U.S. EPA Benchmark Dose Modeling Guidance

J. Allen Davis, MSPH

National Center for Environmental Assessment, U.S. EPA

March 2, 2017

Brussels, Belgium



Disclaimer

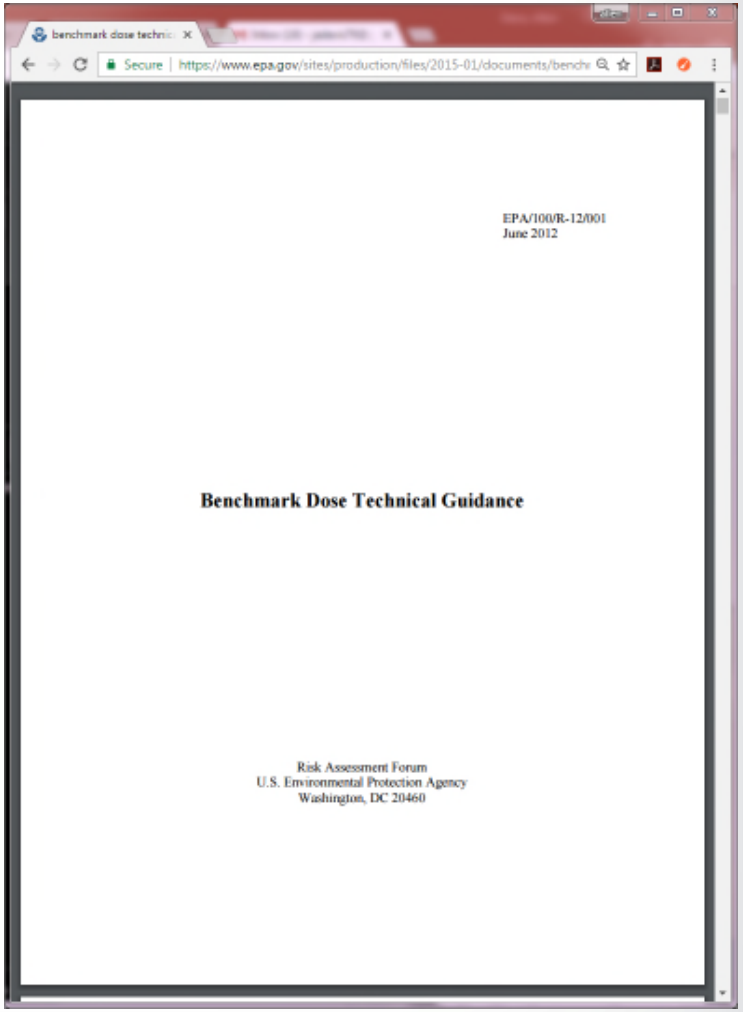
The views expressed in this presentation are those of the author(s) and do not necessarily reflect the views or policies of the US EPA.

EPA's BMD Technical Guidance

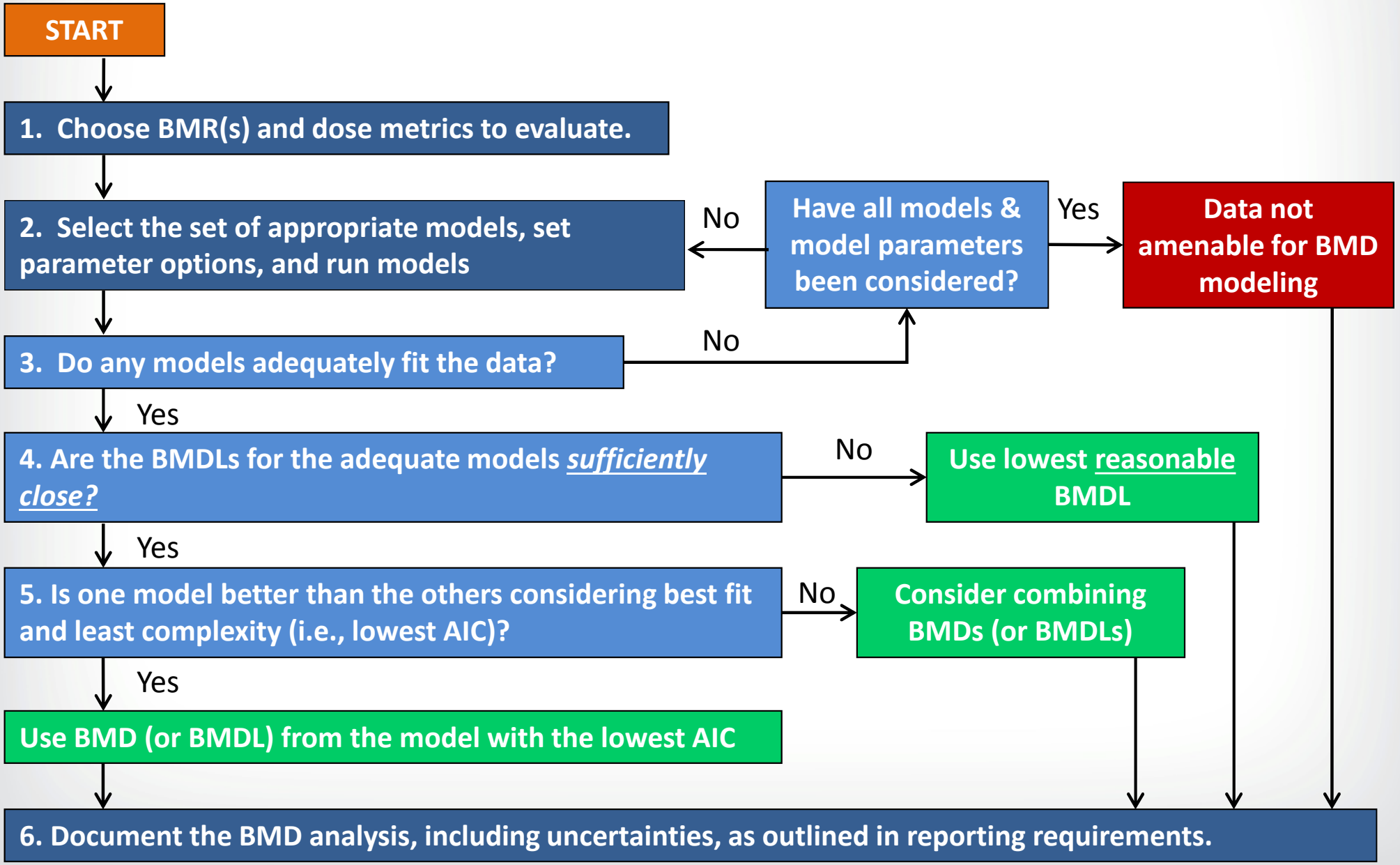
- *Final* version of the EPA's Benchmark Dose Technical Guidance document was published in 2012: https://www.epa.gov/risk/benchmark-dose-technical-guidance

