A new computation method of uncertainty at finite distance for non linear mixed effects models

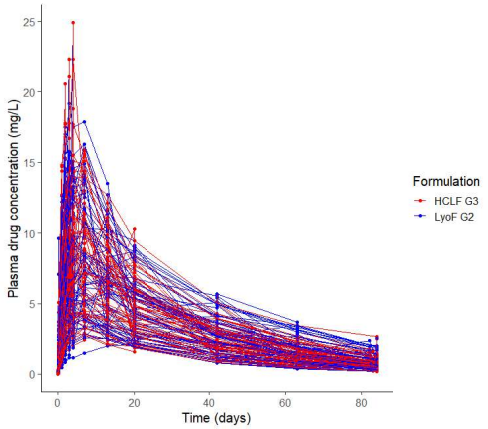
Mélanie Guhl, Julie Bertrand, Emmanuelle Comets

INSERM UMR 1137, *Infection, Antimicrobials, Modelling, Evolution*
Team BIPID

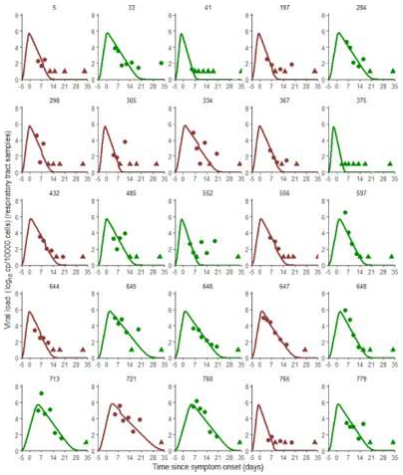
June, 8th 2023

Context

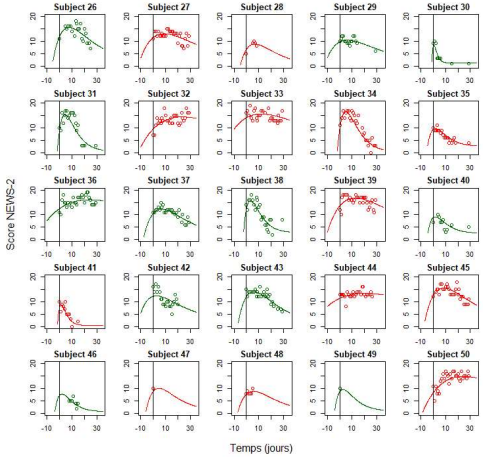
- Longitudinal data are widely collected in clinical trials



Drug concentration profiles[1]



Viral dynamic profiles [2]



NEWS-2 score data

- Non linear mixed effects models (NLMEM) are a powerful tool to model observations of individual $i = 1, ..., N$ at time $j = 1, ..., n_i$

$$y_{ij} = f(\psi_i, t_{ij}) + b\ \epsilon_{ij} \text{ with } \epsilon_{ij} \sim N(0, \Sigma)$$

$$\psi_i = \mu\ e^{\beta COV_i}\ e^{\eta_i} \text{ with } \eta_i \sim N(0, \Omega)$$

$$\theta = \{\mu, \beta, \Omega, b\}$$

[1] Guhl et al., J. Pharmacokinet. Pharmacodyn. 2022

[2] Lingas et al., J. Antimicrob. Chemother. 2022

Uncertainty in the Frequentist paradigm

The population parameter vector θ is not considered as a random variable but as a **fixed parameter**

$\rightarrow \hat{\theta}_{ML}$ Maximum likelihood estimator (MLE)
Estimation methods: Stochastic Approximation Expectation Maximization (SAEM)[3], First-Order Conditional Estimation (FOCE)[4], ...

The standard error (SE) of the MLE can be computed asymptotically based on the **Fisher Information Matrix** (FIM), the inverse of which is the lower bound of the asymptotic variance covariance matrix[5,6]

$$FIM = (-\partial^2_\theta l(\hat{\theta}_{ML}))$$

$$SE(\hat{\theta}_{ML}) = FIM^{-\frac{1}{2}}$$

Problem: when working at finite distance, using the FIM underestimates the SE of NLMEM leading to inflated type I error of tests[7]

[3] Kuhn, Lavielle, Comput. Stat. Data Anal. (2005)
[4] Beal and Sheiner, Crit. Rev. Biomed. Eng. (1982)
[5] Cramer, Princeton Univ. Press. (1946)
[6] Rao, Bull. Calcutta Math. Soc. (1945)
[7] Dubois et al., Stat. Med. (2011)

Uncertainty in the Frequentist paradigm

Other approaches following estimation of the MLE with a frequentist method (e.g. SAEM) :

- Bootstrap [8]
- Sampling Importance Resampling method (SIR) [9]
  - simulation of $s = 1, \ldots, S^1$ parameter vectors θ in a proposal distribution $p(\theta)$
  - computation of importance ratio (IR) for each sample

$$IR_s = \frac{l(y|\theta_s)/l(y|\hat{\theta}_{ML})}{PDF(\theta_s)/PDF(\hat{\theta}_{ML})}$$

  - resampling of $s = 1, \ldots, S^2$ new samples by weighting the S^1 previous samples with their IR_s

[8] Thai et al., J. Pharmacokinet. Pharmacodyn. 2013

[9] Dosne et al., J. Pharmacokinet. Pharmacodyn. 2016

Bayesian paradigm: *a posteriori* distribution

The population parameter vector θ is here considered as a **random variable** with a prior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

We want to estimate the posterior distribution of θ

Under some regularity conditions on the prior, Bayesian credible sets of a certain level α will asymptotically be confidence sets of level α (Bernstein von-Mises theorem[10])

$\rightarrow$ Ueckert et al. proposed to use the standard deviation (SD) of the posterior distribution as a proxy for the SE[11]

This proposal has been implemented via HMC algorithm in *stan* (method hereafter called Post) and used on a BE simulation study [12]

[10] van der Vaart, Cambridge Univ. Press. 1998
[11] Ueckert et al., PAGE. 2015
[12] Loingeville et al., AAPS J, 2020

Algorithm proposal - Concept

Proposal: use a Metropolis Hastings (MH) algorithm (already embedded in SAEM) and draw *a posteriori* distributions of the estimator of θ to compute its SD within the SAEM algorithm

$\rightarrow$ We end up with a frequentist MLE and a bayesian estimation of the SE in SAEM, computed in parallel

Convergence phase of the SAEM algorithm (K2 last iterations):

- Simulation step
- Stochastic approximation
- Maximisation step
- **Bayesian step**

The Bayesian step uses the frequentist estimations as parameters of the proposal kernel, but does not influence the frequentist estimation

