OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

Rainer Schwabe

“To Estimate or to Predict”

Implications on the Design for Linear Mixed Models

“To Estimate or to Predict”
-
Implications on the Design for Linear Mixed Models

Rainer Schwabe, Maryna Prus

rainer.schwabe@ovgu.de

Bundesministerium für Bildung und Forschung

supported by SKAVOE
German Federal Ministry of Education and Research

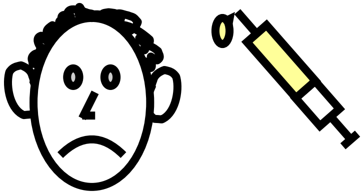
Outline

Prologue: Short introduction

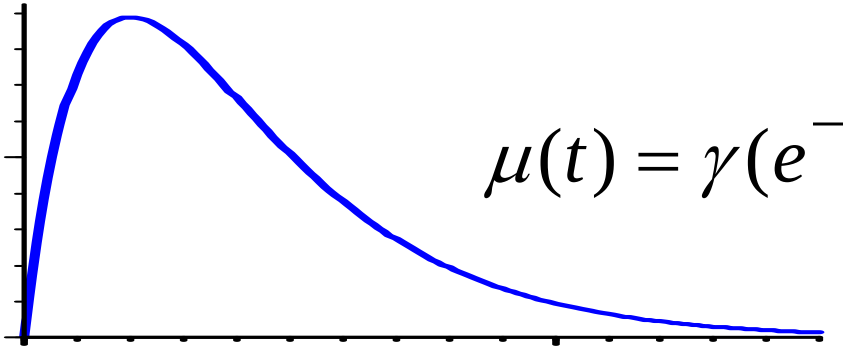
1. Model Description
2. Estimation and “Prediction”
3. Optimal Design
4. A Simple Example
5. Outlook

Prologue: Short introduction

- Example: Pharmacokinetics

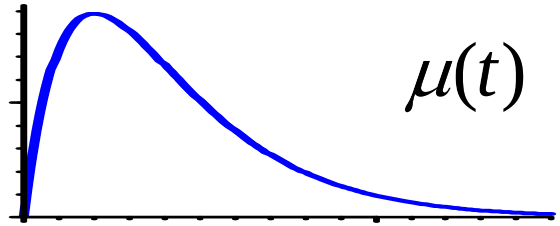


- measure the concentration of a drug in s.o.'s blood over time

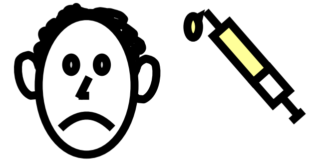


$$\mu(t) = \gamma(e^{-\alpha t} - e^{-\beta t})$$

Example: Pharmacokinetics



$\mu(t)$



- observations

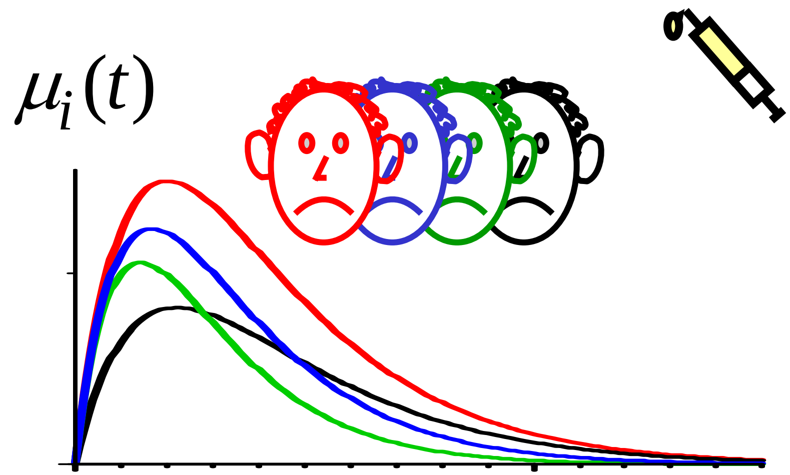
$$Y_j(t_j) = \mu(t_j) + \varepsilon_j$$

- estimate response curve $\mu(t)$,
  AUC (area under the curve),
  $c_{\max}$ (maximal concentration) etc.

