

Research Article

A Novel Method for Combining Outcomes with Different Severities or Gene-Level Classifications

Salomon Sand

Department of Risk and Benefit Assessment, Swedish Food Agency, Uppsala, Sweden

Abstract

Chemical risk assessment is currently based on consideration of health effects individually. The present work discusses a method for combining data by characterizing the dose-related sequence of the development of lower- to higher-order toxicological effects or the range of bioactivity observed at genomic level caused by a chemical/mixture. A reference point profile (RPP) is defined as the relation between benchmark doses for considered effects (or bioactivity measures) and a standardized severity or rank score determined for these effects. For a given dose of a chemical/mixture, the probability for exceeding the RPP can be assessed. An overall toxicological response can also be derived at the same dose by integrating contributions across all effects, with a rationale for severity weighing. Conversely, dose equivalents corresponding to specified responses can be estimated. The variation in RPPs across chemicals suggests that joint consideration of effects under the proposed concept differentiates the consequence of chemical exposure, both at genomic and apical levels, to a higher extent compared to using a specific effect as a basis. This may help to refine the development of points of departure or sets of such values describing a range of health concerns. Analysis and comparison of apical and genomic RPPs, as well as consideration of functional relations between gene sets within such analyses, may aid in the transition towards a new approach method-based risk assessment paradigm that to a higher degree may require methods for combination of effect data compared to relying on specific outcomes.

1 Introduction

For risk characterization of chemical hazards, margin of exposure-related concepts are generally applied where estimates of the human exposure to a chemical are compared, in one way or another, to a reference level, i.e., a health-based guidance value or a reference point (RP) (a term coined by the European Food Safety Authority (EFSA), 2009) also called point of departure (PoD). The RP is based on the critical health effect and is, whenever possible, recommended to be derived by the benchmark dose (BMD) approach (Crump, 1984; Dourson et al., 1985; EFSA, 2005, 2009, 2017; Haber et al., 2018; US EPA, 2005, 2012).

Sand et al. (2018) proposed to expand the current approach by characterizing the dose-related sequence of RPs (e.g., BMDs) associated with lower- to higher-order health effects, representing low to high severities, instead of only using a single RP for

the critical health effect as a basis for risk characterization. Modeling of data describing different severities has previously been performed by regression methods designed to handle categories (discrete data). Hertzberg and Miller (1985) and Hertzberg (1989) introduced such approaches for species extrapolation. Advantages and possibilities of these methods were discussed for different applications related to chemical risk assessment (e.g., Hertzberg and Dourson, 1993; Dourson et al., 1997; Guth et al., 1997; Teuschler et al., 1999; Milton et al., 2017a,b). As an extension, Chen and Chen (2014) suggested to calculate the BMD as a weighted average of the BMDs obtained at dichotomous cutoffs for each severity level.

Many of the methods discussed above may be used to compute the probability for a given severity category at a given dose. Sand et al. (2018) provides this type of output simultaneously across all categories. In addition, a new response metric that integrates prob-

This content is part of the *Special Issue* Development of an Evidence-Based Risk Assessment Framework. doi:10.14573/altex.22S2

Received April 21, 2020; Accepted July 26, 2021;
Epub August 6, 2021; © The Authors, 2021.

ALTEX 39(3), 480-497. doi:10.14573/altex.2004212

Correspondence: Salomon Sand, PhD
Department of Risk and Benefit Assessment
Swedish Food Agency
Box 622, 751 26 Uppsala, Sweden
(salomon.sand@slv.se)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

ability and severity across the entire severity domain is calculated, and dose equivalents for specified levels of this response can also be derived. The new response metric is a comprehensive quantitative measure of chemical toxicity that is expressed in terms of the most severe health effects. Integration across different severities requires weighting, and the developed system in Sand et al. (2018), with nine severity categories, C1 through C9, is therefore mapped to a quantitative severity scale, $S = 0$ to $S = 1$, and this mapping may be modified systematically by a mathematical function.

The present paper describes in detail how the method introduced in Sand et al. (2018) can influence assessments compared to consideration of health effects individually in line with the current framework. The method is also further developed as a step towards application of this concept to data from new approach methodologies (NAMs) using results from transcriptomic BMD analysis as a basis.

2 Materials and methods

2.1 Methodology for combining data on multiple health effects

The method proposed in Sand et al. (2018) characterizes the dose-related sequence of the development of multiple (lower- to higher-order) toxicological health effects caused by a chemical and calculates a response metric that integrates probability and severity across these effects. A description of the methodology, and its further development for genomic data, follows below.

In Sand et al. (2018), a reference point profile (RPP) is defined as the relation between benchmark doses (BMDs) for a set of health effects and a standardized severity score, S , determined for these effects (see Fig. 1). In this paper, toxicological/biological findings related to chemical exposure are generally termed “health effects” or “effects”. Comparison of BMDs across various types of effects requires normalization/standardization of the corresponding response scales. To increase compatibility in this regard, the benchmark response (BMR) associated with the BMD for apical effects is defined as a change in response relative to the (absolute) difference between the maximum and minimum response level. For quantal data, this corresponds to the “extra risk” definition, and in this case the minimum response (background) is ≥ 0 , and the maximum response equals 1 (100%), while the minimum and maximum response theoretically may assume any values for continuous data. For transcriptomic data, the National Toxicology Program (NTP) recommends that BMDs correspond to a response change equal to one standard deviation (SD) at zero dose/control level (NTP, 2018). For certain assumptions (e.g., Crump, 1995), this translates to an extra risk of 10% (for quantal data) and, similarly, represents a normalized BMR definition. Under the consideration of a normalized BMR, the general equation for the RPP is,

$$\log BMD^{ijk} = \mu(S_{ij(k)}) + \varepsilon_{ijk}, \quad (\text{Eq. 1})$$

where $\mu(S_{ij(k)})$ is the central estimate of the (sigmoidal) RPP that applies to a particular value of the standardized BMR, which

should be the same for all included BMDs. As in Sand et al. (2018), a BMR = 0.21 is used here as default for apical effects, but a BMR = 0.1 is also applied for comparison to transcriptomic BMDs that correspond to a change of one control SD.

In the original approach, $S_{ij(k)}$ is a quantitative value of the severity of toxicity associated with health effects (k) classifying in the j^{th} severity category ($j = 1, 2, \dots, 9$) for the i^{th} chemical/mixture. S_{ij} describes the severity of health effects in a relative sense and is therefore not dependent on the BMR associated with the RPP. The assignment of S_{ij} -values is described in detail in Sand et al. (2018). Briefly, health effects are first classified according to a nine-graded categorical severity scale, C1-C9, which is then mapped to a quantitative scale, $S = 0$ to $S = 1$, where each severity category corresponds to an interval of S -values. Derivation of the default mapping is described in Figure 1. This represents a starting point for the analysis that may later be modified by a severity weighting approach described below in association with Equation 3.

In the extended approach for genomic dose-response information, classification of effects/genes according to gene ontology (GO) category rather than severity category is considered. While the described implementation uses GO categories, it may also be applicable to other types of classifications of gene-level BMDs. Here, the default starting point/mapping of effects/genes to $S_{ij(k)}$ is determined using a dose-dependent approach, described in detail in Section 2.3. $S_{ij(k)}$ represents a quantitative rank value associated with the k^{th} health effect/gene ID within the j^{th} GO category for the i^{th} chemical/mixture. The severity weighting approach in Equation 3 below applies in the same manner as before. S represents a fixed rank value, S_{ijk} , in the extended method while Sand et al. (2018) defines S_{ij} as a value within an interval that is common for all associated, k , health effects. Therefore, S is written as $S_{ij(k)}$ in Equation 1 through 4 to be applicable to both approaches. The genomic RPP is illustrated in Figure 2 using triphenyl phosphate (TPHP) as an example.

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 (Sand et al., 2018) or effect group/GO category (extended method) for the i^{th} chemical/mixture) and is assumed to be normally distributed with constant variance, σ^2 . Variability (σ^2) results since BMDs may differ for health effects within the same severity category (Sand et al., 2018) or across GO categories (extended approach). The RPP is estimated using an iterative approach. In each iteration, a value is randomly sampled from the log-normal BMD uncertainty distribution for each BMD and combined with either 1) an S_{ij} -value randomly sampled from the associated S -interval that is assumed to be uniformly distributed (Sand et al., 2018) or 2) an S_{ijk} -value determined as described in Section 2.3 (extended approach). A sigmoidal RPP model is then fitted to each generated data set, which, overall, provides a point estimate and confidence interval for the RPP. More details on the algorithm for RPP estimation are given in Sand et al. (2018) and below for the approach for genomic data.

Three specific RPP models with different geometrical characteristics were considered in Sand et al. (2018), and the present

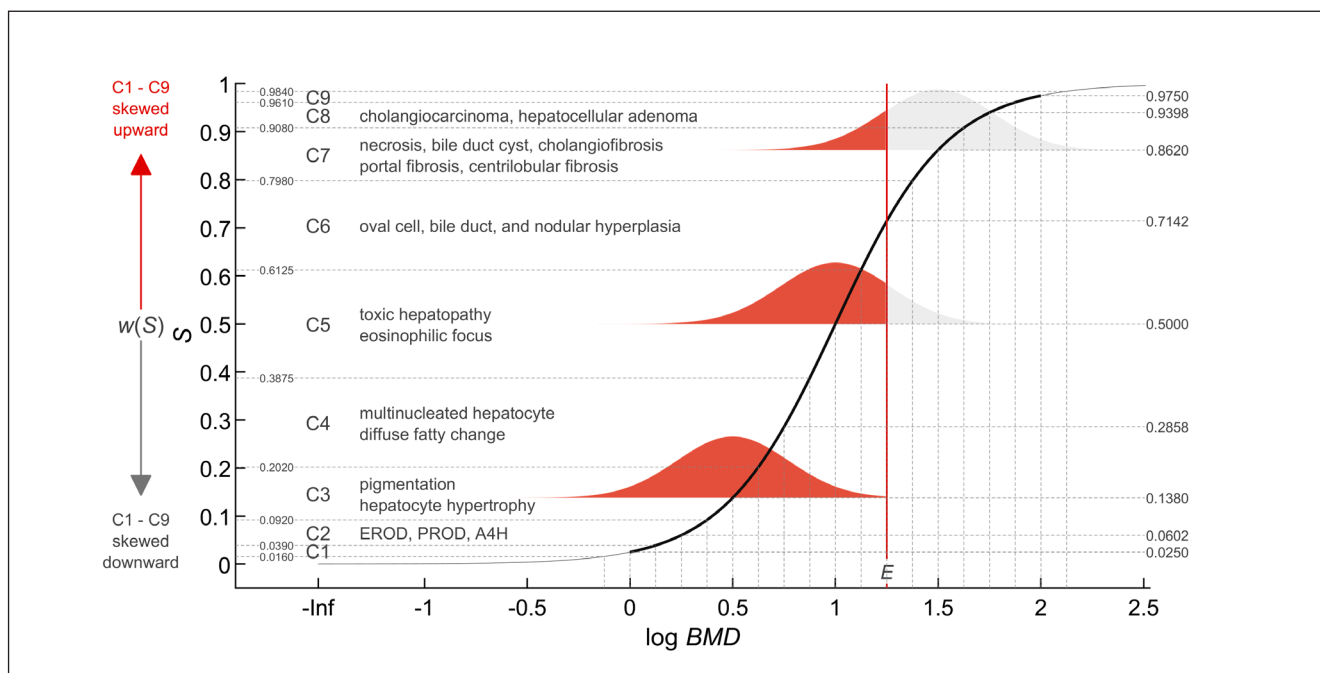


Fig. 1: Technical illustration of the reference point profile (RPP)

