# The algae growth inhibition test – robust initial values for parameter estimation

Anke Schulz

Bayer Schering Pharma AG
Global Drug Discovery Statistics

anke.schulz@bayerhealthcare.com

NCS2008 Leuven

Belgium

September 25th, 2008

# Overview

1 The algae growth inhibition test
2 Four-parameter logistic regression model
3 Initial values for iterative estimation procedure
4 Application

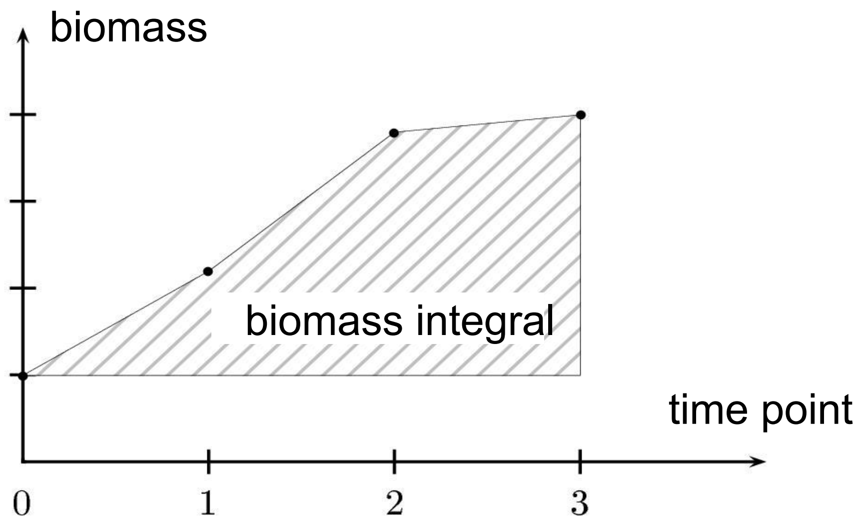
1 The algae growth inhibition test

# Design

- Biological test to investigate the growth inhibition of algae under addition of a substance.
- Test routinely used in toxicology.
- Design:
    - One control group (usually six replicates).
    - Several dose groups of the same substance (usually five dose groups in dilution series with three replicates).
    - Equidistant time points starting at 0h (usually 0, 24, 48 and 72).

# Biomass integral

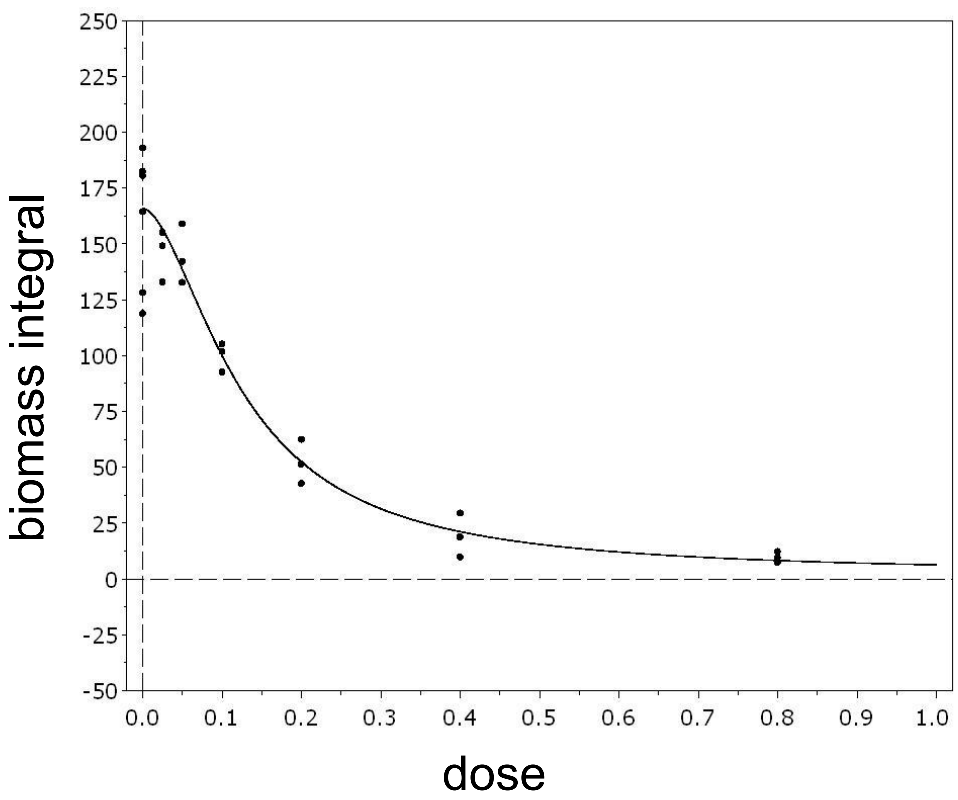
- Algae biomass is a function over time for each replicate.
- The growth is measured by the integral of the biomass.
  → This integral will be the variable of interest.