- Other guidance documents relevant to BMD modeling available at: http://epa.gov/iris/backgrd.html

- EPA's Statistical Working Group periodically updates recommended model practices



BMD Analysis Workflow



BMR Selection: Dichotomous Data

- **An extra risk of 10% is recommended as a standard (not default) reporting level for dichotomous data.**
  - Customarily used because it is at or near the limit of sensitivity in most cancer bioassays and in non-cancer bioassays of comparable size
- **In some situations, use of different BMRs is supported**
  - Biological considerations sometimes support different BMRs (5% for frank effects, >10% for precursor effects)
  - When a study has greater than usual sensitivity, a lower BMR can be used (5% for developmental studies)
  - Results for a 10% BMR should always be shown for comparison when using different BMRs.

BMR Selection: Continuous Data

BMR Type	BMR Calculation
Standard Deviation:	$BMR = mean_0 \pm (BMRF \times SD_0)$
Relative Deviation:	$BMR = mean_0 \pm (BMRF \times mean_0)$
Absolute Deviation:	$BMR = mean_0 \pm BMRF$
Point:	$BMR = BMRF$
Extra (Hill only):	$BMR_{up} = mean_0 + BMRF \times (mean_{max} - mean_0)$ $BMR_{down} = mean_0 - BMRF \times (mean_0 - mean_{min})$

Where:

$mean_0$	=	*Modeled mean response at control dose*
SD_0	=	*Modeled standard deviation at control dose*
BMRF	=	*BMR factor (user input used to define BMR)*
$mean_{max}$	=	*Maximum mean response in dataset*
$mean_{min}$	=	*Minimum mean response in dataset*

BMR Selection: Continuous Data

- Preferred approach is to select a BMR that corresponds to a level change that represents a minimal biologically significant response (i.e., 10% decrease in body weight)

- In the absence of a biological consideration, a BMR of a change in the mean equal to one control standard deviation (1.0 SD) from the control mean is recommended.

- In some situations, use of different BMRs is supported
  - For more severe effects, a BMR of 0.5 SD can be used
  - Results for a 1 SD BMR should always be shown for comparison when using different BMRs.

Selection of a Specific Model

Biological Interpretation	Examples:
	• Dichotomous:
	  • Saturable processes demonstrating Michaelis-Menten kinetics (Dichotomous Hill model)
	  • Two-stage clonal expansion model (cancer endpoints)
	• Continuous:
	  • Can use the Hill or Exponential models for receptor-mediated responses

Policy Decision	• U.S. EPA's IRIS program uses the multistage model for cancer data (i.e., dichotomous data)
	  • sufficiently flexible to fit most cancer bioassay data
	  • provides consistency across cancer assessments
	• U.S. EPA's OPP group uses the Exponential models for modeling acetylcholinesterase inhibition data

Otherwise	However, in the absence of biological or policy-driven considerations, criteria for final model selection are usually based on whether various models mathematically describe the data

Traditional Dichotomous Models

Model name	Functional form	# of Parameters[a]	Low Dose Linearity	Model fits
Multistage	$\gamma + (1-\gamma)\left[1 - \exp\left\{-\sum_{j=1}^{k} \beta_j X^j\right\}\right]$	1+k	Yes, if $\beta_1 > 0$ No, if $\beta_1 = 0$	All purpose
Logistic	$\frac{1}{1+\exp\{-(\alpha+\beta X)\}}$	2	Yes	Simple; no background
Probit	$\Phi(\alpha+\beta X)$	2	Yes	Simple; no background
Log-logistic	$\frac{\gamma+(1-\gamma)}{1+\exp\{-[\alpha+\beta\ln(X)]\}}$	3	No	All purpose; S-shape with plateau at 100%
Log-probit	$\gamma+(1-\gamma)\,\Phi\{\alpha+\beta\ln(X)\}$	3	No	All purpose; plateau S-shape with plateau at 100%
Gamma	$\gamma+(1-\gamma)\left[\int_0^{\beta x} t^{\alpha-1}e^t dt\right] / \Gamma(\alpha)$	3	No	All purpose
Weibull	$\gamma+(1-\gamma)[1-\exp\{-\beta X^{\alpha}\}]$	3	No	”Hockey stick” shape
Dichotomous Hill	$v\times g+\frac{(v-v\times g)}{1+\exp\{-a-b\times\ln(X)\}}$	4	Yes	Symmetrical, S-shape with plateau

[a] Background parameter = γ. Background for hill model = v × g

Continuous Model Forms

Model Name	Functional Form	# of Parameters	Model Fits
Polynomial[a]	$\beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_n X^n$	1 + n	All purpose, can fit non-symmetrical S-shaped datasets with plateaus
Power	$\gamma + \beta X^{\Phi}$	3	L-shaped
Hill	$\gamma + \frac{(v \times X^n)}{(k^n + X^n)}$	4	Symmetrical, sigmoidal, S-shape with plateau
Exponential[b] Model 2 Model 3 Model 4 Model 5	$a \times \exp\{\pm 1 \times b \times X\}$ $a \times \exp\{\pm 1 \times (b \times X)^d\}$ $a \times [c - (c-1) \times \exp\{\pm 1 \times b \times X\}]$ $a \times [c - (c-1) \times \exp\{\pm 1 \times (b \times X)^d\}]$	 2 3 3 4	All purpose (Models 2 & 3) Symmetrical and asymmetrical S-shape with plateau (Models 4 & 5)

