28 Oct 2020

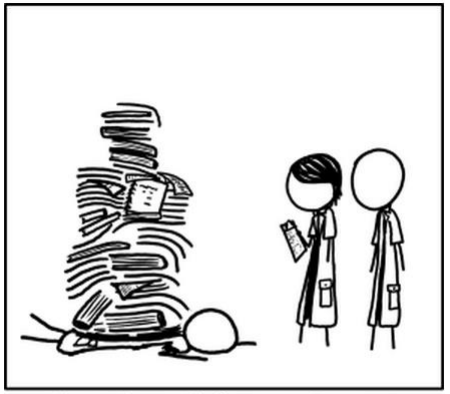
Advanced Topics in Biostatistics: Dose-response modeling

Prof. Dr. Annette Kopp-Schneider
Biostatistics – C060
kopp@dkfz.de

dkfz.
DEUTSCHES KREBSFORSCHUNGSZENTRUM
IN DER HELMHOLTZ-GEMEINSCHAFT

Outline

- Dose-response studies
- Common dose-response models
- Common measures of potency: ED50, ED10
- Parameter estimation
- Dose-response analysis with SigmaPlot
- Data transformations
- Fixing model parameters
- Assessment of model fit
- Experimental design issues
- Recommendations for practical use

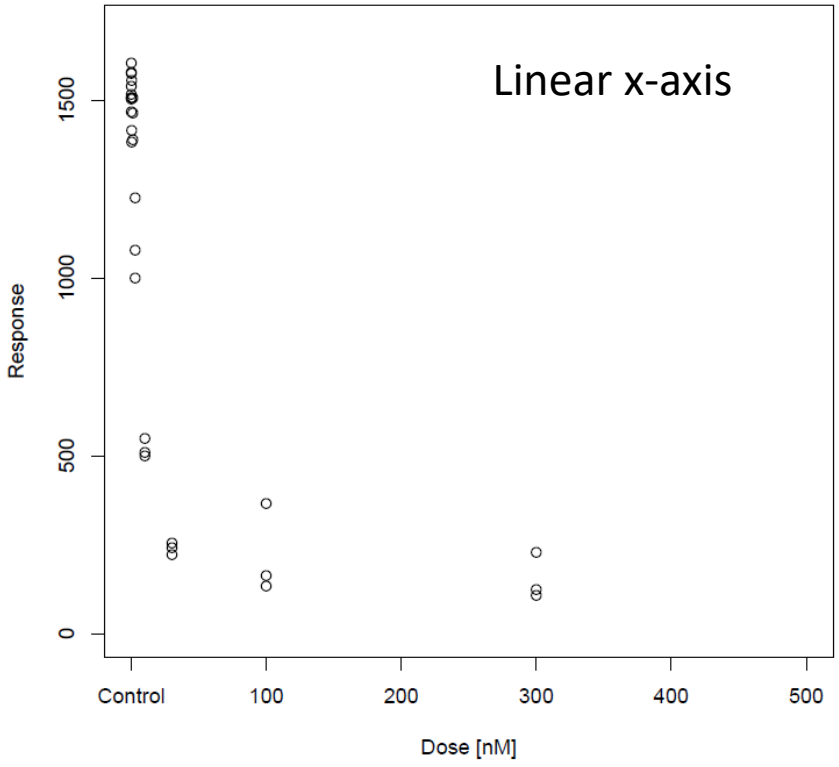


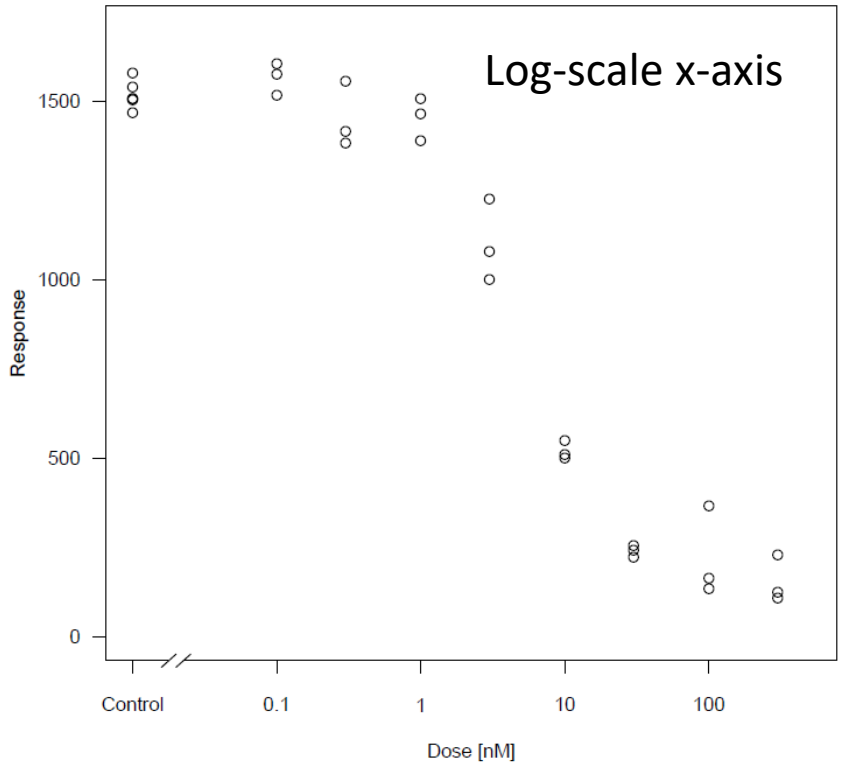
Dose-response studies

- Investigate relationship between different doses/concentrations of a test substance and their effects (responses) on a test system
- Study types, e.g.:
  - *in vitro* experiments
  - bioassays
  - early phase clinical experiments
- Aims:
  - characterize toxicity/effect of test substance
  - risk assessment: determine ‚safe' or ‚hazardous' (e.g. toxic) dose levels for drugs, potential environmental pollutants or other substances to which humans, animals or other organisms are exposed
  - clinical trials: determine ‚optimal' dose to be recommended for treatment of patients with given medical condition
- Dose-response relationships depend on exposure time and exposure route (inhalation, dietary intake,…)

Example

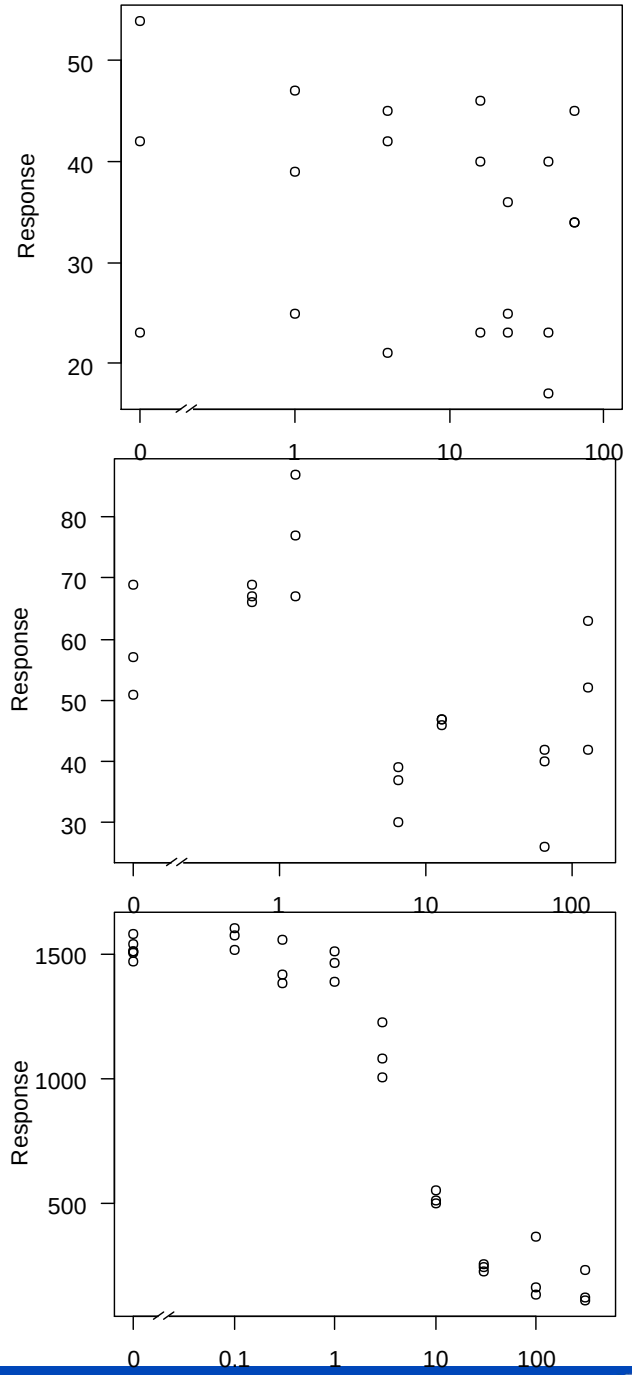
- *In vitro* experiment: recombinant androgen receptor binding assay (inhibition of receptor activity)
- Response: dpm (disintegrations radiolabeled ligand bound per minute)
- Control: 6 replicates, 8 dose levels: 3 replicates each, (+ 6 non-specific binding)





Typical questions

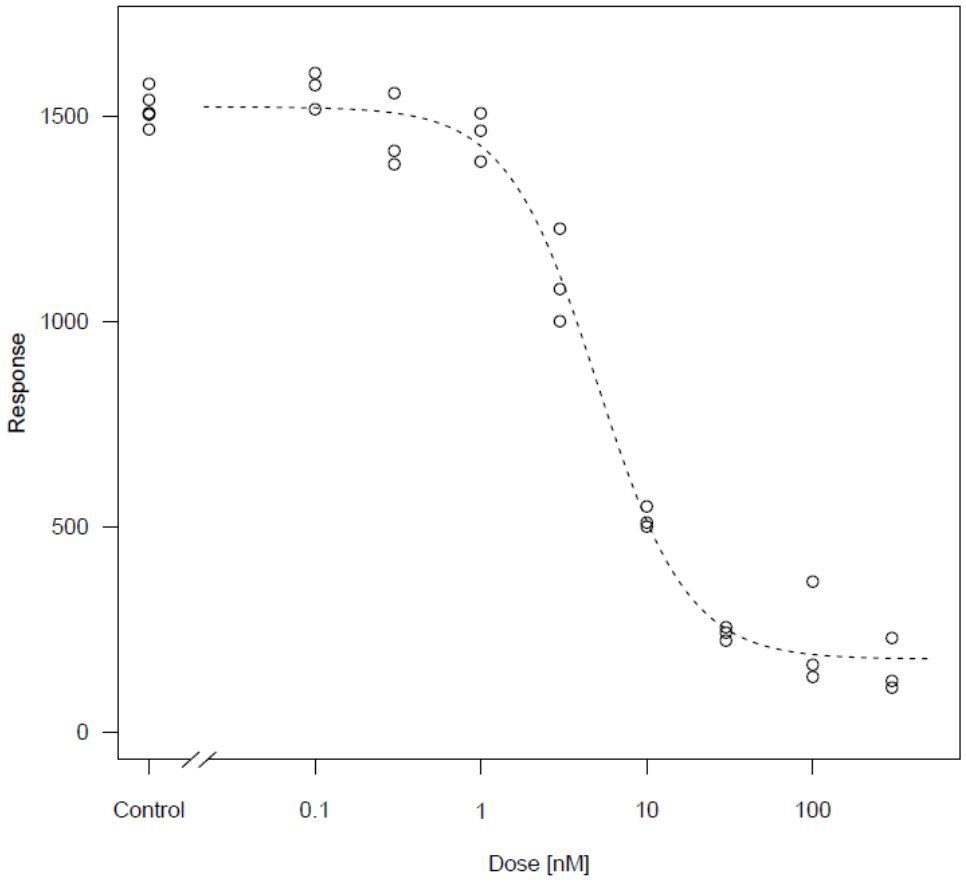
- Is there an effect of dose on response?
- Is a clear dose-response relationship observed?
- If there is a clear dose-response relationship, what is its functional form?



