BY HOW MUCH DO SHAPES OF TOXICOLOGICAL DOSE-RESPONSE RELATIONSHIPS VARY?

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

www.rivm.nl

Wout Slob[1], R. Woodrow Setzer [2]

1 National Institute of Public Health and the Environment (RIVM), The Netherlands wout.slob@rivm.nl
2 National Center for Computational Toxicology, US Environmental Protection Agency setzer.woodrow@epa.gov

Abstract #632a

EPA

www.epa.gov

ABSTRACT

A re-analysis of a large number of historical dose-response data for continuous endpoints showed that the shapes of the dose-response relationships were surprisingly homogenous. The datasets were selected on the sole criterion that they were expected to provide relatively good information on the dose-response shape, and included a variety of endpoints and both in vivo and in vitro studies of various types. Both the four-parameter exponential and Hill model adequately described all toxicological dose-response data we considered. For a given endpoint and study type, dose-response shapes did not differ statistically significantly among chemicals in the in vitro studies considered, while a mild among-chemical variation in the steepness parameter seemed to be present in the in vivo studies. These findings have various practical consequences. For continuous endpoints, model selection in the BMD approach is not a crucial issue. The often-applied approach of using constraints on the model parameters to prevent "infinite" slope at dose zero in fitting a model is not in line with our findings, and appears to be unjustified. Instead, more realistic ranges of parameter values could be derived from re-analyses of large numbers of historical dose-response datasets in the same endpoint and study type, which would then be used as parameter constraints or informative priors in the analysis of future individual datasets. This approach would be particularly useful for weak datasets (e.g. few doses, much scatter). In addition, this approach may open the way to use fewer animals in future studies. Finally, we argue that distinctions between linear, sub/supralinear or thresholded dose-response shapes, based on visual inspection of plots, are not biologically meaningful nor useful for risk assessment.

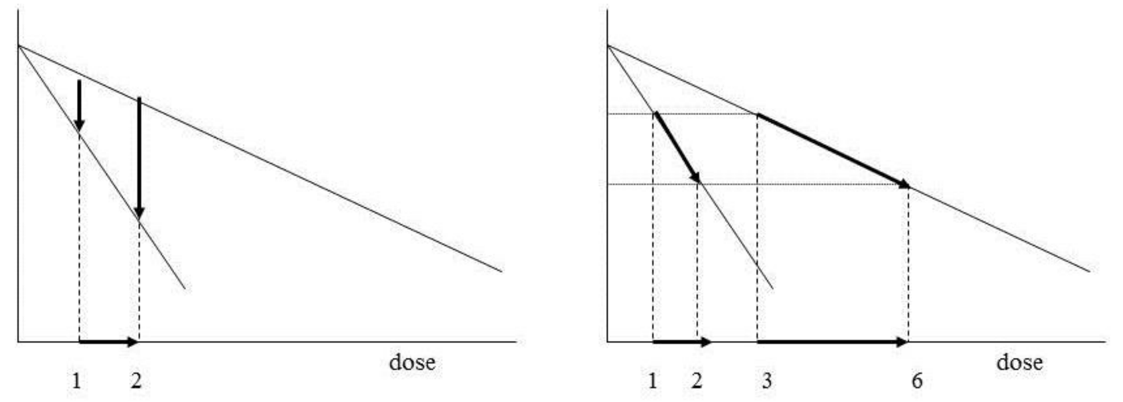
INTRODUCTION

As regulatory toxicology transitions from basing decisions on NOAELS to Benchmark Doses, choices about the shape of dose-response curves have moved into a more prominent position. Currently, the choice of the dose-response models for deriving a point of departure (PoD) is made based on the individual dataset in hand, as if it were the first dose-response dataset ever generated. General insight into the shapes of dose-response relationships based on experience from historical toxicological studies could be highly valuable in the process of model selection, particularly if general patterns of dose-response shapes were to be found.

COMPARING DOSE-RESPONSE SHAPE

Considerations of shape should be independent of the potency or background level for the endpoint. So, before considering shape, we need to scale the response by the background level, and the doses by equipotent doses (doses that yield the same response level). If after this scaling two curves can be superimposed, then they have the same shape.

One simple aspect of shape is the maximum response, relative to control (or background).



The second is more complicated, and is illustrated by reference to the figure below. Consider two linear dose–responses (dose on original scale) for two hypothetical chemicals A and B. In the left panel the change in response for a given additive increment in dose differs, i.e. the steepness in both lines is different. However, as the right panel illustrates, a given change in response is achieved by the same percent change of equipotent doses (indicated as 1 and 3). We say the two lines have the same *log-steepness*, defined as

d log *y* / d log *x*.

Log-steepness is the other aspect of shape that needs to be considered when comparing the shapes of dose-response curves.