---

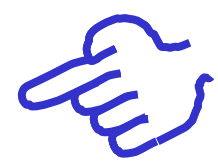
optimal time points of measurements $t_1, \ldots, t_n$?

Random coefficients

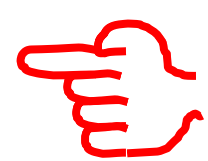


- each individual
  has its **own** curve

- population parameters
  $\Rightarrow$ “typical” curve



- individual parameters
  $\Rightarrow$ “individual” curves



1. Model Description

- linear mean response $\mu(t) = \mathbf{f}(t)^{\mathsf{T}}\boldsymbol{\beta}$

$$Y_j(t_j) = \mathbf{f}(t_j)^{\mathsf{T}}\boldsymbol{\beta} + \varepsilon_j$$

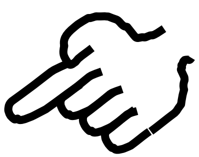
random error

observation j=1,...,n

explanatory variable

ε_j i.i.d.
$\mathrm{Var}(\varepsilon_j) = \sigma^2$

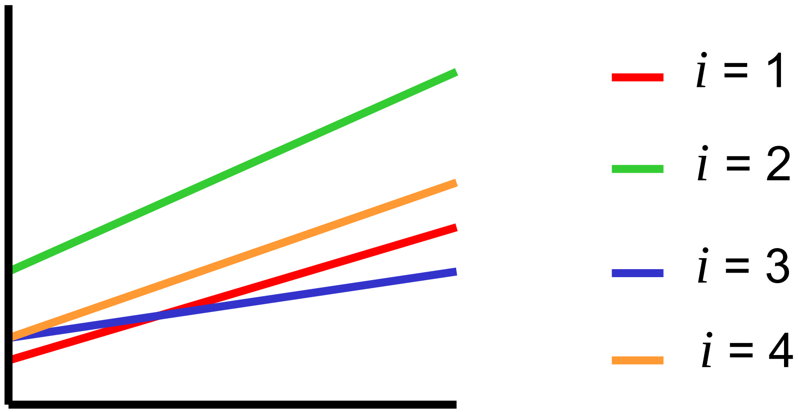
- regression functions $\mathbf{f} = (f_1, \ldots, f_p)$
- parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$



But !

- each individual has its **own** response curve

octodon degus



- individual responses follow a common model

Hierarchical model

- individual level

individual "parameter"

$$Y_{ij} = \mathbf{f}(t_{ij})^{\mathsf{T}} \boldsymbol{\beta}_i + \varepsilon_{ij}$$

individual
i=1,…,n

replication
j=1,…,m_i

explanatory variable

error

$$\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$

- population level

$$\boldsymbol{\beta}_i \sim \mathrm{N}_p\left(\boldsymbol{\beta}, \sigma^2 \mathbf{D}\right)$$

independent

population parameter

Individual observational vector

$$\mathbf{Y}_i = \mathbf{F}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i = \mathbf{F}_i \boldsymbol{\beta} + \mathbf{F}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

individual design matrix

$$\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{im_i} \end{pmatrix} \qquad \mathbf{F}_i = \begin{pmatrix} \mathbf{f}(t_{i1})^{\mathsf{T}} \\ \vdots \\ \mathbf{f}(t_{im_i})^{\mathsf{T}} \end{pmatrix} \qquad \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{im_i} \end{pmatrix}$$

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \sigma^2 \mathbf{I}_{m_i}$$

- individual effect

$$\mathbf{b}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}, \quad \mathrm{E}(\mathbf{b}_i) = \mathbf{0}, \quad \mathrm{Cov}(\mathbf{b}_i) = \sigma^2 \mathbf{D}$$

Individual covariance structure

$$\mathbf{Y}_i = \mathbf{F}_i \boldsymbol{\beta} + \mathbf{F}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

- individual covariance matrix

$$\mathrm{Cov}(\mathbf{Y}_i) = \sigma^2 \left(\mathbf{F}_i \mathbf{D} \mathbf{F}_i^\top + \mathbf{I}_{m_i} \right)$$

observations are **correlated**

- e.g. random intercept

equal correlations

$$\mathrm{Cov}(\mathbf{Y}_i) = \sigma^2 \left(\tau_\alpha \mathbf{1}_{m_i} \mathbf{1}_{m_i}^\top + \mathbf{I}_{m_i} \right)$$

