyl [PCB_{118}], and 2- and 3-component mixtures thereof). The data observed in female Harlan Sprague-Dawley rats at 53 weeks or at the end of the 2-year studies can be found in Supplementary Table 1A and B.

General Description of the RPP

The RPP is defined as the relationship between the RP, eg, the *BMD*, for different (organ-specific) health effects, and the severity of toxicity (S), determined for these health effects. The severity of toxicity is first determined categorically and then mapped to a quantitative scale that range from S = 0 to S = 1. The relation between RP and S is assumed to be sigmoidal on the log-dose scale. S will then monotonously approach 0 (no toxicity) or 1 (maximum severity of toxicity) as the dose/RP approaches zero or becomes very high, respectively, which is regarded to be in line with general toxicological principles.

The RPP can be considered as a cross-section of a dose-severity-response volume where x, y, and z is the dose, the severity of the health effect, and the change in response (ie, using a standardized benchmark response level [BMR] that may range between 0 and 1) associated with the health effect, respectively. The present analysis primarily considers the x-y cross-section at a BMR (z) of 21%, denoted RPP_{21}, describing the BMD_{21} profile (*BMDs* corresponding to a BMR = 0.21). The BMD_{21} represents a change point on the dose-response curve in the low-dose region (Sand *et al.*, 2006, 2012). Since the BMR selection impacts on the analysis RPPs associated with BMRs of 0.1 (RPP_{10}) and 0.4 (RPP_{40}) were also considered. The approach for BMD estimation is described in the Supplementary Material.

Independent from the BMD analysis toxicological findings in the target organ, referred to as "health effects", are categorized by a nine-graded scale, C1 through C9, for increasing severity of toxicity. A hierarchical severity classification scheme has been developed using NFA (2015) as a starting point (see Supplementary Table 2). The scheme organizes general descriptions of health effects in 5 blocks, which are linked to the 9-graded severity scale used for specific health effects. Supplementary Table 2 provides general guidance for selection of C1 to C9 based on toxicological judgment. Details of the RPP concept that uses estimated *BMDs* and associated C1 to C9 categories mapped to S is described in sections below, and also visualized in Figures 1–3 (the classification of studied health effects is also shown in Figure 1, and further commented in the result section).

Default Mapping of Severity Categories to S

Mapping severity categories to S-values involves the questions of how much C1–C9 should cover S, and how C1–C9 should be distributed across S. It is shown in this article that the new

health impact metric introduced in equation 3 reaches ≈ 50% (ie, the "center" of the studied toxicological process/sequence of health effects) at S = 0.71, independent of the RPP model parameters. Motivated by this observation the default mapping described below distributes severity categories symmetrically across S and is designed so that the midpoint of C6 approximates to the S-value of 0.71 (this issue is further discussed in association to equation 3). The default mapping is a reference scenario that can later be modified in a systematical manner by the severity-weight in equation 3 without having to reevaluate the RPP.

1. Each severity category, C1 to C9, is associated with an interval of S-values. First, C1 and C9 categories are centered at S = 0.025 and 0.975, respectively (Figure 1). In this context C1 through C9 health effects are allowed to cover 95% of the quantitative S-domain between 0 and 1, which is denoted CS-95.
2. Symmetrical distribution of severity categories across S is then achieved by constructing S-intervals associated with C1–C9 that correspond to constant dose intervals, q (width of the interval), along the log *BMD*-axis according to a symmetrical RPP (Figure 1). Three common symmetrical cumulative distributions functions (CDFs) were here considered (descriptively): the logistic, normal and Student's t CDFs. The selected S-interval construct is based on the logistic CDF since S-intervals are similar for the 3 CDFs, and results based on the logistic CDF are closest to the mean across CDFs (see Supplementary Table 3).

RPP Models

Besides satisfying general toxicological principles, as discussed as part of the general description of the RPP, a sigmoidal (symmetrical/asymmetrical) RPP was considered due to its flexibility: it can handle both linear and nonlinear relations between log *BMD* and S in the range of C1 to C9. The general equation for the RPP is,

$$\log BMD^{ijk} = \mu(S_{ij}) + \varepsilon_{ijk}, \tag{1}$$

where $\mu(S_{ij})$ is the sigmoidal RPP. S_{ij} is a value of the severity of toxicity within the S-interval linked to health effects (k) classifying in the j'th severity category (j = 1, 2,.9) for the i'th chemical/mixture. The assignment of S_{ij}-values associated severity categories C1–C9 is described in the algorithm for estimating the RPP ("RPP estimation" section). The (error) term, ε_{ijk}, mainly describes the natural variation in log BMD^{ijk} (the *BMD* for the k'th health effect classifying in the j'th severity category for the i'th chemical/mixture) assumed to be normally distributed with constant variance, σ^2. Variability (σ^2) results for example since *BMD*s may differ for health effects within the same severity category. The RPP is illustrated with the "response variable", *BMD*, on the x-axis and the "explanatory variable", S, on the y-axis (Figure 1). Three specific RPP models with different geometrical characteristics were considered,

$$\mu(S_{ij}) = H_i \cdot \left[\frac{S_{ij}}{1-S_{ij}}\right]^{\frac{1}{\lambda_i}}, \quad (1a,\ \text{symmetrical RPP})$$

$$\mu(S_{ij}) = H_i \cdot \left[\log\frac{1}{1-S_{ij}}\right]^{\frac{1}{\lambda_i}}, \quad (1b,\ \text{left}-\text{skewed RPP})$$

$$\mu(S_{ij}) = H_i / \left[\log\frac{1}{S_{ij}}\right]^{\frac{1}{\lambda_i}}, \quad (1c,\ \text{right}-\text{skewed RPP})$$

where H_i, is the RPP (dose) location parameter for the i'th chemical/mixture, and describes potency based on several health effects (H_i decreases with increasing potency). Parameter, λ_i, is the RPP shape for the i'th chemical/mixture, and describes how tightly *BMD*s are separated across the S-domain. Model 1a is a Hill model/curve (as a function of S), which is used as the standard RPP model. Model 1b is a Weibull curve, and model 1c corresponds to the limiting case of the Richards curve when the asymmetry term approaches infinity (Sand *et al.*, 2006). The low-dose geometry/characteristic of model 1b corresponds to the high dose geometry of model 1c, and vice versa. Models 1a–c are used to assess the stability of the method across RPP curves, covering symmetry as well as asymmetry in both directions. For S-intervals in Figure 1, models 1a–c cover assumptions of a q that is constant, decreases, and increases across the log dose scale, respectively (Supplementary Figure 1).

Severity-Probability Profile

The probability to exceed the RPP along the entire S-domain at a given or estimated exposure, E_i, to a chemical or mixture is described by,

$$p(E_i|S_{ij}) = \phi\left[\frac{\log E_i - \log\frac{\widehat{\mu}(S_{ij})}{AF_i}}{\widehat{\sigma}}\right], \tag{2}$$

where ϕ is the cumulative standard normal distribution function; $\widehat{\mu}(S_{ij})$ is based on estimates for $\widehat{H}_i$ and $\widehat{\lambda}_i$; and $\widehat{\sigma}$ is an unbiased estimator of the standard deviation. Similar to traditional risk assessment $\widehat{\mu}(S_{ij})$ can be divided by an overall standard or chemical-specific adjustment factor (AF_i) (WHO/IPCS, 1994). Extrapolation from cell- or animal-based test systems to humans is not addressed in this study, and for simplicity AF_i is set to 1. If substituting E_i with $\widehat{\mu}(S_i)$ and using $AF_i = 1$ equation 2 can be rewritten as,

$$p(E_i|S_{ij}) = \phi\left[\frac{\log Q_i}{\widehat{\lambda}_i \cdot \widehat{\sigma}}\right],$$

where

$$Q_i = \frac{S_i}{1-S_i}/\frac{S_{ij}}{1-S_{ij}} \ (1a); \quad Q_i = \log\frac{1}{1-S_i}/\log\frac{1}{1-S_{ij}} \ (1b);$$

$$Q_i = \log\frac{1}{S_{ij}}/\log\frac{1}{S_i} (1c).$$

Thus, the probability to exceed the RPP at an exposure, E_i, expressed in terms of S_i depends only on the product between the RPP shape (λ_i) and standard deviation (σ).

Risk-based and toxicity-integrated response (RTR), and its dose equivalent (RTD)

The RTR associated with exposure E_i is defined as the integral of the probability of exceeding the RPP (equation 2) multiplied by the severity-weight, w, here described by a beta cumulative distribution function that increases from zero to 1 across S,

$$\mathrm{RTR}(E_i) = \int_0^1 p(E_i|S_{ij}) \cdot w(S_{ij})\ dw,$$
$$w(S_{ij}|a,b) = \int_0^{S_{ij}} t^{a-1}(1-t)^{b-1}\ dt,\ a>0,\ b>0,\ a \text{ or } b=1. \quad (3)$$

The RTR equals the average severity-weighted probability ($p \times w$) of exceeding the RPP at E_i on the interval $S = 0$ to $S = 1$, and approaches 0.5 as E_i becomes large (2 × RTR standardizes this metric to range between 0 and 1). The *RTD* is the RP (or POD) corresponding to a specified RTR. In terms of impact, the RTR corresponds to the (average) probability of exceeding the *BMD* for the most severe health effects. The beta cumulative distribution function was used to describe the severity-weight, w, since this model is defined on the interval [0, 1]. The parameter space for this function, $a > 0$ and $b > 0$, is here constrained (a or $b = 1$) to enables indirect modification of the (default) mapping of severity categories to S at the level of the RTR in a systematic manner, which is further discussed below.

As shown in Figure 2A the relation between S and the RTR is nonlinear even for a linear severity-weight. This results due to the difference between separate (reflected by S) and integrated (reflected by RTR) consideration of dose-related health effects. There exists an S-value, S-25, so that $w(S\text{-}25|a,b) = \sqrt{0.5}$ approximately corresponds to RTR = 0.25 (50%) regardless of the RPP model parameter values (Figs. 2A and 2B). S-25 aligns to the center of the studied toxicological process/sequence [that may be mechanistically related or not] in terms of RTR. As noted, the default mapping of severity categories to S (corresponding to a linear w) is designed so that the midpoint of C6 calibrates to S-25 (Figure 2A). In this study, category C6 is regarded as a region that separates reversible and irreversibility health effects (hyperplasia, Figure 1). Calibration of S-25 to C5 (e.g., toxic hepathopathy) or C7 (eg, necrosis) may be too conservative and too liberal, respectively. As shown in Figure 2B, S-25 increases if $a > 1$ (and $b = 1$), which corresponds to that C1 to C9 are negatively skewed across S, and the opposite results if $b > 1$ (and $a = 1$). The degree of asymmetry (skew) associated with S-25 can be described by $w(0.5|a,b)$ (Figure 2B).