[a] The stand-alone Linear model in BMDS is equal to a first-order polynomial model

[b] Nested family of 4 related models described by Slob (2002) and included in the PROAST software of RIVM

Restricting Model Parameters

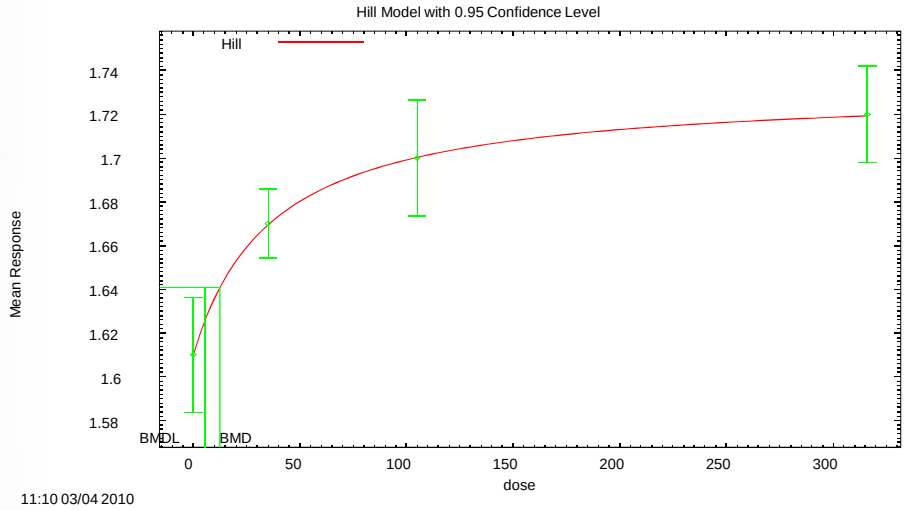
- Model parameters (i.e., slope, background response, etc.) can be bounded to restrict the shape of the dose-response curve
- These restrictions can impact statistical calculations such as the goodness-of-fit p-value and AIC
  - Currently, a parameter estimate that "hits a bound" impacts a model's degrees of freedom (DF) (in BMDS, DF is increased by 1 for *p*-value calculation)
  - When a parameter hits a bound, that parameter is not counted towards the AIC penalization (EPA's Statistical Working Group may modify this approach in the future)
- The use of model restrictions is a topic of ongoing discussion in EPA's Statistical Working Group

Does the Model Fit the Data?

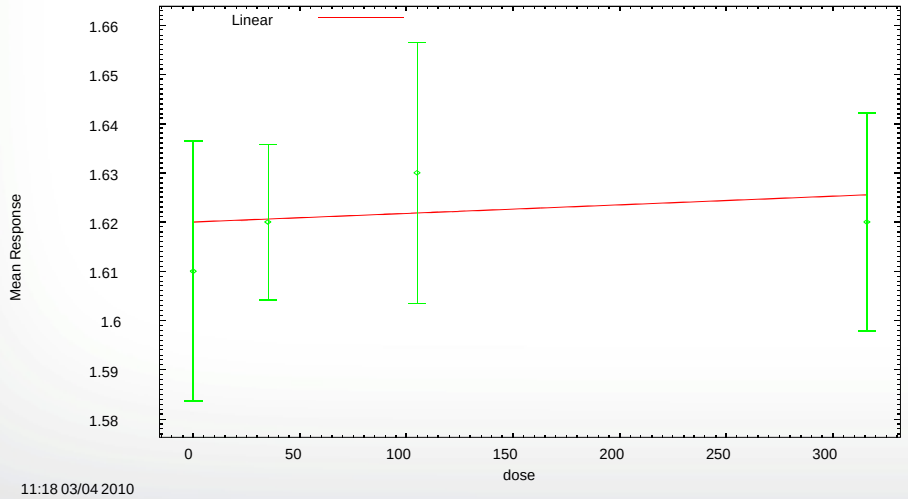
- Tests of interest (response/variance modeling) (continuous models only)
- Global measurement: goodness-of-fit p value ($p > 0.1$)
- Local measurement: Scaled residuals (absolute value < 2.0)
- Visual inspection of model fitting.

Tests of Interest – Differences in Responses/Variances

- Test 1 – Do responses and/or variances differ among dose levels?



The p-value for Test 1 is less than .05. There appears to be a difference between response and/or variances among the dose levels. It seems appropriate to model the data



The p-value for Test 1 is greater than .05. There may not be a difference between responses and/or variances among the dose levels. Modeling the data with a dose/response curve may not be appropriate

Tests of Interest – Variance

- In the current version BMDS, the distribution of continuous measures is assumed to be normal, with either a constant (homogenous) variance or a variance that changes as a power function of the mean value
  - Var(i) = α[mean(i)]$^{\rho}$
  - ρ(rho) = 0, constant variance
  - ρ(rho) $\neq$ 0, modeled variance
- Test 2 – Are variances homogenous?
- Test 3 – Are variances adequately modeled?
- Recommendation is to assume constant variance unless data clearly indicate otherwise

