Summary of Dose-Response Modeling for Developmental Toxicity Studies

Daniel L. Hunt
Shesh N. Rai
Chin-Shang Li
Developmental toxicity studies (DTS) are an important area in the field of toxicology. In a DTS, fetal litters are indirectly exposed to various levels of non-carcinogenic toxic substances through direct exposure to the host animals. Endpoints that are recorded in these studies include fetal weight and length, as well as indicators of abnormality and death. Endpoints are then measured to determine litter responses, which include average weight, malformation and death rate. The dose-response pattern in these studies typically appears to exhibit at least the existence of a threshold effect. The threshold dose-response (DR) model is the default model for non-carcinogenic risk assessment, according to the USEPA, and is encouraged by the agency for the use in the risk assessment process. Several statistical models are proposed to estimate the threshold dose and to account for other important aspects of the developmental toxicity study. Use of these models to different applications will be summarized. The advantages and disadvantages of these models, and the comparison to other alternative models are discussed. We, also, summarize potentials for future research in this field.

**Keywords:** Developmental toxicity, Dose-group variability, Estimation, Splines, Threshold
Developmental Toxicity Study (DTS)

- Randomly assign pregnant animals to exposure levels of toxin
- Outcomes measured on litters (fetuses)
- Fetal endpoints of interest:
  - Live endpoints
    - body weight and length
    - structural malformations
  - Death endpoints
    - Resorption
    - post-natal death
Outcomes in a DTS

- Implantation
  - Resorption
  - Post-Natal Death

Viable Fetus

- Malformation Status
- Weight
Modeling of Data

- Equate implantation site number to litter size
- Categorize malformations, resorptions, and post-natal deaths together as adverse events
- Measure outcome per litter as proportion of adversely affected fetuses
- Predict dose-response relationship by modeling $P(d)$
  - Dose: level of toxic substance ($d$)
  - Response rate: probability toxic response ($P$)
Threshold

Definition: *The largest (non-zero) dose level which yields a toxic response that is equivalent to response at the control level.*

1991 USEPA Guidelines for Developmental Toxicity Risk Assessment:

- “…In general, a threshold is assumed for the dose-response curve for agents that produce developmental toxicity…”

USEPA models of non-carcinogen studies

- NOAEL (no-observed-adverse-effects-level): *highest experimental dose at which response not (statistically) different from control*
- Benchmark dose: *lower statistical confidence limit for dose corresponding to specified increased level of adverse effect over background level (excess risk)*
Threshold Dose-Response Model
Functional Form of Threshold DR Model

\[ P(d) = F\left[ \theta_0 + \theta_1(d - \tau) \times I(d > \tau) \right] \]

OR

\[ P(d) = \begin{cases} 
F(\theta_0), & d \leq \tau \\
F\left[ \theta_0 + \theta_1(d - \tau) \right], & d > \tau 
\end{cases} \]
Model Properties

- Continuous, piecewise function

- Figure slide
  - F would simply be an identity function, appropriate for continuous endpoint, such as weight; probability P could be replaced with W, for weight

- Discrete, binary endpoints
  - F could be logistic or probit function
  - P(d) is the probability of response at dose level d
Linear Spline Model

- Although above-threshold model may not be, below-threshold model is linear.
- However, it is limited to only one pattern.
- To make more robust, one could fit a linear spline model in lieu of the threshold model.
- Spline model gives more flexibility in being able to accommodate below-threshold patterns.
Polynomial Regression B-spline model

- Theory: DeBoor (2001)
- A regression spline of order m (degree m-1) and k interior knots yields a function of the form

\[ s(d; \theta, \varepsilon) = \sum_{i=1}^{m+k} \theta_i B_i(d; \varepsilon) \]

\( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)' \) is the set of interior knots

\( \{B_i(d; \varepsilon) : i=1, \ldots, m+k\} \) is the set of (order m) B-splines

\( \theta = (\theta_1, \ldots, \theta_{m+k})' \) are the regression coefficients of the B-splines
Spline model for developmental study

- Linear \((m=2)\) B-spline with \(k=1\) interior knot \(\varepsilon\)

\[
P_1(d) = G \left[ \sum_{i=1}^{3} \theta_i B_i(d; \varepsilon) \right]
\]

- \(G\) is appropriate function to accommodate the type of data (identity, logistic, probit, etc.)
Example DTS: DEHP administered to pregnant rats (Tyl et al., 1983)
Response Variation

- As noted in the prior slide, there is noticeable degree of variability in the litter responses
- This variability should be accounted for in the modeling process
- Simple modification to the fixed effects models $P(d)$ and $P_1(d)$:
  - Add random effect to model this response variability
Models modified

\[ P(d) = F \left[ \theta_0 + \theta_1 (d - \tau) \times I(d > \tau) + \sigma Z \right] \]

\[ P_1(d) = G \left[ \sum_{i=1}^{3} \theta_i B_i(d; \varepsilon) + \sigma Z \right] \]

