Design of population PK/PD studies: approximation of the individual Fisher information matrix

Sergei Leonov

(fmr GlaxoSmithKline, ftr Vertex)

DAE02: Optimum Design for Mixed Effects Models

August 12, 2011
Outline

- PODE workshops
- Population optimal design software tools
  - Types of problems they address
  - Comparison, whether their outputs match
- Approximation options for Fisher information matrix
Optimal design for population PK/PD models

- PODE Workshop created in 2006
  Population Optimum Design of Experiments
  - Theory of optimal experimental design for nonlinear mixed effects models and its applications in drug development
- Discussion of population optimal design software started in 2007, continued in 2008-2011
Population optimal design software

Five tools available

- PFIM (developed in INSERM, Università Paris 7, France)
- PkStaMp (GlaxoSmithKline, Collegeville, U.S.A.)
- PopDes (CAPKR, University of Manchester, UK)
- PopED (Uppsala University, Sweden)
- WinPOPT (University of Otago, New Zealand)

Main application areas:
  pharmacokinetics (PK) and pharmacodynamics (PD)
Comparison of population design tools

- Key: Fisher information matrix of a properly defined single observational unit (individual patient)

- Calculation of individual matrix $\mu(x, \theta)$
  - Identical under the same assumptions for all tools (benchmark examples)

- Differences: selection of sampling sequences, algorithmic details, libraries of models, types of approximation....

- More from France Mentré
Pharmacokinetics (PK)

- PK: how drug propagates in patient's body
  - Dose $\rightarrow$ concentration
- PK studies: at different phases of drug development
- Models:
  - Compartmental, systems of ODE
  - Non-compartmental (AUC, $T_{\text{max}}$, $C_{\text{max}}$)

Example:
One-compartment model, 1$^{\text{st}}$ order absorption and linear elimination

$$\begin{cases} \dot{f}_0(t) = -k_\alpha f_0(t) \\ \dot{f}_1(t) = k_\alpha f_0(t) - k_e f_1(t) \end{cases}$$

$$f_0(t_i) = f_0(t_i - 0) + D_i, \; f_0(0) = D_0, \; f_1(0) = 0.$$
Pharmacodynamics (PD)

- PK: what body does to the drug
- PD: what drug does to the body, progression of clinical endpoint (concentration $\to$ effect)
  - Drop in blood pressure for hypertensive patients
  - Reduction in the number of “bad” cells
  - Tumor shrinkage
- Popular PD model: $E_{\text{max}}$ (sigmoidal-shaped curve, multi-parameter logistic model)
Design of population PK/PD studies

- What we select/optimize (control):
  - Location of sampling times
  - Number of sampling times per patient
  - Number of patients enrolled

- Optimal population designs:
  - Optimal: with respect to precision of parameter estimates
  - Goal: *find the most informative sampling times*
Nonlinear models, multiple responses

- Predictor \( \mathbf{x} = (x_1, x_2, \ldots, x_k) \) - sequence of sampling times,
- Measurements \( \mathbf{Y} = [y(x_1), \ldots, y(x_k)] \) - vector,
- Response \( \eta(x, \theta) = [\eta(x_1, \theta), \ldots, \eta(x_k, \theta)] \) - vector

Key: \( \mu(x, \theta) \) - information matrix of a \( k \)-dimensional sequence \( x \)
Optimal designs

Information matrix: \( n_i \) patients on seq. \( x_i \) \( \implies M_N(\theta) = \sum_{i=1}^{N} n_i \mu(x_i, \theta) \)

Variance of the MLE: \( \text{Var}(\hat{\theta}) \approx M_N^{-1}(\theta) \)

\[
M(\xi, \theta) = \frac{M_N(\theta)}{N} = \sum_i w_i \mu(x_i, \theta) \quad - \text{normalized information, per observation}
\]

\( \xi = \{w_i, x_i\} \) - normalized design; \( w_i = n_i/N \) - weights

\[
D(\xi, \theta) = M^{-1}(\xi, \theta) \quad - \text{normalized variance-covariance matrix}
\]
Optimal designs (cont.)

Criterion of optimality \( \Psi[D(\xi, \theta)] \rightarrow \min_{\xi} : \) minimization wrt

- weights \( w_i, 0 \leq w_i \leq 1, \sum_i w_i = 1 \) (continuous designs)

- admissible sampling sequences \( x_i \in X \) - design region.