Single group designs

- all individuals
  at the same experimental settings

$$m_i \equiv m$$
$$t_{ij} \equiv t_j$$

$$\mathbf{F}_i \equiv \mathbf{F}$$
$$\mathrm{Cov}(\mathbf{Y}_i) \equiv \sigma^2 \underbrace{\left(\mathbf{F}\,\mathbf{D}\,\mathbf{F}^\mathsf{T} + \mathbf{I}_m\right)}_{=\mathbf{V}}$$

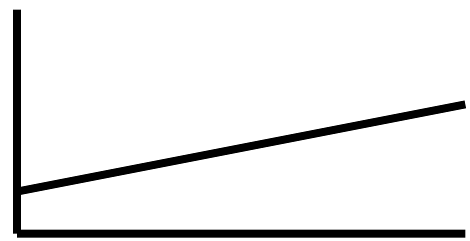
- full observational vector

$$\mathbf{Y} = \left(\mathbf{1}_n \otimes \mathbf{F}\right)\boldsymbol{\beta} + \left(\mathbf{I}_n \otimes \mathbf{F}\right)\mathbf{b} + \boldsymbol{\varepsilon}$$

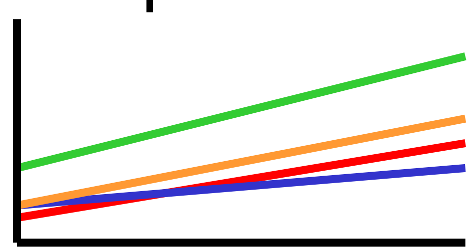
$$\mathrm{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_n \otimes \mathbf{V}$$

2. Estimation and Prediction

- estimation of population parameter $\boldsymbol{\beta}$
  and population response function $\mu(t)$



- prediction of individual effects $\mathbf{b}_i$,
  individual parameters $\boldsymbol{\beta}_i$
  and individual response functions $\mu_i(t)$



Estimation

- BLUE: best linear unbiased estimator of population parameter $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^{\top}\mathbf{F})^{-1}\mathbf{F}^{\top}\overline{\mathbf{Y}}$$

$$\overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_i$$

does not require the knowledge of $\mathbf{D}$ (WLS=OLS)

- BLUE: best linear unbiased estimator of population response function $\mu(t)$

$$\hat{\mu}(t) = \mathbf{f}(t)^{\top}\hat{\boldsymbol{\beta}}$$

Covariance of the estimators

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)$$

$$\mathrm{Var}(\hat{\mu}(t)) = \frac{\sigma^2}{n}\left(\mathbf{f}(t)^\top\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\mathbf{f}(t) + \mathbf{f}(t)^\top\mathbf{D}\,\mathbf{f}(t)\right)$$

- covariance decomposes additively into
  - covariance $\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}$ in the model without random effects $(\mathbf{D} = \mathbf{0})$
  - dispersion $\mathbf{D}$ of the individual parameter

Prediction (regular $\mathbf{D}$)

- BLUP: best linear unbiased predictor of individual effect $\mathbf{b}_i$

$$\hat{\mathbf{b}}_i = \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \mathbf{F}^\top \left(\mathbf{Y}_i - \overline{\mathbf{Y}}\right)$$

Henderson (1959)

- predictor of individual parameter $\boldsymbol{\beta}_i$

$$\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i = \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \left(\mathbf{F}^\top \mathbf{F} \hat{\boldsymbol{\beta}}_{i,\mathrm{OLS}} + \mathbf{D}^{-1} \hat{\boldsymbol{\beta}}\right)$$

- predictor of individual response function $\mu_i(t)$

$$\hat{\mu}_i(t) = \mathbf{f}(t)^\top \hat{\boldsymbol{\beta}}_i$$

MSE of the predictors (regular $\mathbf{D}$)

$$\operatorname{Cov}(\hat{\boldsymbol\beta}_i - \boldsymbol\beta_i)_{i=1,\ldots,n}$$
$$= \sigma^2\left(\left(\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\otimes\left(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1}\right)^{-1} + \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\otimes\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\right)$$

$$\operatorname{Var}(\hat\mu_i(t) - \mu_i(t))$$
$$= \frac{\sigma^2}{n}\mathbf{f}(t)^\top\left((n-1)\left(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1}\right)^{-1} + \left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\right)\mathbf{f}(t)$$