A straightforward way to estimate log-steepness is with the ratio of the BMD_{10} to the BMD_5. It can be shown that the BMD ratio is directly related to the log-steepness.

MODELS AND THEIR PARAMETERS

Rather than scaling the data by background and potency, we choose models that have the scaling built into them as parameters. This makes the statistical resulting statistical inferences more reliable.

Exponential and Hill Models

$$\text{Exponential}: y = a\left[c - (c-1)e^{-bx^d}\right]$$

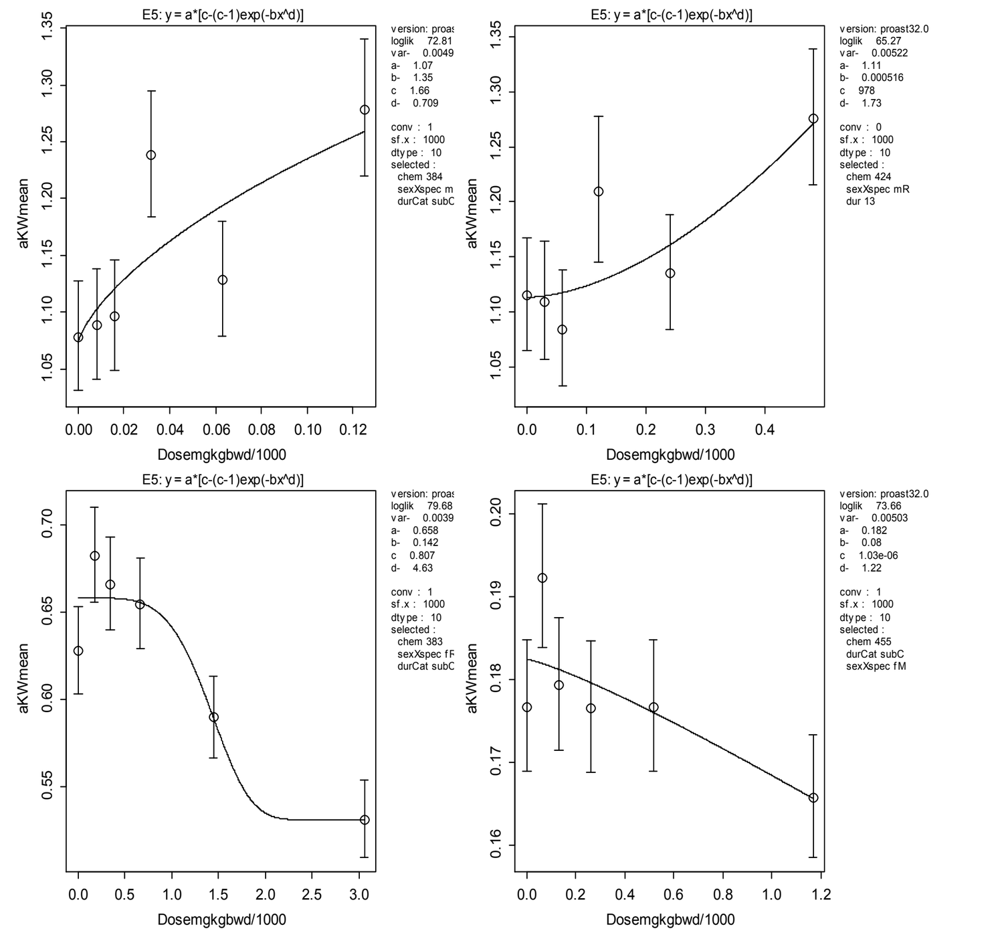
$$\text{Hill}: y = a\left[1 + (c-1)\frac{x^d}{b^d + x^d}\right]$$

Parameter Roles:

a scales *y*
b scales *x* (Hill) or 1 / $b^{1/d}$ scales *x* (Exponential)
c shape (maximum response)
d shape ('sigmoidicity')

We find both models have similar ability to describe the data, and focus on the Exponential model, fit using maximum likelihood, assuming a log-normal likelihood.

Goodness of fit tests of dose-response data using standard statistical methods can overenthusiastically reject fit. It is common to see evidence of dose-group level variability (in addition to the expected variability among animals in the same dose group), as demonstrated in the figure below. We augmented goodness-of-fit testing with direct examination of dose-response fits.



DATA SOURCES

The primary selection criterion for data sets was that there be more than the typical number of dose groups. This could either be an individual study with an exceptionally large number of doses, or a cluster of datasets relating to various toxicity studies that applied a comparable protocol to different compounds, resulting in comparable dose–response data for the same endpoint. There are six such clusters, as summarized in the following Table. For a given cluster of datasets, we analysed the endpoint(s) that were measured in all the constituent studies. Most had more than 5 dose groups.