Algorithm proposal - Bayesian step

- One MH chain is sampled at each iteration k of SAEM, from K1+1 to K1+K2 (convergence phase)
- Chain of length M
- prior on θ: $p(.)$
- At each iteration m ($m = 1, ..., M$)
  - we draw a sample $\theta^{(m)}$ in the kernel $q(.)$
  - $\theta^{(m)}$ is accepted with probability

$$\alpha = min\left(1, \frac{l(y|\theta^{(m)}, \bar{\psi}^{(m)})p(\theta^{(m)})q_\theta(\theta^{(m-1)})}{l(y|\theta^{(m-1)}, \bar{\psi}^{(m-1)})p(\theta^{(m-1)})q_\theta(\theta^{(m)})}\right)$$

with $\bar{\psi}^{(m)}$: mean of 50 samples of $\psi^{(m)}$ in the distribution $p(\psi|\theta^{(m)})$

Iteration of SAEM

Chain used to compute SD

First set of simulations in the context of Bioequivalence (BE) studies

Simulation settings

Our set of simulations is inspired by pharmacokinetic BE studies[13,14,15]
Theophylline data

- 1000 datasets
- 1 compartment of distribution with linear absorption and elimination

$$C(t) = \frac{D}{V}\frac{ka}{\frac{Cl}{V} - ka}\left(\exp(-ka\ t) - \exp(-\frac{Cl}{V}t)\right)$$

μ_{ka}=1.5	β_{ka}=0	ω_{ka}=0.22
μ_{Cl}=0.04	β_{Cl}=log(1.25)	ω_{Cl}=0.11
μ_V=0.5	β_V=log(1.25)	ω_V=0.22

- Proportional error model: b=0.1
- Designs

Rich
N = 150 (75 in each treatment arm)
n = 10: $t \in \{0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12, 24\}$ hours

[13] Dubois et al., Stat. Med. 2011
[14] Loingeville et al., AAPS J, 2020
[15] Mollenhoff et al., Biostatistics, 2022

Simulation settings

Our set of simulations is inspired by pharmacokinetic BE studies[13,14,15]
Theophylline data

- 1000 datasets
- 1 compartment of distribution with linear absorption and elimination

$$C(t) = \frac{D}{V}\frac{ka}{\frac{Cl}{V} - ka}\left(\exp(-ka\ t) - \exp(-\frac{Cl}{V}t)\right)$$

μ_{ka}=1.5	β_{ka}=0	ω_{ka}=0.22
μ_{Cl}=0.04	β_{Cl}=log(1.25)	ω_{Cl}=0.11
μ_V=0.5	β_V=log(1.25)	ω_V=0.22

- Proportional error model: b=0.1
- Designs

Rich	**Sparse**
N = 150	N = 12 (6 in each treatment arm)
n = 10	n = 3: $t \in \{0.25, 3.5, 24\}$ hours

[13] Dubois et al., Stat. Med. 2011
[14] Loingeville et al., AAPS J, 2020
[15] Mollenhoff et al., Biostatistics, 2022

Evaluation

- 95% coverage rates for each element of θ
- Acceptation rates (proportion of accepted samples)

We compare the results obtained with our method with those obtained from:

- the FIM (Asympt)
- Sampling Importance Resampling (SIR)
- Post

Settings

SIR (module in *saemix* [16])

- $S^1 = 1000$ samples
- $S^2 = 500$ resamples

Post

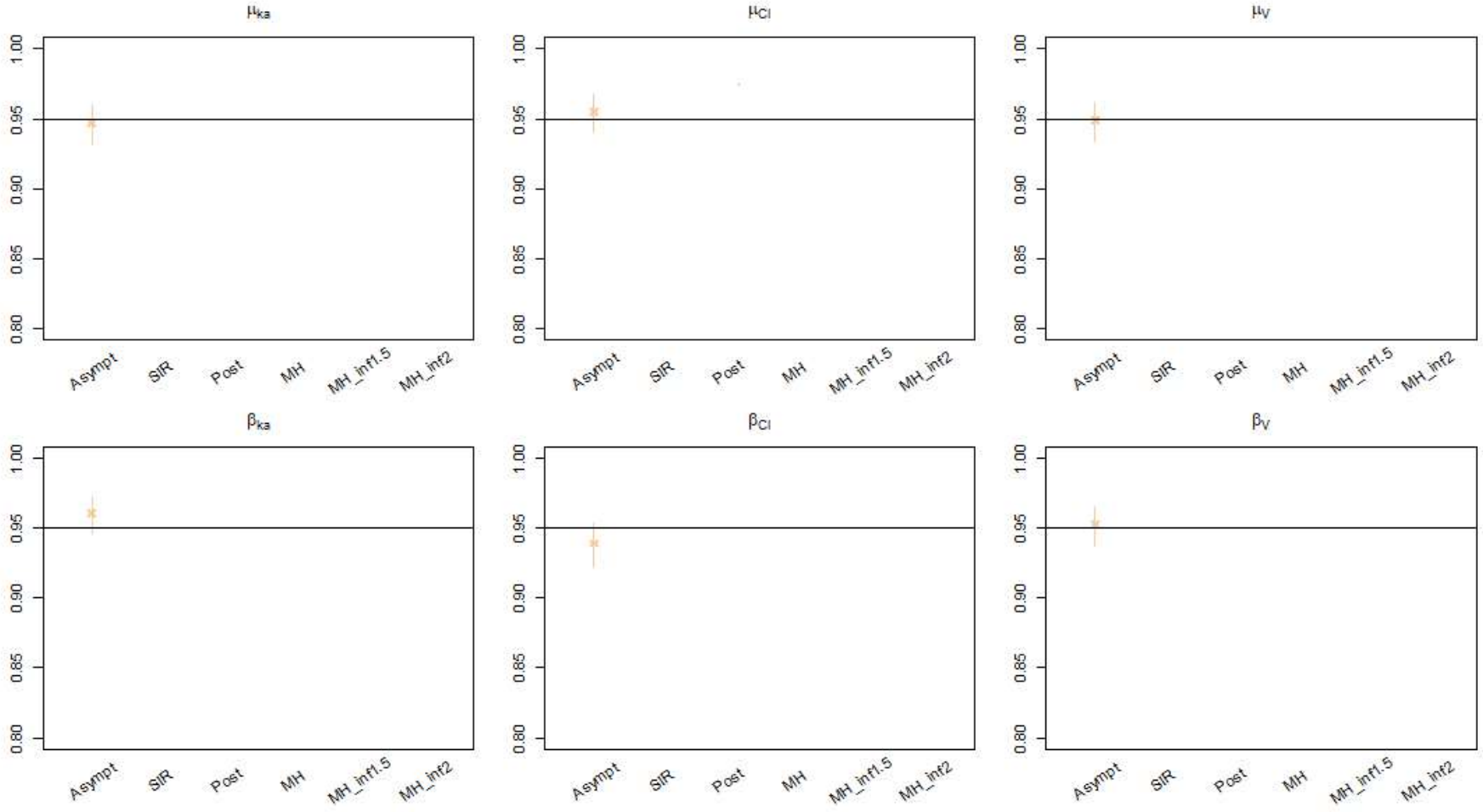
- 3 chains
- 1500 iterations (including 500 iterations of warm up)
- Initial values: estimations of saemix
- Gaussian prior: $p(.) = \mathcal{N}(\theta, \Sigma)$
  Σ diagonal with $\sigma_i = 0.3 * \mu_i$ for all μ, $\sigma_i = 0.5$ otherwise

