Improved conditional approximations of the population Fisher Information Matrix

Joakim Nyberg, Sebastian Ueckert and Andrew C. Hooker

Pharmacometrics research group

Department of Pharmaceutical Biosciences

Uppsala University

UPPSALA
UNIVERSITET

The initial problem

- Analytic solutions exist for a Gaussian linear mixed effect model, but not for a **non-linear**.

- If normality is assumed, the population likelihood could be addressed by:
  - Linearization: First Order (FO), etc.
  - Approximated by a Laplace Integral approximation…
  - Using Monte Carlo techniques
  - EM-algorithm, Gaussian Quadrature etc.

Notation

The model for individual i
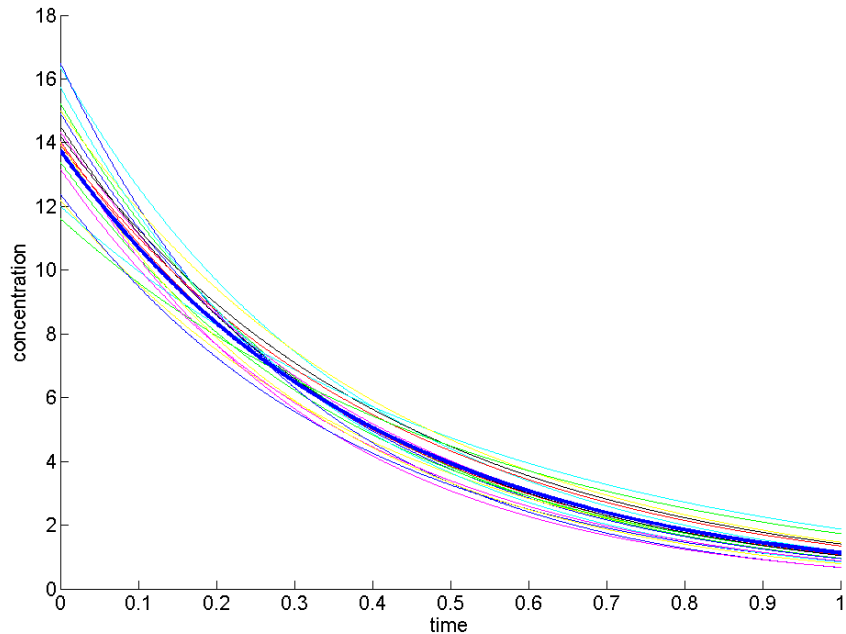
$$y_i = f_i + h_i$$

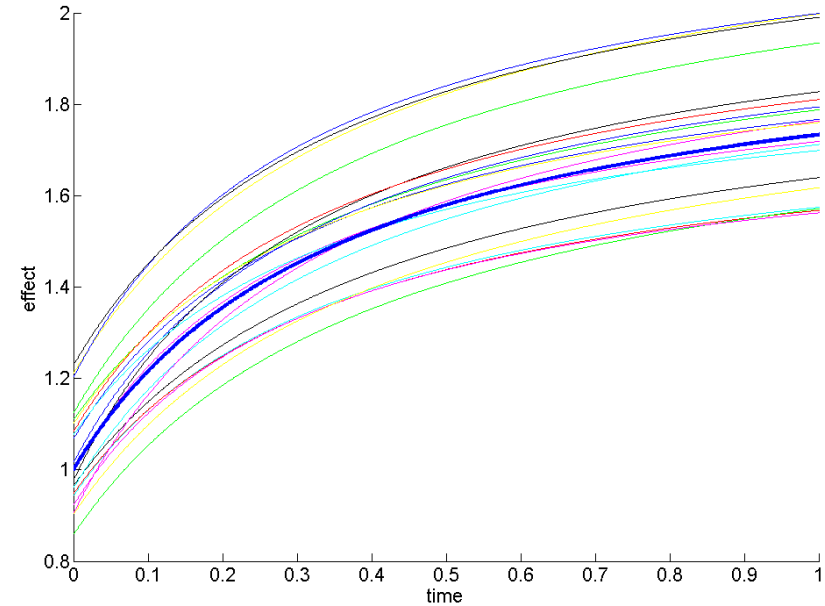structural model   residual model (noise)

Pharmacokinetic

$$f_i = D_i \cdot e^{-\frac{CL_i}{V_i} \cdot t_i}$$
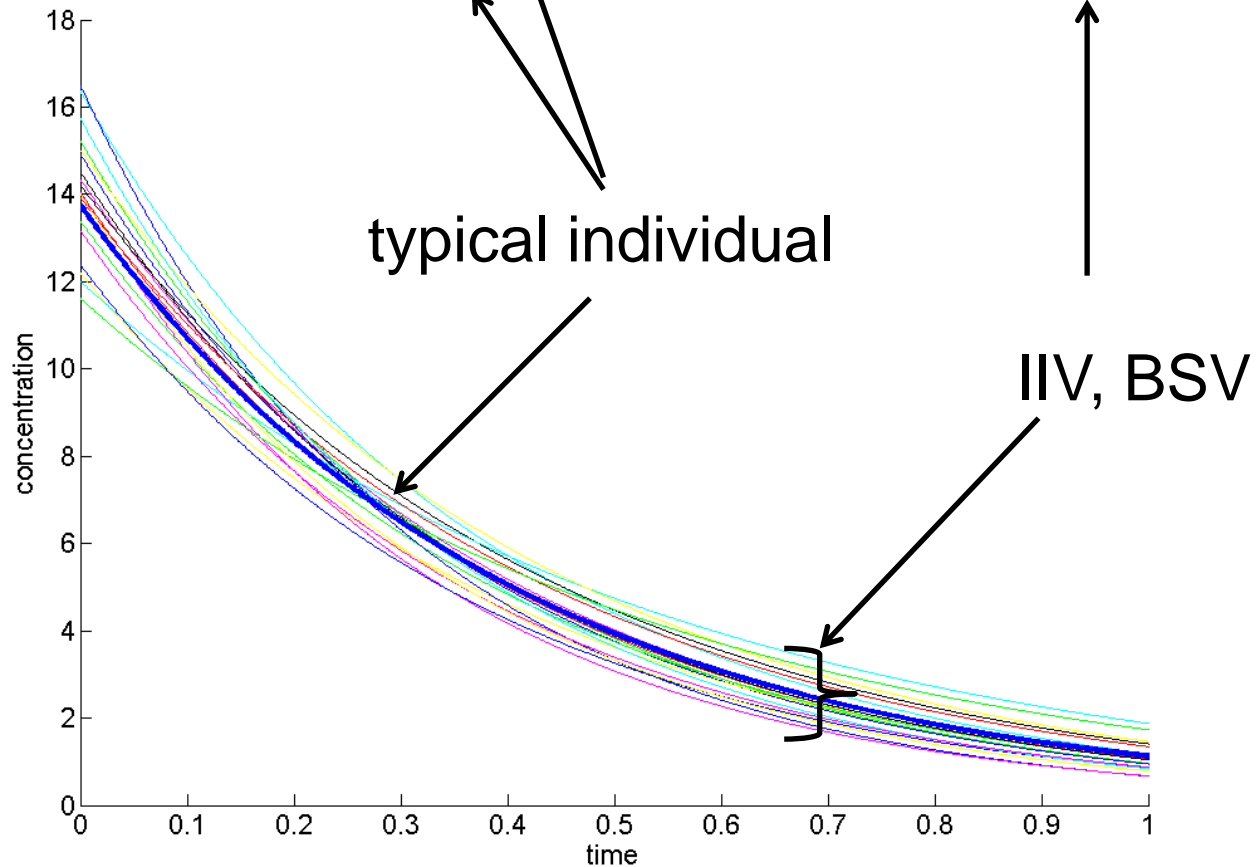
Pharmacodynamic

$$f_i = E_{0,i} + \frac{E_{max,i} \cdot D_i \cdot t_i}{EC_{50,i} + D_i \cdot t_i}$$

