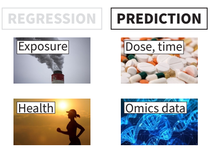
# Detecting structural variations between different concentration-response experiments

**Julia Eichhorn**, Prof. Dr. Kirsten Schorning

TU Dortmund University

REGRESSION PREDICTION




INTEGRATION

**Project P5**

## Structural variations between cytotoxicity experiments

- Conduct same **concentration-response** experiment on different days [1]
- Fit for each experimental day a **non-linear regression model**, e.g.

$$\eta(x, \theta) = \theta_0 + \frac{\theta_1 x^h}{\theta_2^h + x^h} \text{ [2]}$$

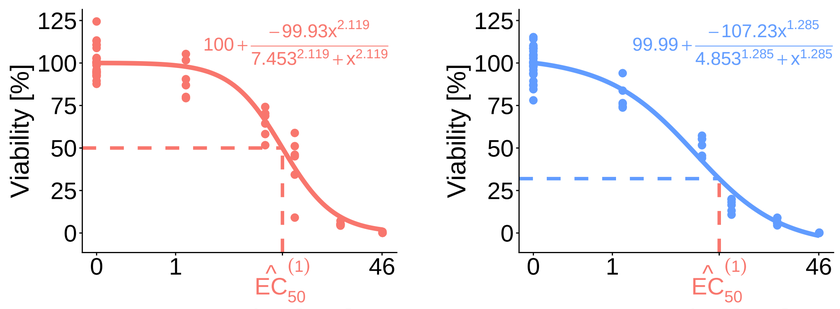



Fig. 1: Same cytotoxicity experiment performed on two experimental days indicating different response values regarding alert concentration $\widehat{\mathrm{EC}}_{50}^{(1)}$ (concentration corresponding to 50% reduction of upper 100% viability)

## Non-linear regression model for two experimental days

1. Use non-linear regression to model cytotoxicity data for two experimental days

$$Y_{ij}^{(k)} = \eta(x_i^{(k)}, \theta^{(k)}) + \varepsilon_{ij}^{(k)},\ i = 1, \dots, l^{(k)}, j = 1, \dots n_i^{(k)}, k = 1, 2$$

- $x_i^{(k)} \in \mathcal{X} = [0, x_{\max}] \subset \mathbb{R}$ design space
- $\eta$ known regression function
- $\theta^{(k)} \in \mathbb{R}^q$ unknown parameter vector
- $\varepsilon_{ij}^{(k)} \sim N(0, \sigma^2)$ independent, overall same variance
- $n^{(k)} = \sum_{i=1}^{l^{(k)}} n_i^{(k)}$ measurements per day, $n = n^{(1)} + n^{(2)}$ total

2. Derive asymptotic distribution of (joint) maximum likelihood estimator (MLE):
   - Individual asymptotics: $\sqrt{n^{(k)}}\left(\hat{\theta}^{(k)} - \theta^{(k)}\right) \xrightarrow{D} N_q\left(0, M^{-1}(\xi^{(k)}, \theta^{(k)})\right), k = 1, 2$

$$\xi^{(k)} = \begin{pmatrix} x_1^{(k)} & \dots & x_{l^{(k)}}^{(k)} \\ \omega_1^{(k)} & \dots & \omega_{l^{(k)}}^{(k)} \end{pmatrix}, M(\xi^{(k)}, \theta^{(k)}) = \frac{1}{\sigma^2} \sum_{i=1}^{l^{(k)}} \frac{1}{\omega_i^{(k)}} \left(\frac{\partial \eta\left(\xi^{(k)}, \theta^{(k)}\right)}{\partial \theta^{(k)}}\right)\left(\frac{\partial \eta\left(\xi^{(k)}, \theta^{(k)}\right)}{\partial \theta^{(k)}}\right)^\top$$

   - For MLE $\hat{\theta} = (\hat{\theta}^{(1)}, \hat{\theta}^{(2)})^T$ of $\theta = (\theta^{(1)}, \theta^{(2)})^T$, $\lambda^{(k)} = \lim_{n\to\infty} \frac{n^{(k)}}{n}$, $\xi = (\xi^{(1)}, \xi^{(2)})^\top$:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_{2q}(0, M^{-1}(\xi, \theta)) \text{ with}$$

$$M^{-1}(\xi, \theta) = \begin{pmatrix} \frac{1}{\lambda^{(1)}} M^{-1}\left(\xi^{(1)}, \theta^{(1)}\right) & 0 \\ 0 & \frac{1}{\lambda^{(2)}} M^{-1}\left(\xi^{(2)}, \theta^{(2)}\right) \end{pmatrix}$$