Quantitative severity-weighting can primarily be performed by selection of S-25 as described in Figure 2B, which reduces this complex problem to (value-based) selection of a single parameter rather than having to individually weight each category versus others. It is also a systematic approach since S-25 corresponds to a given direction and degree of skew in relation to the default symmetrical mapping of severity categories to S. Strictly speaking, besides selection of S-25, severity-weighting could be further elaborated by relaxing the restriction a or $b = 1$ in equation 3 as a second tier (ie, allow w to assume s-shaped forms), and modify each curve in Figure 2B while keeping S-25 constant. This extension represents a rather detailed level; however, and is not illustrated herein. Although a linear weight is regarded as a general default where severity category C6 is associated with RTR = 50% the use of alternatives is assessed.

Algorithm for Estimating the RPP, RTR, and RTD

The RTR/*RTD* is by default based on RPP_{21}, while RTRs/*RTD*s based on the RPP_{10} and RPP_{40} are derived for comparison. RPP_{10} and RPP_{40} are located at lower and higher doses compared with RPP_{21}, respectively. RPPs associated with alternative BMRs differ mainly in RPP location (H). The RTR associated with E_i based on RPP_{10} is therefore higher than the corresponding RTR based on RPP_{40}, and the opposite applies for the *RTD* associated with a specific RTR. The algorithm for estimating the RPP (associated with a specified BMR), the RTR (associated with a given E_i) and the *RTD* (associated with a specified RTR) is described below.

RPP estimation. A 2-step approach was applied for estimating the RPP: *BMD*s were first derived separately, and then used as inputs for characterizing the RPP. The algorithm described below accounts for both the uncertainty in the *BMD* and the sensitivity in the link between severity categories (C1–C9) and S.

1. Log-normal uncertainty distributions for the *BMD* are estimated for each chemical/mixture (i) and health effect (k) selected for RPP characterization. In total, *BMD*s estimated from 100 significant dose-response datasets across 6 chemicals/mixtures were included in the analyses, describing 19 health effects in the liver. The approach for *BMD* estimation is described in the Supplementary Material.

2a. Using a parametric bootstrap approach, a BMD^{ijk} value corresponding to a specific BMR (BMR = 0.21 is default) is randomly generated from the associated distribution estimated in step 1. The BMD^{ijk} is then matched with a S_{ij}-value randomly generated from the uniform S-interval associated with the j'th severity category in Figure 1. This is repeated across all chemicals and health effects associated with a specific RPP analysis.

2b. Model 1a, b, or c is then fitted by the least squares method: the sum of squares of $\log BMD^{ijk} - \log\mu(S_{ij})$ is minimized with respect to H_i and λ_i, which is a usual regression problem with explicit solutions. The estimate of $\hat{\sigma}^2$ is also obtained from which an unbiased estimator of $\hat{\sigma}$ can be derived (Sand *et al.*, 2003).

2c. Steps 2a and b are repeated $n_1 = 1000$ times, providing a 90% CI for the RPP. Median values of $\widehat{H}_i$, $\widehat{\lambda}_i$, and $\hat{\sigma}$ are used to represent a point estimate of the whole RPP, denoted "RPP point estimate", and median values of $\widehat{H}_i$ and $\widehat{\lambda}_i$ are used to represent a point estimate of the central RPP, denoted "central RPP point estimate".

RTR estimation.

3. The RTR in equation 3 associated with exposure level E_i is calculated by numerical integration using the RPP point estimate that provides an RTR point estimate, and using estimates of $\widehat{H}_i$, $\widehat{\lambda}_i$, and $\hat{\sigma}$ across n_1 stored iterations (from step 2c) from which a 90% of the RTR (RTR_L, RTR_U) is derived.

RTD estimation. The *RTD* for a specified RTR is derived by spline interpolation after estimating a number of RTRs, according to step 3, within a specified dose interval, I. A 2-step approach is used that first derives a rough estimate, which is later refined for increased precision.

4. For the present data, and target RTR levels considered, setting the upper limit of I to the dose corresponding to $S = \sqrt{RTR}$ according to the central RPP point estimate, and setting the lower limit (close) to zero, was found to provide a reasonable start interval. In this study, RTR point estimates associated with $n_2 = 25$ doses (E_is) spanning I are derived (ie, step 3 partly evaluated n_2 times) and an intermediary *RTD* point estimate (I-*RTD*) is solved by spline interpolation at the specified RTR. A refined interval, I_{refined}, is then defined with lower and upper limit corresponding to I-*RTD* divided and multiplied by a factor, f, respectively ($f = 4$ in this study). RTR point estimates, and 90% CIs (RTR_L and RTR_U), associated with $n_3 = 25$ doses (E_is) spanning I_{refined} are derived (ie, step 3 is fully evaluated n_3 times).

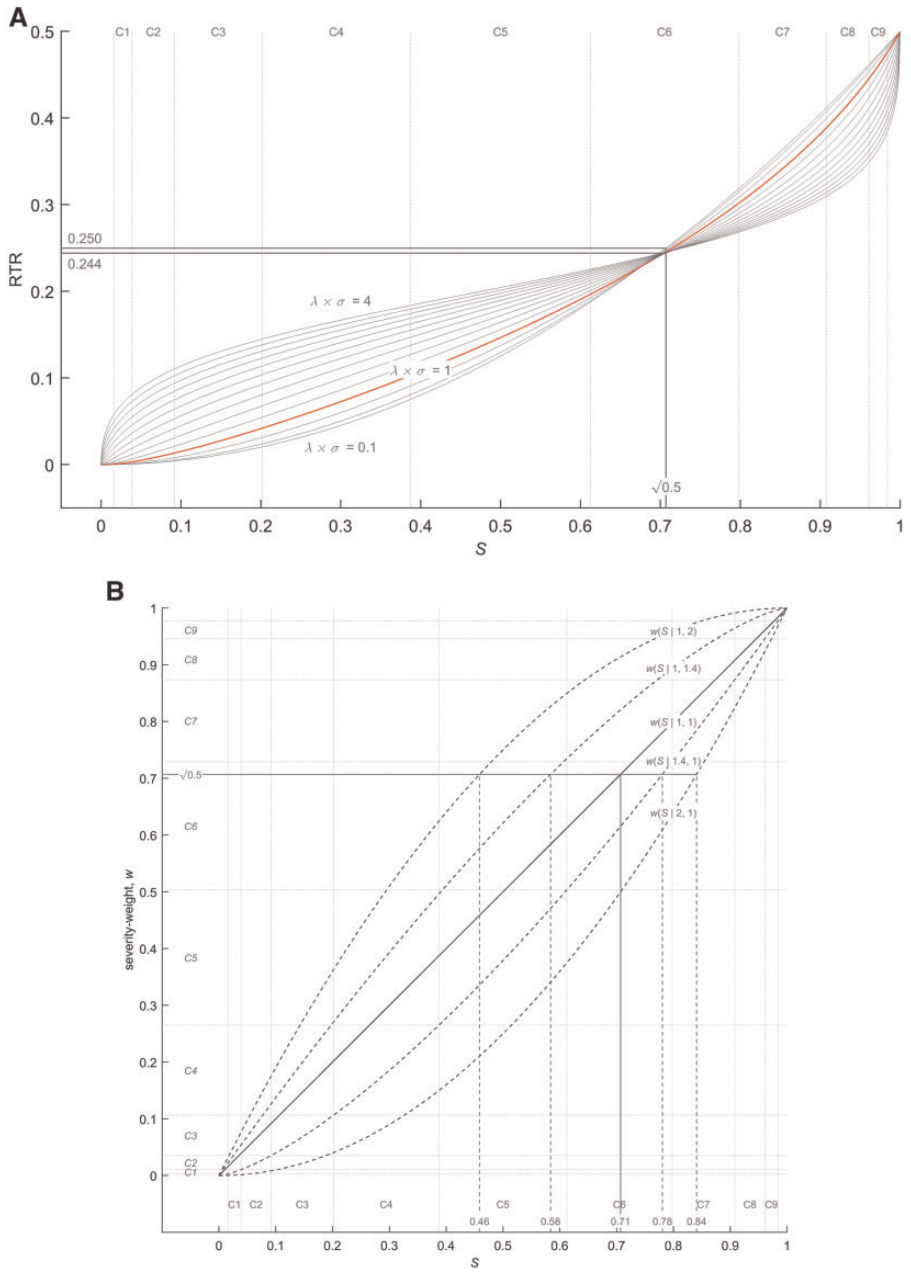

Figure 2. (A) The relation between the severity of toxicity, S, and the RTR, according to RPP model 1a. This relation depends on the product between the RPP shape (λ) and standard deviation (σ): $\lambda \times \sigma = 1$ represents the RPP in Figure 1. The *S*-value (S-25) corresponding to RTR $\approx$ 0.25 (50%) is independent of the RPP model parameters. S-25 is associated with the center of the toxicological process, or sequence of effects, studied, in terms of RTR. For a linear severity-weight $S = \sqrt{0.5}$ (close to the midpoint of C6) corresponds to RTR $\approx$ 0.244 (minimum) for $\lambda \cdot \sigma \approx 2$, and the RTR approaches 0.250 (maximum) as $\lambda \cdot \sigma$ decreases or increase from ≈ 2. (For model 1b the corresponding values are 0.250 and 0.258, and for RPP model 1c the corresponding values are 0.237 and 0.250, data not shown). (B) Illustration of linear (solid line) and nonlinear (dotted curves) severity-weights, $w(S_{ij}|a, b)$, that follow a beta cumulative distribution function. S-25 corresponding to RTR $\approx$ 0.25 (50%) is illustrated for each severity-weight. S-25 is higher than $\sqrt{0.5}$ when $a > 1$ and $b = 1$. This corresponds to a negatively skewed mapping of severity categories to S. The opposite results for $a = 1$ and $b > 1$. The value of w at $S = 0.5$, $w(0.5|a, b)$, describes skewness, which is 0.25, 0.38, 0.5, 0.62, and 0.75 for (a, b) equal to (2, 1), (1.4, 1), (1, 1), (1, 1.4), and (1, 2), respectively. The grid illustrates how *S*-intervals associated with severity category C1 to C9 indirectly becomes modified in the case of $w(S_{ij}|1.4,\ 1)$.

The *RTD* point estimate, and the lower 5th and upper 95th confidence bounds are then solved at the specified RTR by spline interpolation, and controlled to be within $I_{refined}$: ie, the relations $I_{refined}$ versus RTR point estimate, $I_{refined}$ versus RTR_L, and $I_{refined}$ versus RTR_U are each solved at the specified RTR.