MH

- Gaussian prior: $p(.) = \mathcal{N}(\theta, \Sigma)$
  Σ diagonal with $\sigma_i = 0.3 * \mu_i$ for all μ, $\sigma_i = 0.5$ otherwise
- Gaussian kernel: $q(.) = \mathcal{N}(\hat{\theta}_k, inf * \widehat{FIM}_k^{-1})$ with inf=1, 1.5, 2
- M=100

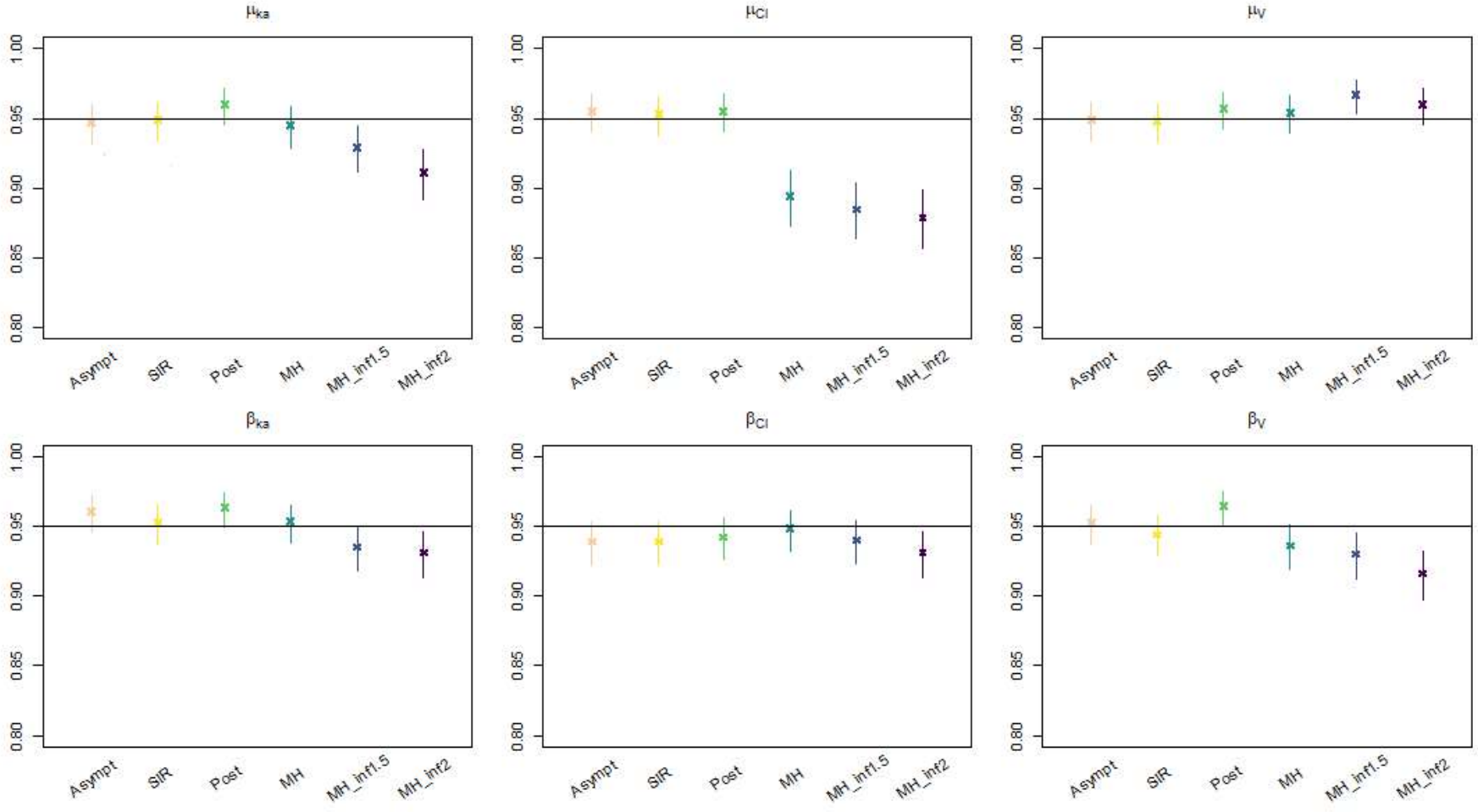
[16] https://github.com/saemixdevelopment

