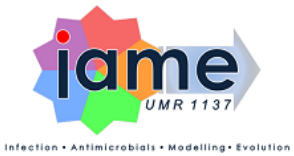
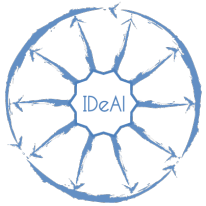
Evaluation of the Fisher information matrix in nonlinear mixed effect models without linearization

Marie-Karelle Riviere, Sebastian Ueckert, and France Mentré

INSERM, IAME, UMR 1137, F-75018 Paris, France;
Univ Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France





Outline

Outline

Longitudinal data

- Longitudinal or repeated measurements
- Several observations per patient across time
- Within and between subject variability

$\Rightarrow$ Non Linear Mixed Effect Models

- Estimation of parameters of the model: by Maximum Likelihood
  $\Rightarrow$ No analytical form of the likelihood
  $\Rightarrow$ Specific software available

Design

- How to design these studies ?
  - Number of subjects
  - Number of points per subject
- Objective: predict the precision in estimation

  $\Rightarrow$ Relative Standard Error (RSE)

FIM in NLMEM

Instead of performing clinical trial simulations
$\Rightarrow$ Evaluation of the Fisher Information Matrix (FIM)

- Inverse FIM = lower bound of the variance-covariance matrix of any unbiased parameters estimator (Cramer-Rao inequality)
- FIM $\Rightarrow$ Evaluate and optimize designs
- In NLMEM: no analytical form
- $\Rightarrow$ First-order (FO) linearization
  Performs very well, but cannot applied to:
  - Complex nonlinear models
  - Discrete data

Mentré et al. Biometrika 84, 429–442 (1997).

Outline

Nonlinear mixed effect models

For continuous data:

$$y_i = f(g(\mu, b_i), \xi_i) + \varepsilon_i$$

For discrete data:

$$p(y_i, \psi | b_i) = \prod_{j=1}^{n_i} h(y_{ij}, g(\mu, b_i), \xi_i)$$

with

- $y_i = (y_{i1}, \ldots, y_{in_i})^T$ response for individual i $(i = 1, \ldots, N)$
- f, h structural model
- ξ_i elementary design for subject i
- g individual parameters vector, function of μ and b_i
- μ vector of fixed effects
- b_i vector of random effects for individual i, $b_i \sim \mathcal{N}(0, \Omega)$
- ε_i vector of residual errors, $\varepsilon_i \sim \mathcal{N}(0, \Sigma)$ and Σ diagonal matrix
- ψ vector of all parameters, $\psi = (\mu, \Omega, \Sigma)$

Fisher Information Matrix (FIM)

Population FIM:

$$M(\psi, \Xi) = \sum_{i=1}^{N} M(\psi, \xi_i)$$

with Ξ population design ($\Xi = (\xi_1, \ldots, \xi_N)^{\mathsf{T}}$)

Individual FIM:

$$M(\psi, \xi) = E\left(\frac{\partial \log(L(y,\psi))}{\partial \psi} \frac{\partial \log(L(y,\psi))}{\partial \psi}^T\right)$$

with the likelihood:

$$L(y, \psi) = \int_b p(y|b, \psi) p(b) db$$

where $p(y|b, \psi)$: conditional density of y given the random effects b

$p(b)$: density of b

Proposed approach

$$M(\psi,\xi) = E\left(\frac{\partial \log(L(y,\psi))}{\partial \psi}\frac{\partial \log(L(y,\psi))}{\partial \psi}^T\right)$$

Proposed approach

$$M(\psi,\xi) = E\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)$$

$$= \int_y \underbrace{\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)}_{A_y} \cdot L(y,\psi)dy$$

Monte Carlo - MC

Proposed approach

$$M(\psi,\xi) = E\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)$$

$$= \int_y \underbrace{\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)}_{A_y} \cdot L(y,\psi)dy$$

Monte Carlo - MC

After calculation... $A_y \iff$

$$\int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k}\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l}\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}db_2$$

Proposed approach

$$M(\psi,\xi) = E\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)$$

$$= \int_y \underbrace{\left(\frac{\partial \log(L(y,\psi))}{\partial\psi}\frac{\partial \log(L(y,\psi))}{\partial\psi}^T\right)}_{A_y} \cdot L(y,\psi)dy$$

Monte Carlo - MC

After calculation... $A_y \iff$

$$\int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2$$

Markov Chains Monte Carlo - MCMC

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)}_{\text{marginal density of } y} dy$$

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)}_{\text{marginal density of } y} dy$$

(I) Draw an R-sample of y from its marginal distribution.

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)}_{\text{marginal density of } y} dy$$

(I) Draw an R-sample of y from its marginal distribution.

(II) For each value of y sampled:

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)dy}_{\text{marginal density of } y}$$

(I) Draw an R-sample of y from its marginal distribution.

(II) For each value of y sampled:

(III) Using MCMC, draw two series of M-samples of b from its conditional density given y.

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)}_{\text{marginal density of } y} dy$$

(I) Draw an R-sample of y from its marginal distribution.

(II) For each value of y sampled:

(III) Using MCMC, draw two series of M-samples of b from its conditional density given y.

(IV) Estimate $\int_{b_1}$ and $\int_{b_2}$ by the mean of the partial derivatives of the conditional log-likelihood taken in the samples of b drawn in step (III).

