Design and Analysis of Experiments in Healthcare

A Symbol of Excellence

Optimal design and parameter estimation for population PK/PD models

Sergei Leonov, Tobias Mielke

ICON Innovation Centre

July 7, 2015

Outline

- Nonlinear mixed effects models: population PK/PD
- Optimal design, Fisher information matrix (FIM)
- Parameter estimation
- Approximations of the information matrix
- Population optimal design software
- Adaptive designs: estimation and optimal design

Design of population PK/PD studies

- Design $\xi = \{\mathbf{x}_i, n_i\}$, what we optimize:
  - Design points: sampling sequences $\mathbf{x}_i$
  - Number of patients n_i on sampling sequence $\mathbf{x}_i$
  - Number of sampling times k_i in sequence $\mathbf{x}_i$
- Why important:
  - Often limited number of samples allowed
  - Special populations (ethical/physiological reasons)
  - Operational issues/costs

Nonlinear mixed effects model: population PK

- γ_i - response parameters of patient i:

  normal, $\gamma_i \sim N(\gamma^0, \Omega)$, or log-normal ($\gamma^0$ - "typical values")

- Data $y(x_{ij}) = \eta(x_{ij}, \gamma_i) + \varepsilon_{ij}, \quad j = 1, \ldots, k_i. \qquad (1)$

  {*or with proportional error*: $y(x_{ij}) = \eta(x_{ij}, \gamma_i)\left[1 + \varepsilon_{ij}^p\right] + \varepsilon_{ij}$ }

  $\varepsilon_{ij} \sim N(0, \sigma^2), \quad \varepsilon_{ij}^p \sim N(0, \sigma_p^2)$

- Combined vector of parameters: $\theta = (\gamma^0;\ \Omega;\ \sigma^2;\ \sigma_p^2)$

Example: one-compartment model, single dose D at $x = 0$:

$$\eta(x, \gamma) = \frac{D k_a}{V(k_a - k_e)}\left(e^{-k_e x} - e^{-k_a x}\right), \quad \gamma = (k_a, k_e, V)^T$$

Key: calculate (*approximate*) individual information matrix $\mu(\mathbf{x}, \theta)$

of a k-dimensional predictor $\mathbf{x}$ (sequence of sampling times)

Optimal design

Information matrix : n_i patients on seq. $\mathbf{x}_i \Longrightarrow \mathbf{M}_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \, \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$

$\mathbf{M}(\xi, \boldsymbol{\theta}) = \dfrac{\mathbf{M}_N(\boldsymbol{\theta})}{N} = \sum_i w_i \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$ - normalized information, per observation

$\xi = \{w_i, \mathbf{x}_i\}$ - normalized design; $w_i = n_i/N$ - weights

Criterion of optimality $\Psi[\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})] \rightarrow \min_{\xi}$: minimization with respect to

- Continuous designs: $0 \leq w_i \leq 1, \ \Sigma_i w_i = 1,$
- Admissible sampling sequences $\mathbf{x}_i \in \mathbf{X}$ - design region.

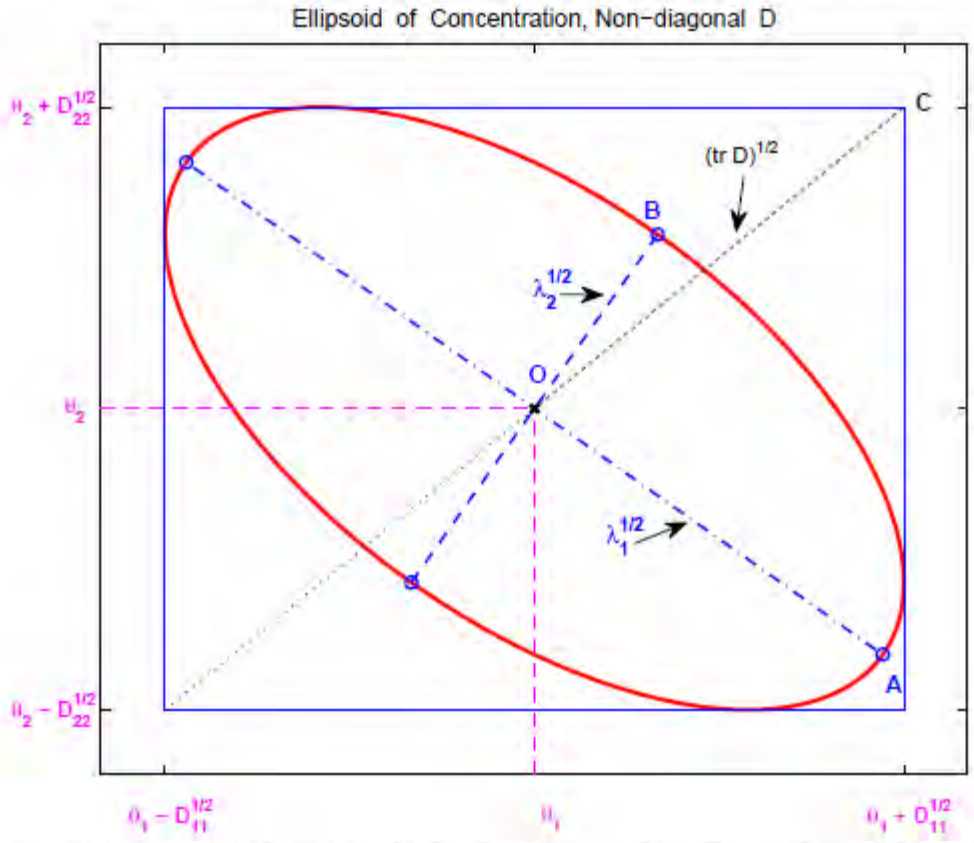
Equivalence Theorem: *Kiefer, Wolfowitz (1960), Fedorov (1972)* - background for 1-st order optimization algorithms (Fedorov-Wynn)