Global Goodness-of-Fit

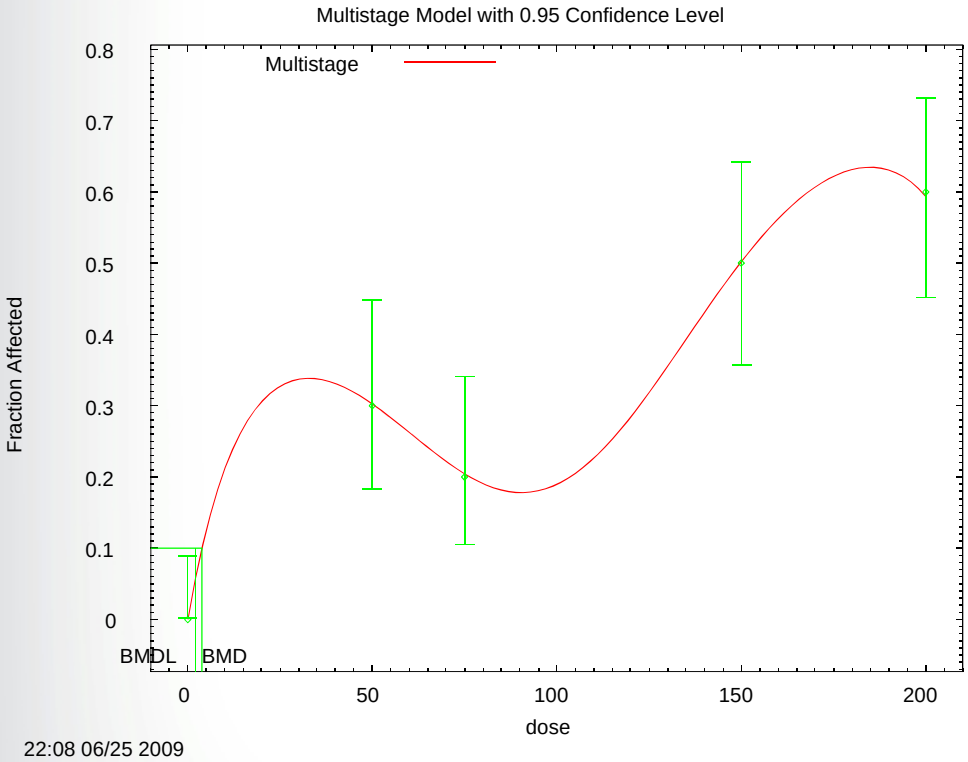
- **BMDS provides a *p*-value to measure global goodness-of-fit**
  - Measures how model-predicted dose-group probability of responses differ from the actual responses
  - Small values indicate poor fit
  - Recommended cut-off value is $p = 0.10$
  - For models selected *a priori* due to biological or policy preferences (e.g., multistage model for cancer endpoints), a cut-off value of $p = 0.05$ can be used
- **What to do when goodness-of-fit is poor?**
  - Consider dropping high dose group(s) that negatively impact low dose fit
  - Use PBPK model if available to calculate internal dose metrics that may facilitate better fit
  - Log-transform doses if appropriate

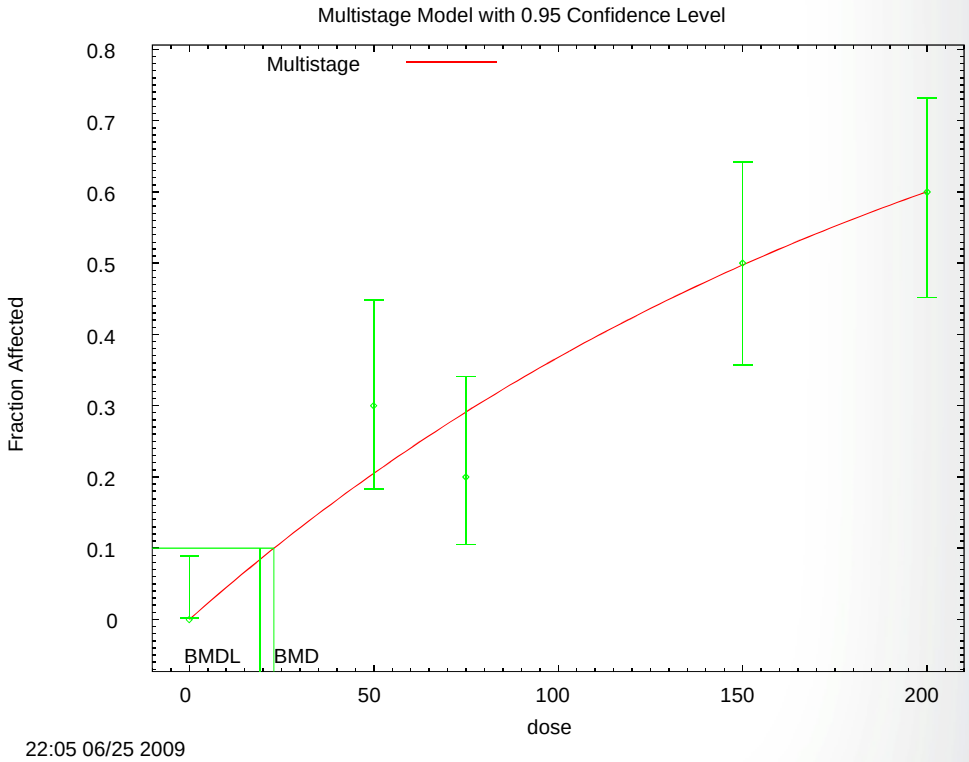
Scaled Residuals

- **Global goodness-of-fit p-values are not enough to assess local fit**
  - Models with large p-values may consistently "miss the data" (e.g., always on one side of the dose-group means)
  - Models may "fit" the wrong (e.g. high-dose) region of the dose-response curve.

- **Scaled Residuals – measure of how closely the model fits the data at each point; 0 = exact fit**
  - Dichotomous data: $\frac{Obs - Exp}{\sqrt{(n * p(1-p))}}$
  - Continuous data: $\frac{Obs\ Mean - Est\ Mean}{\frac{Est\ SD}{\sqrt{n}}}$
  - Absolute values near the BMR should be lowest
  - Question scaled residuals with absolute value > 2

Visual Inspection of Fit





Are BMDL Estimates "Sufficiently Close"?

