Dose-Response:
An International Journal
April-June 2020:1-10

DOI: 10.1177/1559325820919605
journals.sagepub.com/home/dos
SAGE

Toxic Responses Induced at High Doses May Affect Benchmark Doses

Jürg A. Zarn[1], Ursina A. Zürcher[1], and H. Christoph Geiser[1]

Abstract

To derive reference points (RPs) for health-based guidance values, the benchmark dose (BMD) approach increasingly replaces the no-observed-adverse-effect level approach. In the BMD approach, the RP corresponds to the benchmark dose lower confidence bounds (BMDLs) of a mathematical dose–response model derived from responses of animals over the entire dose range applied. The use of the entire dose range is seen as an important advantage of the BMD approach. This assumes that responses over the entire dose range are relevant for modeling low-dose responses, the basis for the RP. However, if part of the high-dose response was unnoticed triggered by a mechanism of action (MOA) that does not work at low doses, the high-dose response distorts the modeling of low-dose responses. Hence, we investigated the effect of high-dose specific responses on BMDLs by assuming a low- and a high-dose MOA. The BMDLs resulting from modeling fictitious quantal data were scattered over a broad dose range overlapping with the toxic range. Hence, BMDLs are sensitive to high-dose responses even though they might be irrelevant to low-dose response modeling. When applying the BMD approach, care should be taken that high-dose specific responses do not unduly affect the BMDL that derives from low doses.

Keywords

risk assessment, threshold, modeling, dose–response, benchmark dose

Introduction

No-Observed-Adverse-Effect Level and Benchmark Dose Approach

Animal toxicity data are typically used to derive human health-based guidance values (HBGVs) such as the acceptable daily intake (ADI) of a substance. For decades, the HBGVs have been based on no-observed-adverse-effect levels (NOAELs).[1] The NOAEL is the highest dose administered to study animals that is not associated with any significant and/or biologically relevant adverse response in any of the investigated endpoints. The subsequent higher dose is termed the lowest observed adverse effect level (LOAEL), which is the lowest dose that induces treatment-related adverse responses. Throughout this article, every pathological observation is referred to as a response, regardless of whether it was considered chemical treatment related or not. Therefore, pathological observations in the control group are also termed as responses, although they are not chemical treatment related.

The NOAEL approach undoubtedly has shortcomings that have been extensively reviewed and contrasted with advantages of an alternative procedure, which is now known as the benchmark dose (BMD) approach.[2-8] The BMD approach principles were initially introduced in 1984.[9] Based on the administered doses and observed responses, a dose–response curve is derived by fitting a mathematical model to the experimental data. For continuous response data, a predefined change relative to the modeled response at dose zero is defined as the critical benchmark response (BMR), typically at 5% (ie, 5% additional risk). For quantal response data, the BMR is typically set at 10% and is defined as an extra risk, that is, a 10% increase adjusted for the background incidence relative to the modeled nonresponding proportion at dose zero. Importantly, the background response at dose zero for both continuous and quantal data is defined by the model and not by the observed response. The BMR values are typically set at 5% for

[1] Risk Assessment Division, Federal Food Safety and Veterinary Office (FSVO), Bern, Switzerland

Received 12 December 2019; received revised 27 February 2020; accepted 24 March 2020

Corresponding Author:
Jürg A. Zarn, Risk Assessment Division, Federal Food Safety and Veterinary Office (FSVO), Schwarzenburgstrasse 155, 3003 Bern, Switzerland.
Email: juerg.zarn@blv.admin.ch

continuous data and 10% for quantal data because these response sizes are commonly triggered by doses within or close to the experimental dose range. For both continuous and quantal data, the model's lower bound of the 95% confidence interval of the dose (benchmark dose lower confidence bound, BMDL), which provides the predefined BMR, is used as the reference point (RP) to derive an HBGV. Preferably, each of the many endpoints investigated in all available studies with the compound of interest should be subjected to BMD modeling to identify the lowest biologically relevant RP. However, this is currently not feasible due to limitations in the available software technology and the study data formats used for submission to the regulatory authorities. Thus, visual inspections and/or statistical preanalyses are conducted to preselect endpoints to be used in BMD modeling.[2] Once a critical endpoint is identified and the corresponding RP determined, the procedure to derive an HBGV via the BMD approach is similar to that via the NOAEL approach. In both approaches, the RP values are divided by an appropriate safety or uncertainty factor (both terms are used interchangeably) to extrapolate the HBGV considered protective for human health, for example, the ADI, from the animal data. The uncertainty factors should take into account differences in sensitivity in the human population to be protected and differences in sensitivity between the human population and the animal strain under study.

The BMD modeling methodologies have been recently reviewed,[2,3] and suggestions for method development and areas of application have been made.[5,10-12]

Subpopulations in Dose Groups

A key paradigm of biological experimentation is that any response, be it treatment related or not, arises from an underlying biological mechanism of action (MOA). This is based on the tenet that nothing happens without reason, that is, every change of state in a system must necessarily have a cause, regardless of whether it is experimentally verifiable or not. Although such a deterministic causality in biology is not proven, it is widely accepted as philosophical basis of all scientific experimentation and emerged from hundreds of years of epistemological philosophy.

The responses observed in toxicological experimentation fall into 2 different categories: first, background responses are triggered by chemical treatment-unrelated animal-inherent biological factors alone or by their interactions with the control environment; second, the background response as defined above is overlaid by treatment-related responses that are triggered by interactions of animal-inherent biological factors with a substance administered to the animals.

Each batch of laboratory animals purchased from a breeder for toxicity testing is almost always composed of subpopulations that differ in certain genetically encoded biological factors, including factors related to toxicokinetics and toxicodynamics.[13-16] The existence of genetically different subpopulations within the animals comprising the animal pool of a particular study becomes obvious by the following 2 observations, which are made in most toxicological studies. First, for many responses, the incidence in the control group is greater than zero but less than 100%. This means that in addition to the subpopulation bearing the chemical-specific MOA, at least one other subpopulation exists bearing an MOA that acts completely independent of the test chemical. Second, even for the highest dose which may exceed the LOAEL by orders of magnitude, usually not all the animals respond. In principle, these everyday observations of different response rates at different doses alone indicate the presence of at least 3 differently sensitive subpopulations in almost every toxicological study: a completely insensitive subpopulation, a subpopulation bearing a treatment-independent MOA, and a subpopulation bearing a treatment-sensitive MOA. It is quite conceivable that not only 1 but sometimes 2 or more MOAs that are sensitive to treatment will act and that they may also be overlaid by MOAs that are indirectly sensitive to treatment. As an example of the evidence of the existence of indirectly treatment-sensitive MOAs, the influence of body weight on various responses should be mentioned here.[17,18] For the interpretation of a dose–response curve, this means that the gross response at any given dose can be composed of subpopulation-specific responses. Since at different doses, different subpopulations may be sensitive, the dose–response curve mainly reflects the composition of the subpopulations. This leads to the conclusion that the use of responses at high doses can be not only unimportant for the response modeling at low doses but even distorting, since different MOAs are active.