1 The algae growth inhibition test

# Data

- Integral of biomass decreases by dose (in dependence on toxicity).

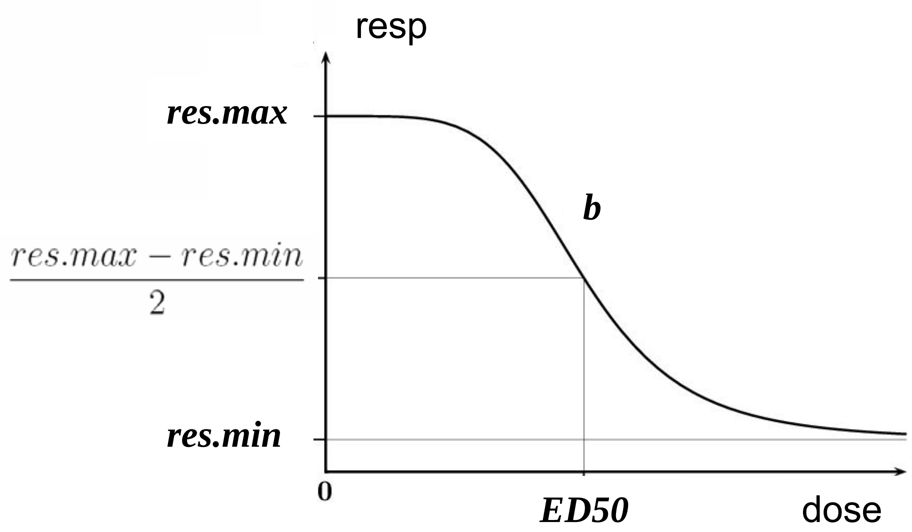


# Logistic regression model

- Sigmoid model:
- Four-parameter logistic regression model:

$$resp(dose) = res.min + \frac{res.max - res.min}{1 + \left(\frac{dose}{ED50}\right)^b} + \epsilon;$$

$$\begin{aligned} \epsilon &\sim N(0, \sigma^2) \\ dose &\in \mathbb{R}_0^+ \\ res.max,\ res.min,\ b &\in \mathbb{R} \\ ED50 &\in \mathbb{R}^+ \end{aligned}$$



2 Four-parameter logistic regression model

# Iterative parameter estimation

- The non-linearity of the model requires an iterative parameter estimation procedure and initial values.
- The convergence behavior depends on the choice of initial values.
- For **small variation**, procedures to obtain initial values developed by (Normolle, 1993) and (Ritz and Streibig, 2005) work very well.
- For **large variation** as found in the data of the algae test:
  - Robust M-estimators proposed by Huber (1964).
  - Our new robust procedure (2006) including an automated determination of initial values.

# Parameters *res.min* and *res.max*

- The initial values for the parameters *res.min* (minimum) and *res.max* (maximum) will be estimated as described by Ritz and Streibig (2005):

$$\begin{aligned} res.min_{ini} &= \text{minimum observation} \\ res.max_{ini} &= \text{maximum observation} \end{aligned}$$

- However, initial values for $res.min_{ini}$ and $res.max_{ini}$ have only a limited impact on the convergence.

