STATISTICAL MODELING AND APPLICATIONS OF PARTICLE SWARM OPTIMIZATION

Seongho Kim, Ph.D

Biostatistics Core, Karmanos Cancer Institute

Department of Oncology, School of Medicine

Wayne State University, Detroit, MI

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability
- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
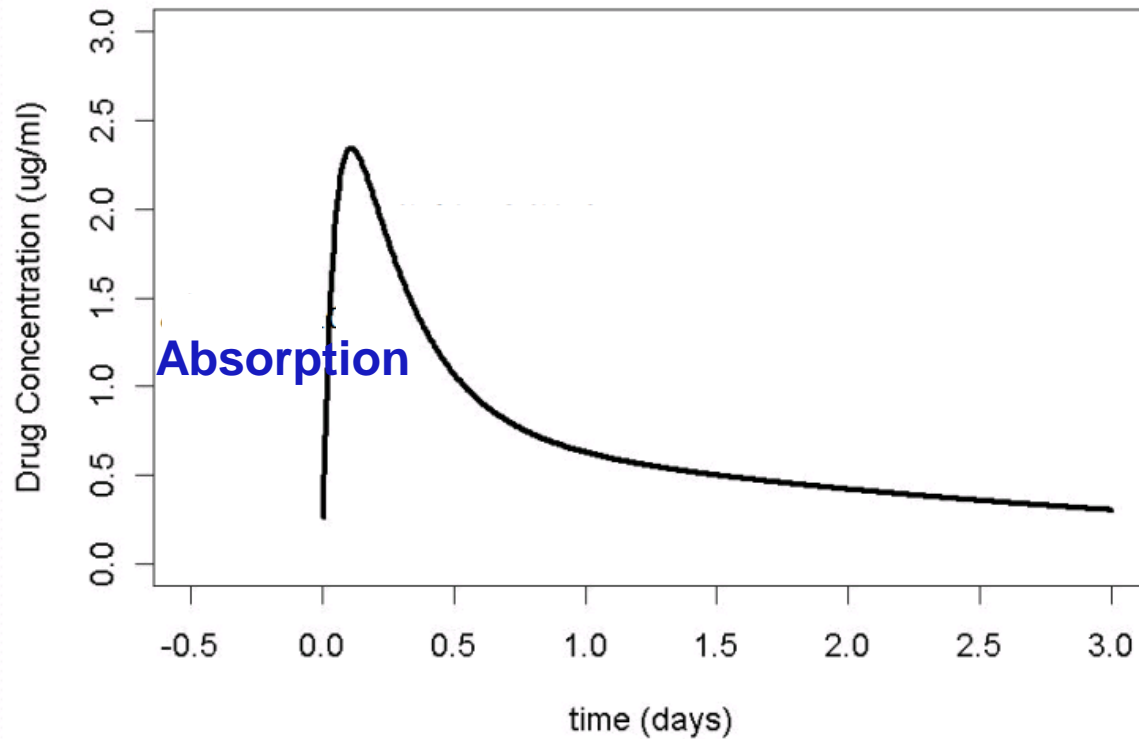 - Adaptive designs with three target response rates

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability
- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage designs
  - Lin and Shih's adaptive designs
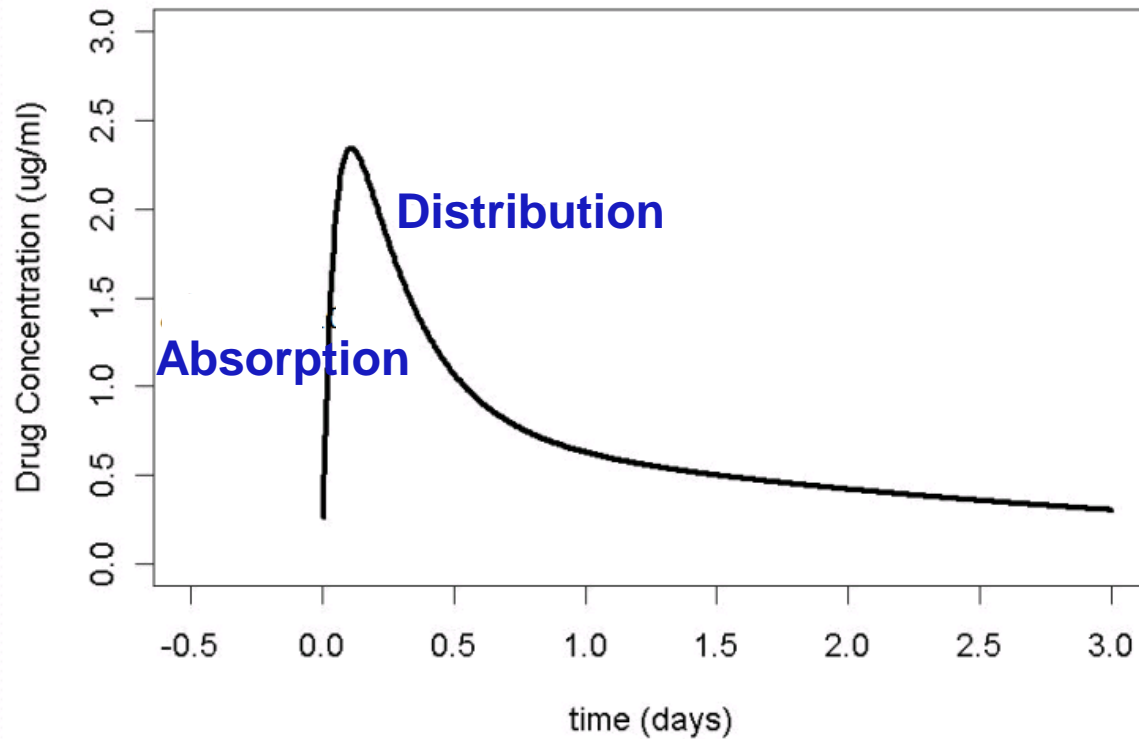 - Adaptive designs with three target response rates
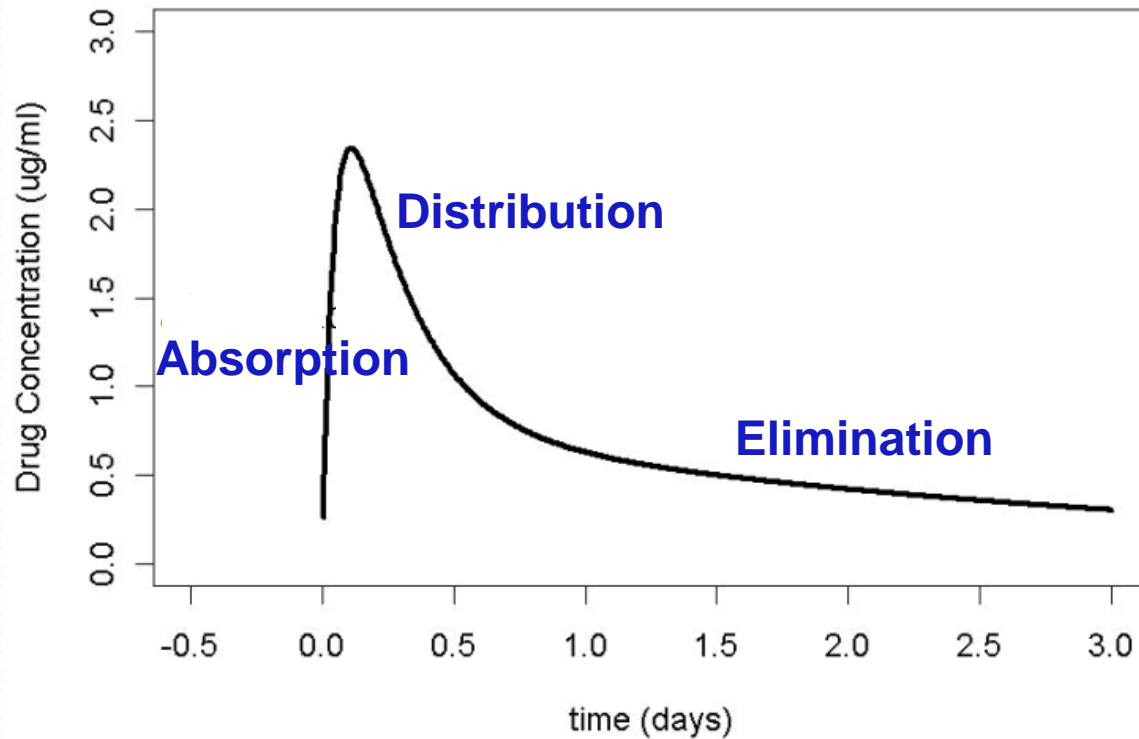
Pharmacokinetics (PK)
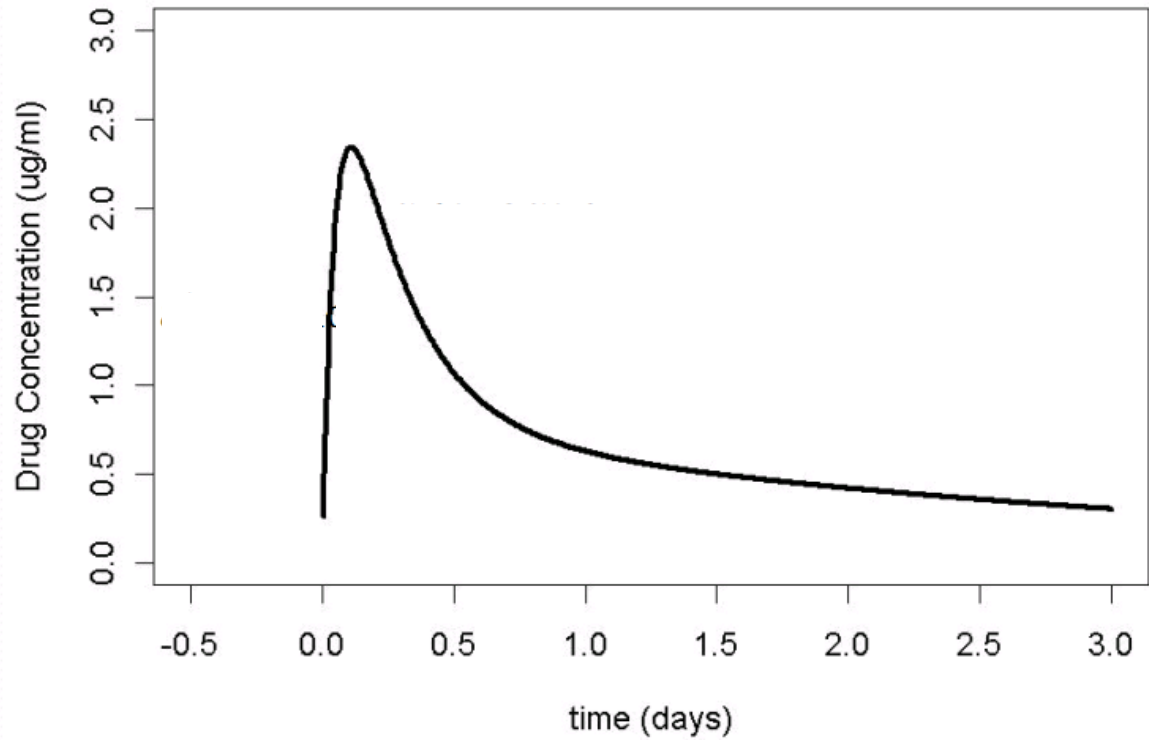
Pharmacokinetics (PK)

Pharmacokinetics (PK)

Pharmacokinetics (PK)

Pharmacokinetics (PK)
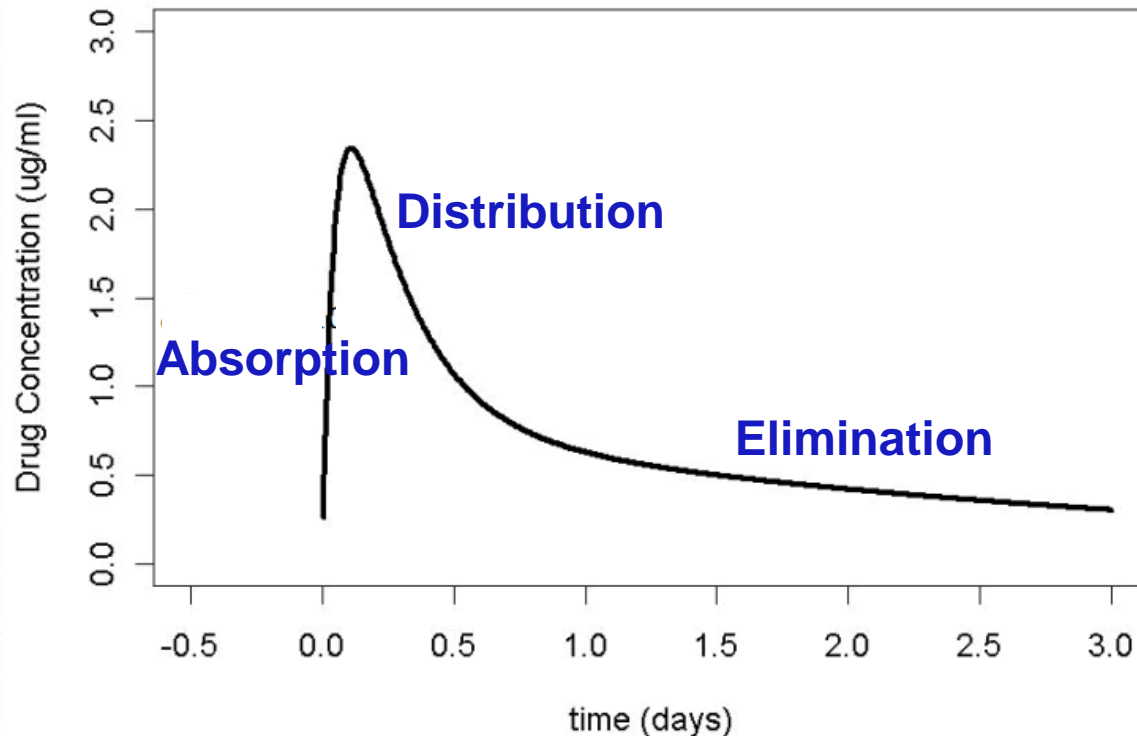
Pharmacokinetic (PK) analysis is to estimate the rates of the absorption, the distribution and the elimination (metabolism and excretion) over time of a drug and its metabolites

Pharmacokinetics (PK) Goal: Therapeutic Optimization

- Achieve concentration profile attaining Efficacy and avoiding Toxicity

How much? How often? How long?: dose regimen
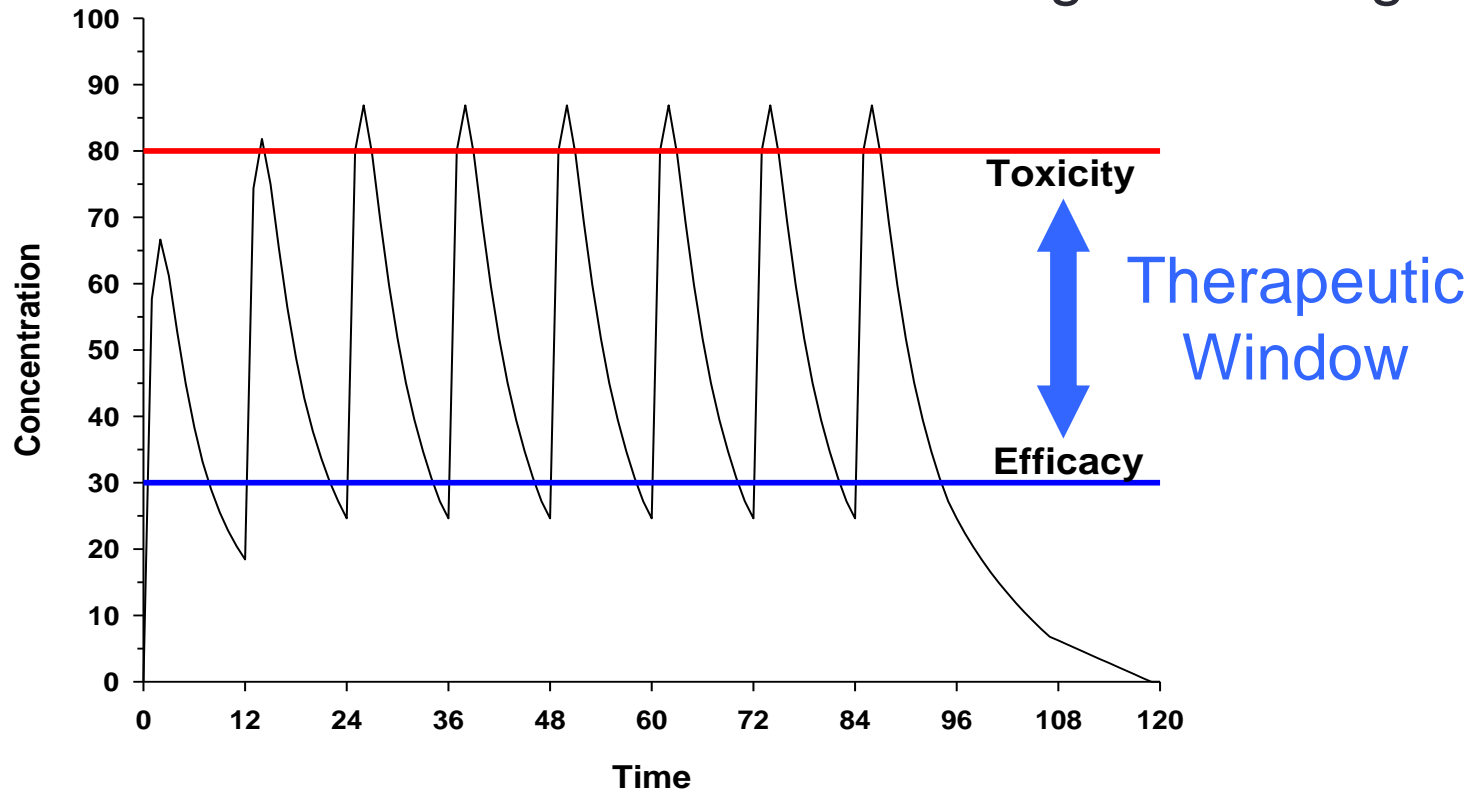
Pharmacokinetics (PK) Study
: Two-Compartment Model

Pharmacokinetics (PK) Study
: Two-Compartment Model
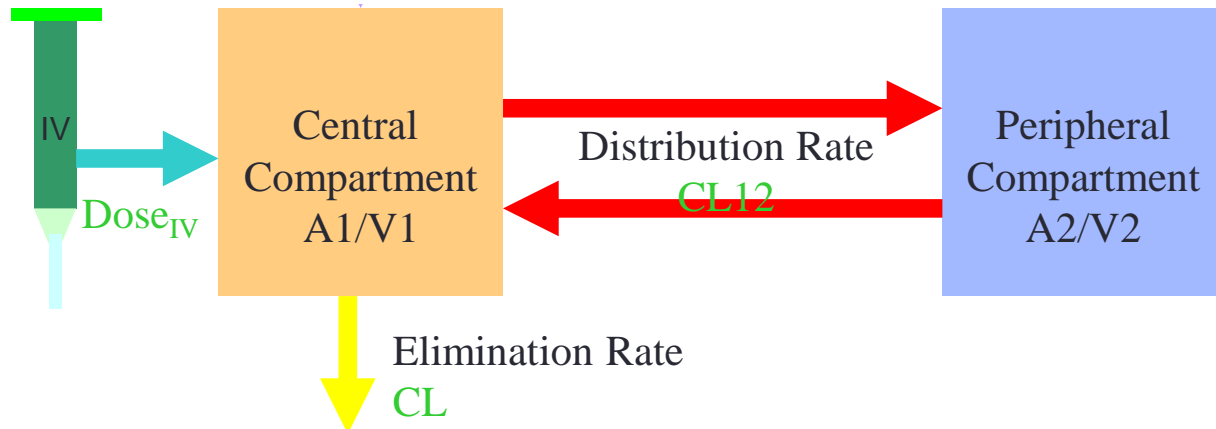
- IV model

Pharmacokinetics (PK) Study : Two-Compartment Model

- IV model

$$\frac{dA_1}{dt} = -CL \times \frac{A_1}{V_1} + CL_{12} \times (\frac{A_2}{V_2} - \frac{A_1}{V_1})$$

$$\frac{dA_2}{dt} = CL_{12} \times (-\frac{A_2}{V_2} + \frac{A_1}{V_1})$$
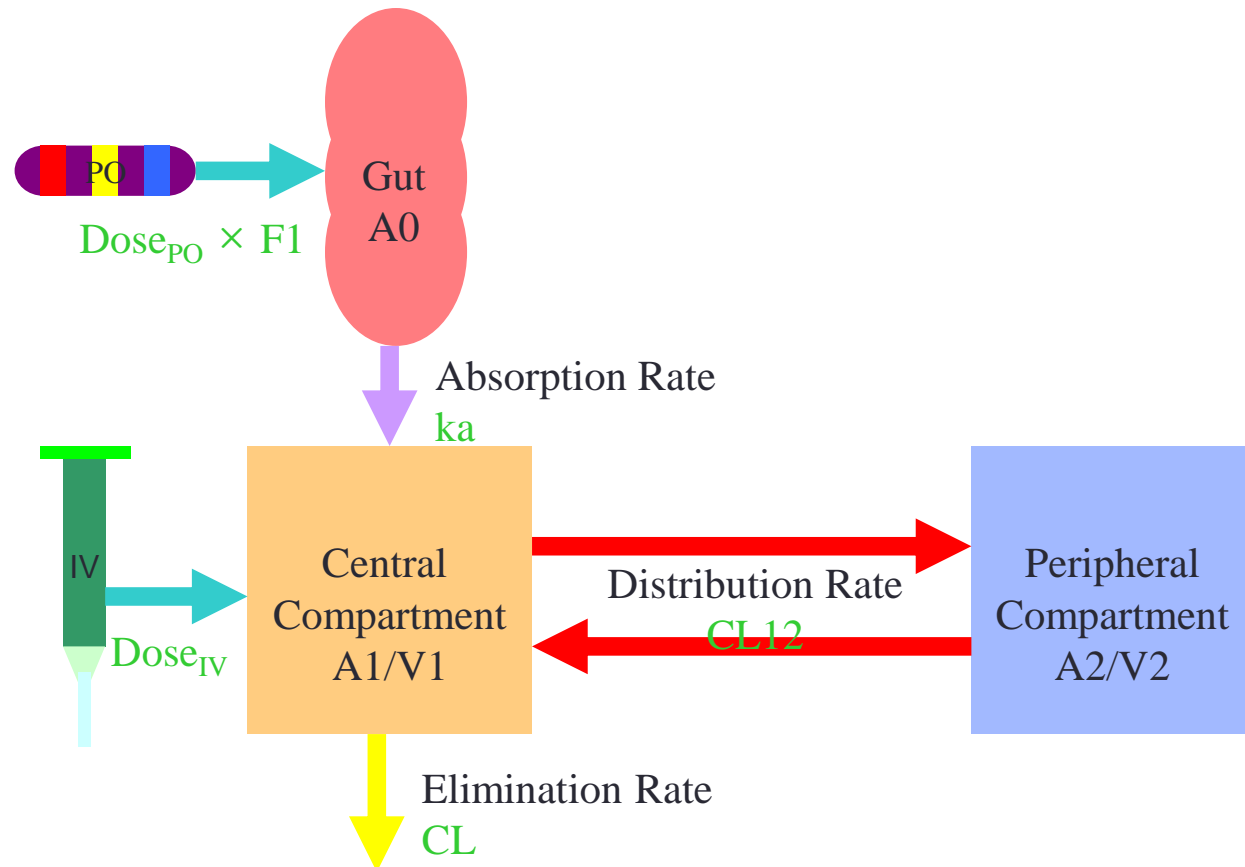
$$[A_1(t), A_2(t)]\,|_{t=0} = (Dose, 0)$$

A_1: Amount of drug in central compartment
A_2: Amount of drug in peripheral compartment

V_1: Volume of distribution in the central compartment
V_2: Volume of distribution in the peripheral compartment

CL: Clearance
CL_{12}: Distribution rate constant

Pharmacokinetics (PK) : Two-Compartment M

- IV model

$$\frac{dA_1}{dt} = -CL \times \frac{A_1}{V_1} + CL_{12} \times (\frac{A_2}{V_2} - \frac{A_1}{V_1})$$

$$\frac{dA_2}{dt} = CL_{12} \times (-\frac{A_2}{V_2} + \frac{A_1}{V_1})$$

$$[A_1(t), A_2(t)]\,|_{t=0} = (Dose, 0)$$

Central Compartment

Peripheral Compartment

CL_{12}: Distribution rate constant

IV

$Dose_{IV}$

Central Compartment A1/V1

Distribution Rate CL12

Peripheral Compartment A2/V2

Elimination Rate CL

Pharmacokinetics (PK)
: Two-Compartment M

- IV model

$$\frac{dA_1}{dt} = -CL \times \frac{A_1}{V_1} + CL_{12} \times (\frac{A_2}{V_2} - \frac{A_1}{V_1})$$