Dose-response models

Response = f(dose) + error

- f typically non-linear
- f often sigmoidal



Assumptions about error:

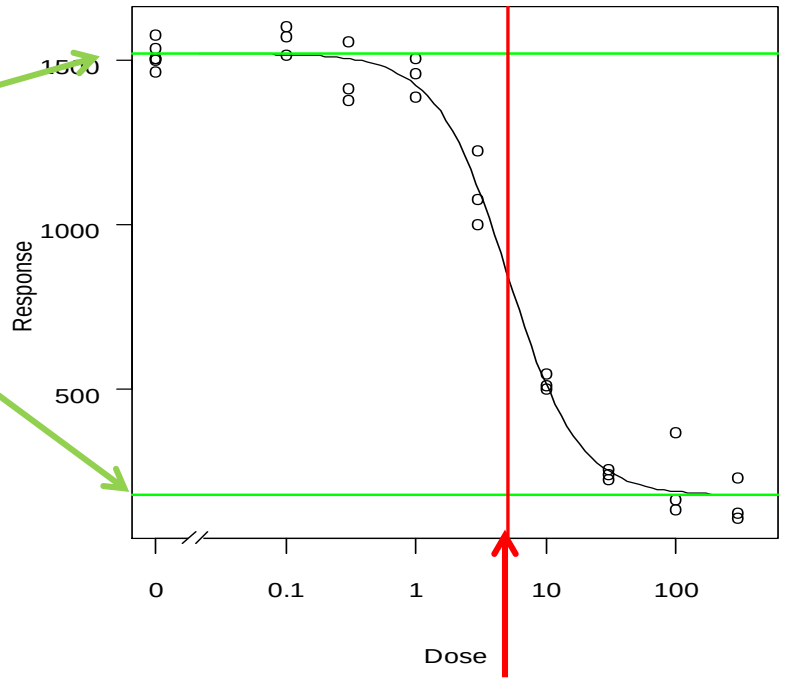
- expected error: 0
- uncorrelated
- normally distributed

Common dose-response model: 4-parameter log-logistic model

$$f(\text{dose}) = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + \left(\frac{\text{dose}}{ED50}\right)^{\text{HillSlope}}}$$

$$= \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + e^{(log(\text{dose}) - log(ED50))\cdot \text{HillSlope}}}$$

$$= \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(log_{10}(\text{dose}) - log_{10}(ED50))\cdot \text{HillSlope}}}$$



ED50

Other dose-response models exist:

- Weibull model
- Gompertz model
- ....

Properties of log-logistic model

- Two parameterizations: either log(ED50) or ED50 is estimated.
- Dose-response curve rescaled to log dose scale to obtain clear visualization.
- log-logistic model function plotted on log scale is symmetric around ED50.
- Different model parametrizations exist (e.g. natural log replaced by log10).
- Hill Slope > 0 usually indicates decreasing dose-response relationship.
- Hill Slope < 0 usually indicates increasing dose-response relationship.

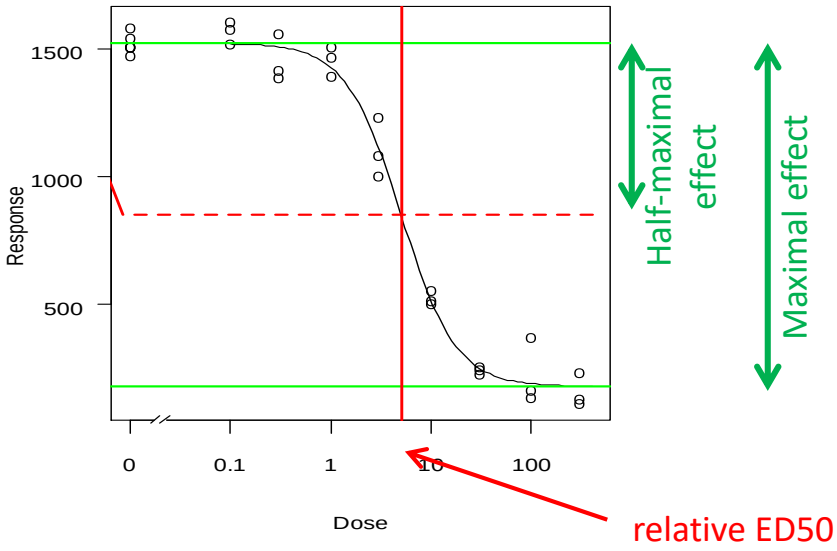
Relative vs. absolute ED50

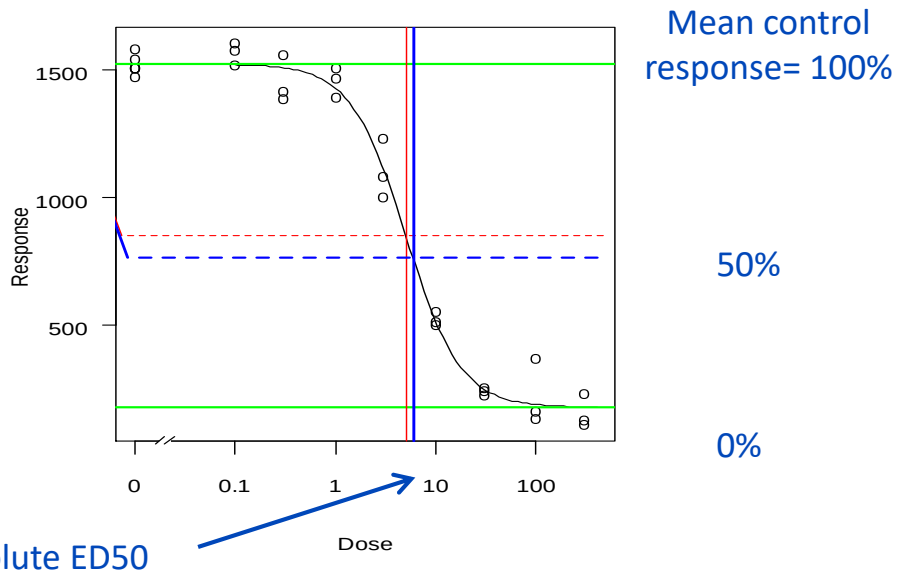
Relative ED50:

Dose producing half-maximal effect, i.e. dose corresponding to response midway between estimates of lower and upper plateau.

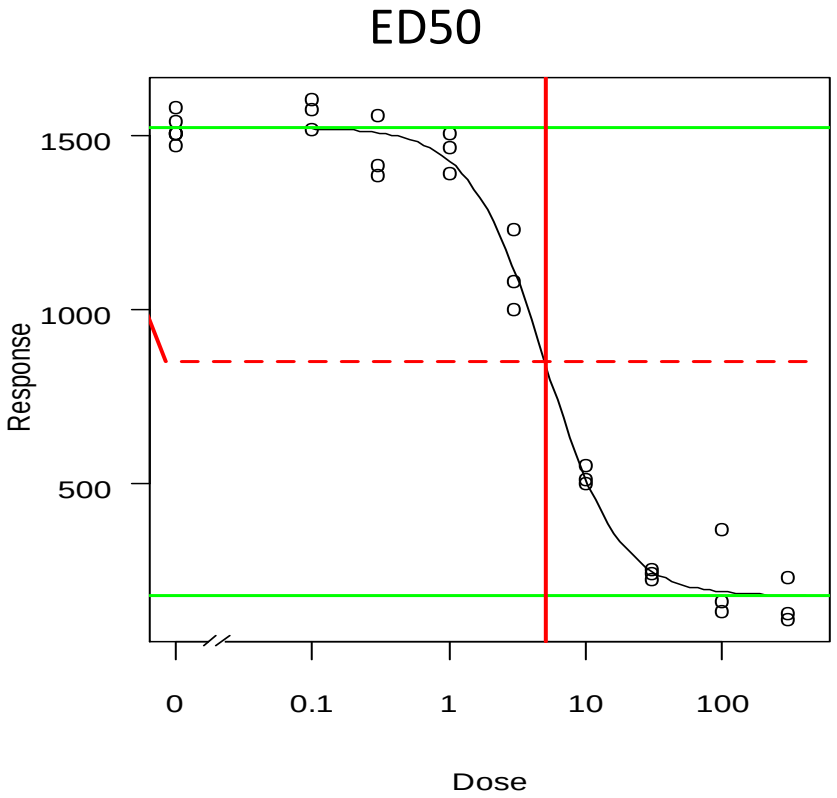
Absolute ED50 (only for decreasing curves):

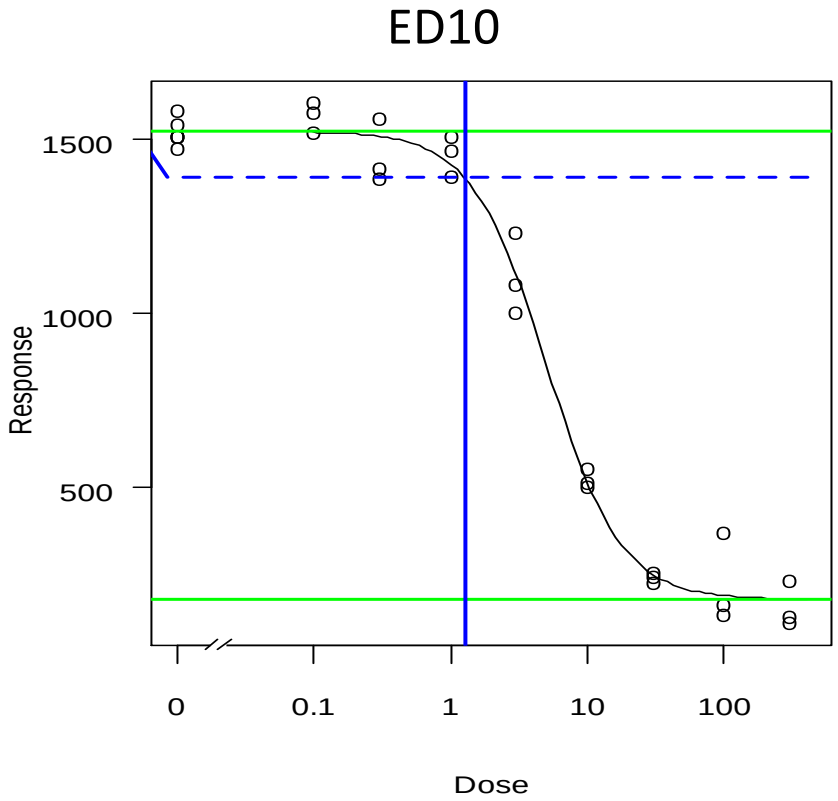
Dose corresponding to 50% of mean control response.