Atkinson, Donev (1992); Fedorov, Hackl (1997); Fedorov, Leonov (2013, Ch. 3)

Optimality criteria (*D - A - E*)

$$\mathbf{D} = \mathbf{M}^{-1} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}, \quad \lambda_{1,2} \; - \; \text{eigenvalues of } \mathbf{D}, \quad |\mathbf{D}| = \lambda_1 \lambda_2$$

$$\text{"Volume"} \; V = \pi R^2 \, |\mathbf{D}|^{1/2}, \quad \text{tr}\, \mathbf{D} = \lambda_1 + \lambda_2 = D_{11} + D_{22}$$



Ellipsoid of concentration for $R = 1$: $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N)^T \mathbf{M} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_N) = R^2$. Dash-dotted/dashed lines: largest/smallest principal axes. $(\overline{OA})^2 = \lambda_1$, $(\overline{OB})^2 = \lambda_2$, $(\overline{OC})^2 = \lambda_1 + \lambda_2 = D_{11} + D_{22}$.

Nonlinear mixed effects model

$$y(x_{ij}) \;=\; \eta(x_{ij}, \boldsymbol{\gamma}_i) + \varepsilon_{ij}, \quad \text{or} \quad \mathbf{y}_i|\boldsymbol{\gamma}_i \;\sim \mathcal{N}\left[\boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\gamma}_i), \sigma^2 I_{k_i}\right]$$

Likelihood for nonlinear η: no closed-form solution $[\boldsymbol{\gamma}_i \sim N(\boldsymbol{\gamma}^0, \boldsymbol{\Omega})]$:

$$L(\boldsymbol{\gamma}^0, \mathbf{y}_i) \sim \int \exp\left[-C||\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\gamma}_i)||^2 - (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}^0)^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}^0)\right] d\boldsymbol{\gamma}_i$$

Simplest approximation: linearize at $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}'$

$$\mathbf{y}_i \approx \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\gamma}') + \mathbf{Z}(\mathbf{x}, \boldsymbol{\gamma}')\,(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}') + \boldsymbol{\varepsilon}_i, \quad \mathbf{Z} = \mathbf{Z}(\mathbf{x}, \boldsymbol{\gamma}') = \left[\frac{\partial \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right]\Bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}'}$$

For example, take $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^0$ (population mean), or $\boldsymbol{\gamma}' = \tilde{\boldsymbol{\gamma}}$ (guess) $\Rightarrow$

Expressions for FIM are different [*Mielke (2012, Ph.D. Thesis; PODE)*]

Normally distributed observations

Gaussian $\mathbf{y}_{ij}|\mathbf{x}_i \sim \mathcal{N}\left[\boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta}),\ \mathbf{S}(\mathbf{x}_i, \boldsymbol{\theta})\right], \quad j = 1, \ldots, n_i,$

where $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) = \left[\eta_1(\mathbf{x}, \boldsymbol{\theta}), \ldots, \eta_k(\mathbf{x}, \boldsymbol{\theta})\right]^T$, $\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$ is a $k \times k$ matrix.

n_i independent observations are taken at predictor levels $\mathbf{x}_i$, $\sum_{i=1}^{n} n_i = N$.

$\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ - information matrix of a k-dimensional sequence $\mathbf{x}$:

$$\mu_{\alpha\beta}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\eta}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \theta_\beta} + \frac{1}{2} \operatorname{tr}\left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta}\right],$$

Muirhead (1982), Magnus and Neudecker (1988)

Linearization at population mean: FIM has both terms
Linearization at a guess: FIM does **not** have trace term

Normally distributed observations (cont.)

