Summary

- ▼ Basic terminology
    - Definition of Toxicology
    - Doses and responses
    - Alert concertration
        - What is it
        - How is it characterized
    - ▼ Studysetting
        - In Vivo
        - In Virto
        - In Situ
        - In Silico
    - VPA (valporic acid example)
- ▼ Isotonic regression (counterpart of antitonic regression)
    - ▼ Problem outline
        - Assuming monotonicity
        - How the optimal solution is characterized

$$\sum_{x \in X} w(x)(g(x) - f(x))^2 \quad \forall f \quad \text{isotonic}$$

    - ▼ PAVA
        - ▼ procedure

          If for $j = i + 1 > i$ it holds, that $\mu(x_i) < \mu(x_j)$ then:

$$\mu^*(x_i, x_j) = \frac{n_i \mu(x_i) + n_j \mu(x_j)}{n_i + n_j}$$

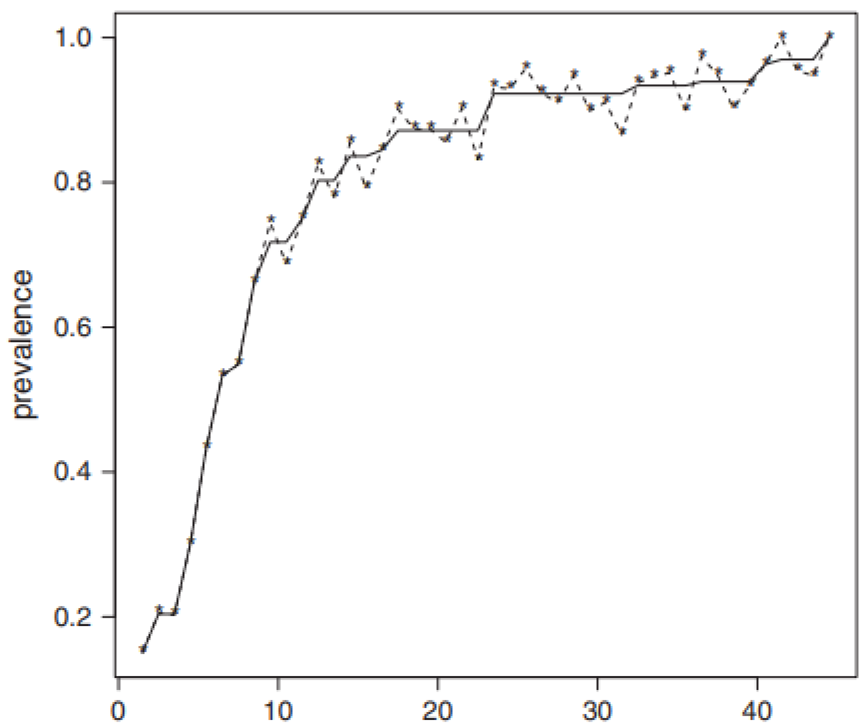
New estimates are replaced with $\mu^*(x_i, x_j)$

- example

  Rubella → Dissease, you can only get once in your life

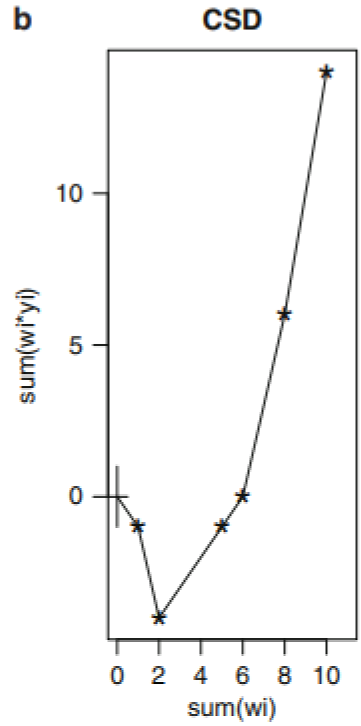
  $\Rightarrow$ So it must follow, that prevelance after age must be monotonically increasing after age

  applied results in:



- Cummulative sum diagramm (CSD)