At least 80% of rat pesticide chronic toxicity/carcinogenicity studies are performed with outbred stocks and at least 15% with inbred strains, whereas approximately 5% of the studies are conducted with strains that were not sufficiently specified.[19] The subpopulations may display subpopulation-specific susceptibilities toward environmental stress and specific treatment. Regarding a quantal response of interest, each animal in a dose group either responds or not and therefore the dose–response curve in principle is a step curve. In a theoretical study with an infinite number of dose groups, the number of steps observed over the entire dose–response curve should then reflect the number of susceptibility subpopulations. Thought through consistently and extrapolated to the individual level, this means that each animal, in principle, can have its own specific sensitivity. The resulting dose–response curve therefore has as many levels as the animals involved in the study.[20] Whether the total dose–response curve shows an apparently continuous or indeed a step curve, therefore, depends on the genetic homogeneity of the animals tested, that is, the number of susceptibility subpopulations, as well as on the homogeneity of the environment the animals are exposed to. Within an ideal population of completely identical animals (clones) in a fully characterized environment, that is, study design variability-related randomness is precluded, individual animals would respond identically toward environmental stress including a specific chemical treatment. If these animals are exposed to a dose below the population-specific dose threshold for the relevant MOA, none of the animals would respond,

whereas all the animals would respond at doses above the threshold. This results in a step curve. Although dose–response curves of actual experiments generally appear to be continuous, there are very good reasons to assume that there are underlying step curves representing different sensitivity subgroups as already discussed by other authors.[16,21-23]

Hence, continuous dose–response curves essentially manifest the presence of subpopulations and/or variability in environmental factors. Since bodyweight is the only stratification factor while randomizing the animals into different treatment groups, the subpopulations are most likely not evenly distributed among the dose groups. Subpopulations that are unevenly distributed between the dose groups, amplified by uncontrollable environmental factors, induce a dispersion of responses, which usually is subsumed under terms such as tolerance or susceptibility distribution. Unevenly distributed subpopulations during the randomization process may explain at least partly the observation of up and down fluctuating responses along the dose range tested and the often insufficient study-to-study reproducibility of results.

From the above considerations, in a study with completely identical animals (clones) in an environment completely identical for each animal, a dose–response curve can only be a step curve. But step curves may remain unrecognized because of the typical study design. With the usually only 4 dose levels including a control group, no experimental information is available on the actual dose–response curve over wide dose ranges between dose groups. In order to model the possible curves, assumptions must hence be made about the basic shape of the curve. And precisely this is an eminently important point in any modeling because by deciding which models to use, others are automatically excluded. Possibly due to their training, most toxicologists spontaneously expect a continuous and monotonous (eg, sigmoidal) dose–response curve. This may be one of the reasons why step curves, nonmonotonous (eg, hormetic) *U*-shaped[24-26] or inverted *U*-shaped dose–response curves are presumably more or less unconsciously not considered at all. However, for the reasons given above, it should not be taken for granted that any dose–response relationship must be continuous and monotonous.

Scope of the Present Study

Most publications about the BMD approach focus on the mathematical/statistical aspects of modeling[27] and the study design effects such as group sizes.[28] In this study, we want to critically discuss BMD modeling in the context of potentially different MOAs associated with different subpopulations inadvertently included in the study contributing to the observed total response of interest.

We assessed the claim that the scientific superiority of the BMD approach over the NOAEL approach derives from the fact that the former includes all responses observed within the tested dose range. This approach implies that the response at all dose levels contains relevant information for modeling the critical low-dose response. The present analysis assumes that the purchased animal batch included 4 subpopulations distinguishable in their responsiveness regarding a given quantal endpoint: one subpopulation with a chemical-independent MOA (responsible for the background response), 2 subpopulations with distinct chemical-dependent MOAs associated with different dose thresholds, and individuals of a fourth subpopulation that were not responsive within the tested dose range.

This article does not compare the 2 approaches (NOAEL and BMD) in the totality of all their aspects in order to arrive at an overall judgment on which is the better one.

Method

In our analysis using fictitious dose–response data, we assumed that a batch of laboratory animals of the same stock, which was purchased from a breeder for a multidose toxicity study, consisted of 4 subpopulations—A, B, C, and D. Under real conditions, the assumption of completely equal subpopulation proportions distributed among dose groups will be probably never met. Thus, disproportional subpopulation allocation to dose groups as a source of randomness may seriously bias dose–response curves. However, due to the additional complexity of this aspect, unequal subpopulation proportions among dose groups were not considered in this analysis. Effects of disproportional subpopulation allocation to dose groups during the randomization process may be investigated in future analyses. Instead, for the sake of clarity of the analyses, we assumed that the proportions of subpopulations A, B, C, and D were equal in all dose groups of the toxicity study. Each subpopulation possessed a specific set of inherent biological factors with MOA_A assigned to subpopulation A, MOA_B to subpopulation B, and MOA_C to subpopulation C. Importantly, the 3 different MOAs affected the same apical quantal response. Animals not belonging to subpopulation A, B, or C were assigned to the nonresponsive subpopulation D. The sum of the fractions of subpopulations A, B, C, and D was equal to 100% for the entire animal batch (scheme in Figure 1A).