1. Maximum likelihood

$$\boldsymbol{\theta}_N^{MLE} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}_N(\boldsymbol{\theta}), \quad \text{log-likelihood } \mathcal{L}_N(\boldsymbol{\theta}) =$$

$$= -\frac{1}{2N}\sum_{i,j}\left\{\ln|\mathbf{S}(\mathbf{x}_i,\boldsymbol{\theta})| + [\mathbf{y}_{ij}-\boldsymbol{\eta}(\mathbf{x}_i,\boldsymbol{\theta})]^T\,\mathbf{S}^{-1}(\mathbf{x}_i,\boldsymbol{\theta})\,[\mathbf{y}_{ij}-\boldsymbol{\eta}(\mathbf{x}_i,\boldsymbol{\theta})]\right\}.$$

$$\{\text{Regularity conditions on } \boldsymbol{\eta}, \mathbf{S}\} \quad +$$

$$\{\text{Existence of the limit } \mathbf{M}(\boldsymbol{\theta}_t) = \lim_{N\to\infty}\frac{1}{N}\sum_i n_i\,\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta}_t)\}$$

$$\Downarrow$$

Asymptotic normality of MLE: $\sqrt{N}(\boldsymbol{\theta}_N^{MLE}-\boldsymbol{\theta}_t) \sim \mathcal{N}\left[0, \mathbf{M}^{-1}(\boldsymbol{\theta}_t)\right].$

Normal observations: variations of LS

2. When $\mathbf{S}(\mathbf{x}_i, \boldsymbol{\theta}) \equiv \mathbf{S}(\mathbf{x}_i)$: GLS

$$\tilde{\boldsymbol{\theta}}_N^{GLS} = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i,j} \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right]^T \mathbf{S}^{-1}(\mathbf{x}_i) \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right].$$

3. When $\mathbf{S}$ depends on $\boldsymbol{\theta}$, it is tempting to use

Bad choice

$$\tilde{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i,j} \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right]^T \mathbf{S}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}) \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right].$$

4. *Iteratively reweighted least squares* (IRLS):

$$\tilde{\boldsymbol{\theta}}_N^{IRLS} = \lim_{N\to\infty} \boldsymbol{\theta}_s, \text{ where } \boldsymbol{\theta}_s = \arg\min_{\boldsymbol{\theta}} V_N^{(1)}(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}),$$

$$V_N^{(1)}(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) = \frac{1}{N} \sum_{i,j} \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right]^T \mathbf{S}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}_{s-1}) \left[\mathbf{y}_{ij} - \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\theta})\right].$$

$$\Downarrow$$

$$\sqrt{N}(\tilde{\boldsymbol{\theta}}_N^{IRLS} - \boldsymbol{\theta}_t) \sim \mathcal{N}\left[0, \tilde{\mathbf{M}}^{-1}(\boldsymbol{\theta}_t)\right],$$

$$\tilde{\mathbf{M}}^{-1}(\boldsymbol{\theta}_t) = \left\{ \lim_{N\to\infty} N^{-1} \left[\sum_i n_i \mathbf{Z}^T(\mathbf{x}_i, \boldsymbol{\theta})\, \mathbf{S}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}) \mathbf{Z}^T(\mathbf{x}_i, \boldsymbol{\theta}) \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} \right\}^{-1}.$$

$$\tilde{\mathbf{M}}(\boldsymbol{\theta}_t) \leq \mathbf{M}(\boldsymbol{\theta}_t) \quad [\mathbf{M}(\boldsymbol{\theta}_t) \text{ - limiting matrix of the MLE}]$$

Normal observations: variations of LS (cont)

5. *Combined iteratively reweighted least squares* (CIRLS):

$$\hat{\boldsymbol{\theta}}_N^{CIRLS} = \lim_{s\to\infty} \boldsymbol{\theta}_s, \text{ where } \boldsymbol{\theta}_s = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_N^{(2)}(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}),$$

$$V_N^{(2)}(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) = V_N^{(1)}(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}) + R_N(\boldsymbol{\theta}, \boldsymbol{\theta}_{s-1}).$$

(R_N include squared deviations of predicted $\mathrm{S}(\mathrm{x}, \theta)$ from observed residual matrices)

CIRLS is equivalent to MLE:

$$\lim_{N\to\infty} P\{\hat{\boldsymbol{\theta}}_N^{CIRLS} \in \boldsymbol{\Theta}_N\} = 1, \ \boldsymbol{\Theta}_N - \text{ stationary points of log-likelihood } \mathcal{L}_N,$$

$$\boldsymbol{\Theta}_N = \left\{\boldsymbol{\theta} : \ \frac{\partial \mathcal{L}_N(\boldsymbol{\theta})}{\partial \theta_j} = 0, \ \ j = 1, \ldots, m\right\}.$$

Fedorov and Leonov (2004), Fedorov and Leonov (2013), Ch. 1.7

Information matrix, sampling sequence x

(1) Gaussian $\mathbf{Y}$: $\mathbf{E}[\mathbf{Y}|\mathbf{x}] = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$, $\mathbf{Var}[\mathbf{Y}|\mathbf{x}] = \mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$

$\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ - information matrix of k-dimensional sequence $\mathbf{x}$:

$$\mu_{\alpha\beta}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\eta}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \theta_\beta} + \frac{1}{2} \operatorname{tr}\left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta}\right]$$