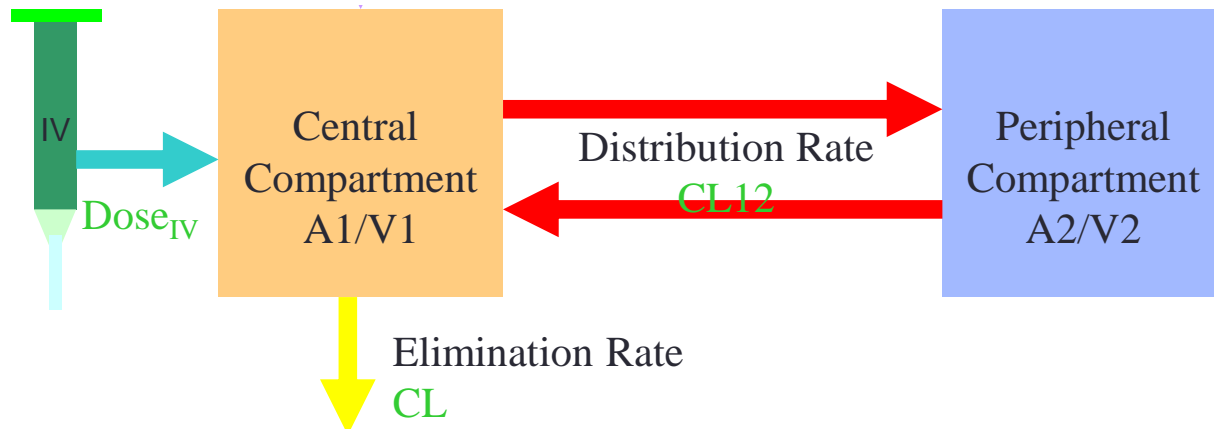$$\frac{dA_2}{dt} = CL_{12} \times (-\frac{A_2}{V_2} + \frac{A_1}{V_1})$$

$$[A_1(t), A_2(t)]\,|_{t=0} = (Dose, 0)$$

Central Compartment

Peripheral Compartment

Concentration

Time

Concentration

Time

CL_{12}: Distribution rate constant

IV

$Dose_{IV}$

Central Compartment A1/V1

Distribution Rate CL12

Peripheral Compartment A2/V2

Elimination Rate CL

Midazolam (MDZ) data: IV infusion
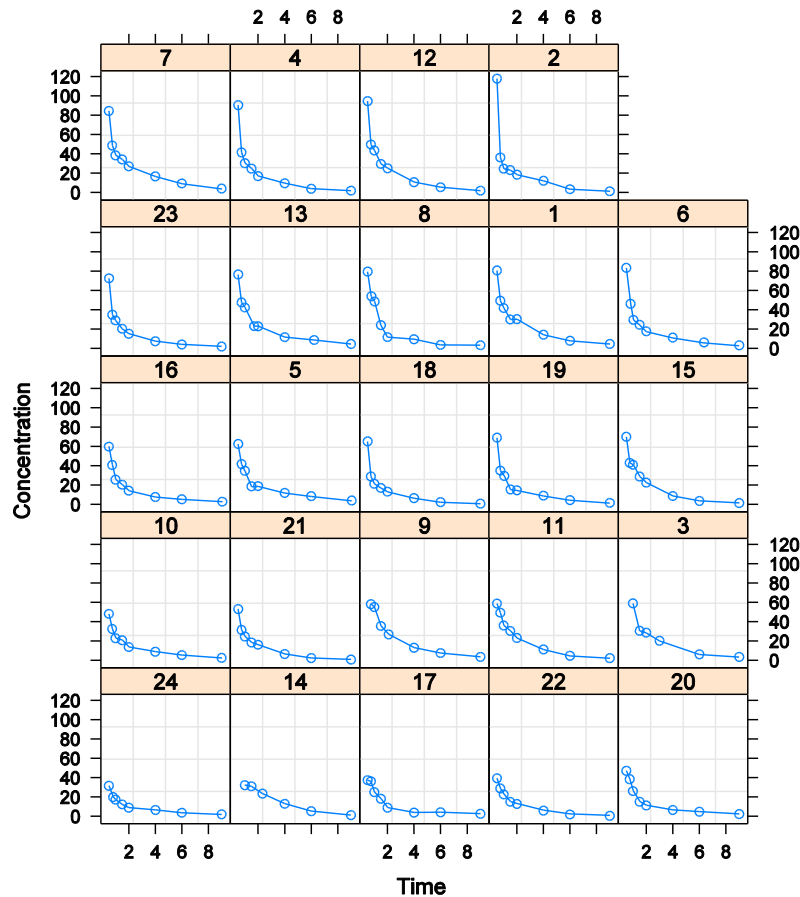
CYP3A is responsible for 60% of drugs' metabolism

CYP3A Substrates → standard CYP3A probe drug

24 volunteers, 18 to 55 years of age, received single dose of 2.74~4.80 mg MDZ intravenously (IV) and Blood samples were collected at 0.5, 0.75, 1, 1.5, 2, 4, 6, and 9 hours after IV MDZ dosing.

Mixed-Effects Model

- Provide a powerful and flexible tool for the analysis of balanced and unbalanced grouped data.

- A mixture of fixed and random factors

  - Fixed-effect (Population-level): the effects of the levels to one another
  - Random-effect (Subject-level): a random sample from a population of effects

Parameter estimation

- Estimation: EM-like Algorithms and Monte Carlo-based Algorithms

- EM Algorithm ➔
  - It can guarantee only up to local optima
  - Approximation needed if either E-step or M-step is intractable (no closed form available)
    - PK/PD models have the nonlinear differential equations
  - It gives us point estimates

- Monte Carlo Algorithm
  - It can guarantee global optima theoretically
  - It can deal with nonlinear functions
  - It can estimate the distribution of parameters (Bayesian approach)

Monte Carlo

- Given: a domain X and a distribution $p(x)$

Monte Carlo

- Given: a domain X and a distribution $p(x)$
- Draw a set of N samples independently
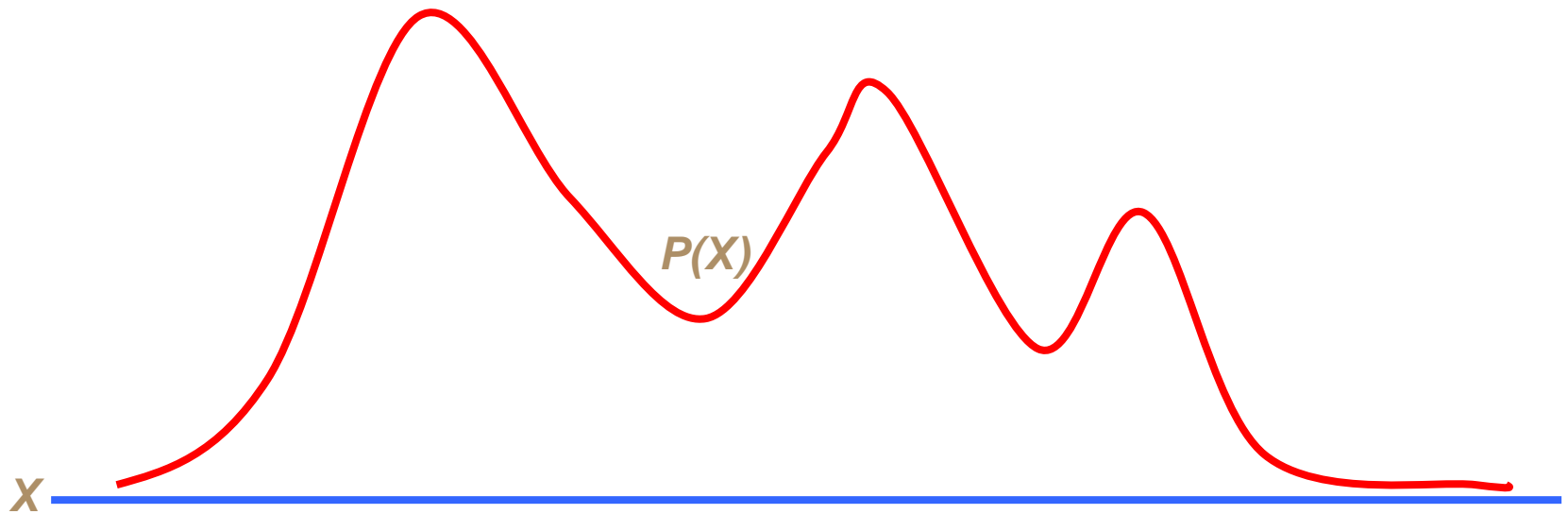
P(X)

X

Monte Carlo

- Given: a domain X and a distribution $p(x)$
- Draw a set of N samples independently
- Approximate the distribution by these samples

$P(X)$

X

Monte Carlo

- Given: a domain X and a distribution $p(x)$
- Draw a set of N samples independently
- Approximate the distribution by these samples

$$p_N(x) = \frac{1}{N} \sum_{i=1}^{N} 1(x^{(i)} \in X) \rightarrow p(x) \text{ as } N \rightarrow \infty$$

P(X)

X
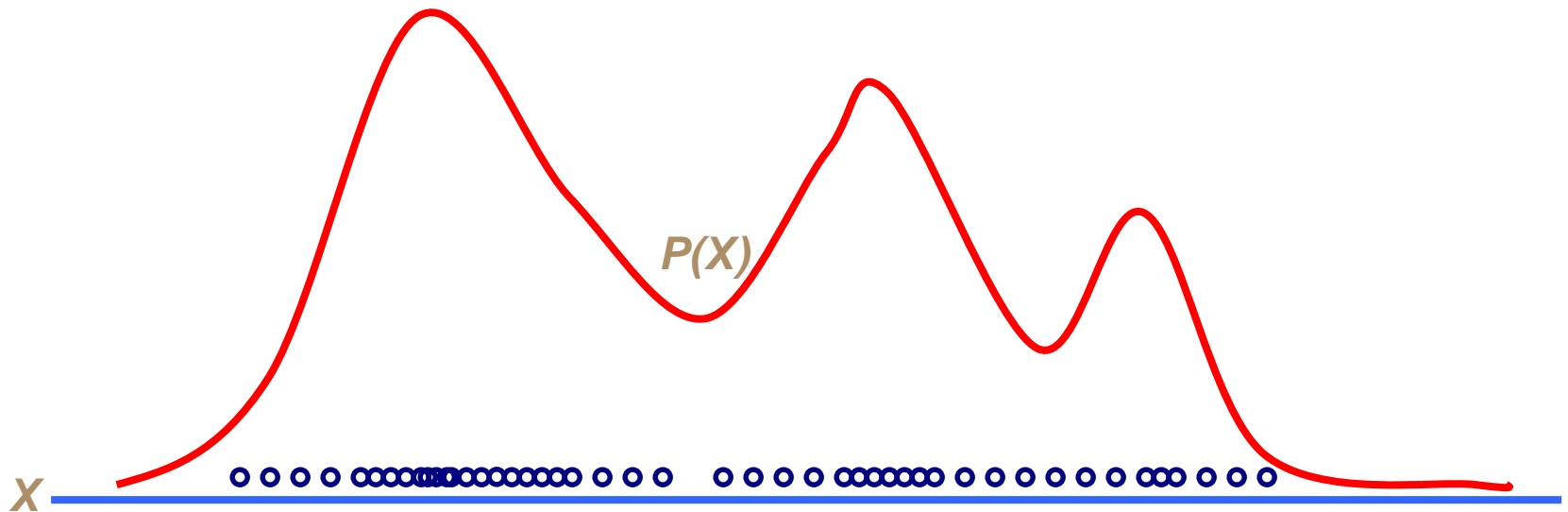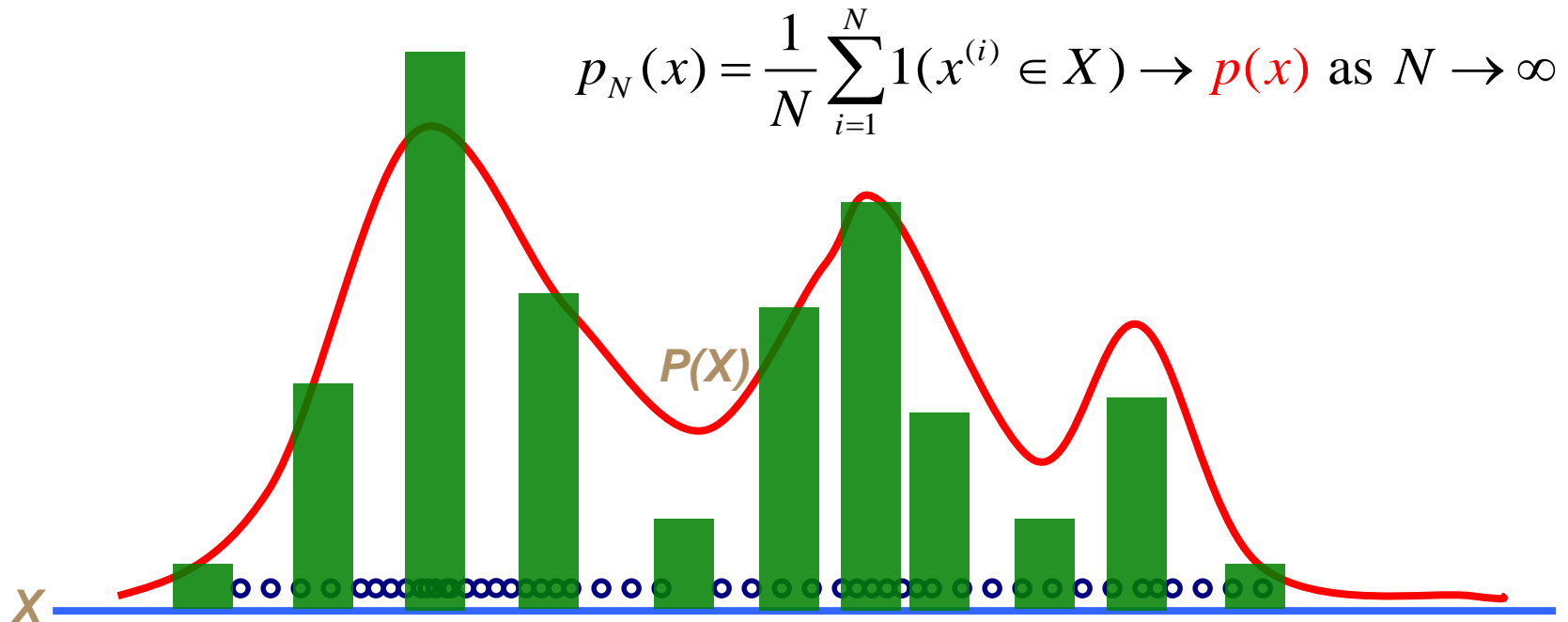
Monte Carlo

- Given: a domain *X* and a distribution *p(x)*
- Draw a set of N samples independently
- Approximate the distribution by these samples

These samples can be used

$\Rightarrow$ to compute expectations:

$$E_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)}) \xrightarrow{N\to\infty} E(f) = \sum_{x} f(x)p(x)$$

$\Rightarrow$ to find a maximum:

$$\hat{x} = \arg_{x^{(i)}} \max[p(x^{(i)})]$$
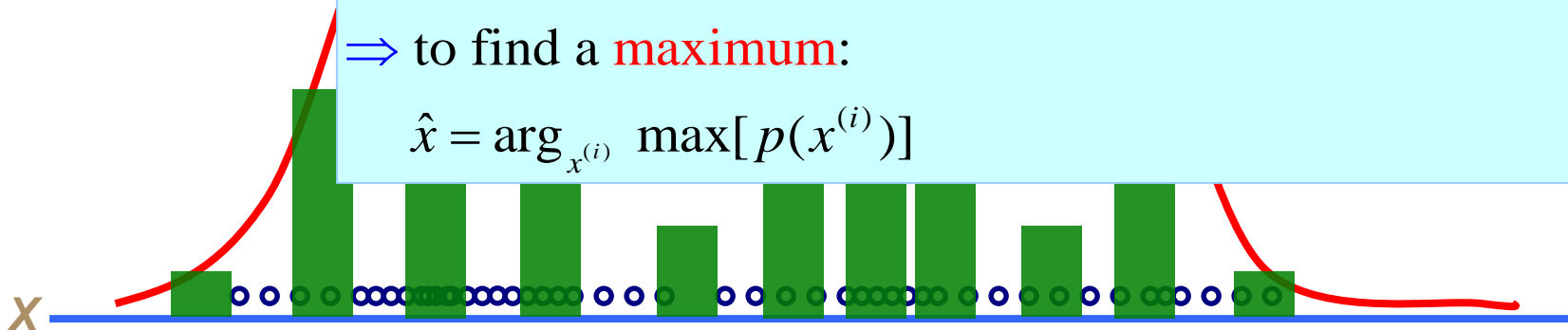
x

Monte Carlo-based Algorithms

- Non-Markovian Methods

Monte Carlo-based Algorithms

- Non-Markovian Methods

- Markovian Methods

Monte Carlo-based Algorithms

- Non-Markovian Methods

  - Rejection Sampling (Smith and Gelfand, 1992)
  - Ratio-of-uniforms method (Wakefield et al., 1994)
  - …

Monte Carlo-based Algorithms

- Non-Markovian Methods

    - Rejection Sampling (Smith and Gelfand, 1992)
    - Ratio-of-uniforms method (Wakefield et al., 1994)
    - …

- Markovian Methods

    - Gibbs Sampling (Geman and Geman 1984)
    - Markov Chain Monte Carlo (MCMC) (Metropolis et al., 1953; Hastings, 1970)
        - Random-walk Metropolis (Roberts, 1995)
        - Independence Metropolis-Hasting (Roberts, 1995)
        - Reversible jump MCMC (Green, 1995)
        - …

Non-Markovian Method

[Rejection Sampling]

$\pi(x)$: intractable target density

$\pi(x) \propto f(x)$

Find a tractable envelope $h(x)$ s.t. $f(x) \leq c \cdot h(x)$

Repeat{

    (*) *Repeat*{

$$y \sim h(\cdot)$$

$$u \sim U(0,1)$$

$$\alpha = \frac{f(y)}{c \cdot h(y)}$$

$$x^{(t+1)} = \begin{cases} y \ \ if \ \ u \leq \alpha \\ \text{go to (*)} \ if \ u > \alpha \end{cases}$$

    }

    $t \leftarrow t+1$ }

Non-Markovian Method

[Rejection Sampling]

$\pi(x)$: intractable target density

$\pi(x) \propto f(x)$

Find a tractable envelope $h(x)$ s.t. $f(x) \le c \cdot h(x)$

It is not easy to find a tractable envelope function.

$Repeat\{$

$(*) \; Repeat\{$

It causes to have much expensive computation times

$\quad y \sim h(\cdot)$

$\quad u \sim U(0,1)$

α does not depend on the previous sample

$\quad \alpha \leftarrow \dfrac{f(y)}{c \cdot h(y)}$

$x^{(t+1)} = \begin{cases} y & if \; u \le \alpha \\ \text{go to } (*) & if \; u > \alpha \end{cases}$

$\}$

$t \leftarrow t+1 \; \}$

Markovian Method

[MCMC]

$\pi(x)$: intractable target density

$\pi(x) \propto f(x)$

Find a tractable proposal $q(\cdot \mid x^{(t)})$

Initialize $x^{(0)}$

Repeat{

$y \sim q(\cdot \mid x^{(t)})$

$u \sim U(0,1)$

$$\alpha = \min\left(1, \frac{f(y)q(x^{(t)} \mid y)}{f(x^{(t)})q(y \mid x^{(t)})}\right)$$

$$x^{(t+1)} = \begin{cases} y & \text{if } u \leq \alpha \\ x^{(t)} & \text{if } u > \alpha \end{cases}$$

$t \leftarrow t+1$ }

Markovian Method

[MCMC]

$\pi(x):$ intractable target density

$\pi(x) \propto f(x)$

Find a tractable proposal $q(\cdot \mid x^{(t)})$

It is not difficult to find a tractable proposal function, e.g., multivariate normal distribution

Initialize $x^{(0)}$

Repeat{

$\quad y \sim q(\cdot \mid x^{(t)})$

$\quad u \sim U(0,1)$

α depends on the previous sample

$$\alpha = \min\left(1, \frac{f(y)q(x^{(t)} \mid y)}{f(x^{(t)})q(y \mid x^{(t)})}\right)$$

$$x^{(t+1)} = \begin{cases} y & \text{if } u \leq \alpha \\ x^{(t)} & \text{if } u > \alpha \end{cases}$$

$\quad t \leftarrow t+1 \ \}$

Nonlinear Mixed-Effects Model

- The first - stage model:

$$[\log y_{ij} \mid \theta_i, \tau^{-1}, t_{ij}] \sim \mathbf{N}(\log f(\theta_i, t_{ij}), \tau^{-1}), \ i = 1, ..., M; j = 1, ..., n_i$$

where $f(\theta_i, t_{ij})$ is a nonlinear function

- The second - stage model :

$$[\theta_i \mid \theta, \Sigma_p] \sim \mathbf{MVN}(\theta, \Sigma_p)$$

Random-Effects

Fixed-Effects

Nonlinear Mixed-Effects Model

- **The first - stage model:**

$$[\log y_{ij} \mid \theta_i, \tau^{-1}, t_{ij}] \sim \mathbf{N}(\log f(\theta_i, t_{ij}), \tau^{-1}), \ i = 1,...,M; j = 1,...,n_i$$

where $f(\theta_i, t_{ij})$ is a nonlinear function

- **The second - stage model :**

$$[\theta_i \mid \theta, \Sigma_p] \sim \mathbf{MVN}(\theta, \Sigma_p)$$

Random-Effects

Fixed-Effects

Nonlinear Mixed-Effects Model

- The first - stage model:

$$[\log y_{ij} \mid \theta_i, \tau^{-1}, t_{ij}] \sim \mathbf{N}(\log f(\theta_i, t_{ij}), \tau^{-1}), \ i = 1,...,M; j = 1,...,n_i$$

where $f(\theta_i, t_{ij})$ is a nonlinear function

- The second - stage model :

$$[\theta_i \mid \theta, \Sigma_p] \sim \mathbf{MVN}(\theta, \Sigma_p)$$

Historic data
Distribution of each parameter

Random-Effects
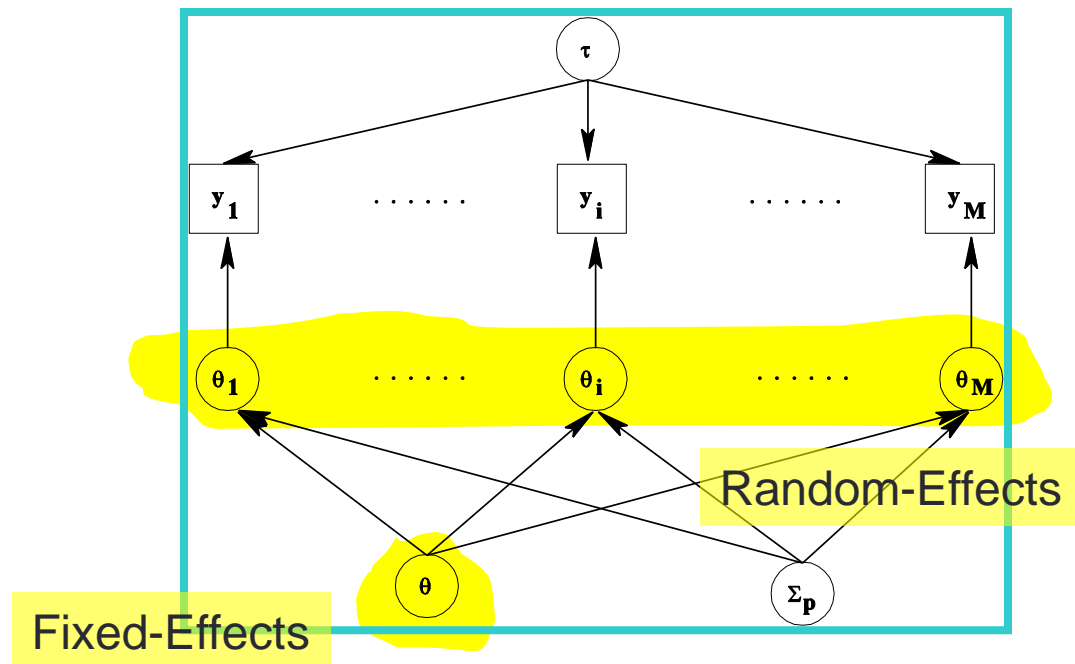
Fixed-Effects

(Bayesian MCMC) Nonlinear Mixed-Effects Model
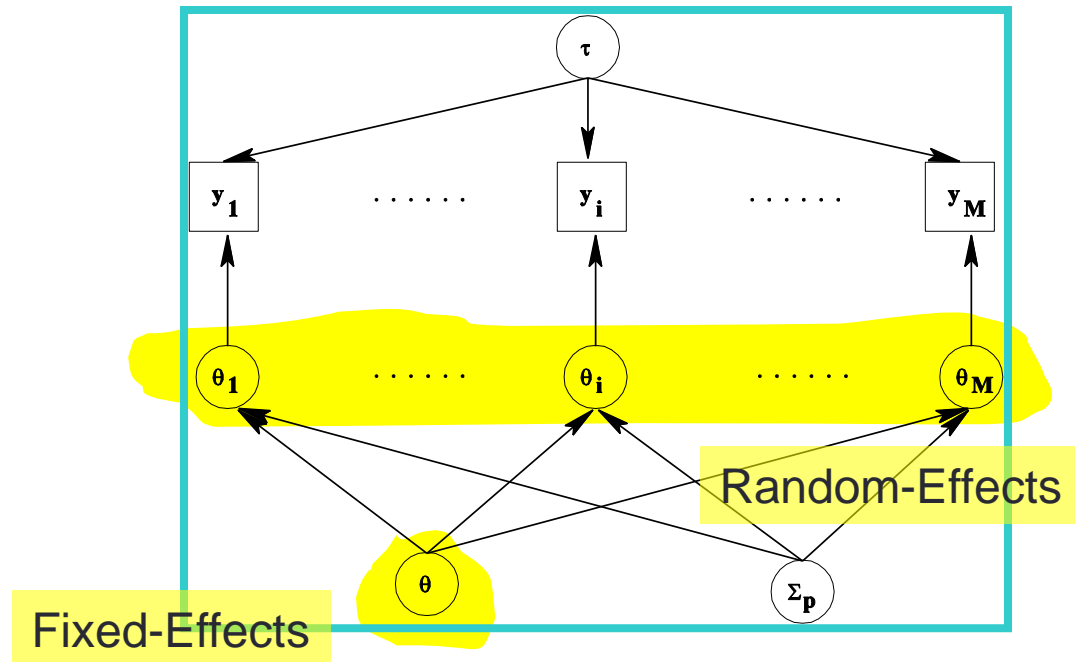
- **The first - stage model:**

$$[\log y_{ij} \mid \theta_i, \tau^{-1}, t_{ij}] \sim \mathbf{N}(\log f(\theta_i, t_{ij}), \tau^{-1}), \ i = 1, ..., M; j = 1, ..., n_i$$

where $f(\theta_i, t_{ij})$ is a nonlinear function

- **The second - stage model :**

$$[\theta_i \mid \theta, \Sigma_p] \sim \mathbf{MVN}(\theta, \Sigma_p)$$

- **The third - stage model :**

$$[\tau] \sim Ga(\frac{v_0}{2}, \frac{\tau_0 v_0}{2})$$

$$[\theta] \sim \mathbf{MVN}(\mathbf{c}, \mathbf{C}_p)$$

$$[\Sigma_p^{-1}] \sim W(\rho, [\rho \mathbf{R}_p]^{-1})$$

Random-Effects

Fixed-Effects

Physiologically based PK drug interaction model
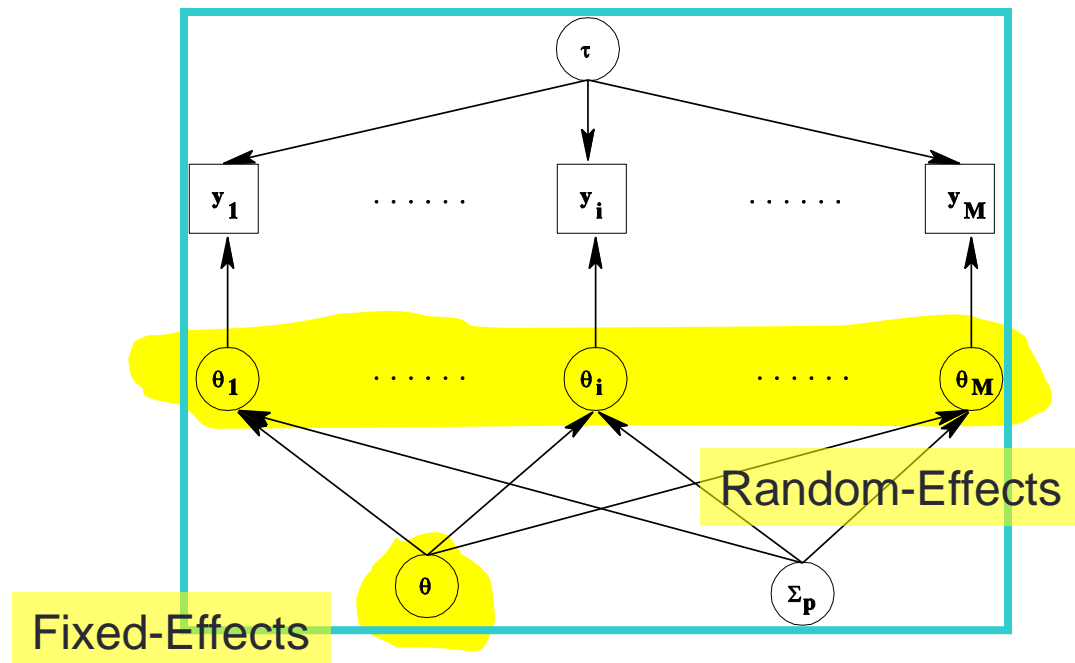
Physiologically based PK drug interaction model

