Interpreting the results of EPA dose-response models

Jeff Gift, PhD
J. Allen Davis, MSPH
National Center for Environmental Assessment, U.S. EPA
October 24, 2017
Peer Review of Draft NTP Approach to Genomic Dose-Response Modeling
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.
Outline

- Existing BMD modeling approaches for continuous data
- Future directions for BMDS
- NTP's BMDE express
Final version of the EPA’s Benchmark Dose (BMD) 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 Key Steps

1. **Benchmark Response (BMR):** Choose BMR(s) to evaluate.

2. **Model Selection:** Run appropriate models and parameter options.

3. **Model Fit:** Determine which models adequately fit the data.

4. **BMDLs:** Are BMDLs for models that fit *sufficiently close* (< 3-fold)?

5. **Akaike Information Criterion (AIC):** If so, use model with lowest AIC; if not use model with lowest BMDL
BMR: Continuous Data

- Preferred approach: Use a benchmark response (BMR) that corresponds to a level of change representing a minimal biologically significant response (i.e., 10% decrease in body wt.)

- 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 cases, use of different BMRs is supported (e.g., 0.5 SD for changes in critical organ systems)
### Continuous Model Selection

<table>
<thead>
<tr>
<th>Biological Interpretation</th>
<th>Example: Hill or Exponential models can be used for receptor-mediated responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy Decision</td>
<td>U.S. EPA’s OPP group uses the Exponential models for modeling acetylcholinesterase inhibition data</td>
</tr>
<tr>
<td>Otherwise</td>
<td>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</td>
</tr>
</tbody>
</table>
# Continuous Model Forms

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Functional Form</th>
<th># of Parameters</th>
<th>Model Fits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial(^a)</td>
<td>(\beta_0 + \beta_1X + \beta_2X^2 + \ldots + \beta_nX^n)</td>
<td>(1 + n)</td>
<td>All purpose, can fit non-symmetrical S-shaped datasets with plateaus</td>
</tr>
<tr>
<td>Power</td>
<td>(\gamma + \beta X^\Phi)</td>
<td>3</td>
<td>L-shaped</td>
</tr>
<tr>
<td>Hill</td>
<td>(\gamma + \frac{(v \times X^n)}{(k^n + X^n)})</td>
<td>4</td>
<td>Symmetrical, sigmoidal, S-shape with plateau</td>
</tr>
<tr>
<td>Exponential(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>(a \times \exp{\pm 1 \times b \times X})</td>
<td>2</td>
<td>All purpose (Models 2 &amp; 3) Symmetrical and asymmetrical S-shape with plateau (Models 4 &amp; 5)</td>
</tr>
<tr>
<td>Model 3</td>
<td>(a \times \exp{\pm 1 \times (b \times X)^d})</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>(a \times [c - (c - 1) \times \exp{\pm 1 \times b \times X}])</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td>(a \times [c - (c - 1) \times \exp{\pm 1 \times (b \times X)^d}])</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) 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.) are generally bounded to restrict the dose-response curve to “reasonable” shapes

• These restrictions can impact statistical calculations such as the goodness-of-fit p-value and Akaike AIC

• 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)
- Global measurement: \( p > 0.1; p > 0.05 \) for preselected models
- Local measurement: Scaled residuals (absolute value < 2)
- Visual inspection of model fitting.
- BMD/BMDL: caution if > 5; serious concern if > 20 (Wizard)
- Use of AIC or BMDL range to choose between adequately fitting models
Tests of Interest – Differences in Responses and/or 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

- 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
  - $\text{Var}(i) = \alpha [\text{mean}(i)]^\rho$
  - $\rho(\text{rho}) = 0$, constant variance
  - $\rho(\text{rho}) \neq 0$, modeled variance
- Test 2 – Are variances homogenous?
- Test 3 – Are variances adequately modeled?
- Recommendation is to assume constant variance unless data clearly indicate otherwise
Goodness-of-Fit

• Global - BMDS provides \( p \)-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
  • Cut-off value is \( p = 0.10 \) (0.05 when selecting a model *a priori*)

• Local - Scaled Residuals measure the *local* fit of the model at each point; 0 = exact fit
  • Continuous data: \( \frac{\text{Obs Mean} - \text{Est Mean}}{\frac{\text{Est SD}}{\sqrt{n}}} \)
  • Absolute values near the BMR should be lowest
  • Question scaled residuals with absolute value > 2
Visual Inspection of Fit

Multistage Model with 0.95 Confidence Level

22:08 06/25 2009

BMDBMDL

Multistage

Fraction Affected

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

0
50
100
150
200

dose

22:05 06/25 2009

BMDBMDL

Multistage

Fraction Affected

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

0
50
100
150
200

dose
What to do when goodness-of-fit is poor?

- Use a different model
- Consider dropping high dose group(s)
- Use alternative dose metric (e.g., internal doses from PBPK model)
- Log-transform doses if appropriate
Are BMDL Estimates “Sufficiently Close”?

- What is “sufficiently close” can vary based on 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**
• New alternative approaches to the “best model” method:
  • Model Averaging – Bayesian or frequentist
  • Use of single, hyper-flexible model (parametric or non-parametric)
  • Preliminary tests for continuous model averaging methods indicate coverage is better than “best model” approach

NTP Proposed Approaches to Genomic Dose-Response Modeling
Four Step Process

1. Filtering the measured features (genes/probe sets)
2. Fitting filtered features to dose-response models
   a. Selecting and running models
   b. Evaluating model fit
3. Parsing the features into predefined gene sets (e.g., Gene Ontology Biological Processes); and
4. Determining potency for each of the adequately populated gene sets by deriving the mean and median potency of the genes in each set.