Common measures of potency/toxicity: Effective doses





Naming conventions

Effective doses EDp : ED10, ED50,…

Effective Concentrations ECp : EC10, EC50,…

Inhibitory Concentration ICp : IC10, IC50,…

Estimation of model parameters

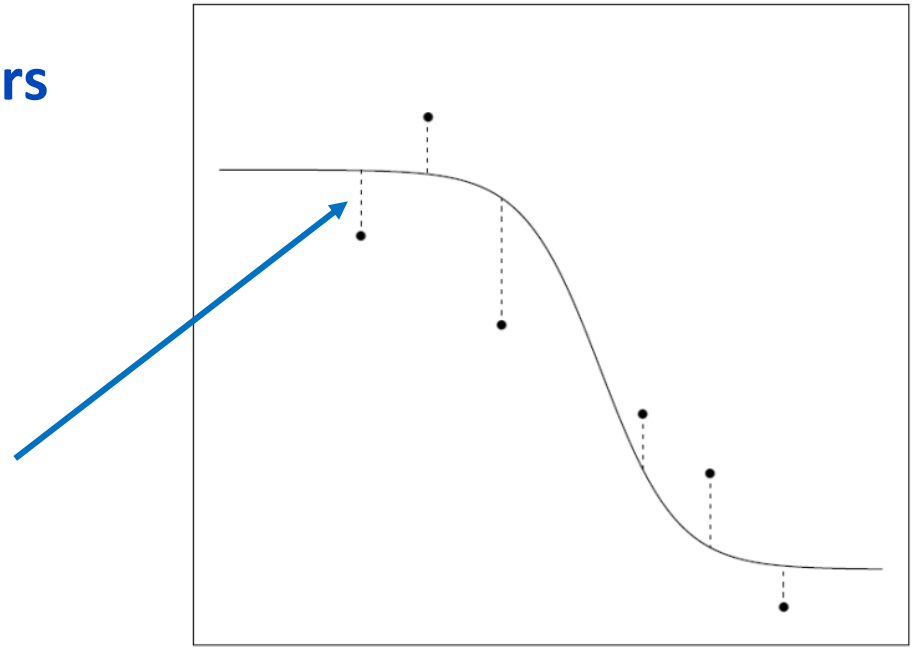
Least squares method:

Minimize sum of squared residuals

Residual=

Observed response – predicted response

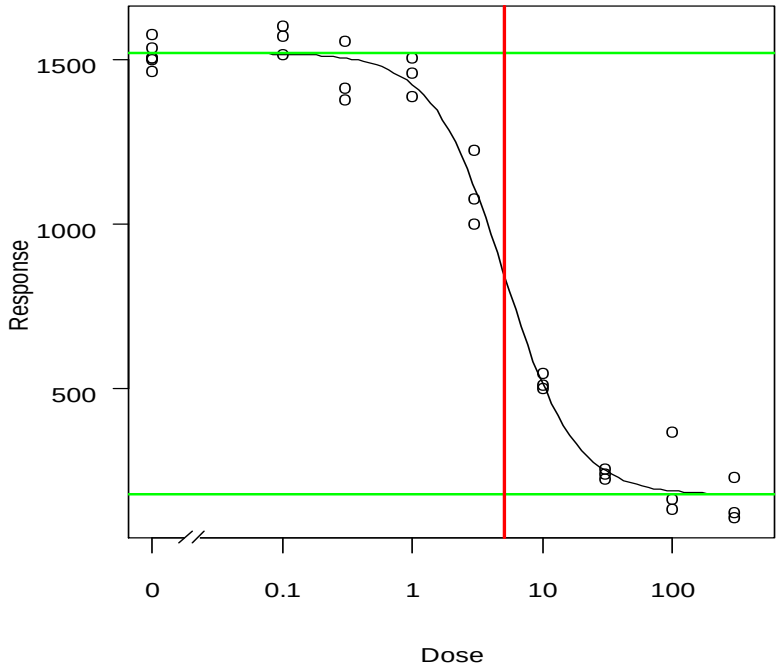
(cf. Estimation in linear regression)



- Non-linear optimization problem → iterative numerical optimization algorithms needed to find optimal parameter values
- At each iteration step, algorithm determines new parameter values based on data, model and current parameter values (until convergence is reached)
- Algorithm requires to pick/estimate initial values for each model parameter as starting point for iterative procedure
- If variance depends on response level → weighted least squares, weights: 1/response or $1/\text{response}^2$

Example: Parameter estimates

TOP = 1522



HillSlope=1.59

BOTTOM = 177

$ED50 = e^{\log(ED50)} = e^{1.615} = 5.03$ nM

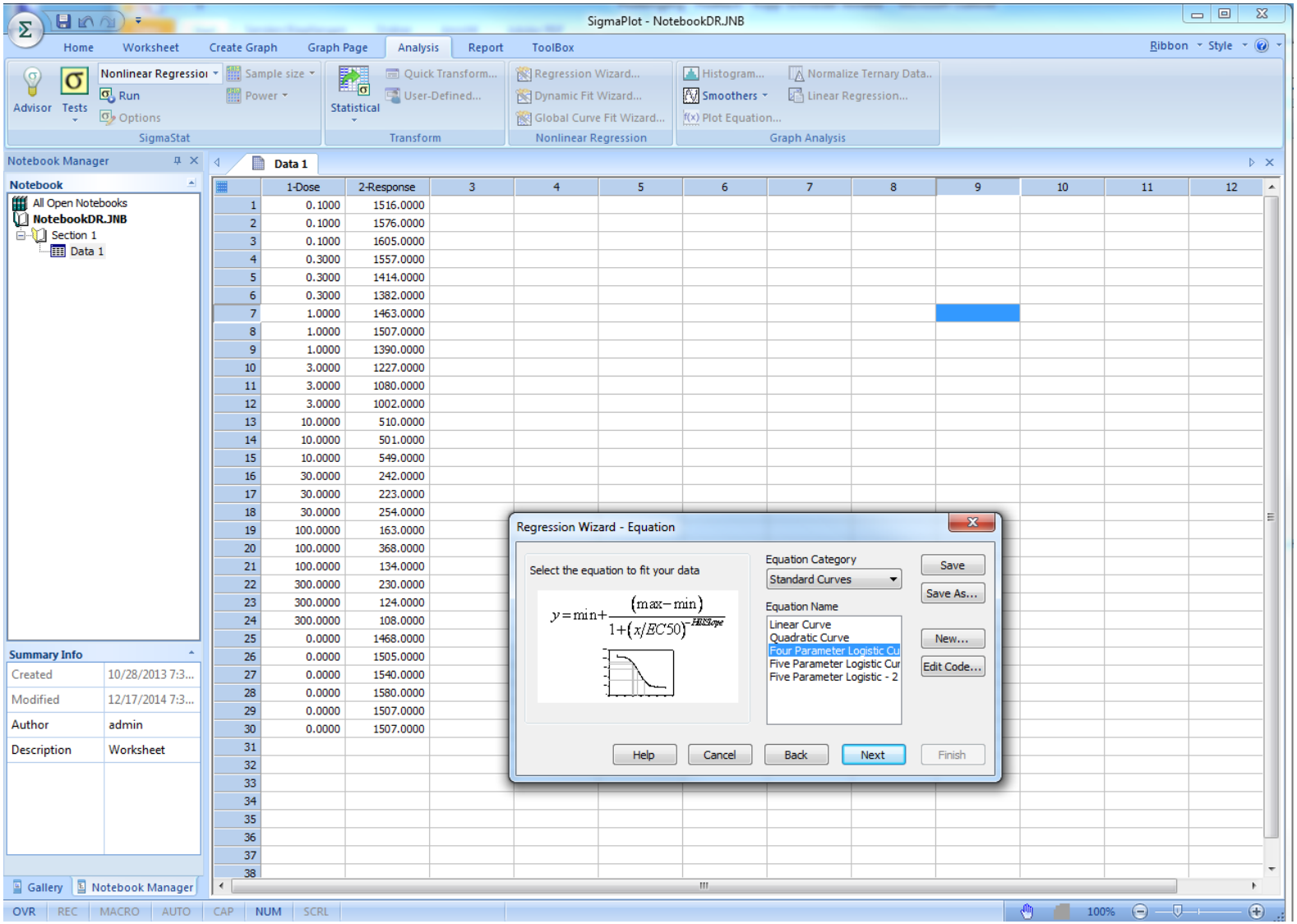
Interpretation: A dose of 5.03 nM produces 50% of the maximal effect

Precision of parameter estimates

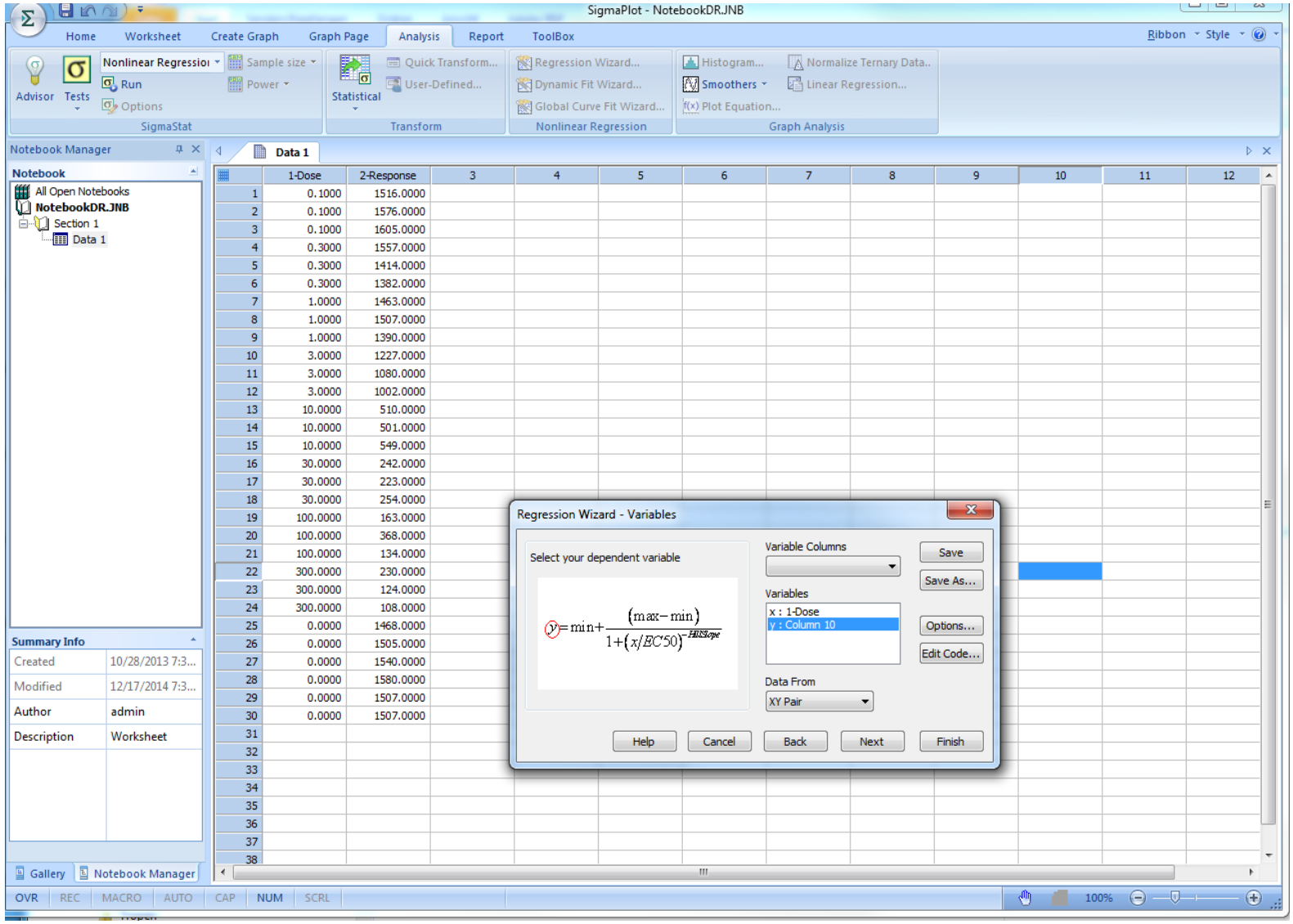
- Estimation of ED50 value from experimental data
- Experimental data vary
- Repetitions of the same experiment result in different ED50 estimates
- Precision of ED50 estimate can be assessed by 95% confidence interval (CI)

- Example: 95% CI for ED50 is [4.26 nM , 5.93 nM]
- Interpretation:
  the interval [4.26 nM , 5.93 nM] covers the true (but unknown) ED50 with 95% probability
  =
  if you would repeat the experiment 100 times and always calculate the 95% CI, the true (but unknown) ED50 is expected to be contained in 95 of the 100 CIs.