95% coverage rates



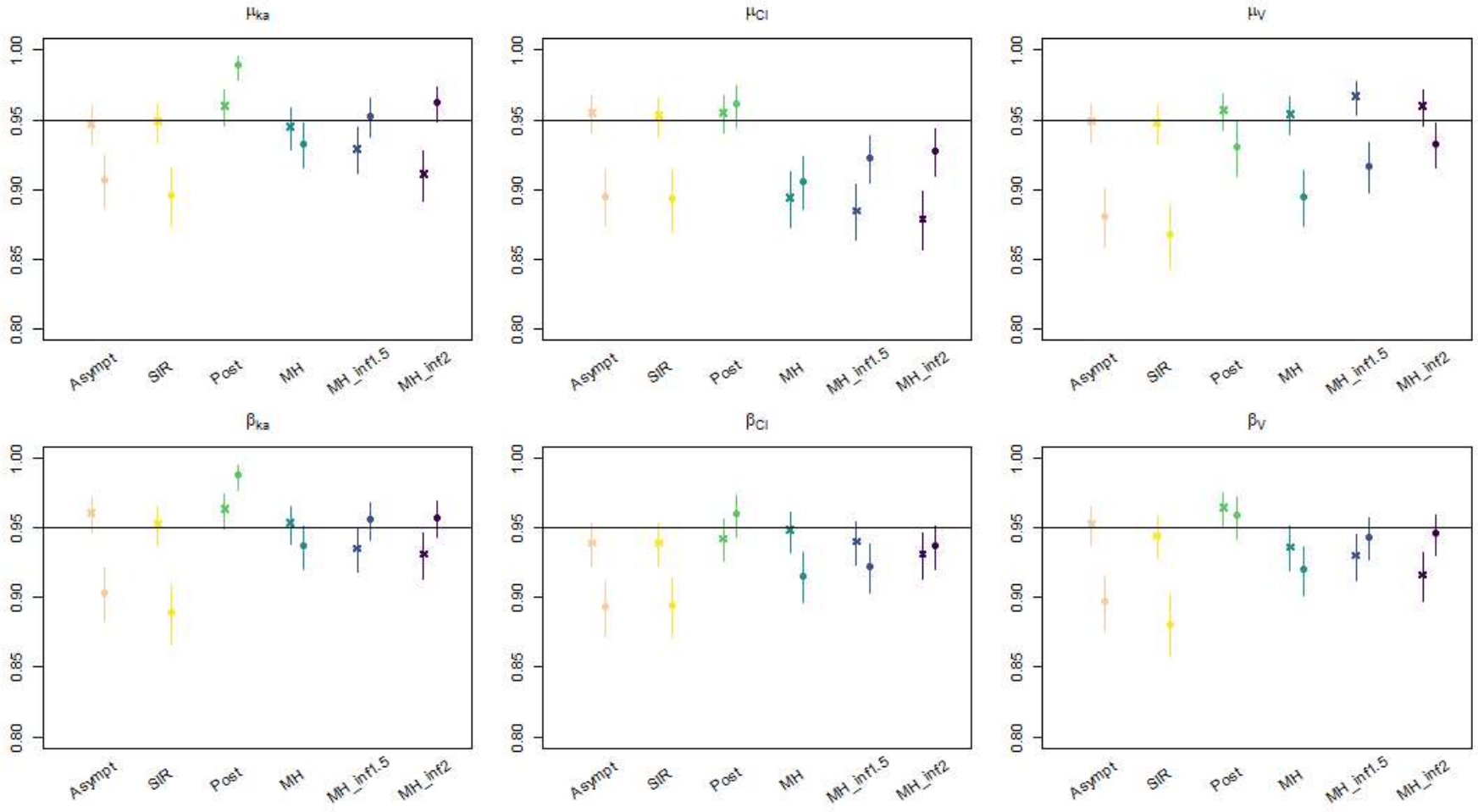
x: N=150, n=10

95% coverage rates



x: N=150, n=10

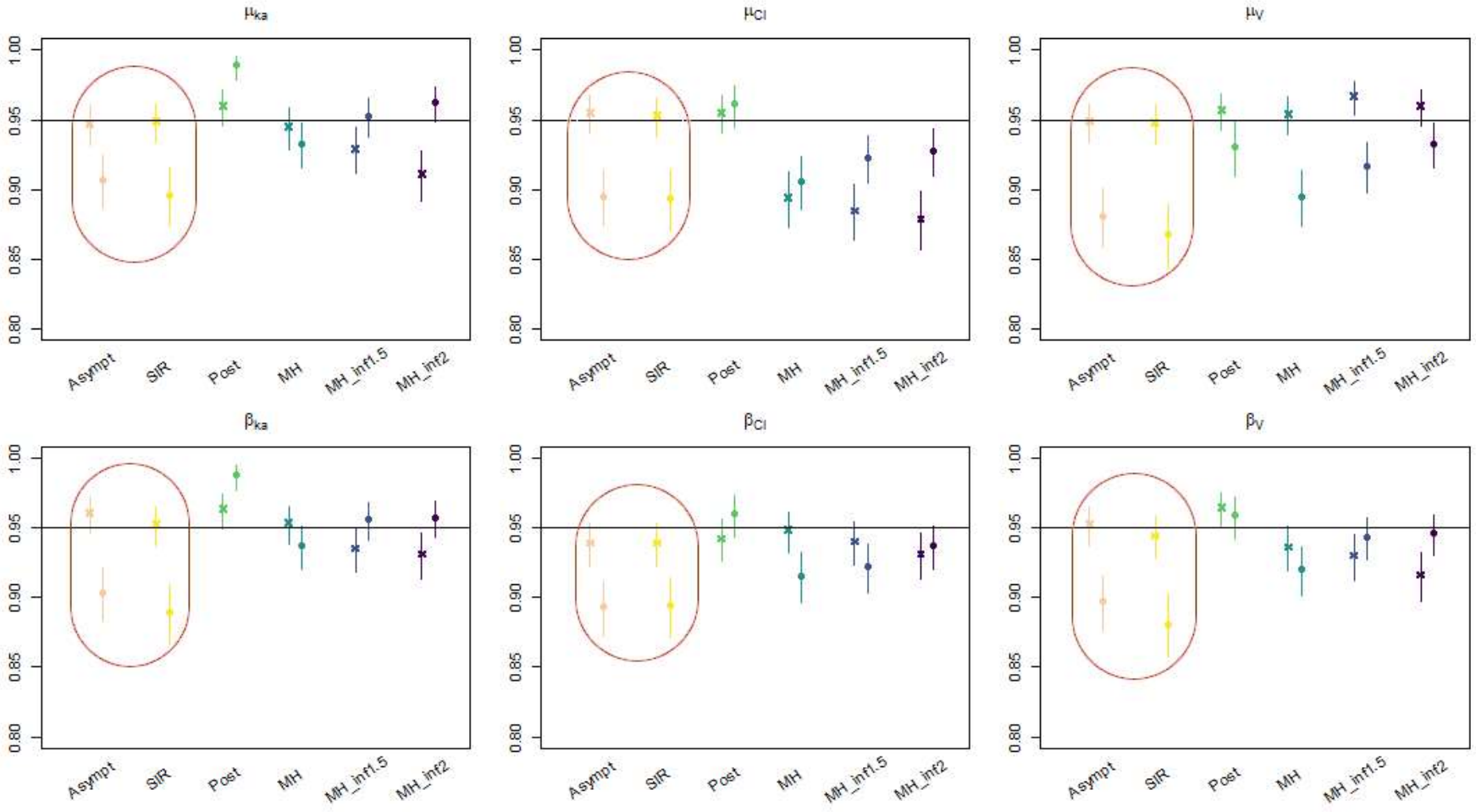
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