## Nested function to detect structural variations

1. Idea: Determine the alert concentration of one experimental day using the inverse regression function $\widehat{\mathrm{EC}}_{50}^{(1)} = \eta^{-1}\left(50, \hat{\theta}^{(1)}\right)$. Evaluate the viability of the second experimental day at this concentration with the regression function $\eta$ (see Figure 1)
2. Derive the **asymptotic distribution** of $f\left(\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\right) = \eta\left(\eta^{-1}\left(50, \hat{\theta}^{(1)}\right), \hat{\theta}^{(2)}\right)$ with $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}$ MLE as plug-in estimators and with Δ- method [3]

$$\sqrt{n}\left(f(\hat{\theta}) - f(\theta)\right) \xrightarrow{D} N\left(0, \left(\frac{\partial}{\partial \theta} f(\theta)\right)^\top M^{-1}(\xi, \theta)\left(\frac{\partial}{\partial \theta} f(\theta)\right)\right) \quad (1)$$

3. Construct **testing procedure** to decide for systematic or non-systematic difference:
   - $H_0 : \theta \in \Theta_0, \Theta_0 = \{(\theta^{(1)}, \theta^{(2)})^\top \in \mathbb{R}^{2q}, f(\theta^{(1)}, \theta^{(2)}) = 50\}$
   - $H_1 : \theta \in \Theta_1, \Theta_1 = \{(\theta^{(1)}, \theta^{(2)})^\top \in \mathbb{R}^{2q}, f(\theta^{(1)}, \theta^{(2)}) \neq 50\}$
   - Use (1) to construct asymptotic confidence interval (CI):

$$f(\hat{\theta}) \pm \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\left(\frac{\partial}{\partial \theta} f(\hat{\theta})\right)^\top M^{-1}(\xi, \hat{\theta})\left(\frac{\partial}{\partial \theta} f(\hat{\theta})\right)}$$

   - Reject $H_0$ if true value $f(\theta) = 50$ is not contained in CI

## Illustration of testing procedure

Simulation Setup

- $N_{Sim} = 10000, \varepsilon \sim N(0, 14.86897^2)$
- The same 6 concentrations are used for both experimental days with $n_i^{(1)} = n_i^{(2)} = 30$, $i = 1, \dots, 6$: $\{0.000, 1.229, 5.432, 9.334, 21.54, 46.4\}$
- Under $H_0$ : $\theta^{(1)} = \theta^{(2)} = (0, 100, 6.5, 1)^\top$

→ simulation shows rejection rate of 5.39%, significance level of testing procedure reached

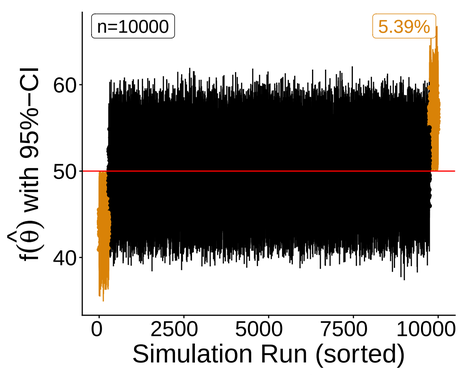



Fig. 2: Color-coded CIs of $f(\hat{\theta})$ under $H_0$

## Using testing procedure with different designs

| Design | Concentrations | | | | | |
|---|---|---|---|---|---|---|
| Bayesian[a] | 0.000 | 1.2290 | 5.432 | 9.334 | 21.54 | 46.4 |
| IfADo | 0.000 | 0.0100 | 0.100 | 1.000 | 10.00 | 46.4 |
| log_eq | 0.000 | 0.0464 | 0.464 | 2.154 | 4.64 | 46.4 |
| $\omega_i$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

Table 1: Overview of considered approximate designs, see [1]

Simulation Setup

- $N_{Sim} = 10000, \varepsilon \sim N(0, 14.86897^2)$
- Same designs for both days, see Table 1
- Increasing $n_i^{(1)} = n_i^{(2)} \in \{3, 6, \dots, 30, 33\}$
- Used $\theta^{(1)} = \theta^{(2)} = (0, 100, 6.5, 1)^\top (H_0)$
  $\theta^{(2)} = (0, 100, 12.5, 1)^\top (H_1)$

Results

- Under $H_0$: for Bayesian and IfADo overall significance level reached, log-eq design shows higher values than expected 5%
- Under $H_1$: Increasing power with increasing sample size, Bayesian design overall has highest power even for e.g. $n_i^{(1)} = n_i^{(2)} = 6$

[a] based on D-optimality criterion, see [1]