- **Often, more than one model or modeling options will result in an acceptable fit to the data, current EPA guidance (2012 TG) is based upon picking a single "best" model**
  - What is "sufficiently close" can vary based on the needs of the assessment, but generally should not be more than 3-fold.
  - If BMDLs are not sufficiently close, *EPA recommends picking the model with the lowest BMDL*
  - If BMDLs are sufficiently close, *EPA recommends selecting the model with the lowest AIC*
  - If multiple models have the same AIC, *EPA recommends combining BMDLs*

Example of BMD Analysis Documentation - Dichotomous

Table B-9. Benchmark dose modeling results for decreased rotorod performance in male Wistar rats exposed to 1,2,4-TMB. (Korsak and Rydzyński, 1996)

Model[a]	Goodness-of-fit		$BMD_{10\%}$	$BMDL_{10\%}$	Basis for Model Selection
	p-value	AIC			
Logistic	0.6024	35.5306	528.905	341.987	Of the models that provided an adequate fit and valid BMDL estimate, the log-logistic model was selected based on the lowest BMDL (BMDLs differed by more than 3-fold).
Log-logistic	**0.9743**	**32.1664**	**193.575**	**93.947**	
Log-probit	0.5825	35.4276	426.494	232.739	
Probit	0.6248	35.4027	489.595	317.868	
Dichotomous Hill	0.9352	34.1023	160.508	--	
Gamma **Weibull** **Linear** **Multistage 2°** **Multistage 3°**	0.9338	32.3299	228.574	129.306	

[a] Decreased rotorod performance was measured as increased percentage of failures per rat, selected model in bold; scaled residuals for selected model for concentrations 0, 123, 492, and 1230 mg/m^3 were 0.000, 0.434, -0.154, -0.089, respectively

Example of BMD Analysis Documentation - Continuous

Table C-12. Summary of BMD modeling results for increased reticulocytes in male Wistar rats exposed to 1,2,3-TMB by inhalation for 3 months; BMR = 1 SD change from control mean(constant variance), (Korsak et al., 2000b)

Model[a]	Goodness-of-fit		BMD_{1SD} (mg/m^3)	$BMDL_{1SD}$ (mg/m^3)	Basis for Model Selection
	p-value	AIC			
Exponential (M2)[b] Exponential (M3)	0.2733	89.08418	1112.25	806.744	Of the models that provided an adequate fit and valid BMDL estimate, the Linear model was selected based on the lowest AIC (BMDLs differed by less than 3-fold).
Exponential (M4)	0.1397	90.67033	900.404	308.017	
Exponential (M5)	n/a[c]	91.37006	540.186	140.925	
Hill	n/a[c]	91.370061	554.848	Not calculated	
Linear[d] Polynomial 2° Polynomial 3° Power	**0.3105**	**88.828645**	**1025.1**	**652.898**	

[a] Constant variance case presented (Test 2 *p*-value = 0.5223). Selected model in bold; scaled residuals for selected model for concentrations 0, 128, 523 and 1,269 mg/m^3 were 0.555, -1.14, 0.793, and -0.212, respectively.

[b] For Exponential model 3, the estimate of d was 1 (boundary). The models in this row reduced to exponential model 2.

[c] χ^2 test had insufficient degrees of freedom (due to estimated model parameters = dose groups). Inspection of scaled residuals indicated appropriate model fit. However, inspection of visual fit indicated uncertain dose-response characteristics, and therefore, these models were excluded from consideration.

[d] For the power model, the power parameter estimate was 1 (boundary). For the polynomial 2° and 3° models, the b2 and b3 coefficient estimates were 0 (boundary). The models in this row reduced to the Linear model.

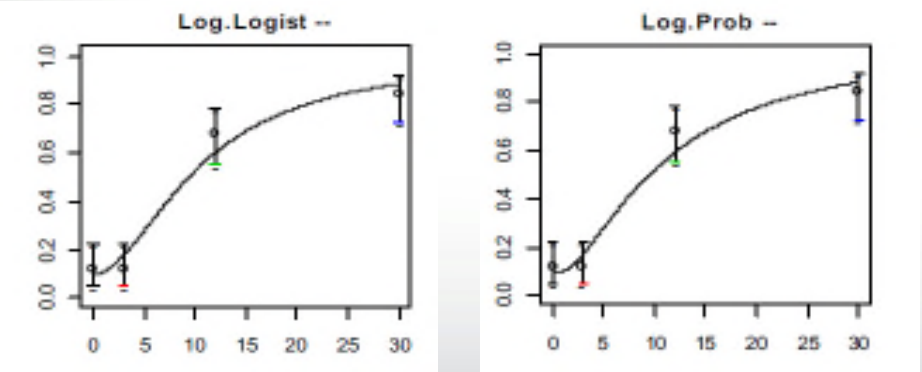
Data Source: (Korsak et al., 2000b).

Example: EFSA Dichotomous Model Selection

Dose (mg/kg day)	No of animals with thyroid epithelial vacuolisation	No of animals in dose group	Sex
0	6	50	F
3	6	50	F
12	34	50	F
30	42	50	F

Model	No of parameters	AIC	$BMDL_{10}$	$BMDU_{10}$[(a)]
Null Model	1	276.38	–	–
Gamma	3	192.99	1.21	2.67[(a)]
Logistic	2	198.47	3.31	4.90[(a)]
Log-Logistic	**3**	189.81[(b)]	1.84	5.00[(a)]
Probit	2	199.07	3.37	NA
Log-Probit	**3**	189.73[(c)]	1.98	5.11[(a)]
Weibull	3	193.55	1.10	4.01[(a)]
LMS (Two stage)	3	194.20	1.35	3.10
Full Model	4	188.04	–	–

AIC: Akaike information criterion; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose.
(a): Calculated by PROAST, as BMDS does not yet provide BMDUs except for the two-stage model.
(b): AIC differs less than two units from lowest AIC.
(c): Model with lowest AIC.

