Bayesian individual dynamic predictions of biomarkers and risk of event in joint modelling (with uncertainty): a comparison between Stan, Monolix and NONMEM

François Riglet, France Mentre and Julie Bertrand

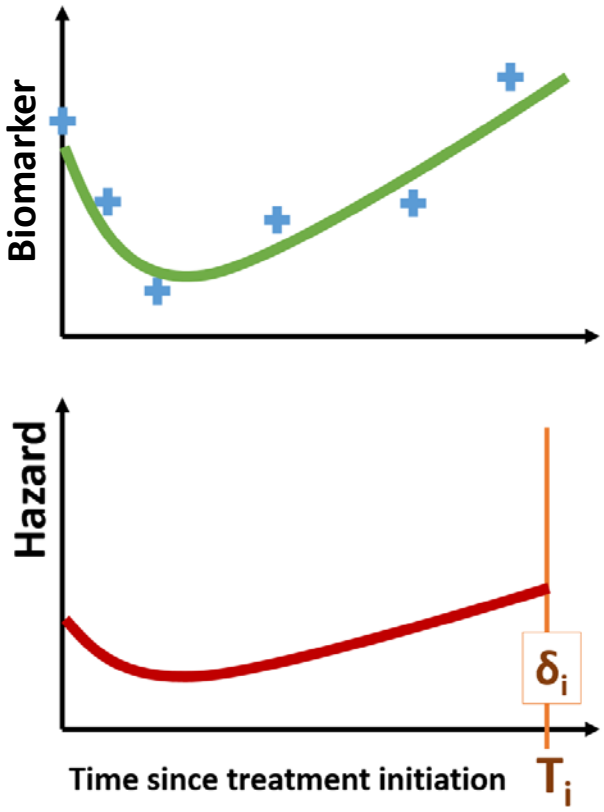
INSERM, IAME, UMR 1137, F-75018 Paris, France; Université Paris Diderot, Sorbonne Paris Cité, France

July 11th, 2019

Joint modeling of longitudinal and survival data

Longitudinal data

- y_i: vector of longitudinal measurements (viral load, bacteria load, lymphocytes, pharmacokinetics, …)
- can be described by a nonlinear model



Time-to-event data

- T_i: observed event time (toxicity, inefficiency, death, …)
- δ_i: event indicator $= \begin{cases} 1 & \text{If event observed} \\ 0 & \text{If event not observed} \end{cases}$

Two objectives

- To characterize the (non-linear) kinetics of a biomarker in presence of a time-to-event
- To characterize the impact of this kinetics on a time-to-event

➔ Reduce bias on biomarker kinetics parameters and potentially those on survival parameters

Nonlinear joint model

2 submodels

→ Longitudinal part – Nonlinear mixed-effects models (NLMEM):

Let $y_i(t)$ be the observed longitudinal data for patient i = {1, …, N}

$$y_i(t) = f(t, \psi_i) + \sigma e_i(t)$$

- f: predictions of the model
- $\psi_i = \mu \times \exp(\eta_i)$: individual parameters
- $e_i \sim \mathcal{N}(0, 1)$: residual errors

with $\mu = v(\mu_1, \ldots, \mu_q)$, Vector of fixed effects
$\eta_i \sim \mathcal{N}(0, \Omega)$, Vector of random effects
where Ω is the variance-covariance matrix of size $q \times q$
with diagonal elements $\{{\omega_1}^2, \ldots, {\omega_q}^2\}$

→ Survival part – Hazard function h for patient i

$$S_i\big(t|f(t,\psi_i)\big) = P(T_i \geq t) = \exp\left[-\int_0^t h_i\left(t|f(t,\psi_i)\right) dt\right]$$

with $h_i(t|f(t,\psi_i)) = h_0(t) \times \exp(\boldsymbol{\beta} \times f(t,\psi_i))$

$$h(t) = \lim_{\Delta t \to 0}\left(\frac{\Pr(t \leq T < t + \Delta t | T \geq t}{\Delta t}\right)$$

➔ Well characterize the association between biomarker kinetics and survival in order to perform dynamic predictions at the individual level and identify high-risk patients

Individual dynamic predictions

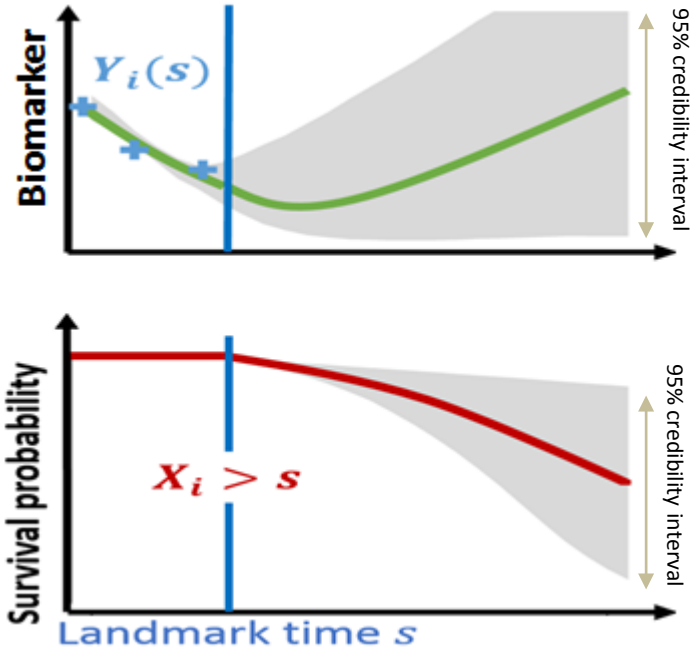
- True joint model is known

→ *Population parameters θ used as priors*

- Landmark time (s): time of interest until which data are observed
- Horizon time (t_{hz}): time of predictions where $t = s + t_{hz}$ with $t_{hz} > 0$

➔ Predict

- $y_i(s+t, Y_i(s), \theta)$
  the longitudinal biomarker predictions with $Y_i(s) = \{y_i(t); 0 \leq t \leq s\}$
- $S_i(s+t|s) = P(T_i > s+t | T_i > s, Y_i(s), \theta)$
  the conditional survival probability with T_i: event time
- For $l = \{1, \ldots, L\}$, L being the number of Monte Carlo samples
  - Draw $\psi_{i,l} \sim \{\psi_i | T_i > s, Y_i(s), \theta\}$
  - Compute $y_{i,l}(s+t)$ and $S_{i,l}(s+t|s)$
  - $\widehat{y_i}(s+t) = median\{y_{i,l}(s+t)\}_{l=1,\ldots,L}$ and $\widehat{\mathrm{CI}}_{0.95}$
  - $\widehat{S_i}(s+t|s) = median\{S_{i,l}(s+t|s)\}_{l=1,\ldots,L}$ and $\widehat{\mathrm{CI}}_{0.95}$



with $\widehat{\mathrm{CI}}_{0.95}$: 95% credibility interval

Rizopoulos (2011) *Biometrics*
Rizopoulos (2012) *Joint Models for Longitudinal and Time-to-Event Data*
Desmée et al. (2017) *BMC Med Res Methodo*