- MSE for prediction: weighted average of
  - » Bayesian covariance $\left(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1}\right)^{-1}$
  - » covariance $\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}$

    in the model without random effects

Prediction (regular $\mathbf{D}$, known $\boldsymbol{\beta}$)

- predictor of individual parameter $\boldsymbol{\beta}_i$

$$\hat{\boldsymbol{\beta}}_i = \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \mathbf{F}^\top \mathbf{Y}_i$$

- MSE of the predictor

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)_{i=1,\ldots,n} = \sigma^2 \mathbf{I}_n \otimes \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1}$$

- MSE = Bayesian covariance $\left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1}$

Gladitz, Pilz (1982)

Prediction (general $\mathbf{D}$)

- BLUP: best linear unbiased predictor of individual effect $\mathbf{b}_i$

$$\hat{\mathbf{b}}_i = \left(\mathbf{D} - \mathbf{D}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\mathbf{D}\right)\mathbf{F}^\top\left(\mathbf{Y}_i - \overline{\mathbf{Y}}\right)$$

- predictor of individual parameter $\boldsymbol{\beta}_i$

$$\hat{\boldsymbol{\beta}}_i = \mathbf{D}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\hat{\boldsymbol{\beta}}_{i,\mathrm{OLS}} + \left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\hat{\boldsymbol{\beta}}$$

- predictor of individual response function $\mu_i(t)$

$$\hat{\mu}_i(t) = \mathbf{f}(t)^\top\hat{\boldsymbol{\beta}}_i$$

MSE of the predictors (general $\mathbf{D}$)

$$\mathrm{Cov}(\hat{\boldsymbol\beta}_i - \boldsymbol\beta_i)_{i=1,\dots,n}$$
$$= \sigma^2\left(\left(\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\otimes\left(\mathbf{D} - \mathbf{D}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\mathbf{D}\right) + \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\right)$$

$$\mathrm{Var}\,(\hat\mu_i(t) - \mu_i(t))$$
$$= \sigma^2\mathbf{f}(t)^\top\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} - \tfrac{n-1}{n}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\right)\right)\mathbf{f}(t)$$

- MSE for prediction: weighted average of
  - covariance $\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}$ in the model without random effects
  - and $\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} - \left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1}\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}$

3. Optimal Design

- "exact design" $(t_1,\ldots,t_n)$

- aim: choose $t_1,\ldots,t_n$ from design region T

  to minimise $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$ or $\mathrm{Var}(\hat{\mu}(t))$ — estimation

  resp. to minimise $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$

  or $\mathrm{Var}(\hat{\mu}_i(t) - \mu_i(t))$ — prediction

IMSE criterion

- minimise the
  Integrated **M**ean **S**quared **E**rror
  for the response function

$$\int_T \mathrm{E}\left((\hat{\mu}(t)-\mu(t))^2\right)\mathrm{d}\,t$$

$$= \frac{\sigma^2}{n}\,\mathrm{trace}\left(\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}+\mathbf{D}\right)\int_T \mathbf{f}(t)\mathbf{f}(t)^\top\,\mathrm{d}\,t\right)$$

estimation

$$\int_T \mathrm{E}\left(\sum_{i=1}^n(\hat{\mu}_i(t)-\mu_i(t))^2\right)\mathrm{d}\,t$$

$$=\sigma^2\,\mathrm{trace}\left(\left((n-1)\left(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1}\right)^{-1}+\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\right)\int_T \mathbf{f}(t)\mathbf{f}(t)^\top\,\mathrm{d}\,t\right)$$

prediction

IMSE criterion for estimation

- linear

$$\operatorname{trace}\left(\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}+\mathbf{D}\right)\int_T \mathbf{f}(t)\mathbf{f}(t)^\top \,\mathrm{d}\,t\right)$$

constant !