Notation – Pharmacokinetic example

$$f_i = D_i \cdot e^{-\frac{CL_i}{V_i} \cdot t_i}$$

$$CL_i = \theta_{CL} \cdot e^{\eta_{CL,i}} \qquad \eta_{CL,i} \sim N\left(0, \omega_{CL}^2\right)$$

$$V_i = \theta_V \cdot e^{\eta_{V,i}} \qquad \eta_{V,i} \sim N\left(0, \omega_V^2\right)$$

typical individual

IIV, BSV

5

$$f_i = D_i \cdot e^{-\frac{CL_i}{V_i} \cdot t_i}$$

$$CL_i = \theta_{CL} \cdot e^{\eta_{CL,i}} \qquad \eta_{CL,i} \sim N\left(0, \omega_{CL}^2\right)$$

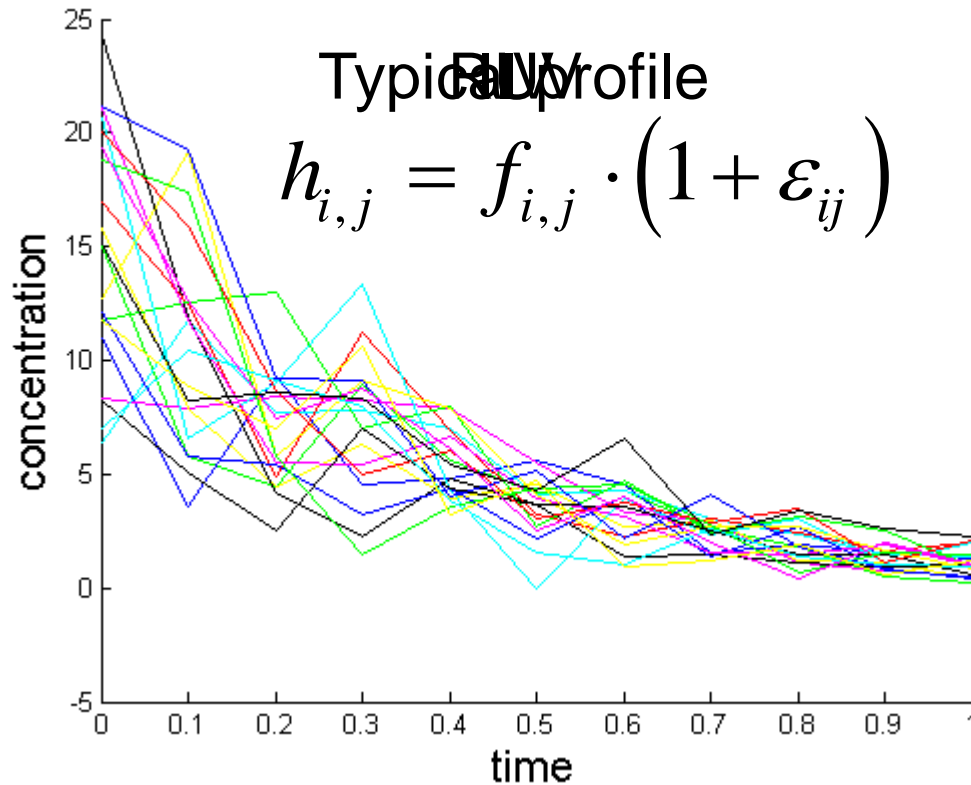$$V_i = \theta_V \cdot e^{\eta_{V,i}} \qquad \eta_{V,i} \sim N\left(0, \omega_V^2\right)$$

Typical IIV profile

$$h_{i,j} = f_{i,j} \cdot \left(1 + \varepsilon_{ij}\right) \qquad \varepsilon_i \sim N\left(0, \sigma^2\right)$$

The model for individual i

$$y_i = f_i + h_i$$

structural model   residual model (noise)

Population parameters

$$\theta, \omega^2 \qquad\qquad \sigma^2$$

Fixed effects   Random effects
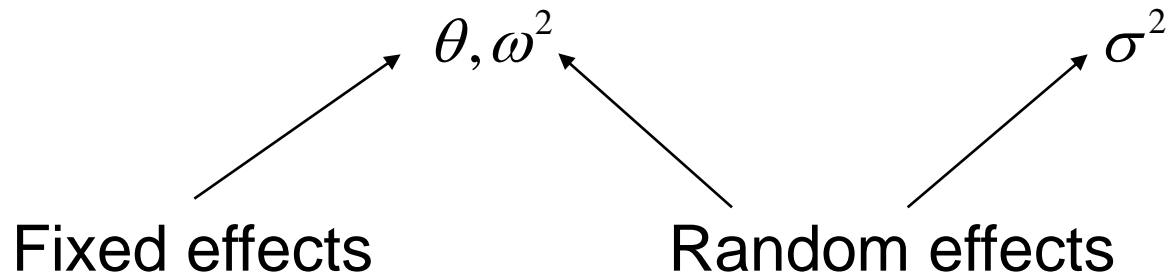
Population likelihood

The sum of the marginal log likelihood for all individuals, given the population parameters:

$$L = \sum \log L_i \left(\theta, \omega^2, \sigma^2 \right)$$

marginal likelihood:

$$L_i = \int_{-\infty}^{\infty} l_i \cdot p\left(\eta \mid \omega^2 \right) d\eta$$

individual likelihood:

$$p\left(y_i \mid \eta_i, \theta, \sigma^2 \right) = l_i = PDF\left[N\left(f\left(\theta, \eta_i, t_i \right), \sigma^2 \right) \right]$$

probability of individual value:

$$p\left(\eta_i, \omega^2 \right) = PDF\left[N\left(0, \omega^2 \right) \right]$$

Calculating Population likelihood by Laplace Integral Approximation

$$L_i\left(\theta,\omega^2,\sigma^2\mid y_i\right) = \int_{-\infty}^{\infty} \underbrace{p\left(y_i\mid\eta_i,\theta,\sigma^2\right)\cdot p\left(\eta_i\mid\omega^2\right)}_{p(x)=\text{joint density}}d\eta$$

p(x)

−∞   ∞
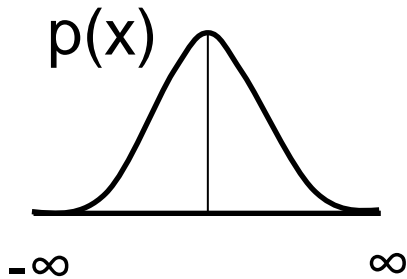
1. Taylor expand ln(p(x)) around peak:

$$\ln\left(p(x)\right) = \ln\left(p(x_0)\right) + (x-x_0)\cdot\ln p'(x_0) + \frac{(x-x_0)^2}{2!}\cdot\ln p''(x_0)$$

but p'(x$_0$)=0 because x$_0$ is at a peak
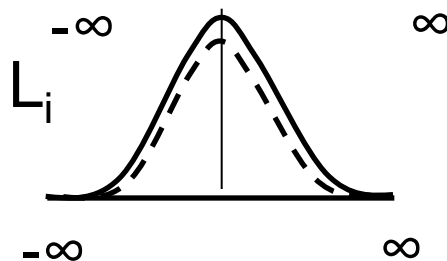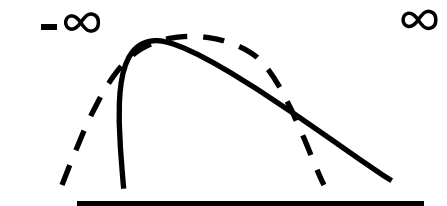
−∞   ∞
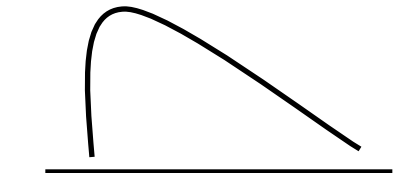
2. Approximate ln(p(x)) by an unnormalized Gaussian:

$$Q(x) \equiv p(x_0)\cdot e^{\frac{(x-x_0)^2}{2!}\cdot\ln p''(x_0)}$$

−∞   ∞

3. Approximate L$_i$ by the normalizing constant of Q:

$$L_i \ \square\ p(x_0)\cdot\sqrt{\frac{2\pi}{-\ln p''(x_0)}}$$

L$_i$

−∞   ∞

* McKay, Information Theory, Inference and Learning Algorithms.
Cambridge University press 2003

