

# Introduction to benchmark dose estimation from multiple endpoints and multiple studies

## Current practice & challenges

Scientific Colloquium N23 - EFSA/EBTC Joint Colloquium  
Evidence integration in risk assessment: the science of combining apples & oranges




October 25 – 26, 2017

Marc Aerts

# Outline



- **Setting the BMD scene**
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- Methodological approaches to multiple studies
- Conclusion: multiple  challenges

# A first example: single study, multiple endpoints

28-day toxicity study with Rhodorsil Silane in rats

$y_1$  = Aspartate aminotransferase (ASAT)

$y_2$  = Alanine aminotransferase (ALAT)

$y_3$  = Haemoglobin

$y_4$  = Lymphocytes

$y_5$  = Mean corpuscular volume

$y_6$  = Mean corpuscular haemoglobin (mch)

$y_7$  = Mean corpuscular concentration (mchc)

$y_8$  = Neutrophiles

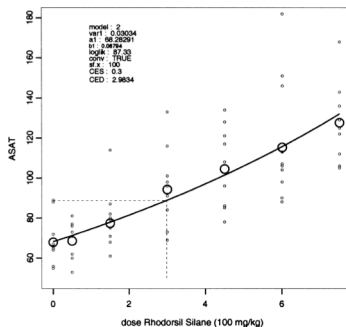
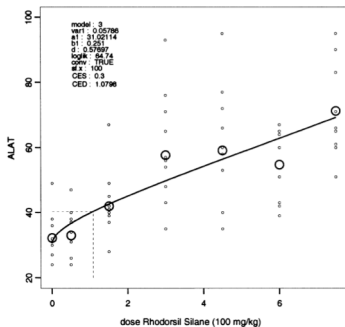
$y_9$  = Red blood cell count

$y_{10}$  = Relative liver weight

$y_{11}$  = Reticulocytes

$y_{12}$  = Spleen weight

$y_{13}$  = Total protein

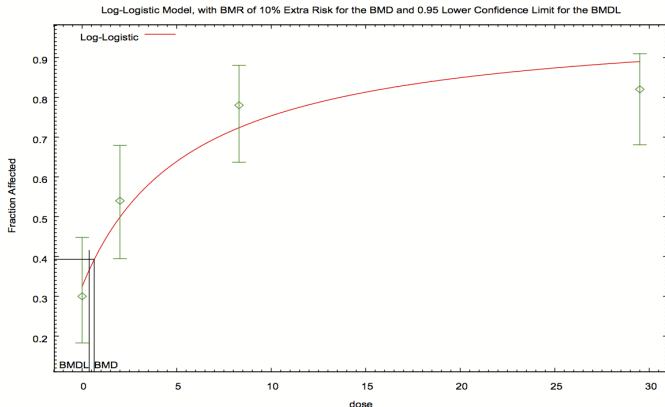


(see Woutersen et al, 2001 & presentation of Wout Slob)

## A second example: two studies, two endpoints

Sunahara et al. (1993), Cho et al. (2008), EFSA CONTAM Panel (2016)

- rats (50 animals/sex per treatment group)
- 3-monochloropropane-1,2-diol (3-MCPD), 4 concentrations
- $y_1$  = Tubular hyperplasia,  $y_2$  = Nephropathy



(see first case of presentation of Matthew Wheeler)

- Single endpoint, single study

$$(d_i, y_i) \quad i = 1, \dots, n$$

- Single study, multiple endpoints

$$(d_i, (y_{1,i}, y_{2,i}, \dots, y_{m,i})) \quad i = 1, \dots, n$$

- Single endpoint, multiple studies

$$(d_i^{(1)}, y_i^{(1)}) \quad i = 1, \dots, n_1$$

$$(d_i^{(2)}, y_i^{(2)}) \quad i = 1, \dots, n_2$$


...