$$=\underbrace{\operatorname{trace}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\int_T \mathbf{f}(t)\mathbf{f}(t)^\top \,\mathrm{d}\,t\right)}_{IMSE\text{ without random effects}}+\overbrace{\operatorname{trace}\left(\mathbf{D}\int_T \mathbf{f}(t)\mathbf{f}(t)^\top \,\mathrm{d}\,t\right)}^{\text{constant !}}$$

IMSE without random effects

- optimal designs for estimation
  in the model without random effects
  remain optimal for estimation
  in the model with random effects

IMSE criterion for prediction

- linear

$$\operatorname{trace}\left(\left((n-1)\left(\mathbf{F}^{\top}\mathbf{F}+\mathbf{D}^{-1}\right)^{-1}+\left(\mathbf{F}^{\top}\mathbf{F}\right)^{-1}\right)\int_{T}\mathbf{f}(t)\mathbf{f}(t)^{\top}\,\mathrm{d}\,t\right)$$

Bayesian *IMSE*

$$=(n-1)\operatorname{trace}\left(\left(\mathbf{F}^{\top}\mathbf{F}+\mathbf{D}^{-1}\right)^{-1}\int_{T}\mathbf{f}(t)\mathbf{f}(t)^{\top}\,\mathrm{d}\,t\right)$$

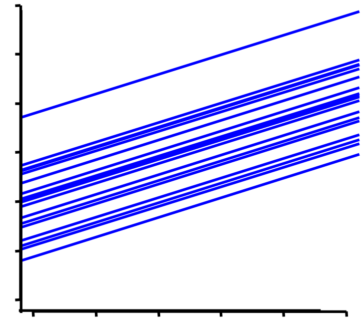
$$+\operatorname{trace}\left(\left(\mathbf{F}^{\top}\mathbf{F}\right)^{-1}\int_{T}\mathbf{f}(t)\mathbf{f}(t)^{\top}\,\mathrm{d}\,t\right)$$

IMSE without random effects

- *IMSE* for prediction: weighted average of
  - Bayesian *IMSE* criterion
  - *IMSE* criterionin the model without random effects

Random intercept (random blocks)

- "parallel" individual curves



$$Y_{ij} = \alpha_i + \mathbf{f}(t_j)^\top \boldsymbol{\beta} + \varepsilon_{ij}$$

- *IMSE* for prediction

constant

IMSE without random effects

$$(n-1)\frac{\tau_\alpha}{m\tau_\alpha + 1}|T| + \operatorname{trace}\left(\left(\mathbf{F}^\top\mathbf{F}\right)^{-1}\int_T \mathbf{f}(t)\mathbf{f}(t)^\top \mathrm{d}t\right)$$

- optimal designs for estimation
  in the model without random effects
  remain optimal for prediction
  in the model with random effects

4. A Simple example

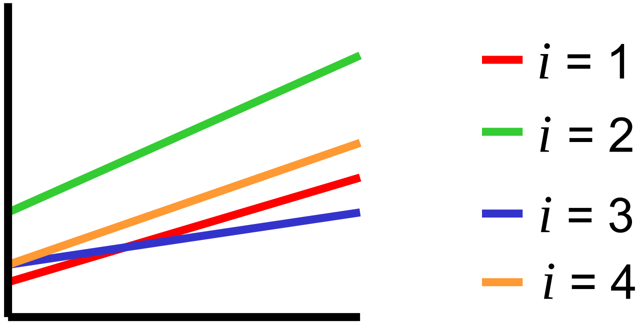
- individual response: simple linear regression on $T=[0,1]$

$$Y_{ij} = \alpha_i + \beta_i t_j + \varepsilon_{ij} \qquad 0 \le t_j \le 1$$

- individual parameter

$$\alpha_i \sim \mathrm{N}(\alpha, \sigma^2 \tau_\alpha), \;\; \beta_i \sim \mathrm{N}(\beta, \sigma^2 \tau_\beta) \quad \text{independent}$$