(2) First-order approximation of variance matrix $\mathbf{S}$, model (1), normal $\boldsymbol{\gamma}$:

$$\mathbf{S}(\mathbf{x}, \boldsymbol{\theta}) \simeq \mathbf{Z}\,\boldsymbol{\Omega}\,\mathbf{Z}^T + \sigma_P^2 \operatorname{Diag}[\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})\,\boldsymbol{\eta}^T(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{Z}\,\boldsymbol{\Omega}\,\mathbf{Z}^T] + \sigma^2 \mathbf{I}_k,$$

$$\mathbf{Z} = \mathbf{Z}(\mathbf{x}, \boldsymbol{\gamma}^0) = \left[\frac{\partial \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}_\alpha}\right]\Bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}$$

No proportional variability ($\sigma_p^2 = 0$): $\mathbf{S} = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}^T + \sigma^2 \mathbf{I}_k$

Population optimal design software

- PODE initiated in 2006
- Discussion of population optimal design tools started at PODE 2007
  - Mentré et al. (2007, 2011, PAGE)
  - Nyberg et al. (2015, *British J. Clin. Pharmacol.*)
    - Comparison of FIM for specific sequences of sampling times
      1. Approximations of FIM (closed-form expressions)
      2. Monte Carlo simulations (simulate data, get empirical variance-covariance matrix)

Comparison of population optimal design tools

BJCP British Journal of Clinical Pharmacology

Br J Clin Pharmacol / 79:1 / 6–17

Methods in Clinical Pharmacology Series

Methods and software tools for design evaluation in population pharmacokinetics–pharmacodynamics studies

Joakim Nyberg,[1] Caroline Bazzoli,[2] Kay Ogungbenro,[3] Alexander Aliev,[4] Sergei Leonov,[5] Stephen Duffull,[6] Andrew C. Hooker[1] & France Mentré[7]

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden, [2]Laboratoire Jean Kuntzmann, Département Statistique, University of Grenoble, Grenoble, France, [3]Centre for Applied Pharmacokinetic Research, School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Manchester, UK, [4]Institute for Systems Analysis, Russian Academy of Sciences, Moscow, Russia, [5]AstraZeneca, Wilmington, DE, USA, [6]School of Pharmacy, University of Otago, Dunedin, New Zealand and [7]INSERM U738 and University Paris Diderot, Paris, France

Correspondence
Professor France Mentré PhD, MD, Université Paris Diderot, INSERM, Paris 75018, France.
Tel.: +33 1 57 27 77 59
Fax: + 33 1 57 27 75 21
E-mail: france.mentre@inserm.fr

Keywords
Fisher information matrix, nonlinear mixed effect models, optimal design, population design, population pharmacokinetics–pharmacodynamics

Received
30 July 2013
Accepted
9 February 2014
Accepted Article Published Online
18 February 2014

Software comparison: warfarin, one-compartment

One-compartment, 1st order absorption, single dose D

Response parameters $\gamma = (k_a, CL, V),\ k_e = CL/V$

Individual parameters: log-normal

$$\gamma_i = \gamma^0 e^{\zeta_i},\ \zeta_i \sim \mathcal{N}(\mathbf{0}, \Omega),\ \Omega - \text{diagonal}$$

Measurements:

$$y_{ij} = \eta(x_{ij}, \gamma_i)\ (1 + \varepsilon_{ij}^p),\ \varepsilon_{ij}^p \sim \mathcal{N}(0, \sigma_p^2)$$

Parameter $\boldsymbol{\theta} = (k_a^0, CL^0, V^0;\ \omega_{k_a}^2, \omega_{CL}^2, \omega_V^2;\ \sigma_p^2)$

Software comparison: warfarin (cont.)

Information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$: block form, *Retout and Mentré (2003)*

$$\boldsymbol{\mu} = \left\{ \begin{array}{cc} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{array} \right\}, \quad \mu_{\alpha\beta}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\eta}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \theta_\beta} + \frac{1}{2} \operatorname{tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta} \right]$$

$$\mathbf{A} = \mathbf{Z}^T \mathbf{S}^{-1} \mathbf{Z} + \tfrac{1}{2} \operatorname{tr} \quad \text{(derivatives wrt } \gamma_\alpha\text{)}$$
$$\mathbf{C} = \tfrac{1}{2} \operatorname{tr} \quad \text{(mixed derivatives wrt } \gamma_\alpha \text{ and } [\omega_\beta^2, \sigma_M^2]\text{)}$$
$$\mathbf{B} = \tfrac{1}{2} \operatorname{tr} \quad \text{(derivatives wrt } [\omega_\beta^2, \sigma_M^2]\text{)}$$

Compared $\mathbf{D}_a = [\boldsymbol{\mu}(\mathbf{x},\boldsymbol{\theta})]^{-1}$ and $\mathbf{D}_e$ (empirical variance-covariance matrix, MC):