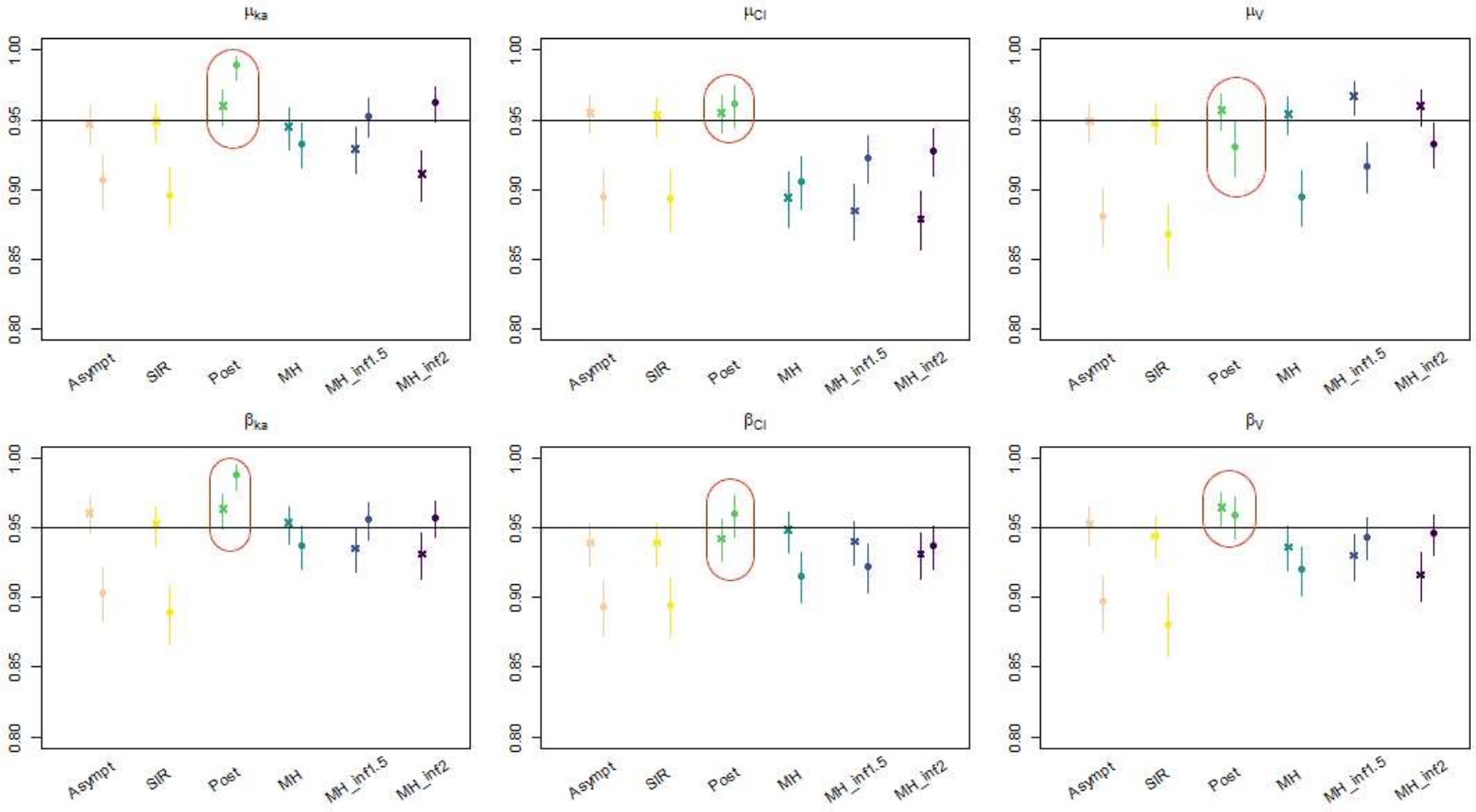
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

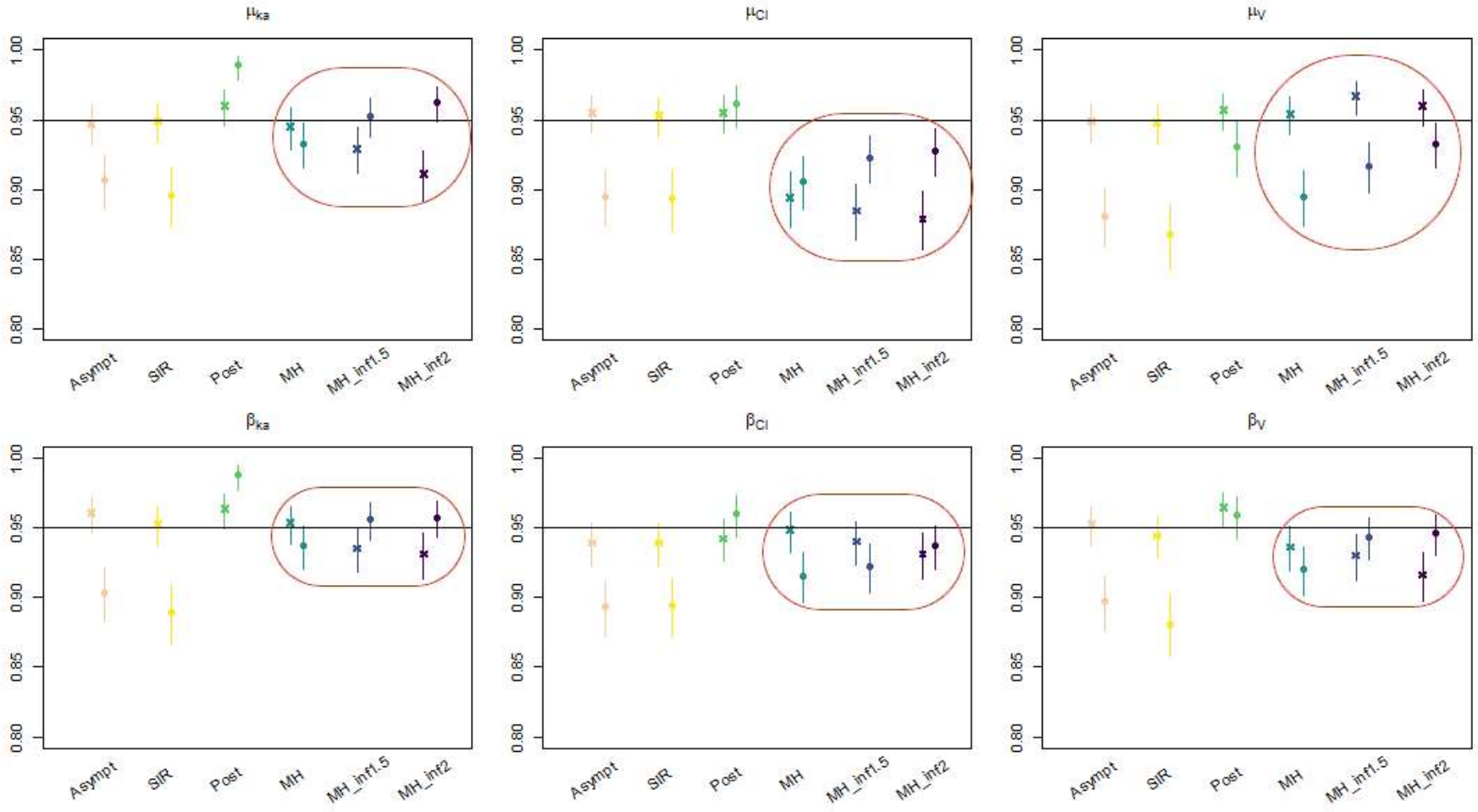
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