- individual curves



IMSE optimal designs

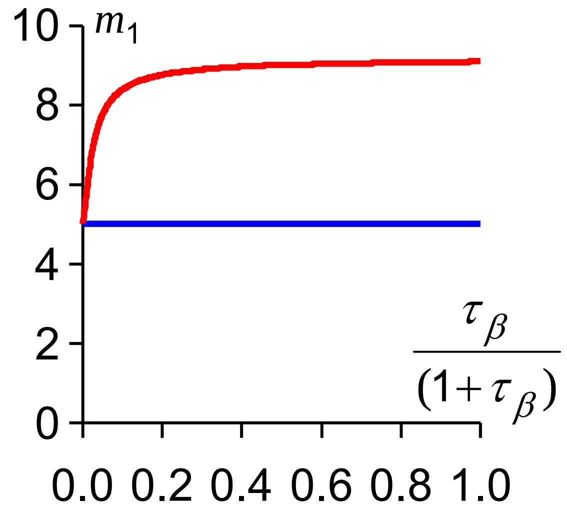
- optimal settings at $t = 0$ or $t = 1$
  - » m_1 observations at $t = 1$

$$\mathbf{F}^{\mathsf{T}}\mathbf{F} = \begin{pmatrix} m & m_1 \\ m_1 & m_1 \end{pmatrix}$$

- *IMSE* optimal proportions

τ_α small	m_1 observations at $t = 1$ τ_β small	τ_β large
estimation	$m/2$	$m/2$
prediction	$m/2$	$\left\lfloor \frac{n-\sqrt{n}}{n-1} \right\rfloor m$

$n = 100$, $m = 10$

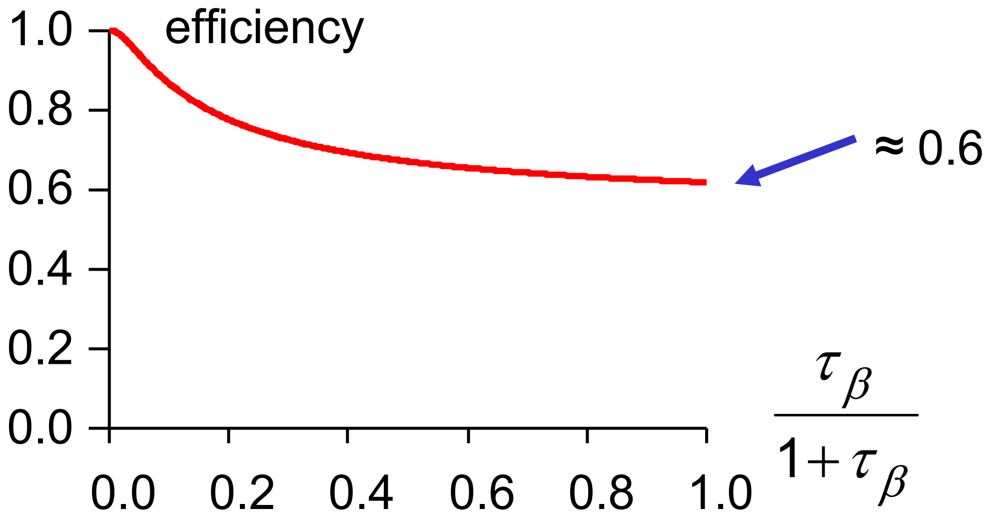


Efficiency

- efficiency of the equi-replicated design $m_1 = m/2$

$n = 100$, $m = 10$

τ_α small



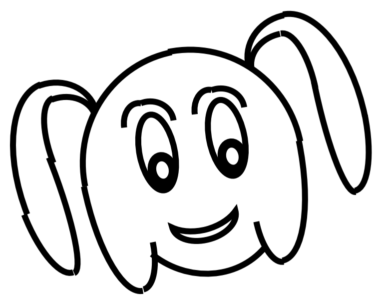
- Bayesian optimal design ($m_1 = m$) singular

$\Rightarrow$ efficiency = 0

5. Outlook

- optimal designs may depend on dispersion matrix $\mathbf{D}$
  - may badly perform, if $\mathbf{D}$ is misspecified
  - $\Rightarrow$ averaged or minimax criteria
- other criteria (D, G) are more sensitive to $\mathbf{D}$
- prediction of individual effects $\mathbf{b}_i$
  - may result in singular designs
- non-linear response ?
- ...

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG



Thank you !