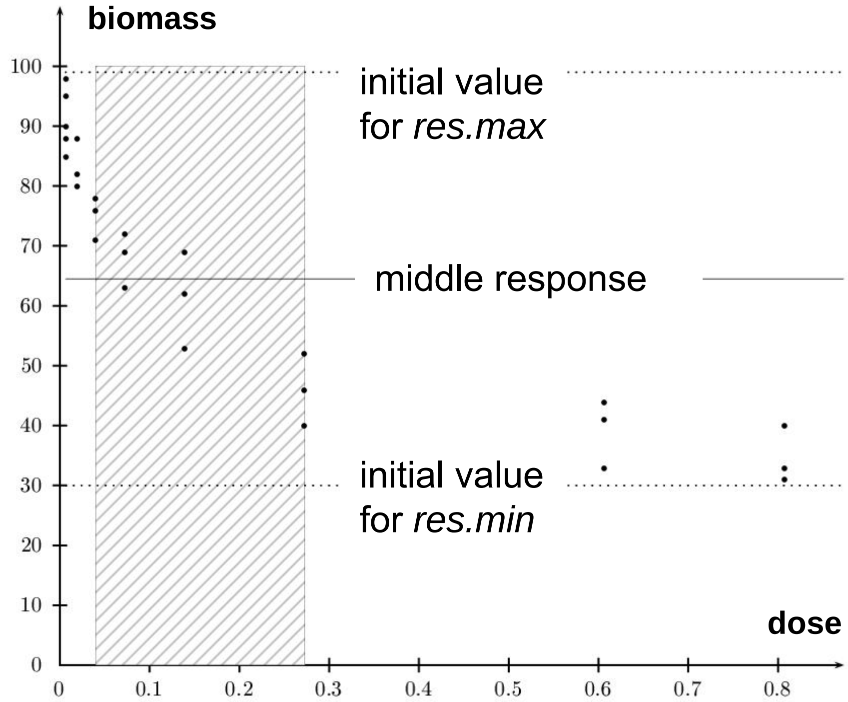
# Parameters *b* and *ED50*

- Estimators for parameters *b* (slope) and *ED50* do have a functional relationship.
- If initial value for *ED50* is inaccurate, i.e. | $ED50_{ini}$ - *ED50*|>>0, parameter estimates for *b* and *ED50* are unreliable.
  Then, the sign of the slope changes.
- Consequently, a reasonable dose interval for $ED50_{ini}$ has to be found.

3 Initial values for iterative estimation procedure

# Parameters $b$ and $ED50$

- Middle response= $\frac{res.max - res.min}{2}$ .
- Dose interval= [largest dose with all observations above the middle response ; smallest dose with all observations below the middle response].
- $ED50_{ini}$ : weighted mean of the dose of the observations of the interval.
  - Weight depends on the distance to the middle response.



3 Initial values for iterative estimation procedure

# Parameters $b$ and $ED50$

- Based on the initial values for $res.max$, $res.min$ and $ED50$, initial value for parameter $b$ is calculated for each observation (Normolle, 1993):

$$b_j = \frac{\log\left(\frac{res.max_{ini} - resp_j}{resp_j - res.min_{ini}}\right)}{\log\left(\frac{dose_j}{ED50_{ini}}\right)}; \qquad b_{ini} = \text{median}(b_1, \ \ldots, \ b_n)$$

- For a first quality check of initial values, the sign of $b_{ini}$ is compared to the sign of the slope parameter of a simple linear regression (negative sign of $b$ corresponds to a positive slope).