MC-MCMC algorithm for FIM evaluation

$$\int_y \int_{b_1} \frac{\partial(\log(p(y|b_1,\psi)p(b_1)))}{\partial\psi_k} \underbrace{\frac{p(y|b_1,\psi)p(b_1)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_1 . \int_{b_2} \frac{\partial(\log(p(y|b_2,\psi)p(b_2)))}{\partial\psi_l} \underbrace{\frac{p(y|b_2,\psi)p(b_2)}{\int p(y|b,\psi)p(b)db}}_{\text{conditional density of } b \text{ given } y} db_2 . \underbrace{L(y,\psi)}_{\text{marginal density of } y} dy$$

(I) Draw an R-sample of y from its marginal distribution.

(II) For each value of y sampled:

  (III) Using MCMC, draw two series of M-samples of b from its conditional density given y.

  (IV) Estimate $\int_{b_1}$ and $\int_{b_2}$ by the mean of the partial derivatives of the conditional log-likelihood taken in the samples of b drawn in step (III).

(V) Using MC, estimate $\int_y$ by the mean according to y of the product of the previous partial derivatives.

Partial derivatives of the conditional log-likelihood

$$\frac{\partial \left(\log(p(y|b,\psi)p(b))\right)}{\partial \psi_k}$$

- By hand. For continuous data:

$$\frac{\partial \left(\log(p(y|b,\psi)p(b))\right)}{\partial \psi_k} = -\frac{1}{2}\left[Tr\left(V_b^{-1}\frac{\partial V_b}{\partial \psi_k}\right) - 2(y-E_b)^T V_b^{-1}\frac{\partial E_b}{\partial \psi_k} - (y-E_b)^T V_b^{-1}\frac{\partial V_b}{\partial \psi_k}V_b^{-1}(y-E_b) + Tr\left(\Omega^{-1}\frac{\partial \Omega}{\partial \psi_k}\right) - b^T\Omega^{-1}\frac{\partial \Omega}{\partial \psi_k}\Omega^{-1}b\right]$$

with $E_b = f(g(\mu, b), \xi)$ and $V_b = \Sigma$

- Numerically **for all types of distributions**

STAN* for MCMC

STAN

- Markov Chain Monte Carlo (MCMC) sampler (as JAGS, BUGS, ...)
  - To sample in posterior distributions
  - Based on constructing a Markov chain that has the desired distribution as its stationary distribution
- STAN uses Hamiltonian Monte Carlo (HMC)
  - non-random walk Monte Carlo method vs. Metropolis-Hastings, Gibbs sampling, ...
  - More complex but more efficient, faster convergence
    $\Rightarrow$ Able to overcome some issues inherent in Gibbs sampling
- STAN calculates the gradient of the log probability function (necessary for HMC)

* Stan Development Team. Gelman, Carpenter, ... Columbia University
2014. Stan: A C++ Library for Probability and Sampling, Version 2.5.0. http://mc-stan.org

Outline

FIM evaluation

We compared 3 approaches:

- Linearization (FO) using PFIM 4.0
- AGQ-based approach (AGQ) implemented in R
- MCMC-based approach (MCMC) implemented in R (using rstan)

with clinical trial simulations (CTS):

- Simulate 1000 datasets Y with $\Psi = \Psi_T$ using R
- For each Y: estimate $\hat{\Psi}$ using Monolix 4.3

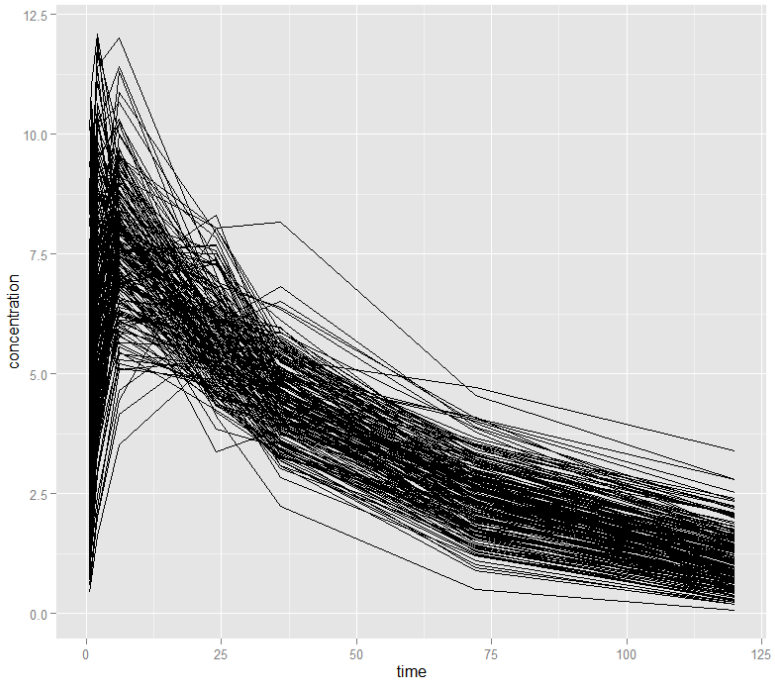
in terms of

- RSE / RRMSE: $RRMSE = \sqrt{\frac{1}{1000}\sum(\hat{\Psi} - \Psi_T)^2}\,/\Psi_T$
- Calculation time

Example 1: PK Warfarin

PKW: One compartment model with first order absorption and elimination:

$$f(\phi = (k_a, V, CL), t) = \frac{70}{V} \frac{k_a}{k_a - \frac{CL}{V}} \left(e^{-\frac{CL}{V} t} - e^{-k_a t} \right)$$