For the chemical-dependent MOAs, the quantal responses were triggered when applied doses exceeded the threshold dose of the respective MOA. Thus, the proportion of responding animals in a dose group corresponded to the proportions of the respective subpopulations (A, B, C, or D) in the dose group. It was assumed that subpopulation A animals always responded via the chemical-independent MOA_A, yielding the background response, whereas animals with the chemical-dependent MOA_B or MOA_C had the threshold at categorical dose level 2 or 9, respectively (Figure 1B). At dose levels <2, including the control group, the total response in a dose group corresponded to the proportion of subpopulation A. At dose levels $\geq$2 and <9, the total response in a dose group corresponded to the sum of the proportions of subpopulations A and B, and at dose levels $\geq$9, it corresponded to the sum of the proportions of subpopulations A, B, and C in the dose groups. In summary, all dose groups of a study were composed of equal proportions of chemical-independently responding subpopulation A, chemical-dependently responding subpopulations B and C, and the never responding subpopulation D.

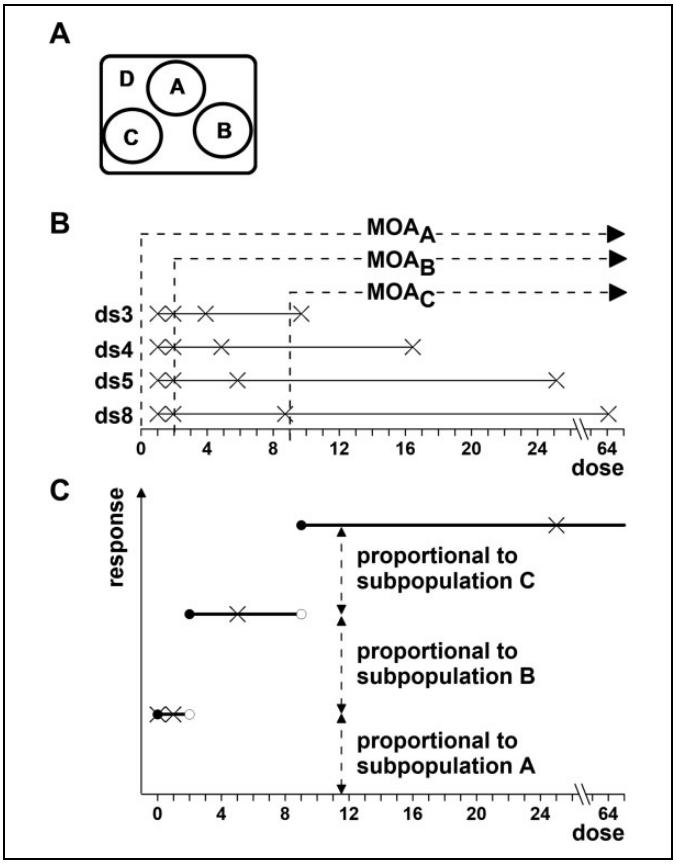

Figure 1. The fictitious study analysis included subpopulation-specific thresholds for MOAs, study designs, and dose–response graph. A, Schematic presentation of the proportions of subpopulations A, B, C, and D in a batch of laboratory animals. B, The dose thresholds from which MOA_A (background response), MOA_B (dose 2), and MOA_C (dose 9) start to operate in the virtual batch of animals composed of the 3 subpopulations A, B, and C are indicated by right-angled dashed arrows. The subpopulations are defined by their capacity to respond through MOA_A, MOA_B, or MOA_C. The 4 lines with the crosses show the 4 study designs that researchers may have chosen to test the toxicity of a chemical. The length of the lines indicates the dose range that was covered and the crosses where the 4 dose levels were set. The dose spacing (abbreviated by "ds") was geometric in all designs with the respective factors 3 (ds 3), 4 (ds 4), 5 (ds 5), or 8 (ds 8). C, The gross true dose–response graph observed when increasing doses of a chemical capable to induce MOA_B or MOA_C is administered at an infinite number of dose groups. The jumps in the discontinuous response curve derive from the dose thresholds for MOA_B or MOA_C; below the threshold dose for a MOA defining a subpopulation, no animals of the respective subpopulation respond and above the threshold dose, all of the respective subpopulation respond. Hence, the extent of the jump reflects the proportion a subpopulation accounts to in the batch of animals studied. The crosses indicate the dose–response data a researcher would observe if he had applied a ds of 5 (see also panel A).

Figure 1C shows a generic dose–response graph based on the above assumptions and an infinitesimally small spacing of the dose groups (ie, an infinite number of dose groups). The graphical presentation of the dose–response relationship generated a step graph with jumps in the total response at the threshold dose levels 2 and 9. This step graph summarizes the effects of the MOAs associated with the different subpopulations, that is, 2 distinct chemical-dependent MOAs, MOA_B, and MOA_C, responsible for the jumps at dose levels 2 and 9, along with MOA_A responsible for the background response. However, in real populations, the diversity of genetic, environmental, and lifestyle conditions may generate enormous numbers of susceptibility subpopulations that are not revealed by the total dose–response curve.[21,22,29] It can be assumed that in real, genetically highly diverse populations consisting of individuals in diverse environmental conditions and genetic backgrounds with unknown thresholds for chemical-dependent responses, the true dose–response relationship cannot be described by a step function. Nevertheless, the more the genetic, environmental, and lifestyle conditions are controlled, as it is possible to a certain degree for in-bred animal strains,[14] the more the discontinuity will appear in the dose–response graphs as a consequence of subpopulation-specific MOAs.[16]

Fictitious animal toxicity data for a quantal response were generated using variable dose spacing (ds) and varying population compositions (ie, variable proportions of subpopulations A, B, C, and D) resulting in differing response sizes (Table 1). The responses in the high-dose groups were varied based on the fixed study designs (studies a-h) with background incidences of 0% (0 of 50 animals) in designs a to d and 10% (5 of 50 animals) in designs e to g and 4% (2 of 50 animals) in design h. Common characteristics of all fictitious animal studies developed are 1 control group and 3 dose groups (termed control, low, mid, and high dose), geometric ds, response in control and low-dose group is equal and set to either 0% (designs a-d) or 4% (design h) or 10% (designs e-g), representing low, medium, and high background responses. The response in the mid-dose group was increased, compared to that in the control and low-dose groups. The increase was chosen such that it constituted the smallest possible increase resulting in a response that differed significantly from the control and the low-dose group (Fisher's exact test, 1 sided, $\alpha = .05$); that is, in an evaluation by regulatory toxicologists, the mid-dose response would be probably attributed to the treatment. Regulatory toxicologists probably would consider the low dose to be the NOAEL. It is important to note that this is not a statement to the effect that a Fisher's test significant on the significance level .05 should be the sole or the best criterion for interpreting an observed difference as a treatment-related response. There are good reasons to apply other criteria as well or instead. For this theoretical work, however, we wanted to use the significant Fisher's test as an objective and transparent criterion for deciding whether a response is treatment related or not. In each of the study designs a to h, the high-dose group responses varied from being equal to the mid-dose group up to a response rate of 100%. Accordingly, the number of responding animals in the high-dose group, starting with a number equal to the number of responding animals in the mid-dose group, was increased in increments of one animal, up to 50 responding animals (100%). All resulting total response rates were BMD modeled. This procedure generated the following numbers of fictitious dose–responses: 46 dose–

Table 1. Design of Fictitious Studies Used for BMDL Calculations.

