EFSA Guidance for BMD analysis

Fitting Models & Goodness of Fit

1st March 2017

European Food Safety Authority

OUTLINE

- **General** principles of model fitting & goodness of fit

- **Continuous** dose-response data
  - Cell proliferation (CP) data
  - Candidate models
  - Illustration

- **Quantal** response data
  - Thyroid epithelial cell vacuolisation (TECV) data
  - Candidate models
  - Illustration

GENERAL PRINCIPLES

- Data vary according to a **distribution**
  - Continuous: log-normal with mean and variance
  - Quantal: binomial with probability parameter
- Model the **mean** parameter as **function of dose**
- **Fit** the model to data and get estimates (ML)
- **Compare** different **candidate models** using a goodness-of-fit criterion (AIC)

CONTINOUS RESPONSE DATA

Cell Proliferation Example

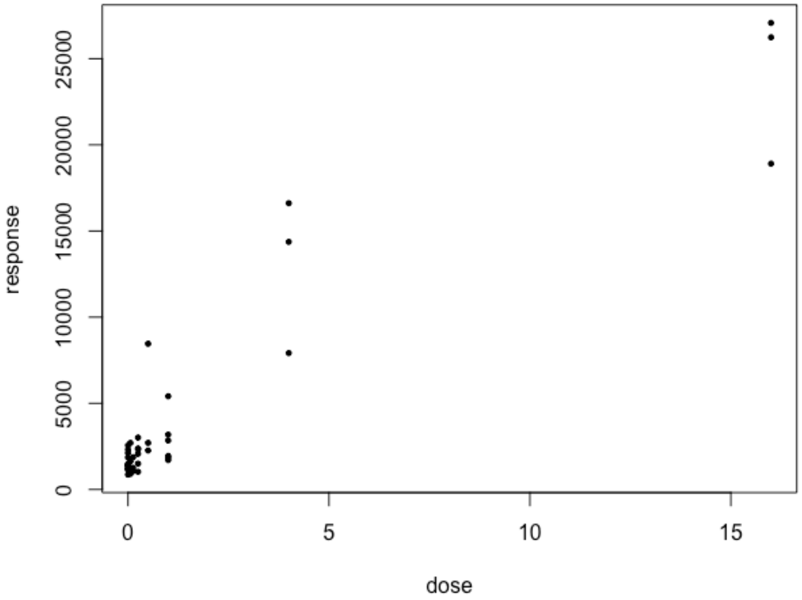
CELL PROLIFERATION

- Local Lymph Node Assay (LLNA) test
- Cell proliferation as an indicator of sensitization
- Related to 15 rubber chemicals, focus on one chemical
- Dose levels and number of observations per level