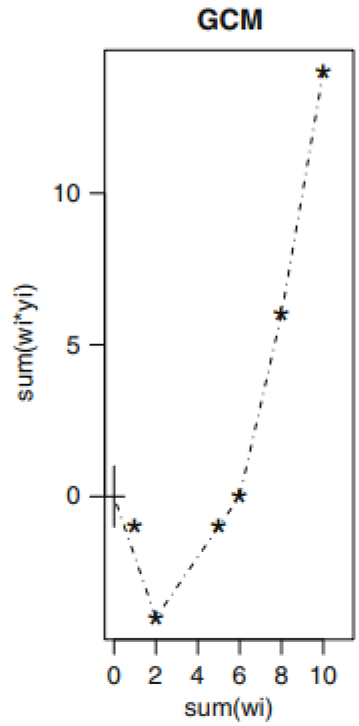
  Let $W_i := \sum_{j=1}^{i} w_j$ where w_i are the weights of the means. Further more let $P_i := \sum_{j=1}^{i} w_j \mu(x_i)$ (sum over all weighted estimates). Then we define the CSD as all (W_i, P_i) for all $i \in \{1, ..., n\}$



If can be easily shown, that the slope between two points of this diagramm is equal to the estimate of the mean $\mu(x_i)$.

- greatest convex minorant (GCM)

  Greatest convex function f, with: $f(W_i) \leq P_i \quad \forall i$



It holds that, estimates $\mu(x_i)$ get pooled iff $P_i > f(W_i)$

- Martix display

  You can show, that $\mu^* = S\mu$ with S beeing a block diagonal matrix.

  Example for a diagonal block martix:

$$\hat{\mu}^* = \boldsymbol{S}\hat{\mu} = \begin{pmatrix} 2/5 & 3/5 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 \\ 0 & 0 & 4/9 & 5/9 \\ 0 & 0 & 4/9 & 5/9 \end{pmatrix} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \end{pmatrix}$$

- Curve and model fitting
  - Sigmoid curves

    Mathematical setting:

- The isotonic regression assumed, that $x_i$is ordinally scalled. Now we are going to asssume that $x_i : \mathcal{A} \rightarrow \mathbb{R} \quad \forall i$ (so x_i is a random variable, that maps into the real numbers)

Why should you do this fit?

- Applying sigmoid curves is the first instance where we are trying to characterize the **alert-concentration.** In this setting the alert-concentration is defined by the mean of the upper and lower asymptote.

Things to consider when looking at research containing sigmoid fits:

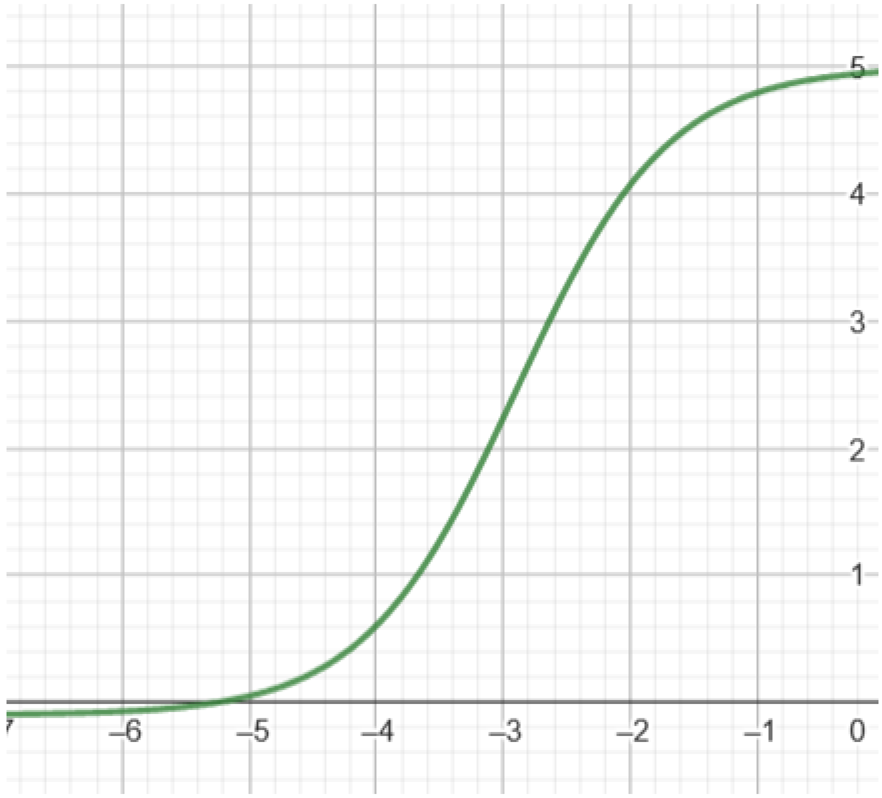
- What parametrization was applied?
    - The interpretation depends on what formular was used.
    - In research, many different names refer to the same model (4pLL) with a different parametrization