−∞   ∞

Calculating Population likelihood by Linearization of the model

$$y_i = f\left(\theta, \eta_i\right) + h\left(\theta, \eta_i, \varepsilon_i\right) \quad \varepsilon_i, \eta_i \sim N\left(0, \Sigma \mid \Omega\right)$$

1. Taylor expand (linearize) h with respect to residual error $\varepsilon_i = 0$:

$$y_i = f\left(\theta, \eta_i\right) + h\left(\theta, \eta_i, 0\right) + \frac{\partial h}{\partial \varepsilon}\left(\theta, \eta_i, 0\right) \cdot \varepsilon$$

2. Taylor expand (linearize) 1. with respect to individual value $\eta_i = \hat{\eta}_i$

$$y_i = \begin{cases} f\left(\theta, \hat{\eta}_i\right) - \dfrac{\partial f\left(\theta, \hat{\eta}_i\right)}{\partial \eta} \cdot \hat{\eta}_i + \dfrac{\partial f\left(\theta, \hat{\eta}_i\right)}{\partial \eta} \cdot \eta + h\left(\theta, \hat{\eta}_i, 0\right) + \\[2em] \dfrac{\partial h\left(\theta, \hat{\eta}_i, 0\right)}{\partial \eta} \cdot \left(\eta - \hat{\eta}_i\right) + \dfrac{\partial h\left(\theta, \hat{\eta}_i, 0\right)}{\partial \varepsilon} \varepsilon + \dfrac{\partial^2 h\left(\theta, \hat{\eta}_i, 0\right)}{\partial \eta \partial \varepsilon} \cdot \left(\eta - \hat{\eta}_i\right) \cdot \varepsilon \end{cases}$$

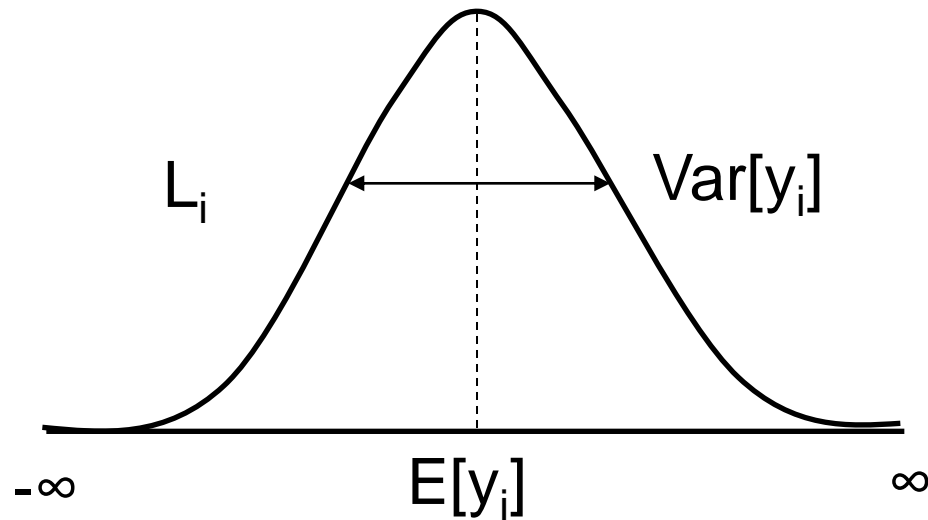3. For simplicity skip interaction term & h(ε=0) and calculate E[y], Var(y):

$$E\left[y_i\right] = f\left(\theta, \hat{\eta}_i\right) - \frac{\partial f\left(\theta, \hat{\eta}_i\right)}{\partial \eta} \cdot \hat{\eta}_i$$

$$Var\left[y_i\right] = E\left[y_i^2\right] - \left(E\left[y_i\right]\right)^2 = \frac{\partial f\left(\theta, \hat{\eta}\right)}{\partial \eta} \cdot \Omega \cdot \frac{\partial f\left(\theta, \hat{\eta}\right)^T}{\partial \eta} + diag\left(\frac{\partial h\left(\theta, \hat{\eta}, 0\right)}{\partial \varepsilon} \cdot \Sigma \cdot \frac{\partial h\left(\theta, \hat{\eta}, 0\right)^T}{\partial \varepsilon}\right)$$

4. Approximate marginal likelihood L_i by assuming normality:

Density function of $y_i \sim N(E[y_i], Var[y_i])$

L_i $\leftrightarrow$ $Var[y_i]$

$-\infty$ $E[y_i]$ ∞

Linearize around typical individual $\eta_i = 0$

$$E[y_i] = f(\theta, \hat{\eta}_i)$$

$$Var[y_i] = \frac{\partial f(\theta, 0)}{\partial \eta} \cdot \Omega \cdot \frac{\partial f(\theta, 0)^T}{\partial \eta} + diag\left(\frac{\partial h(\theta, 0, 0)}{\partial \varepsilon} \cdot \Sigma \cdot \frac{\partial h(\theta, 0, 0)^T}{\partial \varepsilon} \right)$$

Laplace integral approximation

$$L_i \Box \ p(x_0) \cdot \sqrt{\frac{2\pi}{-\ln p''(x_0)}} \cdot e^{\frac{-\ln p'(x_o)^2}{2\ln p''(x_0)}}$$

$$\frac{\partial^2 \ln l_i}{\partial \eta^T \partial \eta} \approx \frac{1}{2} E\left[\frac{\partial \ln l_i}{\partial \eta^T} \times \frac{\partial \ln l_i}{\partial \eta} \right]_{\eta=0}$$

In the end; methods gives exactly the same marginal likelihood!

The FIM for a linear Gaussian Model[*]

$$FIM_{i} = -E\left[\frac{\partial^2 \ln(L_i)}{\partial\Theta\partial\Theta}\right] = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

$$A = \frac{\partial E^T}{\partial\theta} V^{-1} \frac{\partial E}{\partial\theta} + \frac{1}{2} tr\left(\frac{\partial V}{\partial\theta} V^{-1} \frac{\partial V}{\partial\theta} V^{-1}\right)$$

$$B = 2 \cdot tr\left(\frac{\partial V}{\partial[\omega^2,\sigma^2]} V^{-1} \frac{\partial V}{\partial[\omega^2,\sigma^2]} V^{-1}\right)$$

$$C = 2 \cdot tr\left(\frac{\partial V}{\partial[\omega^2,\sigma^2]} V^{-1} \frac{\partial V}{\partial\theta} V^{-1}\right)$$

* Mentré et al, and extended by others (Retout, Hooker, Leonov, Ogungbenro etc.)