dose							
0	0.06	0.125	0.25	0.5	1	4	16
12	6	3	6	3	6	3	3

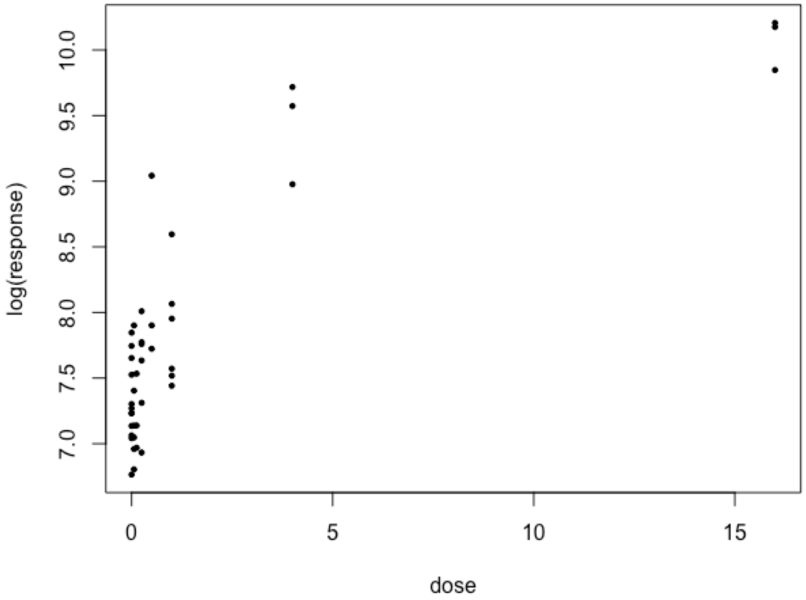
CELL PROLIFERATION

- Data of dose and cell proliferation on original scale



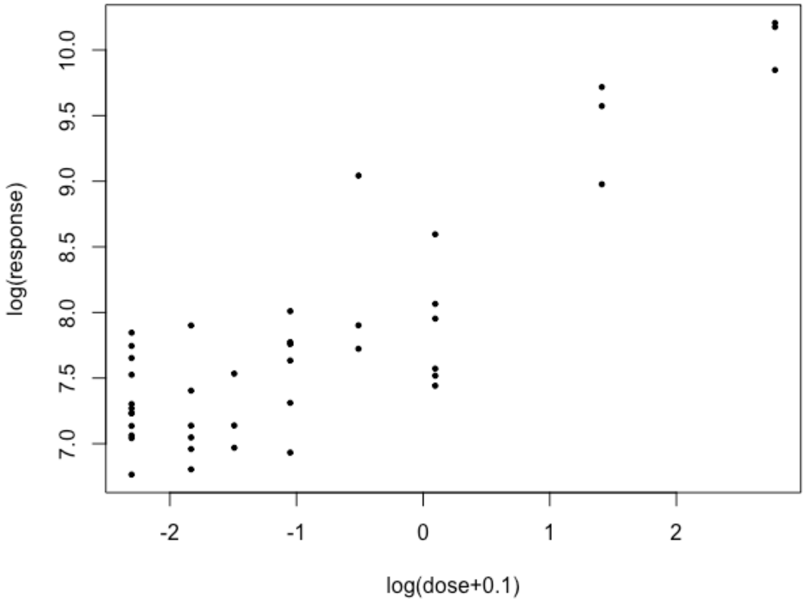
CELL PROLIFERATION

- Data of cell proliferation on log scale



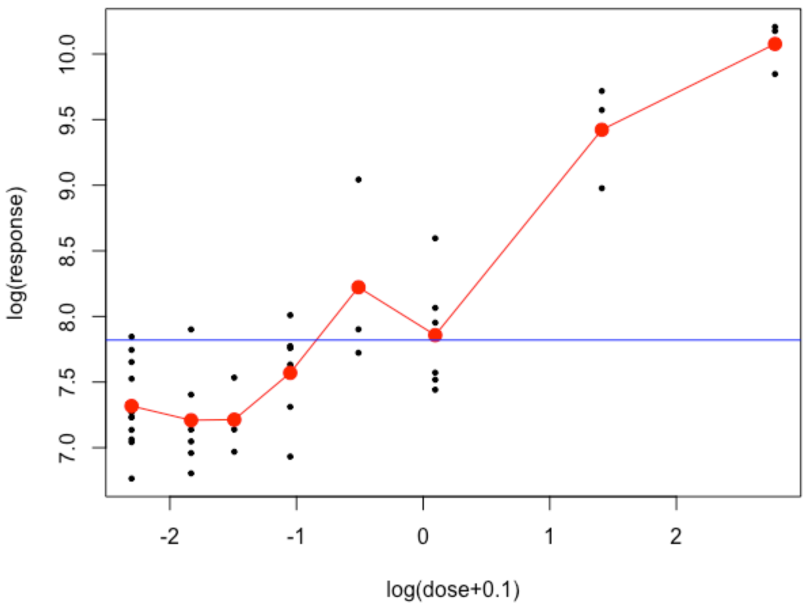
CELL PROLIFERATION

- Data of dose and cell proliferation both on log scale



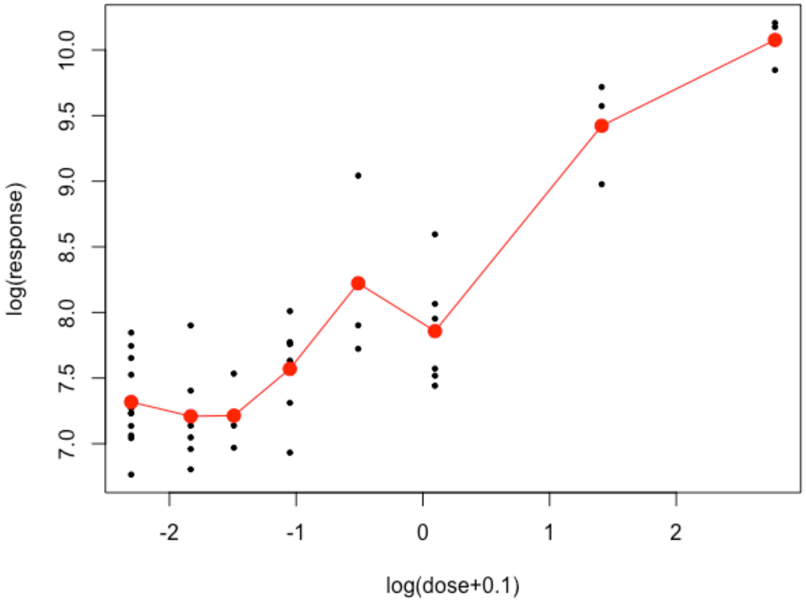
CELL PROLIFERATION

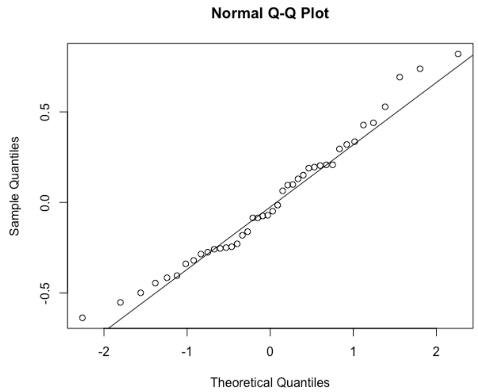
- **Null** model of **no effect**
- **Full** model: each dose level its own mean

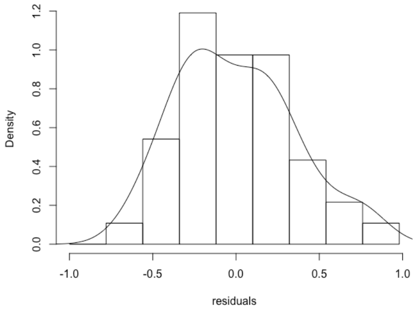


CELL PROLIFERATION

- Equal variances: Levene's test (p=0.89)
- Normal distribution: Shapiro-Wilk (p=0.42)







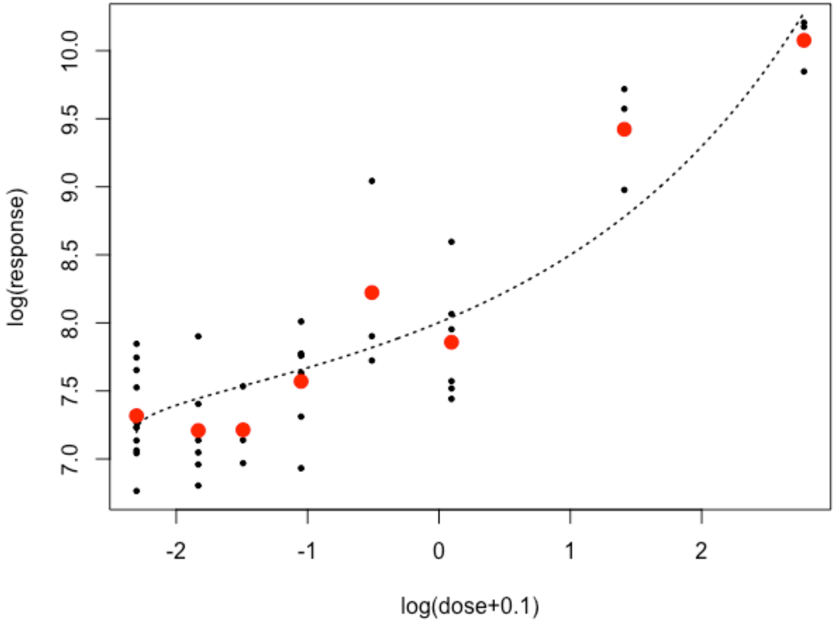
CELL PROLIFERATION

Candidate models for continuous data are models for the **mean** of the normal distribution on log-scale