Example 1: PK Warfarin

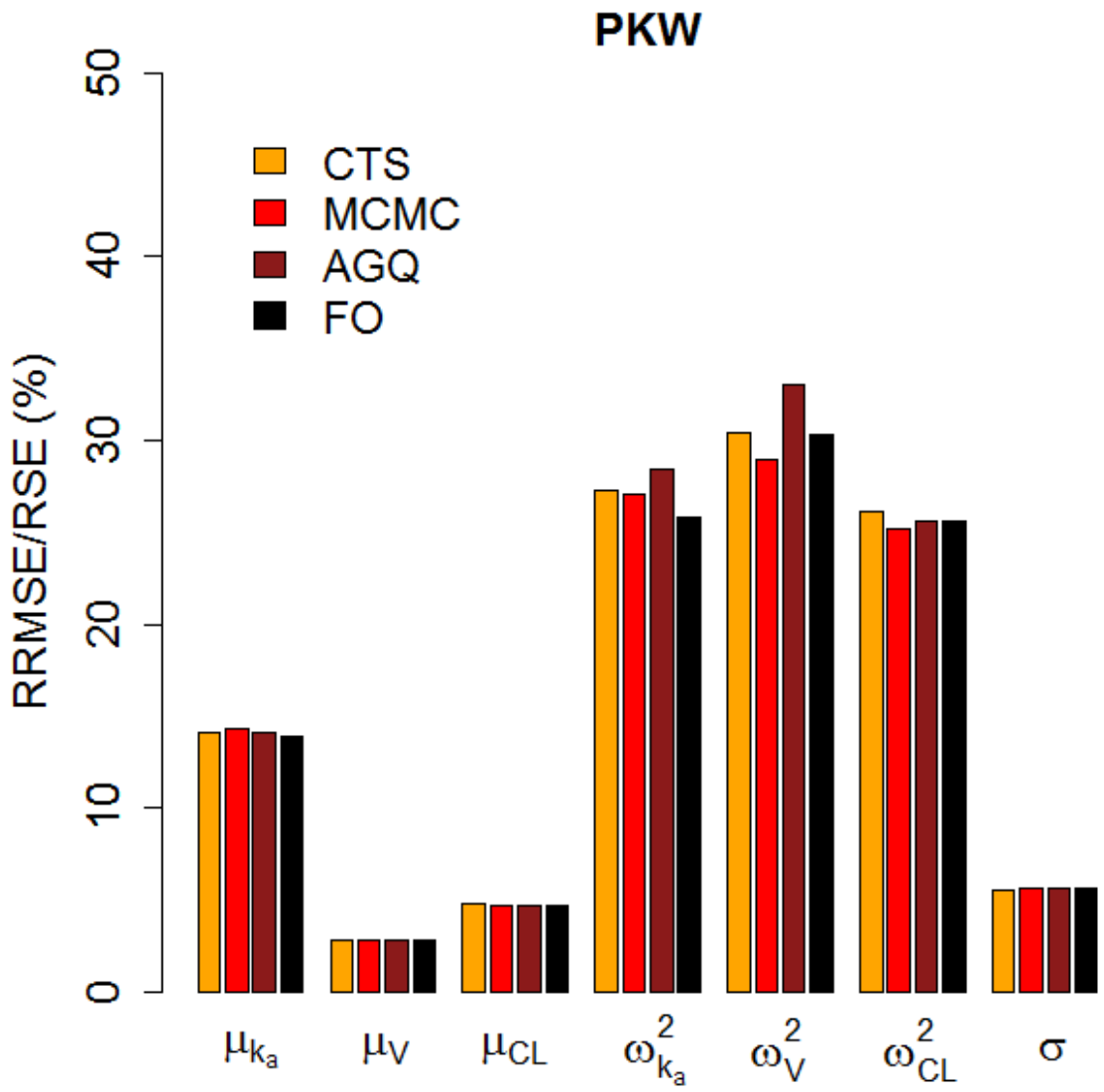
PKW: One compartment model with first order absorption and elimination:

$$f(\phi = (k_a, V, CL), t) = \frac{70}{V} \frac{k_a}{k_a - \frac{CL}{V}} \left(e^{-\frac{CL}{V} t} - e^{-k_a t} \right)$$

- Fixed effects: $(\mu_{k_a}, \mu_V, \mu_{CL}) = (1.00, 8.00, 0.15)$
- Exponential random effects with variances: $(\omega^2_{k_a}, \omega^2_V, \omega^2_{CL}) = (0.60, 0.02, 0.07)$
- Proportional residual error: $\sigma_{\text{slope}} = 0.1$
- 8 times: $t = \xi = (0.5, 1, 2, 6, 24, 36, 72, 120)$
- $N = 32$ patients

Nyberg et al. Br J Clin Pharmacol 79, 6–17 (2015).