$$(d_i^{(N)}, y_i^{(N)}) \quad i = 1, \dots, n_N$$

- Multiple studies, multiple endpoints

**Further:** aggregated data, hierarchical structures, clustered data, ...

# Outline

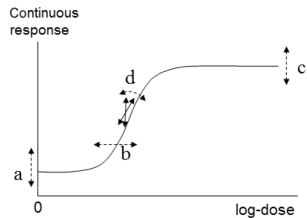
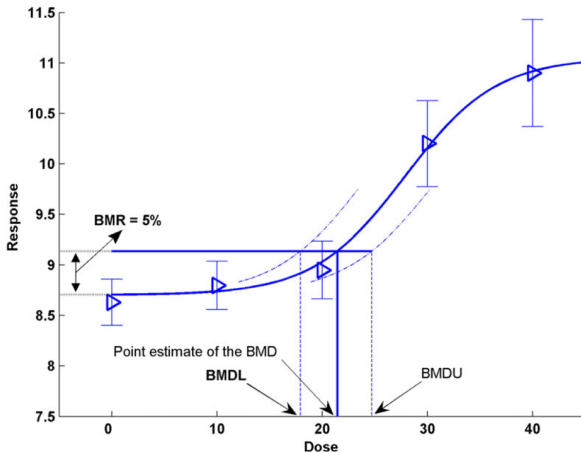
- Setting the BMD scene
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- Methodological approaches to multiple studies
- Conclusion: multiple  challenges

## BMD approach single endpoint, single study

- 1 data  $(d_i, y_i) \quad i = 1, \dots, n$
- 2 dose response model, e.g.  $a\{1 + (c - 1)x^d/(b^d + x^d)\}$
- 3 BMD
- 4 BMDL (and BMDU)

# BMD approach single endpoint, single study

$$a\{1 + (c - 1)x^d/(b^d + x^d)\}$$

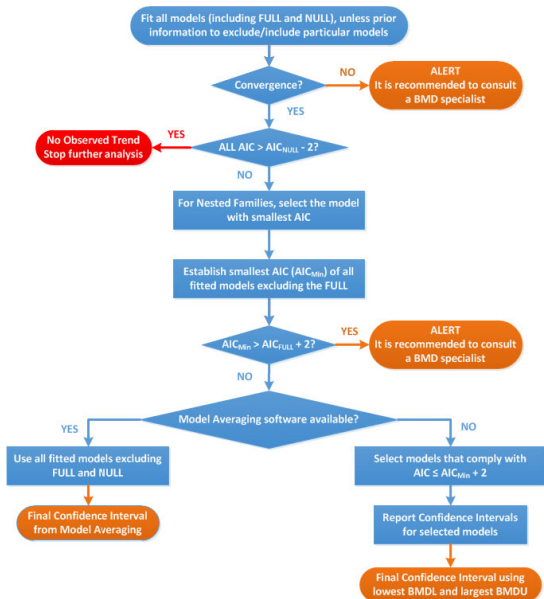




### **Update: use of the benchmark dose approach in risk assessment** (EFSA SC, November 2016)


- BMD more advanced than NOAEL for deriving RP
- Information criterion AIC introduced to measure goodness-of-fit and to define weights for model averaging
- Model averaging is recommended for BMD, BMDL and BMDU
- Flowchart to guide the user step-by-step
- Recommended to report BMDL and BMDU
- A report template is provided

# BMD approach single endpoint, single study



# Outline



- Setting the BMD scene
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- Methodological approaches to multiple studies
- Conclusion: multiple  challenges

### EFSA guidance update 2016, in summary (see appendix)

- Multiple endpoints
  - select endpoint resulting in lowest BMDL
  - taking into account additional information (accuracy and uncertainty) and additional arguments (toxicological, biological)
  - more holistic, case by case, risk assessor's responsibility
  - BMR expressed in terms of percent change allows comparison
- Single endpoint, multiple studies
  - use of covariates