- *Reduced option* (dropping the trace term) → $\mathbf{D}_a$ and $\mathbf{D}_e$ are very close
- *Full option* → visible difference for some elements of $\mathbf{D}$
- Example of overestimation of information when keeping trace term: *Mielke and Schwabe* (2010)
- Discussion of various approximation options: *Fedorov and Leonov* (2013), Section 7.5.6

Software comparison: combined PK/PD, HCV

Drug for treating chronic hepatitis C (HCV) infection

$$\begin{cases} \dot{f}_0(t) &= -k_a f_0(t) &&+ r(t) \\ \dot{f}_1(t) &= k_a f_0(t) &- k_e f_1(t) & \\ \eta_1(t) &= f_1(t)/V_1 & & \end{cases}$$

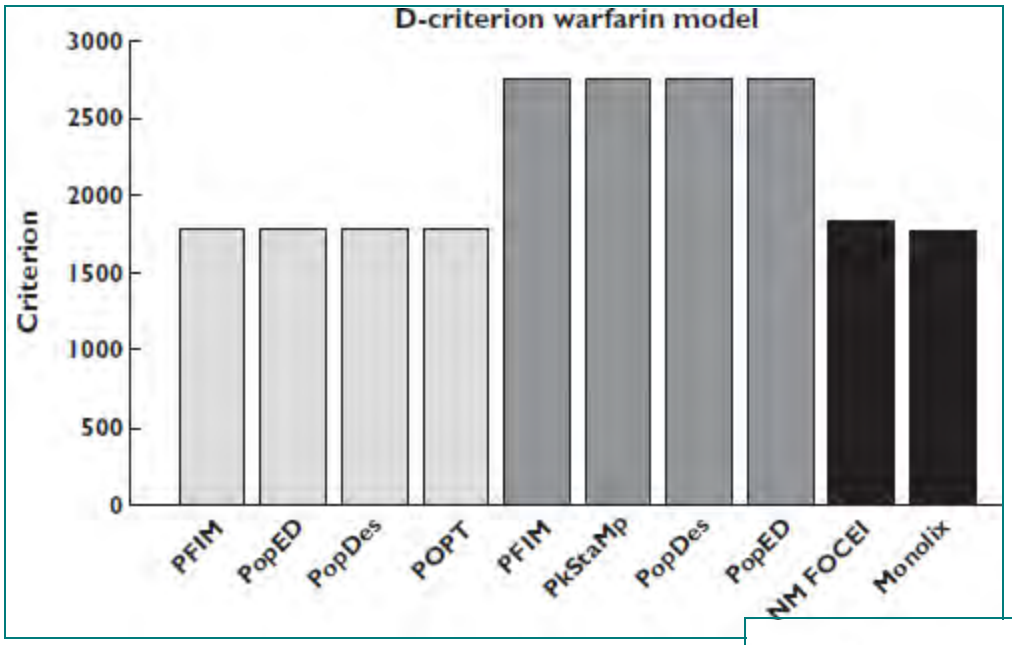
PK: parameters (k_a, k_e, V_1), response η_1 (continuous infusion)

$$\begin{cases} \dot{g}_1(t) &= -C_2 g_1(t) - C_1 g_1(t) g_3(t) + C_3 \\ \dot{g}_2(t) &= -\delta g_2(t) + C_1 g_1(t) g_3(t) \\ \dot{g}_3(t) &= C_4 \left[1 - \frac{1}{1+(EC_{50}/\eta_1)^n}\right] g_2(t) - c g_3(t) \\ \eta_2(t) &= \log_{10} g_3(t) \end{cases}$$

$g_1(t)$ - "target cells", $g_2(t)$ - infected cells, $g_3(t)$ - viral particles (load)

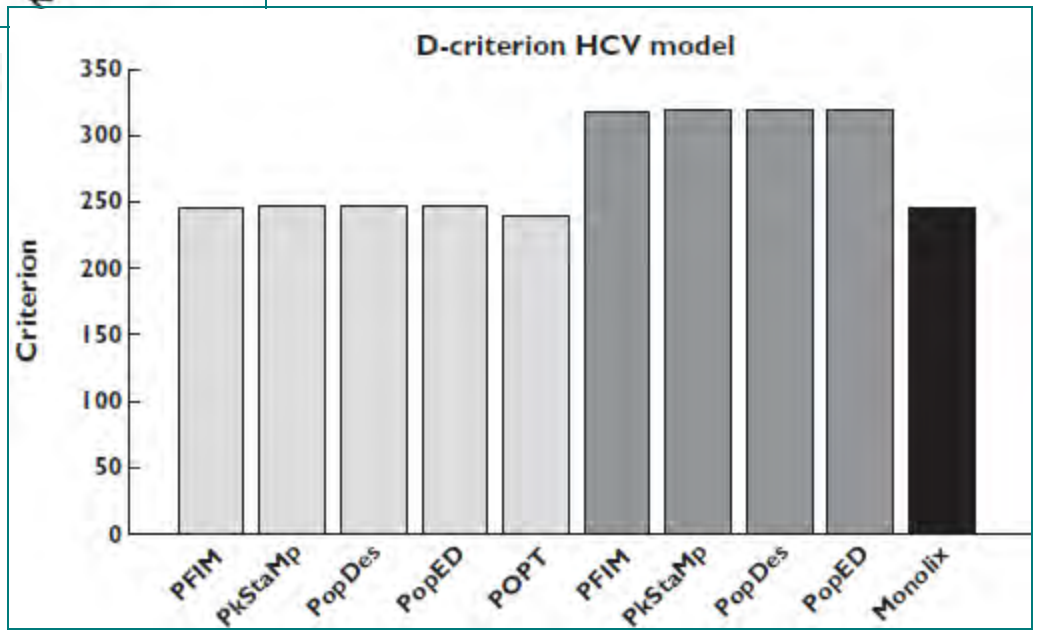
PD: parameters (δ, EC_{50}, n, c), response η_2