Study	Dose Spacing	Proportions of Subpopulations[a] A, B, and C in All Dose Groups	Number of Responding Animals (of 50)				Number of Studies Used for BMDL Calculations
			Control	Low Dose	Mid Dose	High Dose	
a	3	0% A, 10% B, 0→90% C	0	0	5[b]	5[b]→50[b]	46
b	4						46
c	5						46
d	8						46
e	3	10% A, 16% B, 0→74% C	5	5	13[b]	13[b]→50[b]	38
f	5						38
g	8						38
h	3	4% A, 12% B, 0→84% C	2	2	8[b]	8[b]→50[b]	43

Abbreviation: BMDL, benchmark dose lower confidence bound.

→ Indicates that the proportion of subpopulation C, and hence the high-dose response, was successively increased to investigate the effect of high-dose responses on the BMDL. Reading example: Study a has a geometric dose spacing (ds) of 3 with doses of 0 (control), 1 (low dose), 3 (mid dose), and 9 (high dose). The dose groups are composed of 0% of subpopulation A, 10% subpopulation B, and 0% of subpopulation C fraction. Therefore, the fraction of subpopulation D is 90%. In dose groups of 50 animals, this would result in 0, 0, 5 (5 animals of subpopulation B), and 5 (5 animals of subpopulation B) animals in the control, low-dose, mid-dose, and high-dose groups, respectively. The first variation of study a would be to assume again 0% of subpopulation A, 10% subpopulation B, and 2% subpopulation C (one animal) in all dose groups. Therefore, the fraction of subpopulation D would decrease to 88%. This results in 0, 0, 5 (the 5 animals of subpopulation B), and 6 (the 5 animals of subpopulation B and the single animal of subpopulation C) animals responding in the control, low-dose, mid-dose, and high-dose groups, respectively. In all further variations of study a, the subpopulation of C in all dose groups increases stepwise by 1 animal and concomitantly the subpopulation of D decreases by 1 animal. Ultimately, this leads to 0, 0, 5 (5 animals of subpopulation B), and 50 (the 5 animals of subpopulation B and the 45 animals of subpopulation C) animals in the control, low-dose, mid-dose, and high-dose groups, respectively. In total, this procedure creates 46 variations of study.

[a]Proportions of subpopulation D are not presented because they were used to add the sum of the proportions of subpopulations A, B, and C to 100%.

[b]Significant response at a significance level of 0.05 compared to the control group response in Fisher's exact test, one sided.

responses for each of the study designs a, b, c, and d; 38 for each of the study designs e, f, and g; and 43 for study h. To verify that our fictitious data sets are very typical for pesticide toxicology studies, the reader is referred to the evaluation of hundreds of toxicology studies of pesticides evaluated by the Joint Food and Agriculture Organization/World Health Organization Meeting on Pesticide Residue (JMPR).[30]

The BMD analyses were conducted to derive the BMDLs for each dose–response using the PROAST software, version 65.2, released on January 23, 2018, which was accessed with the online tool PROASTweb (https://proastweb.rivm.nl/Analysis/New). The following PROAST web settings were used: quantal analysis, benchmark response 0.1 (extra risk), model averaging, 200 bootstrap runs, AIC criterion of 2. In this analysis, the dose–response results of hypothetical animal studies with ds >8 were not included because the resulting data were extreme and therefore not considered.

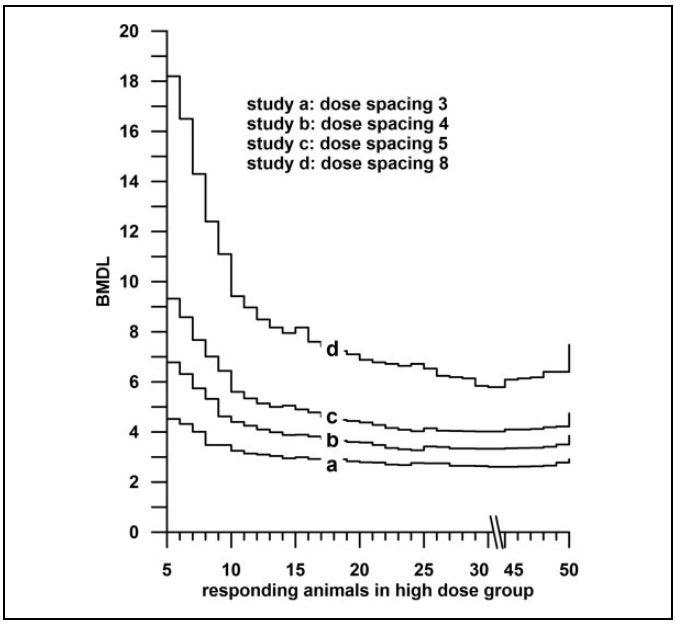

Figure 2. BMDL distributions of studies with background responses of 0% and varying dose spacing. The BMDLs of studies a to d (different dose spacing) are plotted as functions of the incrementally increased number of responding animals in the high-dose group. In all studies, the background and the low-dose responses are set at 0% and the mid-dose responses at 10%. BMDL, benchmark dose lower confidence bound.

Results

Essentially, in all analyses, the total response rates at a given dose level represent the sum of proportions of the subpopulations A, B, and C that are sensitive at the given dose level.