Continuous data		
Exponential family		
3-parameter model(iii)	3	$y = a \exp(bx^d)$
4-parameter model(iv)	4	$y = a\,[c-(c-1)\exp(-bx^d)]$
Hill family		
3-parameter model(iii)	3	$y = a\,[1-x^d/(b^d + x^d)]$
4-parameter model(iv)	4	$y = a\,[1 + (c-1)x^d/(b^d + x^d)]$

CELL PROLIFERATION

- 3 parameter Exponential model $y = a\,\exp(bx^{d})$




```
Coefficients:
         Estimate Std. Error
a       7.1959136 0.11088539
b       0.1110707 0.02044363
d       0.4206736 0.05800638
sigma 0.4392835 0.04792905

'log Lik.' -25.04554
```

MAXIMUM LIKELIHOOD ESTIMATION & INFERENCE

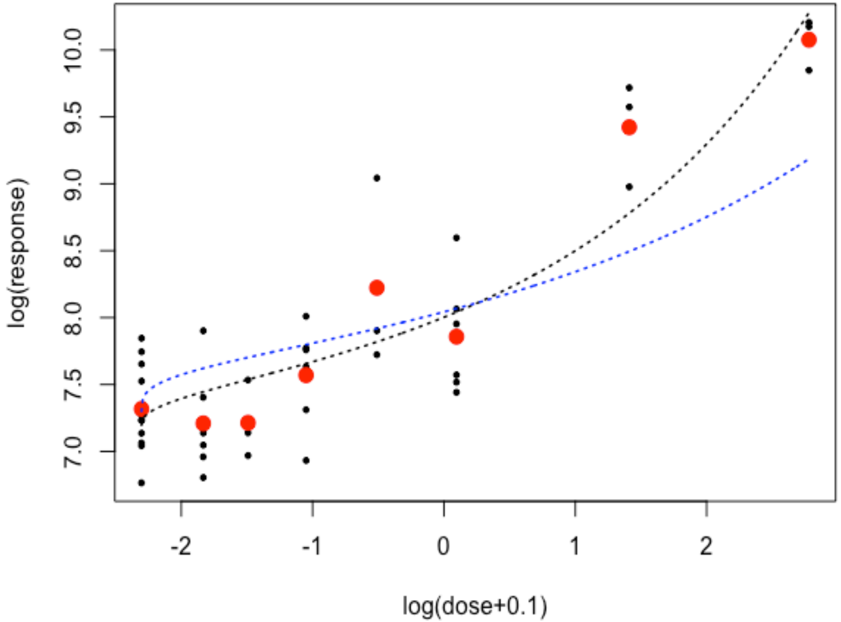
$$y = a \exp(bx^{d})$$

Maximum Likelihood (ML) estimation & inference

- Values for *a*, *b* and *d* that maximize the likelihood
- The likelihood of the data you observed
- Maximize likelihood = maximize log-likelihood

CELL PROLIFERATION

- Maximum Likelihood (ML) estimation & inference



ML (most likely) values for the parameters

```
Coefficients:
       Estimate Std. Error
a     7.1959136 0.11088539
b     0.1110707 0.02044363
d     0.4206736 0.05800638
sigma 0.4392835 0.04792905

'log Lik.' -25.04554
```

Less likely values for the parameters

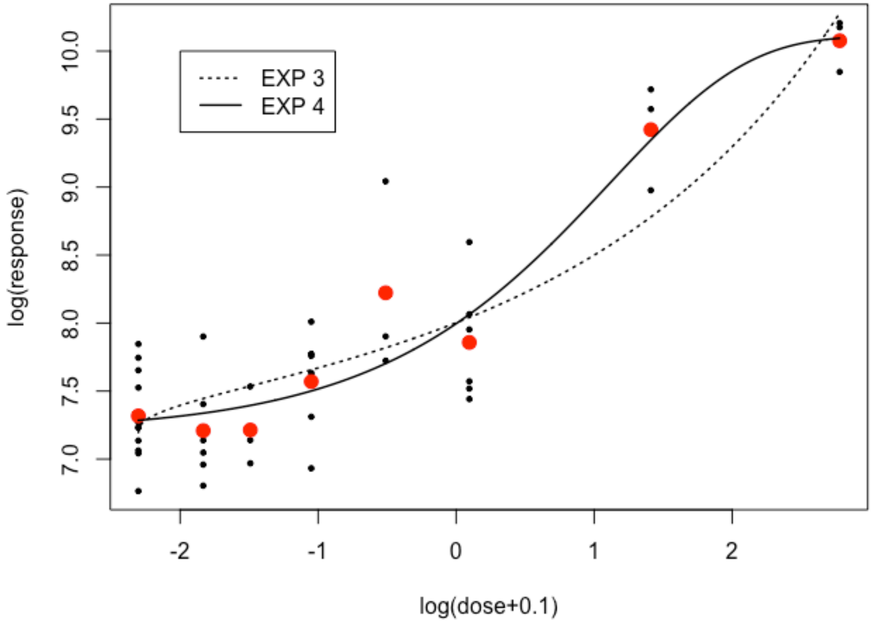
```
a = 7.3
b = 0.1
d = 0.3
log Lik = -37.099
```

MAXIMUM LIKELHOOD ESTIMATION & INFERENCE

- ML is identical to **least squares estimation** in case of the **normal** distribution (possibly on log-scale)
- Difficulty for **non-linear models**: multiple local maxima and **convergence issues**