Conclusion: only some general principles, limited practical guidance

### U.S. EPA technical guidance 2012, in summary (see appendix)

- Specialized models for multiple related outcomes
- Specialized models for multiple endpoints across studies
- Bayesian approaches
- Selection of studies by risk assessor
- Selection of subsets of endpoints as representative
- Simplest approach combines datasets as they were collected simultaneously
- Variability among datasets necessitates more complex modeling
- Limitation to a limited number of flexible models

Conclusion: mentioning specialized methods, some general principles, limited practical guidance

### Is it worthwhile to extend to multiple endpoints & studies?

- more evidence, more information
- confirmed and similar evidence across endpoints and studies




- additional insights
- improved risk assessment - more relevant for human health

### But

- what is the price to pay?
- how far to go with more complete but more complex analysis?
- can we think about a common BMD?

# Outline



- Setting the BMD scene
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- Methodological approaches to multiple studies
- Conclusion: multiple  challenges

Data on  $m$  endpoints (example 1: ASAT, mcv, mch, mchc, ...)

$$(d_i, (y_{1,i}, y_{2,i}, \dots, y_{m,i})) \quad i = 1, \dots, n$$

- **simplify to single ‘outcome’ - be pragmatic...**
- multivariate approach - too complex?



## Simplify to single 'outcome' - be pragmatic...

Data on  $m$  endpoints (example 1: ASAT, mcv, mch, mchc, ...)

$$(d_i, (y_{1,i}, y_{2,i}, \dots, y_{m,i})) \quad i = 1, \dots, n$$

simplifying the genuinely multivariate nature

- at the real end: BMDL
- at the end, one step back: BMD
- at the beginning: data
- in the middle: dose response

### 1. Select endpoint with lowest BMDL

$$\min\{\text{BMDL}_1, \text{BMDL}_2, \dots, \text{BMDL}_m\}$$

---

- protecting against the “most sensitive” outcome
- preselection based on (toxicological...) arguments
- 2 versions:  $m$  univariate or 1 multivariate dose response model
- ? overly protective, too conservative
- + very easy for univariate version
- the more endpoints, the lower the BMDL
- important issues
  - selection ignores accuracy ( $\text{BMDU}_1, \text{BMDU}_2, \dots, \text{BMDU}_m$ )
  - associated BMDU
  - lower bound of CI of any single parameter (statistical interpretation)

### 2. Select endpoint with lowest BMD

$$\text{BMD} = \min\{\text{BMD}_1, \text{BMD}_2, \dots, \text{BMD}_m\}$$

BMDL = lower bound & BMDU = upper bound

---

- protecting against the “most sensitive” outcome
- preselection based on (toxicological...) arguments
- 2 versions:  $m$  univariate or 1 multivariate dose response model
- ? less conservative
- + BMD very easy for univariate version
- + BMDL and BMDU available, with statistical interpretation
- the more endpoints, the lower the BMD(L)
- Determination of BMDL and BMDU less straightforward

### 3. Combine multiple endpoints into one new single endpoint

- 1 Each endpoint  $y_{1,i}, y_{2,i}, \dots, y_{m,i}$  as quantal  
e.g. adverse event  $y < c$  (or  $>$ ) for continuous endpoint  $y$
- 2 Define single endpoint

$$y = \max\{y_1, y_2, \dots, y_m\}$$

- 3 BMD, BMDL and BMDU for single endpoint  $y$

- 
- protecting against the “most sensitive” outcome
  - preselection based on (toxicological...) arguments
  - ? conservative
  - + very easy
  - the more endpoints, the lower the BMD(L)
  - needs individual data & identical dose levels
  - choice of threshold  $c$  for continuous endpoint  $y$

### 3. Combine multiple endpoints into one composite endpoint

- 1 Each endpoint  $y_{1,i}, y_{2,i}, \dots, y_{m,i}$  transformed to a unitless score  $s_{1,i}, s_{2,i}, \dots, s_{m,i}$  based on desirability functions
- 2 Define single, composite endpoint as weighted geometric mean
$$y = \{s_1^{w_1} \times s_2^{w_2} \times \dots \times s_m^{w_m}\}^{1/\sum_i w_i} \quad (\text{Coffey et al 2007})$$
- 3 BMD, BMDL and BMDU for single endpoint  $y$