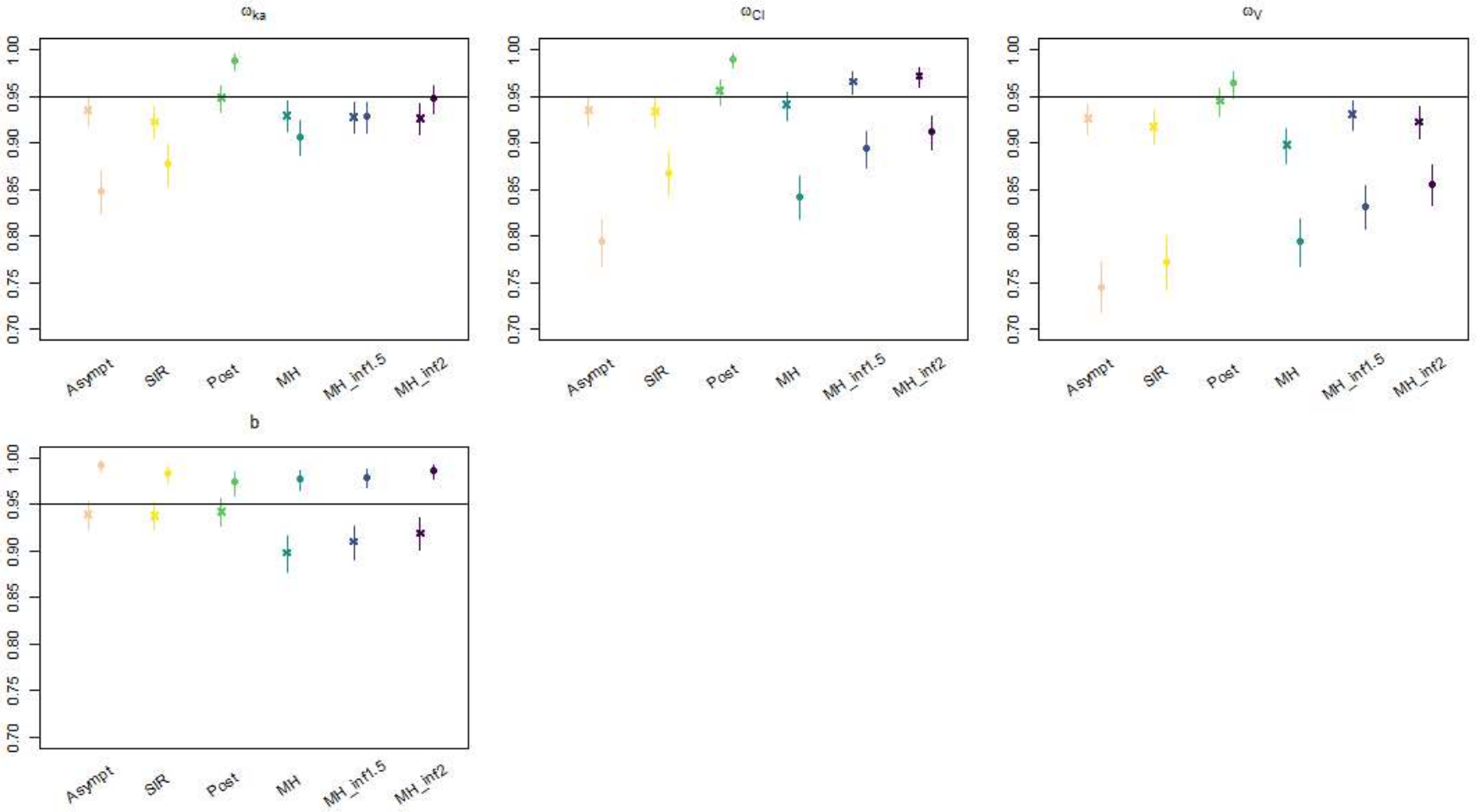
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

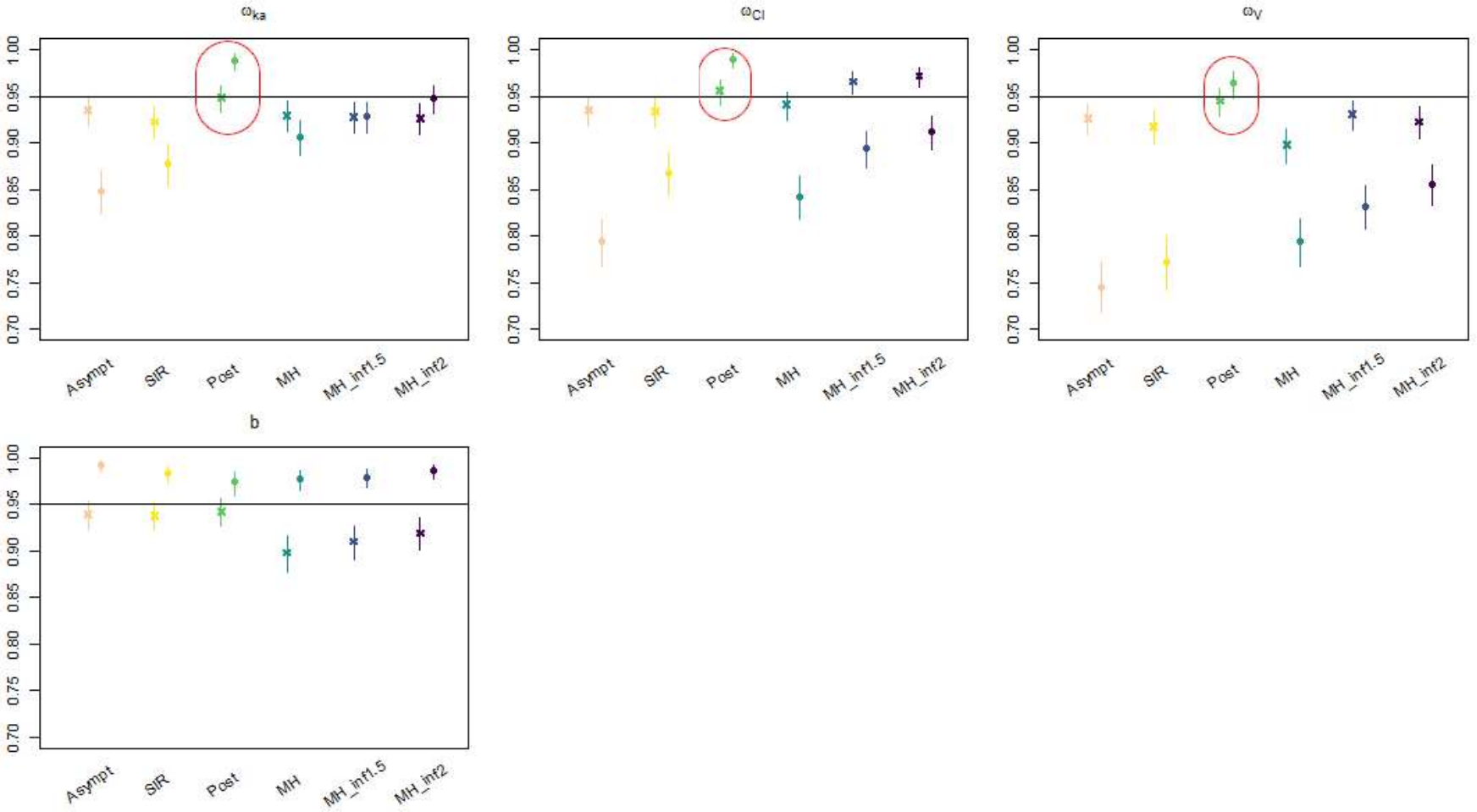
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

