Semi-parametric and non-parametric approaches to concentration-response modelling

Christian Ritz

University of Copenhagen, Denmark

Leuven, September 25 2008
Concentration-response setup

Parameter of interest: effect concentration (such as EC50)

Concentration-response setting:

- biological response $y_i$ to stimulus $x_i$
  (stimulus applied for a range of concentrations)

Response types:

- continuous (length, weight)
- counts (number of fronds, juveniles, offspring, roots)
- quantal (number of organisms responding out of a total)
  (active/inactive, dead/alive, immobile/mobile)
Concentration-response setup

Parameter of interest: effect concentration (such as EC50)

Concentration-response setting:

- biological response $y_i$ to stimulus $x_i$
  (stimulus applied for a range of concentrations)

Response types:

- continuous (length, weight)
- counts (number of fronds, juveniles, offspring, roots)
- quantal (number of organisms responding out of a total)
  (active/inactive, dead/alive, immobile/mobile)
Concentration-response setup

Parameter of interest: effect concentration (such as EC50)

Concentration-response setting:
- biological response $y_i$ to stimulus $x_i$
  (stimulus applied for a range of concentrations)

Response types:
- continuous (length, weight)
- counts (number of fronds, juveniles, offspring, roots)
- quantal (number of organisms responding out of a total)
  (active/inactive, dead/alive, immobile/mobile)
Parametric models

General conditional mean structure:

\[ E(y_i|x_i) = f^P(x_i, \beta) \]

Details:
- \( f^P \) nonlinear mean function in \( \beta \)
  - monotonous: log-logistic, Weibull, . . .
  - non-monotonous: polynomials, biphasic models
- \( \beta \) unknown parameter to be estimated

Methods of estimation:
- least squares
- maximum likelihood
- quasi-likelihood
Limitations

Rough figures obtained from ECVAM:
- 50% fitted nicely by common parametric models
- 20% borderline fits
- 30% no acceptable fit achievable

Problem: *Empirically based* models

Consequences:
- Inadequate summary of the data structure
- Risk of bias in estimates of EC values and other parameters of interest
Limitations

Rough figures obtained from ECVAM:
- 50% fitted nicely by common parametric models
- 20% borderline fits
- **30% no acceptable fit achievable**

**Problem:** *Empirically based models*

**Consequences:**
- Inadequate summary of the data structure
- Risk of bias in estimates of EC values and other parameters of interest
Limitations

Rough figures obtained from ECVAM:

- 50% fitted nicely by common parametric models
- 20% borderline fits
- 30% no acceptable fit achievable

**Problem:** *Empirically based* models

**Consequences:**

- Inadequate summary of the data structure
- Risk of bias in estimates of EC values and other parameters of interest
Non-parametric models

Complete unspecified conditional mean:

\[ E(y_i|x_i) = f^{NP}(x_i) \]

Estimation by local linear regression:

1. choose a bandwidth \( h(x) \)
2. calculate weights \( w_{ij}(x) = W \left( \frac{x_j - x}{h(x)} \right) \)
   (only using \( x_i \)'s in the interval \( [x - h(x), x + h(x)] \))
3. fit weighted linear regression of \( y_{ij} \) versus \( x_{ij} \) with weights \( w_{ij}(x) \)
4. define \( \hat{f}^{NP}(x) \) to be the estimated intercept
More on local linear regression

- How to balance bias-variance trade-off?
- How to choose the bandwidth? Variable bandwidth?
- In practice used for both continuous and quantal data!
- Local likelihood approaches exist (Loader, 1999)

Implementations in R:

- `loess()` in `stats` (standard installation)
- `locfit()` in the `locfit` package
Semi-parametric models

Maybe there exists a compromise:

- imposing some basic concentration-response structure
- leaving enough flexibility for capturing non-standard patterns in the data

Model-robust approach (Nottingham & Birch, 2000):

\[ f^{MR}(x) = \lambda f^{NP}(x) + (1 - \lambda) f^{P}(x, \beta) \]

\( \lambda \in [0, 1] \) controls the mixing of components

Separate estimation of parametric and non-parametric components
Semi-parametric models

Maybe there exists a compromise:
- imposing some basic concentration-response structure
- leaving enough flexibility for capturing non-standard patterns in the data

Model-robust approach (Nottingham & Birch, 2000):

\[ f^{MR}(x) = \lambda f^{NP}(x) + (1 - \lambda)f^P(x, \beta) \]

\[ \lambda \in [0, 1] \] controls the mixing of components

Separate estimation of parametric and non-parametric components
Combining model fits