### Step 1: Filtering the measured features

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way ANOVA, p &lt; 0.05 and fold-change (empirically determined for each technology)</td>
<td>Trend test generally preferred; may exclude more datasets than ANOVA test</td>
</tr>
<tr>
<td>Trend test not recommended because “biologically meaningful, non-monotonic responses can be observed that would be removed by a simple trend test.”</td>
<td>Non-monotonic responses considered unlikely for typical BMD endpoints</td>
</tr>
</tbody>
</table>
### Step 2a: Selecting and Running Models

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hill, power, Linear, Polynomial $2^\circ$ and $3^\circ$, Exponential 2, 3, 4 and 5</td>
<td>Same</td>
</tr>
<tr>
<td>Response data are log2 adjusted and presumed to have normal distribution</td>
<td>Normal or lognormal response distribution allowed for Exponential models (for all models in next BMDS version)</td>
</tr>
<tr>
<td>Model runs assume constant variance (future version to test and compare fit of constant and non-constant variance models)</td>
<td>Compare fit of constant and non-constant variance models; Choice of variance model can impact BMD/BMDL estimation</td>
</tr>
</tbody>
</table>
### NTP Proposed Approach to Genomic Dose-Response Modeling

#### Step 2a: Selecting and Running Models (continued)

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMR = 1.349 X SD at 0 dose; ~10% Δ in transcript abundance vs control</td>
<td>BMR set to biologically relevant change or 1.0 SD by default.</td>
</tr>
<tr>
<td>SD estimated from entire fitted curve, not just the control data</td>
<td>Same; important to consider model estimate of control SD</td>
</tr>
<tr>
<td>For most models, adverse direction determined by BMDS model</td>
<td>Adverse direction determined by BMDS or set by user</td>
</tr>
<tr>
<td>Adverse direction of Polynomial 2° and 3° models determined by which direction results in lowest BMD</td>
<td>BMDS uses linear trend test to auto-detect adverse direction. Care is warranted if data are non-linear or non-monotonic.</td>
</tr>
</tbody>
</table>
# NTP Proposed Approach to Genomic Dose-Response Modeling

## Step 2a: Evaluating Model Fit

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude feature if no model has a complete set of convergent BMD, BMDU and BMDL values</td>
<td>Not a likely scenario for the limited number of datasets analyzed in an IRIS assessment</td>
</tr>
<tr>
<td>Use nested likelihood ratio test to choose between adequate fitting (BMDS p &gt; 0.05) linear and polynomial models</td>
<td>EPA uses AIC to choose between models and has not proposed a nested selection approach such as this for continuous polynomial models</td>
</tr>
<tr>
<td>• Choose more complex model if fit is significantly improved (p &lt; 0.05)</td>
<td></td>
</tr>
<tr>
<td>• Otherwise, choose simpler model</td>
<td></td>
</tr>
<tr>
<td>Compare AIC for chosen Polynomial model with AIC for adequate fitting power, Hill, and exponential models</td>
<td>Consistent with EPA BMD “best model” approach minus “BMDL range” consideration</td>
</tr>
</tbody>
</table>
### Step 2b: If Hill model is best fitting model

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept if Hill model “k” parameter is &gt;1/3 the lowest positive dose</td>
<td>Accept if BMD/BMDL ratio is &lt;20; serious concern if higher</td>
</tr>
<tr>
<td>If Hill model “k” parameter is &lt;1/3 of the lowest positive dose select next best model with fit p-value &gt;0.05</td>
<td>Use caution if BMD/BMDL ratio is &gt;5, “serious concern” (probably reject) if above 20</td>
</tr>
<tr>
<td>If no model has a p &gt;0.05, assign BMD equal to lowest BMD from the probe set with a acceptable Hill model (“k” parameter&gt;1/3 lowest positive dose)</td>
<td>If a model gives a BMD/BMDL ratio of &gt; 20 and no other model has a p &gt;0.1, EPA might relax fit criteria (e.g., to p &gt;0.05) or use another dataset.</td>
</tr>
</tbody>
</table>
### NTP Proposed Approach to Genomic Dose-Response Modeling

**Step 3: Filtering of features before gene set analysis**

<table>
<thead>
<tr>
<th>NTP Proposal</th>
<th>Analogous EPA Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove features with fit p-value &lt; 0.0001</td>
<td>Remove features with fit p-value &lt; 0.05 (pre-selected model) or &lt; 0.1 (no preselected model)</td>
</tr>
<tr>
<td>Remove features with BMDU/BMDL &gt; 40</td>
<td>Remove features with BMD/BMDL &gt; 20</td>
</tr>
</tbody>
</table>
Conclusions

- BMD modeling for traditional continuous endpoints using “best method” approach used for 20+ years; methods are well-defined

- Alternative methods are being researched to address model uncertainty (e.g., model averaging) and provide more accurate modeling results

- BMDExpress leverages BMDS model executables to extend methods to alternative endpoints (i.e., gene expression)

- BMDExpress modeling and model selection criteria are generally consistent with EPA methods in areas of overlapping purpose.

- BMDExpress with well positioned to adapt to updates to BMD modeling approaches (i.e., adoption of model averaging)