Estimation methods for non linear joint model

- Markov Chain Monte Carlo (MCMC) Algorithms
  - Metropolis-Hastings (MH)[1]
  - Hamiltonian Monte Carlo (HMC)[2]

- Implementation in several modeling software
  - No-U-Turn Sampler (NUTS), a more efficient variant of HMC, since 2014[3]:
    → **Stan(2.18.0)**[4]: a programming language used for Bayesian statistical modeling

  - MH since 2016[5]:
    → **NONMEM7.4®**[6]: the most popular software for NLMEM in population pharmacokinetics and pharmacodynamics (PK/PD), developed at the University of California, San Francisco in the late 1970s
    → **Monolix2018R2®**[7]: a software developed for NLMEM in population PK/PD, developed at INRIA in 2005

[1] **Lavielle** (2014) *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*
[2] **Neal et al.** (2011) *MCMC using Hamiltonian Dynamics, Handbook of Markov Chain Monte Carlo*
[3] **Hoffman & Gelman** (2014) *Journal of Machine Learning Research*
[4] https://mc-stan.org/
[5] **Lavielle & Riba** (2016) *Pharm Res*
[6] **Sheiner** (1980) *J Pharmacokinet Biopharm.*
[7] **Lavielle** (2007) *J Pharmacokinet Pharmacodyn.*

Aim of the study

To compare the abilities of:

- **Stan(2.18.0)** (already validated by Desmée et al.[1])
- **Monolix2018R2®**
- **NONMEM7.4®**

to perform Bayesian individual dynamic predictions of biomarker kinetics and risk of death, with uncertainty, using simulated data

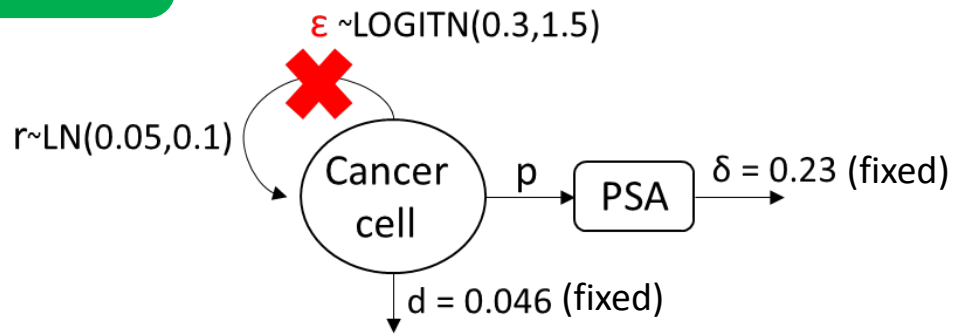
[1] **Desmée et al.** (2017) *BMC Med Res Methodo*

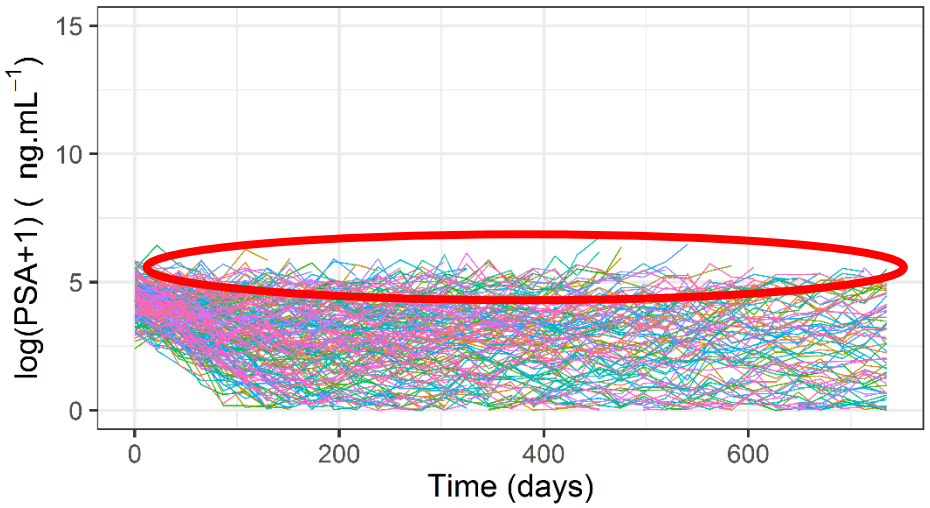
Simulation study design

- Inspired by Desmée et al. (AAPS Journal, 2015) about joint modeling of prostate cancer antigen (PSA) as longitudinal data and overall survival as risk of event

PSA kinetics model

$y_i = PSA(t, \psi_i) + \sigma e_i$ with $\sigma = 0.36$





Survival model

h_0: Weibull

$$h_{i,l}\left(t|PSA(t,\psi_i)\right) = \frac{k}{\lambda} \times \left(\frac{t}{\lambda}\right)^{k-1} \times e^{\beta \times PSA(t,\psi_i)}$$

Survival Parameters	Low Link	High Link	Short Survival
λ (days)	765	**2150**	560
k		**1.5**	
β	0.005	**0.02**	0.02
Survival at the end of the study (%)	25	**40**	5