Software comparison: warfarin and HCV



(▭) block diagonal, (▭) full and (■) simulated

"reduced"



Nyberg et al. (2015)

Approximation options: Monte Carlo

Generate L "patients" according to standard model (1), $\mathbf{Y}_i = \{y_{ij}\}$

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\eta}}(\mathbf{x}, \boldsymbol{\theta}) = \widehat{\mathbf{E}}_{\boldsymbol{\theta}}\mathbf{Y} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{Y}_i \ ,$$

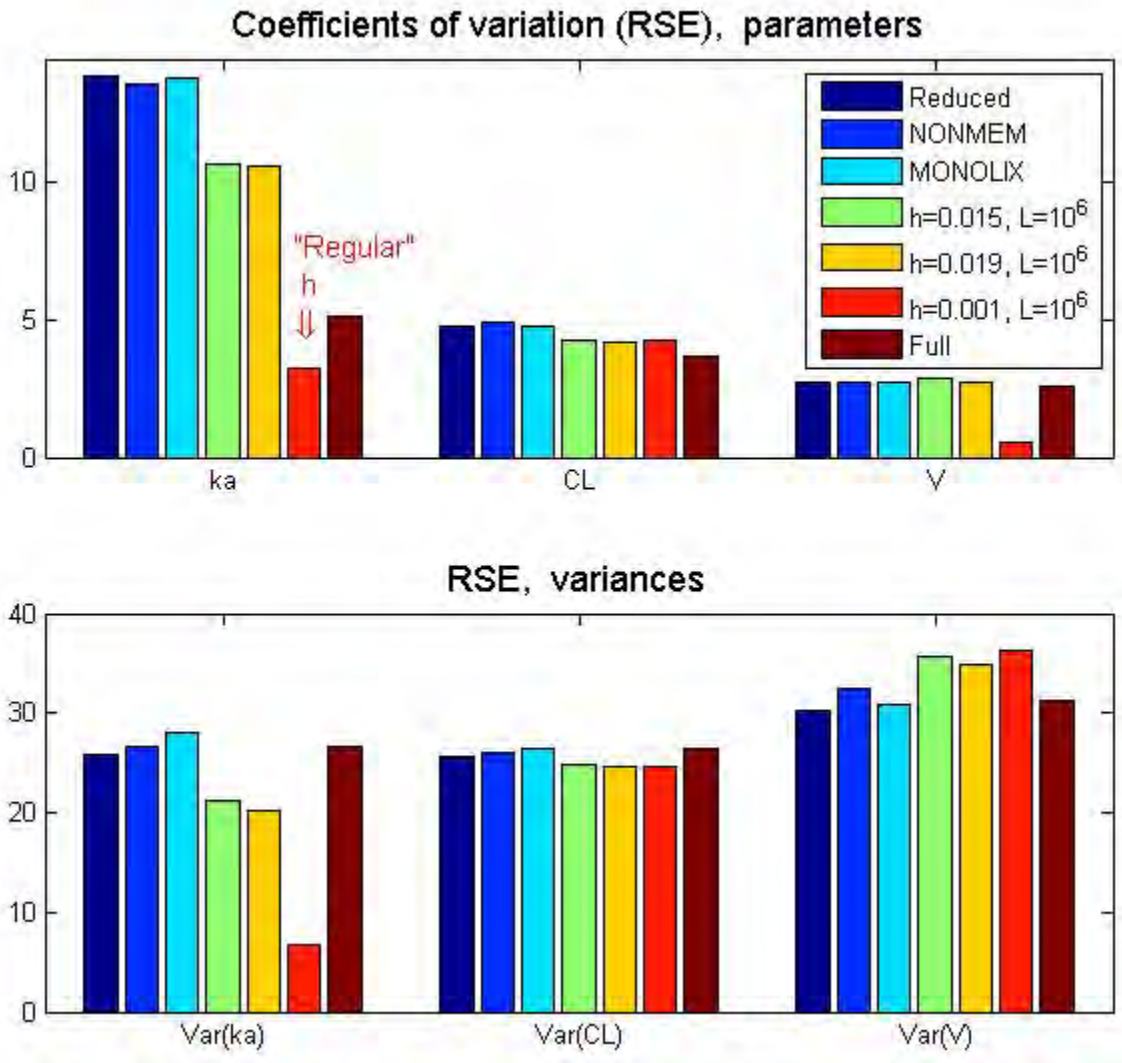
$$\widehat{\mathbf{S}}(\mathbf{x}, \boldsymbol{\theta}) = \widehat{\mathbf{Var}}_{\boldsymbol{\theta}}\mathbf{Y} = \frac{1}{L-1}\sum_{i=1}^{L}[\mathbf{Y}_i - \widehat{\boldsymbol{\eta}}][\mathbf{Y}_i - \widehat{\boldsymbol{\eta}}]^T$$