In Figure 2, the BMDLs for study designs a, b, c, and d, developed using ds 3, 4, 5, and 8, respectively, are plotted as functions of the number of responding animals in the high-dose group. The background and the low-dose responses were set at 0%, and the mid-dose response was set at 10% (ie, the lowest increase of response that results in a significant difference, compared to that of the control, $\alpha = .05$). This setting was associated with 2 general BMDL distribution characteristics. First, a comparison of all dose–responses at a given high-dose–

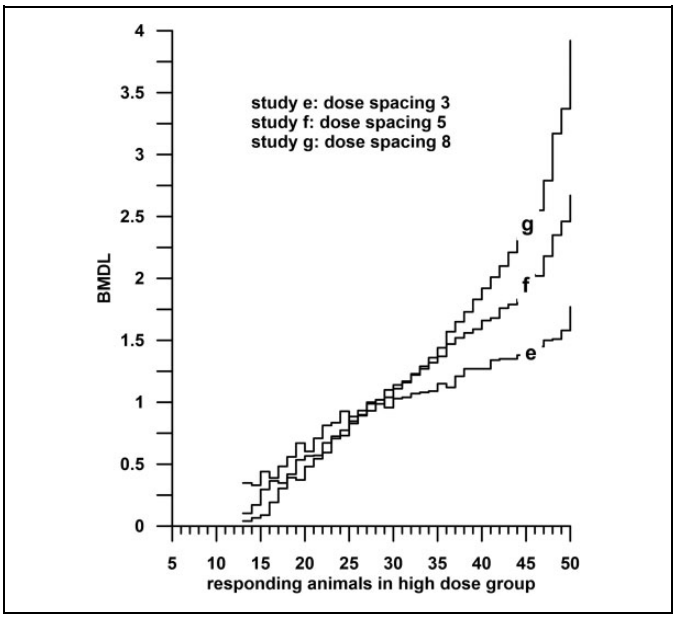

Figure 3. BMDL distributions of studies with background responses of 10% and varying dose spacing. The BMDLs of studies e to g (different dose spacing) are plotted as functions of the incrementally increased number of responding animals in the high-dose group. In all studies, the background and the low-dose responses are set at 10% and the mid-dose responses at 26%, the lowest significant number in the Fisher's exact test. BMDL, benchmark dose lower confidence bound.

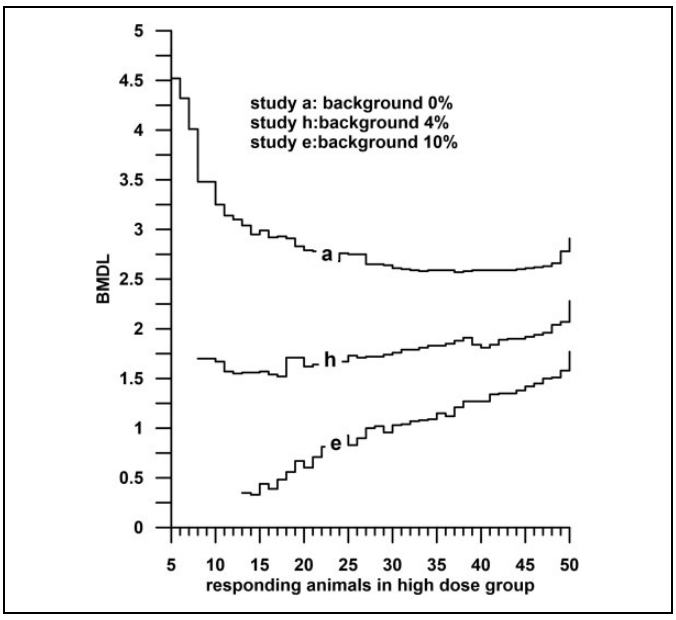

Figure 4. BMDL distributions of studies with varying background responses and fixed dose spacing of 3. The BMDLs of studies a, h, and e (fixed ds at 3) are plotted as functions of the incrementally increased number of responding animals in the high-dose group. In the studies, the background and the low-dose responses were set at 0% in study a, 4% in study h, and 10% in study e with the mid-dose responses at 10% (study a), 16% (study h), and 26% (study e), respectively. BMDL, benchmark dose lower confidence bound.

response rate indicated that the higher the ds, the higher the resulting BMDL. For example, having only animals of subpopulation B but not of subpopulations A and C in the dose groups, 5 (10%) animals of subpopulation B would respond in the mid- and high-dose groups; accordingly, the BMDLs were 4.52, 9.32, and 18.2 using ds of 3, 5, and 8, respectively. Furthermore, when subpopulations B and C were represented with 10% each per dose group, there were 10 responding animals out of 50 (20%) in the high-dose groups; the respective BMDLs were 3.25, 5.6, and 9.42 in studies with ds of 3, 5, and 8, respectively. The second main characteristic of the BMDL distributions was that the higher the proportion of subpopulation C was in a given study design, which increased the responses in the high-dose groups, the lower was the BMDL. Using study design c with a ds of 5, the BMDL was 9.32 for 10% response in the mid- and high-dose group (using only subpopulation B animals), but it was 5.6 after increasing the high-dose group response to 20%. Thus, study designs with narrowly spaced doses and high subpopulation C proportions associated with high response rates in the high-dose groups (due to MOA_C) tended to generate relatively low BMDLs. Moreover, the BMDLs did not further decrease markedly when the responses at the high dose (above the MOA_C threshold) exceeded 50% to 60%.

In Figure 3, the effect of a variable ds on the BMDL distribution is shown for the fictitious study designs e, f, and g (representing studies with ds of 3, 5, and 8, respectively) with a high background response of 10%. In these studies, the resulting BMDLs were markedly lower than in the studies with a background response of 0% (studies a-d). Thus, at 10% background response, the BMDLs increased with increasing responses in the high-dose group and thereby had a tendency opposite to that in the study designs with 0% background incidence (studies a-d). At 10% background response, ds differences only had a marked effect on the BMDLs when the high-dose responses were $\geq$30 animals of 50 ($\geq$60%), with greater ds resulting in higher BMDLs.

Figure 4 shows the effect of different background responses at a constant ds of 3 on the BMDL distribution as a function of the high-dose group responses. Specifically, increasing the subpopulation A proportion from 0 to 2 and 5 responding animals of 50 animals (0%, 4%, and 10% background response, respectively) decreased the BMDLs in fictitious dose–responses with a constant ds of 3. Increasing the high-dose response had opposite effect on the BMDLs at 0% background response (study a), compared to that at 4% and 10% background response (studies h and e, respectively), that is, by increasing the high-dose response, the BMDLs tended to decrease in study a but increased in studies h and e.