Desmee et al. (2015) *AAPS Journal*

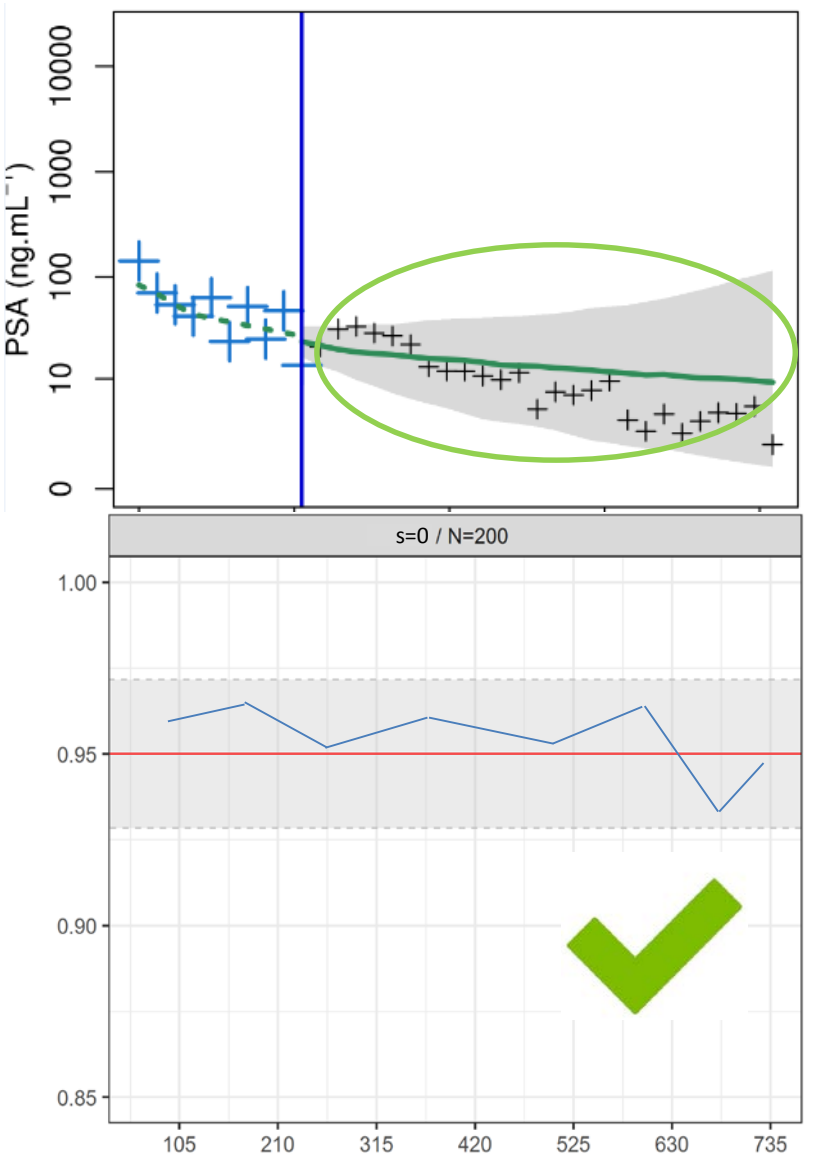
Evaluation at each landmark

- Accuracy and Precision on individual parameters, L = 200 $REE_{i,l}$ per patient i
  - $REE_{i,l} = \frac{\psi_i^* - \widehat{\psi_{i,l}}}{\psi_i^*}$, Relative estimation errors
  - $RBias_i(\%) = 100 \times \left(\frac{1}{L} \times \sum_{l=1}^{L} REE_{i,l}\right)$
  - $RRMSE_i(\%) = 100 \times \sqrt{\frac{1}{L} \times \sum_{l=1}^{L} {REE_{i,l}}^2}$

ψ_i^*: Simulated i parameter
$\widehat{\psi_{i,l}}$: Estimated i parameter

- Individual dynamic prediction plots
  - PSA: $\widehat{y_i}(s+t) = median\{y_{i,l}(s+t)\}_{l=1,\dots,L}$ and $\widehat{\mathrm{CI}}_{0.95}$
  - Survival: $\widehat{S_i}(s+t|s) = median\{S_{i,l}(s+t|s)\}_{l=1,\dots,L}$ and $\widehat{\mathrm{CI}}_{0.95}$

- Coverage rate
  - PSA: $Coverage\ rate(s+t|s) = \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} I_{\{PSA(t,\psi_i^*) \in \widehat{\mathrm{CI}}_{0.95}\}}$
  - Survival: $Coverage\ rate(s+t|s) = \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} I_{\{S(t,\psi_i^*) \in \widehat{\mathrm{CI}}_{0.95}\}}$