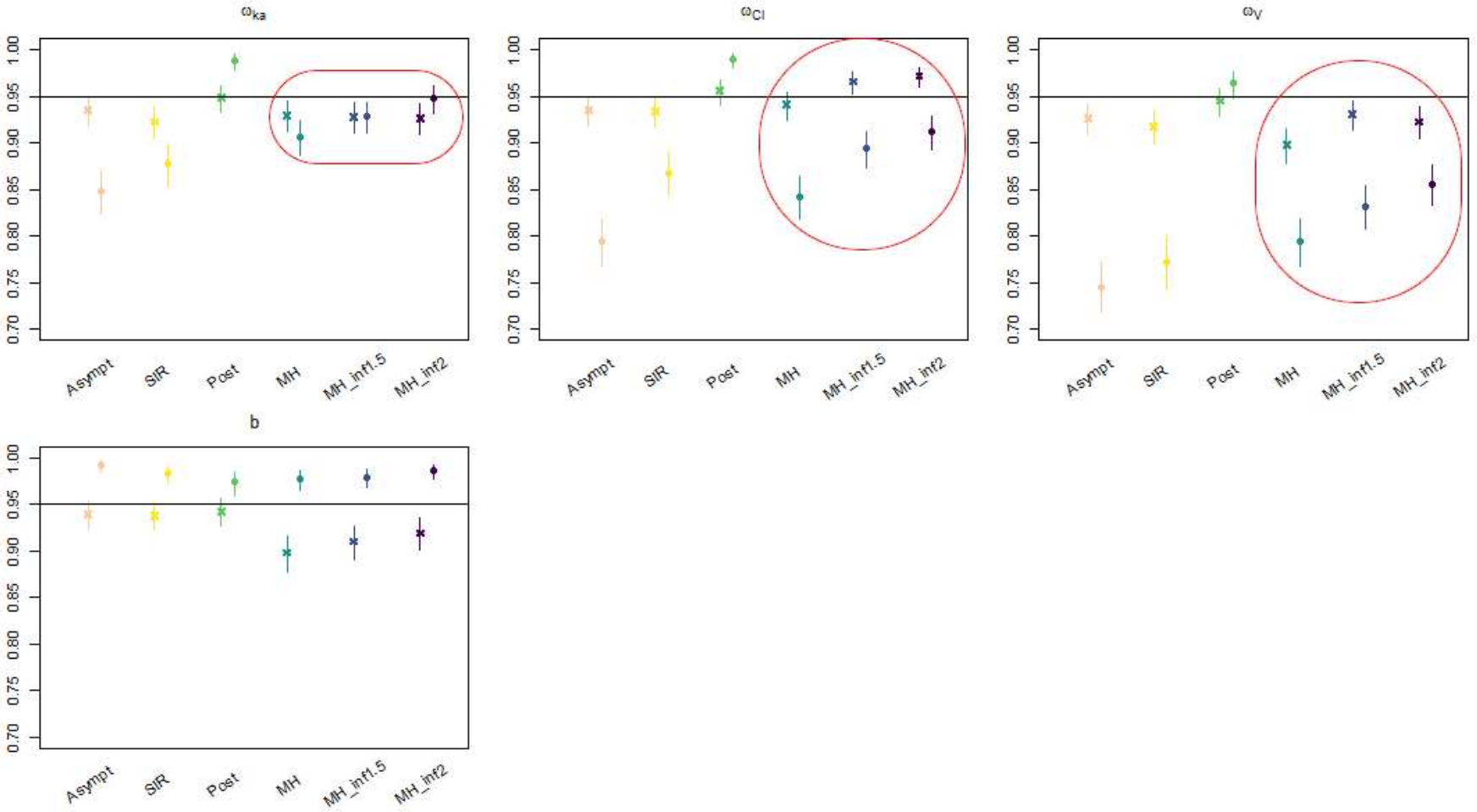
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

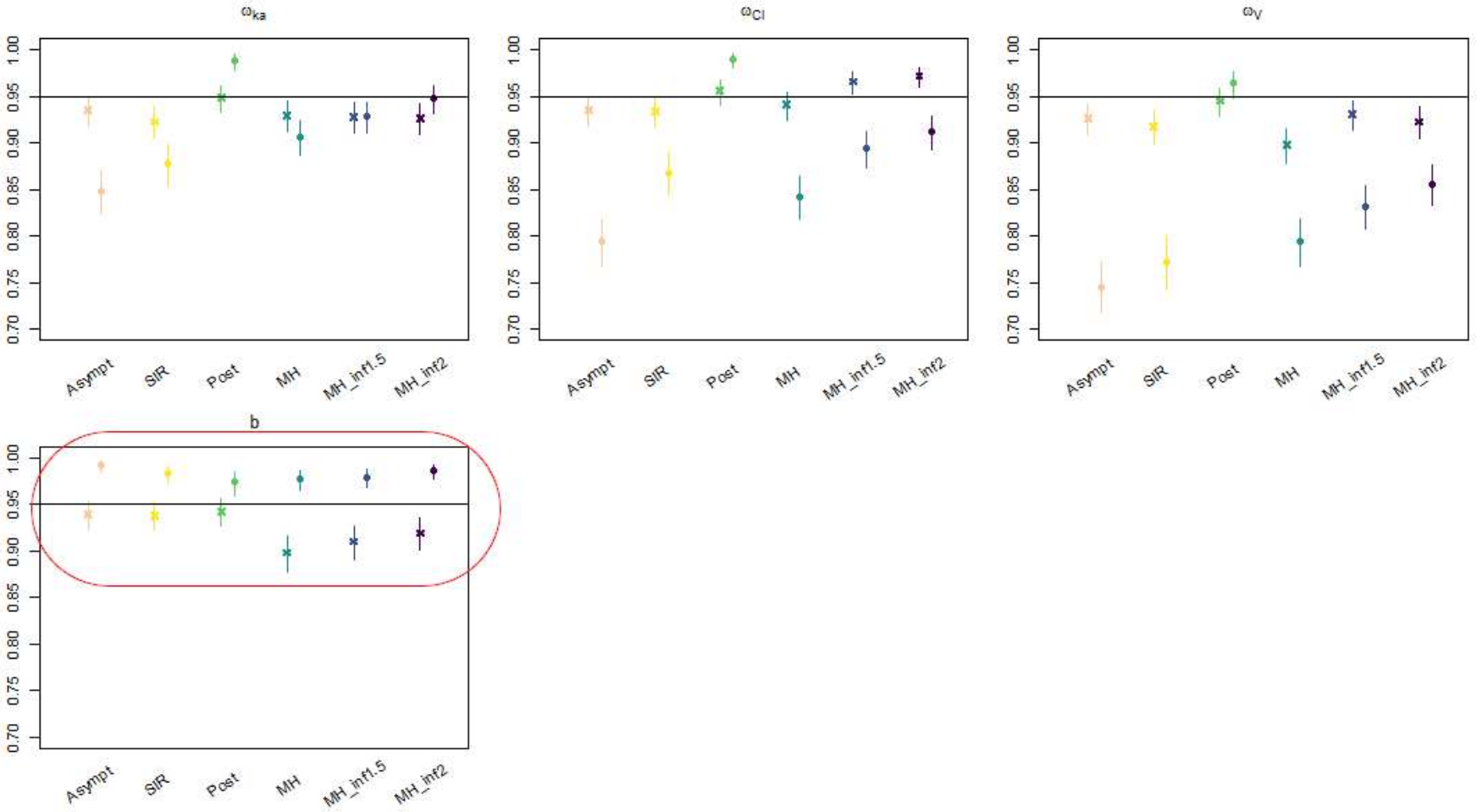
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