Fisher Information Matrix (FIM)

FIM can be calculated in different ways:

Assuming ∂ var(y) w.r.t. the fixed effects≠0

Assuming ∂ var(y) w.r.t. the fixed effects=0

$$FIM_{\text{Full}} = \begin{pmatrix} A^* & C \\ C & B \end{pmatrix}$$

$$FIM_{\text{Reduced}} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

A* is somewhat modified/updated if full is used, i.e.

$$A^* = A + \frac{1}{2} tr\left(\frac{\partial V}{\partial \theta} V^{-1} \frac{\partial V}{\partial \theta} V^{-1} \right)$$

Different between
full and reduced

The new problem

- FO approximation - close to the empirical precision*.

- However, performance of FIM_{Full} is worse than $FIM_{Reduced}$ with the FO approximation**

  - Linearize around typical individual, not enough

  - Second order linearization, not enough

  - How about conditional approximations?

    - FOCE?

* Bazzoli, Retout, Mentré - Fisher information matrix for nonlinear... Statist. Med. 2009
** PODE 2010 (Fedorov & Leonov - Nyberg, Ueckert & Hooker)
** Mielke, Schwabe, *Some Considerations on the FIM.. mODa 9,* 2010, Physica-Verlag HD. p. 129-136.

The FIM FOCE approximation[*]

Take the expectation of the population FIM over the individuals $\hat{\eta}_i \sim N(0,\Omega)$

$$E\left[FIM_i\right] = \int_{-\infty}^{\infty} FIM \ \mathrm{d}\eta \approx \frac{1}{N}\sum_{i=1}^{N} FIM_i\left(\theta,\Omega,\Sigma,\hat{\eta}_i\right)$$

where FIM$_i$ is calculated with the linearization around $\hat{\eta}_i$

$$E\left[y_i\right] = \mathrm{f}\left(\theta,\hat{\eta}_i\right) - \frac{\partial \mathrm{f}\left(\theta,\hat{\eta}_i\right)}{\partial \eta}\cdot\hat{\eta}_i$$

$$Var\left[y_i\right] = E\left[y_i^2\right] - \left(E\left[y_i\right]\right)^2 = \frac{\partial \mathrm{f}\left(\theta,\hat{\eta}\right)}{\partial \eta}\cdot\Omega\cdot\frac{\partial \mathrm{f}\left(\theta,\hat{\eta}\right)^T}{\partial \eta} + diag\left(\frac{\partial \mathrm{h}\left(\theta,\hat{\eta},0\right)}{\partial \varepsilon}\cdot\Sigma\cdot\frac{\partial \mathrm{h}\left(\theta,\hat{\eta},0\right)^T}{\partial \varepsilon}\right)$$

* Retout, Mentré – Further developments of FIM in NLME-models…. J. BioPharm. Stat 2003

The FIM FOCE approximation

+ No need to simulate data, "only" individuals
+ A better approximation of the L_i is used

- Not linearizing around the mode of L_i
- The derivative of the L_i w.r.t. the random effects does not change the $\hat{\eta}_i$

A possible solution to the cons:

Use the same method but linearize around the mode instead

The mode of the marginal likelihood L_i

$$L_i = \int_{-\infty}^{\infty} l_i \cdot p(\eta \mid \Omega) d\eta$$

could be calculated using the
Empirical Bayes Estimate (EBEs) by maximizing:

$$\underset{\hat{\eta}_i}{\arg\max} \left\{ l_i \left(y_i, \hat{\eta}_i, \theta, \Sigma \right) \cdot p\left(\hat{\eta}_i \mid \Omega \right) \right\}$$

But – EBE calculation is dependent on data (l_i)

Approximate the expected data for one individual with η_i

$$\mathrm{E}\left[y_i\right] \approx f\left(\theta, \eta_i\right)$$

and calculate the expected EBE for one individual

$$\widehat{\eta}_{E,i} = \arg\max_{\hat{\eta}_i}\left\{l_i\left(f\left(\theta,\eta_i\right),\hat{\eta}_i,\theta,\Sigma\right)\cdot p\left(\hat{\eta}_i \mid \Omega\right)\right\}$$

Residuals in l_i are now strictly model dependent

$$f\left(\eta_i,\theta\right) - f\left(\hat{\eta}_i,\theta\right)$$

instead of (as previously) data dependent

$$y_i - f\left(\hat{\eta}_i,\theta\right)$$

It is straightforward to get updated expected EBE's when doing numerical differences w.r.t. the population parameters:

$$\hat{\eta}_{E,i+h} = \left[\arg\max_{\hat{\eta}_i} \left\{ l_i \left(f\left(\theta, \eta_i\right), \hat{\eta}_i, \theta, \Sigma \right) \cdot p\left(\hat{\eta}_i \mid \Omega\right) \right\} \right]_{[\theta, \Omega, \Sigma]+h}$$

For each individual contribution to population FIM$_i$:

1. Sample individual η_i from N(0,Ω)

2. Calculate the expected mode $\hat{\eta}_{E,i}$

3. Calculate FIM$_i$ using FOCE approximation around $\hat{\eta}_{E,i}$
    with updating of EBEs when differentiating pop.
    params.

Finally, Monte Carlo integrate over all FIMs:

$$E\left[FIM_i\right] = \int\limits_{-\infty}^{\infty} FIM \ \mathrm{d}\eta \approx \frac{1}{N}\sum_{i=1}^{N} FIM_i\left(\theta,\Omega,\Sigma,\hat{\eta}_{E,i}\right)$$

At PODE 2009 all Population Optimal Design (OD)
Software should evaluate the same simple Warfarin problem…

✓ 1-compartment model, 1st order absorption, oral dose 70 mg

✓ Proportional error model (σ^2=0.01)

✓ 32 subjects with 8 measurements at
   0.5, 1, 2, 6 ,24, 36, 72,120 hours (evaluation)

✓ Fix all parameters except fixed effect ka

Parameters	Fixed effects	ω^2 (IIV, exp)
CL/F (L/h)	0.15	0.07
V/F (L)	8.0	0.02
ka (1/h)	1.0	0.6

Investigation of performance of new FOCE method
Example 1 - results

	Θ_{ka} RSE(%)
"Truth" NONMEM FOCE SSE (1000)	13.59% [13.06-13.88] *
PopED Full FO	6.71%
PopED Full SO	8.94%
PopED Full old FOCE	4.95%
PopED Full FOCE around mode	13.62%
PopED Reduced FO	13.90%
PopED Reduced SO	14.04%
PopED Reduced FOCE	6.49%
PopED Reduced FOCE around mode	12.5% - 13.8%

Example 2 - PK HCV model

- 2 comp PK model, repeated dosing by fast infusion
- Used Linear ODE solver (matrix exponentials)
- 6 PK parameters, 1 res error