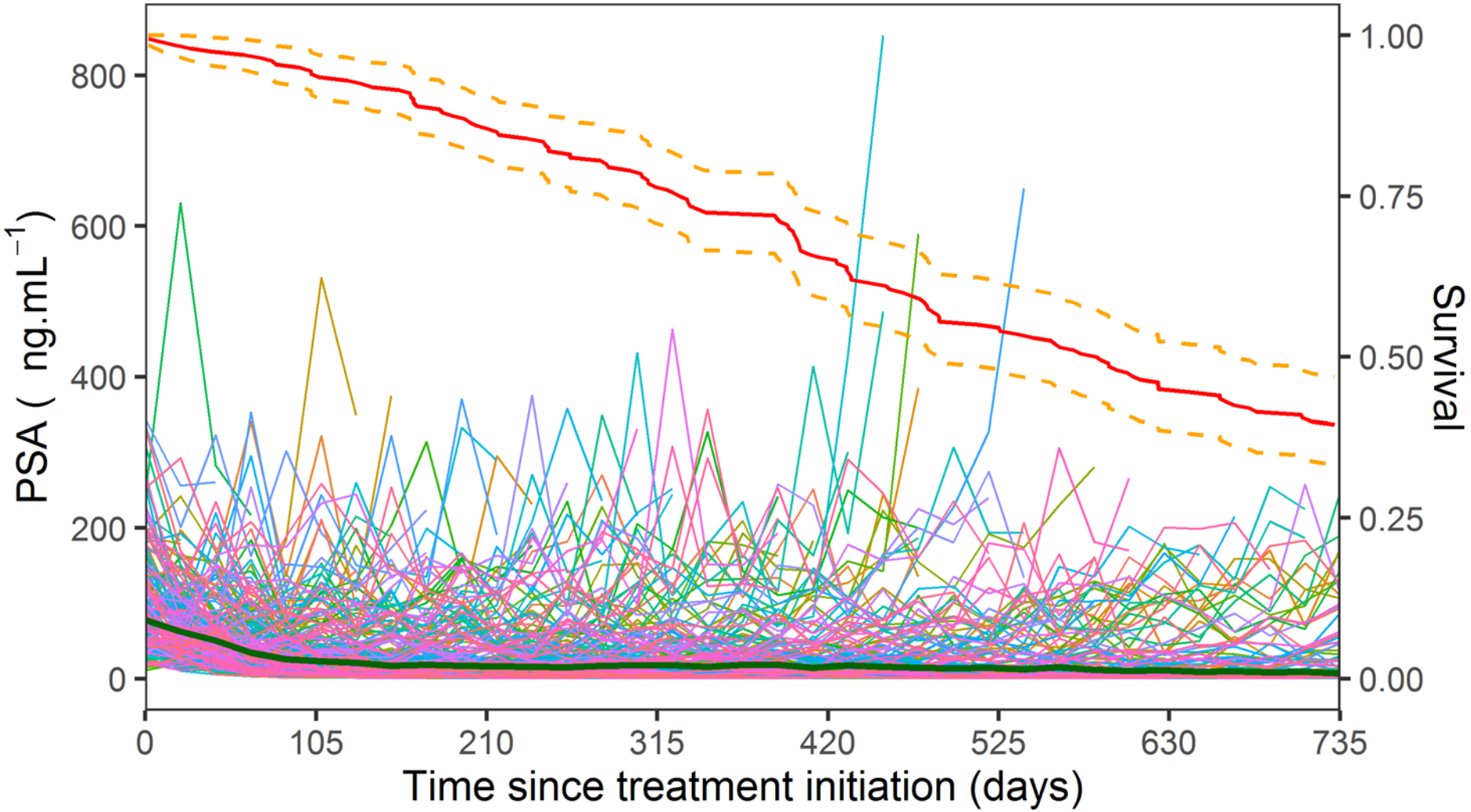


PI95%: 95% prediction interval around 0.95

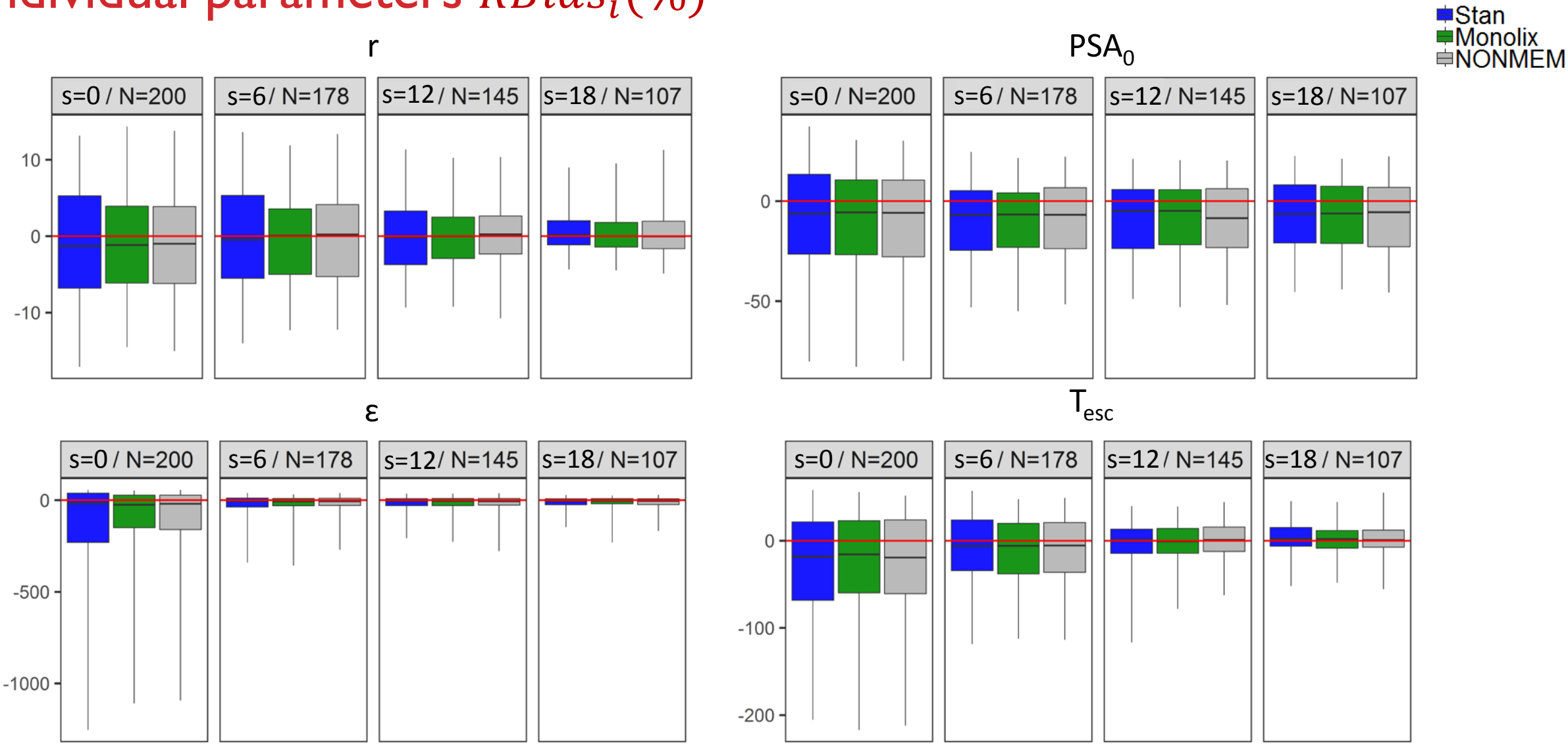
Software settings

Stan(2.18.0)	Monolix2018R2®	NONMEM7.4®
Calibration to reach the target distribution		
Warmup: 500 iterations Chain : 1	ρ_{mcmc} = 1 L_{mcmc} = 500 Chain: 1	CTYPE = 0 NBURN = 500 NITER = 0 Chain = 1
Calibration to obtain L $\psi_{i,l}$		
Thinning: 1 iteration Sampling: L iterations	Simulated parameters per individual: L	EONLY = 1 ISAMPLE = L
Sampling of $\psi_{i,l}$		
→ **L** iterations of the sampling phase using true population parameters, μ and ω, as prior parameters	→ **L** draws from the conditional distribution using true population parameters, μ and ω, as fixed parameters	