The solid S-shaped curve describes the relation between benchmark doses (corresponding to a normalized benchmark response, BMR) for selected apical health effects and the relative severity of toxicity (S) determined for these effects. The RPP applies to the specified value of the normalized BMR and thus represents a cross-section of the dose- S -BMR volume. The relative severity for individual health effects is first determined categorically according to a hierarchical classification scheme: The classification performed in Sand et al. (2018) of considered health effects in the liver is illustrated. The nine-graded categorical scale, C1-C9, is then mapped to a quantitative scale that ranges from $S = 0$ to $S = 1$. The default mapping distributes severity categories symmetrically across S , e.g., with C1, C5, and C9 centered at $S = 0.025$, $S = 0.5$, and $S = 0.975$, respectively (see Sand et al., 2018 for details). The variability is assumed to be normally distributed on the log-scale with constant variance. Red areas describe probabilities, p , for exceeding the RPP at exposure level, E , corresponding to the vertical (red) line. Here, E corresponds to a response (RTR, Equation 3) of 50%, and E intersects the solid curve at $S \approx 0.71$. The midpoint of C6 thus represents the center of the effect/category sequence in terms of RTR. This point of calibration is approximately independent of C1-C9 model parameters (Sand et al., 2018). A non-linear severity-weight, $w(S) \neq S$, can indirectly modify the symmetrical mapping of C1-C9 to S . This allows the midpoint of the system, corresponding to RTR = 50%, to be lower (C1-C9 skewed upward) or higher (C1-C9 skewed downward) than the midpoint of C6, which would also increase or decrease the RTR associated with E , respectively.

work uses a generalized version of one of these models (a generalized Hill model):

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

where H_i is the RPP (dose) location parameter for the i^{th} chemical/mixture that corresponds to a traditional potency measure but based on the consideration of several health effects simultaneously (H_i decreases with increasing potency). Parameter, λ_i , is the RPP shape for the i^{th} chemical/mixture that describes how wide or narrow the range of BMDs is for health effects with different (relative) severities or S_{ijk} ranks, i.e., in the former case the RPP shape/slope is low, indicating a clear relation between BMD and S and *vice versa*. Parameter, ν , is the RPP asymmetry/skewness, where $\nu = 1$ corresponds to a symmetric RPP while $\nu < 1$ and $\nu > 1$ are associated with left- and right-skewed RPPs, respectively. In practice, the RPP is estimated by using a

pre-specified value for ν , so that several RPPs may be defined from Equation 1a, which extends and generalizes the model pool to cover a larger geometrical space compared to Sand et al. (2018). As mentioned, the RPP applies to a particular BMR, and it therefore represents a cross-section of a dose- S response volume where x , y , and z are the dose, the (relative) severity of the health effect or S_{ijk} rank, and the standardized BMR, respectively (see Section 2.2 for further details).

The probability to exceed the RPP along the entire S -domain (the probability to exceed the distribution of BMDs across S) at some exposure, E_i , to a chemical or mixture is described by,

$$p(E_i | S_{ij(k)}) = \phi \left[\frac{\log E_i - \log \frac{\hat{\mu}(S_{ij(k)})}{AF_i}}{\hat{\sigma}} \right], \quad (\text{Eq. 2})$$

where ϕ is the cumulative standard normal distribution function; $\hat{\mu}(S_{ij(k)})$ is based on estimates for the location (\hat{H}_i) and shape ($\hat{\lambda}_i$) of the RPP; and $\hat{\sigma}$ is an unbiased estimator of the SD.

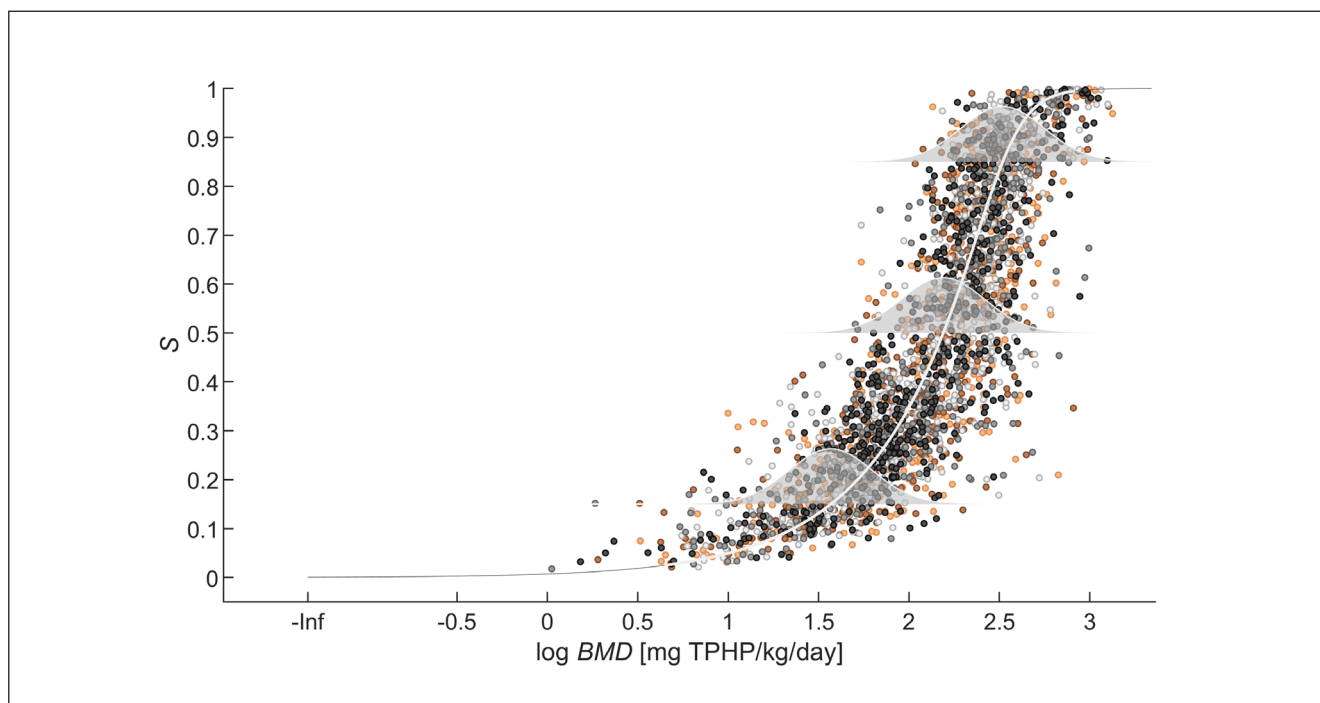


Fig. 2: Genomic RPP for TPHP based on gene-level BMDs derived from a short-term transcriptomic study in male Harlan Sprague Dawley rats

The genomic RPP describes the BMD variability both within and across gene ontology (GO) categories. The central RPP curve (solid white curve) describes the sequence of BMDs associated with GO categories at median. The variation, assumed to be normally distributed on the log dose scale and defined by the RPP SD, describes the variation in BMDs across GO categories. The RPP is estimated using an iterative approach, and the circles correspond to the five iterations (illustrated by different colors, comprising 1% of all iterations) that are closest to the RPP point estimate. They represent snapshots along the BMD sequences associated with different GO categories. The full set of unique BMDs is utilized in each iteration, also accounting for the uncertainty (herein based on reported BMDs, BMDLs, and BMDUs). The GO category assignment associated with a given BMD is, however, randomized across iterations since a given Entrez Gene ID may be part of several GO categories. BMDs are combined with dose-dependent rank values between $S = 0$ and $S = 1$ that describe the fraction of BMDs exceeded within associated GO categories (see Section 2.3). A unique BMD may have several ranks if it is present in several GO categories. As a starting point, individual BMDs are ranked in the same manner across the different GO categories. The paper discusses that methods for comparison of similarity across GO terms may help to refine this ranking and possibly also support parameterization of the systemic weight function, $w(S)$, in Equation 3.

Similar to traditional risk assessment, $\hat{\mu}(S_{ij(k)})$ can be divided by an overall standard or chemical-specific adjustment factor (AF_i). For simplicity, AF_i is set to 1 in this paper.

The risk-based and toxicity integrated response (RTR) associated with exposure E_i is an impact metric defined as the integral of the probability of exceeding the RPP (Equation 2) multiplied by the severity weight, w , described by a beta cumulative distribution function that increases from 0 to 1 across S ,

$$RTR(E_i) = 2 \int_0^1 p(E_i | S_{ij(k)}) \cdot w(S_{ij(k)}) dw = \frac{\int_0^1 p(E_i | S_{ij(k)}) w(S_{ij(k)}) dw}{\int_0^1 w(S_{ij(k)}) dw}, \quad (\text{Eq. 3})$$

$$w(S_{ij(k)} | a, b) = \int_0^{S_{ij}} t^{a-1} (1-t)^{b-1} dt, \quad a > 0, b > 0, \text{ or } a = b = 1.$$

The RTR ranges between 0 and 1 when expressed as a fraction of the maximum value of the $p \times w$ integral that approaches 0.5 as E_i becomes large. Use of this standardized form of the RTR

is a modification compared to Sand et al. (2018), and it translates to a weighted average of the probability, $p(E_i | S_{ij(k)})$, of exceeding the RPP. Within the extended approach for genomic data, the probability, p , might be framed as being related to the degree of perturbing the respective gene-level processes described by the considered categorization of BMDs, herein based on GO biological processes. This interpretation may hold for both methods, depending on the effects used as part of the analysis, e.g., the ranking of apical effects in Sand et al. (2018) covers the adverse outcome pathway discussed for dioxin-like chemicals in rodents. In terms of impact, the RTR corresponds to the (average) probability of exceeding the BMD for outcomes associated with $S = 1$. The risk-based and toxicity integrated dose equivalent (RTD) is the dose that corresponds to a specified value of the RTR. A linear severity weight ($a = b = 1$) is regarded as the default, which is applied throughout this paper. Details for how to estimate the RTR and RTD are described in Sand et al. (2018).



The RTR will reach $\approx 50\%$ for an exposure, E , that intersects with $S = 0.71$ (Fig. 1). Sand et al. (2018) shows that this point of calibrations is approximately independent of the RPP shape (λ) and SD (σ). For the default mapping of severity categories to S , in combination with a linear severity weight, the midpoint of C6 aligns to $S = 0.71$ (see illustration in Fig. 1). Thus, $S = 0.71$ represents the center of the sequence of health effects categorized by severity (C1-C9) or within groupings at genomic level (gene sets) in terms of RTR. In Sand et al. (2018), C6 was considered as the region that separates reversible and irreversible apical health effects. The association of C6 to RTR = 50% under default settings was therefore regarded as a plausible starting point for severity weighting. Technically, this can be modified by using non-linear weights in a systematic manner (a or $b \neq 1$), which indirectly will skew severity categories or the genomic BMD distribution within GO categories upwards or downwards (as indicated in Fig. 1) and thus re-modulate relative severity or ranks. This is more suitable than having to perform weighting at the level of individual effects, and consideration of the center of the effect sequence/s in the way described above can be used to reduce the complex problem of severity weighting to (value-based) selection of a single parameter (a or b). Since the severity weighting acts at the systemic level, the original formulation fits well with the consideration of effects as part of a biological process, which is further emphasized in the extended approach. For application to genomic data, $w(S)$ may be framed as the relative severity of perturbation, i.e., the severity associated with exceeding given fractions, S , of BMDs within considered GO categories (here, $S = 0$ to $S = 1$ enclose the dose range covered by BMDs within and across GO categories, see Section 2.3).

2.2 Generalization

As noted earlier, the RPP is a cross-section of the dose-S-BMR volume at the selected normalized BMR. As an extension of the method, the RPP associated with a given BMR may be dose-adjusted by multiplying the RPP location term, \hat{H}_i , by a factor, H_x , to approximate RPPs associated with other BMRs (Sand et al., 2018). The factor H_x is calculated from the shapes/slopes of dose-response curves obtained in the underlying BMD analysis of individual health effects, and the shapes for all curves are allowed to contribute equally to the shape of the averaged curve. By derivation of RPPs associated with BMRs = 0 to BMR = 1, 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,

$$2D-RTR(E_i) = \int_0^1 RTR \, dBMR, \quad (\text{Eq. 4})$$

where the RTR associated with E_i is calculated using Equation 3 across RPPs spanning the BMR domain, and Equation 4 is then calculated by numerical integration. In standardized form, 2D-RTR ranges between 0 and 1. In terms of impact, it corresponds to the (average) probability of exceeding the dose-response curves (i.e., the BMD associated with BMR levels of 0 through 1) for outcomes associated with $S = 1$. 2D-RTR is thus a proxy/surrogate for the probability of response with re-

gard to such health effects/outcomes. Details on how to estimate 2D-RTR as well as the corresponding RTD are described in Sand et al. (2018).