Sensitivity Analysis of Severity Classification

Sensitivity analyses were performed to investigate how a difference in C1–C9 severity categorization influences the *RTD*, using RPP model 1a, and a linear severity-weight, since this classification involves subjective judgment. In these analyses, *BMD* point estimates in combination with *S*-values centrally located in associated severity categories (central *S*-values, see Figure 1 and Supplementary Table 3) are initially used in the algorithm for *RTD* estimation described previously (ie, results across $n_1 = 1000$ iterations intends in this case to portray different classifications rather than the uncertainty associated with a given classification). Starting from the proposed classification (Figure 1) 25% of

*BMD*s associated with a given RPP dataset are randomly selected in each of the $n_1 = 1000$ iterations of the algorithm. Selected *BMD* are then reclassified as described below.

SA1 (nonsystematic reclassification). Selected *BMD*s in a given iteration are reclassified individually, 1 category upward or downward, at random, and the *RTD* is estimated.

SA2 (systematic reclassification). Selected *BMD*s in a given iteration are reclassified as a group, 1 category upward or downward, at random, and the *RTD* is estimated.

BMD classifications associated with the 5th and 95th percentile of the *RTD* (P5 and P95) are identified in SA1 and SA2, respectively. Complete *RTD* estimation, utilizing *BMD* uncertainty distributions as well as associated *S*-intervals, are then performed for these *BMD* classifications. The potential impact of changing classification for 25% of *BMD*s to adjacent severity categories is described by the relative difference between the *RTD* associated with the proposed classification, and *RTD*s associated with P5 or P95. This is regarded to reflect a minor difference in classification. The ratio between *RTD*s associated with P5 and P95 describe the potential impact of changing classification for up to 50% of *BMD*s to adjacent severity categories, and/or a difference in classification for up to 25% of *BMD*s that span 2 severity categories. This is regarded to reflect at least a moderate difference in classification.

Generalized Methodology for RPP Characterization and Estimation of 2D-RTR

A generalized methodology was developed, which is described in detail in the Supplementary Material. According to Supplementary equation 4 the estimated RPP associated with a given BMR [estimates of $\widehat{H}_i$, $\widehat{\lambda}_i$, and $\widehat{\sigma}$ across n_1 stored iterations derived according to the algorithm for RPP estimation above] may be dose-adjusted by multiplying $\widehat{H}_i$ by a factor, Hx, to approximate RPPs associated with other BMRs. Hx is calculated from the shapes/slopes of dose-response curves obtained in the underlying *BMD* analysis of health effects associated with a given RPP, and the shapes for all curves are allowed to contribute equally to the shape of the averaged curve (see Supplementary Material). Derivation of RPPs associated with BMRs = 0–1 by this indirect/approximate approach provides a description of the dose-severity-response volume. RTRs for a given exposure, E_i, can be integrated across BMR levels resulting in 2D-RTR that accounts for both the severity and response domains,

$$2D - \mathrm{RTR}(E_i) = \int_0^1 \mathrm{RTR}\ \mathrm{dBMR}, \tag{4}$$

where the RTR associated with E_i is calculated according to the algorithm for RTR estimation described above (step 3) across RPPs spanning the BMR domain. As an extension of this algorithm equation 4 is then calculated by numerical integration. The algorithm for *RTD* estimation is generalized in a similar manner. As E_i gradually increases the maximum RTR = 0.5 will be reached for higher and higher BMR values, and eventually RTR is 0.5 across all BMRs. The 2D-RTR, thus, approaches 0.5 when E_i becomes large since the area under the curve is then 0.5×1. As mentioned, in terms of impact the RTR corresponds to the (average) probability of exceeding the *BMD* for the most severe health effects (equation 3). 2D-RTR then corresponds to the (average) probability of exceeding the dose-response curves, ie, the *BMD* associated with BMR levels of 0 through 1, for the most severe health effects. 2D-RTR thus relates to the probability of response for these health effects.

RESULTS

Characterization and Estimation of the RPP

Considered health effects were classified in C2 to C8 severity categories as described in Figure 1. Classification was based on developed guidance in Supplementary Table 2, and is regarded to be in agreement, at a general level, with the dioxin dose-response of key events discussed in Simon *et al.* (2009) and AOP discussed in Becker *et al.* (2015) (see Supplementary Table 4). *BMD*s estimated for all health effects are given in Supplementary Table 5 together with criteria for inclusion/exclusion of datasets for RPP modeling.

Parameter estimates for established RPPs can be found in Supplementary Table 6. Statistical analysis based on RPP_{21} indicate that the median *H* parameter, for standard model 1a, with respect to the mixtures (PCB_{126}: PCB_{118}, and TCDD: PeCDF: PCB_{126}) is similar to that for TCDD. This is expected since mixture doses are expressed in TCDD equivalents (see Supplementary Table 7). The median RPP_{21} (model 1a) shape parameter is quite similar across chemicals/mixtures ($\lambda = 1.2$–1.6) except for PCB_{118} ($\lambda = 2.5$) (Supplementary Table 6). Statistical analysis does not reject the possibility that some other RPPs may also differ in shape/slope versus TCDD, but suggests more clearly that this may be the case for PCB_{118} (see Supplementary Table 7).

Based on this result, ie, mainly resorting to the toxic equivalency factor concept, RPP_{21} (model 1a) for TCDD, PeCDF, and PCB_{126} is illustrated in Figure 3A using a common λ (and σ) across datasets: the relative potency for PeCDF and PCB_{126} versus TCDD is 4.7 and 9.4, respectively, based on median H-ratios for model 1a (Supplementary Table 6). RPP_{21} (model 1a) for PCB_{118} is shown separately in Figure 3B as an example of a higher RPP shape (λ) parameter, whereas an overall RPP_{21} for dioxin-like chemicals, dioxin RPP_{21} (model 1a), is shown in Figure 3C. Dioxin RPP_{21} is based on data for the 3 single chemicals (see Supplementary Figure 2 for separate analysis) and the 2 mixtures (see Supplementary Figure 3 for separate analysis), where *BMD*s for PeCDF and PCB_{126} have been adjusted by the relative potency versus TCDD. RPPs corresponding to BMRs of 0.1 and 0.4, ie, RPP_{10} and RPP_{40}, for dioxin-like chemicals and PCB_{118} are shown in Supplementary Figures 4 and 5, respectively.

The algorithm used for RPP estimation is illustrated in Figure 4 using dioxin RPP_{21} (model 1a) as an example. This RPP is based on information from 83 dose-response curves, and the extensive amount of data results in narrow confidence bounds as shown in Figure 3C. Overall, the ratio between the upper and lower bound of H (for model 1a) is 1.2–1.8 (see Supplementary Table 6). This corresponds to the lower end of the uncertainty within the group of *BMD*s used to establish each RPP, which sometimes was quite large (see Supplementary Table 6: the ratio between the upper and lower bound of the *BMD*, U-BMD, is between 1.3 and 65).

RTD Estimation Using Linear Severity-Weight

*RTD*s, based on RPP_{21} model 1a, corresponding to RTRs in the range of 0.001–0.25 are presented in Table 1. Results across all 3 RPP_{21} models can be found in Supplementary Table 8. *RTD* point estimates are quite consistent across the dioxin-like chemicals/mixtures (TCDD, PeCDF, PCB_{126}, PCB_{126}: PCB_{118}, and TCDD: PeCDF: PCB_{126}) when doses are expressed in TCDD

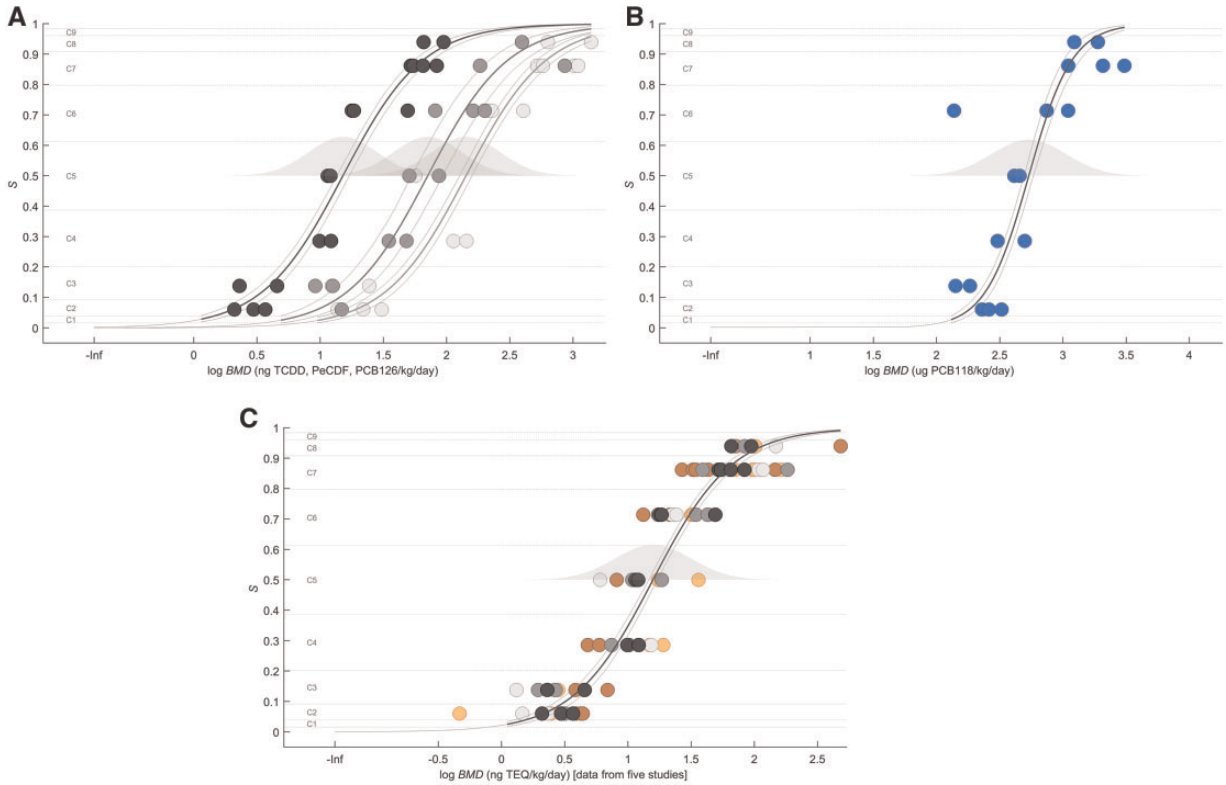

Figure 3. RPP_{21} model 1a for dioxin-like chemicals and mixtures. Circles correspond to BMD_{21} point estimates (x-axis) for liver effects classified in severity categories C2–C8 that are linked to quantitative values of *S* (y-axis) (as an example circles are plotted at a central location in the *S*-intervals associated with C2–C8). Solid curves describe central RPP point estimates with 2-sided 90% CIs, and gray distributions describe the median RPP_{21} variability, across 1000 iterations. (A) RPP_{21} with chemical-specific location, *H*, and common shape, λ (and standard deviation, σ), across chemicals: TCDD (dark), PeCDF (gray), PCB_{126} (light). (B) RPP for PCB_{118}. (C) Dioxin RPP_{21} with common model parameters (*H*, λ, and σ) across 5 chemicals/mixtures: TCDD (dark), PeCDF (gray), PCB_{126} (light), PCB_{126}: PCB_{118} (orange, dark), and TCDD: PeCDF: PCB_{126} (orange, light).

equivalents (Table 1). Also, for each RTR the *RTD* from the combined analyses of the 5 chemicals/mixtures is within the range of *RTD*s from the separate analyses (Table 1).