In summary, the distribution of the 341 BMDLs obtained with the hypothetical study designs a to h ranges from 0.04 (study g, ds 8, no animals of subpopulation C in dose groups, background response 10%) to 18 (study d, ds 5, no animals of subpopulation C in dose groups, background 0%), depending

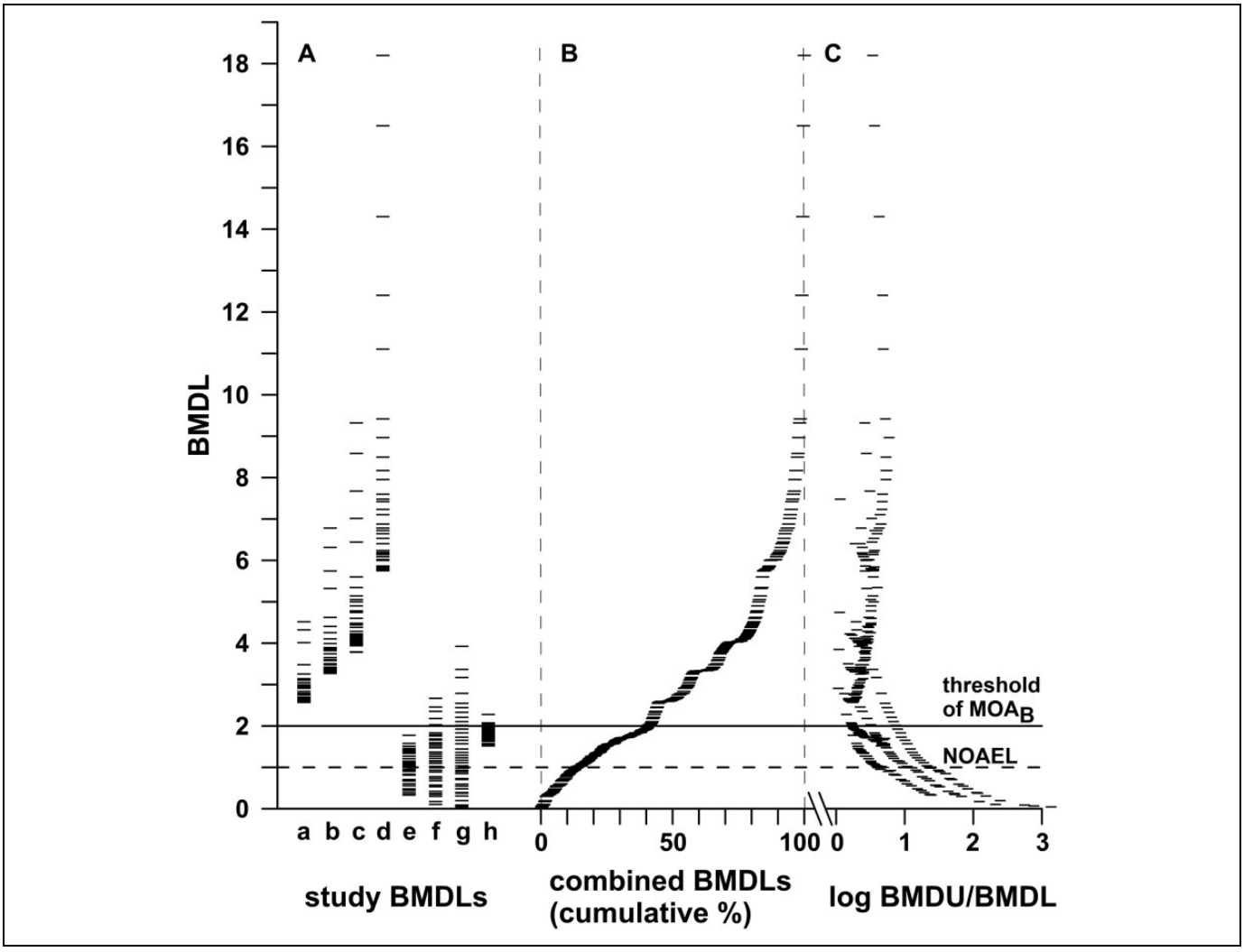

Figure 5. Distribution of BMDLs. In A, the BMDLs of the variations in the fictitious studies a to h (the same BMDLs are also presented in Figures 2–4) are shown. In B, the graph presents the cumulative percentage the BMDLs of all studies a to h (n = 341) account for. In C, the graph presents the BMDU/BMDL ratios of the BMDLs presented in graph B (BMDU is the upper 95% confidence limit of the BMD). The bold horizontal line indicates the dose threshold for MOA_B that divides the entire dose range into a nontoxic and a toxic part. The horizontal dashed line represents the NOAEL of all studies a to h. the graph indicates that ~85% of all BMDLs modeled with studies a to h are above the NOAEL and ca. 60% above the threshold for MOA_B and therefore in the toxic dose range. All BMDLs of studies a to d fall in this toxic dose range. BMDL, benchmark dose lower confidence bound; NOAEL, no-observed-adverse-effect level.

on the ds, the background response, and the high-dose response. In Figure 5A, the distribution of the BMDLs of the individual studies a to h is shown, in panel B, the cumulative distribution of all 341 BMDLs combined from studies a to h and in panel C, the corresponding BMDU/BMDL ratios as quality characteristics for the BMD uncertainty (BMDU is the upper 95% confidence limit of the BMD). Approximately 60% of the BMDLs were at dose levels of $\geq$2, at which MOA_B was inducing significant responses that therefore represented toxic dose levels. The mean and the median BMDU/BMDL ratios were 2.4 and 2.7 for BMDLs within the toxic dose range $\geq$2 (n = 137) and 5.1 and 37, respectively, for the BMDLs within the nontoxic dose range <2 (n = 202). This indicates low BMD model uncertainty for BMDLs in the toxic dose range $\geq$2 and higher model uncertainty in the nontoxic dose range <2. Using the NOAEL approach, the NOAEL value would be at the low-dose level of 1 in all fictitious dose–responses (studies a-h), and the LOAEL would be at the lowest tested dose level of $\geq$2. The individual BMDLs and BMDU/BMDL ratios of all simulations are provided as Supplementary Material.