Summary of clusters of datasets related to various chemicals. The numbers in the cells of the table indicate the number of dose–response datasets.

Cluster A: Subchronic NTP studies	Male rat	Female rat	Male mouse	Female mouse
BW	38	34	22	18
Rel. Liver weight	29	26	15	23
Kidney weight	18	15	5	6
Cluster B: OP ester studies	Male rat	Female rat		
AChE	16	16		
Cluster C: In vivo micronucleus test				
MN frequency	139			
Cluster D: LLNA test	Rubber chemicals		Low molecular weight chemicals	
proliferation	15	10		
Cluster E: WEC test	Lab 1	Lab 2	Lab 4	
Crown-rump length	13	13	12	
Cluster F: In vitro micronucleus test				
MN frequency	5			

Summary of additional datasets for single chemicals that fall outside the clusters of the previous table

Chemical	Endpoints	Number of doses	Sexes	Reference
BBP (developmental)	Foetal weights, PCO, progesterone	10	both	Piersma et al. (2000)
NDMA (carcinogenicity)	Time-to-tumour	16	both	Peto et al. (1991b)
Dichlorobenzene (28 days)	7 endpoints, e.g. Weights, Cholesterol, Prothrombin time	7	both	Appel (2001)
Rhodorsil Silane (28 days)	14 endpoints, e.g. weights, RBC, liver enzymes	7	females	Woutersen et al. (2001)
Silver nanoparticles (28 days)	Spleen weight, IgM	8	both	De Jong et al. (2013)

REFERENCES

MJ Appel (2001). Sub-acute (28-day) oral toxicity study with 1,4-dichlorobenzene in rats, TNO-report V3219, TNO Nutrition and Food research, Zeist, 2001.

W De Jong, FMM Van Och, CF Den Hartog Jager, WW Spiekstra, W Slob, RJ Vandebriel, H Van Loveren (2002). Ranking of Allergenic Potency of Rubber Chemicals in a Modified Local Lymph Node Assay. Toxicol Sci 2002 66: 226-232.

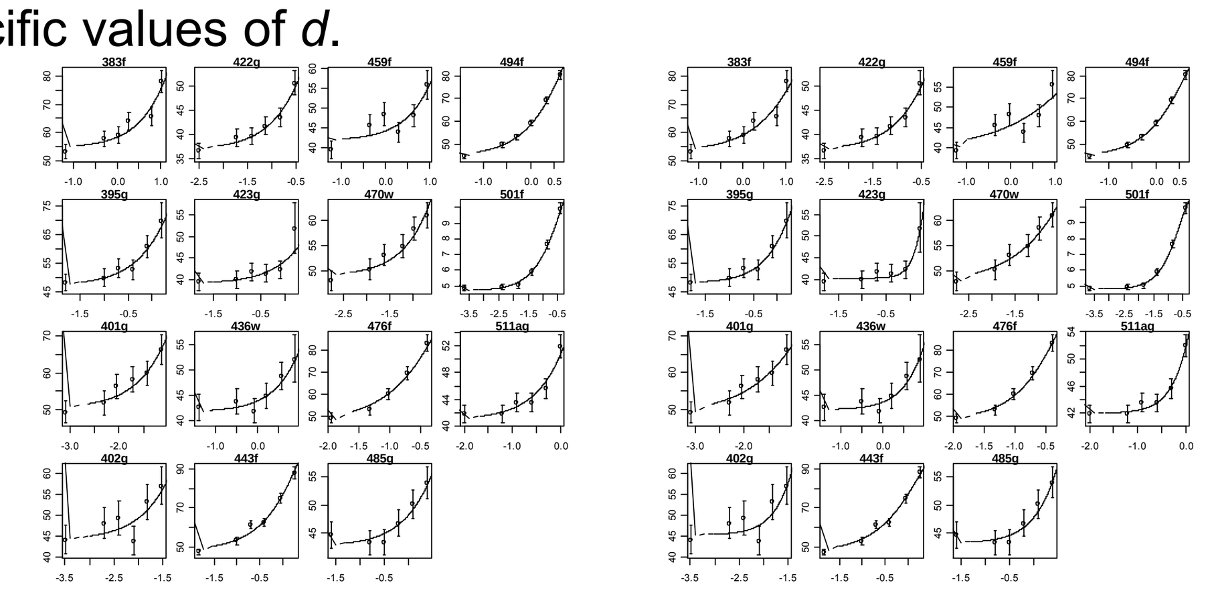
R Peto, R Gray, P Brantom, P Grasso (1991b). Dose and time relationships for tumor induction in the liver and esophagus of 4080 inbred rats by chronic ingestion of N-Nitrosodiethylamine or N-Nitrosodiemtylamine. Cancer Res 51: 6452 – 6469.