Use $\widehat{\boldsymbol{\eta}}$, $\widehat{\mathbf{S}}$ in the formula for $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$

1. "Normal"

2. Be careful with numerical differentiation, choice of step size

Approximation options: Monte Carlo (warfarin)



Leonov, Aliev (2013, mODa 10)

Order of differentiation and integration (MC)

To calculate $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$, need derivatives $\mathbf{Z}(\mathbf{x}, \boldsymbol{\theta}) = [\partial(E\mathbf{Y})/\partial\boldsymbol{\theta}]$

Simplest model: $k = 1$ (single response), $m = 1$ (single parameter),

$$y_i(\theta) = \eta(x, \theta_i) + \varepsilon_i,\ \theta_i = \theta + b_i,\ b_i \sim \mathcal{N}(0, \omega^2), \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

L - no. of MC runs; h - step size for numerical differentiation

MC1. Generate two sets of L mutually independent $\{b_i\}$, $\{\varepsilon_i\}$ and define

$$y_i(\theta + h) = \eta(x, \theta_i + h) + \varepsilon_i, \quad y_i(\theta - h) = \eta(\theta_i - h, x) + \varepsilon_i.$$

Calculate

$$\widehat{Z_1}(x, \theta) = \frac{d\widehat{E}(y)}{d\theta} = \frac{\widehat{E}y_{\theta+h} - \widehat{E}y_{\theta-h}}{2h} = \frac{1}{L}\sum_{i=1}^{L} \frac{\eta(x, \theta_i + h) - \eta(x, \theta_i - h)}{2h} =$$

$$= \frac{1}{L}\sum_{i=1}^{L} \widehat{\eta}'_\theta(x, \theta_i) = \widehat{E}\left(\frac{dy}{d\theta}\right).$$

Order of differentiation and integration (MC)

MC2. Generate four sets of $\{b_{i1}\}$, $\{b_{i2}\}$, $\{\varepsilon_{i1}\}$, $\{\varepsilon_{i2}\}$ and define

$$\theta_{i1} = \theta + b_{i1}, \;\; \theta_{i2} = \theta + b_{i2},$$

$$y_i(\theta + h) = \eta(x, \theta_{i1} + h) + \varepsilon_{i1}, \quad y_i(\theta - h) = \eta(x, \theta_{i2} - h) + \varepsilon_{i2}.$$

Calculate

$$\widehat{Z_2}(x,\theta) = \frac{d\widehat{E}(y)}{d\theta} = \frac{\widehat{E}y_{\theta+h} - \widehat{E}y_{\theta-h}}{2h} =$$

$$= \frac{1}{L}\sum_{i=1}^{L} \frac{\eta(x, \theta_{i1} + h) - \eta(x, \theta_{i2} - h)}{2h} + \frac{1}{L}\sum_{i=1}^{L} \frac{\varepsilon_{i1} - \varepsilon_{i2}}{2h}$$

$$\neq \widehat{E}\left(\frac{dy}{d\theta}\right).$$

Linear mixed model $\eta(x, \theta_i) = \theta_i x$: $E_\theta[\eta(x, \theta_i] = \theta x$,

$$\widehat{Z_1}(\beta, x) \equiv x; \;\; \widehat{Z_2}(\beta, x) = x + \frac{1}{2hL}\sum_{i=1}^{L} \left[(\theta_{i1} - \theta_{i2})x + (\varepsilon_{i1} - \varepsilon_{i2})\right].$$

Approximation options (cont.)

- *Mielke* (PODE 2012): likelihood and approximation of conditional moments
  - Full and reduced options as special cases (linearization at different values of $\boldsymbol{\theta}$)
  - Links to
    - Laplace approximation (*Pinheiro and Bates* (1995, 2002)
    - FO/FOCE estimation methods in NONMEM (*Wang*, 2007)
- *Nguyen and Mentré* (2014): approximation via adaptive Gaussian quadratures
- Population optimal design software: mostly focused on *locally optimal designs*

Locally optimal vs. adaptive designs

- Locally optimal designs: need parameter values
- Adaptive optimal designs: reduce dependence on unknown parameters
  - Estimation and design are performed in stages
  - *Box, Hunter* (1965)
  - Need efficient estimation and optimal design tools!