# Parameters $b$ and $ED50$

- If the sign of $b_{ini}$ does not match the direction of the slope:
  - The sign of $b_{ini}$ is changed.
  - The parameter $ED50_{ini}$ is calculated for each observation based on the initial values for $res.max$, $res.min$ and $b$.
  - Median of all $ED50_j$ is the new initial value for $ED50$.

$$
\begin{aligned}
ED50_j &= dose_j \left( \frac{res.max_{ini} - resp_j}{resp_j - res.min_{ini}} \right)^{-\frac{1}{b_{ini}}} \\
ED50_{ini} &= \text{median}(ED50_1, \ \ldots, \ ED50_n)
\end{aligned}
$$

  - $b_{ini}$ is calculated again.

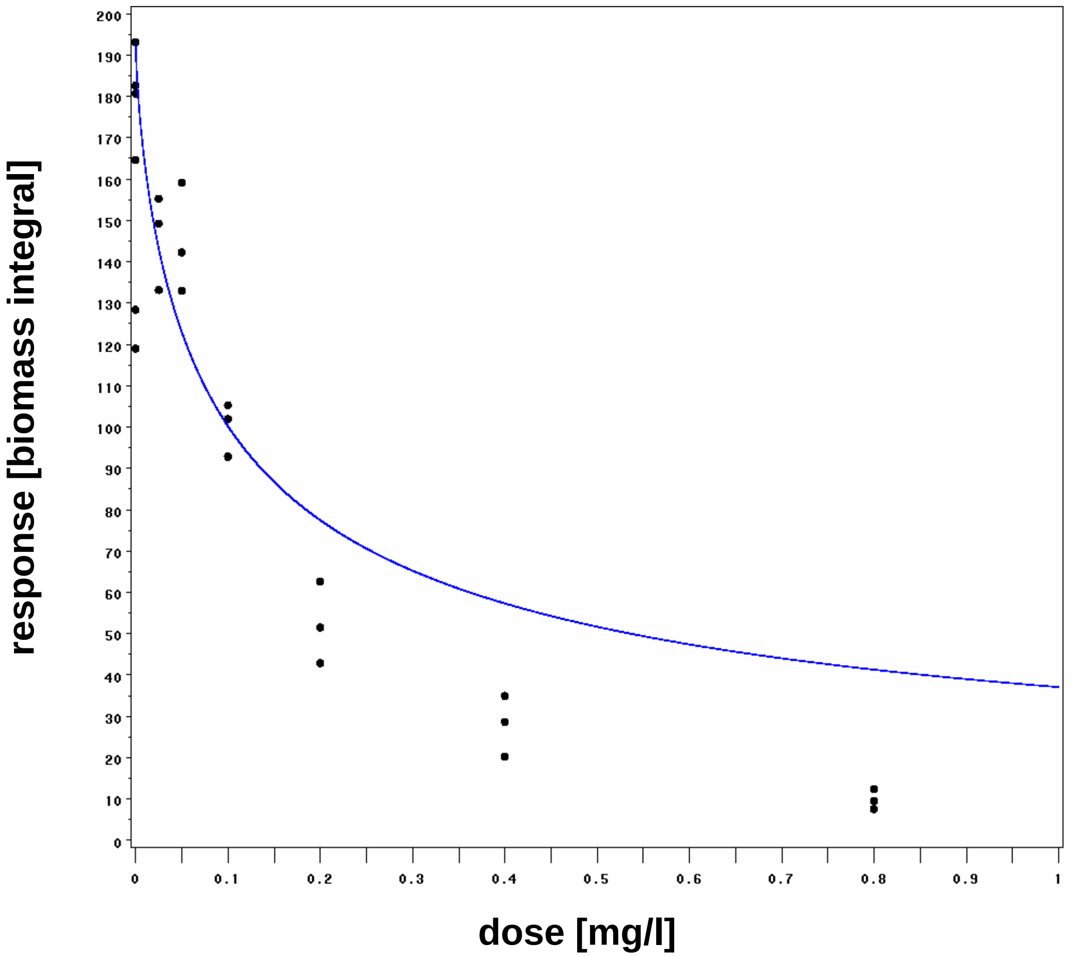
4 Application

# Practical experience

- Our proposed approach is tested on 42 real data sets.
- Common procedures lead to poor results, only 70% of all estimates are reasonable.
- Instead, our initial values yield sensible parameter estimates in all cases.