Table 1 shows that the difference in the *RTD* between RPP_{21} models 1a–c is less than or equal to a factor 1.4, 1.7, 1.3, and 1.6 at RTRs of 0.25, 0.1, 0.01, and 0.001, respectively, indicating a high stability in the results across RPP models (see also Supplementary Table 8). In Supplementary Figure 6 results for dioxin-like chemicals and PCB_{118} are illustrated in more detail, and also contrasted to a more traditional analysis where doses, *D*, associated with specified *S*-values (in the same range as RTRs expressed in %) are estimated directly from the central RPP_{21} point estimate. A higher stability is observed for the proposed method at low doses/responses versus a more traditional approach for RP or effect estimation (Supplementary Figure 6).

The RTD_{10} (corresponding to RTR = 0.10) based on RPP_{21} model 1a is illustrated in Figures 5A–C for dioxin-like chemicals and PCB_{118}, which have different shapes (median λ = 1.38 vs 2.53) but similar standard deviations (median σ = 0.69 vs 0.67) (see Supplementary Table 6). RTD_{10} intersect with the central RPP_{21} point estimate at different *S*-values: ie, at an *S*-value on border between category C4 and C5 for dioxin-like chemicals (Figure 5A), and at an *S*-value centrally located in C4 for PCB_{118} (Figure 5B). Thus, doses of dioxin-like compounds and PCB_{118} corresponding to median *BMD*s for health effects in the same severity category would not correspond to the same RTR (for dioxin-like chemicals the RTR would be 0.067 rather than 0.1 for a dose [8.2 ng TEQ/kg/day] corresponding to an *S*-value centrally located in C4, and differences between the 2 cases increase at lower *S*-values). This indicates a departure from the traditional definition of equipotent doses, the significance of which depends on how much the RPP shape and standard deviation differs between the chemicals being compared. RTR areas corresponding to RTD_{10} are shown in Figure 5C. As can be noted, the product $p \times w$ differ along the *w*-domain between the 2 examples, mainly because of different RPP shapes, although areas under the $p \times w$ curve, ie, the RTRs, are the same.

*RTD*s for dioxin-like chemicals are compared with *RTD*s for PCB_{118} across RTR levels in Figure 5D. Results across RTRs are quite consistent between the 3 RPPs used as basis for *RTD* derivation: RPP_{21} (based on BMD_{21}), RPP_{10} (based on BMD_{10}) and RPP_{40} (based on BMD_{40}). The small deviations from the mean, described by a linear model in Figure 5D, results due to differences between RPP_{21}, RPP_{10}, and RPP_{40} with respect to RPP shape and standard deviation. For dioxin RPP_{21}, RPP_{10}, and RPP_{40} (model 1a) the median λ is 1.38, 1.33, 1.43, and the median σ is 0.69, 0.87, and 0.80, respectively (Supplementary Table 6). For PCB_{118} RPP_{21}, RPP_{10}, and RPP_{40} (model 1a) the median λ is 2.53, 2.72, 2.36, and the median σ is 0.67, 0.79, and 0.62, respectively (Supplementary Table 6).

Linear Versus Nonlinear Linear Severity-Weight

In Table 1 *RTD*s derived using nonlinear severity weights are also compared with the results associated with a linear weight. For $w(S|1, 1.4)$ and $w(S|1.4, 1)$ *S*-values of 0.58 and 0.78, rather than $S = \sqrt{0.5}$, calibrates to RTR $\approx$ 0.25 (50%) (Figure 2B). For a

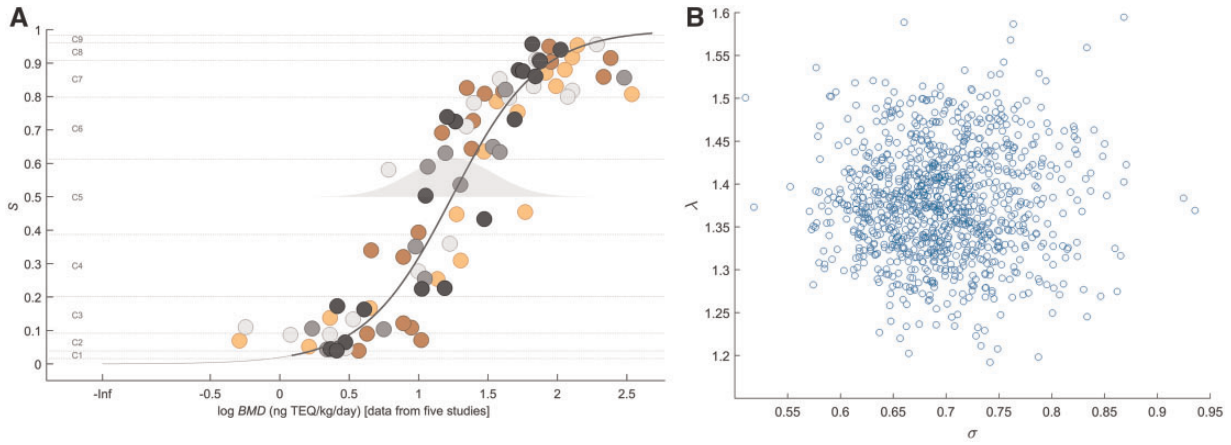

Figure 4. (A) One iteration of the algorithm used for estimating dioxin RPP_{21} model 1a (see Figure 3C) that provides estimates of the RPP shape ($\hat{\lambda} = 1.37$) and standard deviation ($\hat{\sigma} = 0.68$) close to median values (see Supplementary Table 6). Each circle represents a simulated pair of *BMD* and *S*-values. Color codes as in Figure 3C. (B) Correlation between $\hat{\lambda}$ and $\hat{\sigma}$ cross all 1, 000 iterations.

Table 1. *RTD*s Based on RPP_{21} Model 1a

		RTD analysis based on $w(S|1, 1)$					PE-ratio[b]	
Chemical/s	RTR[a]	PE	L05	U95	RPP model uncertainty[c]	Range[d]	$w(S|1, 1.4)$	$w(S|1.4, 1)$
Dioxin-like chemicals[e]	0.25 (50%)	31	28	34	1.3	3.8	1.5	1.3
	0.1 (20%)	11	10	12	1.4	10	1.4	1.4
	0.01 (2%)	2.7	2.3	3.1	1.2	43	1.3	1.6
	0.001 (0.2%)	0.95	0.75	1.2	1.3	122	1.3	1.8
PCB_{118}	0.25 (50%)	781	700	869	1.1	2.0	1.3	1.2
	0.1 (20%)	377	337	418	1.1	4.2	1.2	1.2
	0.01 (2%)	134	110	159	1.1	12	1.2	1.2
	0.001 (0.2%)	66	51	82	1.1	24	1.2	1.3
TCDD	0.25 (50%)	28	25	31	1.3	3.4	—	—
	0.1 (20%)	12	11	14	1.5	7.8	—	—
	0.01 (2%)	3.7	3.0	4.3	1.1	26	—	—
	0.001 (0.2%)	1.5	1.1	1.9	1.4	62	—	—
PeCDF	0.25 (50%)	30	21	45	1.1	4.1	—	—
	0.1 (20%)	10	7.0	13	1.3	12	—	—
	0.01 (2%)	2.1	1.1	3.4	1.0	59	—	—
	0.001 (0.2%)	0.69	0.27	1.4	1.6	182	—	—
PCB_{126}	0.25 (50%)	31	27	35	1.3	4.2	—	—
	0.1 (20%)	11	9.5	13	1.6	12	—	—
	0.01 (2%)	2.5	2.0	3.1	1.2	51	—	—
	0.001 (0.2%)	0.87	0.63	1.2	1.4	149	—	—
PCB_{126}: PCB_{118}	0.25 (50%)	28	23	32	1.2	3.2	—	—
	0.1 (20%)	10	8.3	12	1.3	8.5	—	—
	0.01 (2%)	2.5	1.6	3.4	1.1	35	—	—
	0.001 (0.2%)	0.92	0.52	1.4	1.3	95	—	—
TCDD: PeCDF: PCB_{126}	0.25 (50%)	40	33	47	1.4	4.7	—	—
	0.1 (20%)	13	10	16	1.7	14	—	—
	0.01 (2%)	2.6	1.6	4.1	1.3	71	—	—
	0.001 (0.2%)	0.83	0.41	1.5	1.3	226	—	—

Note: *RTD*s are given in ng TEQ/kg/day and μg/kg/day (PCB_{118}). Point estimates (PE) and 2-sided 90% confidence bounds (L05 and U95) are presented. Complete results across RPP_{21} models 1a–c are given in Supplementary Table 8.

[a]RTR, risk-based and toxicity-integrated response.

[b]PE-ratio: relative (factor) difference between the *RTD* point estimate associated with using parameters (1.4, 1) or (1, 1.4) in the severity-weight function, and the *RTD* point estimate associated with default parameters (1, 1). Complete results across RPP_{21} models 1a-c are given in Supplementary Table 8.

[c]The ratio between the highest and the lowest *RTD* point estimate across RPP models 1a–c (see complete results in Supplementary Table 8).

[d]Range: ratio between the median *BMD* for category C8 health effects (median *BMD*s for cholangiocarcinoma and hepatoellular adenoma according to the central RPP point estimate) and the *RTD* point estimate.

[e]Data on TCDD, PeCDF, PCB_{126}, PCB_{126}: PCB_{118} (mixture), and TCDD: PeCDF: PCB_{126} (mixture) modeled simultaneously (dioxin RPP_{21}, see Figure 3C).