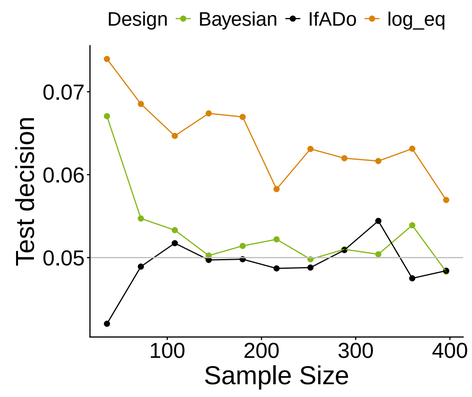



Fig. 3: Comparison of the number of rejected test statistics under $H_0$ for increasing sample sizes for different designs

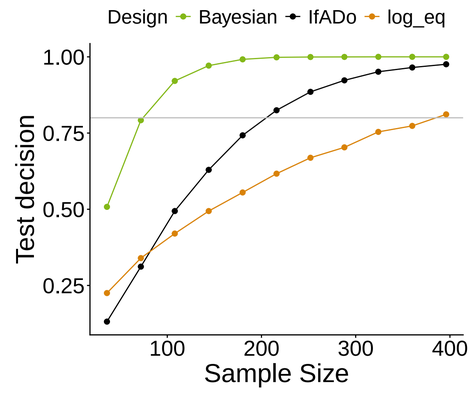



Fig. 4: Comparison of the number of rejected test statistics under $H_1$ for increasing sample sizes for different designs

## Suitable optimality criterion

- Find optimal design $\xi^*$ which maximises the power of testing procedure
- Power is proportional to $p(\xi, \theta) = \left(\left(\frac{\partial}{\partial \theta} f(\theta)\right)^\top M^{-1}(\xi, \theta)\left(\frac{\partial}{\partial \theta} f(\theta)\right)\right)^{-1}$ for $\theta \in \Theta_1$
- Optimal design: $\xi^*$ is called optimal design for detecting structural variations :⇔

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmax}} \min_{\theta \in \widetilde{\Theta}_1} p(\xi, \theta), \text{ where } \widetilde{\Theta}_1 \subset \Theta_1$$

- For fixed $\theta^{(1)}, \theta^{(2)} \in \mathbb{R}^q$ with $\theta^{(1)} \neq \theta^{(2)}$ set $\widetilde{\Theta}_1 = \{(\theta^{(1)}, \theta^{(2)})\}$, the equivalence theorem holds: $\xi^* = (\xi^{(1)*}, \xi^{(2)*})^\top$ is optimal for detecting structural variations in $\widetilde{\Theta}_1$ :⇔

$$\left(\sum_{k=1}^{2} \left(\frac{\partial}{\partial \theta^{(k)}} f(\theta)\right)^\top \frac{1}{\lambda^{(k)}} M^{-1}(\xi^{(k)*}, \theta^{(k)})\left(\frac{\partial}{\partial \theta} \eta(x^{(k)}, \theta^{(k)})\right)\right)^2 \leq p^{-1}(\xi^*, \theta)$$

for all $x = (x^{(1)}, x^{(2)})^\top \in \mathcal{X}^2$

## Criterion values in simulation setup

Calculate $p(\xi, \theta)$ for all designs of Table 1:

Let $\widetilde{\Theta}_1 = \{(\theta^{(1)}, \theta^{(2)})\}$ with $\theta^{(1)} = (0, 100, 6.5, 1)^\top, \theta^{(2)} = (0, 100, 12.5, 1)^\top$:

$p(\xi_{\text{Bayesian}}, \theta) = 103.97 \cdot 10^{-3}$ $p(\xi_{\text{IfADo}}, \theta) = 46.12 \cdot 10^{-3}$ $p(\xi_{\text{log\_eq}}, \theta) = 24.47 \cdot 10^{-3}$

## Conclusion and Outlook

- Parametric testing procedure for structural variations works fine
  → Extension to other alert concentrations and multiple days possible
- Choice of design impacts the power of the testing procedure
  → Next step: Calculate optimal design
- Equivalence theorem can be used to verify optimality
  → Next step: Extension to more complex $\widetilde{\Theta}_1$ and multiple days

## References

[1] L. Schürmeyer, C. Peng, W. Albrecht, T. Brecklinghaus, P. Baur, J. G. Hengstler, and K. Schorning. "Design of optimal concentrations for in vitro cytotoxicity experiments". In: *Archives of Toxicology* (2024)

[2] J. Macdougall. "Analysis of Dose–Response Studies—Emax Model". In: *Dose Finding in Drug Development*. Ed. by N. Ting. Springer New York, 2006, pp. 127–145

[3] A. W. v. d. Vaart. "Delta Method". In: *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998

## Acknowledgments

This work was funded by the German Research Foundation (DFG) as part of RTG 2624 *Biostatistical Methods for High-Dimensional Data in Toxicology* (427806116).

technische universität dortmund