2.3 Extension to genomic data

Based on the analysis in Sand et al. (2018), the method appears robust to minor as well as moderate changes in severity classification of BMDs. Even so, the categorization of health effects is a practical challenge, involving subjective judgement, and, as noted in Sand et al. (2018), all health effects recorded within a study may not necessarily be relevant to include as part of the (primary) RPP. It may be more practical to use a dose-dependent approach that utilizes information across several studies and/or effect groups to support a generic (model-based) ranking. For example, the order of BMDs for the different health effects within studies used in Sand et al. (2018) would jointly produce a categorical rank (e.g., using the average/median rank across studies for each effect) that, relatively speaking, would resemble the severity categorization performed. Such an approach would help to automate the initial ranking of health effects, and toxicological and/or policy judgments can be made at the systemic level, utilizing $w(S)$ as discussed, for differentiation of different types of effect sequences/processes. This type of design may also be more practical for data from NAMs, for example, since determination of severity for individual measures of bioactivity may be problematic. Previously published gene-level BMDs from short term studies in rodents mapped to GO categories have been used to illustrate how the concept can be extended along these lines. The ranking approach and estimation of the RPP are described in subsections below.

Ranking approach for derivation of S -values

Using all BMD point estimates within GO categories for a given chemical, $x = 100$ doses corresponding to percentile values between $P_A = 100/h$ and $P_B = 1 - P_A$ are computed for each category, where h is the highest number of BMDs within GO categories. All non-unique values are then removed, i.e., for categories with few BMDs, only part of the percentile range, P_A to P_B , may be computed with unique values. Using the matlab function “prctile”, $P_A = 100/(2 \times h)$ would correspond to the limit for the lowest unique percentile that can be calculated for h number of input values, so the suggested approach is a bit more stringent than this. Doses corresponding to P_A through P_B are then averaged across the GO categories. This produces the dose vector, xm , which reflects an average version of the within GO category BMD distribution. Initial analysis indicated that xm is left skewed across evaluated chemicals/data sets. A generalized logistic distribution, which is a flexible model for extreme valued data, is therefore considered with probability density function,

$$f(xm) = \frac{v \cdot e^{\left(\frac{-(xm-\mu)}{\sigma_{xm}}\right)}}{\sigma_{xm} \left[1 + e^{\left(\frac{-(xm-\mu)}{\sigma_{xm}}\right)}\right]^{v+1}}, \quad (\text{Eq. 5})$$

where μ , σ_{xm} , and v describe the location, scale (shape 1), and asymmetry (shape 2), respectively (observe that σ is used to represent the RPP SD, not to be confused with σ_{xm}). The associated

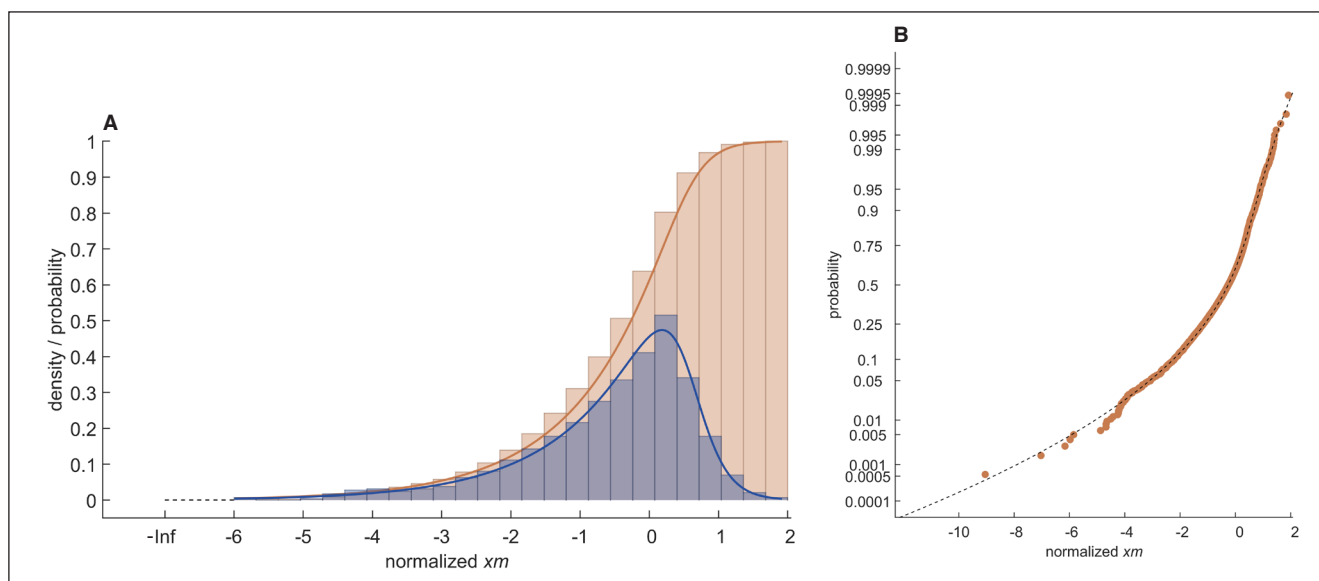


Fig. 3: Generalized logistic probability density and cumulative distribution functions fitted to a normalized version of xm and the corresponding probability profile of nine data sets covering five chemicals combined

(A) The generalized logistic probability density and cumulative distribution functions fitted to a normalized version of xm (xm reflects the within GO category BMD distribution, on average). Nine data sets covering five chemicals have been combined: TPHP, DBE-47, DE-71, MCHM, and PPH (two data sets for each chemical are included except for TPHP). For a given data set, the normalized xm is defined as:

$$xmn_i = \frac{xm_i - c_i}{sv_i}, \quad sv_i = \sqrt{\frac{6}{\pi^2} \cdot \text{var}_i \cdot c_i} = \text{mean}(xm_i) + sv_i \cdot E.$$

where sv_i and c_i are estimates of the scale and location/mode, respectively, of an extreme value distribution where var_i is the variance of xm_i (on log scale) and E is Euler's constant. The extreme value distribution with two parameters provides a reasonable fit across a large part of xmn . However, the generalized logistic distribution, with an additional parameter for asymmetry/skewness, fits well across the whole range as further described by the probability profile in part (B), indicating that the model is also compatible at the very lowest xmn values (not shown in part A). Maximum likelihood estimates (with 95% confidence intervals) of the location, scale (shape 1), and asymmetry (shape 2) are 0.58 (0.50-0.66), 0.25 (0.21-0.28), and 0.20 (0.16-0.24), respectively. The estimate of the asymmetry term used as the generic part of the ranking approach is not dependent on the constant term ($6/\pi^2$) in sv_i or the $sv_i \times E$ product in c_i .

cumulative distribution function (CDF), here denoted $f(xm)$, is effectively the same as the RPP in Equation 1a. Equation 5 is fitted to a normalized version of xm combined across chemicals, illustrated in Figure 3. The model (Fig. 3) nicely describes the considered data and is used to derive a generic representation of the RPP asymmetry/skewness parameter, v , which is estimated to 0.2 with a 95% confidence interval of 0.16-0.24. By pre-specifying v to 0.2, Equation 5 is then fitted to xm , normalized to its mean on log scale, for chemicals individually. The resulting $f(xm)$ provides an initial estimate of the RPP slope, λ , through σ_{xm} (describing BMD variability within GO categories), which, apart from the generic element, v , guides the derivation of S_{ijk} -values. More specifically, for a given chemical BMD, point estimates within each GO category are normalized to the mean BMD for that category (on log scale), and S_{ijk} -values (fixed ranks) associated with BMD^{ijk} within the j^{th} GO category are determined according to $f(xm)$. Thus, the BMD range within GO categories is regarded to be chemical-specific, guided by σ_{xm} , and in line with the interpretation of λ in Equation 1a, while the manner in which BMDs/genes within GO categories are distributed across S is generic, guided by v . To assess model uncertainty, $v = 0.16$ and 0.24 are considered be-

sides $v = 0.2$. Also, a sensitivity analysis is performed related to the number of percentile values, x , used for derivation of xm , and the scaling of xm and BMDs towards the mean.

Notice that a given BMD will have several ranks if it belongs to several GO categories. Here, $S = 0.5$ represents the median log BMD across GO terms, according to $f(xm)$, and as a starting point a linear severity-weight, $w(S)$, is applied across categories. It can be noted that Sand et al. (2018) distributes effect categories, C1-C9, across S using a logistic CFD that corresponds to using $v = 1$ in Equation 5.

Simulation of BMD values across GO categories

The uncertainty in BMD^{ijk} is assumed to be normally distributed on the log scale, with a mean corresponding to the log BMD point estimate and an SD approximated by $\hat{s}_{BMD} = \frac{\log BMDU - \log BMDL}{2 \cdot 1.645}$, where the range between the upper (log BMDU) and lower (log BMDL) bound on the BMD represents the two-sided 90% confidence interval.

Each unique BMD uncertainty distribution within a given data set is randomly assigned to one of the GO categories it belongs to. A BMD^{ijk} value is randomly sampled from the uncertainty dis-



tribution and matched with its S_{ijk} -value (derived as described above) for the assigned GO category. This is performed across all unique BMDs and results in a set of BMD_{ijk} and S_{ijk} -values reflecting the variability across GO categories and ranks, representing a generated RPP data set. In this process, the random selection of a GO category, among the group of categories that includes BMD_{ijk} , is weighted inversely proportional to the number of BMDs within each of the categories so that a category with many BMDs is down-weighted vs a category with a few BMDs. This is done to cover the full BMD variability across categories, which otherwise would not be the case. Also, if an Entrez Gene ID is associated with more than one unique BMD or if a unique BMD is associated with more than one Entrez Gene ID, this variation will be reflected within and between iterations, respectively. Figure 2 illustrates 1% of all iterations for TPHP as an example.

Estimation of the genomic RPP

Model 1a is fitted, using a constant $\nu = 0.20, 0.16$ or 0.24 (matched to that used above), by the least squares method: The sum of squares of $\log BMD_{ijk} - \log \mu(S_{ijk})$ is minimized with respect to H_i and λ_i , which is a usual regression problem with explicit solutions. The estimate of σ^2 is also obtained, from which an unbiased estimator of σ can be derived (Sand et al., 2003). A 90% confidence interval for the genomic RPP is derived using $n = 500$ iterations. Median values of \hat{H}_i , $\hat{\lambda}_i$, and $\hat{\sigma}$ are used to represent a point estimate of the whole RPP, denoted “RPP point estimate”, describing the most central GO category, and the variation (at median) across categories (see Fig. 2).

In summary, the extended approach differs from Sand et al. (2018) with respect to the derivation of S -values and that they represent fixed ranks for BMD_{ijk} (a given BMD has, however, different ranks across categories). Also, several effect sequences (GO categories) are included in the same analysis using a novel simulation approach, but this is more a function of differences in data types (BMDs from traditional vs genomic data). If a data set just consists of one GO category, the genomic RPP will only reflect that category. The response metric, RTR, and its dose equivalent, RTD , are calculated as described earlier in association with Equations 2-4.

Refinement of S -values and severity weighting

Comparison of GO categories using tools for computation of similarity scores may potentially help to refine the determination of S_{ijk} values as well as supporting parameterization of the systemic weight function, $w(S)$, in Equation 3. To illustrate this potential, the web application (MegaGO) from Verschaffelt et al. (2021) was used to calculate similarity between GO terms with the Lin semantic similarity metric (Lin, 1998). The tool provides a score between 0 and 1 for each of the three GO domains, but for the present data a score for “biological process” results exclusively since this is the only domain covered.