High Link scenario data

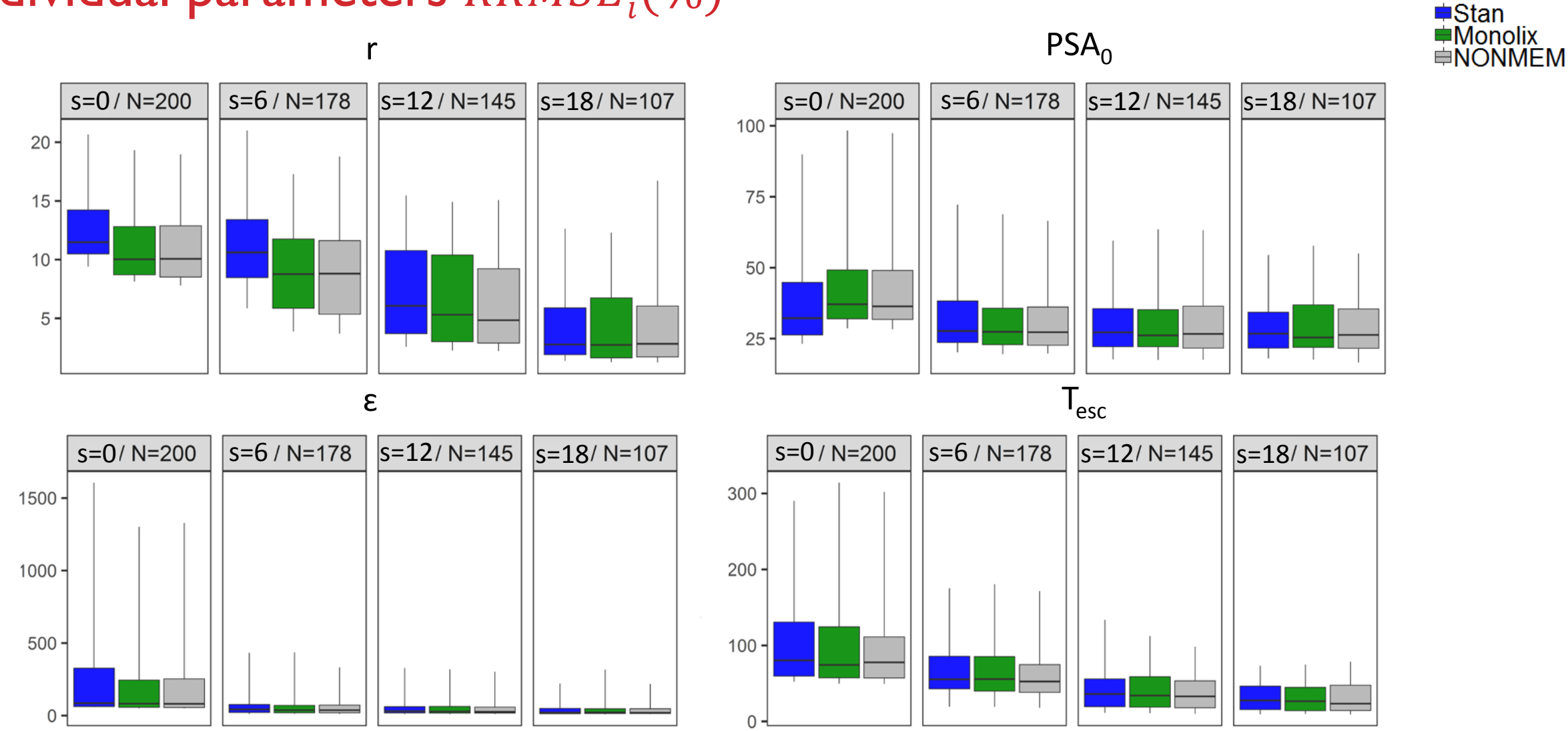


Individual parameters $RBias_i(\%)$



- Low bias on PSA kinetic parameters in early landmarks to be corrected as data are accumulated
- Similar results with all software

Individual parameters $RRMSE_i(\%)$



- Parameters increase in precision as data are accumulated
- Similar results with all software

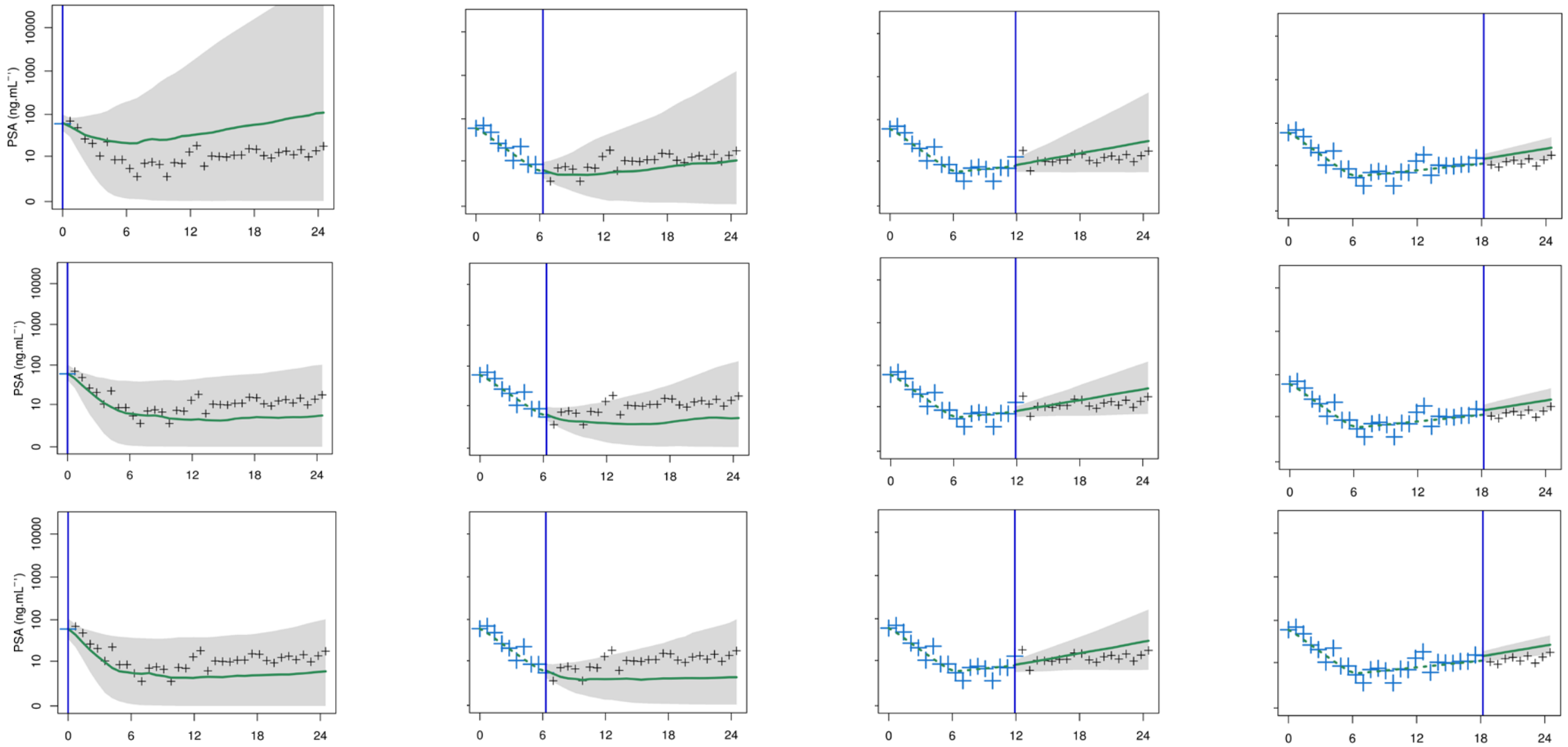
Individual plot: PSA dynamic predictions

Landmark s=0	Landmark s=6	Landmark s=12	Landmark s=18

Stan(2.18.0)