Ketoconazole — Inhibitor (KTZ)   Substrate (MDZ) — Midazolam
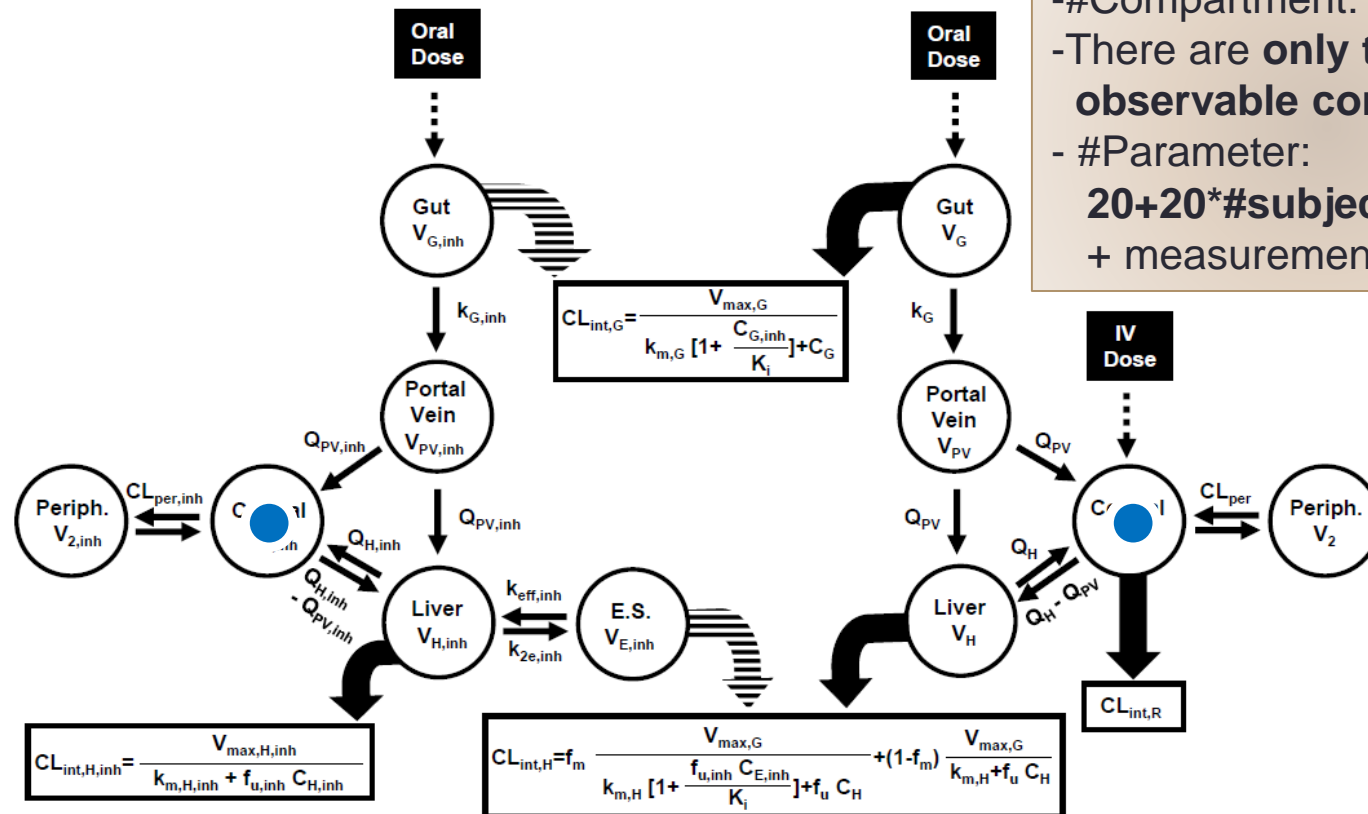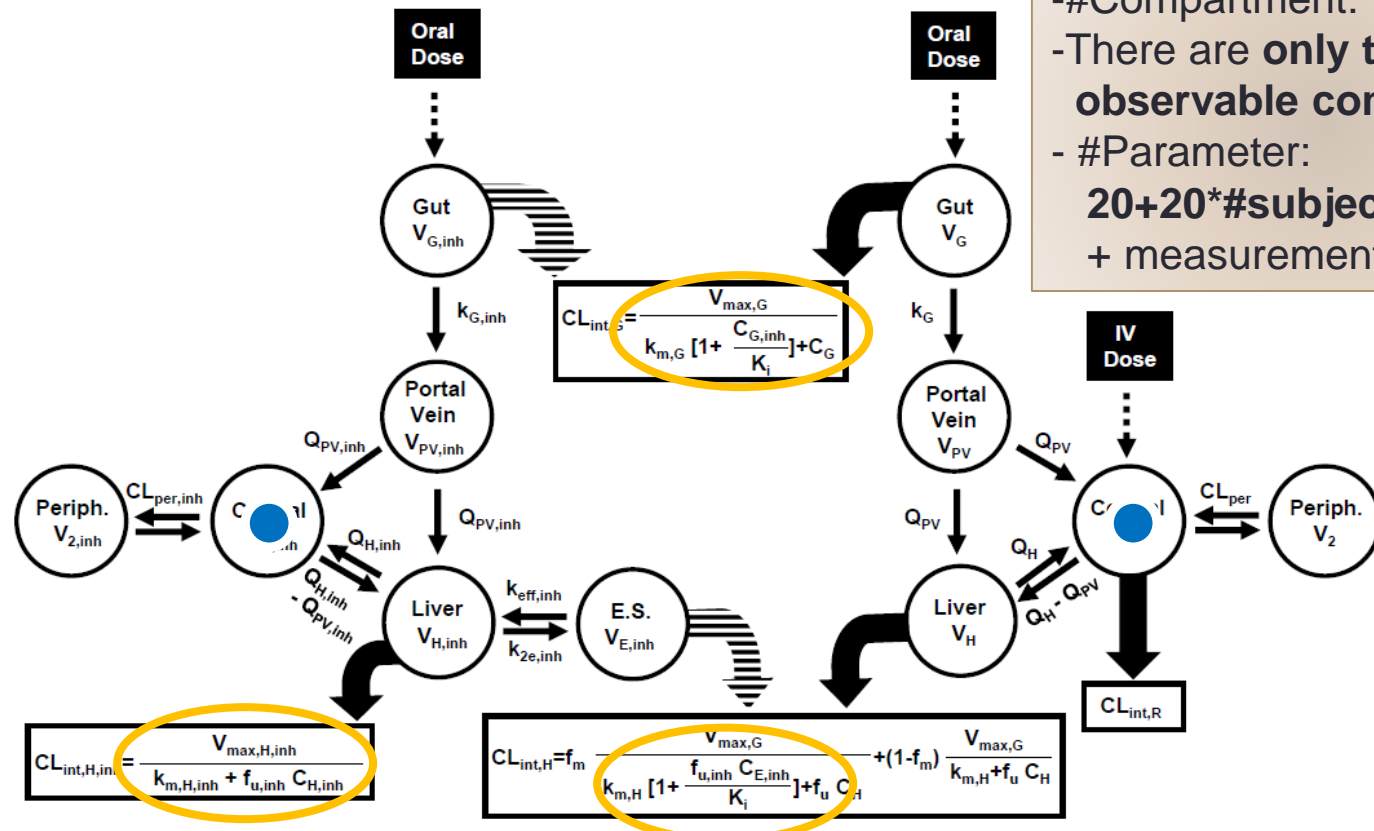
-#Compartment: 11
-There are **only two observable compartments**
- #Parameter:
  20+20*#subject
  + measurement error

Physiologically based PK drug interaction model

Practical Issues for Bayesian MCMC

- Global optima or local optima?
  - High-dimensionality makes it difficult to reach global optima

- The speed of convergence is slow
  - Proposal function (variance-covariance matrix)
  - Starting points (initial values)

- High correlation due to unidentifiable parameters (identifiability)
  - Michaelis-Menten kinetics equation

Three challenges of PK analysis

- Global Optimization: global maximum of the likelihood
  - What is an efficient approach to finding global optima?

- Convergence Rate: the speed of convergence
  - How to improve the speed of convergence?

- (Statistical) Identifiability of PK models
  - What can we do with the statistically unidentifiable parameters?

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability
- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
  - Adaptive designs with three target response rates

An Efficient Global Search Algorithm

- NONMEM (Beal and Sheiner, 1980)
  - The most popular approaches to a population pharmacokinetics/pharmacodynamics (PK/PD) analysis for nonlinear mixed-effects models
  - Local optimization using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm

- A global search algorithm for nonlinear mixed-effects models to meet the challenges of the local optimization in NONMEM

Nonlinear mixed-effects models

- First-stage

$$log y_{ij} = log f(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \ldots N; j = 1, \ldots, n_i \qquad (1)$$

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i

- Second-stage

$$\phi_i = A_i\beta + B_i b_i$$

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i

Nonlinear mixed-effects models

- First-stage

$$logy_{ij} = logf(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \dots N; j = 1, \dots, n_i \qquad (1)$$

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i

- Second-stage

$$\phi_i = A_i\beta + B_i b_i$$

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i

Laplacian estimation method
First-order estimation method
First-order conditional estimation methods

Particle Swarm Optimization (PSO)

$$\widehat{\beta} = argmx[l(\beta; Y, X)]$$

global best
local best

Kennedy and Eberhart, 1995; Eberhart and Kennedy, 1995
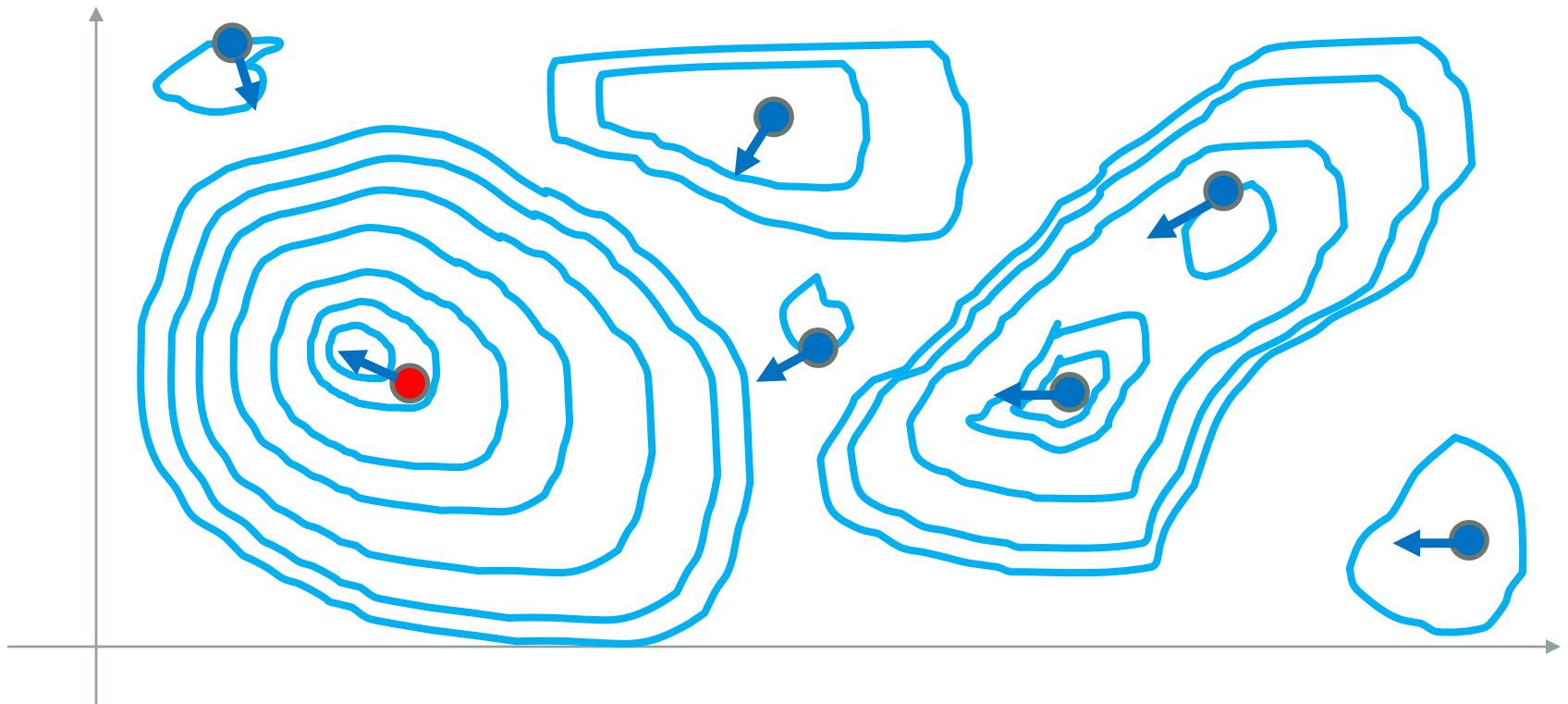
Particle Swarm Optimization (PSO)

- *k*-th iteration

$$v_{k+1}^p = w_k v_k^p + c_1 r_1 \left(x_{lbest}^p - x_k^p \right) + c_2 r_2 \left(x_{gbest} - x_k^p \right) \qquad (4)$$

$$x_{k+1}^p = x_k^p + v_{k+1}^p \qquad (5)$$

- *p = 1,…,P; P*: the population size
- $x_{(lbest)}^p$ and x_{gbest}: local best and global best, respectively
- v_{k+1}^p: the velocity
- w_k: inertia weight

$$w_k = w_{max} - \frac{k}{K} (w_{max} - w_{min})$$

- *c₁, c₂*: cognitive and social coefficient, respectively
- *r₁, r₂*: two random sequences in [0,1]
- *K*: total number iteration number

Nonlinear mixed-effects models

- First-stage

$$logy_{ij} = logf(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \dots N; j = 1, \dots, n_i \qquad (1)$$

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Second-stage

$$\phi_i = A_i\beta + B_i b_i$$

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i
  - $b_i \sim N(0, \Psi)$

Nonlinear mixed-effects models

- First-stage

$$log y_{ij} = log f(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \dots N; j = 1, \dots, n_i \qquad (1)$$

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Second-stage

$$\phi_i = A_i \beta + B_i b_i$$

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i
  - $b_i \sim N(0, \Psi)$

Nonlinear mixed-effects models

- First-stage

$$logy_{ij} = logf(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \ldots N; j = 1, \ldots, n_i$$ (1)

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Second-stage

$$\phi_i = A_i\beta + B_i b_i$$

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i
  - $b_i \sim N(0, \Psi)$

It requires expensive computation time

PSO-based mixed-effects modeling

- High-dimensional parameter space
  - Parameters to estimate
    - β: fixed-effects
    - σ^2: measurement error (variance-covariance matrix)
    - Ψ: inter-individual variance-covariance matrix
    - b_i: random-effects
  - E.g.: 5 fixed-effects, 5 random-effects, 10 subjects => $5 + 5 \times 10 = 50$ without variance estimation
- Expensive computation
- Slow speed of convergence

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO
    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO

    Reduce the population size

    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO

    Reduce the population size

    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization
- Sacrifice random-effects
  - Random-effects by NONMEM
  - Fixed-effects and others by NONMEM+PSO

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO
    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization

- Sacrifice random-effects
  - Random-effects by NONMEM
  - Fixed-effects and others by NONMEM+PSO

Reduce the population size

Reduce the PSO parameter space

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO
    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization

Reduce the population size

- Sacrifice random-effects
  - Random-effects by NONMEM

Reduce the PSO parameter space

  - Fixed-effects and others by NONMEM+PSO

- Multivariate population
  - Grid population using univariate uniform distribution
  - Random-grid population using multivariate uniform distribution

PSO-based mixed-effects modeling

- Hybrid approach
  - NONMEM + PSO
    - NONMEM: exploitation by a local optimization
    - PSO: exploration by a global optimization

Reduce the population size

- Sacrifice random-effects
  - Random-effects by NONMEM
  - Fixed-effects and others by NONMEM+PSO

Reduce the PSO parameter space

- Multivariate population
  - Grid population using univariate uniform distribution
  - Random-grid population using multivariate uniform distribution

Increase the population diversity

Nonlinear mixed-effects models

- First-stage

$$logy_{ij} = logf(\phi_i, t_{ij}) + \epsilon_{ij}; i = 1, \dots N; j = 1, \dots, n_i$$

  - N: the number of subjects
  - n_i: the number of observations from the ith subject
  - y_{ij}: the drug concentration at time t_{ij}
  - f: a nonlinear function of a subject-specific parameter vector ϕ_i
  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Second-stage

$$\phi_i = A_i\beta + B_ib_i$$

Global + Local Optimization

  - A_i and B_i: known design matrices for fixed-effects β and random-effects b_i
  - $b_i \sim N(0, \Psi)$

Local Optimization

Convergence of PSO+NONMEM

- First-order stability analysis (expected value)
  - Trelea (2003)
  - The expected value of the position of each particle converges to its equilibrium

$$\frac{c_1 x_{lbest} + c_2 x_{gbest}}{c_1 + c_2}$$

$$iff$$

$$w < 1, c = \frac{c_1 + c_2}{2} > 0, 2w - c + 2 > 0$$

- PSO+NONMEM: $w \in [0.4, 0.9], c_1 = c_2 = 2$ ➜ $2 \cdot 0.4 - 2 + 2 > 0$

Convergence of PSO+NONMEM

- Second-order stability analysis (variance)
  - Jiang et al. (2007); Poli (2009); Poli et al. (2007)
  - The variance of the position of each particle converges to zero

$$iff$$

$$\frac{c_1 + c_2}{2} < \frac{12(w^2 - 1)}{5w - 7}$$

- PSO+NONMEM: $w \downarrow 0.4, c_1 = c_2 = 2$ ➔ $2 < 2.016$

Subject-specific parameter estimation: local or global?

Theorem 1. *Let us consider a linear mixed-effect model, i.e.*

$$f(\phi_i, t_{ij}) = A_i \beta + B_i b_i$$

in (1). If the parameters β, σ^2, and Ψ converge to their global optima, $\hat{\beta}, \widehat{\sigma^2},$ and $\widehat{\Psi}$, and $\hat{b}_i$ is a local optimum of the individual random-effect b_i, then $\hat{b}_i$ is its global optimum, given the ith observation y_i, $\hat{\beta}, \widehat{\sigma^2},$ and $\widehat{\Psi}$.

Subject-specific parameter estimation: local or global?

Theorem 2. Let us consider a nonlinear mixed-effect model, i.e., $f(\cdot)$ is nonlinear in (1). Suppose the parameters β, σ^2, and Ψ converge to their global optima, $\hat{\beta}, \widehat{\sigma^2}$, and $\widehat{\Psi}$, and $\hat{b}_i$ is a local optimum of the individual random-effect b_i. If $f(\cdot)$ is either concave and nonincreasing in b_i or convex and nondecreasing in b_i, then $\widehat{b}_i$ is its global optimum, given the ith observation y_i, $\hat{\beta}, \widehat{\sigma^2}$, and $\widehat{\Psi}$.

