Bayesian approaches for pharmacogenetic models with JAGS and Stan

Julie Bertrand, Maria De Iorio, David J. Balding

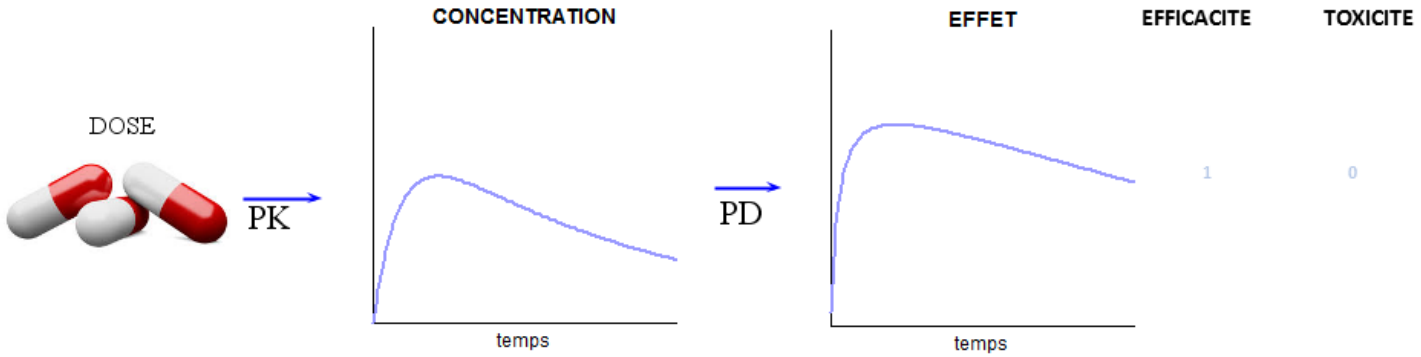
Inserm
Institut national
de la santé et de la recherche médicale

lame
UMR 1137
Infection • Antimicrobials • Modelling • Evolution

UCL

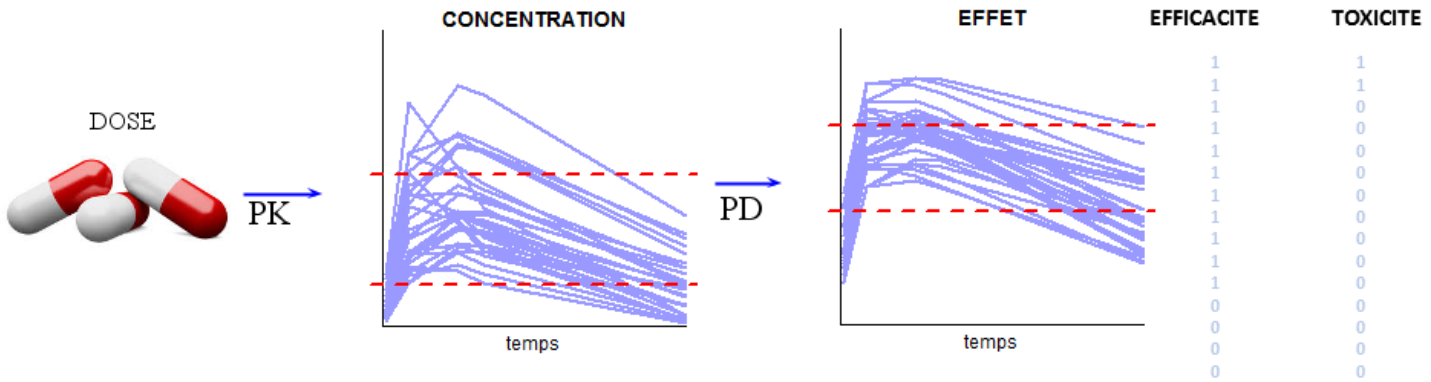
Pharmacological and genetic variability

- Clinical pharmacology: study the interaction between the organism and the drug
  - pharmacokinetics (PK) and pharmacodynamics (PD)