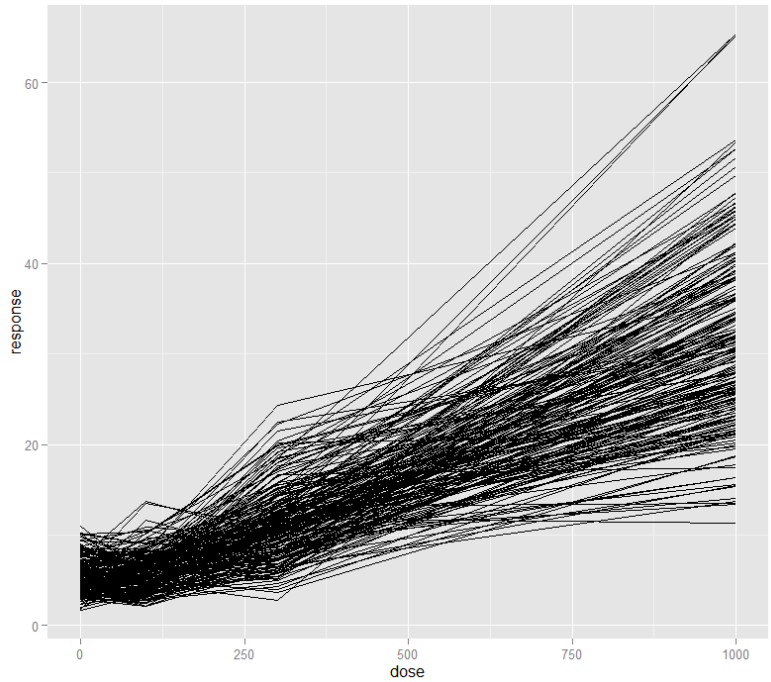
Example 1 - RSE/RRMSE



Example 2: Sigmoïd E_{max} model

SC1: Sigmoïd E_{max} model:

$$f(\phi = (E_0, E_{max}, ED_{50}, \gamma), d) = E_0 + \frac{E_{max} d^{\gamma}}{ED_{50}^{\gamma} + d^{\gamma}}$$



Example 2: Sigmoïd E_{max} model

SC1: Sigmoïd E_{max} model:

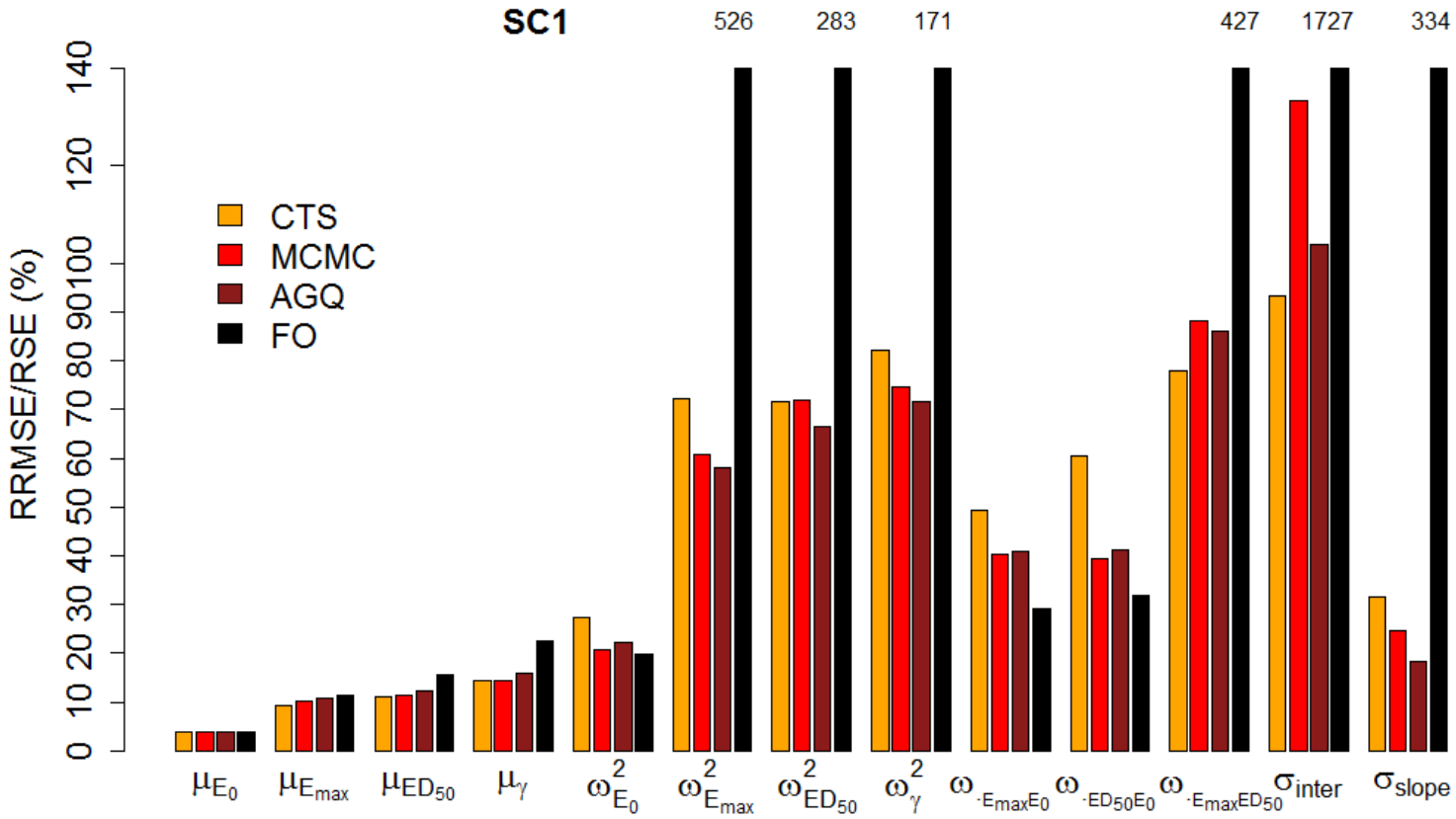
$$f(\phi = (E_{max}, ED_{50}), d) = E_0 + \frac{E_{max} d^\gamma}{ED_{50}^\gamma + d^\gamma}$$