- France Mentré will talk more about this model later today

Results Example 2 PK HCV model

D-criterion $|FIM|^{(1/p)}$ with sig^2, only PK
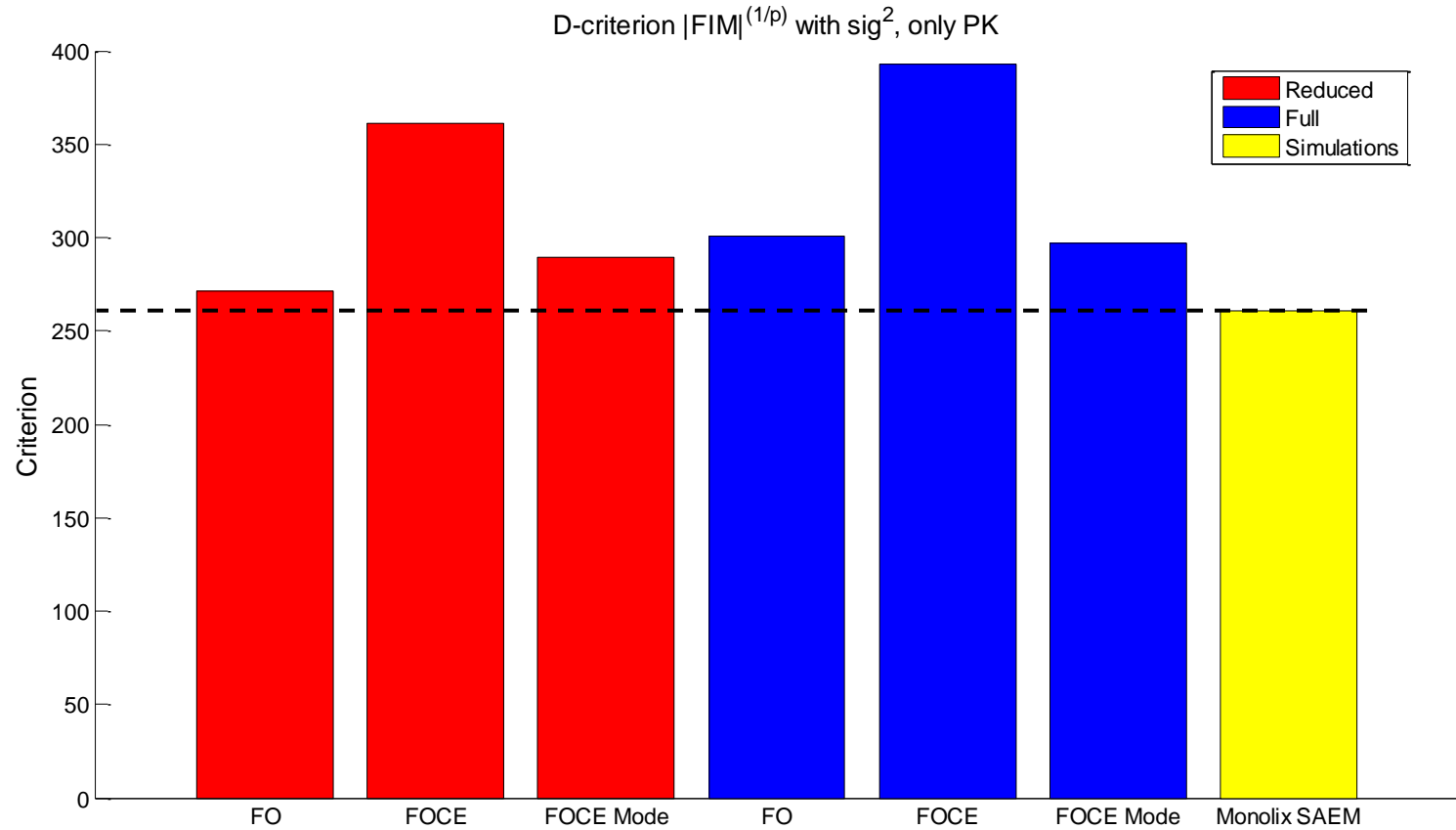
- Simulations in Monolix by C. Bazzoli (1000 sim/est)
- FOCE method used 100 individual samples (Latin hypercube sampled)

Example 3 – "Very" nonlinear Emax model

structural model

$$f_i = \frac{E_{max} \cdot t_i^{\gamma}}{EC_{50,i}^{\gamma} + t_i^{\gamma}}$$

$$EC50_i = \theta_{EC_{50}} \cdot e^{\eta_{EC_{50},i}}$$

$$\theta_{EC_{50}} = 30$$

$$\gamma = 4.5 \text{ FIX}$$

$$E_{max} = 1 \text{ FIX}$$

residual model (noise)