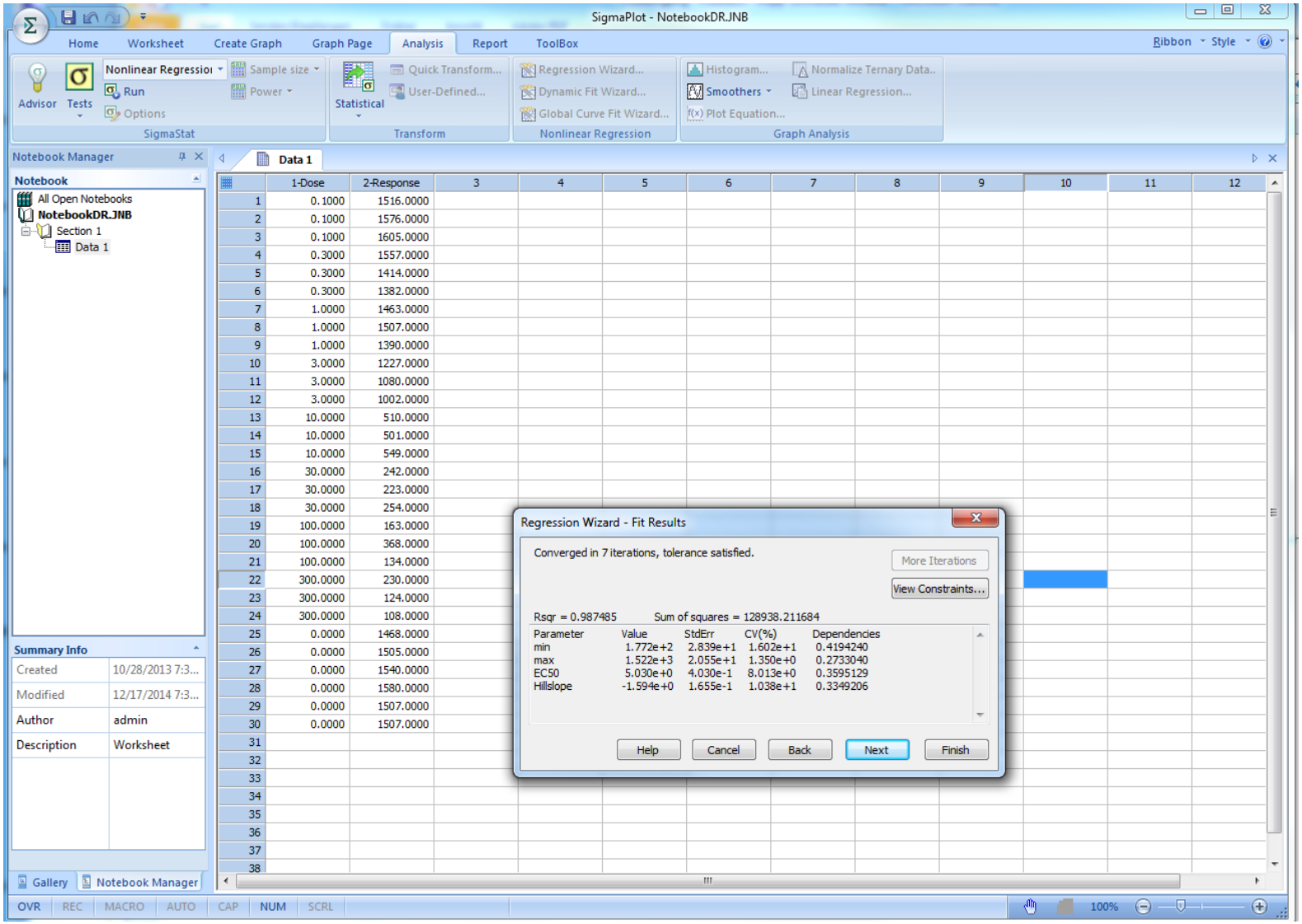
Dose-response analysis with SigmaPlot (1)



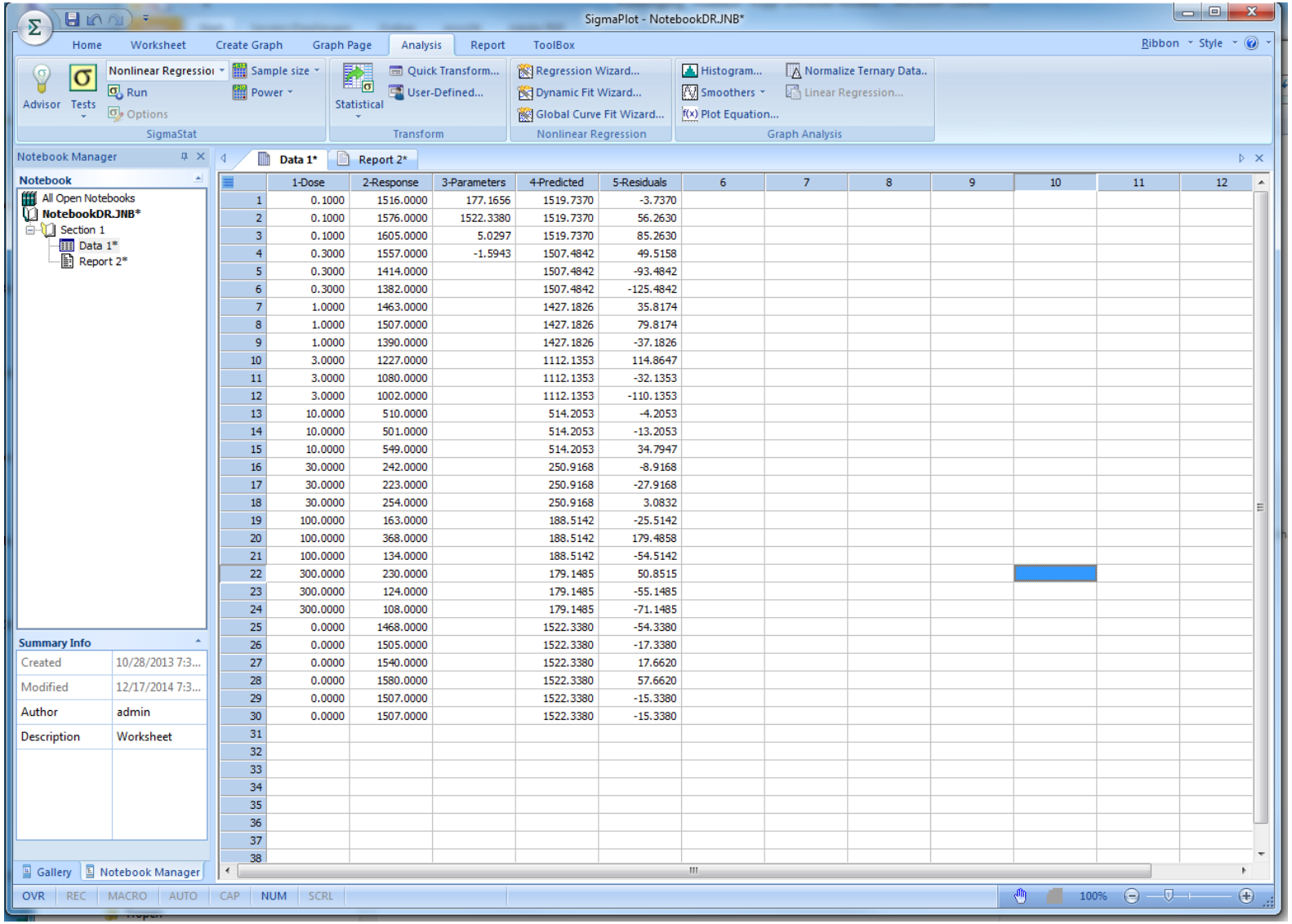
Dose-response analysis with SigmaPlot (2)



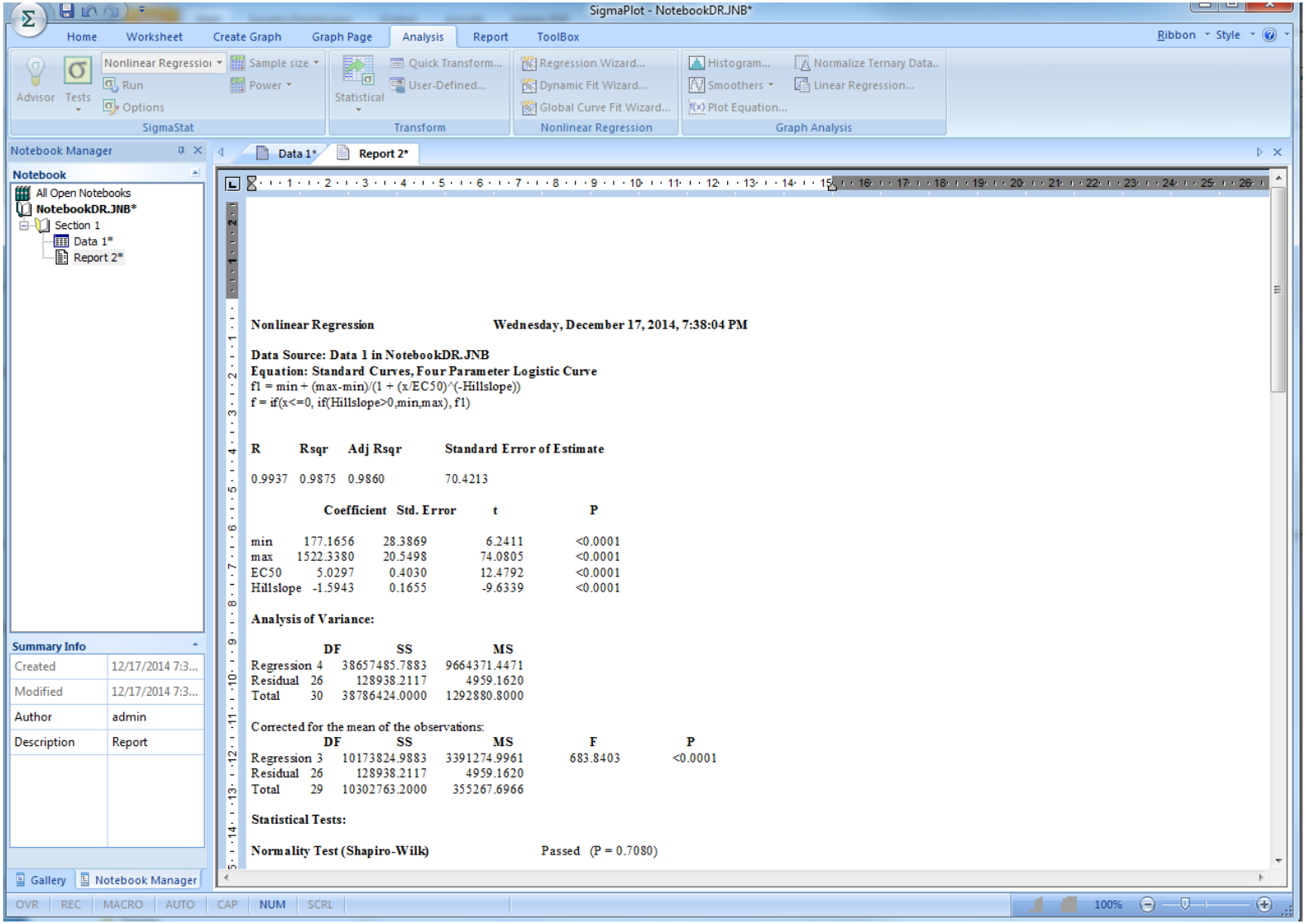
Dose-response analysis with SigmaPlot (3)



Dose-response analysis with SigmaPlot (4)



Dose-response analysis with SigmaPlot (5)



Dose-response analysis using WebApp (1)