- Fixed effects: $(\mu_{E_0}, \mu_{E_{max}}, \mu_{ED_{50}}, \mu_\gamma) = (5, 30, 500, 3)$
- Exponential random effects with variance-covariance:

$$\Omega = \begin{pmatrix} 0.09 & 0.06 & 0.06 & 0 \\ 0.06 & 0.09 & 0.06 & 0 \\ 0.06 & 0.06 & 0.09 & 0 \\ 0 & 0 & 0 & 0.09 \end{pmatrix}$$

- Combined residual error: $(\sigma_{\text{inter}}, \sigma_{\text{slope}}) = (0.2, 0.2)$
- 4 doses: $d = (0, 100, 300, 1000)$
- $N = 100$ patients

Example 2 - RSE/RRMSE

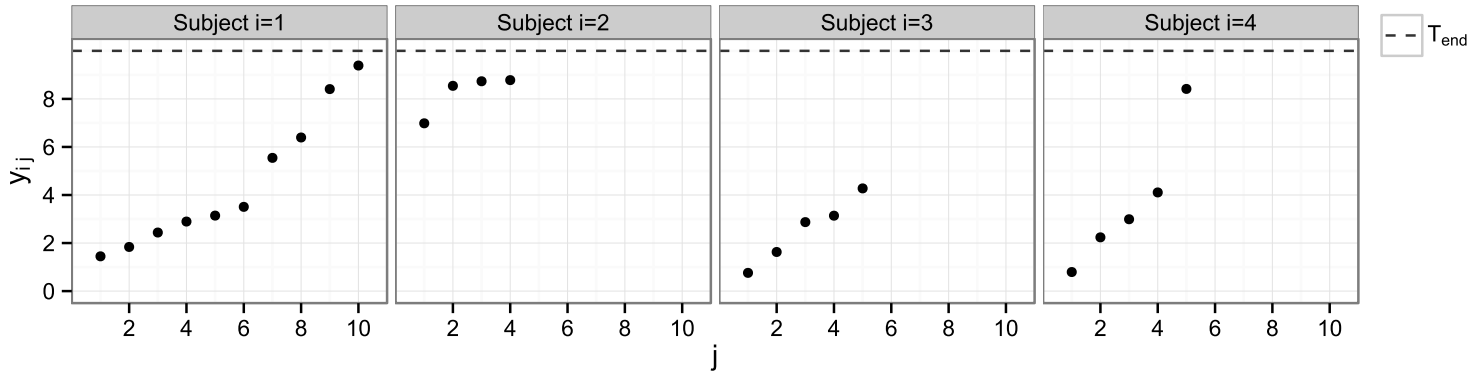


Example 3: Repeated time-to-event

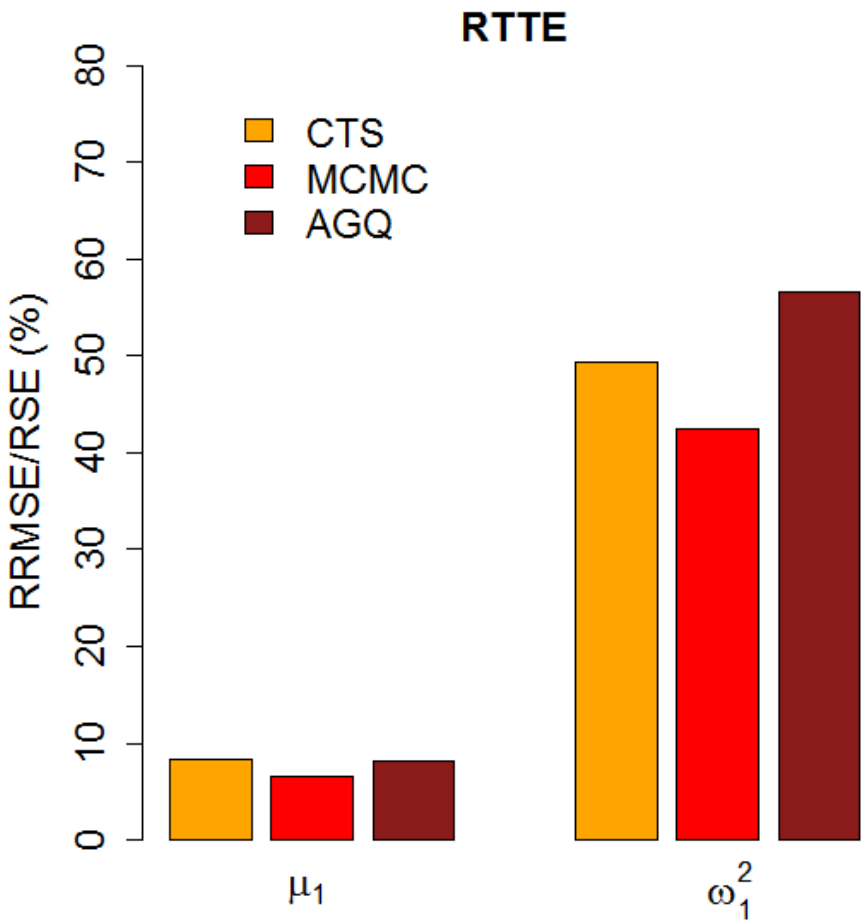
RRTE: Exponential distribution for repeated time-to-event with constant hazard:

$$P(y|b) = \lambda_1 \exp(-\lambda_1 t)$$

- Fixed effects: $\mu_1 = 1.0$
- Exponential random effects: $\lambda_1 = \mu_1 \exp(b)$ with variances: $\omega_1^2 = 0.1$
- Censoring time: 10
- $N = 50$ patients