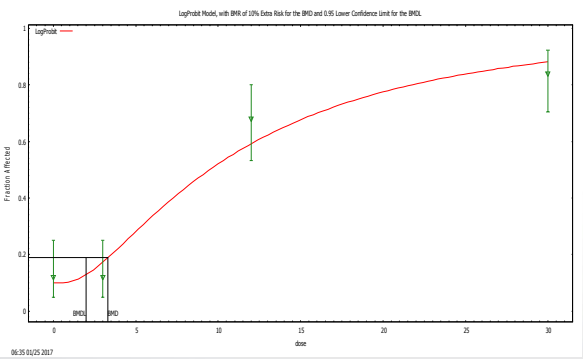


- All AIC > AIC_{null} – 2 (274.38)? No
- AIC_{min} = 189.73
- AIC_{min} > AIC_{full} + 2 (190.04)? No
- AIC_{min} + 2 = 191.73
- Models with AIC ≤ AIC_{min} + 2 = log-logistic and log-probit
- Lowest BMDL = 1.84 (log-logistic)
- Largest BMDU = 5.11 (log-probit)

Example: EPA Dichotomous Model Selection

Dose (mg/kg day)	No of animals with thyroid epithelial vacuolisation	No of animals in dose group	Sex
0	6	50	F
3	6	50	F
12	34	50	F
30	42	50	F

Model	# parameters	p-value	AIC	BMD	BMDL
Gamma	3	0.0095	192.99	2.688	1.422
Logistic	2	0.0007	198.47	4.029	3.313
Log-logistic	3	0.0535	189.81	3.200	1.841
Probit	2	0.0006	199.07	4.003	3.370
Log-probit	3	0.0566	189.73	3.313	1.976
Weibull	3	0.0076	193.55	2.291	1.383
2° Multistage	3	0.0063	194.20	1.684	1.346



- No model has goodness-of-fit p-value > 0.10
- Relaxing the requirement to p > 0.05: log-logistic and log-probit fit adequately
- No scaled residual > |2|
- Visual inspect is OK
- BMDLs within a range of 3 (1.841 and 1.976)
- Model with lowest AIC = log-probit

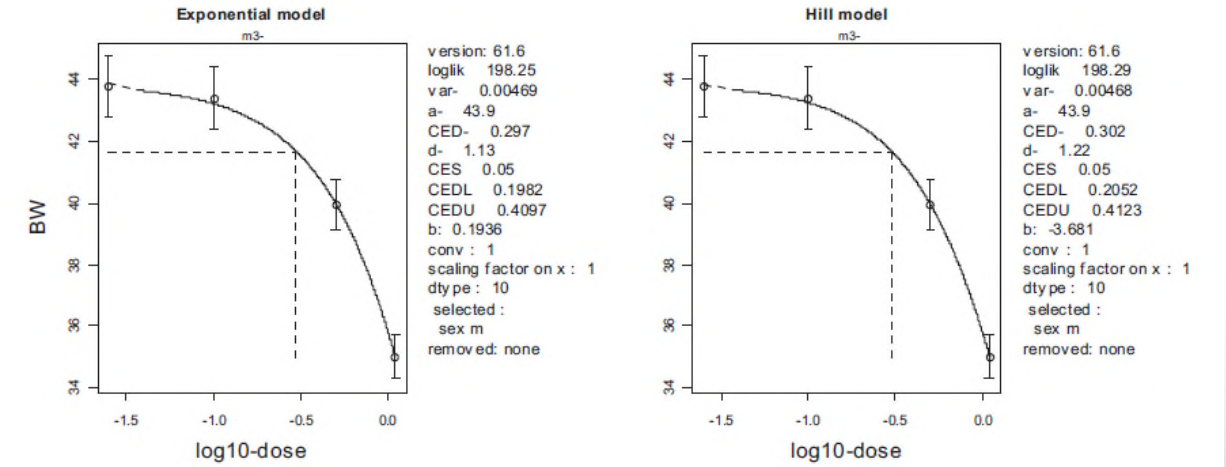
Example: EFSA Continuous Model Selection

Dose (mg/kg bw per day)	Body weight, group mean (g)	SD	n	Sex
0	43.85	2.69	37	M
0.1	43.51	2.86	35	M
0.5	40.04	3.00	43	M
1.1	35.09	2.56	42	M

bw: body weight; SD: standard deviation.

Model	No. of parameters (variance excluded)	AIC		$BMDL_{05}$ (mg/kg)		$BMDU_{05}$ (mg/kg)	
		Exponential	Hill	Exponential	Hill	Exponential	Hill
Null model	1	−234.06					
Model 3	3	−388.50[(a)]	−388.58[(a)]	0.198	0.205	0.410	0.412
Model 5	4	−386.72	−386.72				
Full model	4	−386.72					

AIC: Akaike information criterion; BMDL: lower confidence limit of the benchmark dose; BMDU: upper confidence limit of the benchmark dose.
(a): Selected model, based on lowest AIC.



- All AIC > AIC_{null} – 2 (-236.06)? No
- Model 3 for Exponential and Hill models chosen based on lowest AIC
- Lowest BMDL = 0.198 (Exponential)
- Largest BMDU = 4.12 (Hill)

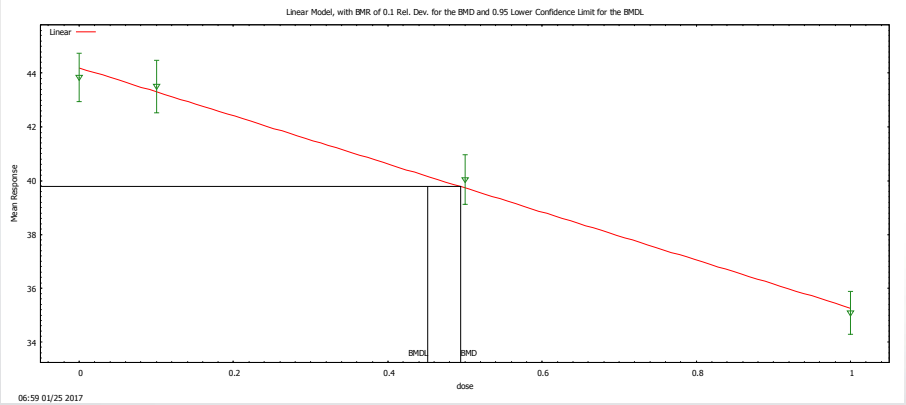
Example: EPA Continuous Model Selection

Dose (mg/kg bw per day)	Body weight, group mean (g)	SD	n	Sex
0	43.85	2.69	37	M
0.1	43.51	2.86	35	M
0.5	40.04	3.00	43	M
1.1	35.09	2.56	42	M

bw: body weight; SD: standard deviation.