The flowchart of the proposed P-NONMEM

Kim and Li **2011**. *J. of PK. and PD.*
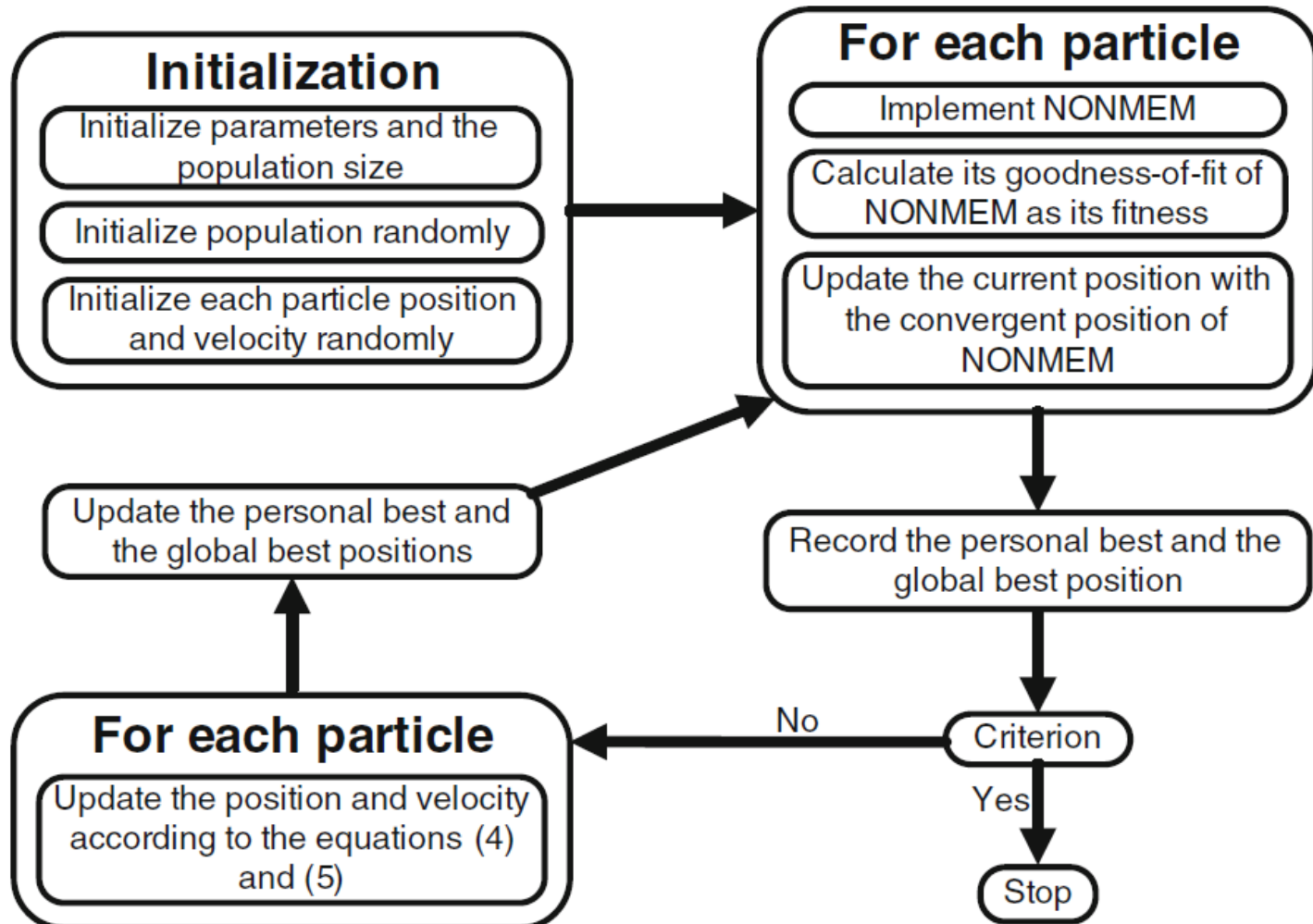
Simulation studies

PK example

$$\begin{cases} \frac{dA1(t)}{dt} = -\frac{Qh \cdot fu \cdot \frac{Vmax}{Km + \frac{A1(t)}{V1}}}{Qh + fu \cdot \frac{Vmax}{Km + \frac{A1(t)}{V1}}} \cdot \frac{A1(t)}{V1} + CL12 \cdot \left(-\frac{A1(t)}{V1} + \frac{A2(t)}{V2}\right) \\ \frac{dA2(t)}{dt} = CL12 \cdot \left(-\frac{A1(t)}{V1} + \frac{A2(t)}{V2}\right) \end{cases}$$

$$f(\Phi_i, t_{ij}) = \frac{A1(t)}{V1}$$

$$\Phi_i = A_i\beta + B_i b_i = A_i\beta + b_i$$

$$\beta = (logV1, logV2, logVmax, logKm)'$$

$$b_i = (V1_i)$$

$$A_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
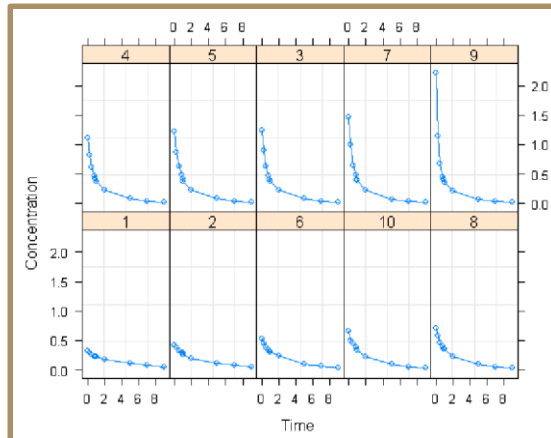
PD example

$$f(\Phi_i, x_{ij}) = \frac{Emax \cdot x_{ij}}{C50 + x_{ij}}$$

$$\Phi_i = A_i\beta + B_i b_i = A_i\beta + b_i$$

$$\beta = (logEmax, logC50)'$$

$$b_i = (Emax_i, C50_i)$$

$$A_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
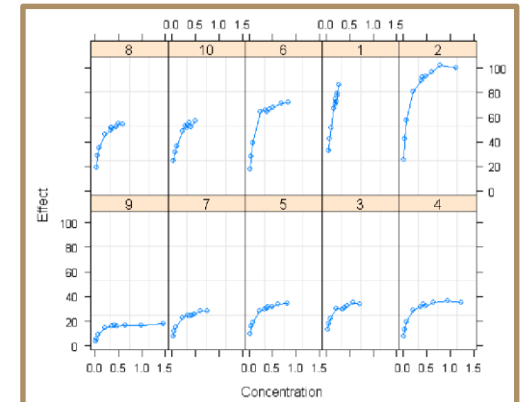
After 1-th Iteration
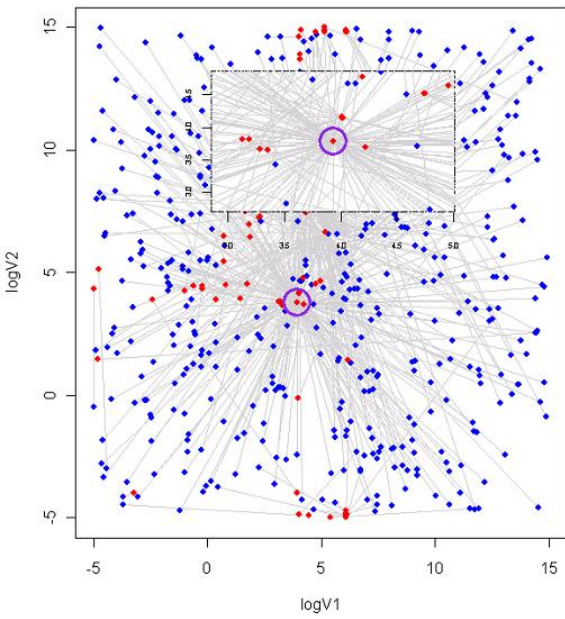
PK example
P=729 (=3^6)
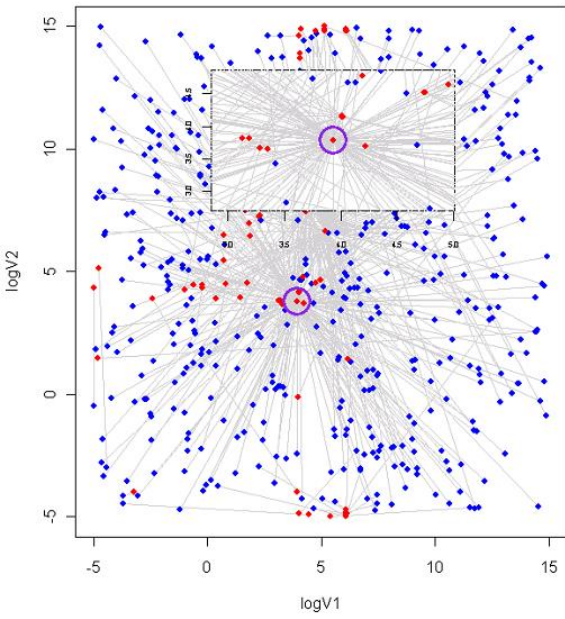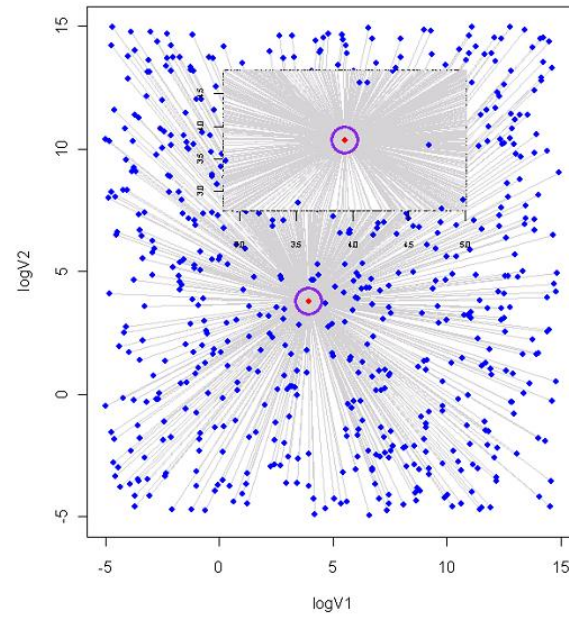
PK example
P=729 (=3⁶)

After 1-th Iteration
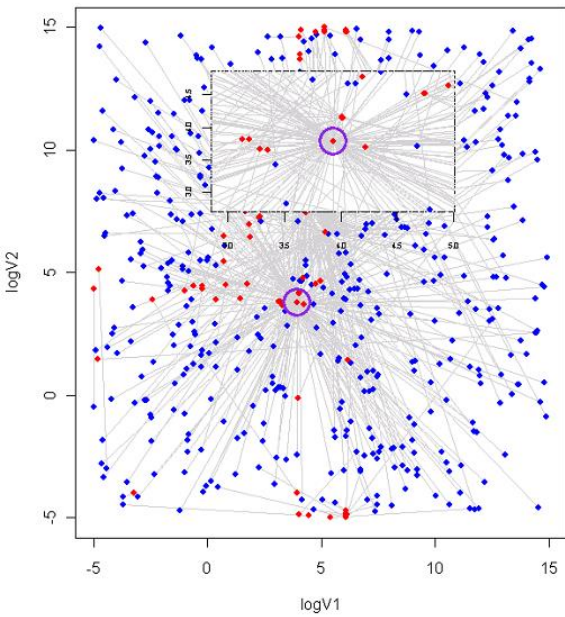
After 50-th Iteration
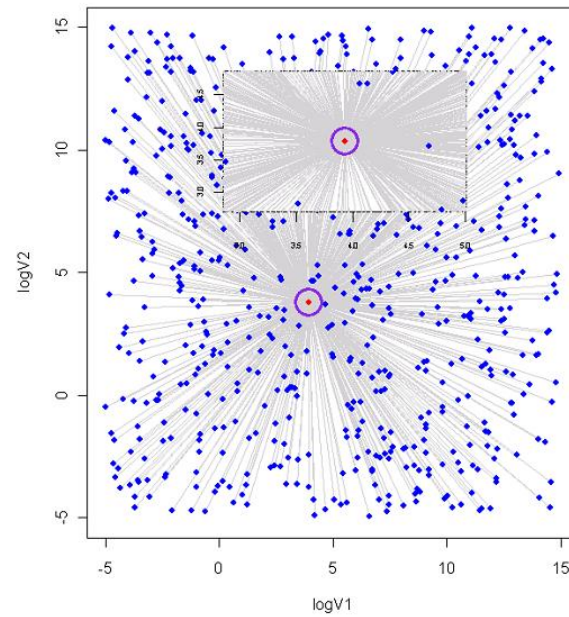
PK example
P=729 (=3^6)

PD example
P=243 (=3^5)

After 1-th Iteration

After 1-th Iteration

After 50-th Iteration

PK example
P=729 (=3^6)

After 1-th Iteration

After 20-th Iteration
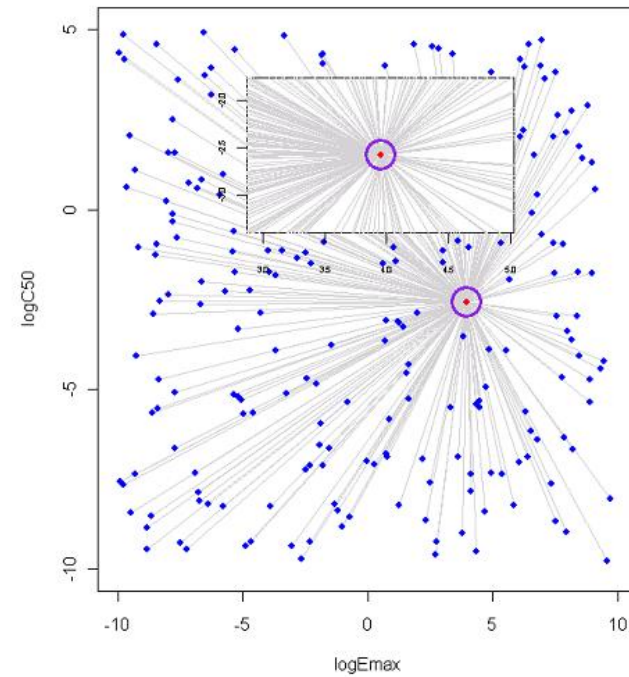
PD example
P=243 (=3^5)

After 1-th Iteration
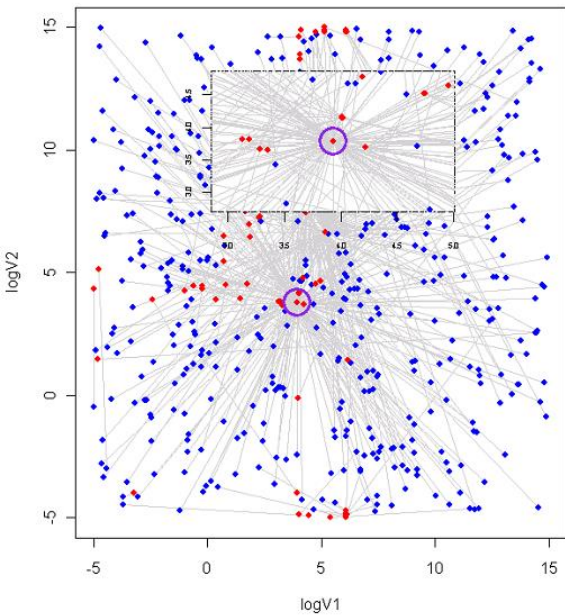
After 50-th Iteration
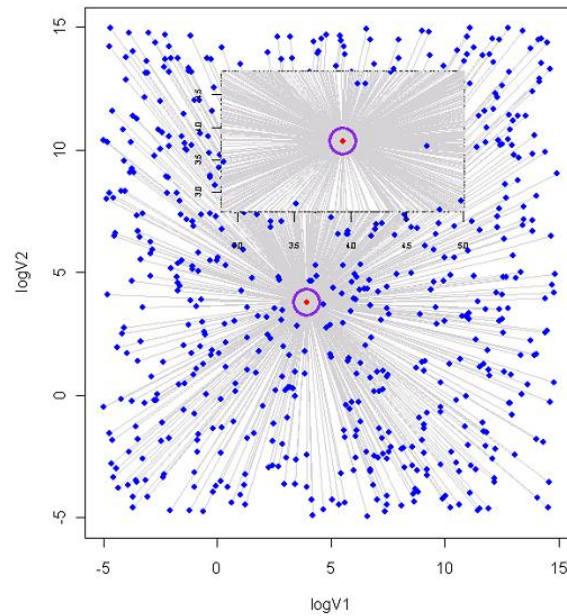
PK example
$P=729\ (=3^6)$

$$10^6 >> 3^6$$

$$10^5 >> 3^5$$

PD example
$P=243\ (=3^5)$
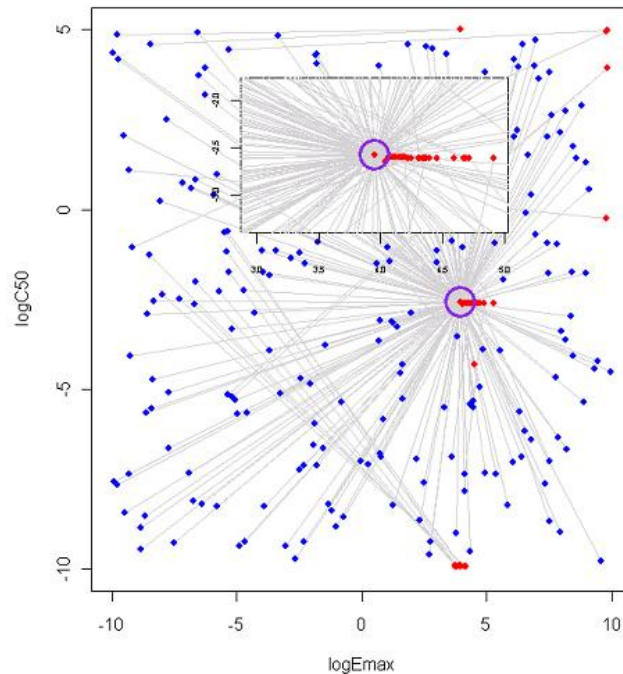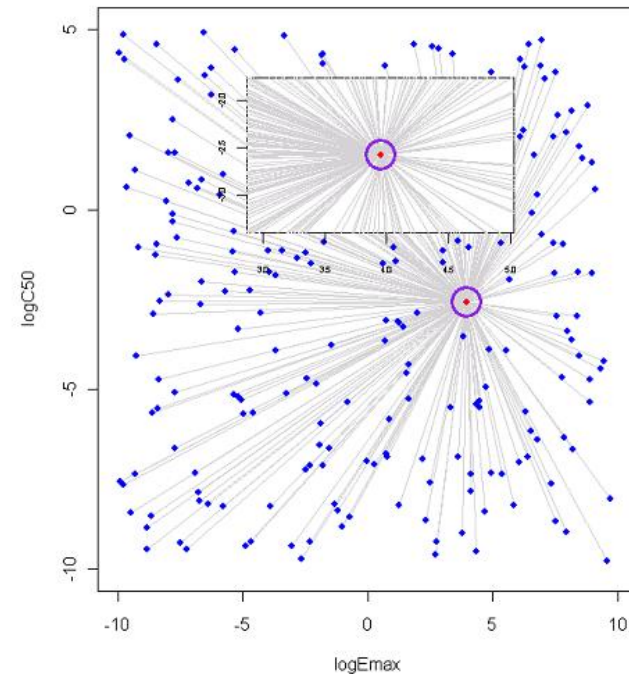
After 1-th Iteration

After 20-th Iteration

Subject-level parameter Absolute relative bias

Mean Squared Error of Prediction

PK example
P=729 (=3^6)

PK example *P*=729 (=3⁶)

PD example *P*=243 (=3⁵)

Subject-level parameter Absolute relative bias

Mean Squared Error of Prediction
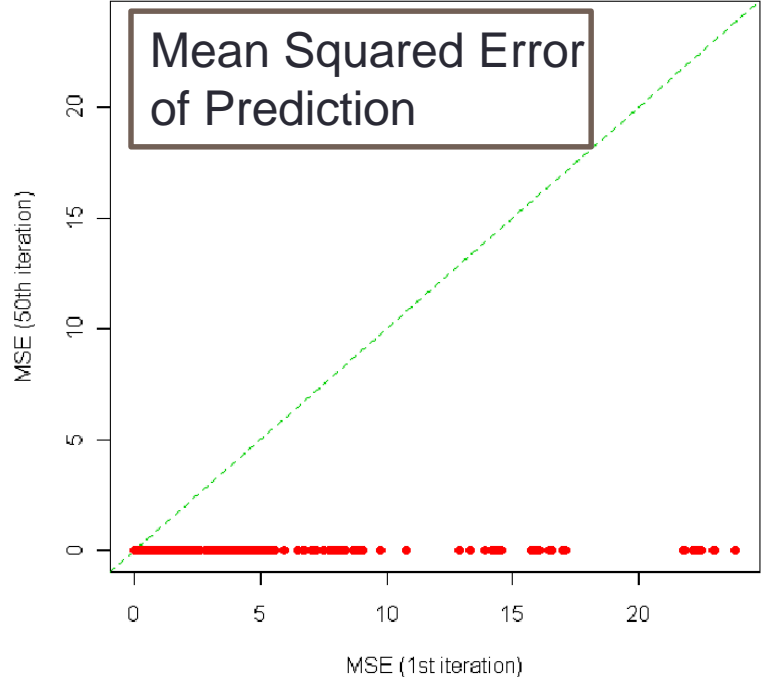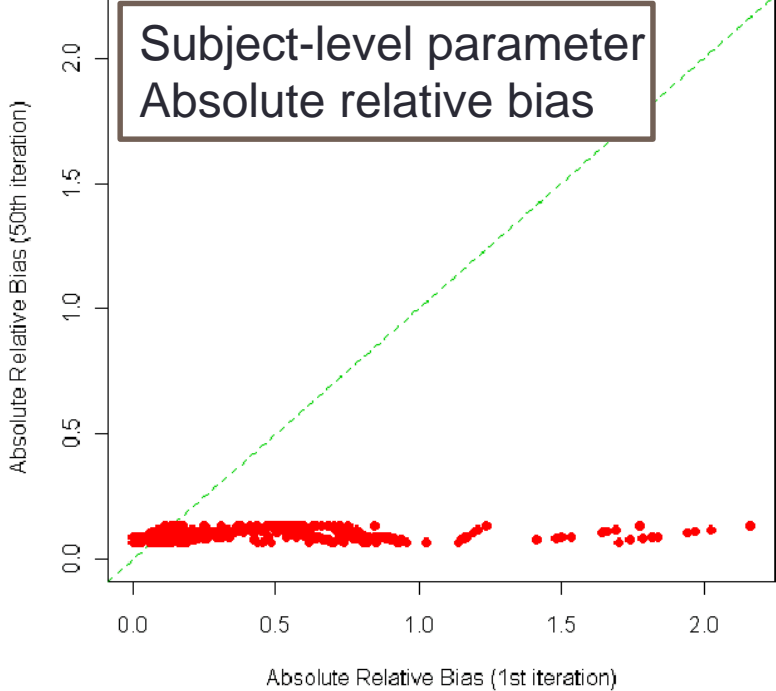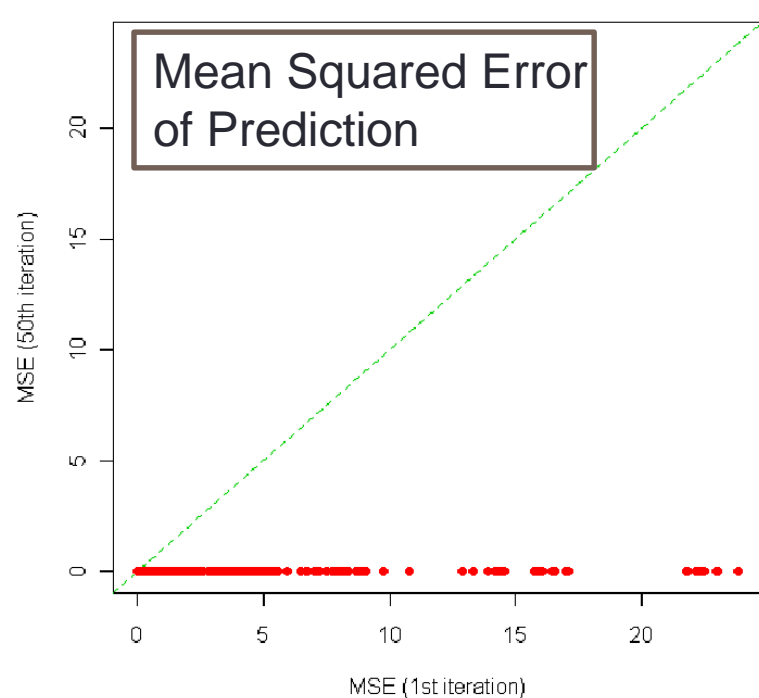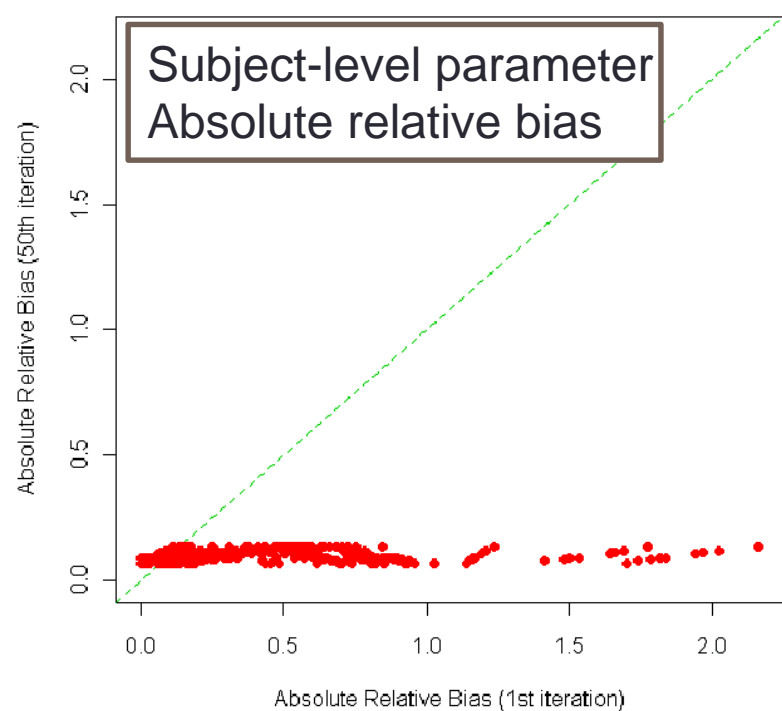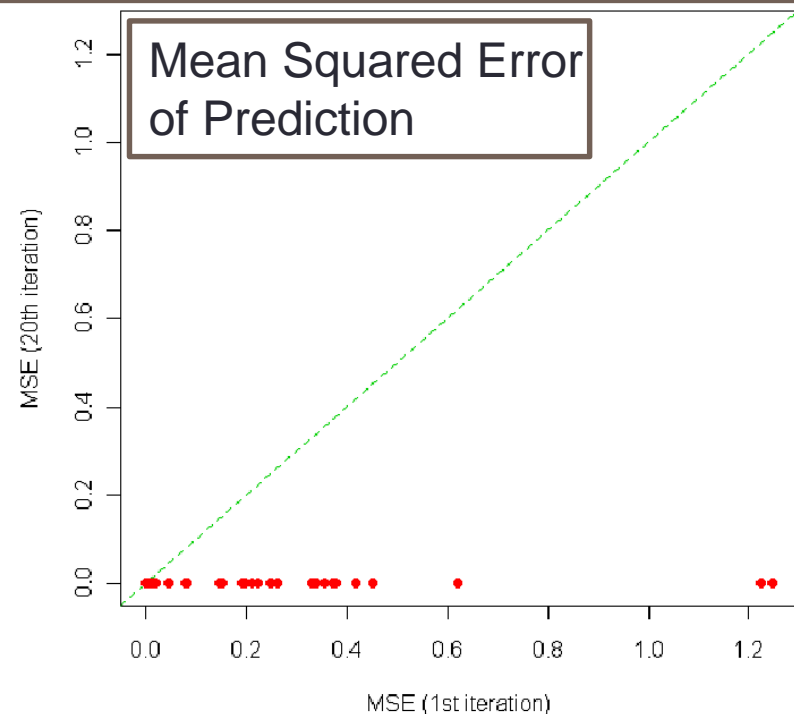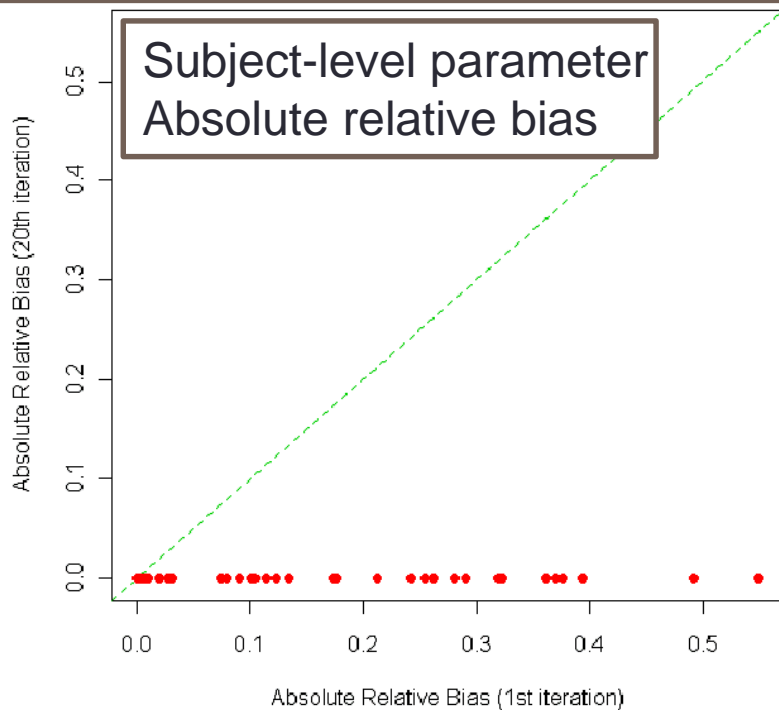
Summary

- The proposed P-NONMEM is not sensitive to initial value selection.