CELL PROLIFERATION

- 4 parameter Exponential model $y = a\,[c-(c-1)\exp(-bx^{d})]$



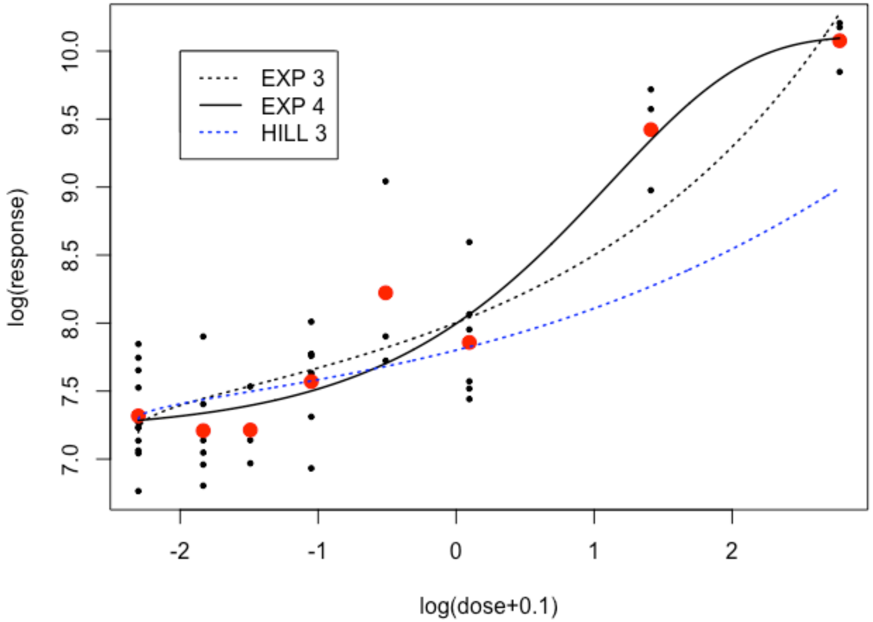
Coefficients:

	Estimate	Std. Error
a	7.2850093	0.09654781
b	0.3227275	0.08541427
c	1.3875325	0.03960234
d	1.0104150	0.22725460
sigma	0.3935590	0.04293928

'log Lik.' -20.42961

CELL PROLIFERATION

- 3 parameter Hill model $y = a\,[1 - x^d/(b^d + x^d)]$



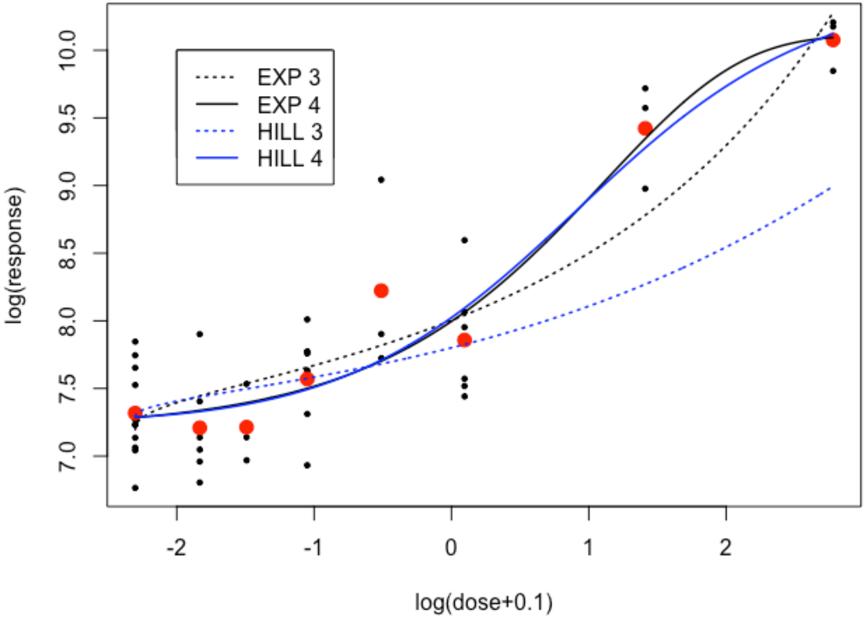

```
Coefficients:
        Estimate     Std. Error
a       7.2953731   0.11004240
b     183.7769316  57.44395880
d       0.4884650   0.06802027
sigma   0.4643575   0.05335079

'log Lik.' -27.06396
```

![efsa European Food Safety Authority]

CELL PROLIFERATION

- 4 parameter Hill model $y = a\,[1 + (c-1)x^d/(b^d + x^d)]$




```
Coefficients:
       Estimate  Std. Error
a      7.2881474 0.09865037
b      2.5468752 1.00250825
c      1.4354407 0.07563080
d      1.1524649 0.33549528
sigma  0.3963275 0.04324196

'log Lik.' -20.7236
```

European Food Safety Authority

AKAIKE INFORMATION CRITERION

- **AIC** stems from information theory
- Measures the **information lost** by replacing the true unknown data generating model by the model used
- Estimates the **Kullback-Leibler distance**
- The **smaller AIC** the better the model
- Defined as

AIC = -2 loglikelihood + 2 #par

- The lower -2 loglikelihood the closer the fit to the data
- **Penalization** for overfitting, for too complex models
- AIC balances accuracy with complexity