▼ 4pLL (4 parameter log logistic regression)

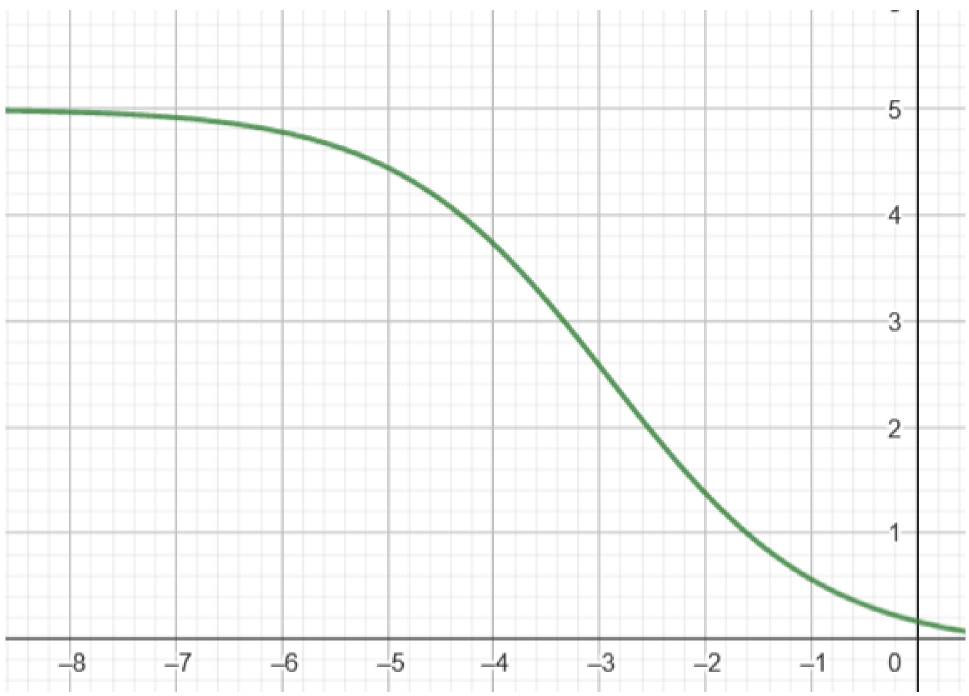
We assume, the relation between the responses Y_{ij} and the dosages x_i can be described by:

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{1 + e^{(\theta_2 - x_i)/\theta_3}} + \epsilon_{ij}$$

- x_i **log scaled**
- θ_1, θ_4: lower and upper asymptote
- θ_3: slope of the function
    - **if the sign changes the roles of upper and lower asymptote reverse!!**
    - for $\theta_3 > 0$:



- for $\theta_3 < 0$:



- θ_2: is the dosis at which the effect reaches $\frac{|\theta_1 - \theta_4|}{2}$

You can reduce the model by assuming fixed values for the different parameters

- For instance, if the response is the relative amount of living cells, we can safely assume, that the upper asymptote is 1. Thus the regression in use is only a 3pLL.

Fit is done nummerical but not further discussed in the lecture.

▼ (sigmoidal) Emax Model

Another way of parametrization for the 4pLL is:

$$Y_{ij} = E_0 + \frac{x_i^n E_{max}}{x_i^n + ED50^n} + \epsilon_{ij}$$

- E_0: the baseline effect of the compound
- E_{max}: The maximum difference between the response and E_0 ($\theta_4 - \theta_1$)
- n: the slope parameter (can also be equal to $-\theta_3$)
- $ED50$: just like θ_2 in 4pLL