$$h_i = \varepsilon_i$$

$$\eta_{EC_{50},i} \sim N\left(0, \omega_{EC_{50}}^2 = 0.1\right)$$

$$\varepsilon_i \sim N\left(0, \sigma^2 = 0.1\right)$$

$$t_i \in [0,100]$$

Example 3 – Experimental setup

- 200 individuals with the same design

- 2 samples per individual

- Local D-optimal design

Example 3 – Simulations from Emax model

Simulation from Emax-model, n=100
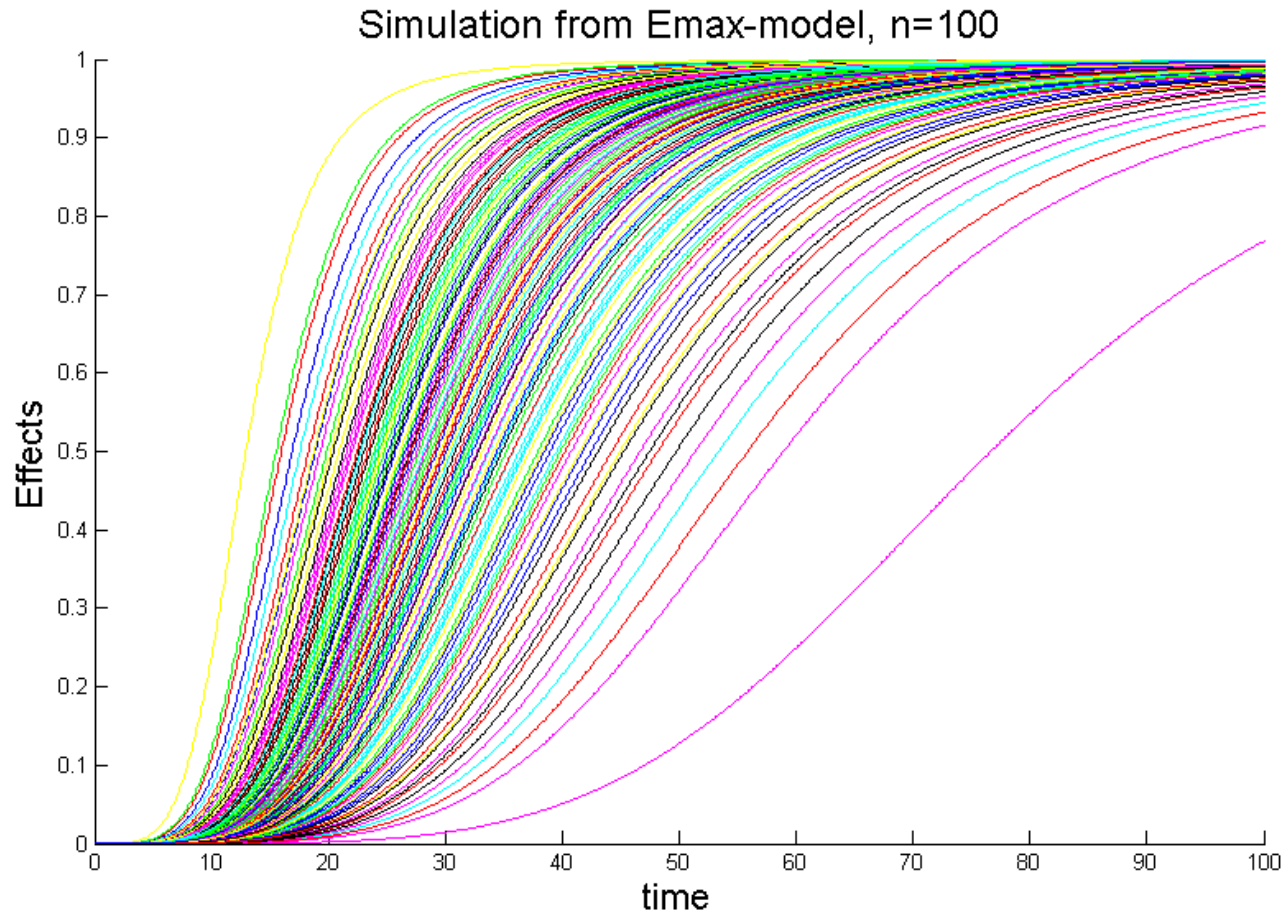