- Even when the initial values are far away from their global optimal, P-NONMEM almost guarantees the global optimization.

- P-NONMEM guarantees the global optimization for fixed effect and variance parameters.

- Under certain regularity conditions, it also leads to global optimization for random effects

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability
- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
  - Adaptive designs with three target response rates

Mathematical Identifiability

- For PK models the corresponding equations are

$$\dot{x}(t) = f(x(t), B(\theta)u(t), \theta, t), \quad x(0) = x_0$$

$$y(t) = G(x(t), B(\theta)u(t), \theta, t)$$

(1)

$x(t)$: the state variables

$x(0)$: the initial conditions

$u(t)$: the input to the system

$B(\theta)$: the matrices depending on θ

$y(t)$: observations

- A single parameter θ of Equation (1) is ***globally identifiable*** if there exists a <u>unique</u> solution for θ

- A parameter with <u>countable or uncountable</u> number of solutions is ***locally identifiable or unidentifiable***

Mathematical Identifiability

- For PK models the corresponding equations are

$$\dot{x}(t) = f(x(t), B(\theta)u(t), \theta, t), \quad x(0) = x_0$$

$$y(t) = G(x(t), B(\theta)u(t), \theta, t)$$

(1)

$x(t)$: the state variables

$x(0)$: the initial conditions

$u(t)$: the input to the system

$B(\theta)$: the matrices depending on θ

$y(t)$: observations

- A single parameter θ of Equation (1) is ***globally identifiable*** if there exists a unique solution for θ

> **Given a model formation and <u>noise-free (perfect)</u> data**
> **Which parameters of the model are identifiable?**
> **(Bellman and Astrom, 1970)**

- A para_____ ____ __ of solutions is ***locally identifiable or unidentifiable***

Statistical Identifiability

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t) + \varepsilon(t)$$

$\mathbf{x}(t)$: the state variables

$\mathbf{x}(0)$: the initial conditions

$\mathbf{u}(t)$: the input to the system

$\mathbf{A}(\theta)$, $\mathbf{B}(\theta)$: the matrices depending on θ

$\mathbf{y}(t)$: observations

Statistical Identifiability

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t) + \varepsilon(t)$$

$\mathbf{x}(t)$: the state variables

$\mathbf{x}(0)$: the initial conditions

$\mathbf{u}(t)$: the input to the system

$\mathbf{A}(\theta), \ \mathbf{B}(\theta)$: the matrices depending on θ

$\mathbf{y}(t)$: observations

Given (a perfect model structure and) experimental (noise) data
Is it possible to uniquely and accurately estimate the parameters?

Statistical Identifiability

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t), \quad \mathbf{x}(0) = \mathbf{x_0}$$

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{x}(t), \mathbf{B}(\theta)\mathbf{u}(t), \theta, t) + \varepsilon(t)$$

$\mathbf{x}(t)$: the state variables

$\mathbf{x}(0)$: the initial conditions

$\mathbf{u}(t)$: the input to the system

$\mathbf{A}(\theta), \ \mathbf{B}(\theta)$: the matrices depending on θ

$\mathbf{y}(t)$: observations

**Given (a perfect model structure and) experimental (noise) data
Is it possible to uniquely and accurately estimate the parameters?**

- Estimates of the statistical identifiability are highly depending on **the quality of the data**

Michaelis-Menten (MM) Kinetics

- MM Kinetics Equation: $V(t) = \dfrac{dC(t)}{dt} = \dfrac{Vmax \cdot C(t)}{Km + C(t)}$

  - **$V(t)$**: the overall velocity of the reaction
  - **$Vmax$**: the maximum velocity
  - **Km**: MM constant
  - **$C(t)$**: the concentration
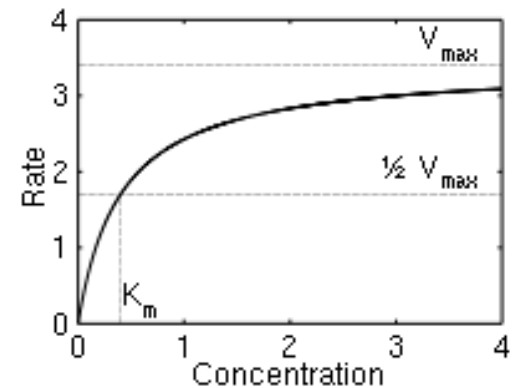
from wikipedia

Michaelis-Menten (MM) Kinetics

- MM Kinetics Equation: $V(t) = \dfrac{dC(t)}{dt} = \dfrac{Vmax \cdot C(t)}{Km + C(t)}$

  - **$V(t)$**: the overall velocity of the reaction
  - **$Vmax$**: the maximum velocity
  - **Km**: MM constant
  - **$C(t)$**: the concentration
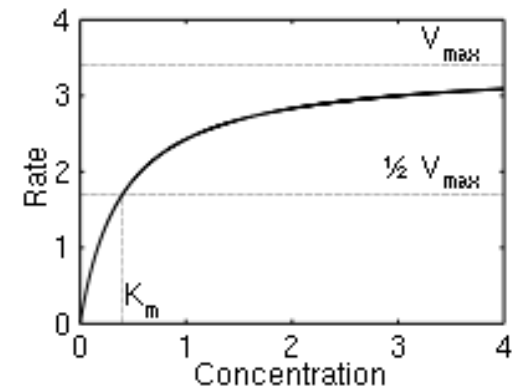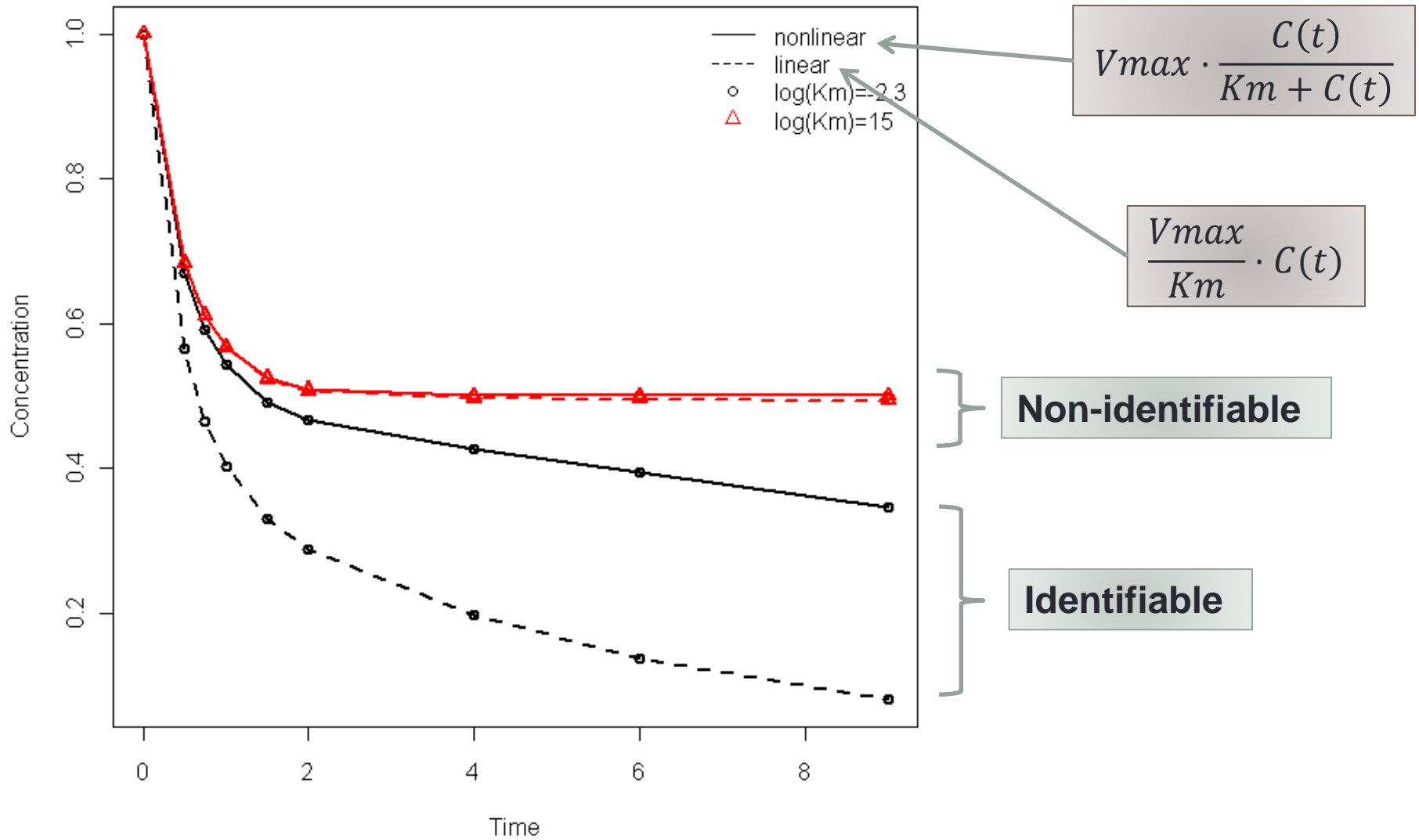
from wikipedia

$$V(t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx \frac{Vmax \cdot C(t)}{Km} \text{ if } C(t) \ll Km.$$

$$V(t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx Vmax \text{ if } C(t) \gg Km.$$

$$Vmax \cdot \frac{C(t)}{Km + C(t)}$$

$$\frac{Vmax}{Km} \cdot C(t)$$

Non-identifiable

Identifiable

What's the matter?

- Derivative-based optimization
  - Singularity
  - NONMEM uses derivative-based optimizations
- Bayesian approach
  - No theoretical concern for both mathematical and statistical identifiability
    - No singularity issue due to priors (Lindley, 1971)
  - Poor convergence of MCMC (Poirier, 1998; Gelfand and Sahu, 1999; Eberly and Carlin, 2000)

What's new?

- NONMEM is replaced with a derivative-free local optimization

- Convergence criteria
  - The local best-quartile method
  - The global best-variance method
  - The local best-quartile-variance method

Kim and Li **2014**. *CMPB.*

The local best-quartile method

- Suppose θ^k is the $S \times p$ matrix of the population (local best) of size S and the p parameters at kth iteration, i.e.,

$$\theta^k = \begin{bmatrix} \theta_{11}^k & \cdots & \theta_{1p}^k \\ \vdots & \ddots & \vdots \\ \theta_{S1}^k & \cdots & \theta_{Sp}^k \end{bmatrix},$$

  where θ_{ij}^k is the local best of ith particle of jth parameter at kth iteration, $1 \leq i \leq S$, $1 \leq j \leq p$.

- The difference between the first and third quartiles for each parameter is calculated based on $\hat{\theta}^k$, i.e.,

$$d_j^k = \left| Q_1^{kj} - Q_3^{kj} \right|, \textit{ where } j = 1, \ldots, p,$$

  obtaining the maximum difference of all the parameters as the following

$$d^k = \max_{j=1,2,\ldots,p} d_j^k.$$

- The $p \times p$ correlation matrix of θ^k, i.e.,

$$\Omega^k = \begin{bmatrix} 1 & \cdots & \omega_{1p} \\ \vdots & \ddots & \vdots \\ \omega_{p1} & \cdots & 1 \end{bmatrix}$$

- Its maximum and minimum eigenvalues, λ_{max}^k and λ_{min}^k, are estimated to calculate the ratio of two eigenvalues, $\rho^k = \left| \frac{\lambda_{min}^k}{\lambda_{max}^k} \right|$.

- If at least one parameter has $d_j^k = 0$, then the eigenvalues cannot be obtained, so we will assign zero to ρ^k in this case.

The global best-variance method

- Suppose ψ^k is the $k \times p$ matrix consisting of the global best for each parameter up to kth iteration,

$$\psi^k = \begin{bmatrix} \psi_1^1 & \cdots & \psi_p^1 \\ \vdots & \ddots & \vdots \\ \psi_1^k & \cdots & \psi_p^k \end{bmatrix},$$

where ψ_j^i is the global best of jth parameter at $i$the iteration and l^k is the vector of the loglikehood of each global best of size k such as $l^k = (l^1, l^2, \ldots, l^k)$. Then the reduced matrix is obtained based on the user-defined window size, w.

$$\psi_w^k = \begin{bmatrix} \psi_1^{k-w+1} & \cdots & \psi_p^{k-w+1} \\ \vdots & \ddots & \vdots \\ \psi_1^k & \cdots & \psi_p^k \end{bmatrix},$$

where $k \geq w > 0$, and the reduced loglikehood vector is $l_w^k = (l^{k-w+1}, l^{k-w+2}, \ldots, l^k)$.

- $SD(\psi_w^k) = \max_{j=1,\ldots,p} SD_w(\psi_j^k);$
- $SD(l_w^k) = \sqrt{Var(l^{k-w+1}, l^{k-w+2}, \ldots, l^k)},$

$$\text{where } SD_w(\psi_j^k) = \sqrt{Var\left(\psi_j^{k-w+1}, \psi_j^{k-w+2}, \ldots, \psi_j^k\right)}.$$

The local-quartile-variance method

- $SD\left(d_w^k\right) = \sqrt{Var(d^{k-w+1}, d^{k-w+2}, \ldots, d^k)};$
- $SD\left(\rho_w^k\right) = \sqrt{Var(\rho^{k-w+1}, \rho^{k-w+2}, \ldots, \rho^k)}.$

Convergence diagnostics

If $SD\left(\rho_w^k\right)$ is less than equal to the user-defined cutoff value (α) with the window size of w, LPSO will be considered as converged to a global optimum. Furthermore, if d^k is greater than the user-defined cutoff value (β), the model is considered as non-identifiable, where k is the number of iterations to converge which is identified by $SD\left(\rho_w^k\right)$. The general guideline for α and β is 0.001 and one, respectively.

Simulation

- ## The constants of PSO
    - $(c_1, c_2, w_{max}, w_{min}, K) = (2,2,0.9,0.3,5000)$
    - #particles of each parameter: 10 for PSO and 5 for LPSO.
    - The parameter boundaries = $(-20, 20)$.
    - The true values are $\theta^{true} = (0, -2.3)$ for the identifiable case and $(0,15)$ for non-identifiable case.

- For both PSO and LPSO, the same seed number was used to generate the initial population.

$$Vmax \cdot \frac{C(t)}{Km + C(t)}$$

$$\frac{Vmax}{Km} \cdot C(t)$$

Non-identifiable

Identifiable

(a) After 1st iteration

Fig. 4 – The scatter-box plots between log V_{max} and log K_m for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.

Fig. 4 – The scatter-box plots between log V_{max} and log K_m for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.

(b) After 500th iteration

Fig. 4 – The scatter-box plots between log V_{max} and log K_m for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.

(c) After 5000th iteration

Identifiable
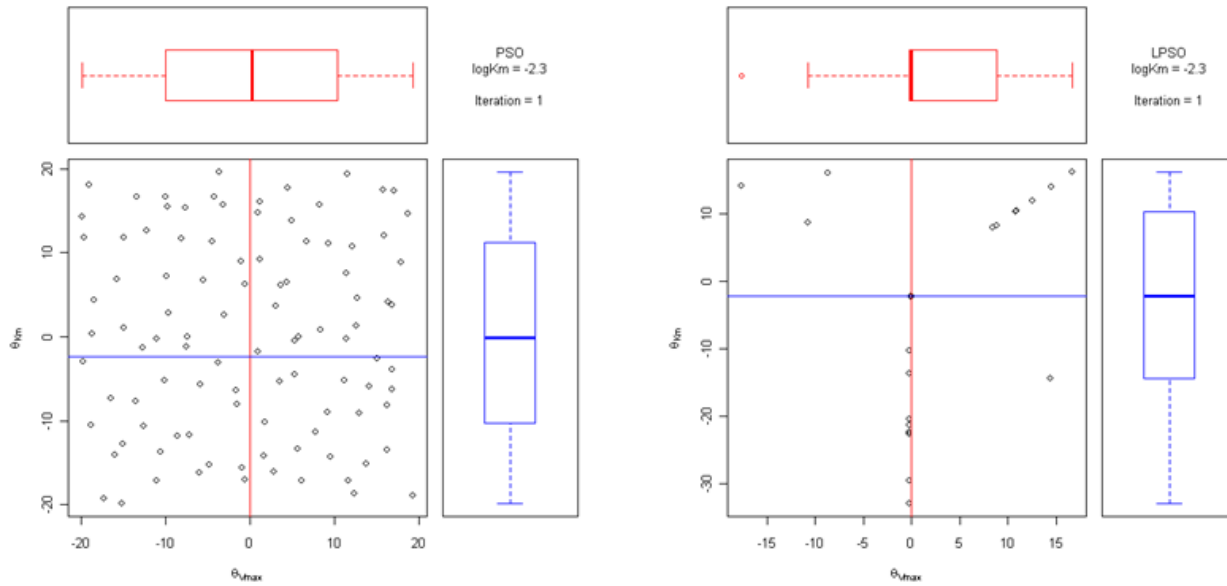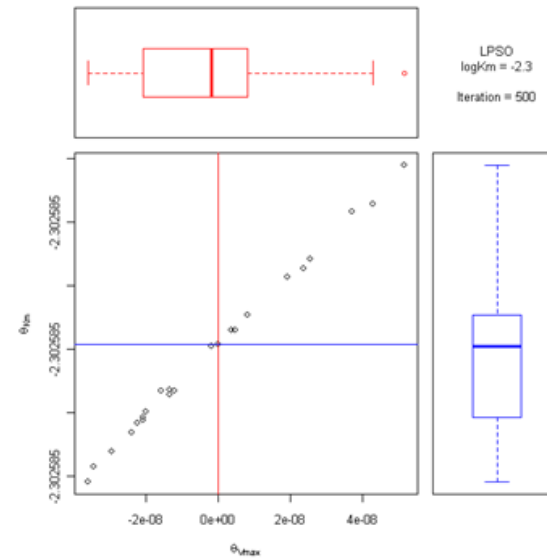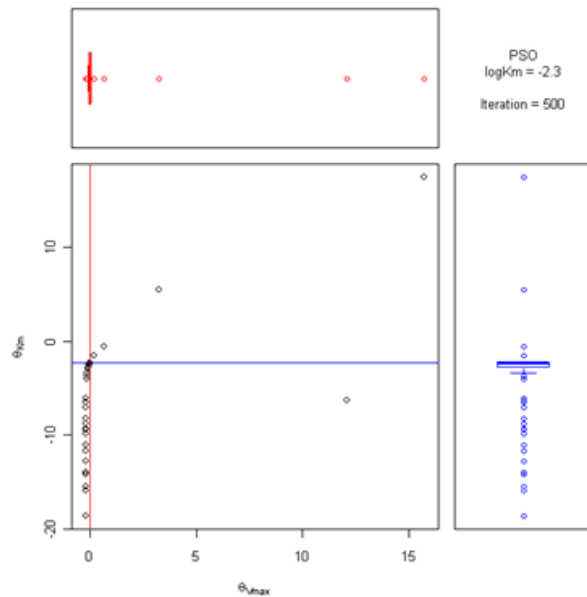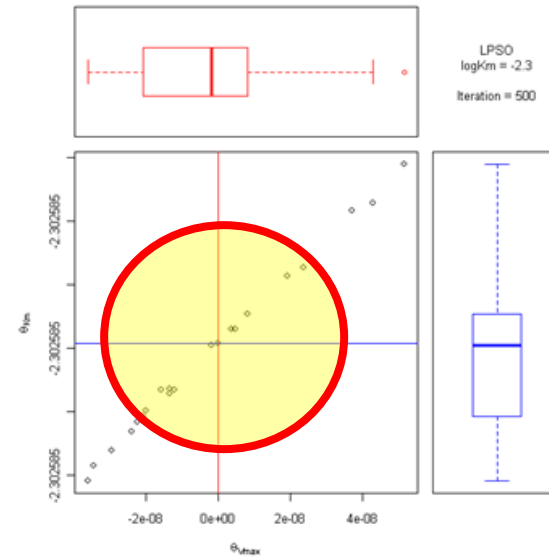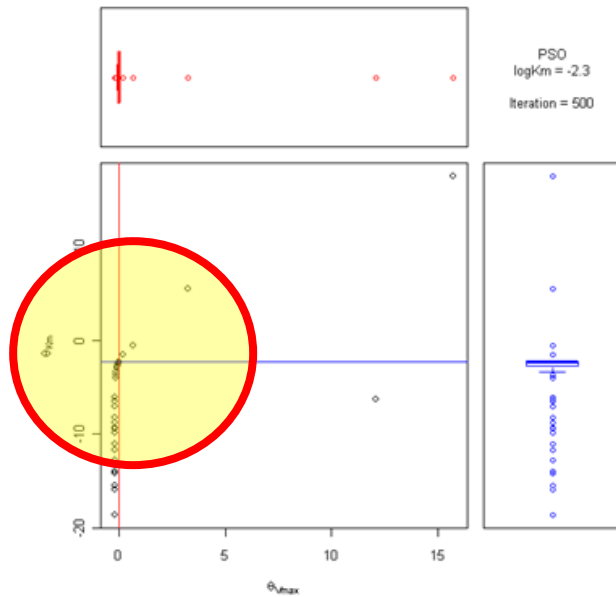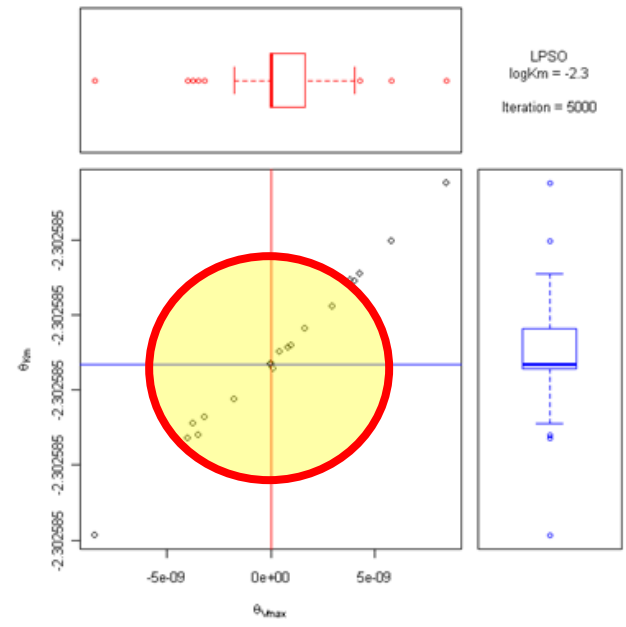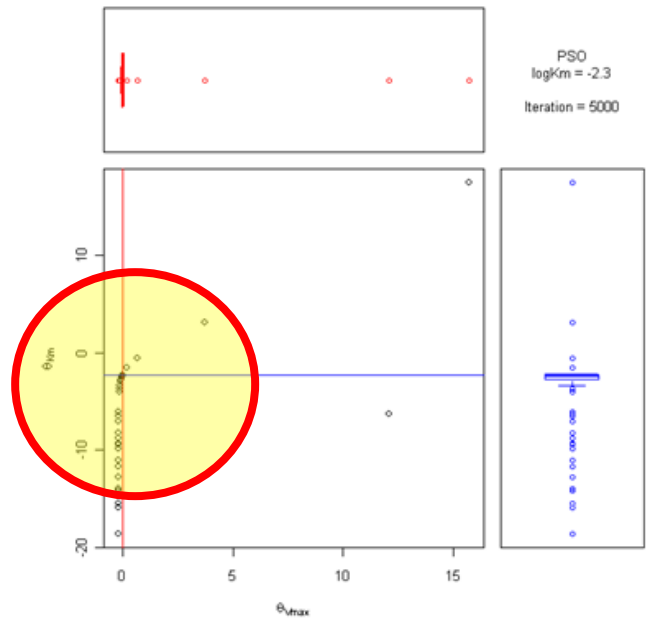
Non-identifiable
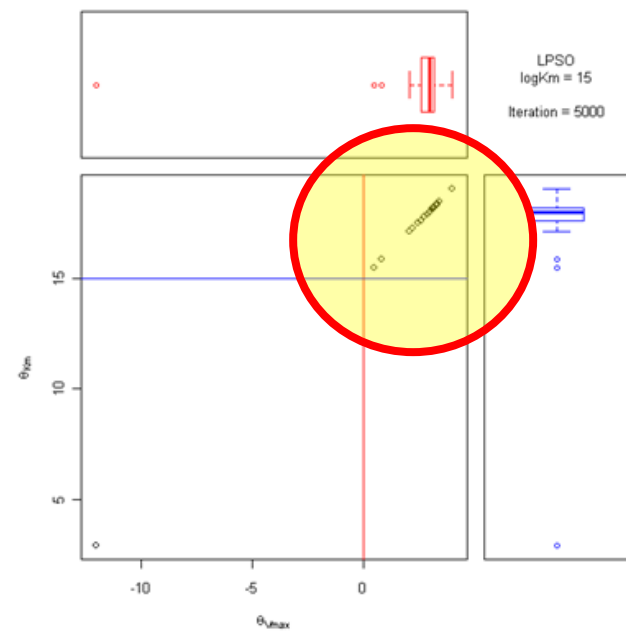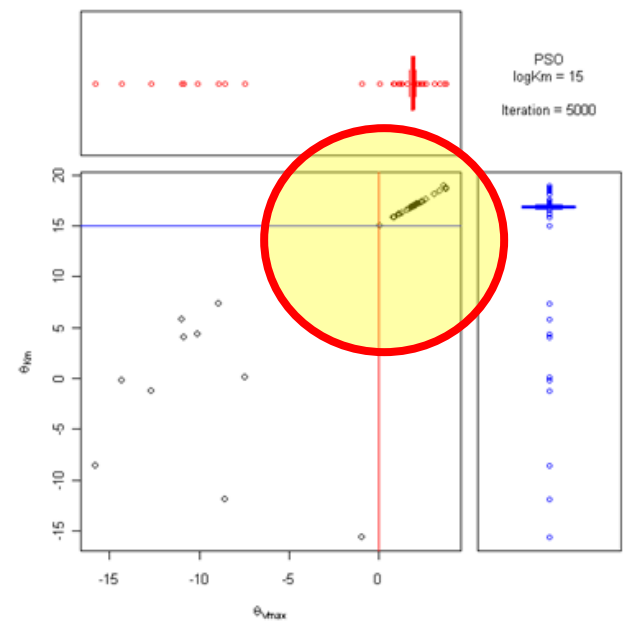
Fig. 4 – The scatter-box plots between log V_{max} and log K_m for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.