CELL PROLIFERATION

Model	Log-likelihood	# par
Null	-56.006	2
Full	-15.659	9
EXP 3	-25.046	4
EXP 4	-20.430	5
HILL 3	-27.064	4
HILL 4	-20.724	5

par = number of parameters in mean model + variance parameter

CELL PROLIFERATION

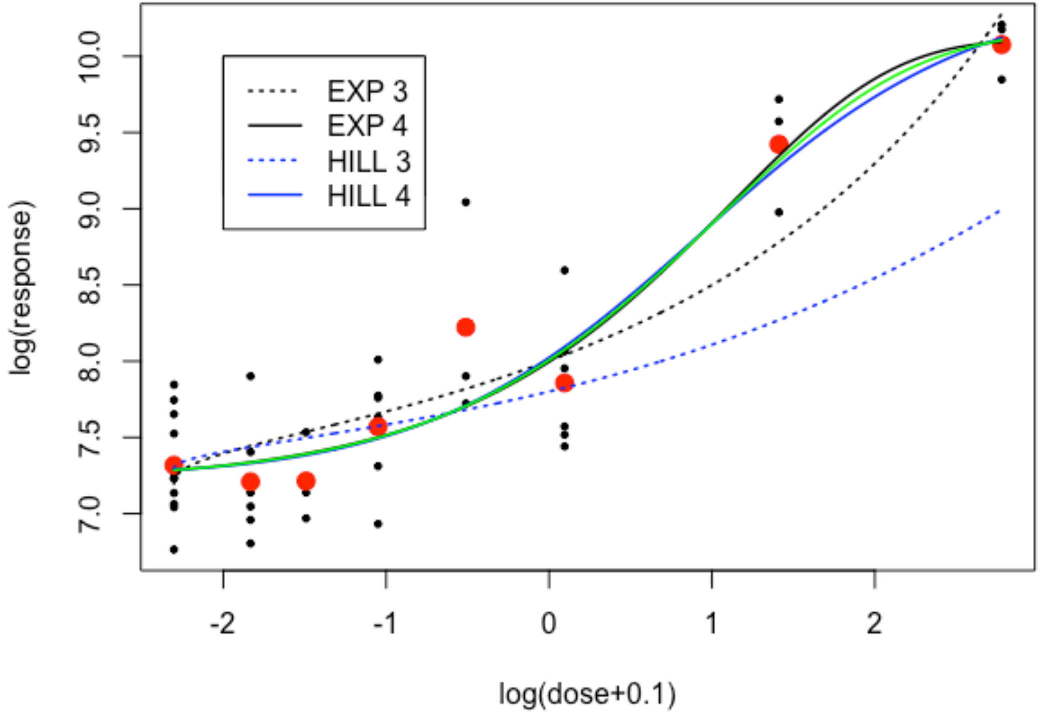
Model	Log-likelihood	# par	AIC
Null	-56.006	2	116.012
Full	-15.659	9	49.319
EXP 3	-25.046	4	58.091
EXP 4	-20.430	5	50.859
HILL 3	-27.064	4	62.128
HILL 4	-20.724	5	51.447

$AIC_{Min} = 50.859 \quad < \quad AIC_{FULL} + 2 = 51.319$

All $\quad AIC < AIC_{NULL} - 2 = 114.012$

CELL PROLIFERATION

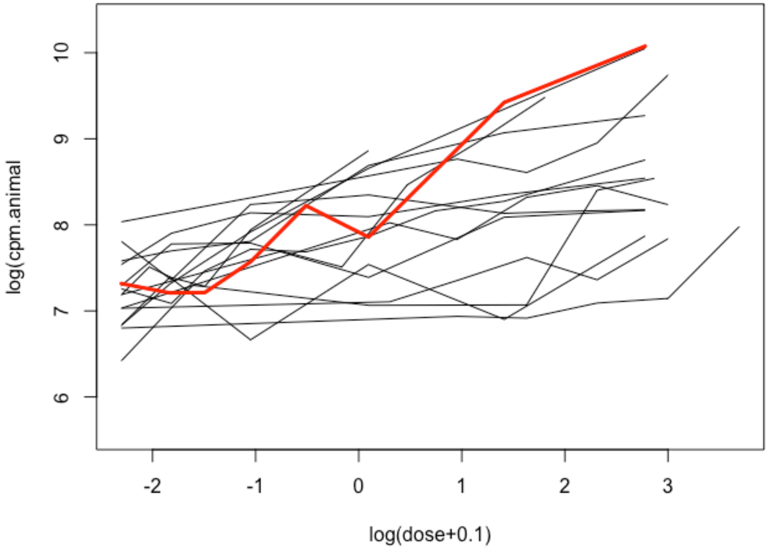
Model averaged fitted curve



Model	AIC	Weight
EXP 3	58.091	0.0056
EXP 4	**50.859**	**0.5693**
HILL 3	62.128	0.0007
HILL 4	**51.447**	**0.4243**

CELL PROLIFERATION

Just one compound, but there are 15 compounds



Approach: include compound as **covariate**, leading to a 4 parameter exponential model, with compound specific values for a and b and common values for c and d

QUANTAL RESPONSE DATA

thyroid epithelial cell vacuolisation

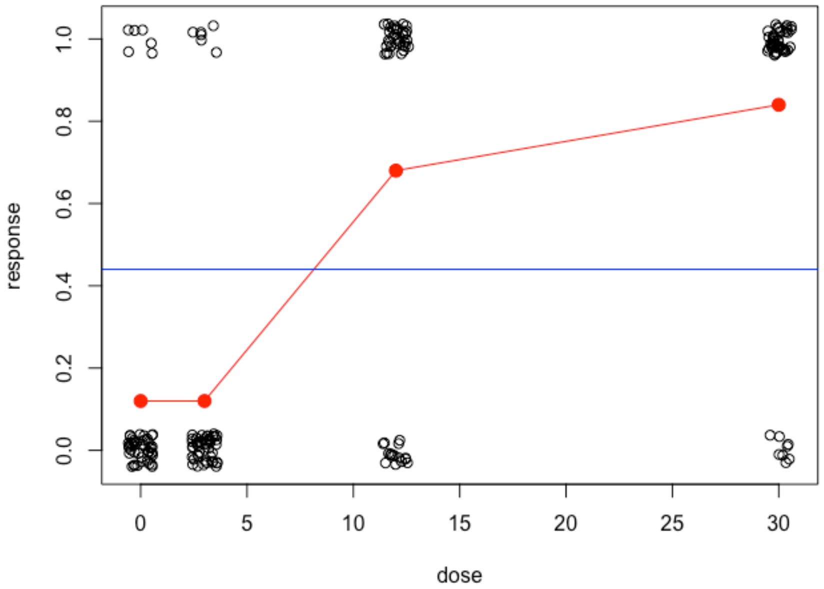
THYROID EPITHELIAL CELL VACUOLISATION

- 2 year study in rats
- Three doses of a substance
- Changes in thyroid epithelial cell vacuolisation