Example 3
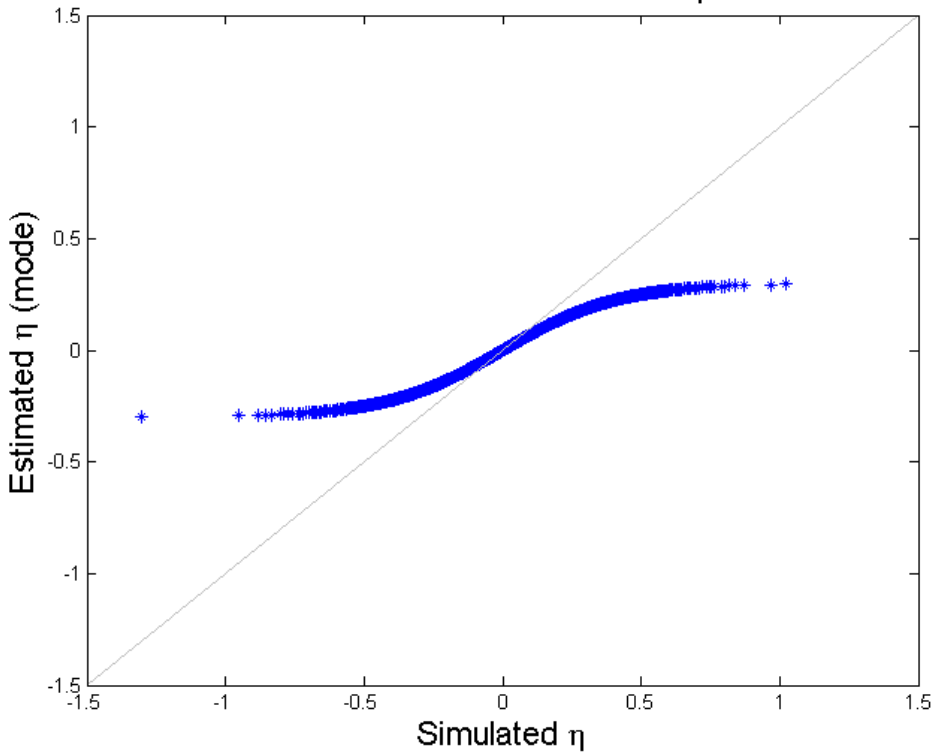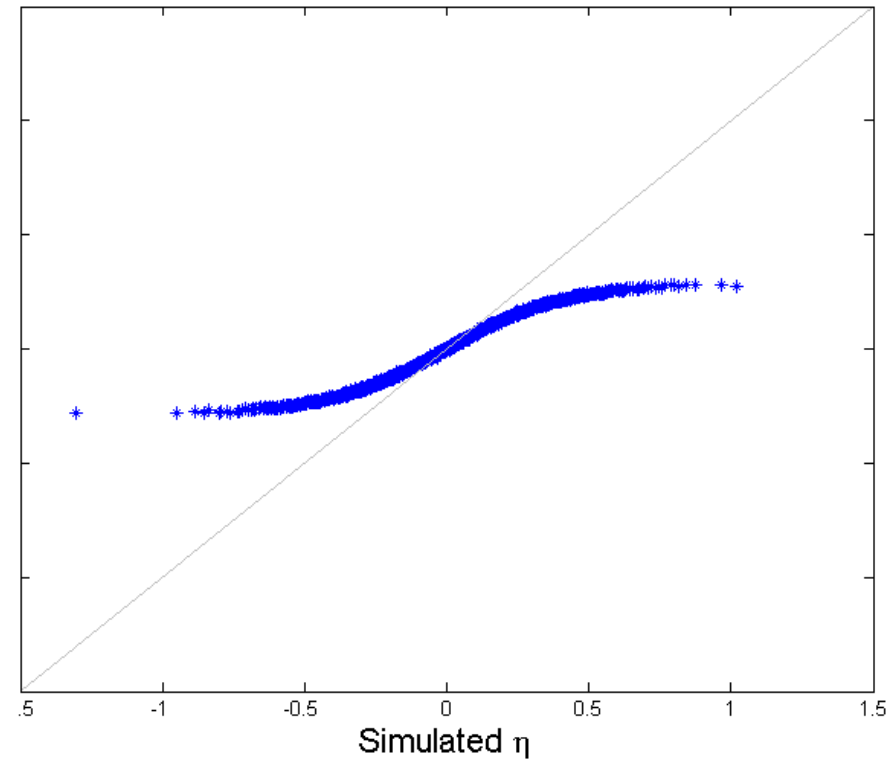# Local D-optimal designs with different approximations
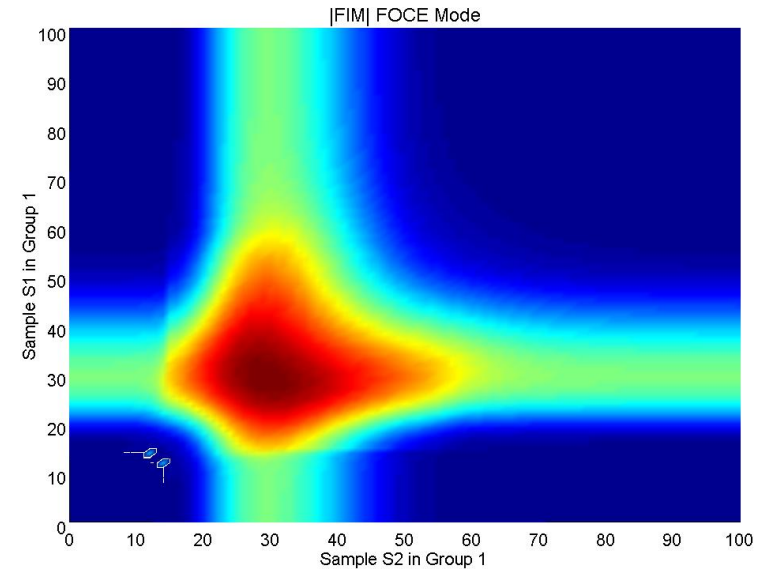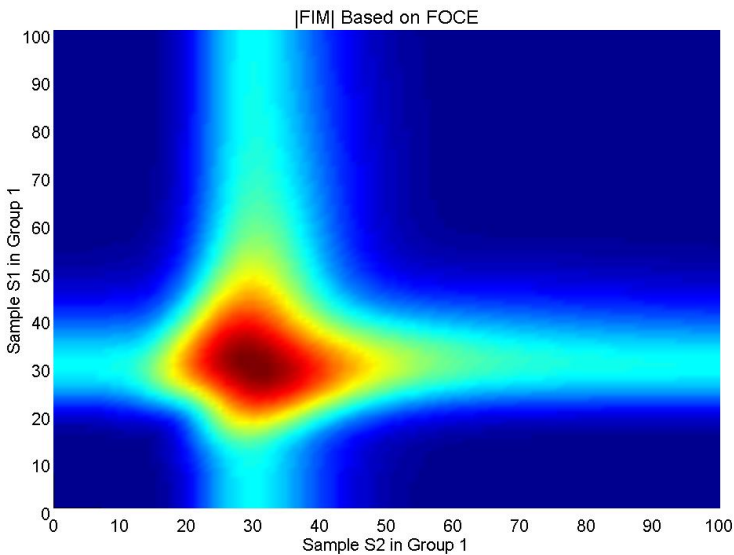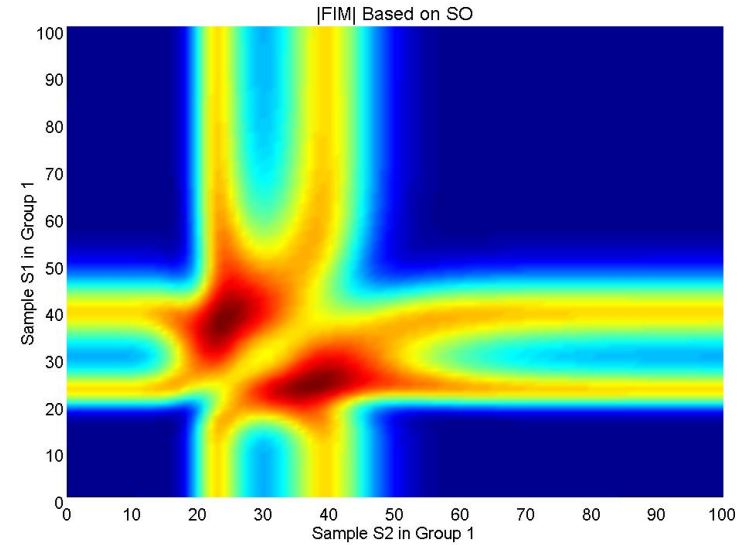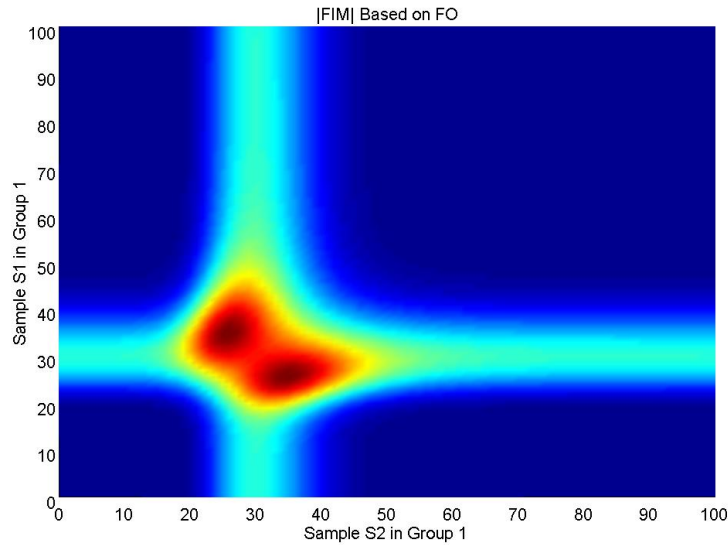
Example 3
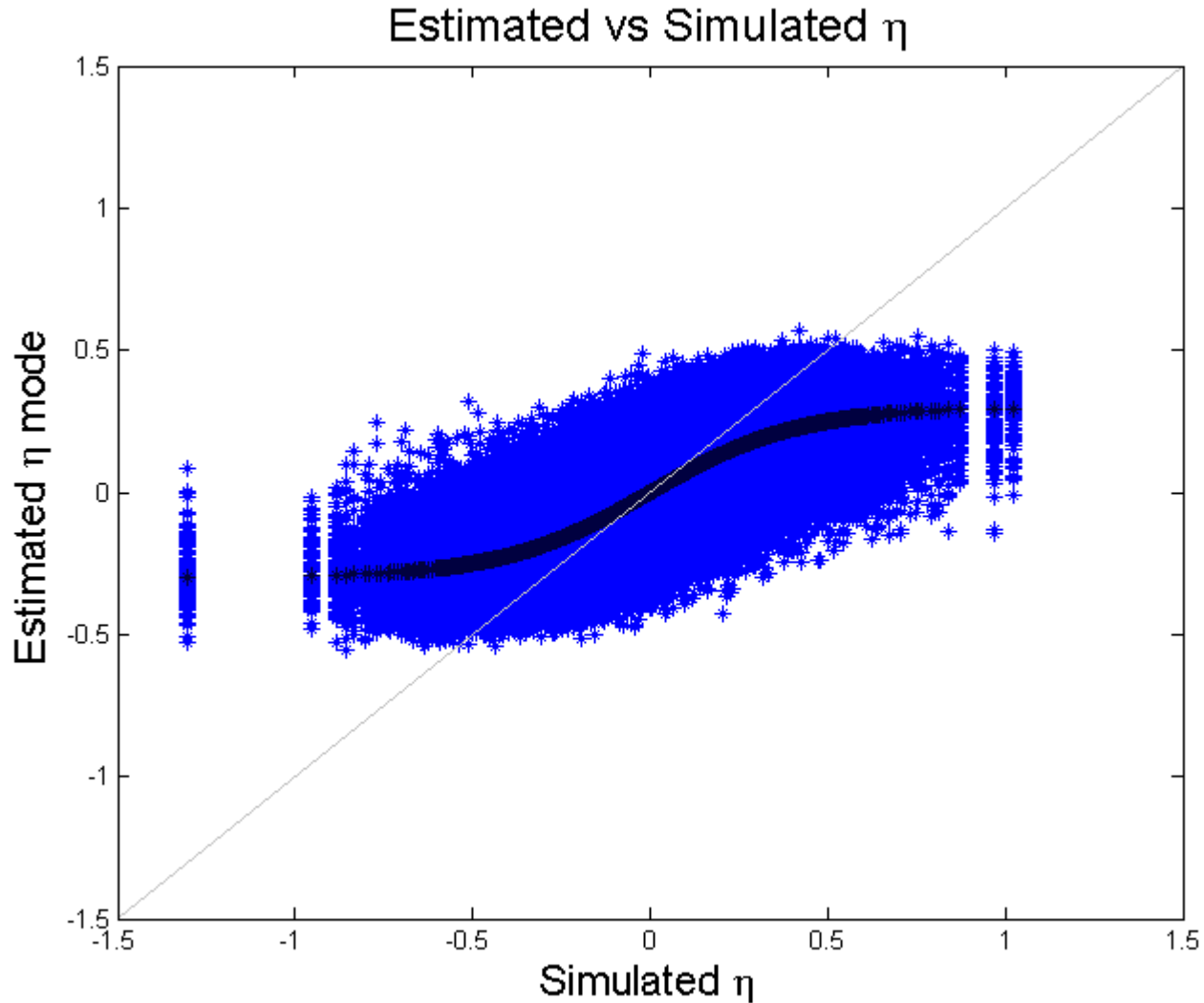## Behavior of expected EBE at (30, 30)