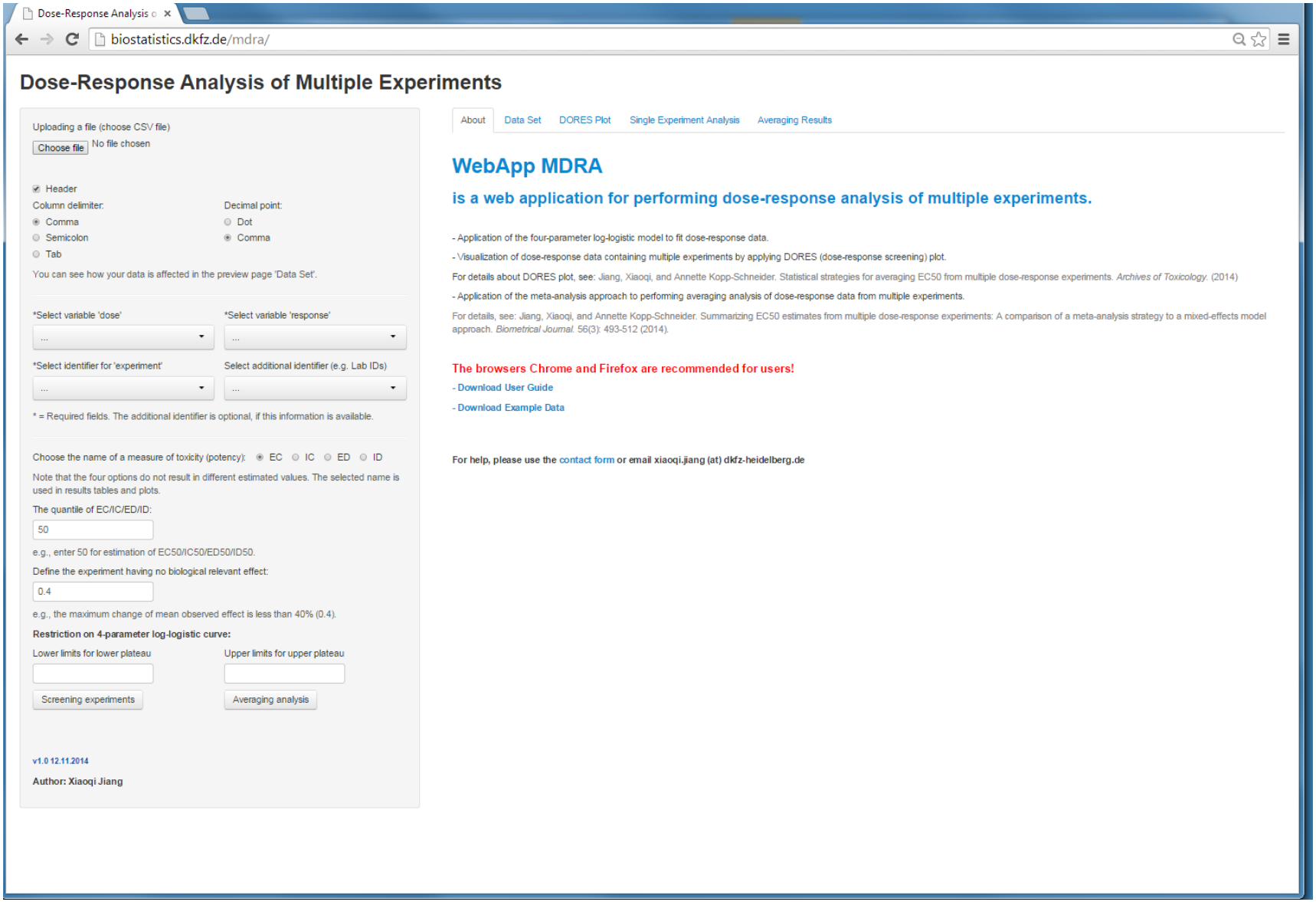
See Website: http://www.dkfz.de/en/biostatistics/software.html

…. Dose-response modeling

…. Web application for analysis of dose-response studies

Or directly http://biostatistics.dkfz.de/mdra/

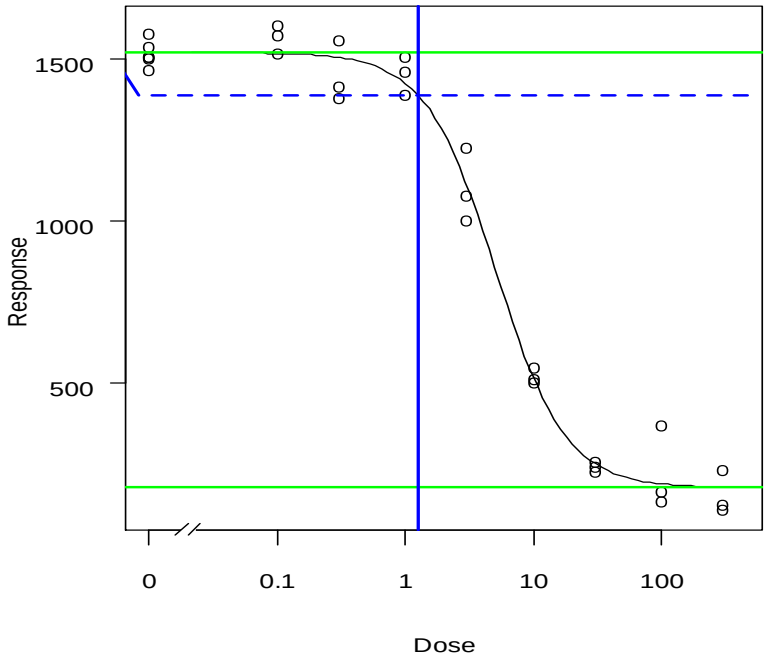
Dose-response analysis using WebApp (2)



Estimation of other quantiles (EDp)

For $0 \leq p \leq 100$:

$$EDp = ED50 \cdot \left(\frac{100}{100 - p} - 1 \right)^{1/\text{HillSlope}}$$



Example:

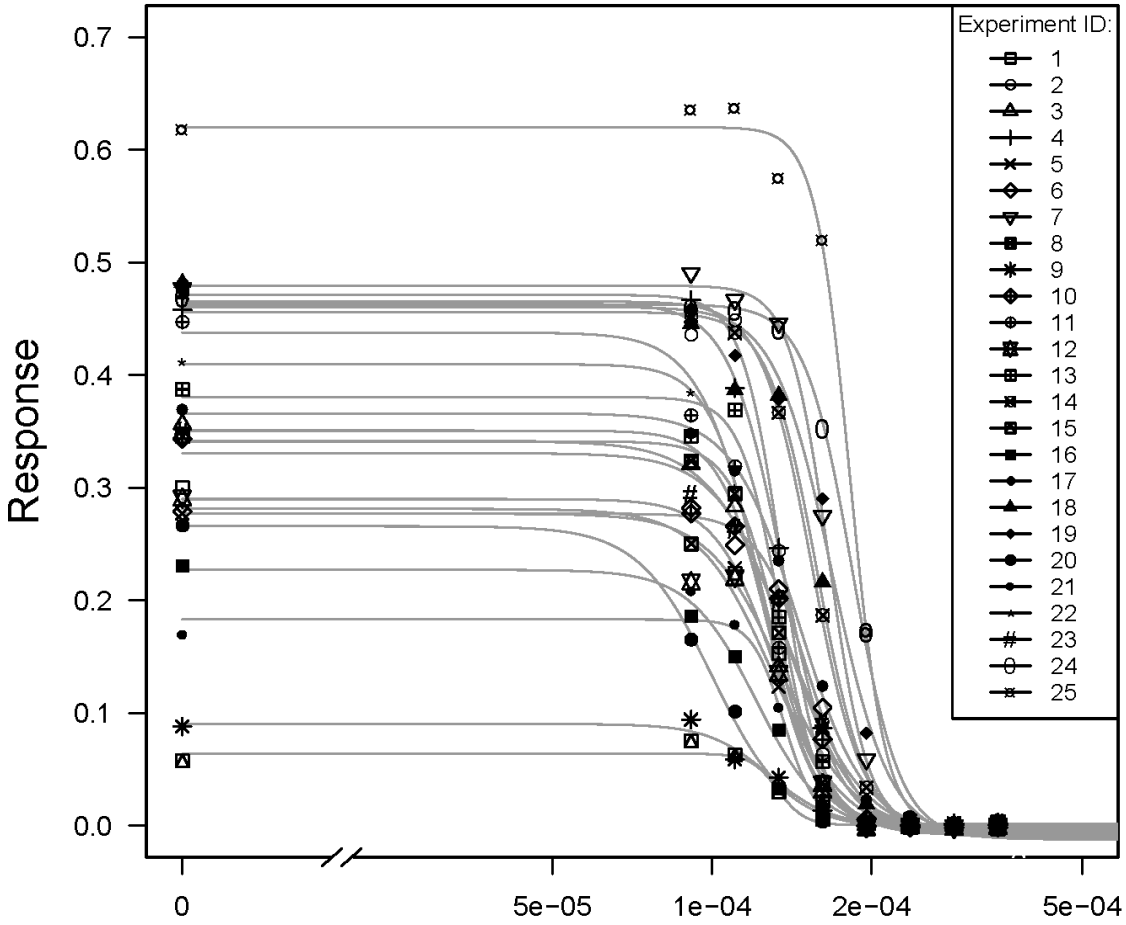
$$ED10 = ED50 \cdot \left(\frac{100}{100 - 10} - 1 \right)^{1/\text{HillSlope}} = 5.03 \cdot \left(\frac{100}{100 - 10} - 1 \right)^{1/1.59} = 1.27$$

Interpretation: Dose of 1.27 nM results in 10% of the maximal effect

Data transformation in dose-response analysis (1)

Scale response range for better data visualization/interpretation

→ More convenient comparison of experiments



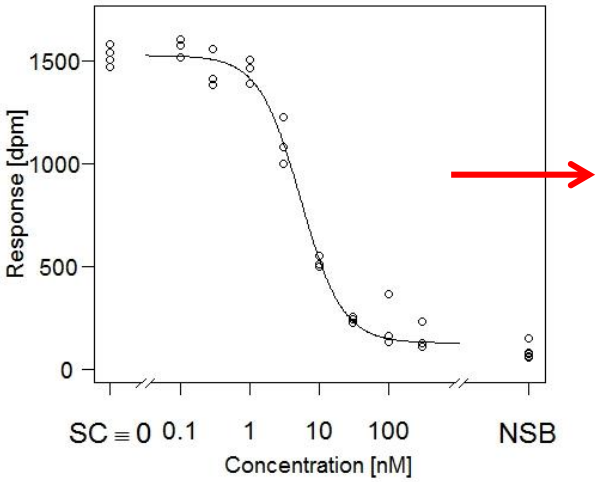
Data transformation in dose-response analysis (2)

Typical transformations of response:

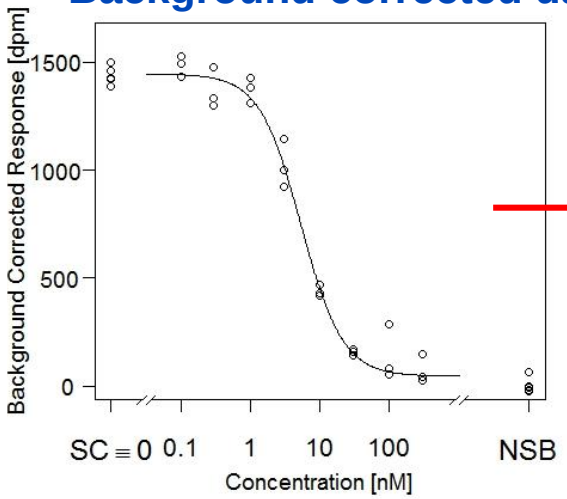
(a) background correction

(b) normalization: divide by mean of (background corrected) control

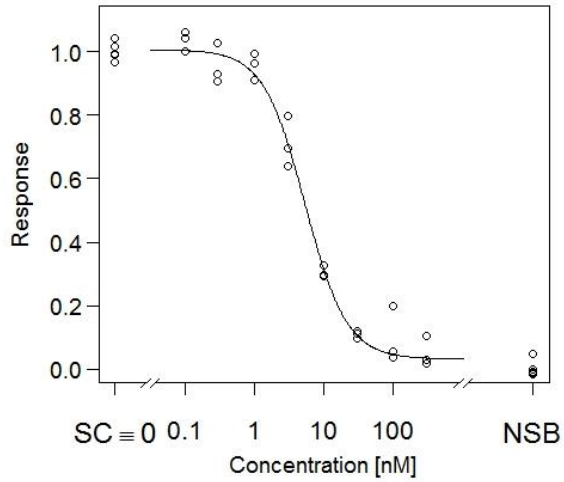
Original data



Background corrected data



Background corrected and normalized data



Evaluation of normalized response data: Common approaches (1)