Dose (mg/kg day)	No of animals with thyroid epithelial vacuolisation	No of animals in dose group
0	6	50
3	6	50
12	34	50
30	42	50

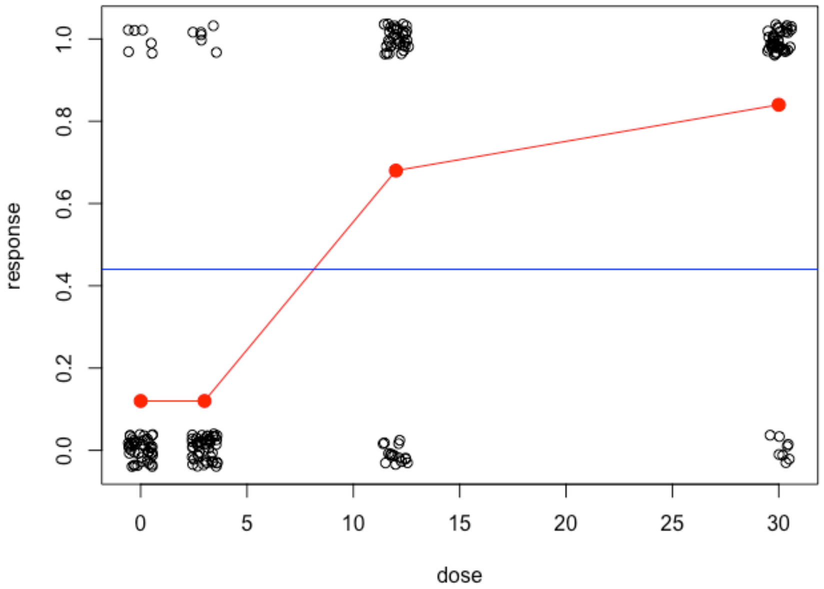
THYROID EPITHELIAL CELL VACUOLISATION

- **Null** model of **no effect**
- **Full** model: each dose level its own probability



THYROID EPITHELIAL CELL VACUOLISATION

- **Binomial** distribution essentially the only choice
- Extensions: beta-binomial model & hierarchical data models



THYROID EPITHELIAL CELL VACUOLISATION

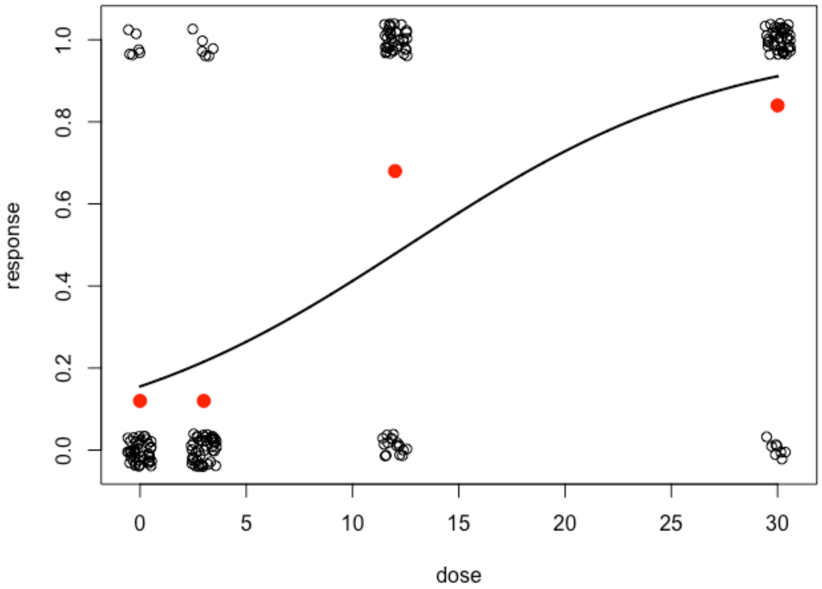
Candidate models for quantal data

Quantal data

Logistic	2	$y = 1/(1 + \exp(-a-bx))$
Probit	2	$y = \mathrm{CumNorm}(a + bx)$
Log-logistic	3	$y = a + (1-a)/(1 + \exp(-\log(x/b)/c))$
Log-probit	3	$y = a + (1-a)\ \mathrm{CumNorm}(\log(x/b)/c)$
Weibull	3	$y = a + (1-a)\ \exp((x/b)^c)$
Gamma	3	$y = a + (1-a)\ \mathrm{CumGam}(bx^c)$
LMS (two-stage) model	3	$y = a + (1-a)(1-\exp(-bx-cx^2))$

THYROID EPITHELIAL CELL VACUOLISATION

- Logistic model $y = 1/(1 + \exp(-a - bx))$

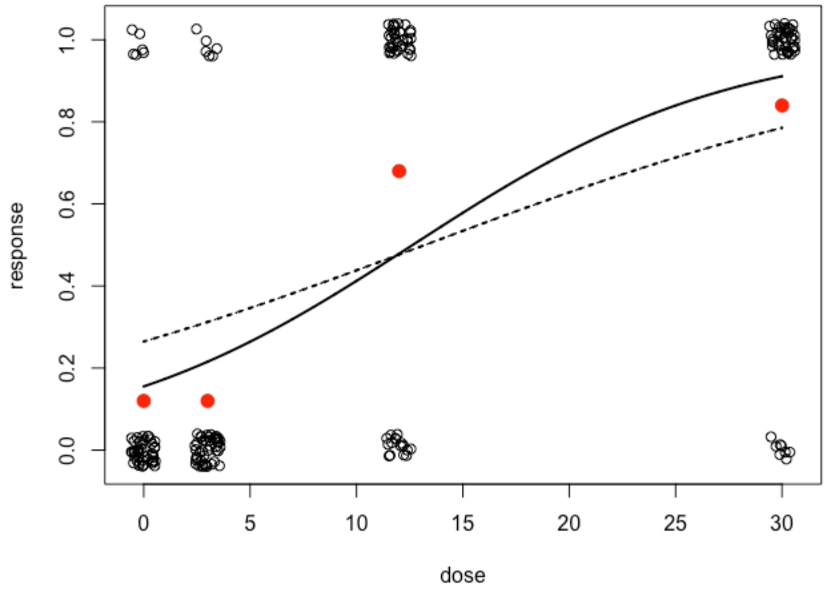



```
Coefficients:
     Estimate Std. Error
a -1.6950777 0.25327280
b  0.1340676 0.01914376


'log Lik.' -97.23388
```