- 
- + relatively easy
  - + weights to rank importance of each endpoint
  - + sensitive to toxicity evident in only a few endpoints
  - degree of subjectivity
  - needs individual data & identical dose levels

### 4. Combine in single response model using covariate

- 1 Make sure all endpoints are of the same type and 'scale'
- 2 Fit single dose response model with endpoint specific covariate effects, e.g.

$$(a + \alpha_{\text{endpoint } i}) \{1 + (c - 1)x^d / (b^d + x^d)\}$$

- 3 Determine BMD(L/U), ideally in common to endpoints

- 
- preselection based on (toxicological...) arguments
  - ? protecting against the "most sensitive" outcome
  - ? what if no common BMD
  - + can host a common BMD  $\Rightarrow$  test for common BMD
  - same type and 'scale'
  - more complex model building with more limitations

## Methodological approaches to multiple endpoints

Data on  $m$  endpoints (example 1: ASAT, mcv, mch, mchc, ...)

$$(d_i, (y_{1,i}, y_{2,i}, \dots, y_{m,i})) \quad i = 1, \dots, n$$

- simplify to single 'outcome' - be pragmatic...
- **multivariate approach - too complex?**

### 5. Multivariate dose response model

Data on  $m$  endpoints (example 1: ASAT, mcv, mch, mchc, ...)

$$(d_i, (y_{1,i}, y_{2,i}, \dots, y_{m,i})) \quad i = 1, \dots, n$$

Multivariate model based on joint distribution of  $(y_1, y_2, \dots, y_m)$

---

- accounts for, potentially models association between endpoints
- accommodates (models for) the association between endpoints
- ? what if no common BMD
- + more flexible modelling
- + allows (to test for) a common BMD
- more complex model building fitting
- complexity increases with number  $m$  of endpoints
- needs individual data & identical dose levels



### 6. Dimension reduced multivariate dose response model

- 1 Determine principle components (or extensions)

$$(d_i, (\text{pc}_{1,i}, \text{pc}_{2,i}, \dots)) \quad i = 1, \dots, n$$

- 2 Fit e.g. bivariate dose response model

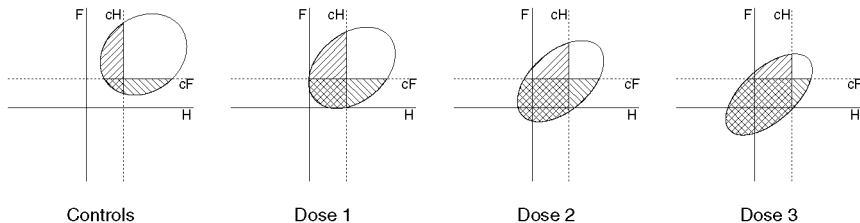
- 
- a few pc's with little loss of information
  - pc 1 and pc 2 might represent a joint, biological effect
  - ? what if no common BMD
  - + appealing if  $m$  large with highly correlated endpoints
  - + allows (to test for) a common BMD on scale of pc's
  - more complex model building fitting
  - needs individual data & identical dose levels

## 7. One example for two continuous endpoint $y_1$ and $y_2$

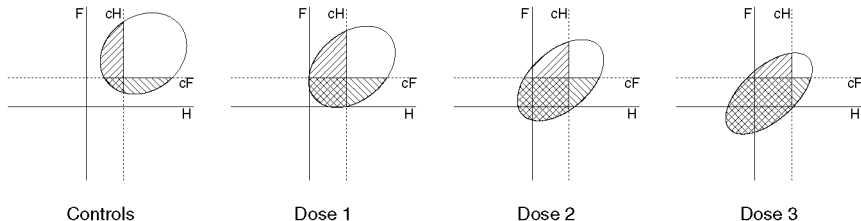
- 1 consider a bivariate (log)normal distribution
- 2 define adverse events  $y_1 < c_1$  and  $y_2 < c_2$  such that

$$P(0) = P(\text{any adverse event for the control group}) = 0.05$$

with the constraint that  $\frac{c_1 - \mu_1(0)}{\sigma_1(0)} = \frac{c_2 - \mu_2(0)}{\sigma_2(0)}$