Estimated vs Simulated η

Estimated (with residuals) vs Simulated η

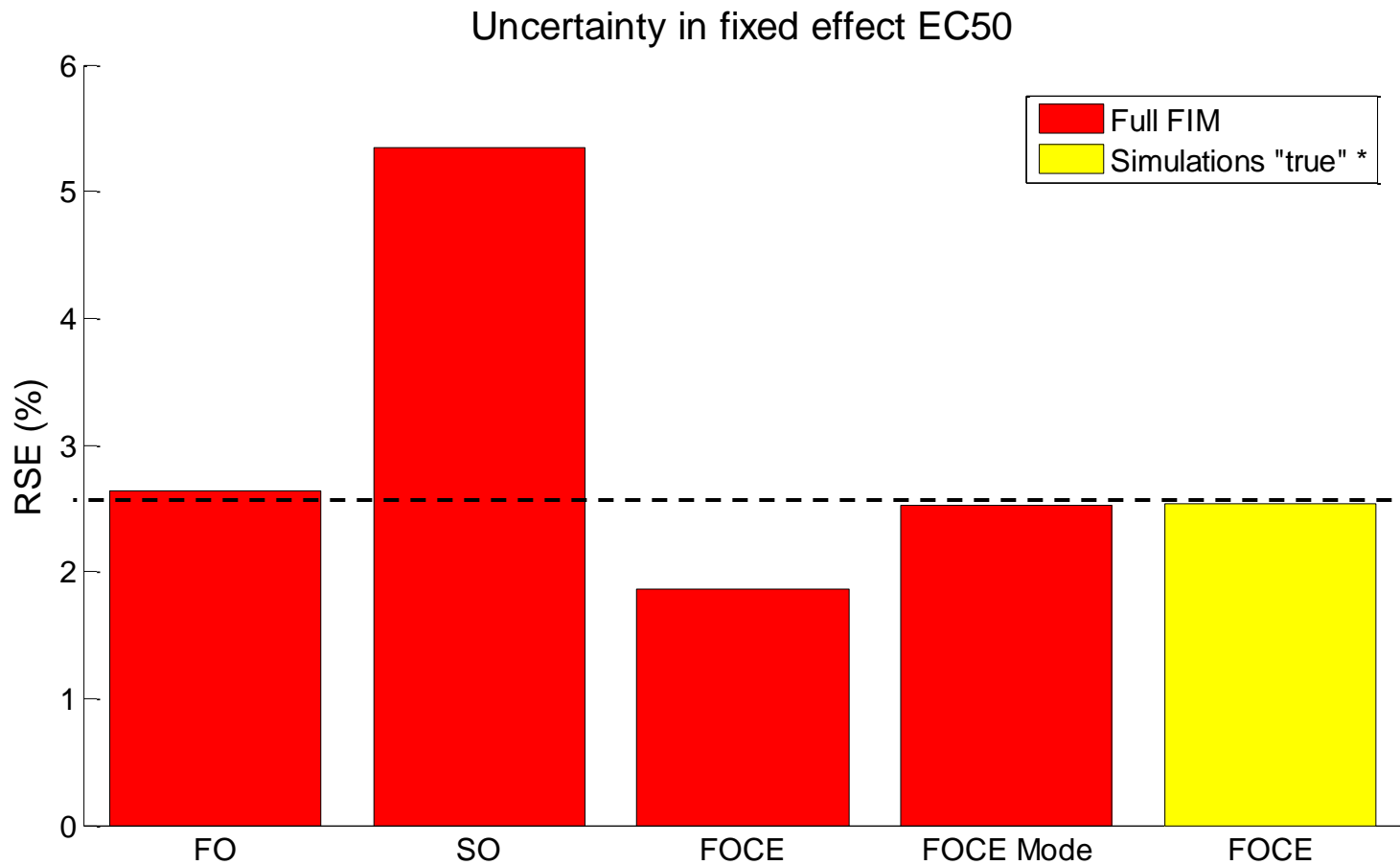* Each expected EBE calculate with 1000 data sets

* 1000 data sets where simulated for each η sample => 1000 EBES/simulated η

Results, FIM evaluated at (30, 30)

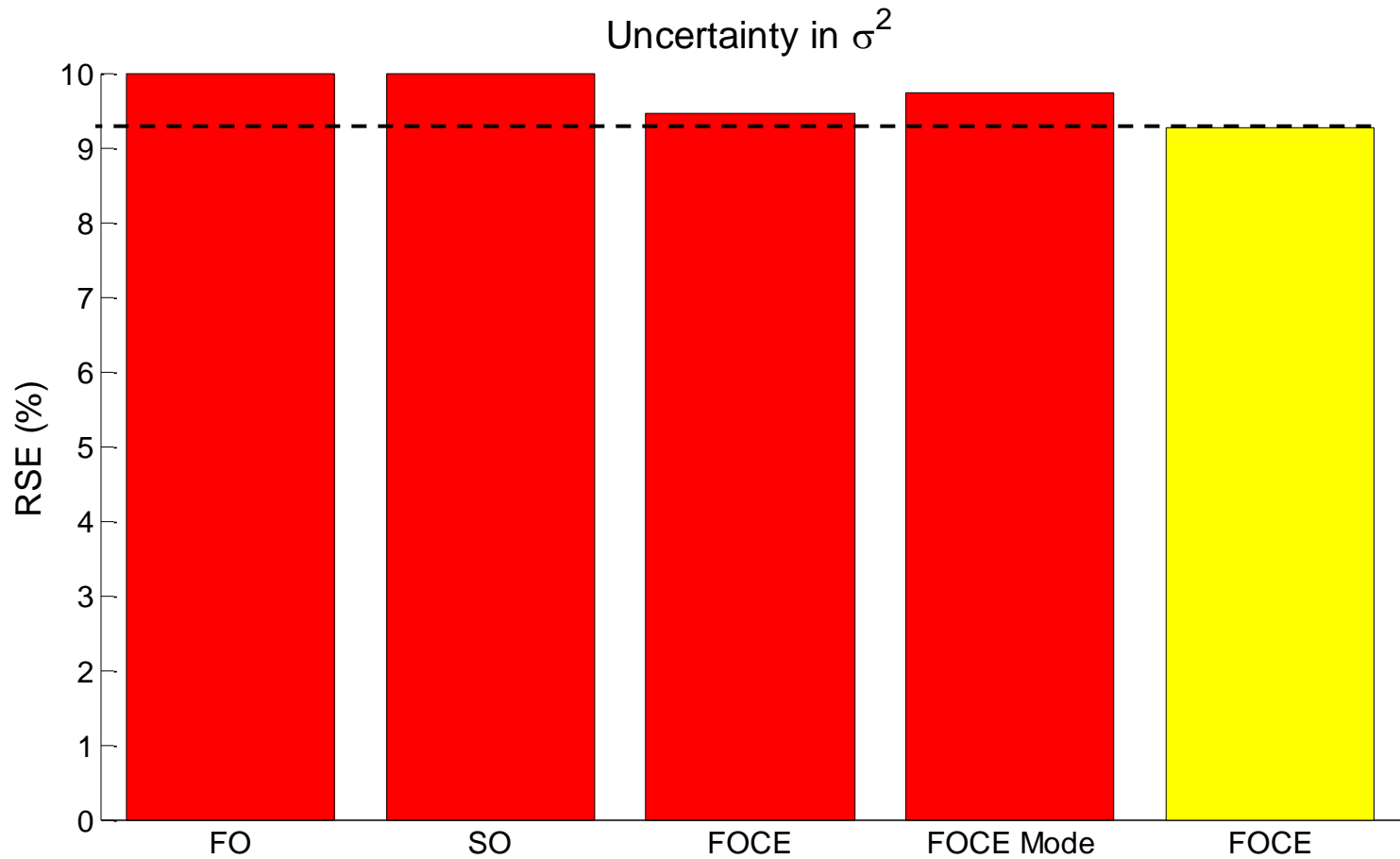* 1000 Simulations/estimations in NONMEM with FOCE

Uncertainty in random effect ω^2_{EC50}

Uncertainty in σ^2

Pros and cons of FIM FOCE around mode

+   Address the full/reduced differences (at least in the examples we investigated)

+   Possibly more realistic predictions of the precision (RSE)

+   The optimal design might be more accurate

-   Slower than previous methods (FO, FOCE)

-   Could suffer from shrinkage if very sparse (or uninformative) designs

-   No closed form solution as in the FO based FIM

What I would do?

- Use the FO based reduced FIM whenever possible
  - After optimizing, evaluate with FOCE Mode FIM to possibly get better predictions of precision.

- When not; Evaluate a design with both FO Full and FOCE Mode Full
  - if similar, use the FO Full FIM to optimize
  - otherwise use FOCE Mode Full FIM to optimize

Take home message – Design is important, regardless of approximation method….