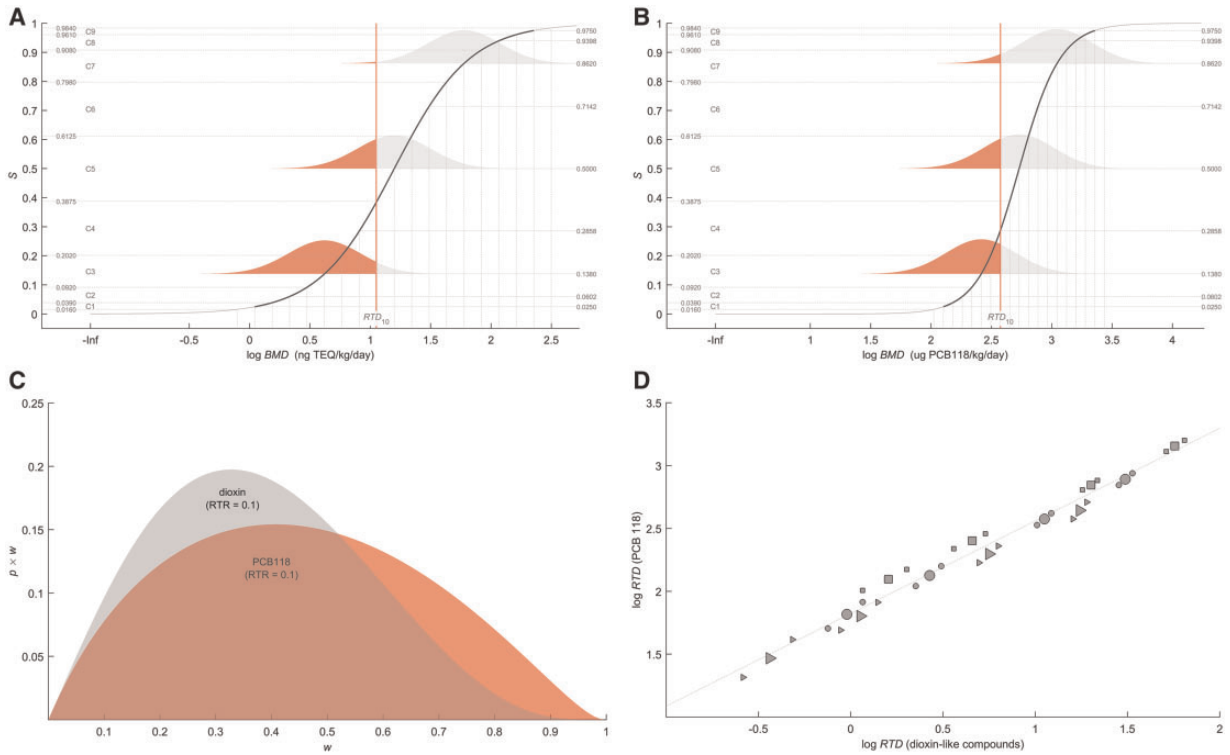

Figure 5. (A) RPP_{21} (model 1a) for dioxin-like chemicals and (B) RPP_{21} (model 1a) for PCB_{118} where vertical lines represent RTD_{10} point estimates. (C) RTRs (area under curve) associated with RTD_{10}. (D) Comparison of *RTD*s for dioxin-like chemicals (units in ng TEQ/kg/day) and *RTD*s for PCB_{118} (units in μg/kg/day). Circles, triangles, and squares represent *RTD*s, corresponding RTRs of 0.25, 0.1, 0.01, and 0.001, derived from RPP_{21}, RPP_{10}, and RPP_{40} (model 1a), respectively. The relative potency is not constant since slopes of the underlying dose-response curves differ, and also due to differences in RPP shapes (Figs. 5A and 5B). Large symbols represent *RTD* point estimates, and small symbols represent *RTD* lower and upper bounds. The coefficient of determination, R2, is 0.98 for a linear model (dotted line) fitted to *RTD* point estimates.

more conservative weight, $w(S|1, 1.4)$, the *RTD* point estimate based on model 1a decreases with a factor 1.3–1.5 for dioxin-like compounds, and a factor 1.2–1.3 for PCB_{118} in the dose region considered (Table 1). For $w(S|1.4, 1)$, which is less conservative than the linear severity-weight, *RTD* point estimates increase with similar factors of 1.3–1.8 and 1.2–1.3 for dioxin-like compounds and PCB_{118}, respectively (Table 1). Complete results across models 1a–c are given in Supplementary Table 8.

Sensitivity Analysis, SA1 and SA2

Results from sensitivity analyses of severity classification based on RPP_{21} model 1a (using linear weight) are summarized in Table 2 with complete results available in Supplementary Table 9. When compared with the proposed classification the *RTD* associated with SA1 and SA2 can systematically increase or decrease by up to a factor between 1.1 and 1.5 (Table 2 and Supplementary Table 9): this specifically illustrates the potential effect of reclassifying 25% of *BMD*s to adjacent severity categories. The ratio between *RTD* point estimates for *BMD* classifications corresponding to the 95th and 5th percentiles in SA1 and SA2 is 1.1–2 depending on dataset and RTR level (Table 2): this indicates a high stability in the *RTD* also with respect to a moderate difference in severity classification. The effect of systematic reclassification of *BMD*s (SA2) is slightly higher than nonsystematic reclassification (SA1), and the effect of reclassification is slightly more pronounced at lower RTRs (Table 2). The classifications used as basis for *RTD* results in SA1 and SA2 associated with RTR = 0.001 are presented in Supplementary Table 10 for each dataset.

Generalized Methodology

In Figure 6 results associated with RPP_{10} and RPP_{40} are compared with results from the generalized methodology for RPP characterization, which (in this case) adjusts the default RPP_{21} to specified BMRs of 0.1 and 0.4 using an indirect/approximate approach (see Supplementary Material). In the BMR region considered the performance of the approximate approach is high with respect to the *RTD* (Figure 6A) as well as the RTR (Figure 6B). Using the approach evaluated in Figures 6A and 6B the RTR associated with an exposure, *E*, can be estimated across the entire BMR domain for derivation of 2D-RTR as illustrated in Figure 6C. Additional results are shown in Table 3 where 2D-RTR is evaluated for a range of exposures. Results become similar when using RPP_{21}, RPP_{10}, or RPP_{40} as the basis for adjusting RPP model 1a across BMR levels using the generalized methodology in the process of estimating the 2D-RTR: point estimates differ only by a factor 1.1–1.3 (Table 3). Thus, at the level of the 2D-RTR results are quite independent of RPP used as a basis in the analysis.

In Figure 6B, RTR point estimates associated with adjusted RPP_{21} (model 1a) differ from RTR point estimates associated with RPP_{10} or RPP_{40} (model 1a) by a factor 1–1.7. It may be noted that the differences reported in Table 3 are even lower: ie, 2D-RTR point estimates based on RPP_{21} and RPP_{10}/RPP_{40} differ by a factor 1–1.2. The former results (Figure 6B) describe comparison at specific BMRs of 0.1 and 0.4, respectively, while the latter analysis (Table 3) extends comparison to the level of the entire BMR domain. The results in Figure 6B and Table 3 thus indicate that the higher-order 2D-RTR at a given dose may be even more robust compared with the RTR.

Table 2. Sensitivity Analyses SA1 and SA2 Using RPP_{21} Model 1 With Linear Severity-Weight

Chemical/s	RTR[a]	PE/PE5[b]		PE95/PE5[c]	
		SA1	SA2	SA1	SA2
Dioxin-like chemicals[d]	0.1	1.1	1.2	1.1	1.4
	0.001	1.1	1.3	1.2	1.5
PCB_{118}	0.1	1.1	1.1	1.2	1.3
	0.001	1.2	1.2	1.3	1.4
TCDD	0.1	1.2	1.2	1.3	1.5
	0.001	1.3	1.4	1.6	1.8
PeCDF	0.1	1.2	1.2	1.4	1.5
	0.001	1.4	1.5	1.8	2.0
PCB_{126}	0.1	1.2	1.3	1.3	1.5
	0.001	1.4	1.4	1.7	1.9
PCB_{126}: PCB_{118}	0.1	1.2	1.2	1.3	1.4
	0.001	1.3	1.4	1.5	1.7
TCDD: PeCDF: PCB_{126}	0.1	1.2	1.3	1.4	1.5
	0.001	1.4	1.4	1.8	2.0

Note: In the sensitivity analyses *BMD* point estimates in combination with central S-values (see Supplementary Table 3) for associated severity categories are initially used for RPP characterization. Starting from the proposed classification (Figure 1) 25% of *BMD*s associated with a given dataset are randomly selected in each of the 1000 iterations of the algorithm for *RTD* estimation. Selected *BMD*s in a given iteration are reclassified individually (SA1) or as a group (SA2), 1 category upward or downward, at random, and the *RTD* (for a specified RTR) is estimated for this classification. *BMD* classifications associated with the 5th and 95th percentile of the *RTD* (P5 and P95) are identified in SA1 and SA2, respectively. Complete *RTD* estimation, utilizing *BMD* uncertainty distributions as well as associated S-intervals, are then performed for these *BMD* classifications. Extended results are available in Supplementary Tables 9 and 10.

[a]RTR: risk-based and toxicity-integrated response.

[b]PE/PE5 is the ratio between *RTD* point estimates for the proposed classification and the classification associated with P5 in SA1 and SA2, respectively.

[c]PE95/PE5 is the ratio between *RTD* point estimates for classifications corresponding to P95 and P5 in SA1 and SA2, respectively. Complete results are given in Supplementary Table 9.

[d]Data on TCDD, PeCDF, PCB_{126}, PCB_{126}: PCB_{118} (mixture), and TCDD: PeCDF: PCB_{126} (mixture) modeled simultaneously (dioxin RPP_{21}, see Figure 3C).

RTD results associated with 2D-RTRs of 0.05 and 0.01 are given in Table 4 with more detailed results in Supplementary Table 8. Extension of the algorithm for *RTD* estimation (for 2D-RTR) is illustrated in Supplementary Figure 7. In concordance with results in Table 3 the *RTD*s for dioxin-like chemicals and PCB_{118} are similar regardless of using RPP_{21}, RPP_{10}, or RPP_{40} as the basis in the generalized methodology. Model uncertainty in terms of differences between RPP models 1a-c is also low, and in the same range as observed for *RTD*s associated with (1D) RTRs in Table 1. In addition, Supplementary Figure 7 indicates that uncertainty intervals associated with 2D-RTR are narrower than for RTR.

DISCUSSION

SST is characterized by the sequence of dose-response curves for lower- to higher-order toxicological health effects, here described by the RPP (Figure 1). This concept extends the current risk assessment framework that uses a specific RP for establishment of health-based guidance values: the shape (λ) and standard deviation (σ) of the RPP add new dimensions to the characterization of chemical toxicity and risk. The present data indicated that estimation of dose equivalents, *RTD*s, corresponding to low values of the RTR, can be performed without observing large differences between RPP models (Tables 1 and 4 and Supplementary Table 8). Additional analyses showed that the method is robust to minor as well as moderate changes in severity classification of *BMD*s (Table 2 and Supplementary Tables 9 and 10). Also, while the severity-weight will/should influence the results the approach is not highly sensitive to this parameter (Table 1 and Supplementary Table 8).