## Multivariate 'special' methods - too complex for practice?



(from Yu and Catalano, 2008, *Risk Analysis*)

- 3 BMD is the dose satisfying the typical extra risk

$$\frac{P(d) - P(0)}{1 - P(0)} = q$$

- 4 BMDL and BMDU as lower and upper bounds

### 8. Structural equations model (Budtz-Jørgensen 2007)

- observed endpoints as manifestations of one (or more) underlying latent variable(s)  $\eta$
- BMD focuses on the probability of an adverse event defined on the latent scale, of an abnormal latent response  $P(\eta > c)$

### 9. Modelling mechanistic processes

9.1 Bayesian hierarchical mechanistic models (Choi et al 2010)


9.2 Physiologically-based pharmacokinetic PBPK models (Lipscomb et al 2012)

- based on mechanistic biological processes
- based on taxonomy and mode of action of chemicals
- facilitating integration of data
- elucidation of common and divergent endpoints

## Summary

- Several approaches can be considered, from simple to complex
- Using all endpoints, “representative” subset, single combined
- Issues & choices
  - aggregate versus individual data
  - hierarchical data: clustered and/or longitudinal
  - other covariates, e.g. gender, time in dose-duration response models
  - model choice and model averaging
  - Bayesian versus likelihood or frequentist methods
- Balancing accuracy and complexity, interpretability
- Not all methods studied sufficiently in the BMD setting
- Availability of software
- Guidelines: pragmatic, stepwise, ...

# Outline

- Setting the BMD scene
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- **Methodological approaches to multiple studies**
- Conclusion: multiple  challenges

## Adding information, but also complexity...

- multiple studies on single endpoint
- but other designs, other sample sizes, other species, other covariates, ...
- multiple endpoints & multiple studies, basic ideas
  - apply multiple endpoints approaches across studies
  - apply multiple endpoints approaches for each study, combine across studies

## Standard meta-analysis methods

- Simple pooling “compatible” datasets as if collected simultaneously
- Meta-analysis: fitting aggregated data and/or pooling datasets but accounting for and quantifying variability across studies

(Meta-analysis of dose-effect relationship of cadmium for benchmark dose evaluation, EFSA 2009)

- Limitations & challenges
  - same endpoint, different species
  - different designs, different scales
  - individual and aggregated data
  - model averaging
  - limitations in types of heterogeneity across studies
  - common BMD across endpoints and across studies



## CatReg


- effects assigned to ordinal categories of severity
- up to two independent variables related to exposure (e.g., concentration and time)
- allows for meta-analysis of data from multiple toxicity studies simultaneously as long as the responses have been converted into ordinal data using the same category descriptions

## Bayesian methods

- Bayesian model averaging for random and fixed effects meta-analysis (metaBMA: Heck et al, 2017)
- Bayesian network models as an intermediate to PBPK models (Hack et al, 2010)

# Outline

---

- Setting the BMD scene
- Benchmark dose approach single endpoint
- Current guidance on multiple endpoints & multiple studies
- Methodological approaches to multiple endpoints
- Methodological approaches to multiple studies
- **Conclusion: multiple  challenges**

## Conclusion: multiple 🌐 challenges

- ? Consensus of an acceptable starting point
- ? Sufficient knowledge and expertise available
- ? Guidelines to go further, depending on case, data, ... (flowchart)
- ? Keep consistency with single study, single endpoint guidelines
- ? Guarantee optimal use of all data structures
- ? Accessible and user friendly software