Monolix2018R2®

NONMEM7.4®



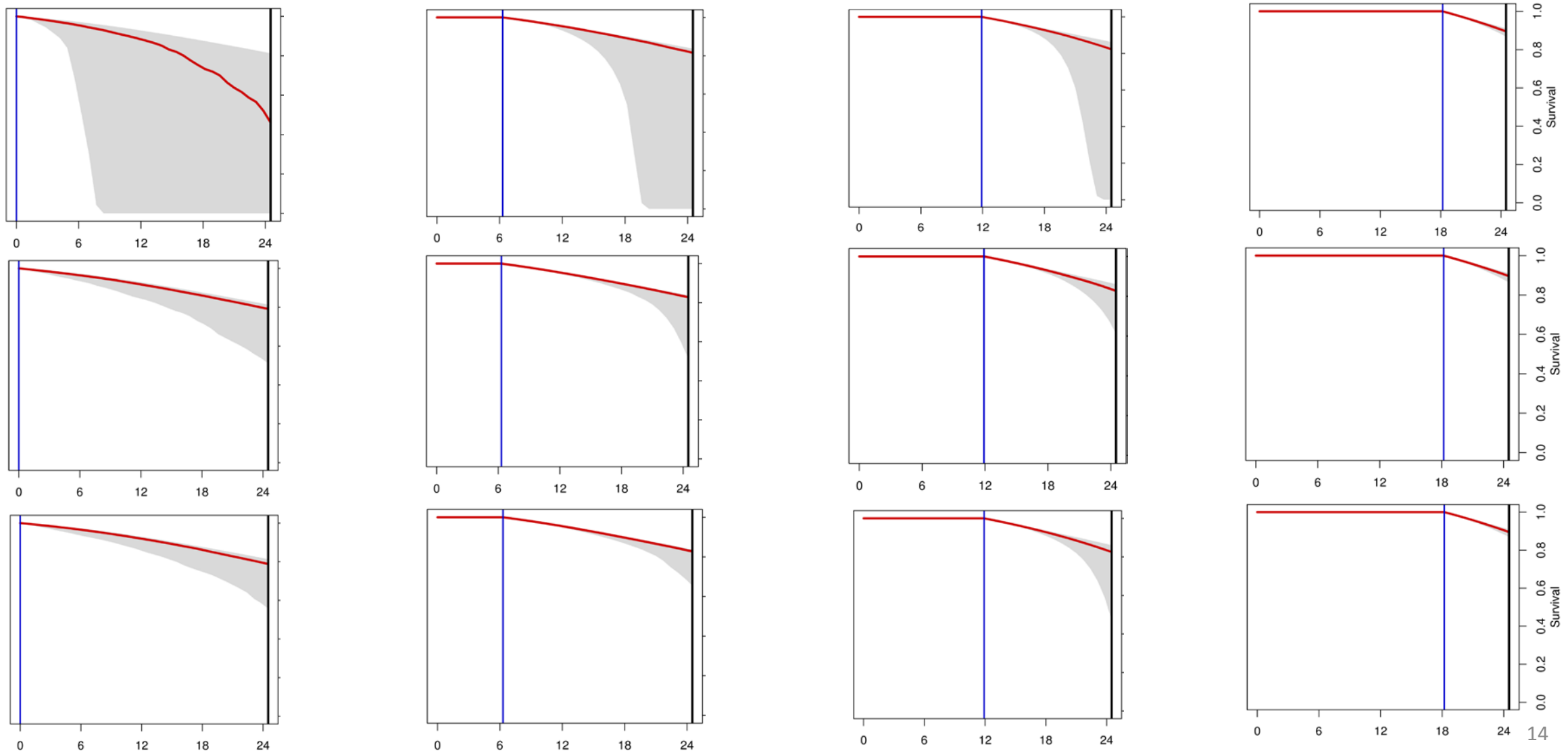
Individual plot: Survival dynamic predictions

Landmark s=0	Landmark s=6	Landmark s=12	Landmark s=18

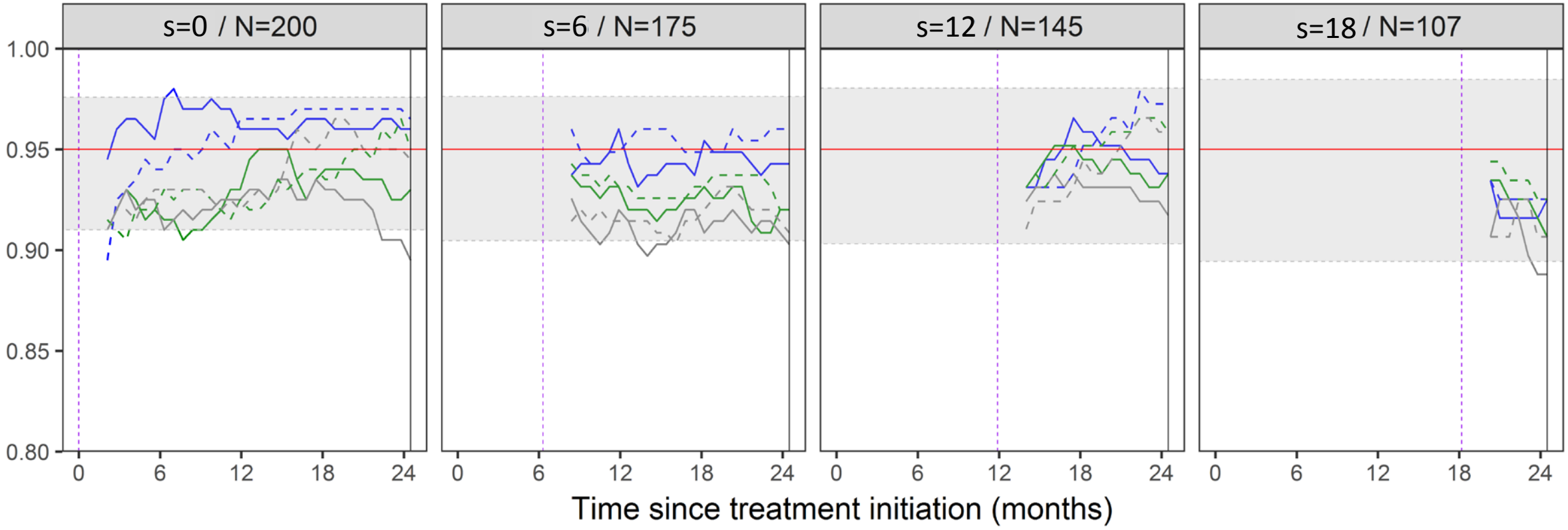
Stan(2.18.0)