Example 3 - RSE/RRMSE

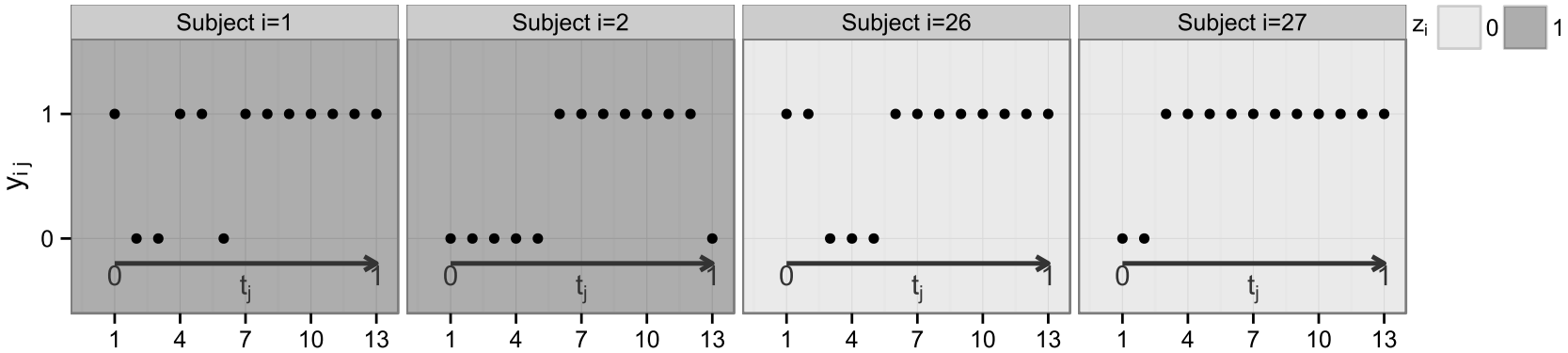


Example 4: Longitudinal binary

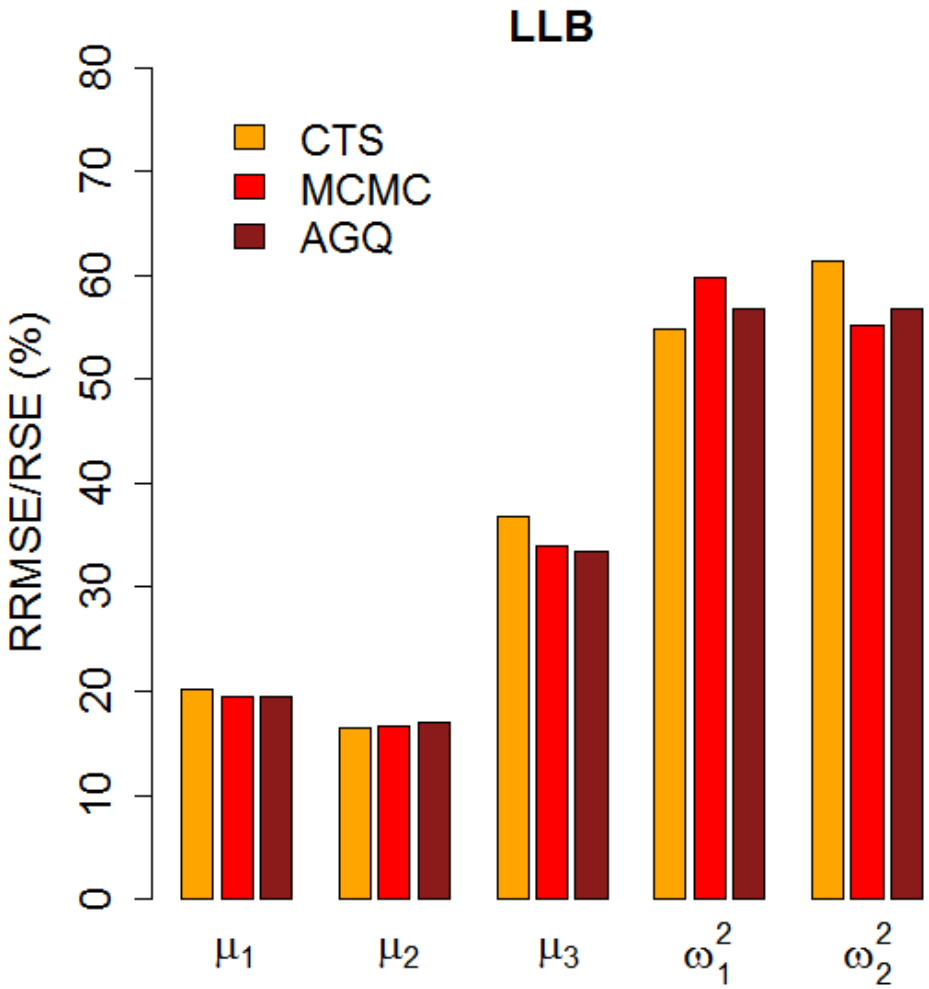
LLB: Probability of response at time t:

$$P(y = 1|b) = \frac{\exp(\beta_1 + \beta_2(1 - \mu_3\delta)t)}{1 + \exp(\beta_1 + \beta_2(1 - \mu_3\delta)t)}$$

- Fixed effects: $(\mu_1, \mu_2, \mu_3) = (-1.0, 4.0, 0.4)$
- Additive random effects with variances: $(\omega_1^2, \omega_2^2) = (0.5, 4.0)$
- 2 groups: $\delta = 0$ and $\delta = 1$
- 13 time points equally spaced between 0 and 1 time units for each patient
- $N = 25$ patients per group