(c) After 5000ᵗʰ iteration

Fig. 4 – The scatter-box plots between log V_{max} and log K_m for PSO and LPSO. The plots in the left and right columns are for PSO and LPSO, respectively, and the first and second rows are for identifiable and non-identifiable cases. The solid lines in the plot indicate the true values for each parameter.

Summary

- A novel version of PSO is proposed with enhancing the convergence of the local best using a derivative-free local optimization algorithm, which is called LPSO.

- LPSO converges to a global optimum much faster than PSO does.

- Since PSO is a derivative-free algorithm and a derivative-free local optimization is combined, the proposed LPSO becomes a derivative-free global optimization algorithm so that LPSO can be applied to the parameter estimation regardless of the identifiability.

- Several convergence diagnostic measures are proposed and evaluated through both the simulation studies and clinical PK data analysis.

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability

- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
  - Adaptive designs with three target response rates

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability

- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
  - Adaptive designs with three target response rates

Clinical Trial Phases

Drug is accepted for marketing

| Phase I Safety | Phase II Activity | Phase III Evaluation | Phase IV Long-term monitor |

Phase	No. of Patients	Length of Phase	Goal	Success rate
1	Small (20-100)	Short (several months)	-Dose finding trials -Safety, dosages (Maximum tolerated dose (MTD)), efficacy	70%
2	Small (30-500)	Short (several months to 2 years)	-Screening trials -Effectiveness and short-term safety	33%
3	Large (500-3000)	Longer (1-4 years)	-Safety and effectiveness	25-30%
4	Huge (>3000)	Long-long (>20 years)	-post-marketing monitor -long-term safety and rare adverse effects	70-90%

Phase II Trials

- Provide initial assessment of efficacy or 'clinical activity'
  - Screen out ineffective drugs
  - Identify promising new drugs for further evaluation
- Further define safety and toxicity
- Minimize cost of the trial
  - Minimize number of patients exposed to an ineffective treatment
  - Enroll as few patients as "necessary" to show benefit or failure

Phase II study

- Single-arm phase II study (Phase IIA)
  - Response rate is often used as its primary end point
  - Small number of patients enrolled
  - Reliance on historical controls for an estimation of expected response rate
  - Gehan's design (1961); Simon's two-stage designs (1989); predictive probability design (2008), etc.
- Randomized phase II trial (Phase IIB)
  - Simon et al's ranking and selection randomized design (1985); randomized discontinuation design (2002); Bayesian adaptive designs, etc.

Simon's Two-Stage Designs

• X: the number of responders
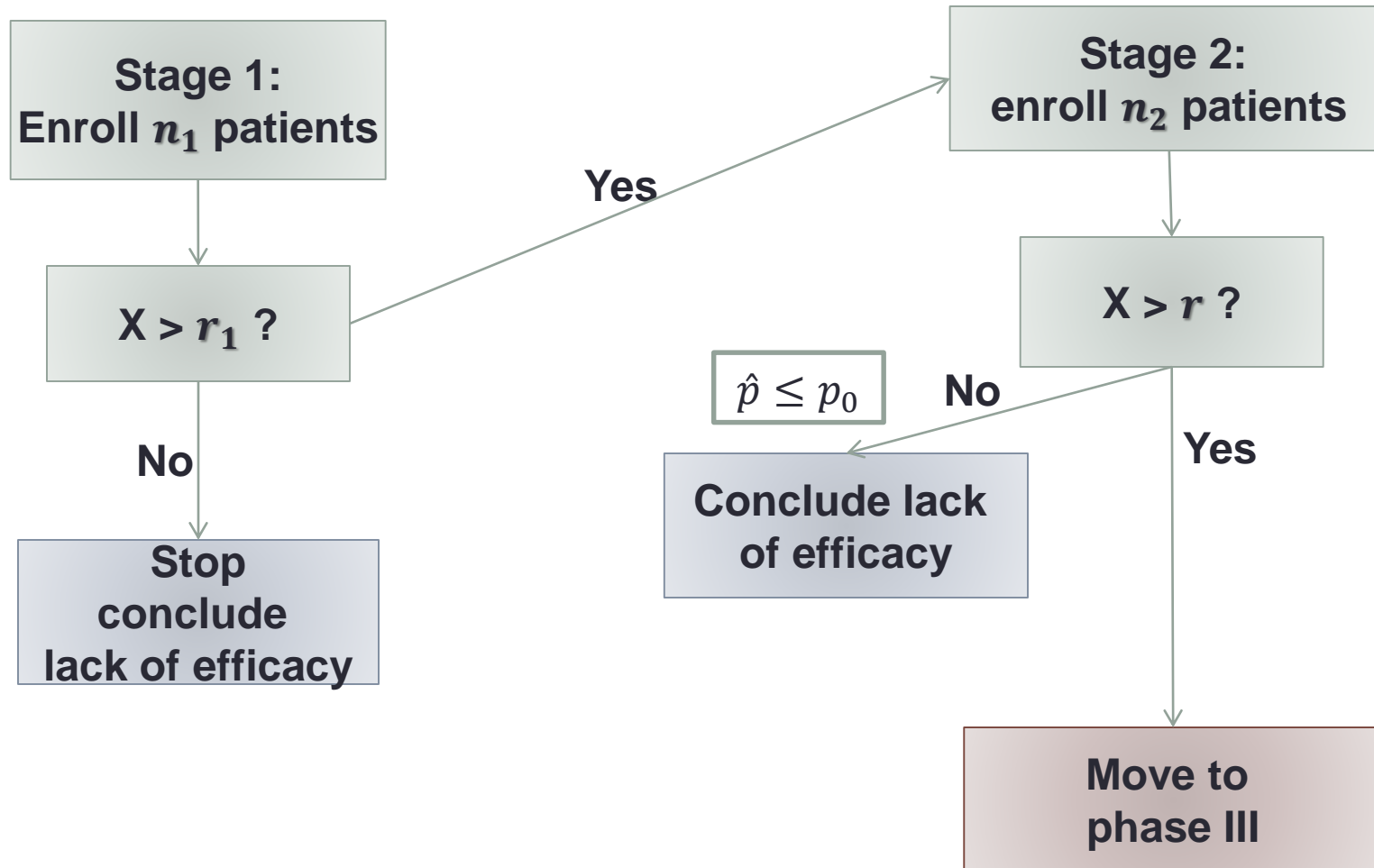
Simon's Two-Stage Designs

- X: the number of responders
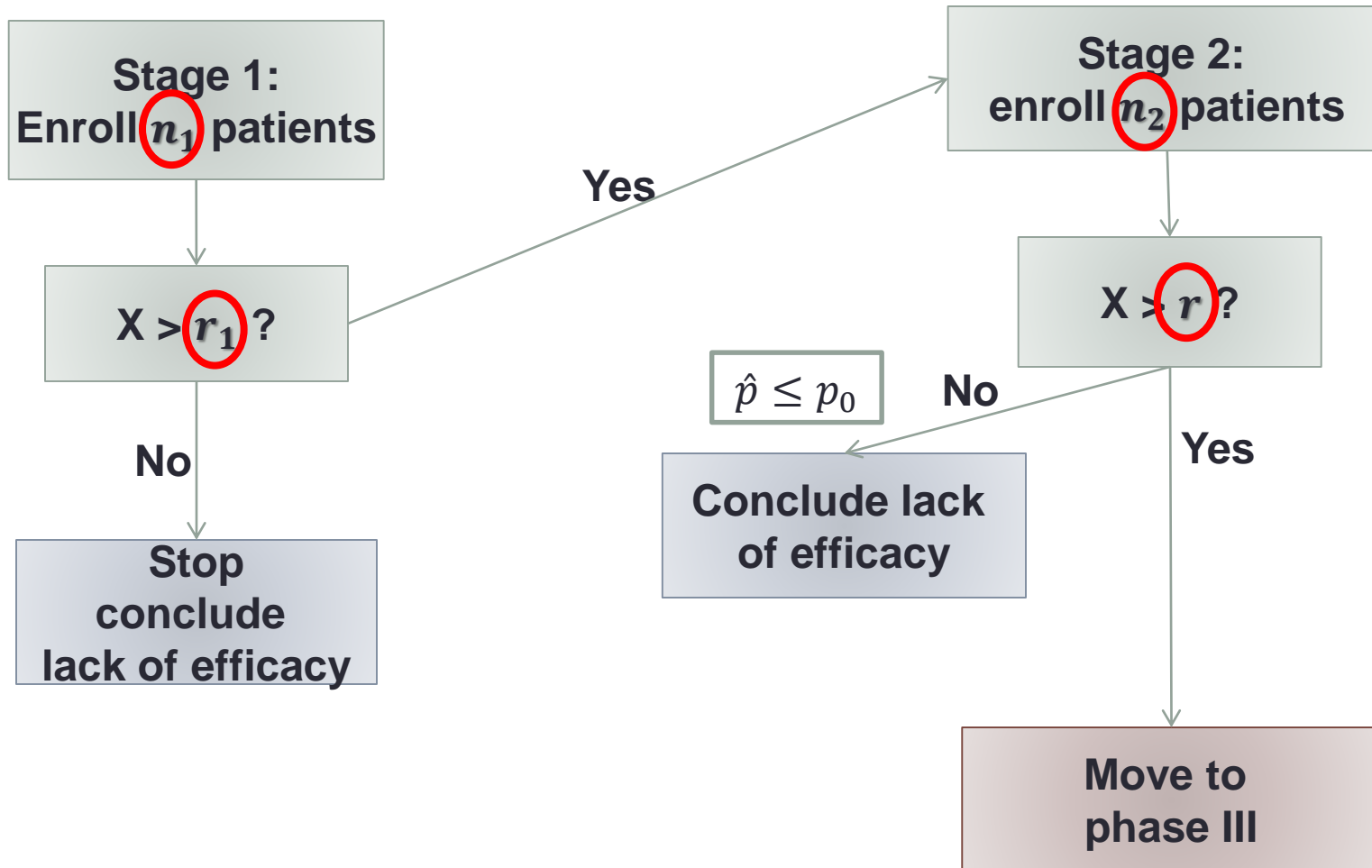
Simon's Two-Stage Designs

- $H_0: p \leq p_0 \ vs. \ H_1: p \geq p_1, where \ p_0 < p_1$
  - $b(x, m, p) \, and \, B(x, m, p)$: the pmf and cdf for $x \sim Bin(m, p)$
  - $G(r_1, n_1, r, n, p)$: the prob of failing to reject H_0

$$= B(r_1, n_1, p) + \sum_{x=r_1+1}^{\min(r, n_1)} b(x, n_1, p) B(r - x, n_2, p)$$

  , where $n = n_1 + n_2$
  - $E(N|p) = n_1 + \left(1 - B(r_1, n_1, p)\right) n_2$: the expected sample size
- Greedy search (look for all cases) given α, β
  - $G(\boldsymbol{r_1, n_1, r, n}, p_0) \geq 1 - \alpha; G(\boldsymbol{r_1, n_1, r, n}, p_1) \leq \beta$
  - Optimum design (min EN_0)
  - Minimax design (min {max$N$})

Adaptive Two-Stage Designs

- Allow the sample size at the second stage to depend on the results at the first stage
  - Lin and Shih (2004)
  - Banerjee and Tsiatis (2006)

Lin and Shih (2004)

- p_0: the maximum uninteresting response rate
- p_1, p_2: two choices of the target response rates, where $p_0 < p_1 < p_2$

- n_1 patients will be enrolled to the first stage
- #(patients) for the second stage will depend on the number of observed responses in the first stage

Lin and Shih (2004)

- X: the number of observed responders
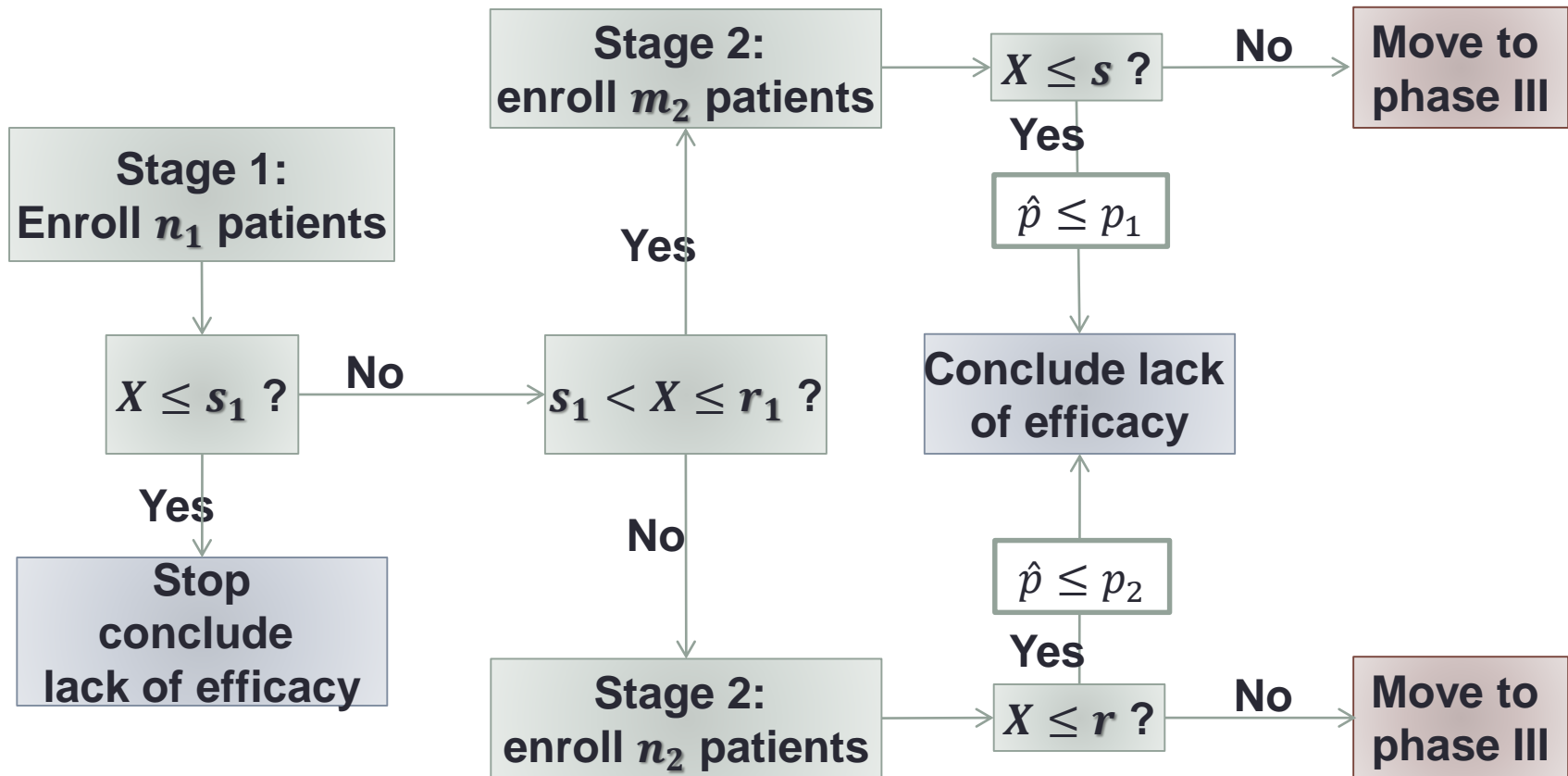
Lin and Shih (2004)

- X: the number of observed responders
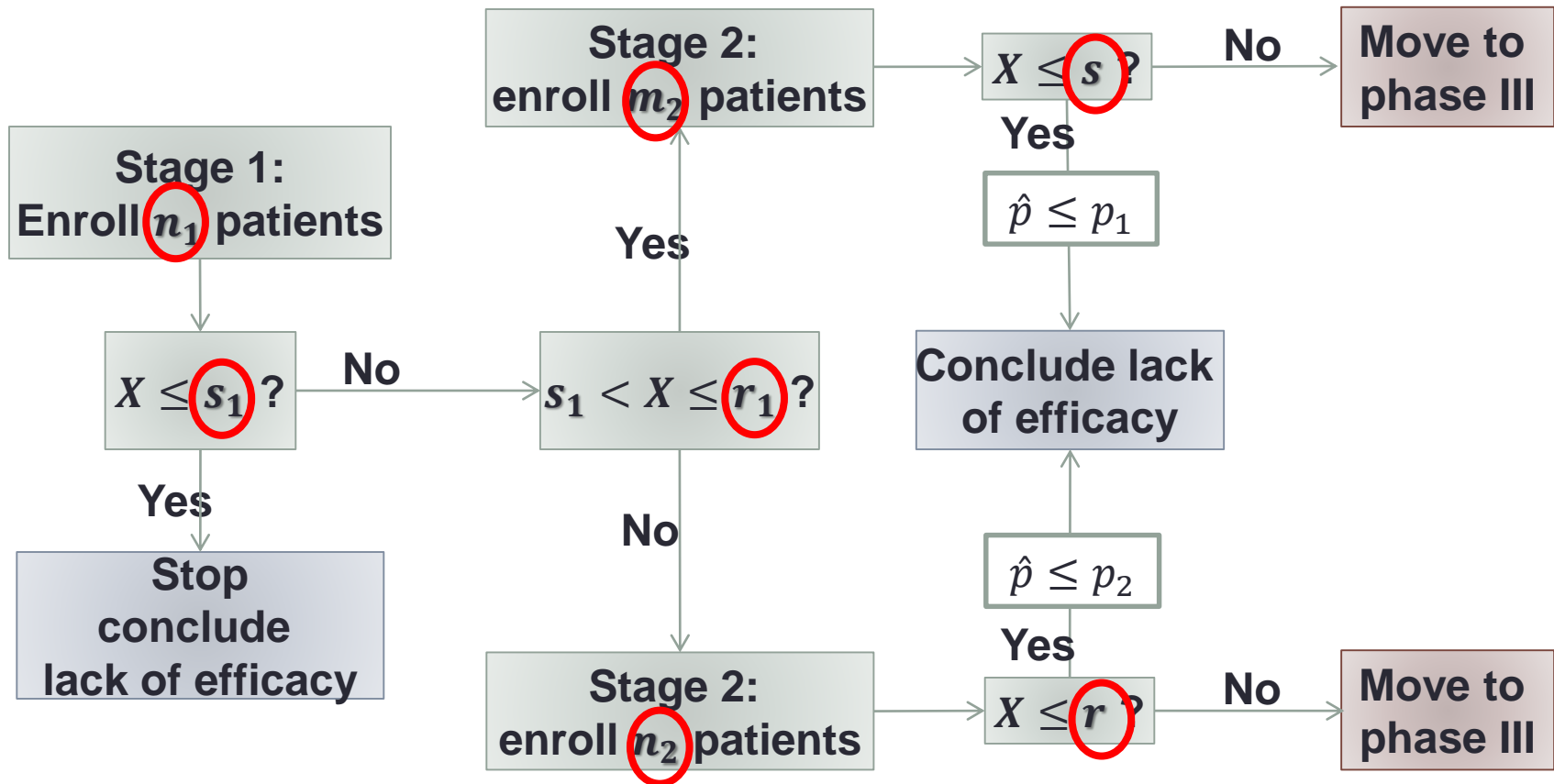
Lin and Shih (2004)

- $H_0: p \leq p_0 \ vs. \ H_1: p \geq p_1 \ or \ H_1: p \geq p_2 \ where \ p_0 < p_1 < p_2$
  - $b(x, m, p) \ and \ B(x, m, p)$: the pmf and cdf for $x \sim Bin(m, p)$
  - $G(s_1, r_1, n_1, s, m, r, n, p)$: the prob of failing to reject H_0

$$= B(s_1, n_1, p) + \sum_{x=s_1+1}^{\min(r_1, s)} b(x, n_1, p) B(s - x, m_2, p)$$

$$+ \sum_{x=r_1+1}^{\min(r, n_1)} b(x, n_1, p) B(r - x, n_2, p)$$

, where $m = m_1 + m_2$; $n = n_1 + n_2$
  - $E(N|p) = n_1 + \{(B(r_1, n_1, p) - B(s_1, n_1, p))m_2 + (1 - B(r_1, n_1, p))n_2\}$: the expected sample size

- Greedy search (look for all cases) given α, β_1, β_2
  - $G(s_1, r_1, n_1, s, m, r, n, p_0) \geq 1 - \alpha$
  - $G(s_1, r_1, n_1, s, m, r, n, p_1) \leq \beta_1$
  - $G(s_1, r_1, n_1, s, m, r, n, p_2) \leq \beta_2$

Optimality criteria

- $O1: \min\{EN_0\}$

- $O2: \min\left\{\max_i EN_i\right\}$

- $O3: \min\{\max(n, m)\}\ and\ \min\{EN_0\}$

- $O4: \min\{\max(n, m)\}\ and\ \min\left\{\max_i EN_i\right\}$

Motivation

- A single arm two-stage phase II trial to see the effect of head and neck cancer (HNC) on the incidence of obstructive sleep apnea (OSA).

- The maximum incidence rate of snoring and sleep apnea on healthy patients is **16.5% (i.e., $p_0 = 0.165$)**.

- Neither historical nor preliminary data available, except that the incidence rate of OSA will be higher in HNC patients.

- An empirical range of the target response rates, from **24.38% to 39.00%.**

- Simon's two-stage design (80% power and 5% level) ➔ the required sample sizes range from **30 to 197**

- Due to wide range of the target response rates, **Lin and Shih's approach will not be able to cover the great uncertainty**.

Extension to three choices?

- "We do not extend the selection to more than two prefixed possible response rates mainly **due to the complexity in the numerical solutions**, and also because it is **usually adequate** for practitioners to contemplate between two (high./low) choices of p_*."-Lin and Shih (2004)

Extension to three choices?