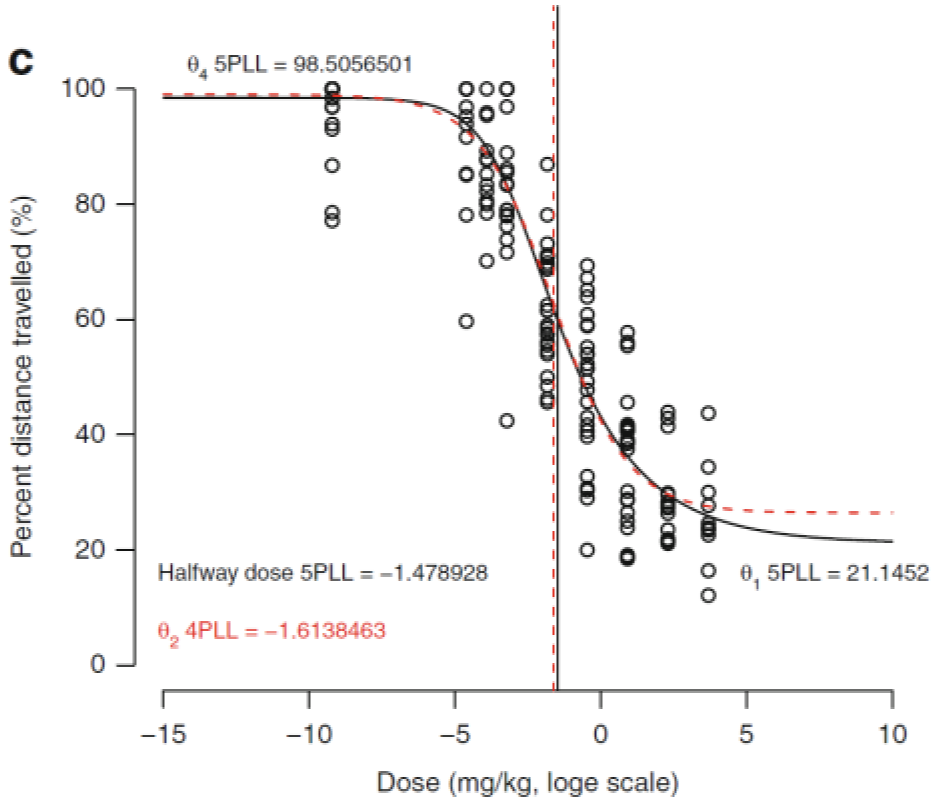
This model is a bit better to interpret

▼ 5pLL

New formular:

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{(1 + e^{(\theta_2 - x_i)\theta_3})^{\theta_5}} + \epsilon_{ij}$$

- With this model you sacrafice interpretability for more accuracy.
- You gain more controll over the tails of the sigmoid:

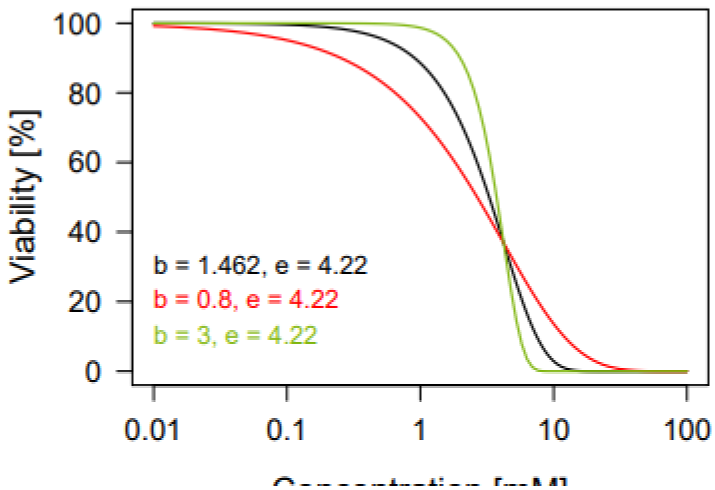


▼ Other

Another way of generalizing the 4pLL is to define a function f such that f : $\mathbb{R} \rightarrow \mathbb{R}$. Now the 4pLL can be written as:

$$f(x, a, b, c, d) = c + (d - c)f(b(log(x) - log(e))$$

- For $f = \phi$ we get the **log-normal regression**
- For $f = \exp(-\exp(x))$ we get a **weibull type 1 regression**
    - This regression has a slow descent at upper asymptote and a faster descent at the lower asymptote:



- for $f = 1 - \exp(\exp(x))$ we get a **weibull type 2 regression**
  - This regression has the oppsosite asymmetry compared to weibull type 1

▼ Application

Things to consider, when applying the model:

- **machine error:**
  - Because the fit is done numerically, we need to consider the machine epsilon and errors when calculating with floating point numbers.
  - **avoid very small values combined with very big values**
    - Possible solution: estimate $-\frac{1}{\theta_3}$ for the slope instead of θ_3

Applying the regression in R

- use the **gnls** function from the **nlme-package**

```
gnls.model001 <- gnls(ratio~(th1+(th4-th1)/(1+(exp((lmpk-th2)*th3)))),
                      data=data2b, params=list(th1+th2+th3+th4~1),
                      start=c(90,-0.2,1,28), control=gnlsControl(nlsTol=0.1))
```