Discussion

We analyzed the distributions of BMDLs generated by modeling fictitious quantal dose–response data of a series of differently designed fictitious studies. The analysis indicated that the BMDLs of quantal dose–response data were affected by the responses at the highest doses after applying the background response and the ds. Some BMDL values were substantially below the lowest threshold for toxicity at dose level 2, the

threshold for MOA_B, whereas other values were substantially above the MOA_B threshold and therefore in a toxic dose range.

One of the disadvantages of the NOAEL approach is that even if a study is high powered and well conducted, a wide ds may generate a potentially unnecessarily low NOAEL.[2] However, according to the analysis summarized in Figure 5, a substantial proportion of the BMDL values may be in the toxic dose range, depending on the chosen study design. As Haber et al[13] note, sometimes the BMDL is closer to the LOAEL than to the NOAEL, or even higher than the LOAEL, exactly as it appears in our study. From a public health perspective, using either the NOAEL or the BMD approach for evaluating an animal toxicity study in which the ds is wide and the observed response at the high dose is moderate has opposite outcomes according to our analyses: a low HBGV may be derived when the NOAEL approach is applied, whereas the BMD approach may generate a less protective HBGV than the NOAEL because it was derived from an RP in a toxic dose range. Here, it is important to note that the HBGV is conceptually the human analog to the RP (NOAEL or BMDL) from animal studies. This is derived from the assumption that a human subpopulation exists that is more sensitive than the animals used in the study according to the safety factor used in the HBGV derivation.

The not toxicity but study design–related scattering of BMDLs questions the reliability of an RP derived from a mathematical model coshaped by responses originating from dose ranges where possibly additional MOAs operate that do not operate at lower doses. Using high-dose–response data for modeling responses at low doses implicitly assumes that high-dose MOAs are relevant for low-dose responses. If this hypothesis is not validated endpoint by endpoint, the use of the BMD approach is based on untested assumptions. The bypassing of MOA considerations is a setback to efforts to drive regulatory toxicology to incorporate mechanistic knowledge into chemical risk assessment. In the hazard identification and characterization steps of a sound chemical risk assessment using the NOAEL approach, the MOAs that may be active exclusively at the high doses and hence not relevant at lower doses are routinely discussed.[31] An integrated assessment that considers the whole data package including all studies available for a particular chemical may provide evidence that treatment-related responses observed at high doses are not relevant at lower doses[32] because the derived dose–response curve does not disclose the contribution of MOAs with different thresholds.[17,18,33,34] Thus, responses observed at higher doses may prove not to be relevant at lower doses. In the BMD approach, the examination of the dose dependency of MOAs is essentially circumvented because relevant guidance documents[2,8] state that the inclusion of all dose–response data is the critical advantage of the BMD approach over the NOAEL approach, but no indications are provided on how to include MOA considerations. However, the United States Environmental Protection Agency and Haber et al state that strong high-dose responses should be included only with caution (without specifying what exactly this means) and after careful analysis regarding the relevance of the implicated MOA.[3,8] To the best of our knowledge, these are the only references that take up this critical MOA-related point in the context of BMD versus NOAEL discussion. In principle, all dose–responses from a toxicological study subjected to BMD modeling require a thorough examination to determine whether the integration of high- and low-dose responses into a common mathematical model is biologically justified. Otherwise, the BMD approach bases merely on the assumption that a well-fitting mathematical model provides the justification that all dose–response data are biologically relevant for the response modeling at low doses. However, especially in the context of regulatory toxicity studies, the underlying MOAs are not known for all dose–response relationships that often comprise several hundred endpoints investigated in a toxicity study. This is acknowledged by statements that the BMD approach mainly is a mathematical/statistical approach[2,3] that interprets deviations of the observed responses from the model only statistically as an expression of variability-related randomness and sampling errors. But based on the principle that nothing happens without reason, also responses that differ from the model used have causes. However, these observations are trivialized by their statistical interpretation in the context of modeling without investigating the underlying causes of the responses. It is important to emphasize that the biological relevance of responses deviating from the model can only be decided if all MOAs and the subpopulation distributions are precisely known.

The current BMD modeling only considers continuous and monotonous functions as possible models for continuous and quantal data and excludes the consideration of step and nonmonotonous hormetic dose–response curves from the outset. As shown in Figure 1C, the typical 4-dose study provides a very incomplete data series to decide what the actual dose–response curve between the observed responses of the widely spaced dose groups actually looks like. The assumption that a dose–response curve must be continuous and monotonous is speculative. It ignores research in hormesis with this assumption contradictory results and fundamental considerations about the usual design of regulatory studies. Our study covers only a small part of the extremely large number of all theoretically conceivable dose–response relationships resulting from a study design with 4 dose groups of 50 animals each. And yet, already within this limited framework a considerable dispersion of BMDLs is observed, which does not depend on the toxicity of the chemical in the relevant low-dose range but on high dose and therefore possibly irrelevant toxicity, background responses, and study design. Importantly, in our analysis, we assumed that the proportions of the subpopulations in the animal batch received from a breeder were retained in the dose groups (assuming perfect randomization), which was a simplification that is probably never true especially when outbred animal stocks are used in toxicity testing. Therefore, if a large number of fictitious dose–response relationships were designed under the assumption of imperfect randomization, the dispersion of BMDLs would probably be significantly greater after modeling these data. In addition, BMD analyses would be interesting for fictitious studies based on dose-dependent

counteracting MOAs for hormetic dose–response relationships. It is certainly justified to assume that the dispersion of BMDLs would increase significantly if further simulations were carried out. In view of the complexity of BMDL-influencing factors that are not toxicity related, we suggest that BMD modeling should always be accompanied by a biological justification focusing on among others possibly involved MOAs that explains why the researcher considers the response data appropriate for modeling-based analysis.

The NOAEL is like the BMDL sensitive to dose spacing but is insensitive to exclusive high-dose effects and less sensitive to the level of background response. From a public health perspective, we consider the NOAEL approach to be at least as suitable as the BMD approach for deriving RP from animal studies if no detailed information on MOAs and subpopulation distribution between dose groups is available.

We are aware that all concepts and models used in regulatory toxicology have strengths and weaknesses. Of course, this also applies to the BMD and the NOAEL approach. We, therefore, strongly advocate that the BMD approach is updated to integrate insights into the mechanistic basis of toxicity.

Acknowledgments

The authors thank Andrea Zarn-Twardek for conceptually important discussions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jürg A. Zarn https://orcid.org/0000-0003-2216-560X

Supplemental Material

Supplemental material for this article is available online.