Since the RTR increases from 0 to 0.5 across the RPP, the proposed concept suggests that the level of protection, or risk, embedded in individual RPs used for traditional risk assessment (that theoretically may be located in any severity category along the RPP) can potentially differ to a high extent. As noted in the introduction, the question whether or not extra safety measures are needed depending on the severity of the critical effect was addressed by the Swedish National Food Agency that have applied a severity-adjusted margin of exposure approach for risk assessment at the national level (NFA, 2017). That approach performs value-based adjustment of the health-based guidance value depending on the severity of the critical effect so it schematically/theoretically represents a value in the lower end of the RPP.

Characterization of the RPP is a further advancement, compared with using adjustment factors, that provides a mean for describing the dose region below the *BMD* for a severe effect like cancer. Estimation of doses/*BMD*s associated with very low response levels, eg, corresponding to very low cancer incidences, is very uncertain. Instead, *BMD*s in the observable region of response for lower-order health effects are here effectively used for characterizing the same (low) dose region. This can improve low-dose assessment considering that the upper end of the RPP, in this case study represented by a *BMD* for cancer corresponding to C8 or C9, would represent a starting point for linear extrapolation or assessment factor application. For example, when the method is applied for chemicals that are both genotoxic and carcinogenic the extra safety factor of 100 recommended by EFSA (2005) that accounts for uncertainties related to the process of carcinogenesis might be reduced. For the present data $RTD_{0.001}$ is about a factor 20–200 below the (median) *BMD* for cancer, respectively (Table 1). Observe that the range of *RTD*s selected in this study were mainly used for assessing low-dose stability, and it needs to be further discussed what *RTD*s would be appropriate for practical risk assessment.

An apparent challenge with the proposed concept concern severity-weighting of health effects on a quantitative scale. Within the current framework debate regarding the definition of the RP for establishment of human exposure guidelines (eg, EFSA, 2009; Murrell *et al.*, 1998; Sand *et al.*, 2006; Slob and Pieters, 1998) is to some extent related to this issue. The *BMD* has been suggested to be expressed as corresponding to a nonadverse change in response for the critical health effect (EFSA, 2009), and this change in response may differ across effects. Under this definition the severity of the health effect would be embedded in the selected response level associated with the *BMD*. Van der Voet *et al.* (2009) suggested a generalization of this concept in such a way that a set of response levels corresponding to "low", "moderate", or "severe" health impact criterions may be defined for a given effect. At the conceptual level the proposed idea addresses this type of problem from a different angle by considering severity at the level of the overall health effect, and using a range of health effects (representing different severities) for which standardized *BMD*s (in the observable region of response) are derived.

The concept of acceptable/nonadverse response levels for individual health effects, or similar, is complicated when considering health effects in a joint context. As described in the

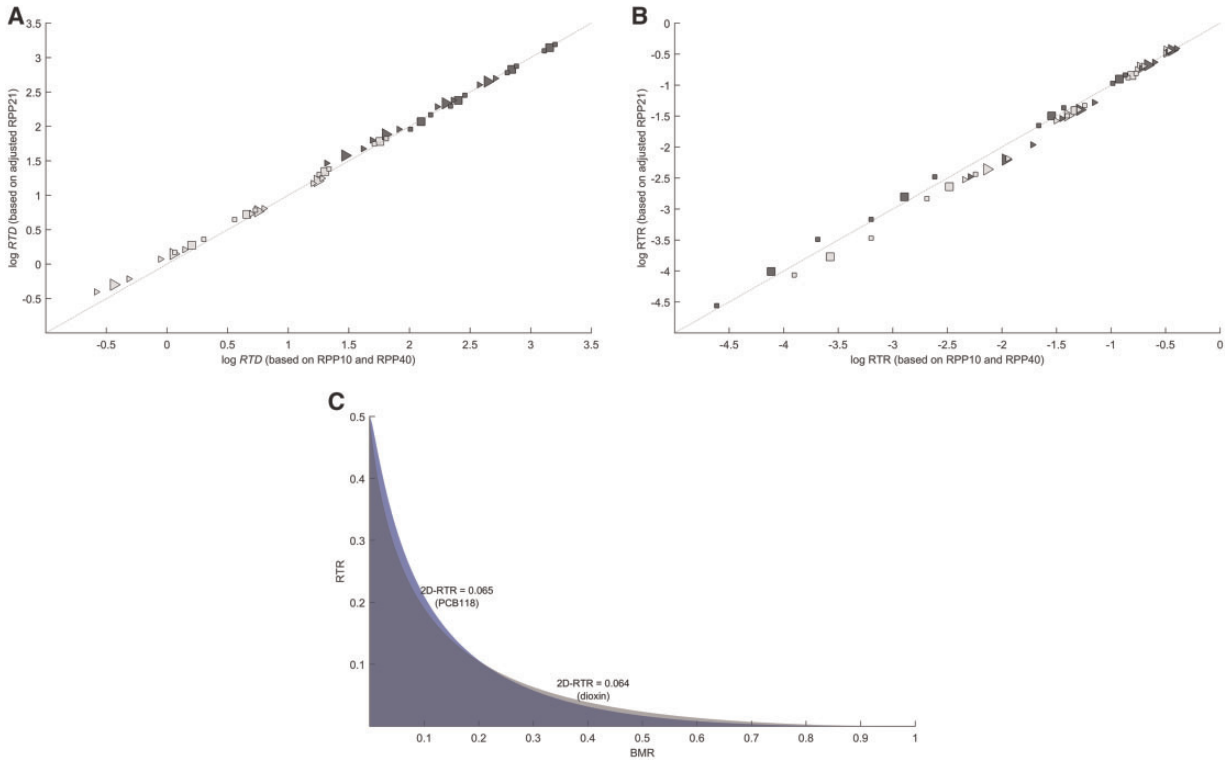

Figure 6. Results associated with the generalized methodology for RPP characterization that adjusts the estimated RPP_{21} (model 1a) to describe RPPs associated with specified BMR levels using an indirect/approximate approach (see Supplementary Material). (A) Correlation between *RTD*s based on RPP_{10}/RPP_{40} (model 1a) and corresponding *RTD*s from the approximate approach. Triangles and squares are results associated with BMRs of 0.1 and 0.4, respectively. Light symbols represent dioxin-like chemicals, and dark symbols represent PCB_{118}. Large symbols are point estimates, and small symbols are lower and upper bounds. *RTD*s correspond to RTRs of 0.25, 0.1, 0.01, and 0.001. (B) Correlation between RTRs based on RPP_{10}/RPP_{40} and corresponding RTRs from the approximate approach. Symbol representation as in (A). RTRs correspond to exposures (*RTD* point estimates in Table 1) of 31, 11, 2.7, and 0.95 ng TEQ/kg/day for dioxin-like chemicals, and 781, 377, 134, and 66 μg/kg/day for PCB_{118}. (C) RTRs estimated across BMR = 0–1 according to the generalized methodology for RPP characterization. RTRs are associated with exposures, $E = 11$ ng TEQ/kg/day for dioxin-like chemicals and $E = 377$ μg/kg/day for PCB_{118} (RTD_{10} in Table 1). The RTR integrated across BMR levels, 2D-RTR, is 0.064 (dioxin-like chemicals) and 0.065 (PCB_{118}), respectively. This is regarded to describe impacts similar to that associated with an (average) probability of response of about 6% for the most severe (C8–C9) health effects.

results, in relation to Figures 5A and 5B, the traditional view of equipotent doses can be challenged by the proposed method since such doses may not correspond to the same RTR. Conversely, the response level regarded as nonadverse for a particular (lower-order) health effect will correspond to a dose associated with an RTR that depends on how this effect is aligned (in terms of dose) with respect to other effects produced by the same chemical, reflected by shape and standard deviation of the RPP that may differ across chemicals (Figs. 5A and 5B). The consideration of health effect in an integrated context can therefore provide a different conclusion. As described in Figure 2 severity-weighting schemes that differ systematically (symmetric, left skew, or right skew) will also differ with respect to the *S*-value that calibrates to RTR ≈ 50%. Since this *S*-value is approximately independent of RPP model parameters severity-weighting is suggested to be addressed by specifying the center of the studied toxicological process (or dose-related sequence of health effects) in terms of RTR.

As noted previously the RPP can be regarded to represent a cross-section of the dose-severity-response volume. The present analysis mainly considers the *x*-*y* cross-section at a response level (*z*) of 21%: ie, the BMD_{21} profile. Based on BMD_{21} the definition of a transitional dose value was developed by Simon *et al.* (2014), and later discussed as an approximation of a threshold in the context of the AOP for dioxin-like chemicals (Becker *et al.* 2015). RPPs based on other BMR levels were also assessed in the present analyzes: ie, RPP_{10} and RPP_{40} associated with BMR = 0.1 and 0.4, respectively. Results showed that the RTR-based relative potency (for a range of RTRs) was similar across RPP_{21}, RPP_{10}, and RPP_{40} (Figure 5D), and also that *RTD* and RTR results associated with RPP_{10}, and RPP_{40} could be well approximated by shifting the estimated RPP_{21} along the dose scale using a generalized approach for RPP characterization (Figs. 6A and 6B).

It was also illustrated how the RTR metric can be integrated across BMR levels for derivation of the 2D-RTR that accounts for both the severity and response domains (Figure 6C). This enables an interpretation of the RTR that indirectly relates to the probability (of response) for the most severe health effects (equation 4), rather than the probability of exceeding a particular *BMD* for these effects (equation 3). In principle, 2D-RTR multiplied with the number exposed individuals provides a surrogate for the number of cases with respect to the higher-order (C8/C9) health effects, which may also provide a link for calculation of the burden of disease, or similar, and/or associated RPs. Further analysis is required to assess the potentials of 2D-RTR. Although this development adds complexity results indicated at the 2D-RTR may be even more robust than the RTR, and results were quite independent on the BMR level associated with the RPP used as basis in the generalized approach (Table 3 and Figure 6B). Using 2D-RTR may enhance the proposed method also for this reason besides providing improved risk interpretation.

