Other Approaches for BMD Analysis
The WHO Approach

Matthew W. Wheeler Ph.D.
Angelika Tritscher Ph.D.
Introduction

- The purpose of this talk is to give an overview of the WHO process of Benchmark Dose (BMD) estimation.

- Cover Following:
  - Short History of past guidance/assessments.
    - What has been done.
  - Current Practice
    - What was done at the 83rd JECFA Meeting
  - Issues Raised
    - Things that prevent a harmonization of BMD methodology
  - Future directions.
    - What the most recent committee would like to see in the future
History

- Prior to 2006 health based guidance values were primarily derived from the NOAEL, which defined the point of departure (POD).

- In 2006, at the 64th meeting of JECFA (FAO & WHO, 2006) the BMD was used to define a point a POD and calculate a margin of exposure (MOE).
  - This was done for several genotoxic carcinogens.
In 2009, EHC 240 (FAO & WHO) was published, which gave guidance on dose-response modeling and the development of health based guidance.

- Health-based guidance values may be derived from either NOAELs or BMDs (BMDLs).
- For BMD estimation it was emphasized that:
  - The data should have a sufficient number of doses.
  - The dose response-model should fit the data adequately by some predefined criterion (e.g. P-value > 0.1).
  - POD should be estimated within the range of the observed data (e.g. 10%).
  - BMD Sensitivity to model choice should be investigated.
History

At the 72\textsuperscript{nd} Meeting of JECFA the committee reaffirmed a commitment to the BMD methodology making the following recommendations:

- In accordance with the 64\textsuperscript{th} meeting, the lowest BMDL is to be chosen.
- For 2-year animal toxicology studies, a BMD is to be estimated using a benchmark response (BMR) of 10\%, though other levels may be more appropriate for human data.
- Recommended the use of appropriate software for BMD/BMDL computation. In practice, this implies use of US EPA BMDS software package or RIVM’s PROAST package.
- Recommended the establishment of a working group to address several issues to harmonize BMD modeling (discussed below).
At its 83rd meeting, the committee reaffirmed the need to establish a working group adding the group should:

- Investigate Model Averaging and other model weighting strategies.
- Develop recommendations on which models to use and when to use constraints.
- Recommend an approach to combine statistical and biological knowledge.
- Review use of goodness-of-fit criteria.
- Review new developments in the USEPA BMDS and PROAST software.

What follows is a brief version of the rationale.
3-MCPD Esters

Consider the data from Cho et al (2008), which was the basis for the 83rd committee's decision.

Renal Tubule hyperplasia

<table>
<thead>
<tr>
<th>Dose (ppm)</th>
<th>N</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>1.97</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td>8.27</td>
<td>50</td>
<td>21</td>
</tr>
<tr>
<td>29.5</td>
<td>50</td>
<td>36</td>
</tr>
</tbody>
</table>
Per WHO guidance all US EPA BMDA models were fit to the data. Additionally, unconstrained models (more on that later) and model averaging was investigated.

<table>
<thead>
<tr>
<th>Model</th>
<th>Restricted Model</th>
<th>BMD</th>
<th>BMDL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>Yes</td>
<td>5.62</td>
<td>4.60</td>
<td>0.00</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>Yes</td>
<td>1.21</td>
<td>0.87</td>
<td>0.61</td>
</tr>
<tr>
<td>Probit</td>
<td>Yes</td>
<td>5.36</td>
<td>4.47</td>
<td>0.00</td>
</tr>
<tr>
<td>Log-Probit</td>
<td>Yes</td>
<td>3.98</td>
<td>2.96</td>
<td>0.01</td>
</tr>
<tr>
<td>Weibull</td>
<td>Yes</td>
<td>2.14</td>
<td>1.66</td>
<td>0.07</td>
</tr>
<tr>
<td>Multistage 2°</td>
<td>Yes</td>
<td>2.14</td>
<td>1.66</td>
<td>0.07</td>
</tr>
<tr>
<td>Quantal-Linear</td>
<td>Yes</td>
<td>2.14</td>
<td>1.66</td>
<td>0.07</td>
</tr>
<tr>
<td>Gamma</td>
<td>Yes</td>
<td>2.14</td>
<td>1.66</td>
<td>0.07</td>
</tr>
<tr>
<td>Gamma</td>
<td>No</td>
<td>0.526</td>
<td>0.07</td>
<td>0.92</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>No</td>
<td>0.831</td>
<td>0.22</td>
<td>0.57</td>
</tr>
<tr>
<td>Log-Probit</td>
<td>No</td>
<td>0.917</td>
<td>0.27</td>
<td>0.54</td>
</tr>
<tr>
<td>Weibull</td>
<td>No</td>
<td>0.63</td>
<td>0.13</td>
<td>0.81</td>
</tr>
</tbody>
</table>
3-MCPD Esters

The question for this analysis: Should you use the unrestricted models, or should you use the restricted models:

- Case for unrestricted model: From a statistical point of view, confidence intervals are only correct if the estimate is not on the bound of the parameter space (i.e. restricted models are problematic from a statistical point of view).
  - Result: Confidence intervals not correct!