Optimal mixing parameter $\lambda$ determined from:

$$\text{PRESS}^* = \sum_{i=1}^{n} g_i(\hat{f}_{-i}^{MR}(x_i), \lambda)$$

using leave-one-out predictions: $\hat{f}_{-i}^{MR}(x_i)$

Least squares criterion (common choice):

$$g_i(z, \lambda) = w_i(y_i - z)^2 / g_0(\lambda)$$

($g_0$ some weight function)
Implementation

- **R package:** mrdrC
- also available as a GUI:
  - http://130.75.68.4:8080/deploy/doseresponse/
Quantal data ($\hat{\lambda} = 0.65$)
Continuous data ($\hat{\lambda} = 1$)
## Simulation: continuous data - null

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Replicates</th>
<th>EC</th>
<th>True</th>
<th>Mean</th>
<th>Width</th>
<th>Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-logistic model</td>
<td>Parametric</td>
<td>1</td>
<td>1.46</td>
<td>1.53</td>
<td>2.73</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.97</td>
<td>2.51</td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.10</td>
<td>2.23</td>
<td>92.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.46</td>
<td>1.49</td>
<td>1.17</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.95</td>
<td>1.11</td>
<td>95.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.09</td>
<td>1.06</td>
<td>94.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1.46</td>
<td>1.48</td>
<td>0.88</td>
<td>97.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.94</td>
<td>0.84</td>
<td>97.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.07</td>
<td>0.82</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td>Semi-parametric (0.23)</td>
<td></td>
<td>1</td>
<td>1.46</td>
<td>1.36</td>
<td>1.66</td>
<td>85.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.91</td>
<td>1.18</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.25</td>
<td>1.32</td>
<td>78.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.46</td>
<td>1.39</td>
<td>0.93</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.88</td>
<td>0.67</td>
<td>76.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.08</td>
<td>0.72</td>
<td>83.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1.46</td>
<td>1.40</td>
<td>0.68</td>
<td>77.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>1.92</td>
<td>1.89</td>
<td>0.57</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>3.06</td>
<td>3.07</td>
<td>0.60</td>
<td>85.5</td>
<td></td>
</tr>
</tbody>
</table>
### Simulation: continuous data - alternative

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Replicates</th>
<th>EC</th>
<th>True</th>
<th>Mean</th>
<th>Width</th>
<th>Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormesis model</td>
<td>Parametric</td>
<td>1</td>
<td>4.46</td>
<td>2.26</td>
<td>13.32</td>
<td>80.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>5.68</td>
<td>25.26</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>30.58</td>
<td>98.15</td>
<td>93.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>4.46</td>
<td>2.08</td>
<td>6.64</td>
<td>62.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>5.40</td>
<td>12.90</td>
<td>91.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>29.05</td>
<td>47.46</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>4.46</td>
<td>2.00</td>
<td>4.97</td>
<td>45.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>5.28</td>
<td>9.79</td>
<td>90.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>28.84</td>
<td>35.97</td>
<td>83.9</td>
<td></td>
</tr>
<tr>
<td>Semi-parametric</td>
<td></td>
<td>1</td>
<td>4.46</td>
<td>3.12</td>
<td>13.62</td>
<td>89.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>7.39</td>
<td>22.74</td>
<td>92.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>32.99</td>
<td>61.04</td>
<td>86.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>4.46</td>
<td>2.96</td>
<td>6.18</td>
<td>65.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>6.62</td>
<td>8.97</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>30.61</td>
<td>52.39</td>
<td>86.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>4.46</td>
<td>2.85</td>
<td>4.50</td>
<td>57.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>6.86</td>
<td>6.24</td>
<td>6.16</td>
<td>75.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>35.05</td>
<td>29.00</td>
<td>42.88</td>
<td>81.3</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Key points:

- semi-parametric approach potentially useful
- more concentrations and less replicates desirable
- for common designs inferior to parametric approach
- model selection criteria useful for choosing between parametric and semi-parametric models
Key points:

- semi-parametric approach potentially useful
- more concentrations and less replicates desirable
- for common designs inferior to parametric approach
- model selection criteria useful for choosing between parametric and semi-parametric models
Conclusion

Key points:

- semi-parametric approach potentially useful
- more concentrations and less replicates desirable
- for common designs inferior to parametric approach
- model selection criteria useful for choosing between parametric and semi-parametric models
Conclusion

Key points:

- semi-parametric approach potentially useful
- more concentrations and less replicates desirable
- for common designs inferior to parametric approach
- model selection criteria useful for choosing between parametric and semi-parametric models


Grant: 2006/S 237-252824 Lot 3
European Centre for the Validation of Alternative Methods (ECVAM)
Institute for Health and Consumer Protection
EU Joint Research Centre
Ispra, Italy