Monolix2018R2®

NONMEM7.4®



High Link scenario coverage rate



PI95%: 95% prediction interval around 0.95

PSA

Survival

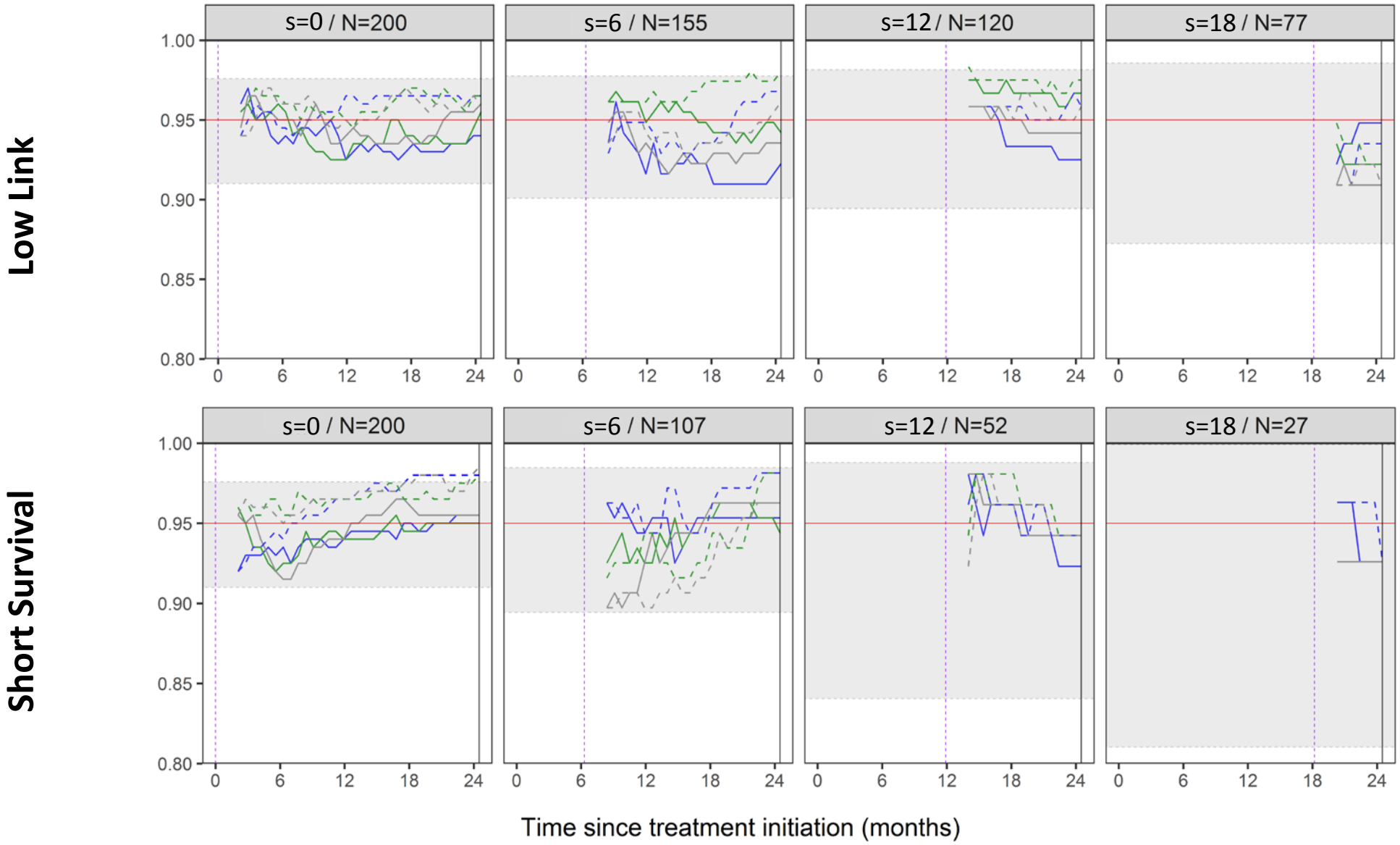
Landmark time s

Stan

Monolix

NONMEM

Other scenarios coverage rate



→ Good coverage with all software and for the different scenarios

Relative computation time (1 unit = 20 seconds)

Software	Landmark time s			
	0	6	12	18
Stan(2.18.0)	1.0	2.9	4 .0	4.1
Monolix2018R2®	13.3	16.8	15.5	24.6
NONMEM7.4®	6.8	11.1	16.3	23.7

- Computation times: Stan(2.18.0) << Monolix2018R2® ≈ NONMEM7.4®

Discussion

- Comparable individual dynamic predictions using **Stan(2.18.0)**, **Monolix2018R2®** and **NONMEM7.4®**
  - Validation of the use of MH implemented in **Monolix2018R2®** and **NONMEM7.4®** to generate individual dynamic predictions
  - Same results for the different scenarios: High Link, Low Link and Short Survival
  - Stan much faster than the other

- Limitations:
  - Uncertainty on population parameters not taken into account
  - Only analytical solution for PSA
    - ODE will be more efficient to describe PSA evolution over time
    - ODE solvers could be different between each software → Same results ?

Discussion

- Software handling:
  - **Stan(2.18.0)** is flexible and allows full Bayesian estimation
  - **Monolix2018R2®** is the easiest to use for someone without modeling knowledge
  - **NONMEM7.4®** is the most popular for nonlinear mixed effect modeling

Acknowledgments

- Grant: Sanofi
- To my PhD supervisor
  - France Mentre
  - Julie Bertrand
- To my INSERM colleagues
- Robert Bauer, ICON

