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 β
  - monotonous: log-logistic, Weibull, ...
  - non-monotonous: polynomials, biphasic models
- β 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 $w_{i'}(x) = W\left(\frac{x_{i'}-x}{h(x)}\right)$
   (only using x_is in the interval $]x - h(x), x + h(x)[$)
3. fit weighted linear regression of $y_{i'}$ versus $x_{i'}$ with weights $w_{i'}(x)$
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 λ determined from:

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

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

Least squares criterion (common choice):

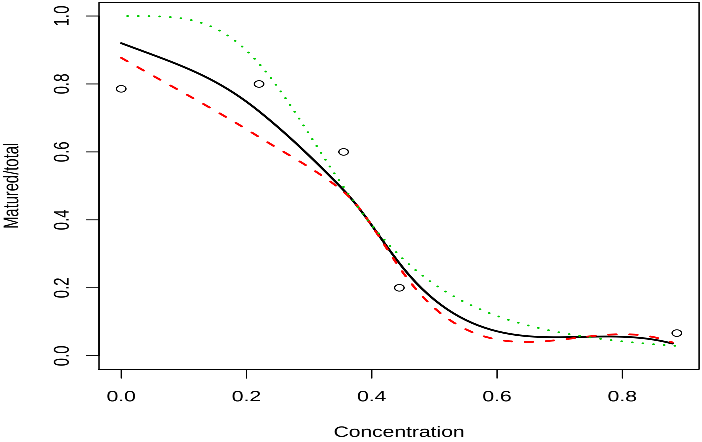
$$g_i(z, \lambda) = w_i(y_i - z)^2 / g_0(\lambda)$$

(g_0 some weight function)

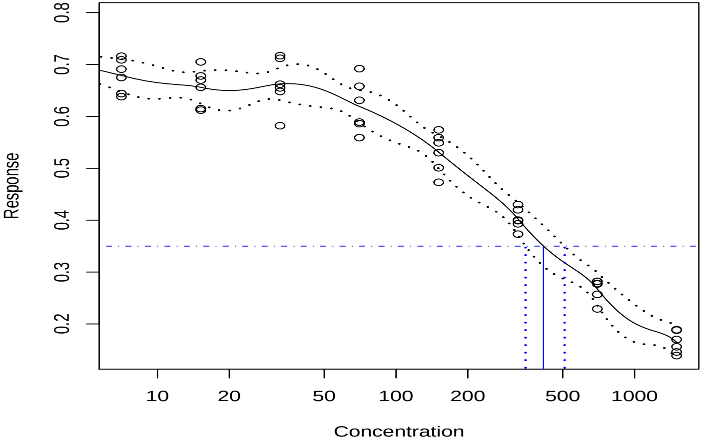
Implementation

- **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

Model	Method	Replicates	EC	True	Mean	Width	Coverage (%)
Log-logistic	Parametric	1	10	1.46	1.53	2.73	95.3
model			20	1.92	1.97	2.51	94.5
7 concs			50	3.06	3.10	2.23	92.8
		2	10	1.46	1.49	1.17	95.3
			20	1.92	1.95	1.11	95.2
			50	3.06	3.09	1.06	94.1
		3	10	1.46	1.48	0.88	97.3
			20	1.92	1.94	0.84	97.1
			50	3.06	3.07	0.82	94.4
	Semi	1	10	1.46	1.36	1.66	85.1
	-parametric		20	1.92	1.91	1.18	84.2
	(0.23)		50	3.06	3.25	1.32	78.6
		2	10	1.46	1.39	0.93	76.2
			20	1.92	1.88	0.67	76.5
	(0.14)		50	3.06	3.08	0.72	83.5
		3	10	1.46	1.40	0.68	77.6
			20	1.92	1.89	0.57	79.0
	(0.11)		50	3.06	3.07	0.60	85.5

Simulation: continuous data - alternative

Model	Method	Replicates	EC	True	Mean	Width	Coverage (%)
Hormesis model	Parametric	1	10	4.46	2.26	13.32	80.9
7 concs			20	6.86	5.68	25.26	95.3
			50	35.05	30.58	98.15	93.1
		2	10	4.46	2.08	6.64	62.4
			20	6.86	5.40	12.90	91.8
			50	35.05	29.05	47.46	85.5
		3	10	4.46	2.00	4.97	45.3
			20	6.86	5.28	9.79	90.1
			50	35.05	28.84	35.97	83.9
	Semi	1	10	4.46	3.12	13.62	89.1
	-parametric		20	6.86	7.39	22.74	92.2
	(0.29)		50	35.05	32.99	61.04	86.8
		2	10	4.46	2.96	6.18	65.9
			20	6.86	6.62	8.97	78.8
	(0.61)		50	35.05	30.61	52.39	86.2
		3	10	4.46	2.85	4.50	57.8
			20	6.86	6.24	6.16	75.9
	(0.68)		50	35.05	29.00	42.88	81.3

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

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

References and acknowledgment

Loader, C. (1999). Local Regression and Likelihood. Springer, New York

Nottingham, Q. J. & Birch, J. B. (2000). A semiparametric approach to analysing dose-response data, *Statist. Med.*, **19**, 389-404

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