## APPENDIX: Current EFSA & EPA guidance

---

- EFSA guidance update 2016
- U.S. EPA technical guidance 2012

### EFSA guidance update 2016, p 8

An **overall study BMDL**, i.e. the critical BMDL of the study, is **selected from** the obtained set of **BMD confidence intervals** for the different potentially critical endpoints. In principle, the BMD approach could be applied **to every endpoint measured in the relevant studies**. The critical effect would then be selected in an analogous way as in the NOAEL approach, that is, **not only as the endpoint resulting in the lowest BMDL, but also taking additional toxicological arguments into account**, just as in the case of the NOAEL approach. However, it is recommended to make use of one of the strengths of the BMD approach, and select the study BMDL based on considering the complete BMD confidence intervals for the endpoints considered and **combine the information on uncertainties in the underlying data with biological considerations**.

### EFSA guidance update 2016, p 18-19

As a statistical consideration, one might consider to select a BMR higher than 5% for endpoints that tend to show a relatively large within-group variation (in terms of coefficient of variation), and/or a relatively high maximum response (if known, based on experience with that endpoint over a larger number of studies (Slob and Setzer, 2014)). Increasing the BMR (in terms of a percent change) for data showing a relatively large maximum response is somewhat similar to using a BMR defined as a change equal to 1 SD (Slob, 2016); an important difference is that the **BMR expressed in terms of a percent change allows for comparison among studies and populations that differ in within-group variation.**

### EFSA guidance update 2016, p 25-26

Besides fitting dose-response models to single data sets, it is possible to fit a given model to a **combination of data sets** which differ in a specific aspect, such as sex, species or exposure duration, but are similar otherwise. In particular, the **response parameter (endpoint) needs to be the same**. By fitting the dose-response model to the combined data set, with the **specific factor included in the analysis as a so-called covariate**, it can be examined in what sense the dose-responses in the subgroups differ from each other, based on statistical principles (like AIC).

Combining data sets in a dose-response analysis with covariate(s) may have two reasons. The first is that it provides a **powerful method** for examining and quantifying potential differences in dose-response between the subgroups. ... .. The second reason for combining data sets and applying the covariate approach is **to improve the precision** of the estimated BMD(s), i.e. to obtain a smaller BMD confidence interval.



### EFSA guidance update 2016, p 29

The BMD confidence interval should be derived **for all data sets considered relevant** (potentially leading to the RP), resulting in a set of confidence intervals indicating the uncertainty ranges around the true BMD for the endpoints considered. This set of BMD confidence intervals concisely reflects the information provided by the available data and **provides the starting point for the risk assessor to derive the RP**. One way to proceed is to **simply select the endpoint with the lowest BMDL and use that value as the RP**. However, this procedure **may not be optimal** in all cases, and the risk assessor might decide to use **a more holistic** approach, where all relevant aspects are taken into account, such as the BMD confidence intervals (rather than just the BMDLs), the biological meaning of the relevant endpoints, and the consequences for the HBGV or the MOE. This process will differ from **case to case** and it is the **risk assessor's responsibility** to make a substantiated decision on what BMDL will be used as the RP.

### U.S. EPA technical guidance 2012, p 10

Many noncancer health effects are characterized by multiple endpoints that are not completely independent of one another. Lefkopoulou et al. (1989), Chen et al. (1991), Ryan (1992a, b), Catalano et al. (1993), Zhu et al. (1994), Krewski and Zhu (1995), and Fung et al. (1998) have worked on this issue using developmental toxicity data and have shown that, in most cases, the **BMDL derived from a multinomial modeling approach is lower than that for any individual endpoint**. This approach has not been applied to other health effects data but should be kept in mind when **multiple related outcomes** are being considered for a particular health effect.

### U.S. EPA technical guidance 2012, p 11

Most approaches to BMD modeling have focused on modeling single or multiple responses from a single study. **Categorical regression modeling** (Dourson et al. 1985; Hertzberg 1989; Hertzberg and Miller 1985; Guth et al. 1997; Simpson et al. 1996a, b) is one method that allows the results for **multiple endpoints across studies** to be used to make an **overall assessment** of the toxicity of a compound based on a larger database. Although so far this method has **not been widely used for BMD** computation, it shows promise as a way to more quantitatively and rigorously combine information from a rich database.