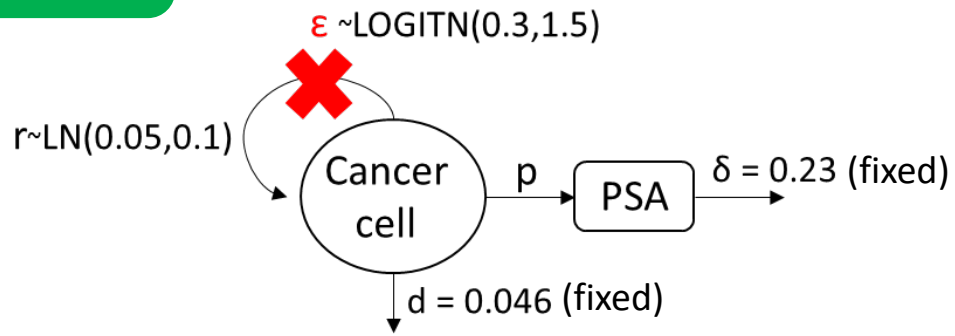
BACK-UP

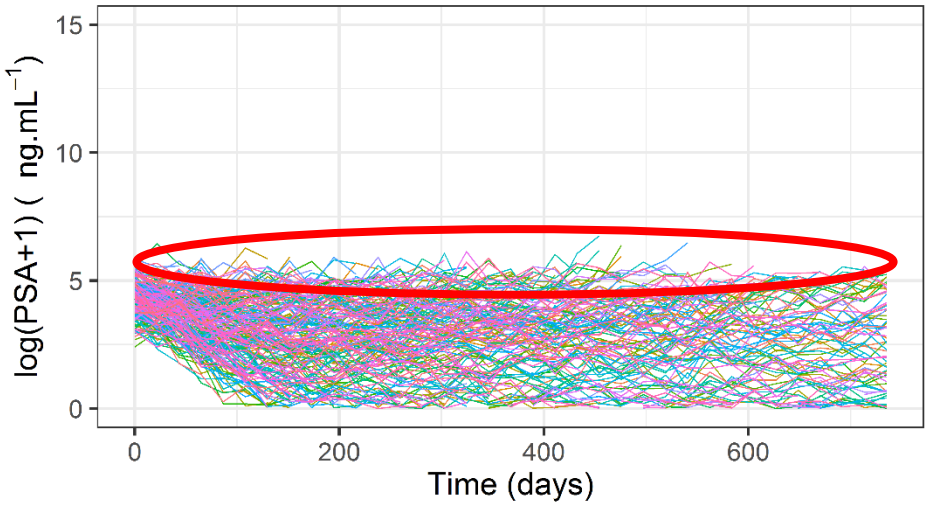
Simulation study design

- Desmee et al. AAPS Journal article about Joint Modelling of prostate cancer antigen (PSA) and overall surival

PSA kinetics model

$y_i = PSA(t, \psi_i) + \sigma e_i$ with $\sigma = 0.36$





Survival model

h_0: Weibull

$$h_{i,l}\left(t|PSA(t,\psi_i)\right) = \frac{k}{\lambda} \times \left(\frac{t}{\lambda}\right)^{k-1} \times e^{\beta \times PSA(t,\psi_i)}$$

Survival Parameters	Low Link	High Link	Short Survival
λ (days)	765	**2150**	560
k		**1.5**	
β	0.005	**0.02**	0.02
Survival at the end of the study (%)	25	**40**	5

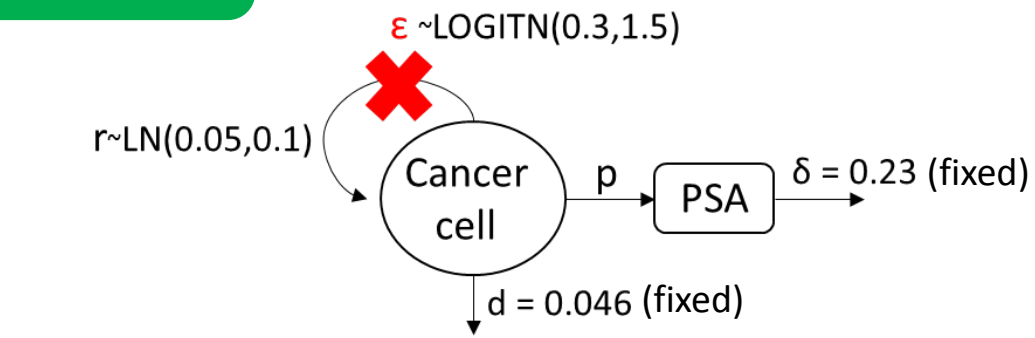
Desmee et al. (2015) *AAPS Journal*

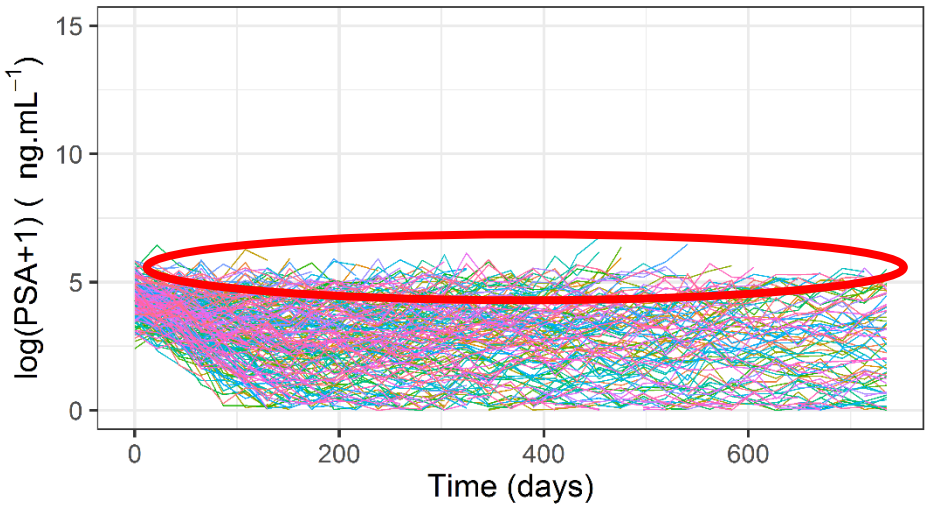
Simulation study design

- Desmee et al. AAPS Journal article about Joint Modelling of prostate cancer antigen (PSA) and overall surival

PSA kinetics model

$y_i = PSA(t, \psi_i) + \sigma e_i$ with $\sigma = 0.36$





Survival model

h_0: Weibull

$$h_{i,l}\left(t|PSA(t,\psi_i)\right) = \frac{k}{\lambda} \times \left(\frac{t}{\lambda}\right)^{k-1} \times e^{\beta \times PSA(t,\psi_i)}$$

Survival Parameters	Low Link	High Link	Short Survival
λ (days)	765	**2150**	560
k		**1.5**	
β	0.005	**0.02**	0.02
Survival at the end of the study (%)	25	**40**	5

Desmee et al. (2015) *AAPS Journal*

Software settings

Stan(2.18.0)	Monolix2018R2®	NONMEM7.4®
Calibration		
Thinning : 1 iteration Warmup : 500 iterations Sampling : L iterations Chain : 1	$\rho_{mcmc} = 1$ $L_{mcmc} = 500$ Chain : 1	CTYPE = 0 NBURN = 500 NITER = 0 Chain = 1 ---------- EONLY = 1 ISAMPLE = L
Sampling of $\psi_{i,l}$		
L iterations of the sampling phase (using true population, μ and ω ,as prior parameters)		**L** draws from the conditional distribution (using true population parameters, μ and ω, as fixed parameters)