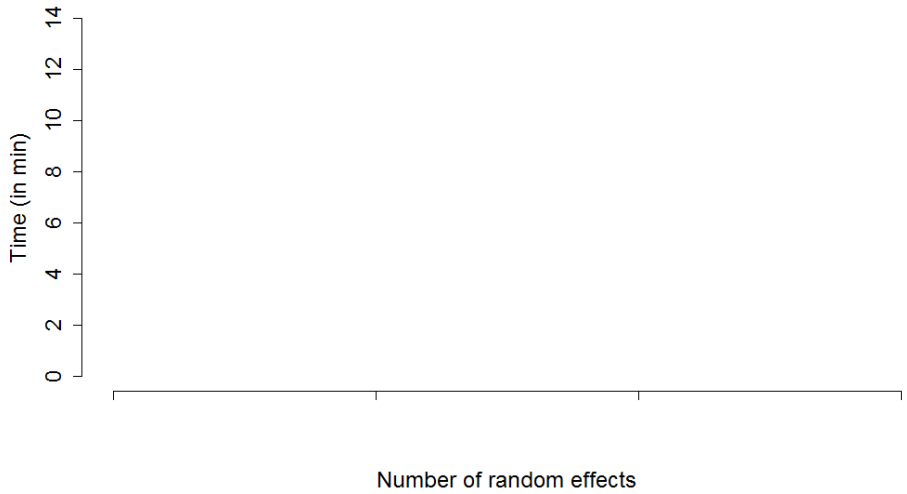


Example 4 - RSE/RRMSE



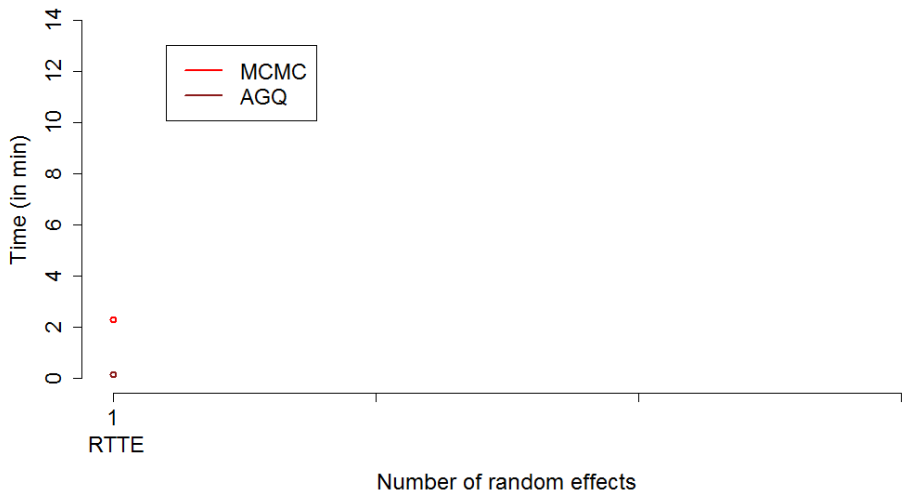
Comparison: calculation time

		PKW	SC1	RTTE	LLB
	CTS	>5h	>5h	>5h	>5h
Time:	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



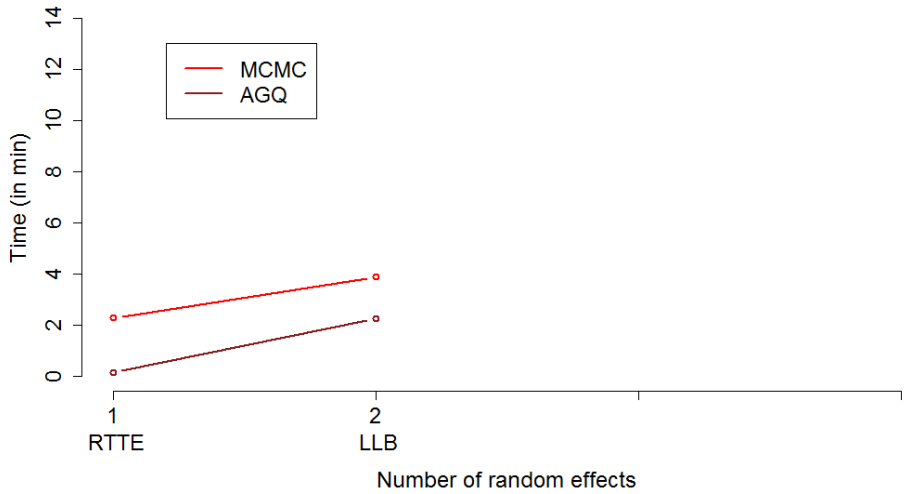
Comparison: calculation time

		PKW	SC1	RTTE	LLB
Time:	CTS	>5h	>5h	>5h	>5h
	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