U.S. EPA technical guidance 2012, p 29

**Bayesian approaches** to BMD calculation express the uncertainty in the BMD estimate with a probability distribution (in Bayesian parlance, the posterior distribution), in contrast to the confidence limits employed by the more commonly used frequentist approach (Hasselblad and Jarabek 1995). Although the Bayesian approach has not yet found wide application, it has some potentially useful features. The Bayesian approach facilitates **combining results from different datasets** to provide a **more robust estimate** as well as an **evaluation of the uncertainty** in that estimate that would take into **account the variability among studies**. This type of approach may lead to improvements over the more widely used methods, which only quantify the uncertainty inherent in a single study.

U.S. EPA technical guidance 2012, p 14

Following a complete review of the toxicity data, the **risk assessor selects the studies for BMD analysis**, based on the human exposure situation being addressed, the quality of the studies, the reporting adequacy, and the relevance of the endpoints. The process of selecting studies for BMD analysis is intended to identify those studies for which modeling is feasible, so that BMDs can be calculated. All relevant studies should be considered for modeling. In some cases, the selection process will identify a single study or very few studies for which calculations are appropriate. In other cases, there may be a number of studies, or studies with a number of endpoints reported, which may require a large number of BMD calculations. In these latter cases, **it may be possible to select a subset of endpoints as representative of the effects in a target organ or study. This selection can be made on the basis of sensitivity or severity, which may be more easily compared within a single study in the same target organ than across studies.**

### U.S. EPA technical guidance 2012, p 15

Typically **all endpoints** within a study that the risk assessor has judged to be relevant to the exposure **should be considered for modeling**. This will help ensure that no endpoints with the potential of having the most sensitive effect for risk assessment applications, usually having the lowest BMDL, are excluded from the analysis. ... ..  
... Selected endpoints from different studies that have the potential to be used in the determination of a POD(s) should all be modeled, especially if different UFs may be used for different studies and endpoints. The risk assessor selects the BMDL(s) to serve as the POD(s) using scientific judgment and principles of risk assessment as well as the results of the modeling process. Note that it is sometimes desirable to carry through risk estimate derivations for multiple endpoints for comparisons and other purposes.

U.S. EPA technical guidance 2012, p 18

Datasets that are statistically and biologically compatible may be combined prior to dose-response modeling, resulting in increased confidence, both statistical and biological, in the calculated BMD. **The simplest approach to combining datasets is to treat the data as if they were all collected simultaneously.** If it is plausible that the multiple datasets represent a homogeneous picture of the dose-response (for example, the responses at doses common to two or more datasets are essentially the same and statistically undifferentiable), then this is a justifiable approach. ... .. More likely, there will be **some variability among datasets, requiring more elaborate modeling** to combine information properly. There is as **yet too little practical, as well as theoretical, experience** with this situation to provide specific guidance in the matter, other than to say that statistically appropriate methods and biological judgment must be used and justified if datasets are combined for modeling. One technique for statistically accommodating variability among studies is **categorical regression analysis** (Simpson et al. 1996a, b), although this method requires a large number of studies for the chemical of interest.

U.S. EPA technical guidance 2012, p 26

The initial selection of a group of models to fit to the data is governed by the nature of the measurement that represents the endpoint of interest and the experimental design used to generate the data. In addition, certain constraints on the models or their parameter values sometimes need to be observed and may influence model selection. Finally, it may be desirable **to model multiple endpoints at the same time**. The **diversity of possible endpoints** and shapes of their dose-response relationships for different agents **precludes specifying a small set of models** to use for BMD computation. This will inevitably lead to **the need for judgment when selecting the final model** and BMD/BMDL for dose-response assessment. As experience using BMD methodology in dose-response assessment accumulates, it may be possible to **narrow the number of models to a few that are sufficiently flexible and non-redundant** to be specified for certain scenarios.