2.4 Data and assessment

For estimation of apical RPPs, data from the NTP on dioxin-like chemicals and their mixtures were used. These data (changes in enzyme activity, non-neoplastic and neoplastic liver lesions)

were observed in female Harlan Sprague-Dawley rats at 53 weeks or at the end of the 2-year studies (NTP, 2006a-e, 2010). They are also tabulated as part of the supplementary material in Sand et al. (2018) together with estimated BMDs corresponding to a 21% change in response (Sand et al., 2006, 2012). Based on this material, it is investigated how the proposed method compares with a traditional approach based on individual health effects. The RTR and 2D-RTR associated with BMDs for specific health effects are evaluated as a basis for these comparisons. It is also studied whether information corresponding to the lower severity categories provides adequate estimates of the RTD . The results in Sand et al. (2018), based on data in severity categories C2-C8 (Fig. 1), are contrasted to RTD derived herein for the same data sets but where data in categories C7 and C8 have been omitted.

For estimation of genomic RPPs, gene-level BMDs from 5-day toxicogenomic studies in rodents for TPHP, 4-methylcyclohexanemethanol (MCHM), and propylene glycol phenyl ether (PPH) were used (NTP, 2021). Gene-level BMDs for 2,2,4,4-tetrabromodiphenyl ether (PBDE-47) and technical pentabromodiphenyl (DE-71) from Dunnick et al. (2018, 2020) were also considered, where dams were dosed from gestation day 6 through postnatal day 21, and from GD 6 through postnatal day 4, respectively. In all studies, the effects of chemical exposure were assessed in the liver, and also in the kidney for MCHM and PPH. Overall, this provided nine data sets for evaluation.

The classification of gene-level BMDs in GO biological process categories performed within the respective study was utilized. In line with NTP (2018), individual BMDs larger than the highest dose were excluded, and GO categories for which the mean BMD (on log scale) was 10 times smaller than the lowest dose were also excluded. Moreover, BMDs for which the ratio between the BMDU and BMDL was larger than a factor 100 were excluded. NTP (2018) recommends a factor 40 as the breaking point, but a less stringent criterion was used herein to better allow evaluation of the impact of uncertainty in the RPP analysis. Also, GO categories populated with fewer than 5 BMD values were not considered.

As part of the evaluations, the lowest mean and median log BMD across included GO categories was used to represent the type of potency estimate that previously has been discussed for this type of data (e.g., NTP 2018). Like the apical RPP analysis, the RTR associated with this type of PoD (mean/median log BMD) was evaluated across chemicals to assess how consideration of individual BMDs vs the whole RPP may affect potency estimation. $RTDs$ corresponding to RTRs of 0.001, 0.01, 0.1, and 0.5 were also derived for the different data set. The uncertainty in $RTDs$ was compared to that resulting from selection of the lowest mean/median log BMD. For DE-71, a comparative analysis of genomic and apical RPPs was performed using data from transcriptomic analysis described above, and data from a long-term NTP study in the same species (NTP, 2016). Apical BMD analysis was performed for quantal liver lesions as described in Sand et al. (2018) using $BMR = 0.1$ (extra risk) to match gene-level BMDs that are associated with a response change equal to one SD at control level.

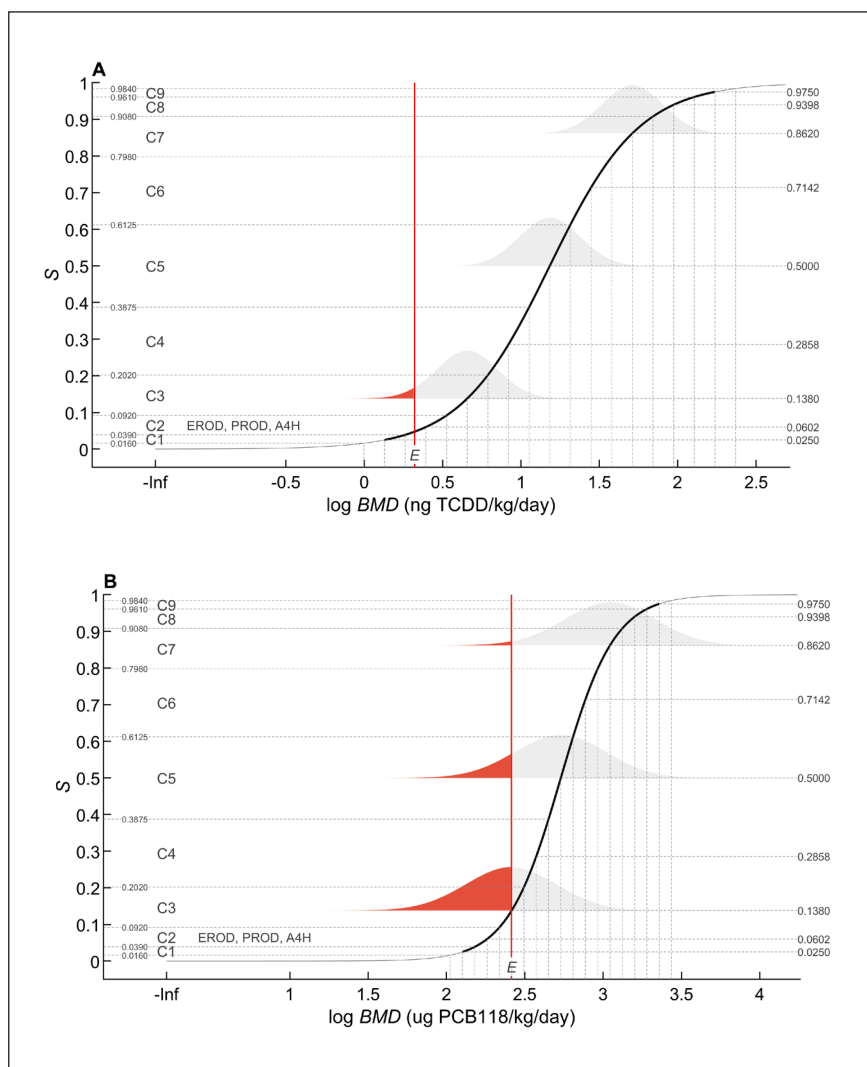


Fig. 4: Reference point profiles (RPPs) for (A) TCDD and (B) PCB₁₁₈

The point estimates of \hat{H}_i , $\hat{\lambda}$, and $\hat{\sigma}$ are 15.3, 1.51, and 0.45 for TCDD, and 537, 2.53, and 0.67 for PCB₁₁₈, respectively. Vertical lines indicate BMDs (corresponding to a 21% change in response) for EROD activity classified in category C2. While these BMDs would normally be regarded as equipotent doses, the associated probability profiles across S differ (red areas) and the RTR (Equation 3) is about 0.5% for TCDD and 10% for PCB₁₁₈. Note that the effects at C2 (EROD, PROD, A4H) represent continuous data, while the effects at C3–C8 represent quantal data. The definition of the BMR as a percent change (21%) in response in relation to the difference between the minimum (estimated background) and the maximum response (estimated response as the dose approaches infinity, which is always 1 for quantal data) levels provides a form of mathematical standardization across the two data types. However, the continuous and quantal responses still have a different interpretation, i.e., mean response vs probability of response. Even though there is no perfect match between the continuous and quantal BMDs within the RPP, the estimated RTRs (0.5 vs 10%) nevertheless indicate different impacts for the same effect dose (the continuous BMDs) across the two chemicals.

3 Results

The consequence of combining data on multiple apical health effects according to Sand et al. (2018) in relation to considering effects individually is illustrated in Figure 4 for TCDD and PCB₁₁₈. The new response metric, RTR (Equation 3), is evaluated here at doses corresponding to the BMDs for EROD activity, classified as severity category C2 (Fig. 1). BMDs defined in terms of a specific response level for a given health effect would normally be regarded as equipotent doses. However, the point estimate of the RTR associated with these BMDs is about 0.5% and 10% for TCDD and PCB₁₁₈, respectively, i.e., the RTRs differ by around a factor 20. In Table 1, both RTR and 2D-RTR associated with BMDs for all three liver enzyme parameters used in the RPP analysis (EROD, PROD, and A4H) are given. RTR point estimates differ by a factor 7–21 between TCDD and PCB₁₁₈ depending on health effect. For 2D-RTR, differences between the chemicals are less pronounced and somewhat more similar across health effects: 2D-RTR point estimates differ by a factor 3–7 be-

tween TCDD and PCB₁₁₈ depending on health effect. If the RPP SD for TCDD was the same as for PCB₁₁₈, i.e., considering a difference in RPP shapes only, RTR and 2D-RTR point estimates would differ by a factor 4–10 and a factor 3–5, respectively, between chemicals depending on health effect (data not shown).

Since the proposed method, particularly in the case of traditional toxicity data, is quite data-intensive, the effect of removing data for the highest severity categories was assessed as a supplement to the previous analyses in Sand et al. (2018). In Figure 5, *RTDs* associated with RTR of 2% derived from “reduced data” (severity categories C2 to C6) and “complete data” (severity categories C2 to C8) are shown for all chemicals/mixtures considered in Sand et al. (2018). Removal of C7 and C8 reduces information (no. of BMD values) by at most 39% (two-component mixture, PCB₁₂₆:PCB₁₁₈) and at least 23% (PeCDF), while data reduction is between 29% and 35% for other chemicals/mixtures (Fig. 5). *RTDs* associated with RTR = 2% are similar for the reduced vs complete data (Fig. 5). Point estimates generally differ by about a factor 1.1 and at most by a factor 1.3 (two-component mixture).

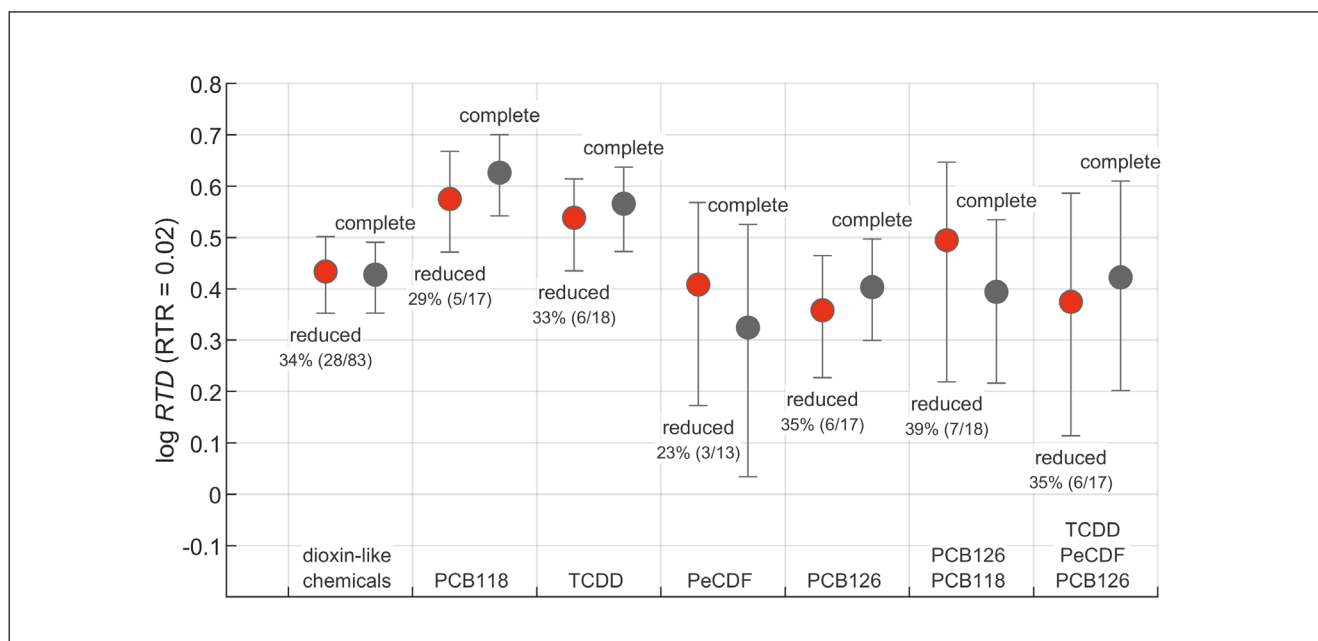


Fig. 5: *RTDs* corresponding to *RTR* = 2%

Results for “reduced data” where severity categories C7 and C8 have been omitted are compared to results for “complete data” published in Sand et al. (2018). BMDs for PeCDF and PCB₁₂₆ have been adjusted by the relative potency vs TCDD before estimation of the *RTD* (see Sand et al., 2018). Results are given in ng TEQ/kg/day except for PCB₁₁₈ (μg/kg/day). For PCB₁₁₈, results have also been adjusted with -1.5 log units to enable graphical illustration together with the other chemicals.