4 Application

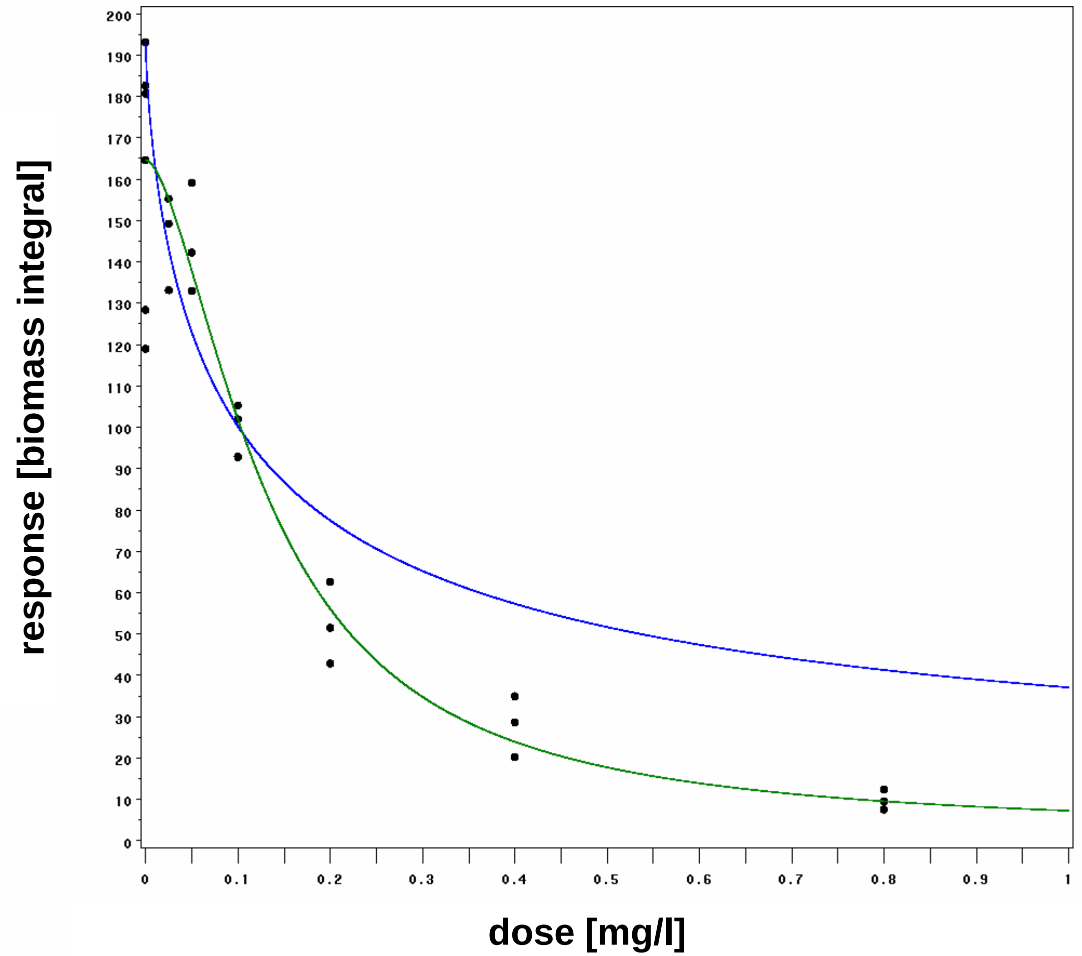
# Example




Blue: Model of initial values

4 Application

# Example

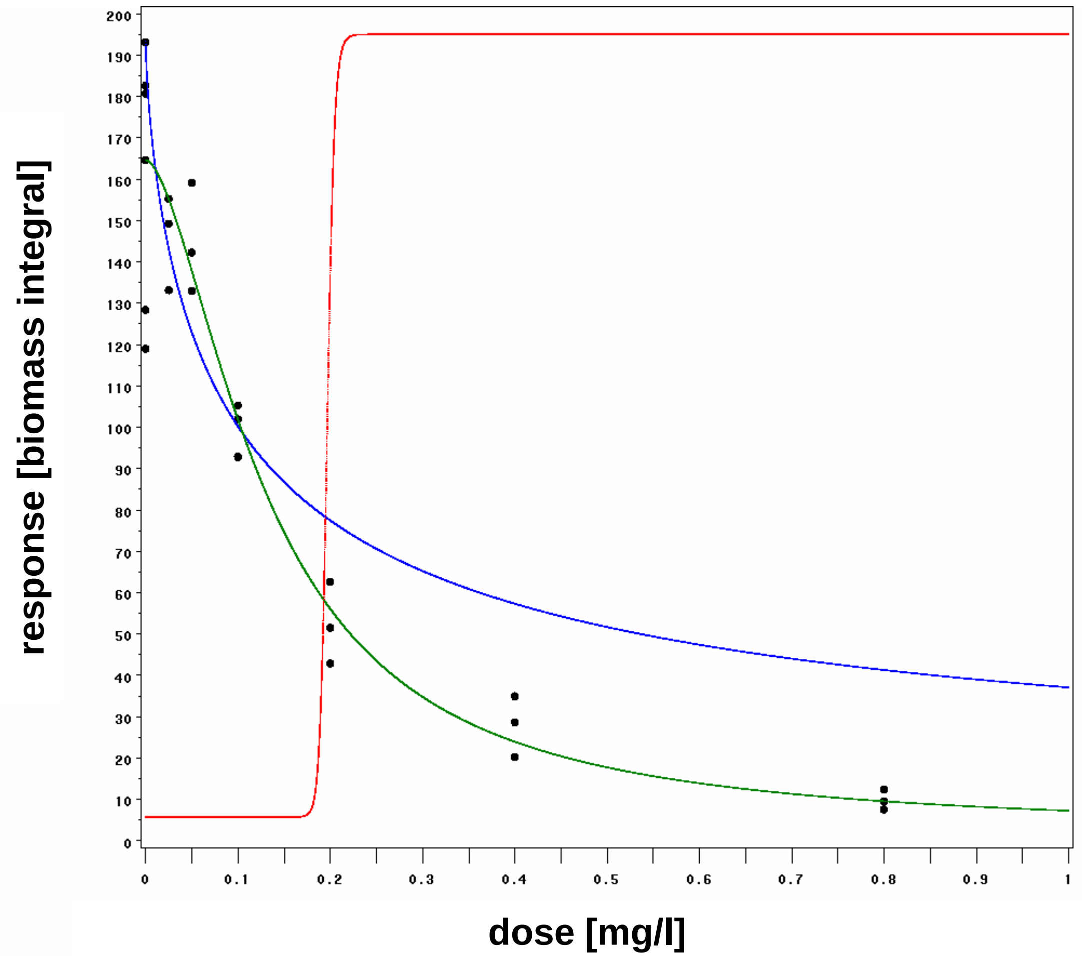



Blue: Model of initial values

Green: Estimated model

## 4 Application

# Example




Blue: Model of initial values

Green: Estimated model

Red: Model of initial values (Normolle, 1993)

# References


- Huber, P.J. (1964). ‘Robust estimation of local parameter’. *Annals of Mathematical Statistics* 35: 73-101 .
- Normolle, D.P. (1993). ‘An Algorithm for Robust non-linear Analysis of Radioimmunoassays and other Bioassays’. *Statistics in Medicine* 12: 2025-2042
- Ritz, C. & Streibig, J.C. (2005). ‚Bioassay Analysis using R’. *Journal of Statistical Software* 12, Issue 5.
- Schulz, A (2006). ‘Das vierparametrische logistische Modell und seine Anwendung bei einem ökotoxikologischen Problem’. Master Thesis. Humboldt-Universität zu Berlin, Germany.