Locally D-optimal designs: \( \Psi = |D(\xi, \theta)| \)

Equivalence Theorem: Kiefer, Wolfowitz (1960), Fedorov (1972) -

Backward step: Atwood (1973)
Optimality criteria, ellipse \((\theta - \theta^*)^\top D^{-1} (\theta - \theta^*) \leq 1\)

**D-criterion:** \(|D| = \lambda_1 \cdot \lambda_2 = (OA \cdot OB)^2\); area \((V) = \pi (\lambda_1 \cdot \lambda_2)^{1/2}\)

**E-criterion:** \(\lambda_1 = (OA)^2\)

**A-criterion:** \(\text{tr } D = \lambda_1 + \lambda_2 = (OC)^2 = D_{11} + D_{22}\)
Mixed effects model model

- $\gamma$ - response parameters (rate constants)

- $\gamma_i$ - parameters of patient $i$ (sampled from population):
  
  normal, $\gamma_i \sim N(\gamma^0, \Omega)$, or log-normal ($\gamma^0$ - “typical values”)

- Data $y(x_{ij}) = \eta(x_{ij}, \gamma_i) \left[ 1 + \varepsilon^p_{ij} \right] + \varepsilon^a_{ij}, \quad j = 1, \ldots, k_i$. (1)

  $\varepsilon^a_{ij} \sim N(0, \sigma^2_a)$, $\varepsilon^p_{ij} \sim N(0, \sigma^2_p)$

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

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

$$
\eta(x, \gamma) = \frac{Dk_a}{V(k_a - k_e)} \left( e^{-k_ex} - e^{-k_ax} \right), \quad \gamma = (k_a, k_e, V)^T
$$
Information matrix for sequence $\mathbf{x}$

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

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

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

$\mathbf{S} = \mathbf{S}(\mathbf{x}, \theta), \quad \eta = \eta(\mathbf{x}, \theta)$ [Muirhead (1982), Magnus and Neudecker (1988)]

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

$$
\mathbf{S}(\mathbf{x}, \theta) \simeq \mathbf{F} \Omega \mathbf{F}^T + \sigma_p^2 \text{Diag}[\eta(\mathbf{x}, \theta) \eta^T(\mathbf{x}, \theta) + \mathbf{F} \Omega \mathbf{F}^T] + \sigma_A^2 \mathbf{I}_k,
$$

$$
\mathbf{F} = \mathbf{F}(\mathbf{x}, \gamma^0) = \left. \left[ \frac{\partial \eta(\mathbf{x}, \theta)}{\partial \gamma_\alpha} \right] \right|_{\gamma = \gamma^0} - (k \times m_\gamma) \text{ matrix}
$$

Retout, Mentré (2003), Gagnon, Leonov (2005)
Design region \( X \)

**PkStaMp: Sampling Times Allocation (STand-Alone Application), Matlab Platform**

Selection of sampling sequences:

- Option 1: specify
  - All candidate times \((x_1, x_2, \ldots, x_K)\)
  - Number of sampling times per patient \(k \in [k_{\text{min}}, k_{\text{max}}]\)
  - Lag between samples: \(x_{i,j+1} - x_{i,j} \geq \Delta\)

- Option 2: pre-specify an arbitrary set of candidate sequences in a file

\[
\text{Design region } X = \{ x_i = (x_{i,1}, \ldots, x_{i,k_i}) \}
\]
Typical screen: one-compartment, 1st order absorption
More complex setting: cost-based designs

Measurements at $x_i$ associated with cost $c(x_i)$,

$$\sum_i n_i c(x_i) \leq C \implies M_C(\theta) = \sum_i \frac{n_i}{C} \mu(x_i, \theta) = \sum_i \tilde{w}_i \tilde{\mu}(x_i, \theta),$$

Information matrix normalized by total cost $C$,

$$\tilde{w}_i = \frac{n_i c(x_i)}{C}; \quad \tilde{\mu}(x_i, \theta) = \frac{\mu(x_i, \theta)}{c(x_i)} \implies \text{same framework}, \quad \text{same algorithms}$$


In PkStaMp: (a) Cost $c(x)$ proportional to # of samples in sequence $x$, or (b) Entered by user for each candidate sampling sequence
More complex models: nonlinear kinetics

Two-compartment model, 1\textsuperscript{st} order absorption, Michaelis-Menten elimination: no analytical solution (ODE solver)

\[
\begin{align*}
\dot{f}_0(t) &= -k_a f_0(t) \\
\dot{f}_1(t) &= k_a f_0(t) - (k_{12} + k_e) f_1(t) - \frac{(V_m/V)f_1(t)}{k_m + f_1(t)/V} + k_{21} f_2(t) \\
\dot{f}_2(t) &= k_{12} f_1(t) - k_{21} f_2(t),
\end{align*}
\]