Table 2 presents parameter estimates associated with the genomic RPP for five evaluated chemicals (nine data sets). Results associated with two scenarios are shown: 1) one based on the proposed method using BMD uncertainty distributions as input, and 2) another that uses BMD point estimates where the step of simulating BMDs from the uncertainty distributions is omitted, keeping everything else equal. As shown, the median location, H , and shape, λ , across iterations is more or less the same, as expected, while the SD is larger for the proposed protocol compared to using BMD point estimates only. Overall, results reflect a range of combinations of RPP slopes and SDs, indicating difference of such features at the genomic level. The range of this variation in RPP structure is illustrated by BDE-47 and MCHM in Figure 6.

Table 3 includes information on the *RTR* associated with the lowest mean and median log BMD across GO categories for the evaluated chemicals. This analysis is similar to that based on apical effects in Table 1. The point estimate of the *RTR* associated with the considered GO category specific potency estimates (lowest mean and median log BMD) range between 0.05% and 10% depending on the chemical/data set. Similar to the results for apical effects (Tab. 1), consideration of multiple BMDs, in this case across GO categories described by the genomic RPP, can affect estimates of potency.

Table 3 also presents *RTDs* associated with *RTRs* between 0.001 and 0.5. Results are based on using BMD uncertainty distribution as input for RPP estimation. The *RTDs* would increase if the uncertainty in BMDs were not accounted for since

only the RPP SD (besides differences in parameter confidence intervals) practically differed between the two scenarios in Table 2. Thus, using the proposed method, the *RTD* will be less conservative as the uncertainty in the BMD reduces. Comparing ratios between the lower bound on the *RTD* across the two scenarios in Table 2 shows that results become a factor 1.3–1.6 (*RTR* = 0.001) to a factor 1–1.1 (*RTR* = 0.5) more conservative under the proposed method (Tab. 3). While the method accounts for uncertainty in an appropriate manner (less conservative when less uncertain), the uncertainty in BMDs does not appear to greatly affect the results in quantitative terms. When using a GO category-specific potency estimate, i.e., the lowest mean/median log BMD, the uncertainty in terms of the mean BMDU to mean BMDL ratio ranges between 2.6 and 8.6, which are not extremes considering the range of corresponding ratios for all GO categories within a data set (Tab. 3). In contrast, the ratio between the lower and upper bound on the *RTD* under the proposed method, also accounting for RPP model uncertainty (RPP asymmetry terms, $\nu = 0.20, 0.16$ or 0.24), is a factor 1.1 to 2.9 across evaluated *RTRs*.

A sensitivity analysis was performed since some settings associated with the extended method may be subject to modification. As noted in Section 2.3, derivation of the dose vector, xm , involves generating $x = 100$ doses corresponding to linearly distributed percentile values. This was compared to using $x = 50$ instead. Also, xm and BMDs are normalized to the mean, and normalization to the median was performed as an alternative. *RTDs* associated with *RTRs* of 0.001 to 0.5 under the proposed proto-



Tab. 1: The RTR evaluated at doses corresponding to BMDs for individual health effects

Chemical	Health effect	BMD ^a	RTR (Equation 3)			2D-RTR (Equation 4)		
			point estimate	L05	U95	point estimate	L05	U95
TCDD	EROD	2.1	0.0048	0.0028	0.0090	0.012	0.0099	0.016
	PROD	3.0	0.012	0.0074	0.020	0.020	0.017	0.026
	A4H	3.7	0.020	0.013	0.032	0.028	0.023	0.035
PCB ₁₁₈	EROD	260	0.10	0.076	0.13	0.084	0.073	0.096
	PROD	229	0.077	0.056	0.10	0.072	0.062	0.082
	A4H	327	0.16	0.13	0.19	0.11	0.098	0.13

Note: The point estimate and 90% confidence interval (L05 and U95) for RTR and 2D-RTR, evaluated at doses corresponding to the BMD (point estimate) for individual health effects (EROD (7-ethoxyresorufin-O-deethylase), PROD (7-pentoxoresorufin-O-deethylase) and A4H (acetanilide-4-hydroxylase)) classifying in severity category C2 (Fig. 1). Results are associated with an RPP asymmetry parameter, ν , in Equation 1a equal to 1. ^a BMDs are taken from Sand et al. (2018, Supplementary Table 5), and are associated with a 21% change in response with respect to given health effects. Units are in ng/kg/day for TCDD and $\mu\text{g/kg/day}$ for PCB₁₁₈.

Tab. 2: Genomic RPP model parameter estimates

Chemical	BMD-input ^a	location, H (mg/kg/day)			shape, λ			standard deviation, σ		
		PE	LB	UB	PE	LB	UB	PE	LB	UB
TPHP <i>liver</i>	uncertainty distributions	348	333	364	4.2	4.1	4.4	0.53	0.50	0.57
	point estimates	348	339	356	4.2	4.2	4.3	0.31	0.30	0.33
BDE-47 <i>liver</i> <i>pnd 22</i>	uncertainty distributions	29	26	32	3.1	3.0	3.3	0.78	0.69	0.88
	point estimates	29	27	31	3.1	3.1	3.2	0.54	0.49	0.59
BDE-47 <i>liver</i> <i>pnd 4</i>	uncertainty distributions	38	33	44	3.3	3.0	3.6	0.73	0.61	0.85
	point estimates	38	34	42	3.3	3.1	3.4	0.48	0.42	0.55
DE-71 <i>liver</i> <i>pnd 22</i>	uncertainty distributions	35	33	38	4.2	4.0	4.5	0.68	0.63	0.75
	point estimates	35	34	37	4.2	4.1	4.3	0.49	0.46	0.52
DE-71 <i>liver</i> <i>pnd 4</i>	uncertainty distributions	40	37	44	8.0	7.2	8.9	0.53	0.45	0.60
	point estimates	40	38	43	8.0	7.5	8.5	0.34	0.29	0.41
MCHM <i>liver</i>	uncertainty distributions	402	372	437	15	13	18	0.30	0.25	0.34
	point estimates	401	392	410	15	15	16	0.09	0.08	0.12
MCHM <i>kidney</i>	uncertainty distributions	363	332	402	6.2	5.4	7.0	0.50	0.44	0.58
	point estimates	364	346	380	6.2	5.9	6.5	0.28	0.25	0.31
PPH <i>liver</i>	uncertainty distributions	955	856	1061	7.1	6.1	8.2	0.42	0.36	0.52
	point estimates	949	888	1026	7.1	6.7	7.4	0.29	0.25	0.34
PPH <i>kidney</i>	uncertainty distributions	758	667	864	7.3	6.0	9.3	0.46	0.39	0.56
	point estimates	761	720	801	7.3	6.9	7.7	0.30	0.27	0.34

Note: RPP model parameter estimates (H, λ , and σ) are given as medians (PE) with two-sided 90% confidence bounds (LB and UB) based on 500 iterations. The RPP asymmetry parameter, ν , in Equation 1a is set to 0.2. ^a According to the proposed method, the RPP is estimated using BMD uncertainty distributions as input, but for comparison results associated with using BMD point estimates as input are also shown.



Tab. 3: *RTDs* (mg/kg/day) associated with different *RTRs*, including comparison to the lowest mean/median log BMD (mg/kg/day) across gene ontology (GO) categories

Chemical	PoD-type	RTR	Point estimate	Lowest LB	Highest UB	LB-ratio ^a	UB/LB ^b	Sensitivity analysis ^c
TPHP <i>liver</i>	<i>RTD</i>	0.001	4.6	3.3	6.2	1.3	1.9	1.2
		0.01	18	15	21	1.2	1.4	1.1
		0.1	71	67	75	1.2	1.1	1.1
		0.5	232	220	244	1.0	1.1	1.1
	lowest mean BMD	0.038	40	19	85	-	4.6 (2.0-16)	-
	lowest median BMD	0.035	38					
BDE-47 <i>liver</i> <i>pnd 22</i>	<i>RTD</i>	0.001	0.080	0.051	0.12	1.4	2.3	1.1
		0.01	0.50	0.38	0.63	1.4	1.6	1.0
		0.1	3.2	2.8	3.6	1.3	1.3	1.1
		0.5	17	15	18	1.0	1.2	1.1
	lowest mean BMD	0.028	1.1	0.40	3.3	-	8.2 (2.1-49)	-
	lowest median BMD	0.058	2.1					
BDE-47 <i>liver</i> <i>pnd 4</i>	<i>RTD</i>	0.001	0.14	0.079	0.23	1.5	2.9	1.2
		0.01	0.79	0.57	1.1	1.4	1.9	1.1
		0.1	4.7	3.9	5.5	1.3	1.4	1.2
		0.5	22	20	26	1.1	1.3	1.1
	lowest mean BMD	0.035	2.1	0.70	6.0	-	8.6 (2.2-50)	-
	lowest median BMD	0.039	2.2					
DE-71 <i>liver</i> <i>pnd 22</i>	<i>RTD</i>	0.001	0.39	0.29	0.51	1.4	1.8	1.1
		0.01	1.5	1.3	1.8	1.3	1.4	1.0
		0.1	6.2	5.6	6.9	1.2	1.2	1.0
		0.5	23	21	25	1.0	1.2	1.1
	lowest mean BMD	0.015	2.0	0.95	4.1	-	4.3 (2.2-34)	-
	lowest median BMD	0.046	3.8					
DE-71 <i>liver</i> <i>pnd 4</i>	<i>RTD</i>	0.001	3.0	2.2	3.8	1.4	1.7	1.2
		0.01	6.1	5.0	7.4	1.4	1.5	1.1
		0.1	14	12	16	1.2	1.3	1.1
		0.5	32	30	35	1.0	1.2	1.1
	lowest mean BMD	0.00048	2.3	1.1	5.1	-	4.8 (2.1-21)	-
	lowest median BMD	0.027	8.5					
MCHM <i>liver</i>	<i>RTD</i>	0.001	98	80	116	1.4	1.4	1.0
		0.01	145	128	162	1.4	1.3	1.0
		0.1	224	207	241	1.3	1.2	1.0
		0.5	356	333	383	1.1	1.2	1.0
	lowest mean BMD	0.051	195	122	312	-	2.6 (1.6-3.5)	-
	lowest median BMD	0.10	224					
MCHM <i>kidney</i>	<i>RTD</i>	0.001	16	11	22	1.5	2.0	1.2
		0.01	41	33	50	1.4	1.5	1.0
		0.1	108	97	121	1.3	1.3	1.1
		0.5	272	251	302	1.0	1.2	1.1
	lowest mean BMD	0.047	78	30	199	-	6.6 (1.6-12)	-
	lowest median BMD	0.068	91					



Chemical	PoD-type	RTR	Point estimate	Lowest LB	Highest UB	LB-ratio ^a	UB/LB ^b	Sensitivity analysis ^c
PPH <i>liver</i>	RTD	0.001	64	44	86	1.4	2.0	1.1
		0.01	145	114	174	1.3	1.5	1.1
		0.1	337	293	373	1.2	1.3	1.0
		0.5	743	679	808	1.0	1.2	1.1
	lowest mean BMD	0.0083	136	63	295	-	4.7 (1.5-14)	-
	lowest median BMD	0.046	251					
PPH <i>kidney</i>	RTD	0.001	53	32	76	1.6	2.4	1.0
		0.01	116	87	143	1.4	1.7	1.0
		0.1	265	232	294	1.2	1.3	1.1
		0.5	594	542	661	1.1	1.2	1.1
	lowest mean BMD	0.011	118	40	347	-	8.6 (1.6-19)	-
	lowest median BMD	0.0035	81					

Note: Point estimates for the RTD based on an RPP model with asymmetry term, $\nu = 0.2$, and 90% confidence intervals defined by the lowest lower bound (lowest LB) and the highest upper bound (highest UB) across three estimated RPP models with $\nu = 0.2, 0.16$, and 0.24 are shown. RTDs correspond to RTR between 0.001 and 0.5 and are based on 500 iterations. The lowest mean and median log BMD across GO categories is also presented for each data set, where LB and UB correspond to mean log BMDL and mean log BMDU, respectively, for the relevant GO category. The RTR (point estimate) associated with lowest mean and median log BMD is also given.