Model	# parameters	p-value	AIC	BMD	BMDL
Exp2	2	0.2512	483.14	0.229	0.207
Exp3	3	0.7957	482.44	0.320	0.225
Exp4	3	0.2515	483.14	0.229	0.198
Exp5	4	n/a	484.38	0.338	0.226
Hill	4	n/a	484.38	0.337	0.225
Linear	2	0.474	481.87	0.247	0.226
Poly2	3	0.574	482.69	0.330	0.231
Power	3	.721	482.50	0.314	0.232

Test 2 & 3 = 0.752 indicating constant variance



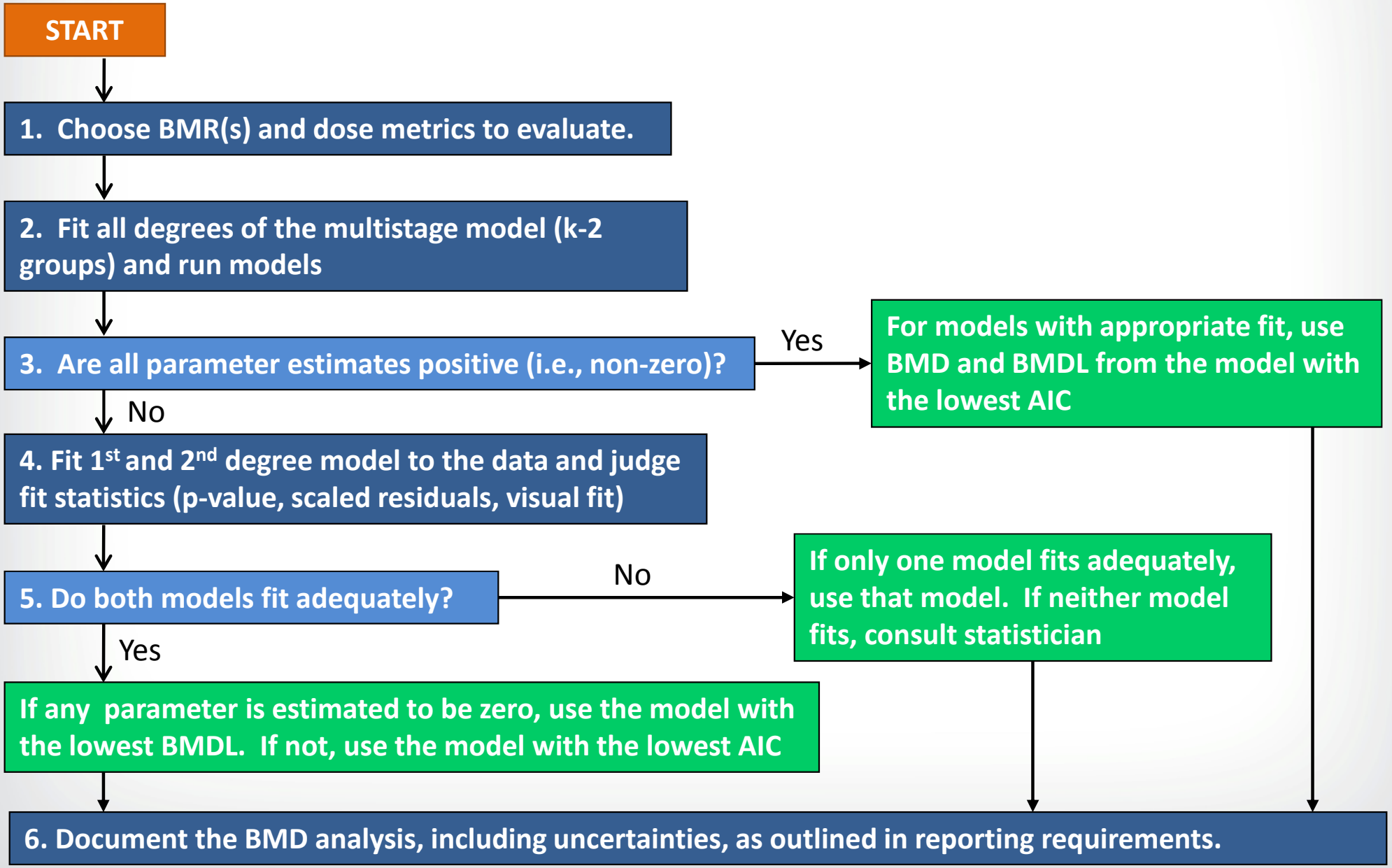
- All tests of fit (variance, global goodness-of-fit, and scaled residuals) indicate all models fit the data
- Exp5 and Hill model have no degrees of freedom to calculate p-value as # dose groups = # estimated parameters
- Linear model selected due to lowest AIC (BMDLs for all models within a range of 3)
- BMD = 0.247; BMDL = 0.226
- Compare to EFSA results: BMDL = 0.198

Updating EPA modeling procedures

- **EPA modifies its modeling procedures when necessary based on sound science without updating formal technical guidance**
  - Updated modeling procedures & recommendations conveyed to the public via publications, chemical assessments, or meeting presentations

- **Examples include:**
  - A new analysis workflow for modeling cancer data with the multistage model
  - Approximation methods to account for litter effects when only summary statistics are available
  - Use of AIC weights in model selection
  - Use of historical controls
  - Variance lack of fit

BMD Cancer Analysis Workflow



Modified Cancer Example

Dose	Incidence	N
0	0	71
1.1	4	73
6.1	5	73
12.9	11	71
28.7	31	67

Model order[a]	Goodness of fit				$BMD_{10\%}$ (mg/kg-d)	$BMDL_{10\%}$ (mg/kg-d)
	p-value	Scaled residuals	AIC	Coefficients *		
Three df=2	0.1602	-1.007 1.584 -0.321 -0.189 0.047	231.261	γ = 0.0140719 β_1 = 0.01066274 β_2 = 0 β_3 =1.25521E-005	9.04	5.32
Two df=2	0.1350	-1.095 1.596 -0.270 -0.389 0.187	231.771	γ =0.016603 β_1 = 0.00768476 β_2 = 0.00044062	9.03	5.01
One df=3	0.0717	-0.456 1.870 -1.047 -1.118 0.977	231.647	γ = 0.00291998 β_1 = 0.0179367	5.87	4.62