- "We do not extend the selection to more than two prefixed possible response rates mainly **due to the complexity in the numerical solutions**, and also because it is **usually adequate** for practitioners to contemplate between two (high./low) choices of p_*."-Lin and Shih (2004)

Lin and Shih (2004)
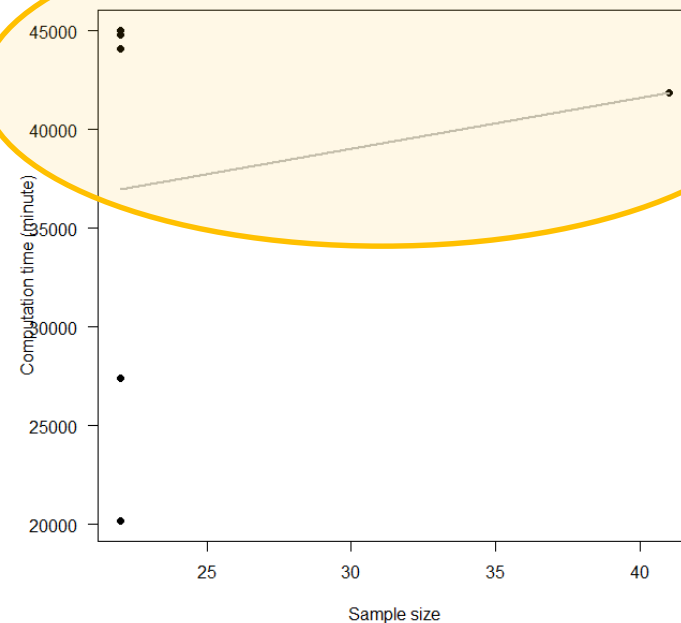
Sample size vs. Computation time

Simon's two-stage
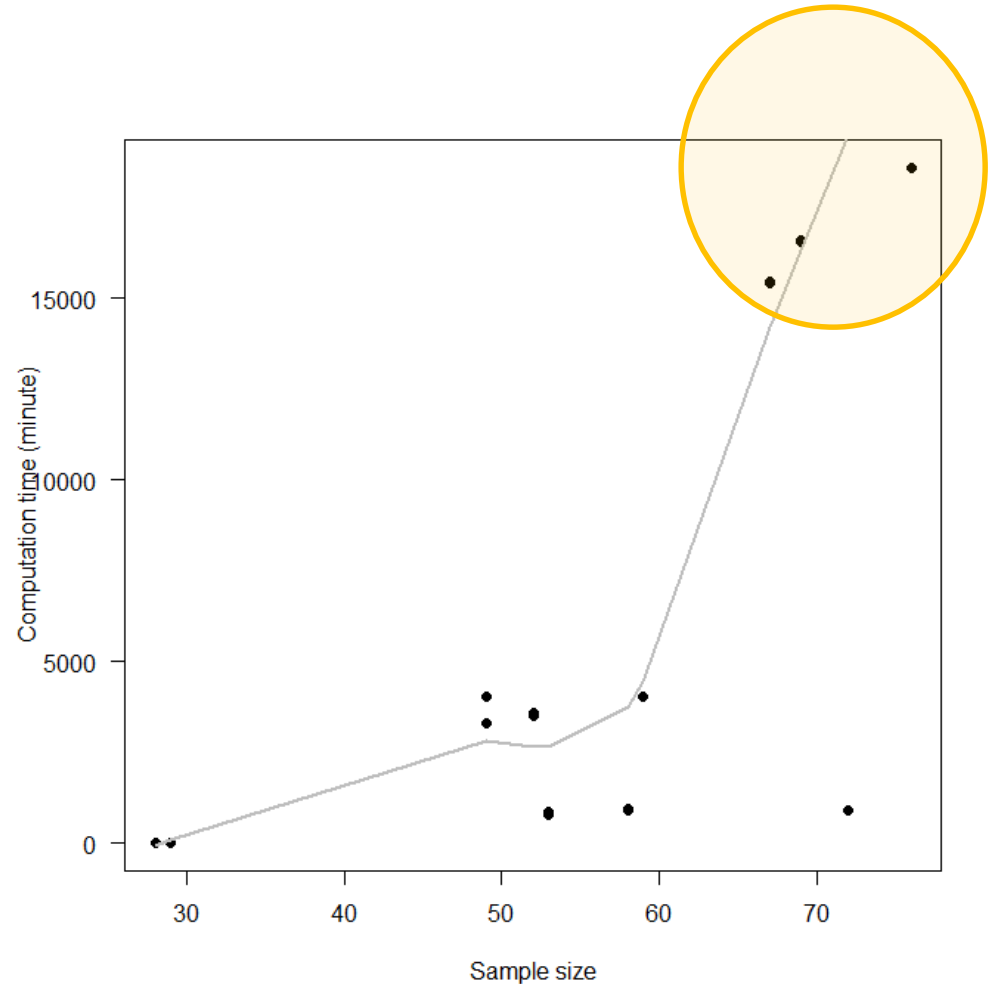
How to reduce the computational burden?

- Nonlinear
- No closed form solution
- Not differentiable

Particle Swarm Optimization (PSO)

$$v^p_{k+1} = w_k v^p_k + c_1 r_1 \left(x^p_{lbest} - x^p_k \right) + c_2 r_2 \left(x_{gbest} - x^p_k \right)$$

$$x^p_{k+1} = x^p_k + v^p_{k+1}$$

Discrete Particle Swarm Optimization (DPSO)

- *k*-th iteration

$$v_{k+1}^p = w_k v_k^p + c_1 r_1 \left(x_{lbest}^p - x_k^p \right) + c_2 r_2 \left(x_{gbest} - x_k^p \right)$$

$$x_{k+1}^p = x_k^p + v_{k+1}^p$$

- *p = 1,…,P; P*: the population size
- x_{lbest}^p and x_{gbest}: local best and global best, respectively
- v_{k+1}^p: the velocity
- w_k: inertia weight

$$w_k = round \left(w_{max} - \frac{k}{K} (w_{max} - w_{min}) \right)$$

- $r_1 c_1, r_2 c_2$: two random sequences in {0,1,2,…,n}; $n = c_1 \, or \, c_2$
- *K*: total number of iterations

Outline

- Pharmacokinetics (PK) analysis
  - Global optimization
  - Identifiability

- Two-stage single-arm phase 2 clinical trial designs
  - Simon's two-stage and Lin and Shih's adaptive designs
  - Adaptive designs with three target response rates

Extension of Lin and Shih (2004)
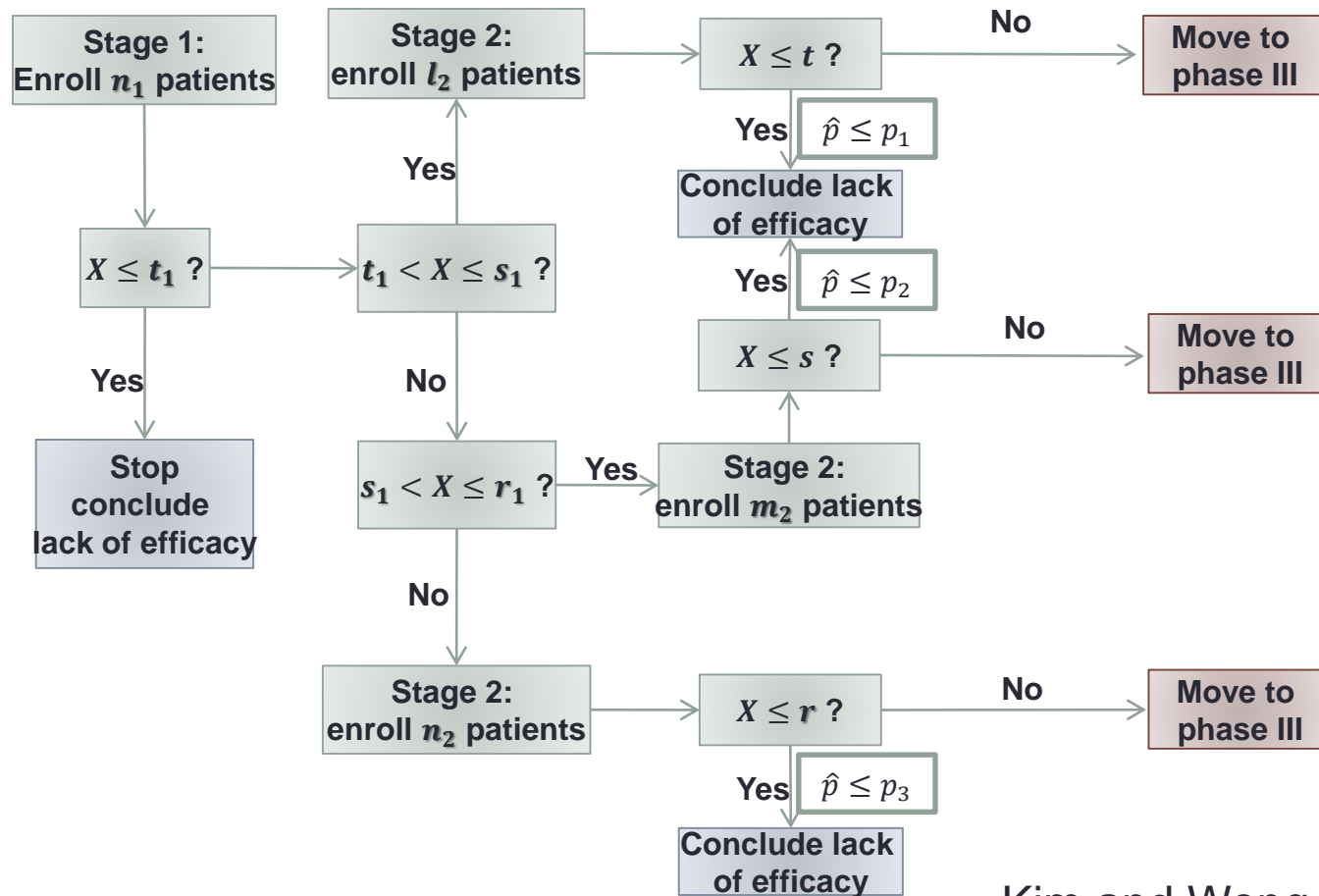
- X: the number of observed responders

Kim and Wong **2017**. *SMMR.*

Extension of Lin and Shih (2004)

- X: the number of observed responders

Kim and Wong **2017**. *SMMR.*
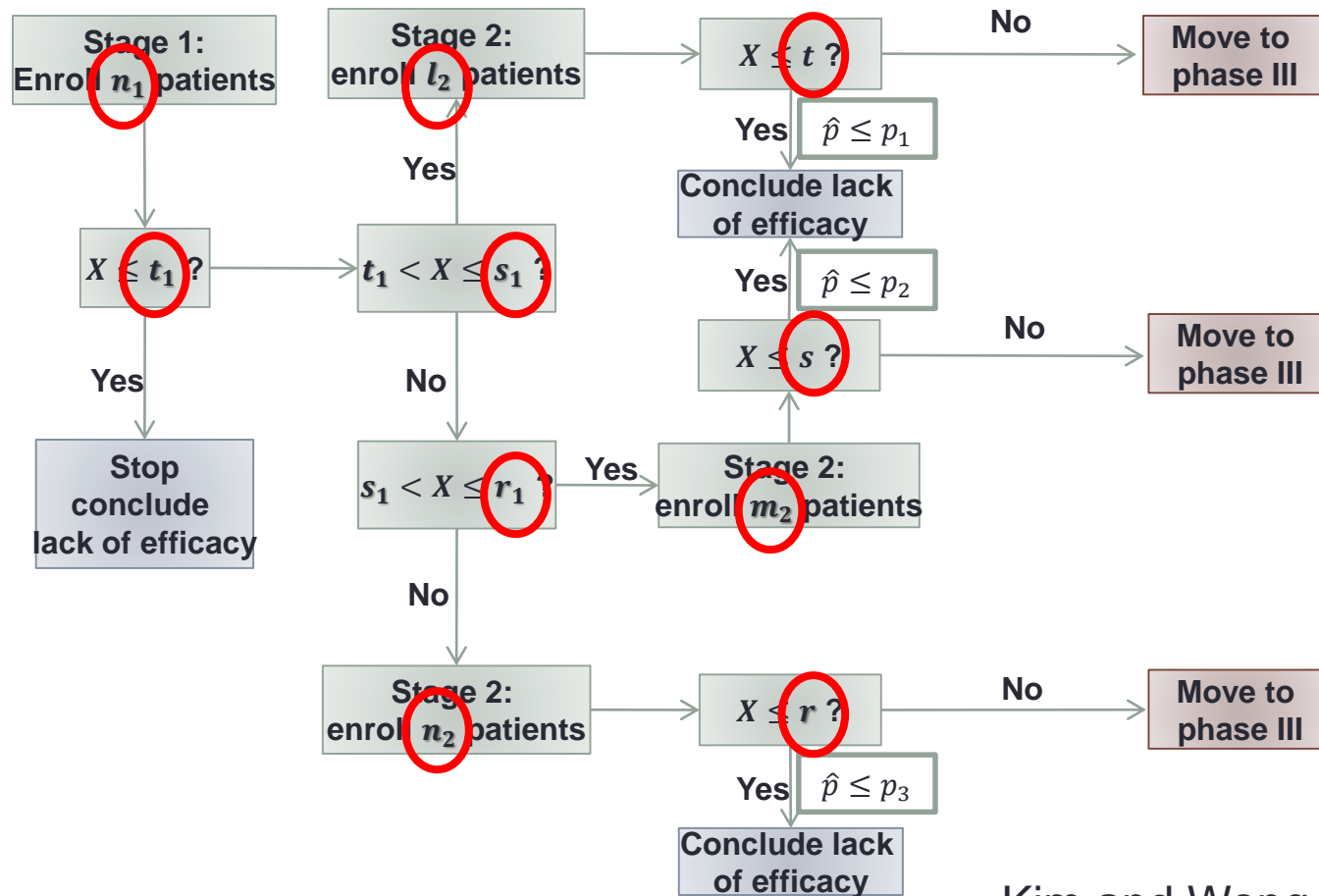
Extension of Lin and Shih (2004)

- $H_0: p \leq p_0 \ vs. \ H_1: p \geq p_1 \ or \ H_1: p \geq p_2 \ or \ H_1: p \geq p_3, where \ p_0 < p_1 < p_2 < p_3$
  - $b(x, m, p) and \ B(x, m, p)$: the pmf and cdf for $x \sim Bin(m, p)$
  - $G(s_1, r_1, n_1, s, m, r, n, p)$: the prob of failing to reject H_0

$$= B(t_1, n_1, p) + \sum_{x=t_1+1}^{\min(s_1, t)} b(x, n_1, p) B(t - x, l_2, p)$$

$$+ \sum_{x=s_1+1}^{\min(r_1, s)} b(x, n_1, p) B(s - x, m_2, p) + \sum_{x=r_1+1}^{\min(r, n_1)} b(x, n_1, p) B(r - x, n_2, p)$$

, where $l = l_1 + l_2; \ m = m_1 + m_2; \ n = n_1 + n_2$

- $E(N|p) = n_1 + \{(B(s_1, n_1, p) - B(t_1, n_1, p))l_2 + (B(r_1, n_1, p) - B(s_1, n_1, p))m_2 + (1 - B(r_1, n_1, p))n_2\}$: the expected sample size

- Greedy search (look for all cases) given $\alpha, \beta_1, \beta_2, \beta_3$
  - $G(s_1, r_1, n_1, s, m, r, n, p_0) \geq 1 - \alpha$
  - $G(s_1, r_1, n_1, s, m, r, n, p_1) \leq \beta_1$
  - $G(s_1, r_1, n_1, s, m, r, n, p_2) \leq \beta_2$
  - $G(s_1, r_1, n_1, s, m, r, n, p_3) \leq \beta_3$

Optimality criteria

- $O1: \min\{EN_0\}$

- $O2: \min\left\{\max_i EN_i\right\}$

- $O3: \min\{\max(l, n, m)\} \text{ and } \min\{EN_0\}$

- $O4: \min\{\max(l, n, m)\} \text{ and } \min\left\{\max_i EN_i\right\}$

Finding initial values

- G-DPSO
  - Within this smaller domain, we searched for an appropriate set of initial values using the same strategy as the greedy search did for the rest of the parameters.

- D-DPSO
  - Use when the number of target responses is two or more.
  - Find the initial set of values using an optimal set of values decided by the case with the one less number of target response.

Table 1. Various adaptive two-stage optimal designs with one target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\|p_0)$	$E(N\|p_1)$
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960
			G-DPSO	0/10	3/29	0.953	0.199	17.624	26.960
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C3	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C4	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
		C3	GS	6/31	15/53	0.950	0.198	40.436	51.983
			G-DPSO	6/31	15/53	0.950	0.198	40.436	51.983
		C4	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
0.55	0.70	C1	GS	15/26	48/76	0.952	0.195	42.021	69.735
			G-DPSO	15/26	48/76	0.952	0.195	42.021	69.735
		C2	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO[a]	20/35	43/67	0.953	0.200	45.802	64.662
		C3	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO	20/35	43/67	0.953	0.200	45.802	64.662
		C4	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO[b]	20/35	43/67	0.953	0.200	45.802	64.662

a,b: The number of particles was increased from 10 to 20 and the population size was increased from 10,000 to 70,000.

Two-stage single-arm phase 2 clinical trial designs

Table 1. Various adaptive two-stage optimal designs with one target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\mid p_0)$	$E(N\mid p_1)$
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960
			G-DPSO	0/10	3/29	0.953	0.199	17.624	26.960
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C3	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C4	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
		C3	GS	6/31	15/53	0.950	0.198	40.436	51.983
			G-DPSO	6/31	15/53	0.950	0.198	40.436	51.983
		C4	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
0.55	0.70	C1	GS	15/26	48/76	0.952	0.195	42.021	69.735
			G-DPSO	15/26	48/76	0.952	0.195	42.021	69.735
		C2	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO[a]	20/35	43/67	0.953	0.200	45.802	64.662
		C3	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO	20/35	43/67	0.953	0.200	45.802	64.662
		C4	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO[b]	20/35	43/67	0.953	0.200	45.802	64.662

a,b: The number of particles was increased from 10 to 20 and the population size was increased from 10,000 to 70,000.

Table S1. Computation time corresponding to **Table 1**; Various adaptive 2-stage optimal designs with one target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	Computation time (minute)
0.05	0.20	C1	GS	0.09
			G-DPSO	2.51
		C2	GS	0.1
			G-DPSO	2.32
		C3	GS	0.09
			G-DPSO	2.35
		C4	GS	0.09
			G-DPSO	2.39
0.20	0.35	C1	GS	15.15
			G-DPSO	2.87
		C2	GS	14.25
			G-DPSO	2.66
		C3	GS	13.04
			G-DPSO	2.59
		C4	GS	13.43
			G-DPSO	2.72
0.55	0.70	C1	GS	310.23
			G-DPSO	3.44
		C2	GS	258.06
			G-DPSO	9.28
		C3	GS	256.76
			G-DPSO	2.93
		C4	GS	257.66
			G-DPSO	9.40

Table S1. Computation time corresponding to **Table 1**; Various adaptive 2-stage optimal designs with one target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	Computation time (minute)
0.05	0.20	C1	GS	0.09
			G-DPSO	2.51
		C2	GS	0.1
			G-DPSO	2.32
		C3	GS	0.09
			G-DPSO	2.35
		C4	GS	0.09
			G-DPSO	2.39
0.20	0.35	C1	GS	15.15
			G-DPSO	2.87
		C2	GS	14.25
			G-DPSO	2.66
		C3	GS	13.04
			G-DPSO	2.59
		C4	GS	13.43
			G-DPSO	2.72
0.55	0.70	C1	GS	310.23
			G-DPSO	3.44
		C2	GS	258.06
			G-DPSO	9.28
		C3	GS	256.76
			G-DPSO	2.93
		C4	GS	257.66
			G-DPSO	9.40

Table 2. Various adaptive two-stage optimal designs with two target responses when $\alpha = 0.05$, $\beta_1 = 0.20$ and $\beta_2 = 0.10$.

p_0	p_1	p_2	Optimal criteria	Method	$s_1/r_1/n_1$	s/m	r/n	$1-\alpha$	β_1	β_2	$E(N|p_0)$	$E(N|p_1)$	$E(N|p_2)$
0.05	0.20	0.25	C1	GS	0/1/9	3/30	4/41	0.957	0.199	0.094	17.548	33.383	36.119
				G-DPSO	0/1/10	3/28	3/31	0.953	0.199	0.088	17.481	27.940	29.254
				D-DPSO	0/1/8	4/41	3/36	0.952	0.200	0.107	18.821	32.980	34.532
			C2	GS	0/2/16	3/29	3/22	0.955	0.199	0.082	22.978	24.097	23.249
				G-DPSO	0/2/14	3/29	3/21	0.954	0.200	0.084	21.444	23.925	22.982
				D-DPSO	0/2/14	3/29	3/21	0.954	0.200	0.084	21.444	23.925	22.982
			C3	GS	0/4/9	4/41	5/22	0.960	0.174	0.084	20.831	36.333	37.668
				G-DPSO	0/5/11	3/28	6/25	0.956	0.199	0.083	18.330	26.505	27.179
				D-DPSO	0/5/11	3/28	6/25	0.956	0.199	0.083	18.330	26.505	27.179
			C4	GS	0/2/16	3/29	3/22	0.955	0.199	0.082	22.978	24.097	23.249
				G-DPSO	0/2/15	3/28	3/23	0.957	0.197	0.078	21.796	24.533	24.007
				D-DPSO	0/2/15	3/28	3/23	0.957	0.197	0.078	21.796	24.533	24.007
0.20	0.35	0.40	C1	G-DPSO	4/9/20	17/62	10/36	0.952	0.199	0.069	35.487	53.869	53.499
				D-DPSO	5/10/22	19/72	11/36	0.951	0.199	0.078	35.311	60.004	60.179
			C2	G-DPSO	8/12/37	16/57	13/41	0.950	0.196	0.059	42.913	46.949	44.252
				D-DPSO	5/10/29	16/56	11/36	0.951	0.196	0.055	43.094	46.415	42.638
			C3	G-DPSO	6/11/27	16/58	13/43	0.950	0.200	0.063	35.833	51.406	50.886
				D-DPSO	3/10/21	15/53	11/36	0.950	0.200	0.058	41.131	50.629	49.683
			C4	G-DPSO	8/12/38	16/56	13/43	0.952	0.200	0.057	43.830	47.337	45.224
				D-DPSO	3/10/21	15/53	11/36	0.950	0.200	0.058	41.131	50.629	49.683
0.55	0.70	0.75	C1	G-DPSO	15/20/26	48/76	27/39	0.951	0.200	0.053	41.807	63.720	61.521
				D-DPSO	15/20/26	48/76	27/39	0.951	0.200	0.053	41.807	63.720	61.521
			C2	G-DPSO	24/28/41	47/73	30/45	0.951	0.194	0.039	48.872	55.463	50.268
				D-DPSO	11/17/24	45/70	30/45	0.951	0.199	0.043	57.956	59.756	54.718
			C3	G-DPSO	13/20/25	43/67	25/42	0.951	0.194	0.038	47.730	62.880	61.206
				D-DPSO	14/20/26	43/67	23/39	0.950	0.199	0.041	45.163	59.975	56.926
			C4	G-DPSO	25/30/44	44/68	32/49	0.950	0.197	0.037	51.862	56.540	52.457
				D-DPSO	10/16/21	43/67	28/41	0.950	0.200	0.044	51.907	60.628	57.152

Table S2. Computation time corresponding to **Table 2**; Various adaptive 2-stage optimal designs with two target responses when $\alpha = 0.05$, $\beta_1 = 0.20$ and $\beta_2 = 0.10$.

p_0	p_1	p_2	Optimal criteria	Method	Computation time (minute)
0.05	0.20	0.25	C1	GS	697.41
				G-DPSO	3.49
				D-DPSO	5.07
			C2	GS	335.84
				G-DPSO	3.49
				D-DPSO	4.89
			C3	GS	456.9
				G-DPSO	3.48
				D-DPSO	4.95
			C4	GS	734.74
				G-DPSO	3.4
				D-DPSO	4.92
0.20	0.35	0.40	C1	G-DPSO	10.6
				D-DPSO	5.86
			C2	G-DPSO	10.24
				D-DPSO	5.54
			C3	G-DPSO	11.44
				D-DPSO	5.6
			C4	G-DPSO	11.07
				D-DPSO	5.58
0.55	0.70	0.75	C1	G-DPSO	147.31
				D-DPSO	5.89
			C2	G-DPSO	131.2
				D-DPSO	5.64
			C3	G-DPSO	197.07
				D-DPSO	5.86
			C4	G-DPSO	193.18
				D-DPSO	5.68

Table S2. Computation time corresponding to **Table 2**; Various adaptive 2-stage optimal designs with two target responses when $\alpha = 0.05$, $\beta_1 = 0.20$ and $\beta_2 = 0.10$.

p_0	p_1	p_2	Optimal criteria	Method	Computation time (minute)
0.05	0.20	0.25	C1	GS	697.41
				G-DPSO	3.49
				D-DPSO	5.07
			C2	GS	335.84
				G-DPSO	3.49
				D-DPSO	4.89
			C3	GS	456.9
				G-DPSO	3.48
				D-DPSO	4.95
			C4	GS	734.74
				G-DPSO	3.4
				D-DPSO	4.92
0.20	0.35	0.40	C1	G-DPSO	10.6
				D-DPSO	5.86
			C2	G-DPSO	10.24
				D-DPSO	5.54
			C3	G-DPSO	11.44
				D-DPSO	5.6
			C4	G-DPSO	11.07
				D-DPSO	5.58
0.55	0.70	0.75	C1	G-DPSO	147.31
				D-DPSO	5.89
			C2	G-DPSO	131.2
				D-DPSO	5.64
			C3	G-DPSO	197.07
				D-DPSO	5.86
			C4	G-DPSO	193.18
				D-DPSO	5.68

Table 3. Various adaptive two-stage optimal designs with three target responses when $\alpha = 0.05$, $\beta_1 = 0.20$, $\beta_2 = 0.10$, and $\beta_3 = 0.05$.