^a Ratio between the lower bound on the RTD derived using BMD point estimates as input for RPP estimation and the lower bound on the RTD derived using BMD uncertainty distributions as input for RPP estimation. Results are associated with an RPP model with asymmetry term, $\nu = 0.2$. ^b Ratios between the lowest LB and the highest UB based on the consideration of results across three RPP models ($\nu = 0.2, 0.16$, and 0.24) using BMD uncertainty distributions as input and ratios between mean log BMDU and mean log BMDL for the GO category with the lowest mean log BMD. The lowest and highest ratio between mean log BMDU and mean log BMDL across all GO categories is given in parenthesis. ^c The largest relative difference between RTDs across three scenarios: 1) the proposed approach that uses $x = 100$ percentile values for derivation of the dose vector, xm , and performs normalization to the mean, 2) using $x = 50$ percentile values and normalization to the mean, and 3) using $x = 100$ percentile values and normalization to the median. The largest difference is assessed across point estimates, lower and upper bounds associated with an RPP model with asymmetry term, $\nu = 0.2$.

Tab. 4: Similarity score according to the application by Verschaffelt et al. (2021) between sets of GO categories that differ in dose location with regard to log mean BMD described by the percentile range (P_{x-y})

Chemical	P ₀₋₁₀ vs P ₉₀₋₁₀₀	P ₅₋₁₅ vs P ₈₅₋₉₅	P ₃₅₋₄₅ vs P ₅₅₋₆₅	P ₄₀₋₅₀ vs P ₅₀₋₆₀
TPHP <i>liver</i>	0.55 (125/125) ^a	0.59 (125/125)	0.76 (125/125)	0.79 (125/125)
BDE47 <i>liver pnd 22</i>	0.31 (13/13)	0.33 (13/13)	0.62 (13/12)	0.70 (12/13)
BDE47 <i>liver pnd 4</i>	0.44 (30/30)	0.46 (30/30)	0.67 (30/31)	0.60 (30/31)
DE71 <i>liver pnd 22</i>	0.32 (20/20)	0.38 (21/21)	0.64 (20/20)	0.61 (20/21)
DE71 <i>liver pnd 4</i>	0.43 (57/57)	0.46 (57/56)	0.66 (57/58)	0.68 (57/57)
MCHM <i>liver</i>	0.40 (32/32)	0.48 (32/33)	0.70 (32/32)	0.73 (32/33)
MCHM <i>kidney</i>	0.30 (37/35)	0.41 (40/37)	0.68 (37/36)	0.77 (37/37)
PPH <i>liver</i>	0.38 (28/28)	0.41 (28/27)	0.62 (28/27)	0.71 (29/28)
PPH <i>kidney</i>	0.55 (48/48)	0.47 (49/51)	0.64 (48/48)	0.67 (46/46)

Note: P_{x-y} identifies GO categories with mean log BMD classifying in the given percentile range considering mean log BMD values across all GO terms. ^a Number of GO categories (gene sets) within the respective percentile range. Verschaffelt et al. (2021) application accessed June 2021.

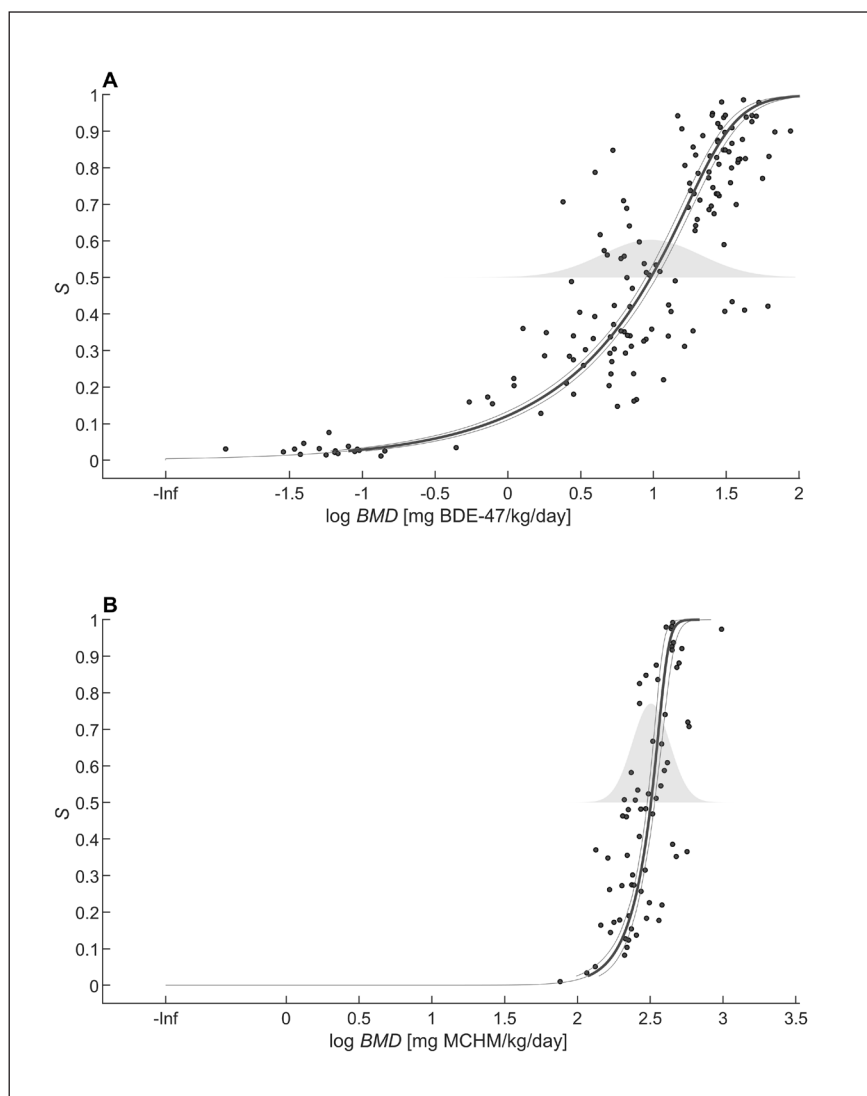


Fig. 6: Genomic reference point profile for BDE-47 (pnd 22) (A) and MCHM (B) based on gene-level BMDs derived from neonatal and adult rat liver transcriptomic data, respectively

Circles correspond to the iteration of the approach that is closest to the RPP point estimate (based on $n = 500$ iterations) illustrated by the solid curve and associated normal distribution. Model parameter estimates are given in Table 2. The lower RPP slope and higher RPP SD for DBE-47 vs the higher RPP slope and lower RPP SD for MCHM cover the range of variation in genomic RPP structures across the data sets analyzed (Tab. 2).

col were then compared to corresponding *RTDs* derived under the two alternative scenarios. The ratio between the maximum and minimum *RTD* across all three scenarios, considering point estimates as well as lower/upper bounds, is below 1.2 in all cases (Tab. 3), and the mean/median ratio across data sets and RTRs is 1.07/1.06. Thus, the method is stable with regard to the assessed variations in settings.

For DE-71 the genomic RPP was contrasted to that based on apical data from a long-term NTP study. As shown in Figure 7, the genomic BMDs associated with exposure from gestation day 6 through postnatal day 22 matches the apical BMDs well so that a single RPP could be used to describe both sets of data in a combined analysis. Here, the dose-dependent rank approach was applied for both types of data. It can be noted that the resulting S -values for the apical BMDs (given in the text to Fig. 7) describe a rank order quite similar to the severity-based ranking in Sand et al. (2018) of these types of liver effects (except for hepatocellular adenoma) described in Figure 1.

To address potentials for further development, Table 4 presents results from a comparison of GO categories using the web application from Verschaffelt et al. (2021). As shown, semantic similarity across the GO terms appears to be dose-dependent, i.e., as the dose separation between groups of GO terms increases, the similarity score decreases. The degree of dose-dependence appears to differ across chemicals and data sets and may depend on which GO terms were affected by the exposure (and for which BMDs could be derived). As noted in Section 2.3, individual BMDs are ranked according to $f(xm)$, and in this process they are distributed similarly across S for the different GO categories. As a secondary step, the default rank could potentially be refined using the type of information presented in Table 4. For example, S_{ijk} -values within GO categories with dose location (log mean BMD) below the central RPP curve may be weighted downwards, while S_{ijk} -values within GO categories with dose location above the central RPP curve may be weighted upwards, guided by the strength of the dose-dependence indicated in Table 4.

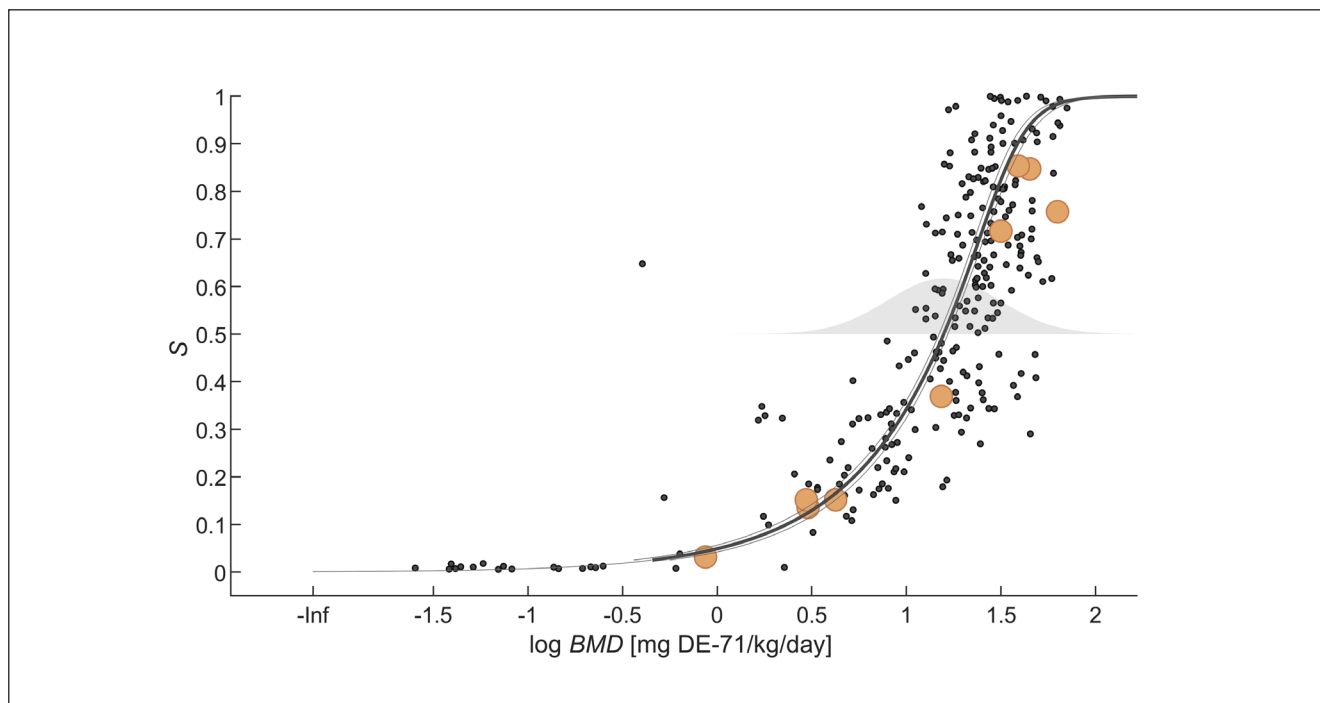


Fig. 7: Genomic and apical reference point profile (RPP) for DE-71 based on BMDs derived from neonatal rat liver transcriptomic data from a short-term study (small circles) and data from a long-term NTP study (large circles) in the same rat species