* From cancer models fitted in BMDS 2.40 by J. Allen Davis, 15 May 2014

The β_2 parameter for the 3° model is estimated on the boundary

Therefore, only the 1° and 2° models are considered further

Both models fit the data adequately ($p > 0.05$, scaled residuals $< |2|$)

As both models fit adequately AND no parameter for these models is on a boundary, the model with the lowest AIC is chosen

Therefore, the 1° model is selected using new workflow

3° model would've been chosen with old workflow based on

Modeling Developmental Toxicity Data with Summary Statistics

- When modeling developmental toxicity data (i.e., nested dichotomous data, clustered data), litter effect must be accounted for
  - Litter effect refers to the propensity of litter-mates to respond more alike one another compared to offspring from different litters
  - Not taking clustering into account leads to underestimated variances and higher BMDLs

- When individual dam data is available, litter effects (specifically intra-litter correlation) can be accounted for by use of BMDS' nested dichotomous models; but what to do when individual animal data is not available?

Modeling Developmental Toxicity Data with Summary Statistics

Risk Analysis DOI: 10.1111/risa.12867

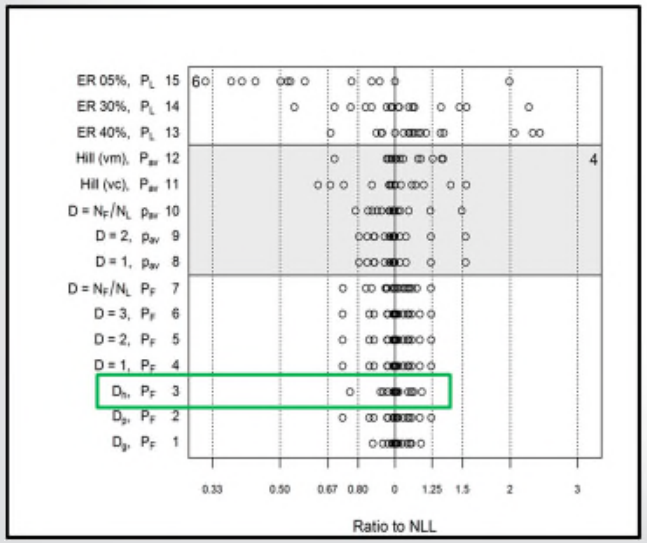
Dose-Response Modeling with Summary Data from Developmental Toxicity Studies

John F. Fox,* Karen A. Hogan, and Allen Davis

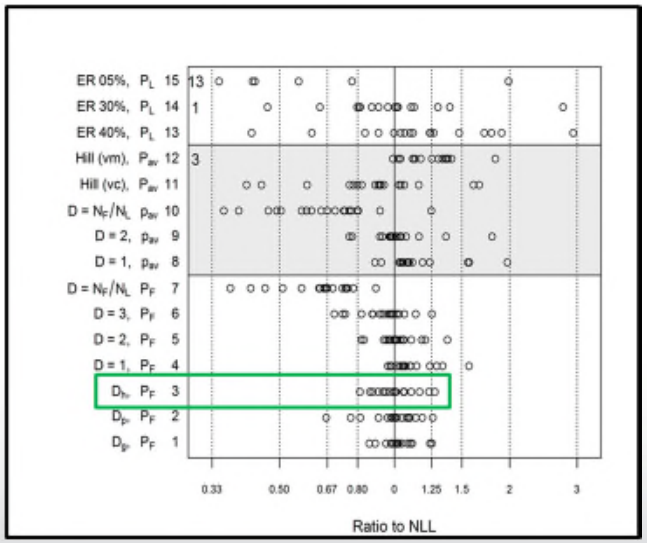
Table 1. Linear least squares (LS) and orthogonal regression (OR) estimates by species, for the relation $\log_e(D) = a + b*\log_e(P_F)$, for cases with $P_F > 0$.[a]

	Method	Studies	n, Dose groups	a	b	σ_{res}^2
Mice	LS	21	88	1.5938	0.2866	0.2078
Mice	OR	21	88	1.6943	0.3132	0.1863
Rats	LS	25	101	1.6852	0.3310	0.1248
Rats	OR	25	101	1.8327	0.3690	0.1090
Rabbits	LS	10	43	1.0582	0.2397	0.1452
Rabbits	OR	10	43	1.1477	0.2739	0.1299

BMD estimates



BMDL estimates



- Applying Rao-Scott Transformation (give a value of D):
  - Use dose-group totals for offspring: number of offspring (N_F), number of affected fetuses (X_F)
  - Divide number of offspring (N_F) by D
  - Divide number of affected fetuses (X_F) by D
- Rao-Scott transformed data can now be modeled with BMDS dichotomous models

Use of AIC Weights in Model Selection

- EPA's Statistical Working Group is proposing using AIC weights for model selection
  - $AIC\ difference\ (\Delta_i) = AIC_i - AIC_{min}$
  - $AIC\ weight\ (w_i) = \frac{\exp(-\Delta_i/2)}{\sum_{r=1}^{R}\exp(-\Delta_r/2)}$
  - One proposal is to exclude models with $w_i < 0.10$

Model Type	Test 4 p-value	AIC	AIC weight	BMD	BMDL
Gamma	0.896	101.89	0.085	197	101
Dichotomous Hill	0.606	102.98	0.050	235	133
Logistic	0.992	99.721	0.252	159	120
Log-logistic	0.606	102.98	0.050	235	133
Probit	0.996	99.627	0.264	146	113
Log-probit	0.669	102.63	0.059	223	133
Weibull	0.985	101.61	0.098	179	93.1
Multistage 3°	0.895	103.59	0.037	167	65.0
Multistage 2°	0.988	101.61	0.098	174	74.9
Quantal-Linear	0.131	106.79	0.007	55.1	42.9

- Using existing methods: quantal-linear model would be selected
  - BMDLs differed by > 3-fold
  - BMDL = 42.9
- Using modified methods: Probit model selected
  - Only Probit and Logistic model had w_i >0.10
  - Probit model selected (lowest AIC)
  - BMDL = 113
  - 2.6-fold higher than