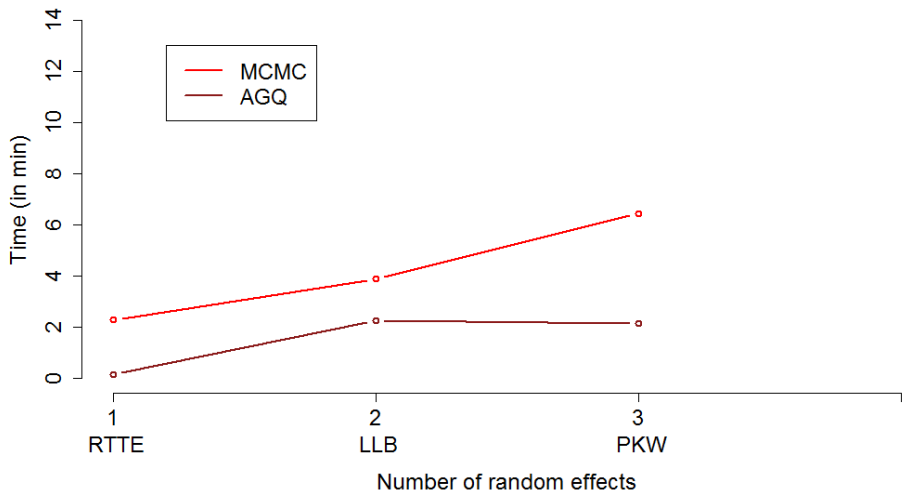
Comparison: calculation time

		PKW	SC1	RTTE	LLB
Time:	CTS	>5h	>5h	>5h	>5h
	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



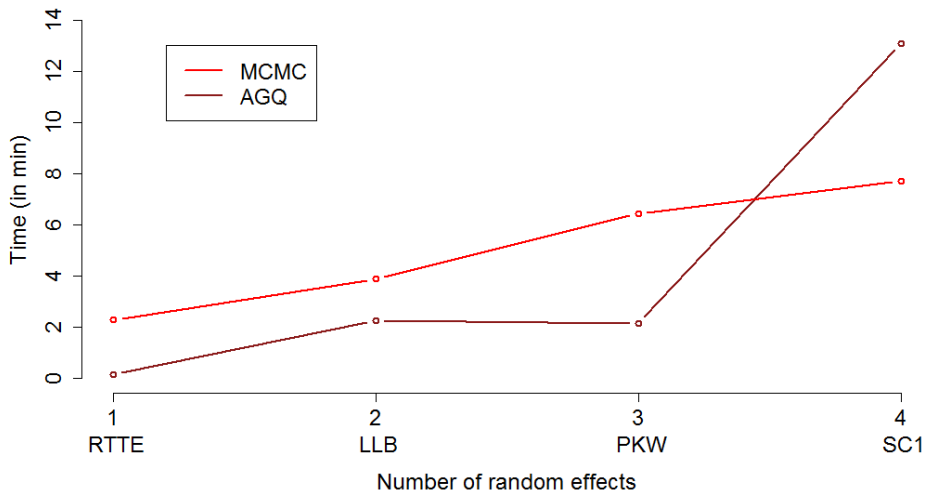
Comparison: calculation time

		PKW	SC1	RTTE	LLB
	CTS	>5h	>5h	>5h	>5h
Time:	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



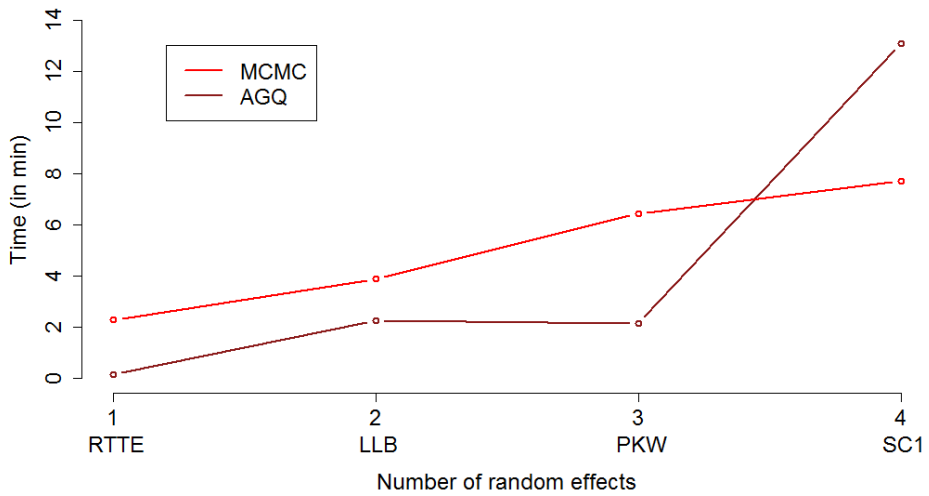
Comparison: calculation time

		PKW	SC1	RTTE	LLB
Time:	CTS	>5h	>5h	>5h	>5h
	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



Comparison: calculation time

		PKW	SC1	RTTE	LLB
Time:	CTS	>5h	>5h	>5h	>5h
	MCMC	≈ 6min	≈ 8min	≈ 2min	≈ 4min
	AGQ	≈ 2min	≈ 13min	≈ 10s	≈ 2min
	FO	<5s	<5s	-	-



- AGQ: time increases exponentially with the number of random parameters
- MCMC: time increases linearly

Outline

Conclusion

- Need of new methods for FIM evaluation
- Developed a MC-MCMC-based method for evaluating FIM
- Advantages:
  - Adapted for discrete and continuous models
  - No model linearization
  - Very high agreement with clinical trial results
- Drawbacks:
  - Much slower than FO approximation
- Perspectives:
  - Publish R package *MIXFIM* on CRAN (*rstan* 07/2015)
    - Show convergence, stochastic error
    - Investigate design optimization
  - Use model averaging to account for model uncertainty

Thank you for your attention!

This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115156, resources of which are composed of financial contributions from the European Union's Seventh Framework Program (FP7/2007 –2013) and EFPIA companies' in kind contribution. The DDMoRe project is also supported by financial contribution from Academic and SME partners. This work does not necessarily represent the view of all DDMoRe partners. This work has also received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.