The random component \( \sigma Z \) is such that \( Z \sim N(0,1) \) and \( \text{Var}(Z) = \sigma^2 \)
### DEHP data (Tyl et al., 1983)

<table>
<thead>
<tr>
<th>Dose of DEHP</th>
<th>Number Litters</th>
<th>Number Fetuses</th>
<th>Average Litter Size</th>
<th>Number Affected Fetuses</th>
<th>Proportion Affected Fetuses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>30</td>
<td>396</td>
<td>13.2</td>
<td>75</td>
<td>.189</td>
</tr>
<tr>
<td>0.025</td>
<td>26</td>
<td>320</td>
<td>12.3</td>
<td>37</td>
<td>.116</td>
</tr>
<tr>
<td>0.05</td>
<td>26</td>
<td>319</td>
<td>12.3</td>
<td>80</td>
<td>.251</td>
</tr>
<tr>
<td>0.10</td>
<td>24</td>
<td>276</td>
<td>11.5</td>
<td>192</td>
<td>.696</td>
</tr>
<tr>
<td>0.15</td>
<td>25</td>
<td>308</td>
<td>12.3</td>
<td>302</td>
<td>.981</td>
</tr>
</tbody>
</table>
Threshold and Spline Models fit to DEHP data
Comparing Models

- From deBoor (2001), alternate representation of the spline model
  \[ s(d,\theta,\varepsilon) = \theta_1 + \theta_2 d + \theta_3 (d-\varepsilon)_+ \]

- If \( \theta_2 = 0 \), the model becomes
  \[ s_1(d,\theta,\varepsilon) = \theta_1 + \theta_3 (d-\varepsilon)_+ \]

- which is equivalent to the threshold model
Likelihood Ratio (LR) testing

- Dealing with parametric, nested models, hence LR test for significance

- Result
  - p-value = 0.42
  - Indicative of non-significance (of the spline model)

- Note the limited # dose groups below threshold
Factors to Improve Study Power

- To better estimate threshold & other effects:
- More dose groups (below threshold)
- Larger sample size (at dose groups below threshold)
- Adequate dose spacing

Authors
- Sielken and Stevenson, 1998
- Teeguarden et al., 2000
<table>
<thead>
<tr>
<th>Dose of DYME (mg/kg/day)</th>
<th>Number of Litters</th>
<th>Number of Fetuses</th>
<th>Average Litter Size</th>
<th>Number of Affected Fetuses</th>
<th>Proportion of Affected Fetuses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
<td>297</td>
<td>14.1</td>
<td>17</td>
<td>.057</td>
</tr>
<tr>
<td>62.5</td>
<td>20</td>
<td>242</td>
<td>12.1</td>
<td>20</td>
<td>.083</td>
</tr>
<tr>
<td>125</td>
<td>24</td>
<td>312</td>
<td>13.0</td>
<td>35</td>
<td>.112</td>
</tr>
<tr>
<td>250</td>
<td>23</td>
<td>299</td>
<td>13.0</td>
<td>102</td>
<td>.341</td>
</tr>
<tr>
<td>500</td>
<td>23</td>
<td>285</td>
<td>12.4</td>
<td>277</td>
<td>.972</td>
</tr>
</tbody>
</table>
Threshold and Spline Models fit to DYME data
LR testing: DYME data

- Significance of spline over threshold model
- p-value=0.091
- Much lower than for DEHP data
- Reasons? Now, 3 dose groups below threshold instead of 2
- Conclusion? More dose groups needed to adequately assess significance
Considerations when using Spline Approach

- In DTS, there are few dose groups
- The number of noticeable changepoints in the pattern of the DR data will be few (if not non-existent)
- Spline order and #knots should be based on these factors
- Current design of DTS seems to accommodate this
Advantages of these models

- Direct estimation of threshold or changepoint

- Inclusion of random effects into DR function to counter models such as the BB; facilitates estimation by allowing common methods to find parameter estimates

- Direct modeling of DR pattern
Disadvantages

- Threshold can be difficult to estimate

- Current design of DTS accommodates use of threshold and low-order spline model; however, modified design DTS may require more complex models; can lead to estimation difficulties
Future Work

- Modify models to include multiple $\sigma$ parms.
- Easier method for estimating SEs, e.g., bootstrapping
- Higher order spline (perhaps quadratic) model
- More general spline model (higher order, more knots)
References

References, cont.

References, cont.

- National Technical Information Service, Springfield, VA
References, cont.

- Williams DA. 1975. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31: 949-952
Biomedical References for Applications of the Spline Model


QUESTIONS???