Both sets of data could be described using common RPP model parameters, i.e., using group-specific parameters did not result in a significantly better fit. Circles correspond to the iteration of the approach that is closest to the RPP point estimate (based on $n = 500$ iterations) illustrated by the solid curve and associated normal distribution. The point estimates of \hat{H}_i , $\hat{\lambda}_i$, and $\hat{\sigma}$ are 35, 4.2, and 0.68, respectively (the RPP asymmetry term, v , is set to 0.2). The dose-dependent rank approach for derivation of S -values has been applied for both the genomic and apical data (see Section 2.3). Nine apical BMDs describing liver lesions (and passing selection criteria), mainly in female rats, were included, i.e., hepatocyte hypertrophy ($S \approx 0.032$, male), fatty change ($S \approx 0.14$, female), eosinophilic focus ($S \approx 0.15$, male and female), hepatocellular adenoma ($S \approx 0.36$, female), oval cell hyperplasia ($S \approx 0.71$, female), nodular hyperplasia ($S \approx 0.76$, female), hepatocellular carcinoma ($S \approx 0.85$, female), and hepatoblastoma ($S \approx 0.85$, female).

4 Discussion

The method introduced in Sand et al. (2018) and further developed in the present paper characterizes the sequence of BMDs/dose-response curves for toxicological health effects or measures of bioactivity described by the RPP (Fig. 1) or the complete dose-severity-response volume. This enables toxicological responses and their dose equivalents to be derived by combination of dose-response information from multiple effects. The shape (λ) and variability (σ) of the RPP extend the dimensions of quantitative risk assessment beyond the use of a potency measure, represented by the RPP location (H) in the proposed method, or by a single RP (or PoD) under a more traditional approach.

Differences in the RPP shape and/or SD across chemicals can modify conclusions regarding toxicity. Generally speaking, RPs based on individual apical health effects might be associated with a low or high risk (be conservative or non-conservative) under the proposed method. This depends on the RPP parameters in combination with how the RP is aligned to the RPP, e.g., whether it is close or far away from BMDs for other effects, and whether the critical health effect represented is associated with a low

or high severity. In the comparison of TCDD and PCB₁₁₈ using traditional toxicity data (Fig. 4), the BMD for EROD activity is located to the left and to the right of the solid curve at the central S -value ($S = 0.0602$) at C2, respectively. Differences in the location of BMDs for EROD, PROD, and A4H in relation to the respective RPP for the two chemicals explain the discrepancy between RTR results across health effects (Tab. 1).

The results in Figure 4 and Table 1 also indicate that the concept of what change in response may be non-adverse/adverse for individual health effects, which is part of some guidance for how to define the BMD (e.g., EFSA, 2017), may be problematic when considering effects in a joint context. Intuitively, the response level (considering a normalized response scale), regarded as the breaking point between acceptability and adversity for individual health effects, should decrease across severity categories C1 to C9, i.e., as health effects become more severe. However, since the BMD for a specific health effect can correspond to different RTRs (Fig. 4, Tab. 1), critical response values defining the BMD may need to vary across chemicals for the same health effect to result in BMDs that are associated with equivalent RTRs. In the present examples (Tab. 1), the bench-



mark response for any of the liver enzyme parameters would need to be at least 40% to give a BMD for TCDD associated with an RTR/2D-RTR (data not shown), which is similar to the RTR/2D-RTR associated with the BMD for PCB₁₁₈ that corresponds to a 21% benchmark response. Thus, if the separation of a set of health effects across the dose continuum depends on the type of chemical, whether a certain benchmark response level is to be considered low or high becomes a chemical-specific as well as an endpoint-specific issue.

From a more general viewpoint, the joint consideration of multiple effects in line with the proposed method could improve the basis for risk management/prioritization since it allows for assessing the consequences of exposures above the health-based guidance value (HBGV), or similar, in a standardized manner. Exceedance of the HBGV for estimated exposures or defined exposure scenarios for some or several population groups is quite common. For example, this was noted in several risk assessments by the EFSA CONTAM panel in recent years (EFSA, 2020). To allow for a better differentiation of health concerns, a number of *RTDs* corresponding to specific RTRs can be derived, and exposures may then be categorized in terms of the extent they fall below/above these reference values. The separation of such *RTDs* will depend on the shape and SD of the RPP. Considering all six compounds/mixtures evaluated in the previous study, the range of *RTDs* (for standardized RTRs of 0.002 to 0.5 according to a Hill RPP model) covers a factor 12 (PCB₁₁₈) to a factor 50 (three-component mixture) (Sand et al., 2018).

When applied to apical health effects, the proposed method may appear data demanding. The earlier analysis in Sand et al. (2018) represents a data-rich example that allowed for most severity categories to be populated. Removing the top categories (above C6) for these data, however, did not affect the *RTD* much (Fig. 5). An overview of NTP long-term studies in rodents over the last decade shows that the average/median number of non-neoplastic effects for which dose-response data is available in a given study is 10/8 for a specific species and sex (NTP, 2020). The corresponding average/median is 29/19, considering all the available data (rats and mice, males and females) that may be used in combined analyses. For most of these NTP studies, data on neoplastic effects are also available. This indicates a possibility to consider the proposed method for apical effects more broadly, provided that significant dose-response trends are apparent to enable derivation of BMDs and that effects can be differentiated into a number of severity categories.

An extension of the method was introduced to address the possibility to apply the proposed concept to data from NAMs. While not investigated in detail herein, it may also be further studied if the type of dose-dependent, rather than severity-based, initial ranking approach used for genomic data can benefit the method more generally (since this might be more practical). In relation to NAMs, the National Research Council (NRC, 2007) report, aimed at developing a long-range strategic plan to modernize toxicity testing, envisioned a shift from the traditional animal-based system to using human cells or cell lines *in vitro* and computational modeling. More recently, the interagency Tox21 consortium has developed a new strategic and opera-

tional plan with a broader focus (Thomas et al., 2018). This includes consideration of high-throughput transcriptomics (HTT) as a timely and cost-effective screening approach, and the use of short-term studies in rodents in this context is regarded to serve as a bridge between *in vitro* approaches and traditional guideline toxicity studies (Gwinn et al., 2020).

The interest in the short-term study (i.e., the 5-day assay) relates to the observation that gene set BMDs correlate reasonably well with potency estimates traditionally used as PoDs (i.e., BMDLs/NOAELs for the most sensitive apical effects) from cancer- and non-cancer toxicity studies in the same species (e.g., Thomas et al., 2013, NTP, 2018, Gwinn et al., 2020). The NTP (2018) approach to genomic dose-response analysis applies to both *in vivo* and *in vitro* studies, but the latter is less investigated and would also require the additional element of *vitro*-to-*in vivo* extrapolation as a crucial element. Understanding relations between quantitative risk estimates at the genetic and apical levels *in vivo*, however, appears to be a relevant step in the process of developing a framework based on gene-level effects *in vitro*. As noted above, this provides a bridge to facilitate understanding of the different data streams.

The NTP (2018) approach defines gene-level potency in terms of the median BMD, BMDL, and BMDU, providing a central estimate for each gene set. This type of summary measure, herein defined in terms of the mean/median log BMD over all gene sets included in a particular analysis, corresponds to the variability in the genomic RPP at a given *S*-value. In line with earlier discussions, accounting for BMD variability, in this case both within and across GO terms described by the genomic RPP, allows for a more refined differentiation of potency compared to using a summary measure, as shown by the variation in both λ and σ across the studied chemicals (Tab. 2, Fig. 2, 6, 7). This is also indicated by the evaluation of RTRs associated with the lowest mean/median log BMD (Tab. 3), providing a picture similar to that for apical effects (Tab. 1), i.e., that joint analysis of multiple effects/BMDs, in this case across GO categories, can modify conclusions on potency.

Results in this study also show that combination of several BMDs under the proposed method may allow for a less uncertain derivation of exposure guidelines/PoDs. This was to some extent pointed out in the introduction of the concept in Sand et al. (2018), but herein more specifically evaluated across data sets focusing on the transcriptional level. The increase in dose-response information that follows when moving away from the apical response provides an opportunity for determination of less uncertain PoDs by combination of gene set BMDs as suggested. For the evaluated studies, the uncertainty in *RTD* was clearly lower than that associated with the lowest mean/median log BMD, also accounting for RPP model uncertainty (Tab. 3).

NTP (2018) notes that it may be further evaluated whether alternative methods for summarizing gene set BMDs, including consideration of potency rank, could provide better surrogates for apical PoDs. Genomic and apical RPPs for DE71 were compared in this work, and this illustrates how a more complete assessment of the relation between the two types of data can be performed (Fig. 7). The new strategic plan within Tox21 includes

making better use of legacy *in vivo* toxicity data in the process of moving towards using NAMs for chemical risk assessment. Traditional toxicity data is regarded to provide a rich resource that, for example, can help to link the effects observed at the molecular level to those at the tissue-, organ-, and organism-level, also characterizing how the variability in traditional toxicity studies may differ from *in vivo* testing approaches (Thomas et al., 2018). Further comparison of genomic and apical RPPs may help as part of such analysis. In the example for DE-71, the transcriptional and apical levels were adequately characterized by a single RPP (Fig. 7). While the default set-up/rank approach in this case provided similar RPPs, differences may occur in other cases. Therefore, modification of the severity weight, w , may also be an integral part of this type of analysis. Differentiation of the severity weight across data types may then help to provide matching *RTDs*. The introduction of some default uncertainty in the severity weight based on a broader assessment of this issue might help to provide *RTDs* that better encapsulate the corresponding doses associated with apical response.

Knowledge of relations between toxicity pathways may also inform the ranking and/or severity weighting. As a step in this direction, it was illustrated how derivation of the genomic RPP could be supported by approaches for assessing semantic similarities between GO terms using the web application by Verschaffelt et al. (2021) as an example. Interestingly, the similarity score across the genomic RPP was dose dependent (Tab. 4). This observation is based on the present data and the particular approach used for comparing similarity. Information provided at the GitHub repository states¹ that scores < 0.3 , $0.3-0.9$, and > 0.9 indicate “not functionally similar”, “functionally related”, and “highly similar functions”, respectively. Thus, GO terms part of genomic RPPs in this paper appear to be “functionally related” (to various degrees). This may need to be further assessed under a broader and/or more detailed analysis, including how to technically integrate the type of results in Table 4 as part of the overall method.

The consequence of “tilting” the default rank, as exemplified in Section 3, appears mainly to have the effect of reducing the RPP SD, while the RPP location and slope will be less affected. Theoretically, a reduction in the RPP SD implies that a change in impacts/RTRs as a result of a given change in exposure from A to B will increase. Thus, as the relative distance metric between GO terms increases (smaller similarity score), so would a change in the RTR. Also, using some overall measure of similarity, e.g., a score associated with the central region of the RPP, may possibly support parameterization of the systemic weight function, $w(S)$, in Equation 3 that indirectly affects all categories. For example, a stronger overall relation between GO terms may indicate a higher specificity/significance of the exposure that intuitively may indicate a higher concern/severity. If so, this could motivate the use of a more conservative systemic weight.