AH Piersma, A Verhoef, JD te Biesebeek, MN Pieters, W Slob (2000). Developmental toxicity of butyl benzyl phthalate in the rat using a multiple dose study design. Reprod Toxicol 14: 417-425

W Slob and RW Setzer (2014). Shape and steepness of toxicological dose-response relationships of continuous endpoints. Crit. Rev. Toxicol. 44(3): 270-297.

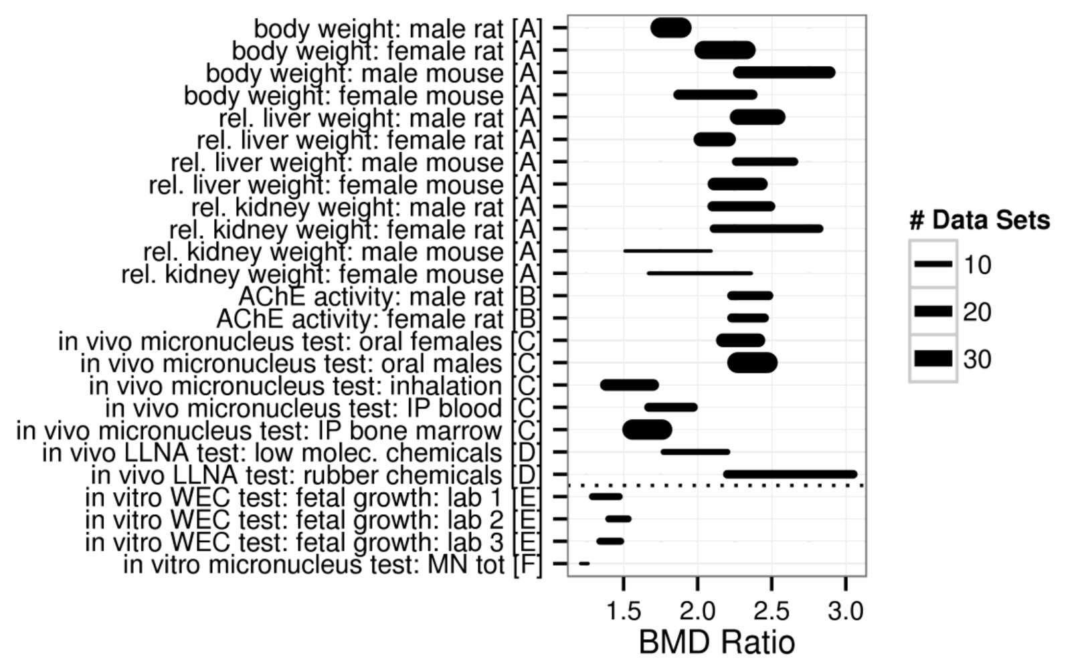
RESULTS

- In all datasets in all clusters the visual fit of the model, taking different background responses and potencies into account, was remarkably good (left panel, below, shows an example: liver weights as a function of dose in 15 subchronic NTP studies). There is statistical evidence that *d* varies somewhat among subgroups, but much of this could be the result of dose-group-level errors. The right panel below shows fits with subgroup-specific values of *d*.



- The estimate of *d* was frequently close to one (left figure below) particularly in *in vivo* studies. Higher values were found in the *in vitro* studies.

Summary of values for parameters *d* and *c* estimated by fitting the four-parameter exponential model to the clusters of datasets, assuming both parameters constant across chemicals in the cluster considered. The horizontal bars reflect the confidence intervals for parameters *c* and *d.* Note that the confidence interval for parameter *c* could not be estimated in some cases due to statistical limitations of the data



Summary of values for the BMD ratio estimated by fitting the four-parameter exponential model to the clusters of datasets, assuming both parameters constant across chemicals in the cluster considered. The horizontal bars reflect the confidence intervals for the BMD ratio

CONCLUSIONS

1. The shapes of toxicological dose-response relationships for continuous endpoints appear to be surprisingly homogeneous, well-described by a four-parameter exponential or Hill models.
2. Maximum response is similar across chemicals for the same endpoint, but can differ across endpoints.
3. Log-steepness was greater in *in vitro* than in *in vivo* studies.
4. These results imply that, for continuous data, choice of model for BMD modeling should not be a major issue; model uncertainty is generally a relatively minor component of BMD uncertainty.
5. The frequency with which estimates of *d* were close to one suggests that this parameter should not be constrained to exceed 1, as is often recommended for BMD modeling.
6. The narrow range of *c* for the same endpoint, and for *d* more generally, suggest that BMD modeling can be made more efficient for small (few doses) datasets by using this information, perhaps as strongly informative priors.