THYROID EPITHELIAL CELL VACUOLISATION

- Probit model $y = \mathrm{CumNorm}(a + bx)$

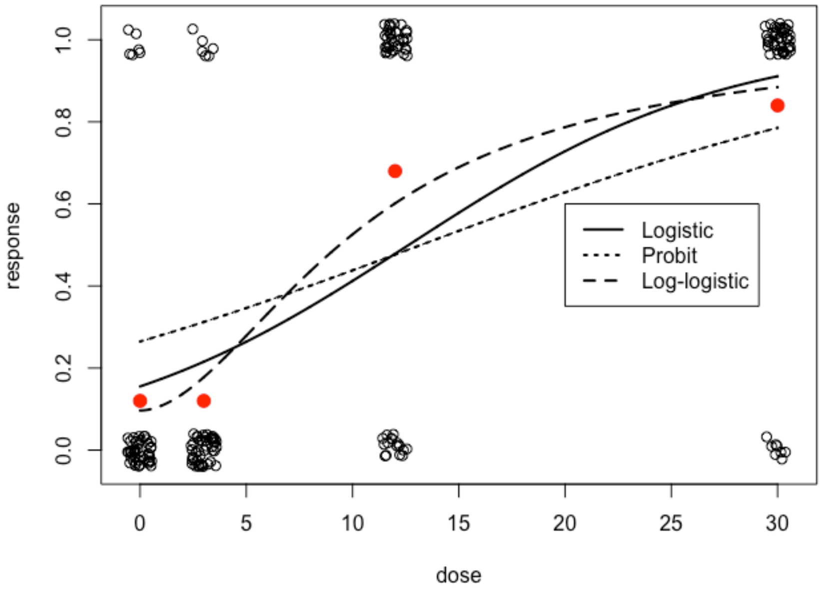



```
Coefficients:
    Estimate   Std. Error
a -1.02188175 0.145057689
b  0.07728999 0.009641039
```

'log Lik.' -97.5356

THYROID EPITHELIAL CELL VACUOLISATION

- Log-logistic model $y = a + (1-a)/(1 + \exp(-\log(x/b)/c))$



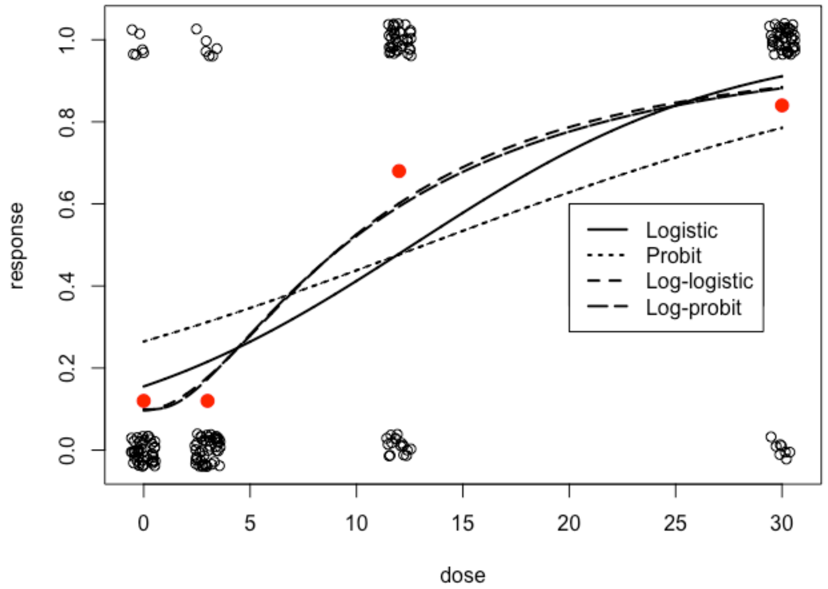
Coefficients:

	Estimate	Std. Error
a	0.0964394	0.03517616
b	10.5572758	1.52570441
c	0.5433158	0.10007674

'log Lik.' -91.90305

THYROID EPITHELIAL CELL VACUOLISATION

- Log-probit model $y = a + (1-a)\,\mathrm{CumNorm}(\log(x/b)/c)$

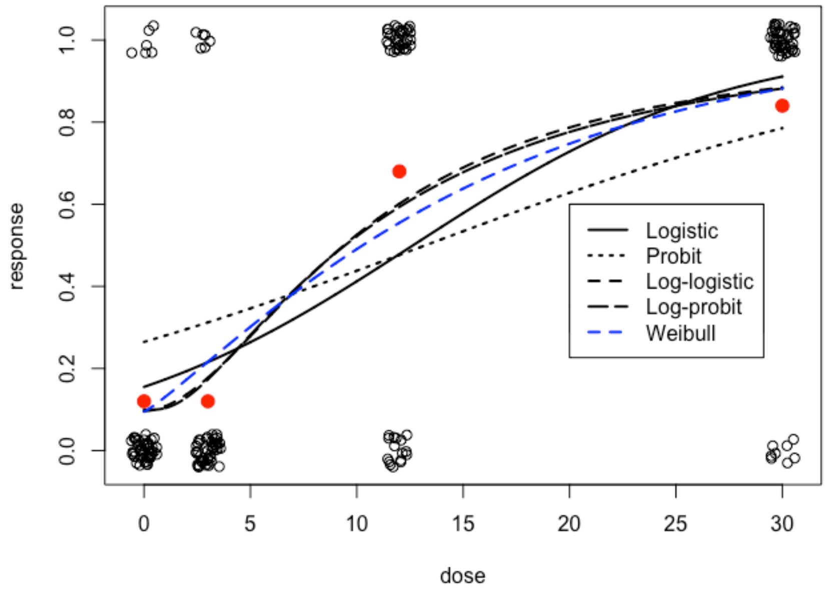



```
Coefficients:
    Estimate Std. Error
a  0.09922882 0.03617311
b 10.72949432 1.54589431
c  0.91708158 0.15473914

'log Lik.' -91.86382
```