Dose

\[\begin{array}{c}
0 \\
\downarrow k_3 \\
\downarrow k_{10} \\
\downarrow k_M, V_M
\end{array}\]

\[\begin{array}{c}
1 \\
V_1 \\
\downarrow k_{12} \\
\downarrow k_{21}
\end{array}\]

\[\begin{array}{c}
2 \\
V_2
\end{array}\]
More complex models: combined PK/PD model

One-compartment PK and Emax PD model

Final PK/PD Model

\[ E = E_o \cdot \left( 1 - \frac{C_p}{IC_{50} + C_p} \right) \]

- \( k_a \): first-order absorption rate constant (h^{-1})
- \( V/F \): apparent volume of distribution (L)
- \( CL/F \): apparent systemic clearance (L/h)
- \( E_o \): PD endpoint at baseline (nM/min/mL)
- \( IC_{50} \): Drug X plasma concentration causing 50% inhibition of PD endpoint (ng/mL)

PK and PD compartments measured, in general, at different times
Another benchmark test: HCV

Proposed by France Mentré, Spring 2011: combination drug for treating chronic hepatitis C (HCV) infection

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

**PK:** parameters \((k_a, k_e, V_1)\), response \(\eta_1\) (continuous infusion term \(r(t)\))

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

\(g_1(t)\) - “target cells”, \(g_2(t)\) - infected cells, \(g_3(t)\) - viral particles (load)

**PD:** parameters \((\delta, EC_{50}, n, c)\), response \(\eta_2\)
HCV example: user-defined option

Design to be tested
- 12 sampling times for each patient
- Same times for PK and PD endpoints

Parameterization
- Log-parameters
- Normal population distribution
- Diagonal population covariance matrix

User-defined option and last 2+ years of PkStaMp development: collaboration with Dr. Alexander Aliev (Institute for Systems Analysis, Russian Academy of Sciences, Moscow)
- “Arbitrary” system of ODE, and/or
- “Arbitrary” closed-form solution
- “Arbitrary” number of compartments
Model specification

User model definition:

Short name: HCV Inf Log

Model description:
Combined PK (1st order absorption/Inf) and viral dynamics PD, log-parameters

Output (measured compartments):
Central: {\(A(2)/\exp(P(3))\)}
Viral load: \(\log_{10}(A(5))\)

Compartments

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Right-hand side in ODE</th>
<th>Administering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Depot</td>
<td>-exp(P(1))*A(1)</td>
<td>Doses &amp; Infusi</td>
</tr>
<tr>
<td>2</td>
<td>Central</td>
<td>exp(P(1))*A(1) - exp(P... None)</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>Target cell</td>
<td>20000 - 1e-7*A(3)*A(5)</td>
<td>Doses</td>
</tr>
<tr>
<td>4</td>
<td>Infected</td>
<td>1e-7*A(3)*A(5) - exp(P... Doses</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Viral load</td>
<td>100*(1-(A(2)/exp(P(3)))... Doses</td>
<td></td>
</tr>
</tbody>
</table>

Compartment 2 properties

Name: Central

Right-hand side of differential equation:
\[ \exp(P(1)) * A(1) - \exp(P(2)) * A(2) \]

Measured output:
\[ A(2)/\exp(P(3)) \]
Parameters, dosing (1), sampling (2)
PODE 2009-2010 comparison

Goal: compare FIM for a particular model/sampling sequence

Model: one-compartment, 1st order absorption, single dose $D = 70$ mg

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

Individual parameters $\gamma_i = \gamma^0 e^{\xi_i}, \ \xi_i \sim \mathcal{N}(0, \Omega)$

$\gamma^0 = (1, 0.15, 8), \ \Omega = \text{diag}(0.6, 0.07, 0.02)$

Measurements: $y_{ij} = \eta(\gamma_i, x_{ij}) [1 + \varepsilon_{M,ij}]$

$\{x_{ij}\} \equiv x = (0.5, 1, 2, 6, 24, 36, 72, 120)$ hours

$\varepsilon_{M,ij} \sim \mathcal{N}(0, \sigma^2_M), \ \sigma^2_M = 0.01; \ i = 1, \ldots, 32; \ j = 1, \ldots, 8$

Combined parameter $\theta = (k_a^0, CL^0, V^0; \ \omega^2_{k_a}, \omega^2_{CL}, \omega^2_V; \ \sigma^2_M)$
Information matrix $\mu(x, \theta)$: block form, Retout and Mentré (2003)

$$
\mu = \begin{pmatrix} A & C \\ C^T & B \end{pmatrix},
$$

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

$\mu(x, \theta)$ - information matrix of a single ($k$-dimensional) sequence $x$:

$$
\mu_{\alpha\beta}(x, \theta) = \frac{\partial \eta}{\partial \theta_\alpha} S^{-1} \frac{\partial \eta}{\partial \theta_\beta} + \frac{1}{2} \text{tr} \left[ S^{-1} \frac{\partial S}{\partial \theta_\alpha} S^{-1} \frac{\partial S}{\partial \theta_\beta} \right],
$$
PODE 2009-2010 comparison (cont.)