- Case for restricted model: Toxicologically an unrestricted model makes no sense. For example, if you assume a 1-hit cancer model, then an unrestricted model violates this assumption.
  - Result: Estimates are unrealistic!
Based upon these considerations we argued that the unrestricted models were far too conservative because:

- Other data (Sunahara, 1993) for males and female rats point to a similar BMD/BMDL as the log-logistic.
- The quantal linear model is typically thought of as the “most conservative” model, and the constrained log-logistic model is more conservative than the values it provides.
- Model averaging (Wheeler and Bailer 2007) gives a nearly identical result to the quantal linear model.
Issues raised

As the POD chosen by JECFA is 0.87, which is an order of magnitude greater than the value chosen by EFSA using the EXACT same data and the EXACT SAME model suite, this analysis raises several important issues:

– Which approach is correct?
– Can model averaging harmonize these approaches?
  • If so which approach do we use?
– What does one do when a new method suggests a different POD to the exact same dataset, which was already used in a previous decision?
Issues Raised

Merging statistical and toxicological knowledge

For the discussion assume the Weibull model for dichotomous responses:

\[ \Pr(x|d) = \gamma + (1 - \gamma)[1 - \exp(-\beta d^\alpha)], \]

with \( BMD = \left[ \frac{-\log(1 - BMR)}{\beta} \right]^{1/\alpha} \)

The pesky parameter is \( \alpha \). If it is greater 1, then we are O.K. from a statistical and toxicological standpoint.

If the parameter is constrained to be 1 and is estimated to be 1, we are ok from a toxicological standpoint, but not so from a statistical standpoint. Here computing the BMDL is questionable.
Issues Raised

Crudely, you can think of the statistical argument as follows:

- Statisticians assume normality. Confidence intervals are based upon normality assumptions.
- If you place a bound on this value, then normality does not apply.
- This includes all functions of the derived parameters like the BMD.
- Using our traditional assumptions, without normality, we can’t produce reliable confidence intervals.
Issues Raised
Issues Raised

- As a result, our BMDL, which is used in risk assessment for statistical uncertainty, does not really quantify the statistical uncertainty of the experiment!

- You might say it is a 95% lower bound, but in reality it is nowhere even close.
  - I have seen cases using the default EPA methodology where it is a 16% lower bound !?
Issues Raised

- A toxicologist would say that the models considered below the restriction are nonsense, and thus should not be used based upon other knowledge.

- This argument has credibility.

- Let’s revisit the 3-MCPD Esters analysis using the Weibull model. Here I stress the BMDL is computed by finding “new values” for the Weibull distribution that modify the maximum likelihood by a predefined value.

- That is: we refit the model until it is different enough from the original by some statistical criterion and use that model to calculate the BMDL.
Weibull Lower Bound Model

This model says that the majority of the increase in the response probability happens at ultra low doses! Alternatively 50% of the RTH increase seen at the dose 1.97 would be seen at the dose 0.1?!? The implication is even more ridiculous using the Gamma model, which is the lowest BMDL.
Issues Raised

- Can model averaging (e.g., Wheeler and Bailer 2007) save one from this situation?
  - Unfortunately, though it provides a reasonable answer in the case of 3-MCPD Esters, as long as unrestricted models are used, you can’t get rid of this behavior using that approach.

- There are better solutions. The method the EPA is exploring is fully Bayesian, which would allow crucial toxicologist input into the model’s parameters while keeping the models unrestricted.
  - This method is very promising.
Issues Raised

The next issue is more practical:

- What do you do when a new statistical method suggests a different POD?
- Is there a range where we are OK and do nothing?
- Does this make the science less transparent?
- One may argue current practice is to be health protective, which is not an empirical principle. If better science argues for a less conservative estimate, is that OK?
Conclusion

- There is much to discuss regarding this approach.

- Harmonizing the BMD approach between all stakeholders would increase the reproducibility of risk assessment science.

- Thank you!