References

1. Herrman JL, Younes M. Background to the ADI/TDI/PTWI. *Regul Toxicol Pharmacol*. 1999;30(2):109-113.
2. European Food Safety Authority. Update: use of the benchmark dose approach in risk assessment. *EFSA J*. 2017;15(1):e04658.
3. Haber LT, Dourson ML, Allen BC, et al. Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Crit Rev Toxicol*. 2018;48(5):387-415.
4. Sand S, Victorin K, Filipsson AF. The current state of knowledge on the use of the benchmark dose concept in risk assessment. *J Appl Toxicol*. 2008;28(4):405-421.
5. Slob W. A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. *Crit Rev Toxicol*. 2017;47(4):342-351.
6. Starr TB, Goodman JI, Hoel DG. Uses of benchmark dose methodology in quantitative risk assessment. *Regul Toxicol Pharmacol*. 2005;42(1):1-2.
7. Travis KZ, Pate I, Welsh ZK. The role of the benchmark dose in a regulatory context. *Regul Toxicol Pharmacol*. 2005;43(3):280-291.
8. United States Environmental Protection Agency. *Benchmark Dose Technical Guidance*. Washington, DC: United States Environmental Protection Agency; 2012.
9. Crump KS. A new method for determining allowable daily intakes. *Fundam Appl Toxicol*. 1984;4(5):854-871.
10. Slob W. Benchmark dose and the three Rs. part I. getting more information from the same number of animals. *Crit Rev Toxicol*. 2014;44(7):557-567.
11. Slob W. Benchmark dose and the three Rs. part II. Consequences for study design and animal use. *Crit Rev Toxicol*. 2014;44(7):568-580.
12. Slob W, Setzer RW. Shape and steepness of toxicological dose-response relationships of continuous endpoints. *Crit Rev Toxicol*. 2014;44(3):270-297.
13. Dornbos P, LaPres JJ. Incorporating population-level genetic variability within laboratory models in toxicology: from the individual to the population. *Toxicology*. 2018;395:1-8.
14. Festing MFW. Genetically defined strains in drug development and toxicity testing. In: Proetzel G, Wiles MV, eds. *Mouse Models for Drug Discovery: Methods and Protocols*. New York, NY: Springer; 2016:1-17.
15. Harrill AH, McAllister KA. New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. *Environ Health Perspect*. 2017;125(8):086002.
16. Lutz WK. Susceptibility differences in chemical carcinogenesis linearize the dose-response relationship: threshold doses can be defined only for individuals. *Mutat Res*. 2001;482(1-2):71-76.
17. Keenan KP.Effects of diet and overfeeding on body weight and survival in the rodent bioassay: the impact on pharmaceutical safety assessment. *Int J Toxicol*. 1998;17(suppl 2):101-117.
18. Keenan KP, Ballam GC, Soper KA, Laroque P, Coleman JB, Dixit R. Diet, caloric restriction, and the rodent bioassay. *Toxicol Sci*. 1999;52(2 suppl):24-34.
19. Zarn JA, O'Brien CD. Current pesticide dietary risk assessment in light of comparable animal study NOAELs after chronic and short-termed exposure durations. *Arch Toxicol*. 2018;92(1):157-167.
20. Lutz WK. Dose-response relationships in chemical carcinogenesis reflect differences in individual susceptibility. consequences for cancer risk assessment, extrapolation, and prevention. *Hum Exp Toxicol*. 1999;18(12):707-712.
21. Lutz WK. Differences in individual susceptibility to toxic effects of chemicals determine the dose-response relationship and consequences of setting exposure standards. *Toxicol Lett*. 2002;126(3):155-158.
22. Lutz WK, Gaylor DW, Conolly RB, Lutz RW. Nonlinearity and thresholds in dose-response relationships for carcinogenicity due to sampling variation, logarithmic dose scaling, or small differences in individual susceptibility. *Toxicol Appl Pharmacol*. 2005;207(2 suppl):565-569.

23. Lutz WK, Lutz RW, Andersen ME. Dose-incidence relationships derived from superposition of distributions of individual susceptibility on mechanism-based dose responses for biological effects. *Toxicol Sci*. 2006;90(1):33-38.
24. Agathokleous E, Calabrese EJ. Hormesis: the dose response for the 21st century: the future has arrived. *Toxicology*. 2019;425: 152249.
25. Calabrese EJ.The linear no-threshold (LNT) dose response model: a comprehensive assessment of its historical and scientific foundations. *Chem Biol Interact*. 2019;301:6-25.
26. Calabrese EJ, Agathokleous E, Kozumbo WJ, Stanek EJ III, Leonard D. Estimating the range of the maximum hormetic stimulatory response. *Environ Res*. 2019;170:337-343.
27. Ringblom J, Johanson G, Oberg M. Current modeling practice may lead to falsely high benchmark dose estimates. *Regul Toxicol Pharmacol*. 2014;69(2):171-177.
28. Ringblom J, Kalantari F, Johanson G, Oberg M. Influence of distribution of animals between dose groups on estimated benchmark dose and animal welfare for continuous effects. *Risk Anal*. 2017;37(9):1716-17278.
29. Crump KS. Use of threshold and mode of action in risk assessment. *Crit Rev Toxicol*. 2011;41(8):637-650.
30. World Health Organization. JMPR toxicological monographs. *WHO Evaluations Part II: Toxicology*. https://www.who.int/foodsafety/publications/jmpr-monographs/en/. Accessed January 20, 2020.
31. Boobis AR, Cohen SM, Dellarco VL, et al. Classification schemes for carcinogenicity based on hazard-identification have become outmoded and serve neither science nor society. *Regul Toxicol Pharmacol*. 2016;82:158-166.
32. Meek ME, Boobis A, Cote I, et al. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *J Appl Toxicol*. 2014; 34(1):1-18.
33. Slikker W Jr, Andersen ME, Bogdanffy MS, et al. Dose-dependent transitions in mechanisms of toxicity: case studies. *Toxicol Appl Pharmacol*. 2004;201(3):226-294.
34. Slikker W Jr, Andersen ME, Bogdanffy MS, et al. Dose-dependent transitions in mechanisms of toxicity. *Toxicol Appl Pharmacol*. 2004;201(3):203-225.