The principle of finding exposures below which no significant perturbation of toxicity pathways is observed, in line with the

NRC (2007) vision, appears conceptually similar to the objective within the current paradigm that sets an RP/PoD based on the critical effect, which is ideally defined as the first adverse effect or known precursor (US EPA, 2002). However, moving towards using data from NAMs for risk assessment may be more in line with a data-driven rather than knowledge-driven approach. As discussed in Whelan and Andersen (2013), mechanistic description of the underlying biological system and details on how pathway perturbation leads to adversity might not be absolute prerequisites. Overall, moving away from the apical response may to a higher extent require/emphasize that probability becomes part of the consideration of what constitutes a significant alteration, e.g., in gene expression within the network of toxicity pathways, while the current paradigm considers this from a more absolute standpoint.

The proposed method attempts to be in line with a probability-based framework. Under the default rank, the probability, p , across S may serve as a quantitative description of the extent of perturbation across GO categories in terms of how many BMDs (relatively speaking) are exceeded within and across the categories. Introduction of a non-linear weight, $w(S)$, might then allow this to be severity-adjusted, as noted in Section 2. In its current form, the method allows for the derivation of reference values, *RTDs*, that are standardized in this regard. However, the question of what would constitute an acceptable change (e.g., acceptable RTR) remains, and, generally speaking, this issue will likely pose challenges under a NAM-based framework as it does within traditional risk assessment. As noted earlier, the use of several exposure guidelines that better allow for consideration of the continuum of risk can, however, help in the assessment of chemical exposure. A system that enables evaluation of the gradual increase in the total amount of effect/biological activation might be particularly useful if risk assessment is informed by measures, e.g., at the genomic level, that are more numerous and less directly related to disease compared to apical effects used within the current approach.

5 Conclusion

Systematic combination of data by characterization of the dose-related severity sequence or the sequence of BMDs across gene sets condenses information of the chemical effect domain into a small set of parameters with toxicological interpretation. Based on the previously proposed concept, an extension to genomic dose-response information was developed. Results indicated a variation in RPP shapes and SDs across chemicals, suggesting, as for the apical response, that the method differentiates the consequence of chemical exposure to a higher extent compared to standard approaches. This can help to refine establishment of RPs/PoDs or sets of such values describing various levels of health concerns that, e.g., would permit assessment of risks/impacts at exposures above the traditional human exposure

¹ <https://github.com/MEGA-GO/MegaGO> (accessed 19.06.2021)



guideline. The standardized scale used for grading health effects or measures of bioactivity facilitates comparative analysis across individual chemicals in contrast to the current method, and the attached severity weighting approach allows for summarizing/integrating contributions across multiple outcomes, enriching the quantitative risk assessment metric. Further comparison of apical and genomic RPPs can help to improve understanding of different data streams to facilitate transition to a NAM-based risk assessment paradigm. In this process, analysis at the genomic level may potentially be advanced by considering functional relations between gene sets to refine the ranking of bioactivity and parameterization of the systemic weight function. The proposed concept supports the use of an increasing amount of effect parameters that may result when focus is shifted upstream from adverse apical response. This may more strongly promote or require methods for combination of data and evaluation of the overall impact compared to relying on specific observations as within traditional risk assessment.

References

- Chen, C. and Chen J. J. (2014). Benchmark dose calculation for ordered categorical responses. *Risk Anal* 17, 1435-1447. doi: 10.1111/risa.12167
- Crump, K. (1984). A new method for determining allowable daily intakes. *Fundam Appl Toxicol* 4, 854-871. doi:10.1093/toxsci/4.5.854
- Crump, K. (1995). Calculation of benchmark doses from continuous data. *Risk Anal* 15, 79-89. doi:10.1111/j.1539-6924.1995.tb00095.x
- Dourson, M. L., Hertzberg, R. C., Hartung, R. et al. (1985). Novel approaches for the estimation of acceptable daily intake. *Toxicol Ind Health* 1, 23-41. doi:10.1177/074823378500100404
- Dourson, M. L., Teuschler, L. K., Durkin, P. R. et al. (1997). Categorical regression of toxicity data: A case study using aldicarb. *Regul Toxicol Pharmacol* 25, 121-129. doi:10.1006/rtph.1996.1079
- Dunnick, J. K., Shockley, K. R., Pandiri, A. R. et al. (2018). PBDE-47 and PBDE mixture (DE-71) toxicities and liver transcriptomic changes at PND 22 after in utero/postnatal exposure in the rat. *Arch Toxicol* 92, 3415-3433. doi:10.1007/s00204-018-2292-y
- Dunnick, J. K., Shockley, K. R., Morgan, D. L. et al. (2020). Hepatic transcriptomic patterns in the neonatal rat after pentabromodiphenyl ether (PBDE) exposure. *Toxicol Pathol* 48, 338-349. doi:10.1177/0192623319888433
- 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. doi:10.2903/j.efsa.2005.282
- EFSA (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, 1150. doi:10.2903/j.efsa.2009.1150
- EFSA (2017). Update: Use of the benchmark dose approach in risk assessment. *EFSA J* 15, 4658. doi:10.2903/j.efsa.2017.4658
- EFSA (2020). Overview of full risk assessments by the EFSA CONTAM panel between 2017 and 2020. <https://www.efsa.europa.eu/en> (accessed 09.07.2020)
- Guth, D. J., Carroll, R. J., Simpson, D. G. et al. (1997). Categorical regression analysis of acute exposure to tetrachloroethylene. *Risk Anal* 17, 321-332. doi:10.1111/j.1539-6924.1997.tb00870.x
- Gwinn, W. M., Auerbach, S. S., Parham, F. et al. (2020). Evaluation of 5-day in vivo rat liver and kidney with high-throughput transcriptomics for estimating benchmark doses of apical outcomes. *Tox Sci* 176, 343-354. doi:10.22427/ntp-data-002-00058-0002-0000-7
- Haber, L. T., Dourson, M. L., Allen, B. C. et al. (2018). Benchmark dose (BMD) modeling: Current practice, issues, and challenges. *Crit Rev Toxicol* 48, 387-415. doi:10.1080/10408444.2018.1430121
- Hertzberg, R. C. and Miller, M. (1985). A statistical model for species extrapolation using categorical response data. *Tox Ind Health* 1, 43-57. doi:10.1177/074823378500100405
- Hertzberg, R. C. (1989). Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Phys* 57, 405-409. doi:10.1097/00004032-198907001-00057
- 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 B. M. Ayyub (ed.), *Proceedings, Second International Symposium on Uncertainty Modeling and Analysis, College Park, MD, USA* (254-261). Los Alamitos, CA, USA: IEEE Computer Society Press. doi:10.1109/isuma.1993.366760
- Lin, D. (1998). An information-theoretic definition of similarity. *Proceedings of the 15th International Conference on Machine Learning*, 296-304.
- Milton, B., Farrell, P. J., Birkett, N. et al. (2017a). Modeling u-shaped exposure-response relationships for agents that demonstrate toxicity due to both excess and deficiency. *Risk Anal* 37, 265-279. doi:10.1111/risa.12603
- Milton, B., Krewski, D., Mattison, D. R. et al. (2017b). Modeling U-shaped dose-response curves for manganese using categorical regression. *Neurotoxicology* 58, 217-225. doi:10.1016/j.neuro.2016.10.001
- NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: National Academy of Sciences.
- 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. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr521.pdf
- NTP (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. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr525.pdf
- NTP (2006c). 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. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr521.pdf

- genesis studies of 3,3',4,4',5-pentachlorobiphenyl (PCB₁₂₆) (CAS No. 57465-28-8) in female Harlan Sprague-Dawley rats (Gavage Studies). *Natl Toxicol Program Tech Rep Ser 520*. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr520.pdf
- NTP (2006d). NTP technical report on the toxicology and carcinogenesis studies of a binary mixture of 3,3',4,4',5-pentachlorobiphenyl (PCB₁₂₆) (CAS No. 57465-28-8) and 2,3',4,4',5-pentachlorobiphenyl (PCB₁₁₈) (CAS No. 31508-00-6) in female Harlan Sprague-Dawley rats (Gavage Studies). *Natl Toxicol Program Tech Rep Ser 531*. https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr531.pdf
- NTP (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 (PCB₁₂₆) (CAS No. 57465-28-8) in female Harlan Sprague-Dawley rats (Gavage Studies). *Natl Toxicol Program Tech Rep Ser 526*. <https://ntp.niehs.nih.gov/go/tr526>
- NTP (2010). NTP technical report on the toxicology and carcinogenesis studies of 2,3',4,4',5-pentachlorobiphenyl (PCB₁₁₈) (CAS No. 31508-00-6) in female Harlan Sprague-Dawley rats (Gavage Studies). *Natl Toxicol Program Tech Rep Ser 559*. doi:10.22427/ntp-tox-77
- NTP (2016). NTP technical report on the toxicology studies of a pentabromodiphenyl ether mixture [DE-71 (technical grade)] (CAS No. 32534-81-9) in F344/N rats and B6C3F1/N mice and toxicology and carcinogenesis studies of a pentabromodiphenyl ether mixture [DE-71 (technical grade)] in Wistar Han [CrI: WI(Han)] rats and B6C3F1/N mice. *Natl Toxicol Program Tech Rep Ser 589*. doi:10.22427/ntp-tr-589
- NTP (2018). NTP research report on National Toxicology Program approach to genomic dose-response modeling. *Research Report 5*. doi:10.22427/ntp-rr-5
- NTP (2020). Overview of NTP technical report between 2010 and 2020. <https://ntp.niehs.nih.gov> (accessed 10.07.2020).
- NTP (2021). BMD results from 5-day toxicogenomic studies in rodents for triphenyl phosphate (TPHP), 4-methylcyclohexanemethanol (MCHM), and propylene glycol phenyl ether (PPH). Available at: <https://cebs.niehs.nih.gov/tgxbenchmark/doseresponse/> (accessed Jan. 2021).
- 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. doi:10.1111/1539-6924.00381
- Sand, S., von Rosen, D., Victorin, K. et al. (2006). Identification of a critical dose level for risk assessment: Developments in benchmark dose analysis of continuous endpoints. *Toxicol Sci* 90, 241-251. doi:10.1093/toxsci/kfj057
- Sand, S., Ringblom, J., Håkansson, H. et al. (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. doi:10.1002/jat.2757
- Sand, S., Lindqvist, R., von Rosen, D. et al. (2018). Dose-related severity sequence, and risk-based integration, of chemically induced health effect. *Toxicol Sci* 165, 74-89. doi:10.1093/toxsci/kfy124
- Teuschler, L. K., Dourson, M. L., Stiteler, W. M. et al. (1999). Health risk above the reference dose for multiple chemicals. *Regul Toxicol Pharmacol* 30, S19-S26. doi:10.1006/rtp.1999.1321
- Thomas, R. S., Wesselkamper, S. C., Wang, N. C. Y. et al. (2013). Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol Sci* 134, 180-194. doi:10.1093/toxsci/kft094
- Thomas, R. S., Paules, R. S., Simeonov, A. et al. (2018.) The US Federal Tox21 Program: A strategic and operational plan for continued leadership. *ALTEX* 35, 163-168. doi:10.14573/altex.1803011
- US 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.
- US EPA (2005). Guidelines for carcinogen risk assessment. Final report. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.
- US EPA (2012). Benchmark dose technical guidance. Risk Assessment Forum. EPA/100/R-12/001. Washington, DC.
- Verschaffelt, P., Van Den Bossche, T., Gabriel, W. et al. (2021). MegaGO: A fast yet powerful approach to assess functional gene ontology similarity across meta-omics data sets. *J Proteome Res* 20, 2083-2088. doi:10.1021/acs.jproteome.0c00926
- Whelan, M. and Andersen, M. (2013). Toxicity Pathways – From concepts to application in chemical safety assessment. EUR 26389. Luxembourg (Luxembourg): Publications Office of the European Union; 2013. JRC86467.

Conflict of interest

The author declares there are no conflicts of interest.

Acknowledgements

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