- Use background correction
  → fix BOTTOM ≡ 0, use 3-parameter log-logistic model

- Use (background correction and) normalization
  → fix BOTTOM ≡ 0, TOP ≡ 1 (=100%), use 2-parameter log-logistic model

Fixing parameters leads to…

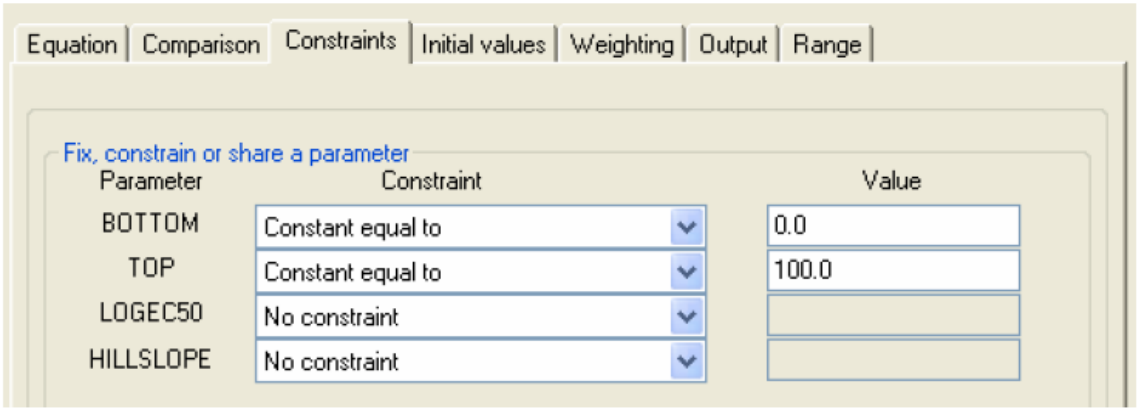
- Less numerical problems with curve fitting
- Smaller Confidence Intervals
- … and potentially **incorrect results**

Evaluation of normalized response data: Common approaches (2)

e.g. from GraphPad Prism ‘Analyzing dose-response data’:

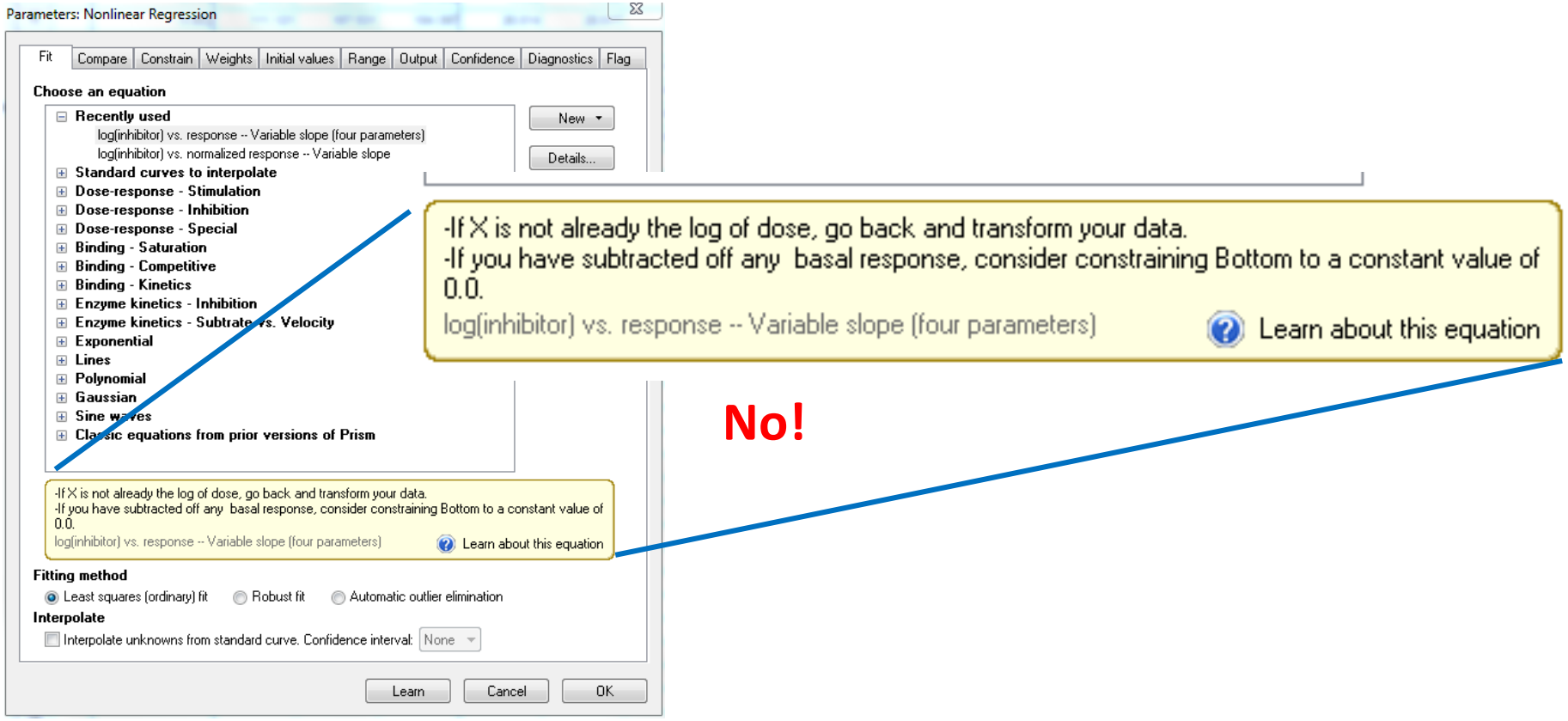
Constraining Curve-Fit Parameters

Since we normalized the original data such that the vertical range extends by definition from 0 to 100, it doesn't make sense to fit the "bottom" and the "top" of the curves. So we'll fix those parameters, leaving only the midpoint (log EC_{50}) and slope (Hill slope) of each curve to be fitted by Prism. Select the **Constaints** tab. Constrain the parameters BOTTOM and TOP to 0 and 100, respectively:



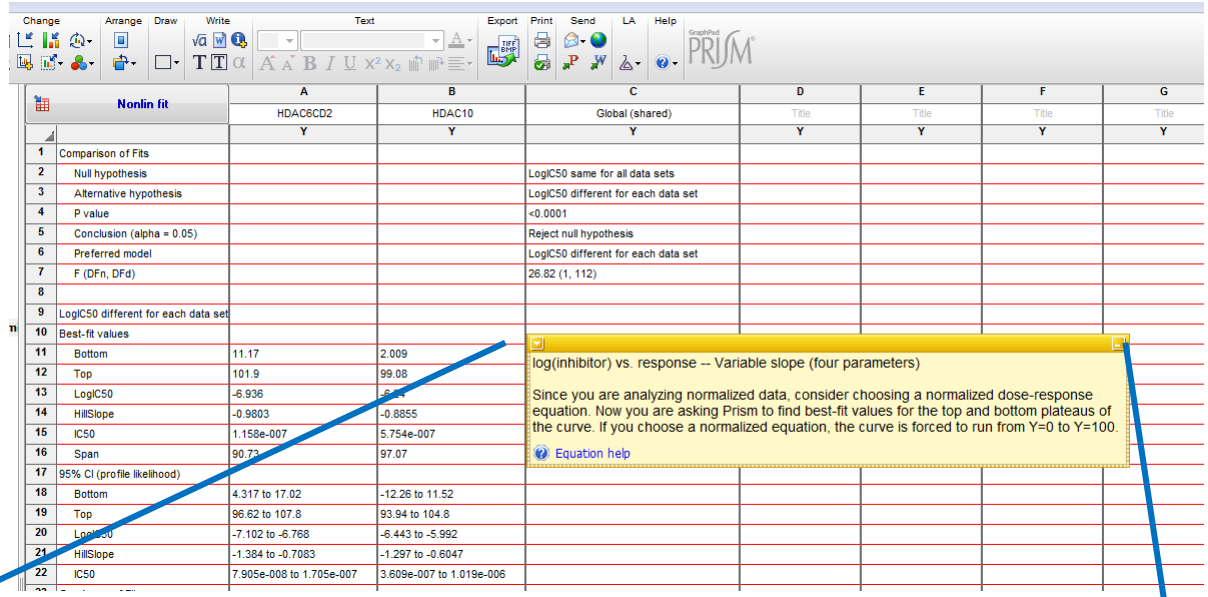
Evaluation of normalized response data: Common approaches (3)

e.g. GraphPad Prism



Evaluation of normalized response data: Common approaches (4)

e.g. GraphPad Prism



Nonlin fit	A	B	C	D	E	F	G
	HDAC6CD2	HDAC10	Global (shared)	Title	Title	Title	Title
	Y	Y	Y	Y	Y	Y	Y
1 Comparison of Fits							
2 Null hypothesis			LogIC50 same for all data sets				
3 Alternative hypothesis			LogIC50 different for each data set				
4 P value			<0.0001				
5 Conclusion (alpha = 0.05)			Reject null hypothesis				
6 Preferred model			LogIC50 different for each data set				
7 F (DFn, DFd)			26.82 (1, 112)				
8							
9 LogIC50 different for each data set							
10 Best-fit values							
11 Bottom	11.17	2.009					
12 Top	101.9	99.08					
13 LogIC50	-6.936	-6.44					
14 HillSlope	-0.9803	-0.8855					
15 IC50	1.158e-007	5.754e-007					
16 Span	90.73	97.07					
17 95% CI (profile likelihood)							
18 Bottom	4.317 to 17.02	-12.26 to 11.52					
19 Top	96.62 to 107.8	93.94 to 104.8					
20 LogIC50	-7.102 to -6.768	-6.443 to -5.992					
21 HillSlope	-1.384 to -0.7083	-1.297 to -0.6047					
22 IC50	7.905e-008 to 1.705e-007	3.609e-007 to 1.019e-006					

log(inhibitor) vs. response -- Variable slope (four parameters)

Since you are analyzing normalized data, consider choosing a normalized dose-response equation. Now you are asking Prism to find best-fit values for the top and bottom plateaus of the curve. If you choose a normalized equation, the curve is forced to run from Y=0 to Y=100.

Equation help

No!

This problem should be fixed now.

Correct approach

Toxicology Letters 213 (2012) 292–298

Contents lists available at SciVerse ScienceDirect

Toxicology Letters

journal homepage: www.elsevier.com/locate/toxlet

The impact of data transformations on concentration–response modeling

Marc Weimer[a,1], Xiaoqi Jiang[a,1], Oriana Ponta[a], Sven Stanzel[a], Alexius Freyberger[b], Annette Kopp-Schneider[a,*]

[a] Department of Biostatistics, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[b] Bayer Pharma AG, GGD GED Toxicology, Department of Pathology and Clinical Pathology, Aprather Weg 18, 42096 Wuppertal, Germany

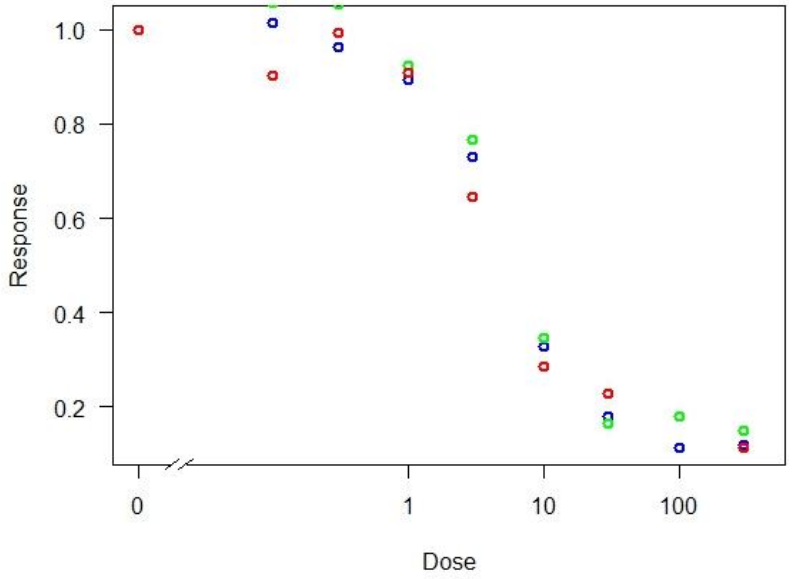
Effects of normalization and fixing parameters

	TOP	BOTTOM	HillSlope	ED50 (95%-CI)
Original data	1522	177	1.59	5.03 nM [4.26, 5.93]
Background corrected and normalized data 4-par model	1.00	0.12	1.59	5.03 nM [4.26, 5.93]
Background corrected data Fix BOTTOM≡0, Use 3-par model	1451	-	1.32	5.76 nM [4.86, 6.82]
Background corrected and normalized data Fix BOTTOM≡0, TOP≡1, Use 2-par model	-	-	1.35	5.88 nM [5.06, 6.84]

When using (background corrected and) normalized data: fit 4-parameter model!