p_0	p_1	p_2	p_3	Optimal criteria	Method	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$1-\alpha$	β_1	β_2	β_3	$E(N|p_0)$	$E(N|p_1)$	$E(N|p_2)$	$E(N|p_3)$
0.05	0.20	0.25	0.30	C1	G-DPSO	0/1/4/10	3/28	3/31	5/28	0.953	0.200	0.088	0.037	17.481	27.841	29.020	29.593
					D-DPSO	0/1/2/9	4/36	3/34	3/31	0.959	0.199	0.093	0.044	18.816	30.463	31.375	31.691
				C2	G-DPSO	0/1/2/13	3/31	3/28	3/20	0.951	0.200	0.088	0.036	21.158	23.725	22.613	21.636
					D-DPSO	0/1/2/14	4/36	3/33	3/19	0.952	0.198	0.094	0.043	24.391	24.899	22.847	21.245
				C3	G-DPSO	0/2/5/11	3/28	3/28	6/21	0.956	0.200	0.084	0.033	18.330	26.458	27.042	27.116
					D-DPSO	0/3/4/11	3/29	5/29	5/22	0.953	0.198	0.085	0.034	18.761	27.101	27.437	27.172
				C4	G-DPSO	0/1/2/15	3/28	3/27	3/24	0.958	0.198	0.077	0.026	21.698	24.904	24.615	24.354
					D-DPSO	0/1/2/14	4/32	3/30	3/22	0.960	0.198	0.079	0.028	22.675	25.189	24.130	23.260
0.20	0.35	0.40	0.45	C1	G-DPSO	5/9/10/24	17/61	10/36	11/32	0.952	0.200	0.064	0.017	36.401	48.570	45.349	40.823
					D-DPSO	5/11/13/22	19/72	12/36	14/32	0.951	0.200	0.078	0.028	35.355	62.126	63.957	61.554
				C2	G-DPSO	5/9/10/29	17/60	13/44	11/34	0.951	0.197	0.061	0.015	44.649	45.199	40.615	37.117
					D-DPSO	3/6/8/19	19/67	11/38	9/31	0.950	0.194	0.060	0.016	43.149	47.817	43.504	39.094
				C3	G-DPSO	4/9/13/22	16/57	10/39	16/36	0.951	0.196	0.059	0.014	37.889	50.724	49.244	46.353
					D-DPSO	3/8/9/17	17/62	12/36	11/28	0.951	0.197	0.070	0.024	37.230	54.484	54.003	50.929
				C4	G-DPSO	5/9/10/26	16/57	13/44	11/34	0.950	0.197	0.059	0.013	38.717	46.656	43.284	39.542
					D-DPSO	4/8/11/24	18/64	11/36	12/28	0.952	0.193	0.068	0.026	44.584	48.279	42.940	37.622
0.55	0.70	0.75	0.80	C1	G-DPSO	15/20/21/26	48/76	29/43	23/31	0.951	0.199	0.052	0.010	41.812	63.491	60.658	51.947
					D-DPSO	10/15/16/19	46/72	23/39	18/29	0.951	0.198	0.047	0.008	44.911	62.696	60.681	54.260
				C2	G-DPSO	17/22/23/32	47/73	32/48	24/35	0.950	0.195	0.041	0.005	51.994	54.810	46.623	39.493
					D-DPSO	8/13/14/17	46/72	27/39	21/31	0.951	0.195	0.048	0.009	52.800	62.503	58.360	51.268
				C3	G-DPSO	15/20/21/27	46/72	27/41	24/35	0.951	0.192	0.041	0.005	44.743	59.648	54.638	46.497
					D-DPSO	12/17/18/22	47/74	26/39	24/32	0.950	0.197	0.054	0.014	44.314	63.001	60.007	52.356
				C4	G-DPSO	17/22/23/32	47/73	32/48	24/35	0.950	0.195	0.041	0.005	51.994	54.810	46.623	39.493
					D-DPSO	7/12/14/16	45/70	26/39	19/31	0.951	0.200	0.044	0.006	55.302	60.783	56.534	50.252

Table 3. Various adaptive two-stage optimal designs with three target responses when $\alpha = 0.05$, $\beta_1 = 0.20$, $\beta_2 = 0.10$, and $\beta_3 = 0.05$.

p_0	p_1	p_2	p_3	Optimal criteria	Method	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$1-\alpha$	β_1	β_2	β_3	$E(N\|p_0)$	$E(N\|p_1)$	$E(N\|p_2)$	$E(N\|p_3)$
0.05	0.20	0.25	0.30	C1	G-DPSO	0/1/4/10	3/28	3/31	5/28	0.953	0.200	0.088	0.037	17.481	27.841	29.020	29.593
					D-DPSO	0/1/2/9	4/36	3/34	3/31	0.959	0.199	0.093	0.044	18.816	30.463	31.375	31.691
				C2	G-DPSO	0/1/2/13	3/31	3/28	3/20	0.951	0.200	0.088	0.036	21.158	23.725	22.613	21.636
					D-DPSO	0/1/2/14	4/36	3/33	3/19	0.952	0.198	0.094	0.043	24.391	24.899	22.847	21.245
				C3	G-DPSO	0/2/5/11	3/28	3/28	6/21	0.956	0.200	0.084	0.033	18.330	26.458	27.042	27.116
					D-DPSO	0/3/4/11	3/29	5/29	5/22	0.953	0.198	0.085	0.034	18.761	27.101	27.437	27.172
				C4	G-DPSO	0/1/2/15	3/28	3/27	3/24	0.958	0.198	0.077	0.026	21.698	24.904	24.615	24.354
					D-DPSO	0/1/2/14	4/32	3/30	3/22	0.960	0.198	0.079	0.028	22.675	25.189	24.130	23.260
0.20	0.35	0.40	0.45	C1	G-DPSO	5/9/10/24	17/61	10/36	11/32	0.952	0.200	0.064	0.017	36.401	48.570	45.349	40.823
					D-DPSO	5/11/13/22	19/72	12/36	14/32	0.951	0.200	0.078	0.028	35.355	62.126	63.957	61.554
				C2	G-DPSO	5/9/10/29	17/60	13/44	11/34	0.951	0.197	0.061	0.015	44.649	45.199	40.615	37.117
					D-DPSO	3/6/8/19	19/67	11/38	9/31	0.950	0.194	0.060	0.016	43.149	47.817	43.504	39.094
				C3	G-DPSO	4/9/13/22	16/57	10/39	16/36	0.951	0.196	0.059	0.014	37.889	50.724	49.244	46.353
					D-DPSO	3/8/9/17	17/62	12/36	11/28	0.951	0.197	0.070	0.024	37.230	54.484	54.003	50.929
				C4	G-DPSO	5/9/10/26	16/57	13/44	11/34	0.950	0.197	0.059	0.013	38.717	46.656	43.284	39.542
					D-DPSO	4/8/11/24	18/64	11/36	12/28	0.952	0.193	0.068	0.026	44.584	48.279	42.940	37.622
0.55	0.70	0.75	0.80	C1	G-DPSO	15/20/21/26	48/76	29/43	23/31	0.951	0.199	0.052	0.010	41.812	63.491	60.658	51.947
					D-DPSO	10/15/16/19	46/72	23/39	18/29	0.951	0.198	0.047	0.008	44.911	62.696	60.681	54.260
				C2	G-DPSO	17/22/23/32	47/73	32/48	24/35	0.950	0.195	0.041	0.005	51.994	54.810	46.623	39.493
					D-DPSO	8/13/14/17	46/72	27/39	21/31	0.951	0.195	0.048	0.009	52.800	62.503	58.360	51.268
				C3	G-DPSO	15/20/21/27	46/72	27/41	24/35	0.951	0.192	0.041	0.005	44.743	59.648	54.638	46.497
					D-DPSO	12/17/18/22	47/74	26/39	24/32	0.950	0.197	0.054	0.014	44.314	63.001	60.007	52.356
				C4	G-DPSO	17/22/23/32	47/73	32/48	24/35	0.950	0.195	0.041	0.005	51.994	54.810	46.623	39.493
					D-DPSO	7/12/14/16	45/70	26/39	19/31	0.951	0.200	0.044	0.006	55.302	60.783	56.534	50.252

Table S3. Computation time corresponding **Table 3**; Various adaptive 2-stage optimal designs with three target responses when α = 0.05, β_1 = 0.20, β_2 = 0.10, and β_3 = 0.05.

p_0	p_1	p_2	p_3	Optimal criteria	Method	Computation time (minute)
0.05	0.20	0.25	0.30	C1	G-DPSO	5.49
					D-DPSO	5.75
				C2	G-DPSO	5.32
					D-DPSO	5.64
				C3	G-DPSO	5.42
					D-DPSO	5.69
				C4	G-DPSO	5.3
					D-DPSO	5.66
0.20	0.35	0.40	0.45	C1	G-DPSO	1592.3
					D-DPSO	6.42
				C2	G-DPSO	1581.77
					D-DPSO	6.32
				C3	G-DPSO	1602.55
					D-DPSO	6.33
				C4	G-DPSO	1584.49
					D-DPSO	6.34
0.55	0.70	0.75	0.80	C1	G-DPSO	41962.35
					D-DPSO	6.49
				C2	G-DPSO	41384.16
					D-DPSO	6.39
				C3	G-DPSO	41835.97
					D-DPSO	6.45
				C4	G-DPSO	41498.29
					D-DPSO	6.38

Table S3. Computation time corresponding **Table 3**; Various adaptive 2-stage optimal designs with three target responses when $\alpha = 0.05$, $\beta_1 = 0.20$, $\beta_2 = 0.10$, and $\beta_3 = 0.05$.

p_0	p_1	p_2	p_3	Optimal criteria	Method	Computation time (minute)
0.05	0.20	0.25	0.30	C1	G-DPSO	5.49
					D-DPSO	5.75
				C2	G-DPSO	5.32
					D-DPSO	5.64
				C3	G-DPSO	5.42
					D-DPSO	5.69
				C4	G-DPSO	5.3
					D-DPSO	5.66
0.20	0.35	0.40	0.45	C1	G-DPSO	1592.3
					D-DPSO	6.42
				C2	G-DPSO	1581.77
					D-DPSO	6.32
				C3	G-DPSO	1602.55
					D-DPSO	6.33
				C4	G-DPSO	1584.49
					D-DPSO	6.34
0.55	0.70	0.75	0.80	C1	G-DPSO	41962.35
					D-DPSO	6.49
				C2	G-DPSO	41384.16
					D-DPSO	6.39
				C3	G-DPSO	41835.97
					D-DPSO	6.45
				C4	G-DPSO	41498.29
					D-DPSO	6.38

Figure 2. Expected sample sizes under the null hypothesis for the 4 criteria C1–C4 with 1, 2 or 3 target alternatives estimated by D–DPSO for 3 scenarios (from left to right): (i) $p_0 = 0.05, p_1 = 0.20, p_2 = 0.25, p_3 = 0.30$, (ii) $p_0 = 0.20, p_1 = 0.35, p_2 = 0.40, p_3 = 0.45$, and (iii) $p_0 = 0.55, p_1 = 0.70, p_2 = 0.75, p_3 = 0.80$. Error rates were set at $\alpha = 0.05, \beta_1 = 0.20, \beta_2 = 0.10$ and $\beta_3 = 0.05$
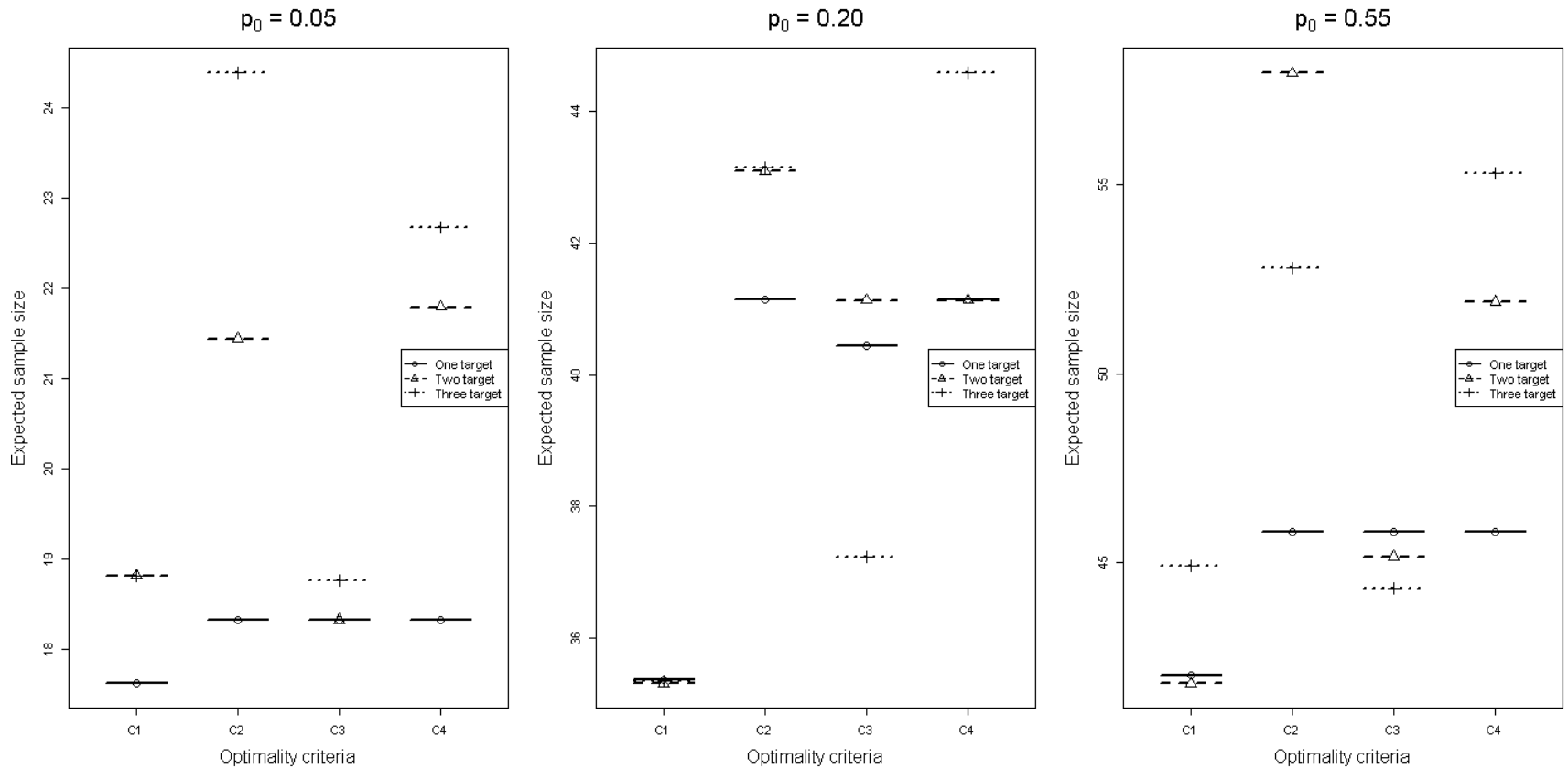
Figure 2. Expected sample sizes under the null hypothesis for the 4 criteria C1–C4 with 1, 2 or 3 target alternatives estimated by D–DPSO for 3 scenarios (from left to right): (i) $p_0 = 0.05, p_1 = 0.20, p_2 = 0.25, p_3 = 0.30$, (ii) $p_0 = 0.20, p_1 = 0.35, p_2 = 0.40, p_3 = 0.45$, and (iii) $p_0 = 0.55, p_1 = 0.70, p_2 = 0.75, p_3 = 0.80$. Error rates were set at $\alpha = 0.05, \beta_1 = 0.20, \beta_2 = 0.10$ and $\beta_3 = 0.05$
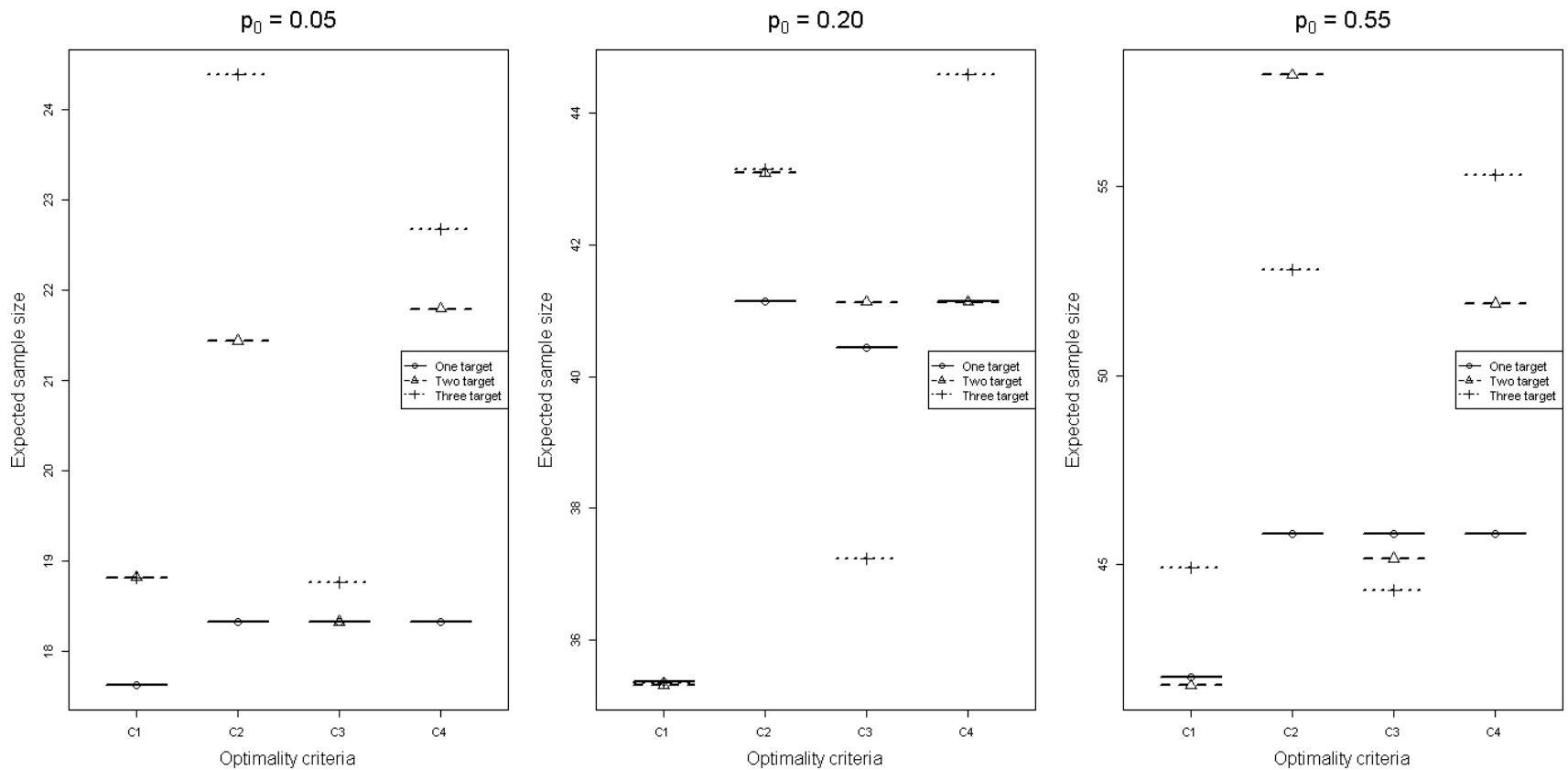
There is **no clear winner** that consistently requires the smallest expected sample sizes

Motivation

- A single arm two-stage phase II trial to see the effect of head and neck cancer (HNC) on the incidence of obstructive sleep apnea (OSA).

- The maximum incidence rate of snoring and sleep apnea on healthy patients is **16.5% (i.e., $p_0 = 0.165$)**.

- Neither historical nor preliminary data available, except that the incidence rate of OSA will be higher in HNC patients.

- An empirical range of the target response rates, from **24.38% to 39.00%.**

- Simon's two-stage design (80% power and 5% level) ➔ the required sample sizes range from **30 to 197**

- Due to wide range of the target response rates, Lin and Shih's approach will not be able to cover the great uncertainty.

Obstructive sleep apnea (OSA)

- To assess the effect of HNC on the incidence of OSA compared to healthy patients.
- p0 = 16.50%, p1 = **24.38%**, p2 = 31.69%, p3 = **39.00%**
- β1 = 0.20, β2 = 0.15, β3 = 0.10

Optimality criteria	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$E(N|p_0)$	$E(N|p_1)$	$E(N|p_2)$	$E(N|p_3)$
OSA								
C1	2/8/9/21	39/188	13/55	10/39	136.92	167.28	157.91	124.73
C2 and C4	1/11/12/25	34/161	20/72	13/36	152.07	158.99	154.07	135.42
C3	1/8/9/17	37/177	12/58	10/35	144.40	166.73	167.70	153.87

Obstructive sleep apnea (OSA)

- To assess the effect of HNC on the incidence of OSA compared to healthy patients.
- $p0 = 16.50\%$, $p1 = $ **24.38%**, $p2 = 31.69\%$, $p3 = $ **39.00%**
- $\beta1 = 0.20$, $\beta2 = 0.15$, $\beta3 = 0.10$

Optimality criteria	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$E(N\mid p_0)$	$E(N\mid p_1)$	$E(N\mid p_2)$	$E(N\mid p_3)$
OSA								
C1	2/8/9/21	39/188	13/55	10/39	136.92	167.28	157.91	124.73
C2 and C4	1/11/12/25	34/161	20/72	13/36	152.07	158.99	154.07	135.42
C3	1/8/9/17	37/177	12/58	10/35	144.40	166.73	167.70	153.87

It can cover the wide variation of the total sample size (35~188)

Phase II study: BREAK-2

- A multicenter, international, single-arm, phase II study (BREAK-2) was carried out to assess the overall response rate of dabrafenib from patients with BRAFV 600E mutationpositive metastatic melanoma (Ascierto et al. (2013)).

- The null hypothesis was set at **p0 = 0.25** and the alternative hypothesis was set at **p1 = 0.40**. The trial wanted to recruit at least **85** patients and the plan was to declare the treatment a success if at least 29 patients responded.

- The efficacy results show that **76** patients with BRAFV 600E mutationpositive metastatic melanoma were enrolled and 45 patients (**59%**) had a confirmed response.

- Although its parent phase I study (Falchook *et al.* (2012)) showed the same type of patients had a response rate of **50%**, this phase II study chose the response rate of **40%** as an alternative hypothesis by lowering the response rate of phase I study.

- However, based on the phase I study, it would be of benefit if the higher response rate was explored in addition to 40% because the final response rate of the phase II study was 59%.

- $p0 = 0.25$, $p1 = 0.40$, $p2 = 0.50$, and $p3 = 0.55$
- $\beta 1 = 0.15$, $\beta 2 = 0.10$, and $\beta 3 = 0.05$

Optimality criteria	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$E(N\mid p_0)$	$E(N\mid p_1)$	$E(N\mid p_2)$	$E(N\mid p_3)$
OSA								
C1	2/8/9/21	39/188	13/55	10/39	136.92	167.28	157.91	124.73
C2 and C4	1/11/12/25	34/161	20/72	13/36	152.07	158.99	154.07	135.42
C3	1/8/9/17	37/177	12/58	10/35	144.40	166.73	167.70	153.87
BREAK-2								
C1	4/10/11/19	26/80	14/44	12/34	51.522	72.216	65.961	58.879
C2 and C4	4/9/12/22	27/83	16/43	13/31	62.079	65.691	50.196	43.063
C3	2/7/9/13	25/77	15/44	14/39	55.526	70.032	66.463	62.186

- p0 = 0.25, p1 = 0.40, p2 = 0.50, and p3 = 0.55
- β1 = 0.15, β2 = 0.10, and β3 = 0.05

Optimality criteria	$s_1/r_1/q_1/n_1$	s/l	r/m	q/n	$E(N\mid p_0)$	$E(N\mid p_1)$	$E(N\mid p_2)$	$E(N\mid p_3)$
OSA								
C1	2/8/9/21	39/188	13/55	10/39	136.92	167.28	157.91	124.73
C2 and C4	1/11/12/25	34/161	20/72	13/36	152.07	158.99	154.07	135.42
C3	1/8/9/17	37/177	12/58	10/35	144.40	166.73	167.70	153.87
BREAK-2								
C1	4/10/11/19	26/80	14/44	12/34	51.522	72.216	65.961	58.879
C2 and C4	4/9/12/22	27/83	16/43	13/31	62.079	65.691	50.196	43.063
C3	2/7/9/13	25/77	15/44	14/39	55.526	70.032	66.463	62.186

- **The total sample sizes at p1 = 0.4 is less than 85 (BREAK-2)**
- **The total sample size at p3 = 0.55 is 31~39, which is at least 46 patients less than 85 (BREAK-2)**

Summary

- A novel and effective nature-inspired stochastic population-based algorithm called discrete particle swarm optimization (DPSO) to find extended two-stage adaptive designs.

- Algorithms based on a greedy search invariably failed to find extended two-stage adaptive designs and an improved version of DPSO, called D-DPSO finds the optimum.

- When the problem is simplified to one or two target response rates, D-DPSO outperformed their peers by a wide margin.

Thank you for being patient!

EM Algorithms

- Expectation (E) step

  - Use current parameters to estimate the missing data

$$Q(\theta) = E_{\mathbf{z}}\left[\log p\left(\mathbf{y}, \mathbf{z} \mid \theta\right) \Big| \mathbf{y}\right] = \int_{-\infty}^{\infty} p\left(\mathbf{z} \mid \mathbf{y}, \theta_n\right) \log p\left(\mathbf{y}, \mathbf{z} \mid \theta\right) d\mathbf{z}$$

- Maximization (M) step

  - Use estimated missing data to perform ML/MAP parameter estimation

$$\theta_{n+1} = \arg\max_{\theta} Q(\theta)$$

- Repeat EM steps, until convergence

EM Algorithms

- Expectation (E) step

  - Use current parameters to estimate the missing data

  $$Q(\theta) = E_{\mathbf{z}}\left[\log p(\mathbf{y}, \mathbf{z} \mid \theta) \,\middle|\, \mathbf{y}\right] = \int_{-\infty}^{\infty} p(\mathbf{z} \mid \mathbf{y}, \theta_n) \log p(\mathbf{y}, \mathbf{z} \mid \theta) \, d\mathbf{z}$$

- Maximization (M) step

  - Use estimated missing data to perform ML/MAP parameter estimation

  $$\theta_{n+1} = \arg\max_{\theta} Q(\theta)$$

- Repeat EM steps, until convergence

EM Algorithms: some limitations

- Its limiting position can strongly depend on its starting position
- Its speed of convergence can be slow
- It can converge to local maxima or saddle points
- Either E-step or M-step is intractable (no closed form available)
  - Pharmacokinetics/pharmacodynamics (PK/PD) models have nonlinear differential equations

- …