Table 3. 2D-RTR Based on RPP Model 1a With Linear Severity-Weight

Chemical/s	E^a	$RPP_X{}^b$	2D-RTR			RPP model uncertainty[c]		
			PE	LU	LB	21 vs 10	21 vs 40	high vs low
Dioxin-like chemicals[d]	31	21	0.14	0.14	0.15	1.0	1.1	1.1
		10	0.14	0.13	0.15			
		40	0.15	0.14	0.16			
	11	21	0.064	0.059	0.069	1.0	1.1	1.1
		10	0.067	0.060	0.074			
		40	0.070	0.065	0.076			
	2.7	21	0.014	0.013	0.017	1.1	1.1	1.1
		10	0.016	0.014	0.020			
		40	0.016	0.015	0.019			
	0.95	21	0.0041	0.0036	0.0048	1.2	1.2	1.2
		10	0.0049	0.0040	0.0061			
		40	0.0048	0.0042	0.0059			
PCB_{118}	781	21	0.13	0.12	0.15	1.0	1.0	1.1
		10	0.14	0.13	0.16			
		40	0.13	0.12	0.14			
	377	21	0.065	0.058	0.073	1.1	1.0	1.1
		10	0.071	0.061	0.083			
		40	0.062	0.056	0.069			
	134	21	0.018	0.015	0.021	1.2	1.1	1.2
		10	0.020	0.016	0.026			
		40	0.017	0.015	0.019			
	66	21	0.0063	0.0053	0.0076	1.2	1.1	1.3
		10	0.0075	0.0058	0.0099			
		40	0.0059	0.0052	0.0068			

Note: The RTR integrated across BMR = 0–1, 2D-RTR, is given as point estimates (PE) and 2-sided 90% confidence bounds (L05 and U95). RPPs associated with BMR levels spanning BMR = 0–1 have been derived using a generalized methodology for RPP characterization that adjust RPP_{21} (based on BMD_{21}), RPP_{10} (based on BMD_{10}), or RPP_{40} (based on BMD_{40}) to specified BMRs (see Supplementary Material). 2D-RTRs associated with *E* are then estimated according to equation 4).

[a]Exposure corresponding to *RTDs* in Table 1.

[b]BMR (%) associated with RPP used as basis in the generalized approach for RPP characterization according to Supplementary equation 4.

[c]Relative differences between 2D-RTR point estimates (PE-ratios) associated with exposure, *E*.

[d]Data on TCDD, PeCDF, PCB_{126}, PCB_{126}: PCB_{118} (mixture), and TCDD: PeCDF: PCB_{126} (mixture) modeled simultaneously (dioxin RPP_{21}, see Figure 3C).

Both in its more simplified (RTR) and extended (2D-RTR) form the proposed method provides a way of deriving starting points for establishment of health-based guidance values that account for health effects in a comprehensive manner. Also, the standardized nature of the approach may enable comparative risk characterization across different chemical exposure scenarios, which for example could facilitate risk-based prioritization. Suggestions regarding dose-response modeling approaches and estimation of (standardized) RPs across different health effects and data types (Murrell *et al.*, 1998; Sand *et al.*, 2011, 2017; Shockley 2015; Slob, 2017; Wignall *et al.*, 2014) may help to refine the method at the technical level.

The *BMD* analysis of health effects may for example be optimized by differentiating the BMR from a statistical perspective since a single BMR may not fit effects that occur at low and high doses equally well for a given experimental design. The RPP associated with a given BMR, or the complete dose-severity-response volume, may then be estimated using the most optimal/certain *BMDs* for each (group of) health effects, and adjusting for the associated BMR using the generalized approach introduced herein. Although different RPP models (1a–c) and severity weights were applied to assess the stability of the method future versions of the proposed algorithm for RTR/*RTD* estimation can be extended to also cover such uncertainties besides the uncertainty in the *BMD* and *S*. A step-wise approach was applied for estimation of the RPP/dose-severity-response volume, and it may be discussed if this can be done more directly: the former was here regarded more practical since relevant dose-response data may not be of the same type/format, which complicates a direct combination of data.

As noted in the introduction SST is more descriptive compared with pathway or mode of action concepts. All responses in molecules and metabolic pathways, as a result increasing dosage, may not be directly associated with the pathophysiological development of a specific tissue lesion, but occur more as a general response reflecting a variety of changes and interactions in the tissue homeostasis. However, these general and broad effects in the target tissue could also affect the pathway-specific toxicity in terms of severity. The SST concept will therefore account for toxicity that is not only limited to a specific pathway, but the total influence of a chemical on tissue homeostasis, and this may more accurately reflect the overall toxicity and risk. Although this may be advantageous a challenge is how the selection of health effects for RPP characterization should be limited. As a starting point organ-specific data was used. Even under this constraint, however, some health effects may be more or less relevant to be considered as part of the studied RPP.

Two health effects were excluded in the present analysis based on their deviation from the RPP (see Supplementary Figure 8). In total 17 dose-response datasets were excluded, mainly due to nonsignificant dose-response trends (see Supplementary Table 5A). If the analyses would have been limited to only consider health effects which were recorded across all NTP studies this criteria alone would exclude 15 of these 17 datasets (including those deviating in Supplementary Figure 8).

Table 4. *RTD* Associated With 2D-RTR Based on Model 1a With Linear Severity-Weight

Chemical/s	2D-RTR[a]	RPP_X[b]	*RTD*			RPP model uncertainty	
			PE	L05	L95	models 1a–c[c]	$RPP_{21,\ 10,\ 40}$[d]
Dioxin-like chemicals[e]	0.05	21	8.6	7.8	9.4	1.3	1.1
		10	8.1	7.1	9.2	—	
		40	7.8	6.9	8.4	—	
	0.01	21	2.0	1.7	2.2	1.2	1.1
		10	1.7	1.4	2.0	—	
		40	1.7	1.5	1.9	—	
PCB_{118}	0.05	21	301	269	333	1.1	1.1
		10	275	233	318	—	
		40	312	286	340	—	
	0.01	21	90	79	101	1.1	1.2
		10	80	66	95	—	
		40	94	85	103	—	
TCDD	0.05	21	9.8	8.5	11	1.3	—
	0.01		2.9	2.5	3.3	1.2	—
PeCDF	0.05	21	7.4	5.1	9.5	1.2	—
	0.01		1.3	0.82	1.8	1.1	—
PCB_{126}	0.05	21	8.6	7.4	9.9	1.4	—
	0.01		2.2	1.8	2.6	1.3	—
PCB_{126}: PCB_{118}	0.05	21	7.4	6.0	8.7	1.2	—
	0.01		1.4	1.1	1.6	1.1	—
TCDD: PeCDF: PCB_{126}	0.05	21	9.9	7.5	12	1.5	—
	0.01		2.2	1.5	2.9	1.4	—

Note: *RTD*s are given in ng TEQ/kg/day and μg/kg/day (PCB_{118}). Point estimates (PE) and 2-sided 90% confidence bounds (L05 and U95) are presented. Complete results across RPP_{21} models 1a-c are given in Supplementary Table 8. RPPs associated with BMR levels spanning BMR = 0–1 have been derived using a generalized methodology for RPP characterization that adjust RPP_{21} (based on BMD_{21}), RPP_{10} (based on BMD_{10}), or RPP_{40} (based on BMD_{40}) to specified BMRs (see Supplementary Material). For a number of doses within a given interval the 2D-RTR is estimated according to equation 4, and the *RTD* is solved by spline interpolation at the specified 2D-RTR. This extended algorithm for *RTD* estimation is illustrated in Supplementary Figure 7.

[a]2D-RTR, risk-based and toxicity-integrated response integrated across BMR = 0–1.

[b]BMR (%) associated with RPP used as basis in the generalized approach for RPP characterization according to Supplementary equation 4.

[c]The ratio between the highest and the lowest *RTD* point estimate across RPP models 1a–c (see complete results in Supplementary Table 8).

[d]Ratio between the highest and lowest *RTD* point estimates using RPP_{21}, RPP_{10}, or RPP_{40} model 1a as the basis for *RTD* derivation using the generalized methodology.

[e]Data on TCDD, PeCDF, PCB_{126}, PCB_{126}: PCB_{118} (mixture), and TCDD: PeCDF: PCB_{126} (mixture) modeled simultaneously (dioxin RPP_{21}, see Figure 3C).

The remaining 2 datasets, "A4H" (with a highly uncertain BMD) and "bile duct cyst/fibrosis" (with a nonsignificant trend) in TR-525 should probably be excluded in any case (see Supplementary Table 5A). In addition, "PROD", "eosinophilic focus", "portal fibrosis", and "centrolobular fibrosis" would have been excluded if health effects recorded across all NTP studies were considered only. However, the *BMD*s for these 4 health effects are quite consistent with those in the same severity category (see Supplementary Table 5A) suggesting that their inclusion/exclusion will only have a minor effect on the estimated RPP.

The proposed method has advantages and some remaining challenges as discussed in the 2 paragraphs above. Ongoing developments aim for example to address the issue of health effect selection more generally. Future improvements on the knowledge of AOPs, or similar, may suggest it to be more relevant to perform categorization based on key events rather than the severity of effect (see Supplementary Table 4 that compares the current severity classification and the AOP suggest for dioxin-like chemicals). Besides applications to traditional toxicity data it may also be investigated if the proposed concept can support risk assessment based on *in vitro* data, in line with the National Research Council's (NCR) long-range strategic plan to modernize toxicity testing (NCR, 2007; Andersen and Krewski, 2010). Although it may not be feasible to assign severity scores to such data categorization related to knowledge on toxicity pathways may be considered instead. It has been hypothesized that the degree of perturbation and adversity would be measured by a composite probability incorporating the contributions from the sequential stages of increasing pathway activation obtained by dose-response characterization of the pathway assays (Bhattacharya *et al.*, 2011). Conceptually, the RTR (eg, Figs. 5C and 6C) may be able to incorporate and synthesize that kind of information.

Further evaluation of the practical applicability of the RPP/SST concept for additional compounds is warranted: eg, while the present data describe quite clear trends between dose and severity this may not always be the case (for less clear trends the RPP becomes steep). Intuitively, the proposed approach is applicable for data-rich compounds. However, studies have discussed that data in the growth/decay region are most critical for characterizing the shape of the sigmoidal curve (eg, Kuljus *et al.*, 2006). It may thus be further investigated if data for the most important severity categories in this context are sufficient for characterizing the RPP, making the approach less data intensive. Also, since the RPP is constrained so that $S = 0$ and $S = 1$ is the minimum and maximum values, respectively, it may also be tested if data only for the lower severity categories (eg, below C6) in fact are sufficient for RPP estimation.

In conclusion, the SST/RPP concept involves systematic consideration of chemically induced health effects by characterizing their dose-related severity sequence. The novel method described, including a design that reduces the complexity of severity-weighting, enables the components of risk to be

integrated across multiple outcomes. As a result an alternative way of describing the (low) dose region below the *BMD* for a severe effect is achieved, and RPs (or points of departures/exposure guidelines) derived by the proposed method are based on the joint consideration of several health effects, in contrast to current approaches. According to the present data the estimation of such RPs was stable across models, and results were robust for alternative severity categorizations of BMDs, at low values of the RTR. In its generalized form the new impact metric, RTR, relates to the probability of response for the most severe health effects.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

FUNDING

This work was conducted as part of activities related to development of risk assessment methodology at the Swedish National Food Agency supported by internal funds.

REFERENCES

Andersen, M. E., and Krewski, D. (2010). The vision of toxicity testing in the 21st century: Moving from discussion to action. *Toxicol. Sci.* **117**, 17–24.

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., *et al.* (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **29**, 730–741.