THYROID EPITHELIAL CELL VACUOLISATION

- Weibull model $y = a + (1-a)\exp((x/b)^c)$



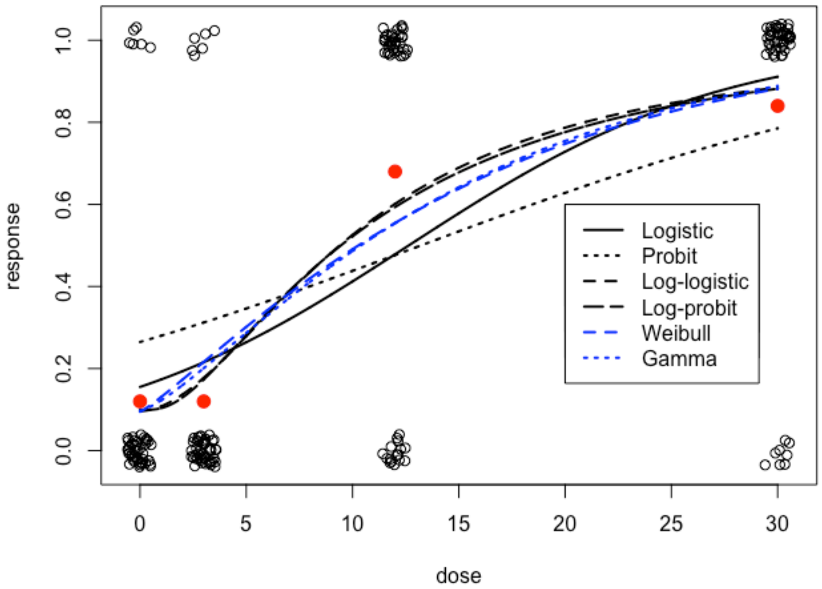

```
Coefficients:
   Estimate   Std. Error
a 0.09475043 0.03550996
b 0.04072331 0.02383889
c 1.15006741 0.19314240

'log Lik.' -93.77242
```

European Food Safety Authority

THYROID EPITHELIAL CELL VACUOLISATION

- Gamma model $y = a + (1-a)\,\mathrm{CumGam}(bx^c)$

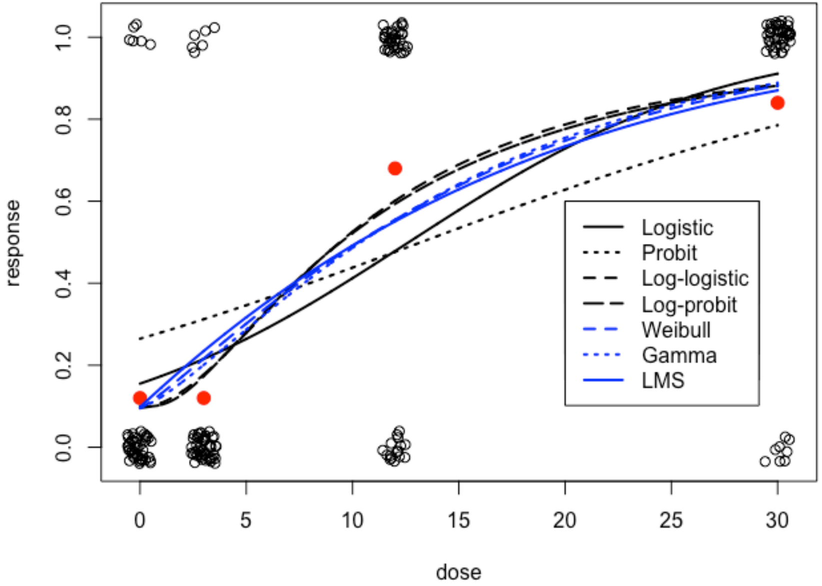



```
Coefficients:
     Estimate   Std. Error
a 0.09758502 0.03628932
b 0.09005540 0.02978308
c 1.38160881 0.40025088
```

```
'log Lik.' -93.49419
```

THYROID EPITHELIAL CELL VACUOLISATION

- LMS model $y = a + (1-a)(1-\exp(-bx-cx^2))$




```
Coefficients:
     Estimate     Std. Error
a 0.0977349769 0.0363010174
b 0.0537970136 0.0137209538
c 0.0003621077 0.0004136556

'log Lik.' -94.19809
```

THYROID EPITHELIAL CELL VACUOLISATION

Model	Log-likelihood	# par	AIC
Null	-137.186	1	276.372
Full	-90.020	4	188.039
Logistic	-97.234	2	198.468
Probit	-97.536	2	199.071
Log-logistic	-91.903	3	189.806
Log-probit	-91.864	3	189.728
Weibull	-93.772	3	193.545
Gamma	-93.494	3	192.988
LMS	-94.198	3	194.396

$AIC_{Min} = 189.728 \quad < \quad AIC_{FULL} + 2 = 190.039$

All $\quad AIC < AIC_{NULL} - 2 = 274.372$

THYROID EPITHELIAL CELL VACUOLISATION

Next presentation

Model Averaging and BMD estimation

STAY CONNECTED!

Subscribe to

www.efsa.europa.eu/en/news/newsletters

www.efsa.europa.eu/en/rss

Engage with careers

www.efsa.europa.eu/en/engage/careers

Follow us on Twitter

@efsa_eu

@plants_efsa

@methods_efsa