A special situation where fixing TOP is legitimate

- Several experiments
- Each experiment: control + several doses, 1 replicate each
- Measurements are normalized to control (fold-change data)
- Here: evaluate with 3-parameter log-logistic model, fixing TOP=1,
  and exclude control data.



Model assessment (1)

- Plot data together with fitted curve: Is the fitted curve close to the data?
- Use diagnostic plots to assess whether model assumptions are violated (like in linear regression): plot residuals
- Compute CIs: wide CIs indicate problems
- Consider R^2:
  Fraction of the total variance (of response) explained by the model equation

  $R^2 = 0$: best-fit curve is no better than a horizontal line going through overall mean response

  $R^2 = 1$: perfect fit

Model assessment (2)

Residual Plot (with SigmaPlot)

SigmaPlot - NotebookDR.JNB*

Home | Worksheet | Create Graph | Graph Page | Analysis | Report | ToolBox | Ribbon | Style

Graph Wizard... | Scatter | Line | Line/Scatter | Bar | Area | Polar | Radar | Ternary | Box | Pie | Bubble Plot | Vector | Contour | 3D Scatter | 3D Line | 3D Mesh | 3D Bar | Add to Gallery

Wizard | 2DGraphs | 3DGraphs | Graph Galler

Notebook Manager

Notebook

- All Open Notebooks
- NotebookDR.JNB*
  - Section 1
    - Data 1*
    - Report 1
    - Graph Page 1*
    - Graph Page 2*

Summary Info

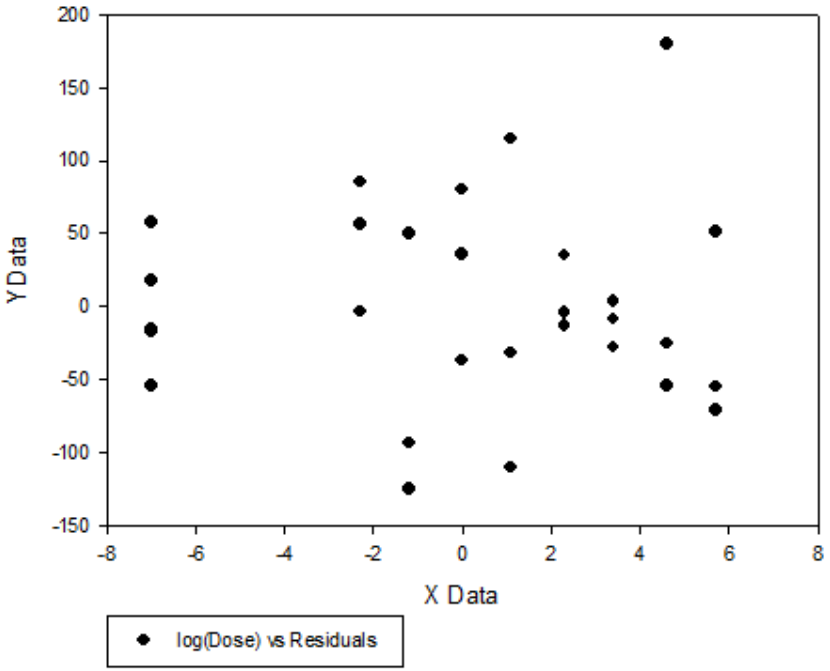
Created	10/28/2013 7:3...
Modified	10/28/2013 10:...
Author	admin
Description	Worksheet

Gallery | Notebook Manager

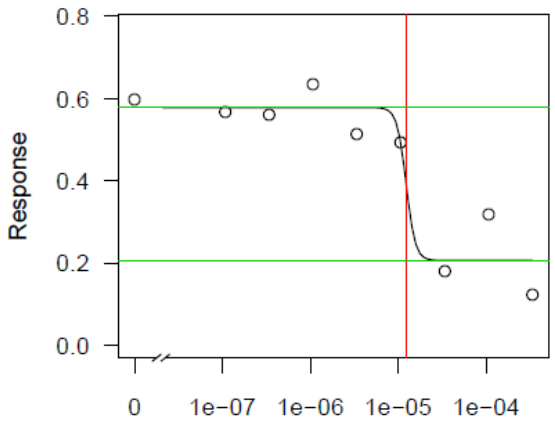
Report 1 | Data 1*

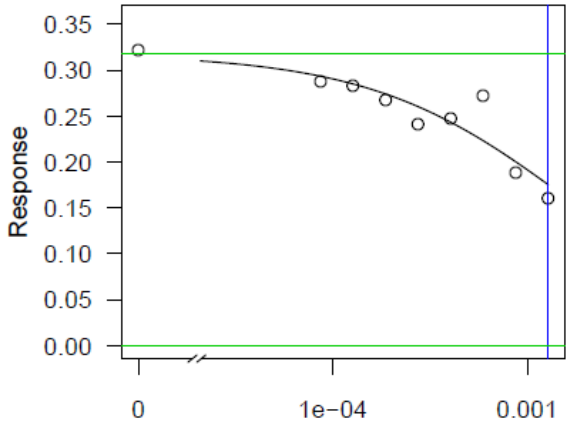
	1-Dose	2-Response	3-Parameters	4-Predicted	5-Residuals	6-log(Dose)
1	0.1000	1516.0000	177.1656	1519.7370	-3.7370	-2.3026
2	0.1000	1576.0000	1522.3380	1519.7370	56.2630	-2.3026
3	0.1000	1605.0000	5.0297	1519.7370	85.2630	-2.3026
4	0.3000	1557.0000	-1.5943	1507.4842	49.5158	-1.2040
5	0.3000	1414.0000		1507.4842	-93.4842	-1.2040
6	0.3000	1382.0000		1507.4842	-125.4842	-1.2040
7	1.0000	1463.0000		1427.1826	35.8174	0.0000
8	1.0000	1507.0000		1427.1826	79.8174	0.0000
9	1.0000	1390.0000		1427.1826	-37.1826	0.0000
10	3.0000	1227.0000		1112.1353	114.8647	1.0986
11	3.0000	1080.0000		1112.1353	-32.1353	1.0986
12	3.0000	1002.0000		1112.1353	-110.1353	1.0986
13	10.0000	510.0000		514.2053	-4.2053	2.3026
14	10.0000	501.0000		514.2053	-13.2053	2.3026
15	10.0000	549.0000		514.2053	34.7947	2.3026
16	30.0000	242.0000		250.9168	-8.9168	3.4012
17	30.0000	223.0000		250.9168	-27.9168	3.4012
18	30.0000	254.0000		250.9168	3.0832	3.4012
19	100.0000	163.0000		188.5142	-25.5142	4.6052
20	100.0000	368.0000		188.5142	179.4858	4.6052
21	100.0000	134.0000		188.5142	-54.5142	4.6052
22	300.0000	230.0000		179.1485	50.8515	5.7038
23	300.0000	124.0000		179.1485	-55.1485	5.7038
24	300.0000	108.0000		179.1485	-71.1485	5.7038
25	0.0000	1468.0000		1522.3380	-54.3380	-7.0000
26	0.0000	1505.0000		1522.3380	-17.3380	-7.0000
27	0.0000	1540.0000		1522.3380	17.6620	-7.0000
28	0.0000	1580.0000		1522.3380	57.6620	-7.0000
29	0.0000	1507.0000		1522.3380	-15.3380	-7.0000
30	0.0000	1507.0000		1522.3380	-15.3380	-7.0000

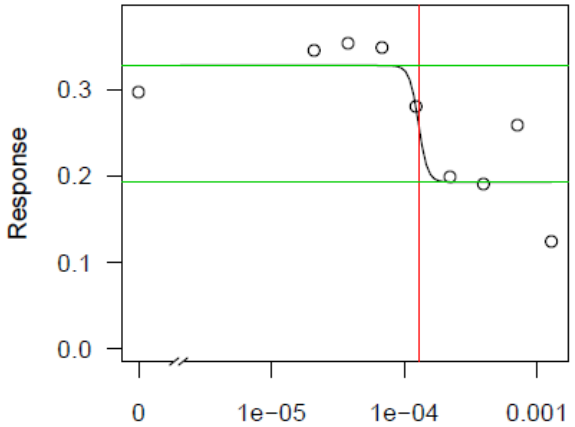
OVR | REC | MACRO | AUTO | CAP | NUM | SCRL | 100%

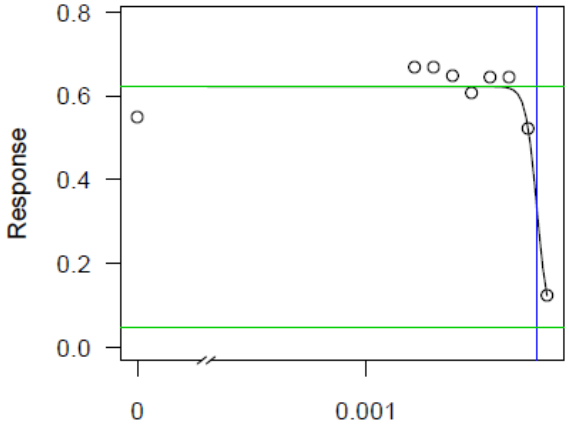


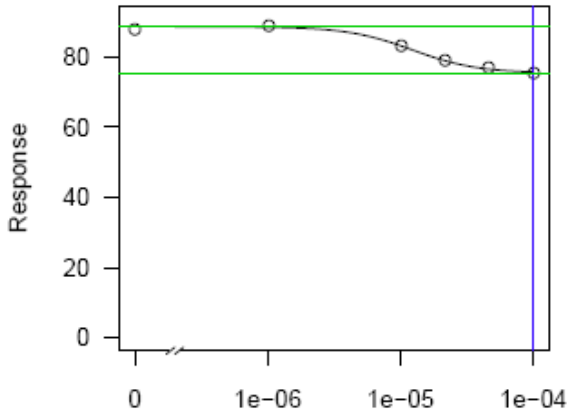
Model assessment (3)











Experimental design of dose-response experiments

First step: **Range finder experiment**

dose levels should span wide dose range

how many dose levels?