Becker, R. A., Patlewicz, G., Simon, T. W., Rowlands, J. C., and Budinsky, R. A. (2015). The adverse outcome pathway for rodent liver tumor promotion by sustained activation of the aryl hydrocarbon receptor. *Reg. Toxicol. Pharmacol.* **73**, 172–190.

Bhattacharya, S., Zhang, Q., Carmichael, P., Boekelheide, K., and Andersen, M. (2011). Toxicity testing in the 21st century: Defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *PLoS One* **6**, e20887.

Bogdanffy, M. S., Daston, G., Faustman, E. M., Kimmel, C. A., Kimmel, G. L., Seed, J., and Vu, V. (2001). Harmonization of cancer and noncancer risk assessment: Proceedings of a consensus-building workshop. *Toxicol. Sci.* **61**, 18–31.

Crump, K. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4**, 854–871.

Dourson, M. L., Hertzberg, R. C., Hartung, R., and Blackburn, K. (1985). Novel approaches for the estimation of acceptable daily intake. *Toxicol. Ind. Health* **1**, 23–41.

Dourson, M. L., Teuschler, L. K., Durkin, P. R., and Stiteler, W. M. (1997). Categorical regression of toxicity data: A case study using aldicarb. *Regul. Toxicol. Pharmacol.* **25**, 121–129.

ECHA (European Chemicals Agency). (2012). Guidance on information requirements and chemical safety assessment. Chapter R.8: Characterisation of dose[concentration]-response for human health. ECHA-2010-G-19-EN.

EFSA (European Food Safety Authority). (2005). Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. *Efsa J.* **3**, 282–231.

EFSA (European Food Safety Authority). (2009). Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. *EFSA J.* **7**, 172.

EFSA (European Food Safety Authority). (2012). Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. *EFSA J.* **10**, 2579.

EFSA (European Food Safety Authority). (2017). Update: Guidance on the use of the benchmark dose approach in risk assessment. *EFSA J.* **15**, 4658–4641.

EPA (U.S. Environmental Protection Agency). (2002). A review of the reference dose and reference concentration processes. Risk Assessment Forum. EPA/630/P-02/002F. Washington, DC.

EPA (U.S. Environmental Protection Agency). (2005). Guidelines for carcinogen risk assessment. Final report. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

EPA (U.S. Environmental Protection Agency). (2012). Benchmark dose technical guidance. Risk Assessment Forum. EPA/100/R-12/001. Washington, DC.

EPA (U.S. Environmental Protection Agency). (2018). Integrated risk information system. Data retrieved on March 24, 2018. Available at https://www.epa.gov/iris. Accessed date March 24, 2018.

FAO/WHO (Food and Agriculture Organization of the United Nations, World Health Organization). (2008). Codex Alimentarius Commission. Procedural Manual, 18th ed. Joint FAO/WHO Food Standards Programme, Rome. 18, pp. 73.

Guth, D. J., Carroll, R. J., Simpson, D. G., and Zhou, H. (1997). Categorical regression analysis of acute exposure to tetrachloroethylene. *Risk Anal.* **17**, 321–332.

Hertzberg, R. C., and Miller, M. (1985). A statistical model for species extrapolation using categorical response data. *Tox. Ind. Health* **1**, 43–57.

Hertzberg, R. C. (1989). Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Phys* **57**, 405–409.

Hertzberg, R. C., and Dourson, M. L. (1993). Using categorical regression instead of a NOAEL to characterize a toxicologist's judgment in noncancer risk assessment. In *Proceedings, Second International Symposium on Uncertainty Modeling and Analysis*, College Park, MD, (B. M. Ayyub, Ed.) IEEE Computer Society Press, Los Alamitos, CA, pp. 254–261.

Kuljus, K., von Rosen, D., Sand, S., and Victorin, K. (2006). Comparing experimental designs for benchmark dose calculations for continuous endpoints. *Risk Anal.* **26**, 1031–1043.

Lehman, A. J., and Fitzhugh, O. G. (1954). 100-fold margin of safety. *Assoc. Food Drug Off. USQ Bull.* **18**, 33–35.

Mattison, D. R., Milton, B., Krewski, D., Levy, L., Dorman, D. C., Aggett, P. J., Roels, H. A., Andersen, M. E., Karyakina, N. A., Shilnikova, N., *et al.* (2017). Severity scoring of manganese health effects for categorical regression. *Neurotoxicology* **58**, 203–216.

Meek, M. E., Palermo, C. M., Bachman, A. N., North, C. M., and Lewis, R. J. (2014). Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *J. Appl. Toxicol.* **34**, 595–606.

Milton, B., Krewski, D., Mattison, D. R., Karyakina, N. A., Ramoju, S., Shilnikova, N., Birkett, N., Farrell, P. J., and McGough, D. (2017). Modeling U-shaped dose-response curves for manganese using categorical regression. *Neurotoxicol.* **58**, 217–225.

Murrell, J. A., Portier, C. J., and Morris, R. W. (1998). Characterizing dose-response I: Critical assessment of the benchmark dose concept. *Risk Anal.* **18**, 13–26.

NCR (National Research Council). (2007). Toxicity testing in the 21st century: A vision and a Strategy. National Academy of Sciences, Washington, DC.

NFA (Swedish National Food Agency). (2015). The risk thermometer - a tool for risk comparison. *National Food Agency Report No. 8*.

NFA (Swedish National Food Agency). (2017). Swedish Market Basket Survey 2015: Per capita-based analysis of nutrients and toxic compounds in market baskets and assessment of benefit or risk. *National Food Agency Report 26*.

NTP (National Toxicology Program). (2006a). NTP technical report on the toxicology and carcinogenesis studies of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) (CAS No 1746-01-6) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **521**, 4–232.

NTP (National Toxicology Program). (2006b). NTP technical report on the toxicology and carcinogenesis studies of 2, 3, 4, 7, 8-pentachlorodibenzofuran (PeCDF) (CAS No. 57117-31-4) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **525**, 1–198.

NTP (National Toxicology Program). (2006c). NTP technical report on the toxicology and carcinogenesis studies of 3, 3′, 4, 4′, 5-pentachlorobiphenyl (PCB_{126}) (CAS No. 57465-28-8) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **520**, 4–246.

NTP (National Toxicology Program). (2006d). NTP technical report on the toxicology and carcinogenesis studies of a binary mixture of 3, 3′, 4, 4′, 5-pentachlorobiphenyl (PCB_{126}) (CAS No. 57465-28-8) and 2, 3′, 4, 4′, 5-pentachlorobiphenyl (PCB_{118}) (CAS No. 31508-00-6) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **531**, 1–218.

NTP (National Toxicology Program). (2006e). NTP technical report on the toxicology and carcinogenesis studies of a mixture of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) (CAS No 1746-01-6) 2, 3, 4, 7, 8-pentachlorodibenzofuran (PeCDF) (CAS No. 57117-31-4) and 3, 3′, 4, 4′, 5- pentachlorobiphenyl (PCB_{126}) (CAS No. 57465-28-8) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **526**, 1–180.

NTP (National Toxicology Program). (2010). NTP technical report on the toxicology and carcinogenesis studies of 2, 3′, 4, 4′, 5-pentachlorobiphenyl (PCB_{118}) (CAS No. 31508-00-6) in female Harlan Sprague-Dawley rats (Gavage Studies). Natl. Toxicol. Program Tech. Rep. Ser. **559**, 1–174.

OECD (Organization for economic co-operation and development). (2013). Guidance document on developing and assessing adverse outcome pathways. Series on Testing and Assessment, No. 184.

Renwick, A. G., Flynn, A., Fletcher, R. J., Muller, D. J. G., Tuijtelaars, S., and Verhagen, H. (2004). Risk-benefit analysis of micronutrients. *Food Chem. Toxicol.* **42**, 1903–1922.

Sand, S., von Rosen, D., and Filipsson, A. (2003). Benchmark calculations in risk assessment using continuous dose-response information: The influence of variance and the determination of a cut-off value. *Risk. Anal.* **23**, 1059–1068.

Sand, S., von Rosen, D., Victorin, K., and Falk Filipsson, A. (2006). Identification of a critical dose level for risk assessment: Developments in benchmark dose analysis of continuous endpoints. *Toxicol. Sci.* **90**, 241–251.

Sand, S., Portier, C. J., and Krewski, D. (2011). A signal-to-noise crossover dose as the point of departure for health risk assessment. *Environ. Health Perspect.* **119**, 1766–1774.

Sand, S., Ringblom, J., Håkansson, H., and Öberg, M. (2012). The point of transition on the dose-effect curve as a reference point in the evaluation of in vitro toxicity data. *J. Appl. Toxicol.* **32**, 843–849.

Sand, S., Parham, F., Portier, C. J., Tice, R. R., and Krewski, D. (2017). Comparison of points of departure for health risk assessment based on high-throughput screening data. *Environ. Health Perspect.* **125**, 623–633.

Shockley, K. R. (2015). Quantitative high-throughput screening data analysis: Challenges and recent advances. *Drug Discov. Today* **20**, 296–300.

Simon, T., Aylward, L. L., Kirman, C. R., Rowlands, J. C., and 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.

Simon, T. W., Simons, S. S., Jr, Preston, R. J., Boobis, A. R., Cohen, S. M., Doerrer, N. G., Fenner-Crisp, P. A., McMullin, T. S., McQueen, C. A., and Rowlands, J. C. (2014). The use of mode of action information in risk assessment: Quantitative key events/dose-response framework for modeling the dose-response for key events. *Crit. Rev. Toxicol.* **44**, 17–43.

Slob, W., and Pieters, M. N. (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: General framework. *Risk Anal.* **18**, 787–798.

Slob, W. (2017). A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. *Crit. Rev. Toxicol.* **47**, 342–351.

Teuschler, L. K., Dourson, M. L., Stiteler, W. M., McClure, P., and Tully, H. (1999). Health risk above the reference dose for multiple chemicals. *Regul. Toxicol. Pharmacol.* **30**, S19–S26.

Van der Voet, H., van der Heijden, G. W. A. M., Bos, P. M. J., Bosgra, S., Boon, P. E., Muri, S. D., and Bruschweiler, B. J. (2009). A model for probabilistic health impact assessment of exposure to food chemicals. *Food Chem. Toxicol.* **47**, 2926–2940.

WHO/IPCS (World Health Organization, International Programme on Chemical Safety). (1994). *Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-Based Exposure Limits. Environmental Health Criteria 170*, World Health Organization, Geneva.

WHO/IPCS (World Health Organization, International Programme on Chemical Safety). (2009). *Principles and Methods for the Risk Assessment of Chemicals in Food*. Environmental Health Criteria 240, World Health Organization, Geneva.

Wignall, J. A., Shapiro, A. J., Wright, F. A., Woodruff, T. J., Chiu, W. A., Guyton, K. Z., and Rusyn, I. (2014). Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ. Health Perspect.* **122**, 499–505.