- Compared $Var_a = [\mu(x, \theta)]^{-1}$ produced by different tools: identical results under the same assumptions
- Compared $Var_a$ and $Var_e$: empirical variance-covariance matrix (Monte Carlo + estimation in NONMEM and Monolix):
  - If block $C$ “excluded” ($C = 0$), and 2nd term in $A$ removed, then analytical results (1st order approximation) and $Var_e$ are very close
  - If block $C$ and 2nd term in $A$ are both kept, then there is a visible difference for some elements of $Var$
Approximation options

Individual parameters, log-normal distribution:

\[ \gamma_i = e^{\xi_i}, \quad \xi_i \sim \mathcal{N}(0, \Omega), \]

- 1st-order approximation, \( \mathbb{E}\xi_i = 0, \quad \text{Var}(\xi_i) = V \implies \)

\[ \mathbb{E}_\xi(e^{\xi_i}) \approx 1, \quad \text{Var}_\xi(e^{\xi_i}) \approx V \]

- Exact moments: \( \mathbb{E}_\xi(e^{\xi_i}) = e^{V/2}, \quad \text{Var}_\xi(e^{\xi_i}) = e^V(e^V - 1) \).

- \( V = 0.6 \implies \mathbb{E}_{1st} = 1, \quad E_{exact} = 1.35; \quad \text{Var}_{1st} = 0.6, \quad \text{Var}_{exact} = 1.50 \)
Approximation options (cont.)

2nd - order approximation for mean/variance

\[ E_\theta[\eta(x, \gamma)] \approx \eta(x, \gamma^0) + \frac{1}{2} \text{tr} [H(\gamma^0)\Omega] , \]

\[ H(\gamma^0) = \left[ \frac{\partial^2 \eta(x, \gamma)}{\partial \gamma_\alpha \partial \gamma_\beta} \right]_{\gamma=\gamma^0} \]

Numerically may be rather tedious

- All derivatives calculated numerically (central differences)
- Derivatives of variance \( S \) require second derivatives of \( \eta \)
- With 2nd order approximation: fourth derivatives…..
Approximation options (cont.)

Calculation of mean/variance via Monte Carlo:

\[ \hat{\eta}(x_j) = \hat{E}_\theta(y_{ij}) = \frac{1}{N} \sum_{i=1}^{N} y(x_{ij}) , \]

\[ \hat{S}(x_j) = \hat{\text{Var}}_\theta(y_{ij}) = \frac{1}{N} \sum_{i=1}^{N} [y(x_{ij}) - \hat{\eta}(x_j)]^2 \]

Numerically straightforward: OK if normal approximation is “reasonable”


Approximation options  (cont.)

Mean response curves for one-compartment model example

- **Solid** - 1st order approximation
- **Dashed** - computed at mean values of log-normal distribution,
- **Dotted** - Monte Carlo average
Approximation options

- Measures of nonlinearity:
  - Curvature measures, intrinsic vs parameter effects

- Simulation studies for PK/PD, Merlé and Tod (2001)
  - Criteria values may be substantially affected by linearization
  - Designs and relative efficiencies are often not

- PODE 2009-2011 comparison/simulation studies:
  - Linearization (1st order) very crude, but performed reasonably well without block C (?)
Optimal design for PK/PD


- There is a rich related literature, mostly non-Bayesian, on design for complex PK and biological models... With a few exceptions, this important work is not in the mainstream statistics literature...
Concluding remarks

Goals of population optimal design

- Find most informative sampling times
- Validate the quality of standard/alternative designs (optimal design as a reference/benchmarking)
- Test robustness of optimal designs (sampling windows)
- Reduce number of samples with “minimal” loss of precision
  - Example: from 16 sampling times – to 8 most informative
    D-efficiency \( \text{Eff} = \left( \frac{|M(\xi_8^*)|}{|M(\xi_{16})|} \right)^{1/m} = 0.85 \) (only 15% lost)

- May incorporate costs

Approximation options?
References


References