- problem-specific
- the more, the better

small number of replicates/dose level

Second step: **Main experiment**

Optimal experimental design for main experiment

Choice of design depends on which measure of potency should be estimated

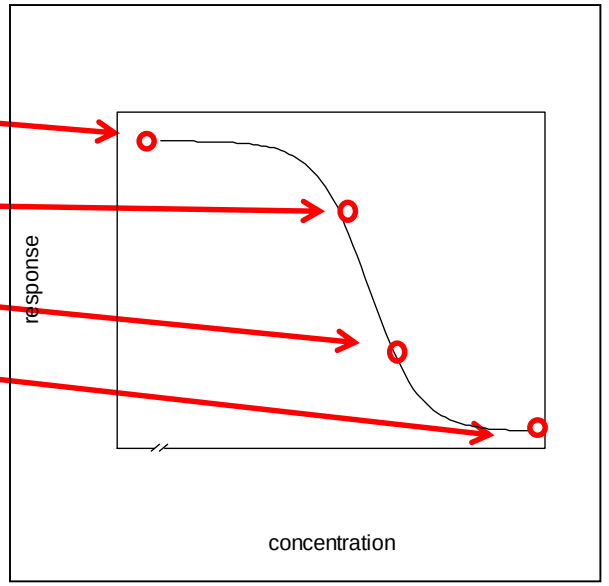
Here: ED50

control
+ 3 dose levels:

1 at start of ‚linear range‘

1 at end of ‚linear range‘

1 to determine lower plateau



25% measurements at each dose level

Optimal experimental design for main experiment using WebApp (1)

See Website: http://www.dkfz.de/en/biostatistics/software.html

…. Dose-response modeling

…. Web application for design of dose-response studies

Or directly http://biostatistics.dkfz.de/DoseResponseDesigns/

Requires rough estimates of ED50 and HillSlope

Optimal experimental design for main experiment using WebApp (2)

Optimal Experimental Design for single substance and interaction trials

This Application allows computation of D-optimal designs for interaction trials in a dose response context. Designs are computed for two singular treatments as well as up to 5 combination treatments. Furthermore, the efficiency of prespecified designs can be checked, and more
robust designs suitable for several parameter conditions can be computed.

For details see:
Holland-Letz, T and Kopp-Schneider, A (2020): *An R-Shiny application to calculate optimal designs for single substance and interaction trials in dose response experiments* (under review)
For the (outdated) previous application from *Optimal experimental designs for dose-response studies with continuous endpoints*, Archives of Toxicology (2015), 89(11), 2059-68, see https://biostatistics-dkfz.shinyapps.io/dosis/

Two different dose response functions can be considered:

$$Log-logistic: y = c + \frac{d-c}{1+\exp^{b(ln(x)-ln(e))}}$$

$$Weibull: y = c + (d-c)\exp^{(-\exp^{(-b(ln(x)-ln(e)))})}$$

Basic settings for design algorithm

Lowest log dose level: -10

Highest log dose level: 10

Number of available dose levels (min 10): 101

Reduction parameter: 0,99

Number of iterations for algorithm (min 50): 500

1. Compute Optimal Designs | 2. Check efficiency of specific designs | 3. Quasi-Bayesian Designs | 4. Compute Optimal Designs for Interactions

This part computes D-optimal designs for a single treatment on the specified design space. One of two available dose response functions can be chosen, and an a priori assumption regarding the assumed slope and ED50 parameters can be made.
Lowering the value for the reduction parameter will try to find a design with fewer support points, but might reduce the efficiency.

Function Parameters

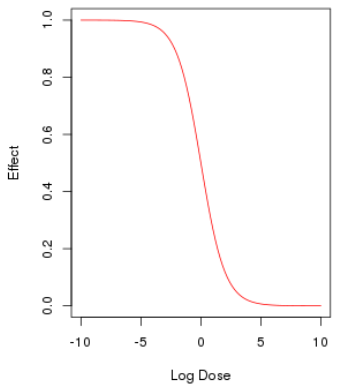
Function:
- (•) Log-logistic
- ( ) Weibull

Slope (Parameter b): 1

ED50 (Parameter e): 1

Compute

Plot of function



Optimal experimental design for main experiment using WebApp (3)

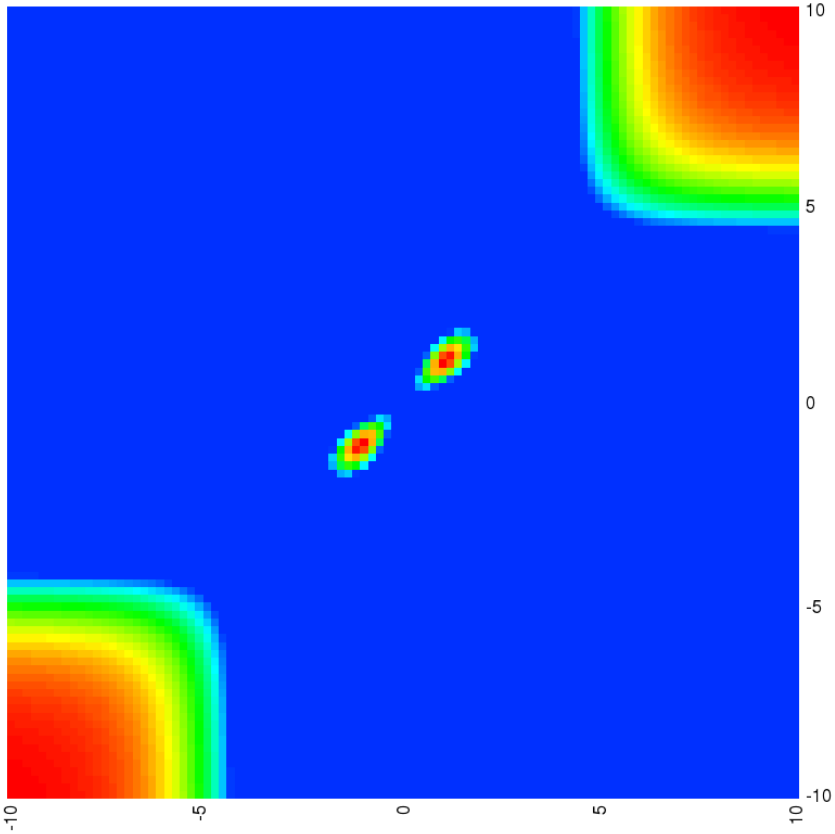
Result

Proposed designs, values of parameters b and e, and resulting D-Efficiency of proposed designs:

	Design	Design	Design	Design	e	b	D-Eff
LogDose	-10.00	-1.00	1.00	10.00	1	1	1.00
Weight	0.25	0.25	0.25	0.25	NA	NA	NA

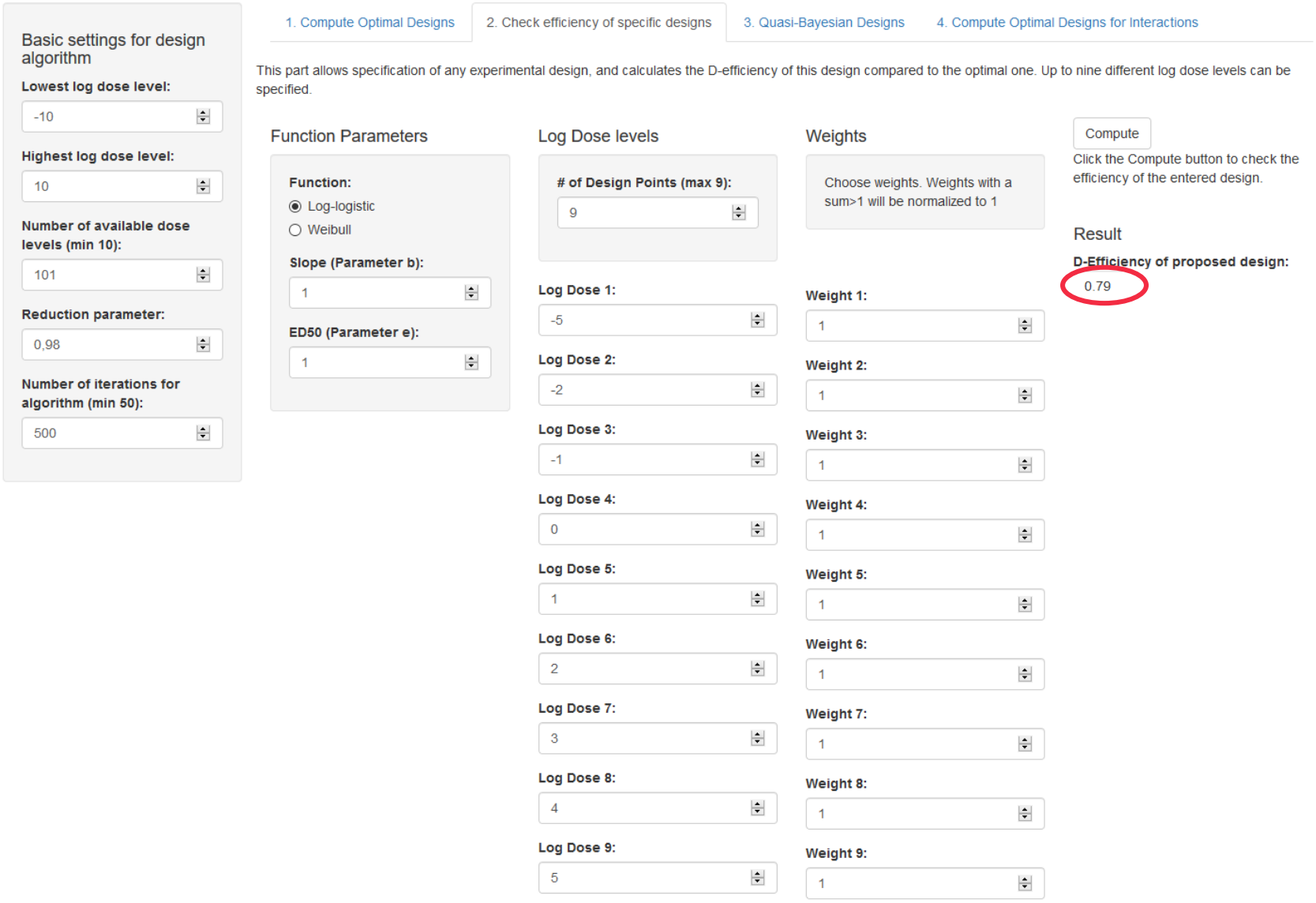
Design Heatmap

Points marked red on the diagonal are potential design points. Pairs of different design points marked red when crossreferenced are interchangeable with negligible loss of efficiency.



/ersion 3.2, 13Oct2020
Contact: t.holland-letz(at)dkfz.de

Optimal experimental design for main experiment using WebApp (4)



Take home: Recommendations for practical use

- Always plot your raw data (on log-scaled dose) before fitting a (dose-response) model.
- Fit dose-response model to all data points instead of mean response per dose level.
- Use control measurements for model fitting (typically).
- Fix model parameters only if there is compelling reason to do so.
- Use 4-parameter dose-response model even if data have been (background-corrected and) normalized.
- Plot data points with fitted curve, inspect residuals.
- Check whether model assumptions are violated.
- Compute confidence intervals of model parameters (e.g. ED50) to assess precision of parameter estimates.

Software

- SigmaPlot
- GraphPad Prism
- R package drc
- WebApps at http://www.dkfz.de/en/biostatistics/software.html
  - MDRA
  - DoseResponseDesigns

References

- Holland-Letz T, Kopp-Schneider A. (2015) Optimal experimental designs for dose-response studies with continuous endpoints. *Archives of Toxicology* **89**(11):2059-68.
- Jiang, X, and Kopp-Schneider A. (2015) Statistical strategies for averaging EC50 from multiple dose-response experiments. *Archives of Toxicology* **89**(11):2119-27.
- Kappenberg F, Brecklinghaus T, Albrecht W, Blum J, van der Wurp C, Leist M, Hengstler JG, Rahnenführer J. (2020) Handling deviating control values in concentration-response curves. *Archives of Toxicology* 2020 online.
- Motulsky, H., Christopoulos, A. (2004). *Fitting Models to Biological Data Using Linear and Nonlinear Regression. A practical guide to curve fitting.* Oxford University Press, New York.
- Ritz, C. (2010). Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry* **29**, 220-229.
- Ritz C, Baty F, Streibig JC, Gerhard D (2015) Dose-Response Analysis Using R. PLoS ONE 10(12): e0146021. doi:10.1371/journal.pone.0146021
- Weimer, M., Jiang, X., Ponta, O., Stanzel, S., Freyberger, A., Kopp-Schneider, A. (2012). The impact of data transformations on concentration-response modeling. *Toxicology Letters* **213**, 292-298.

Next lexture

4 November

Non-parametric methods