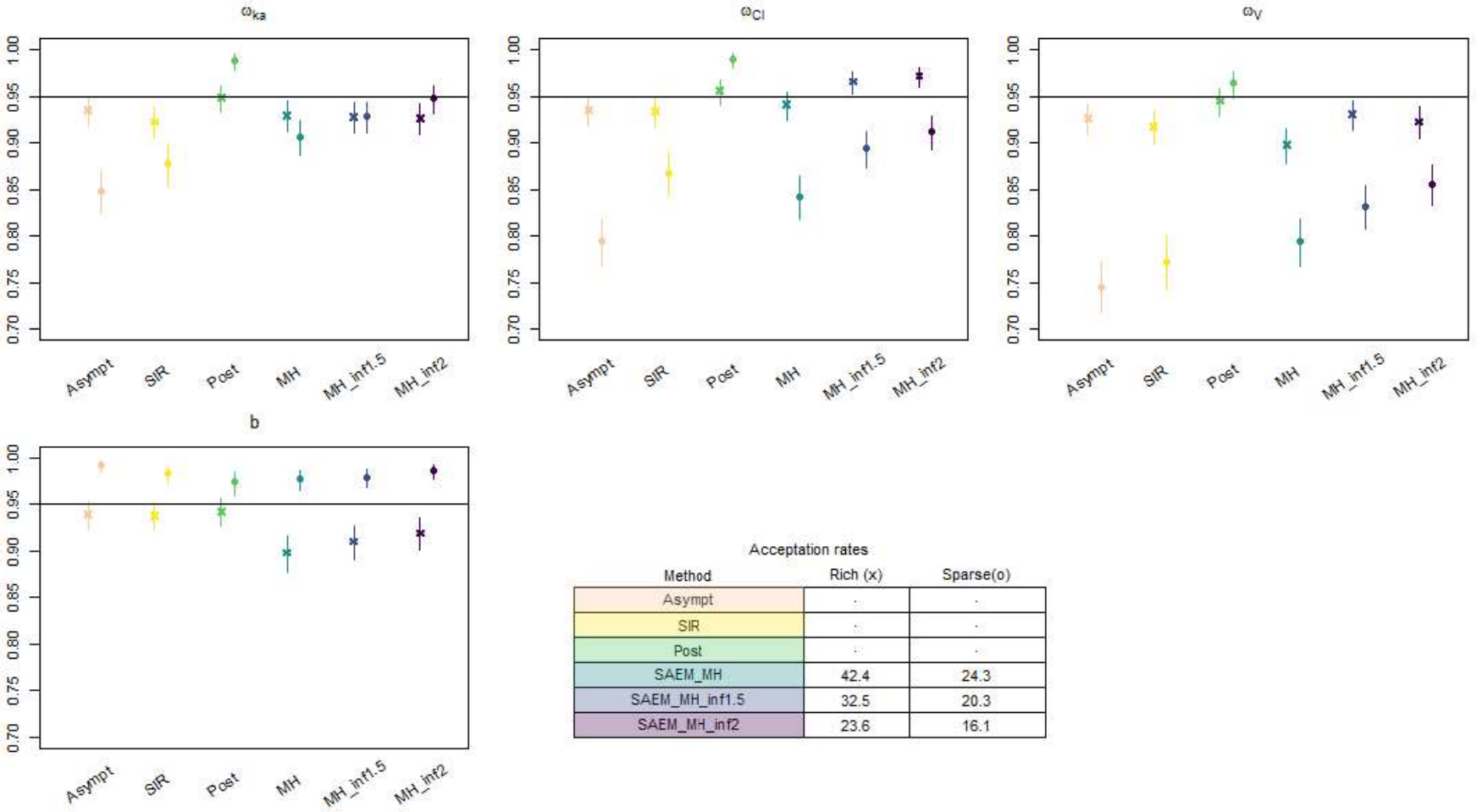
95% coverage rates



x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

95% coverage rates



Acceptation rates

Method	Rich (x)	Sparse(o)
Asympt	.	.
SIR	.	.
Post	.	.
SAEM_MH	42.4	24.3
SAEM_MH_inf1.5	32.5	20.3
SAEM_MH_inf2	23.6	16.1

x: N=150, n=10 / o: N=12, n=3

Post: data sets with $\hat{R}$>1.05 were not used (27% for sparse scenario) - SIR: 15.3% failure for sparse scenario

Summary

- On rich design (N=150,n=10)

→ Controlled coverage rates with Asympt, SIR, Post and MH

- On sparse design (N=12, n=3)

→ Asympt, SIR and MH have coverage rates under the target, especially for variance parameters

→ Post has coverage rates over the target

⇒ More work is needed to develop a reliable method of SE computation on sparse data

- An inflation factor of 2 on the on the kernel variance allows MH to give more controlled coverage rates on small sample sizes
  - Acceptation ratio are decreased
  - best coverage rates are not met at "The asymptotically optimal acceptance rate is 0.234 under quite general conditions"[17]
- On more challenging settings (e.g. high inter individual variability)
  - MH method has coverage rates under the target and below Asympt
  - Acceptation ratio collapse → good tool to diagnose when MH is too challenged
- Inflation of prior distribution or increase of the chains length do not change the results with MH (not shown)

[17] Robert et al. Ann. Appl. Prob. 1997

Perspectives

- Calibration of the kernel
  - decouple fixed effects and variance parameters, univariate conditional distribution, random walk

  $\rightarrow$ Lucie Fayette internship

- Implementation of MH in *saemix* on the CRAN

- Extension of our new method of computations of SE to categorical data: one issue is that we cannot compute the **linearised likelihood** in that case

Thank you