ICON: tools for parameter estimation and design

- NONMEM: estimation for NONlinear Mixed Effects Models
  - Pharma industry standard for estimation in population PK/PD models
  - Under ICON since 2006
- ADDPLAN: fully validated statistical software for design, simulation and analysis of adaptive clinical trials
  - Industry standard, used by many pharma companies
  - Under ICON since 2014 (Innovation Centre)
  - ADDPLAN Classic (Base, MC, PE)
  - ADDPLAN DF (Dose Finding): software for design, simulation and analysis of adaptive dose-finding trials
    - Based on MCP-Mod approach, which was adopted by European regulators in 2014
    - Utilizes both estimation and optimal design modules
    - Dose-response models: linked to PK/PD, thus critical to further develop estimation and optimal design methodology and software

Final comments

- Optimal design: more and more popular within pharma industry
- Adaptive designs: combination of model-based design and estimation techniques
- Adaptive/optimal designs: development of methodology and software
- Many open interesting problems!

References

- Atkinson, A.C., Donev, A. (1992). *Optimum Experimental Design*. Clarendon Press, Oxford.
- Demidenko, E. (2004). *Mixed Models. Theory and Application*. Wiley, Hoboken, NJ.
- Fedorov, V.V. (1972), *Theory of Optimal Experiment*. Academic Press, NY.
- Fedorov, V.V., Leonov, S. (2004). Parameter estimation for models with unknown parameters in variance. *Comm. Statist. Theory Methods*, **33**(11), 2627-2657.
- Fedorov, V.V., Gagnon, R., Leonov, S., Wu, Y. (2007). Optimal design of experiments in pharmaceutical applications. In: Dmitrienko, A., Chuang-Stein, C., D'Agostino, R. (Eds), *Pharmaceutical Statistics Using SAS.* SAS Press, Cary, NC, pp. 151-195.
- Fedorov, V.V., Leonov, S.L. (2013). *Optimal Design for Nonlinear Response Models*. Chapman & Hall/CRC Biostatistics Series, Boca Raton, FL
- Kuhn, E., Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data An.*, 49 (4), 1020-1038.
- Leonov, S., Aliev, A. (2013). Approximation of the Fisher information matrix for nonlinear mixed effects models in population PK/PD studies. In: Ucinski, D., Atkinson, A.C., Patan, M. (Eds), *mODa 10 – Advances in Model-Oriented Design and Analysis.* Springer, Heidelberg, pp.145-152.
- Lestini, J. (2014). Influence of the size of the cohorts in adaptive design for nonlinear mixed effect models: an evaluation by simulation for a pharmacokinetic (PK) and pharmacodynamic (PD) model in oncology. PODE 2014, Basel. http://www.maths.qmul.ac.uk/~bb/PODE/PODE2014_Slides/JLestini_PODE_2014.pdf
- Lindstrom, M.J., Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673-687.
- Magnus, J.R., Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, NY.

References

- Mentré, F., Mallet, A., Baccar, D. (1997). Optimal design in random-effects regression models, *Biometrika*, **84** (2), 429-442.
- Mentré, F., Nyberg, J., Ogungbenro, K., Leonov, S., Aliev, A., Duffull, S., Bazzoli, C., Hooker, A. (2011). Comparison of results of the different software for design and evaluation in population pharmacokinetics and pharmacodynamics. In: *Abstracts of the Annual Meeting of the Population Approach Group in Europe (PAGE).* ISSN 1871-6032, www.page-meeting.org/?abstract=2066
- Mielke, T. (2012). Impact of information approximations on experimental designs in nonlinear mixed effects models. PODE 2012, Paris. http://www.maths.qmul.ac.uk/~bb/PODE/PODE2012_Slides/TMielke.pdf
- Mielke, T., Schwabe, R. (2010). Some considerations on the Fisher information in nonlinear mixed effects models. In: Giovagnoli, A., Atkinson, A.C., Torsney, B. (Eds), *mODa 9 – Advances in Model-Oriented Design and Analysis*. Physica-Verlag/Springer, Berlin, ; pp. 129–136.
- Nguyen, T.T., Mentré, F. (2014). Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature. *Comput. Stat. Data An*, **80**, 57-69.
- Nyberg, J., Bazzoli, C., Ogungbenro, K., Aliev, A., Leonov, S., Duffull, S., Hooker, A., Mentré, F. (2015). Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics. *Brit. J. Clin. Pharmacol.,* **79** (1), 6-17.
- Pinheiro, J.C., Bates, D,M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comp. Graph. Stat.*, **4**(1), 12-35.
- Pinheiro, J.C., Bates, D,M. (2002). *Mixed Effects Models in S and S-Plus*. Springer, New York, 2002.
- Retout, S., Mentré, F. (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *J. Biopharm. Stat*., **13** (2), 209-227.
- Wang, Y. (2007). Derivation of various NONMEM estimation methods. *J. Pharmacokin. Pharmacodyn.*, **34**(5), 575- 593.