Pharmacological and genetic variability

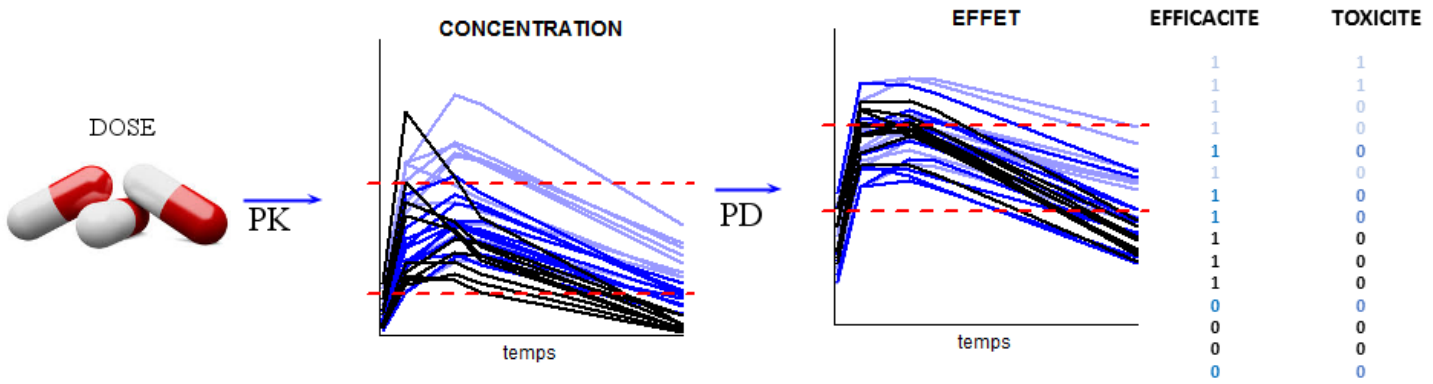
- Clinical pharmacology: study the interaction between the organism and the drug and its variability
  - pharmacokinetics (PK) and pharmacodynamics (PD)



Pharmacological and genetic variability

- Clinical pharmacology: study the interaction between the organism and the drug and its variability
  - pharmacokinetics (PK) and pharmacodynamics (PD)

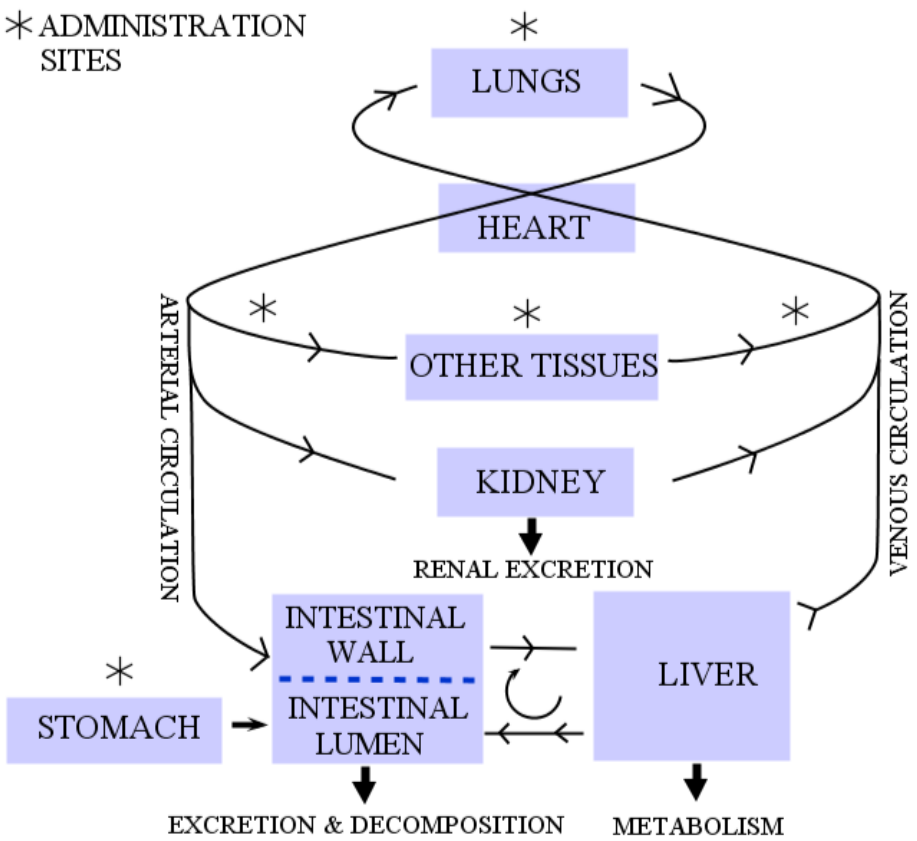
TT	GT	CT	AG	AG	CT	GG	TT	TT	TT
TT	TT	CT	AG	AG	CT	GG	GT	GT	GT
TT	TT	TT	GG	GG	CC	GG	GG	GG	GG
TT	TT	CT	AG	AG	CT	GG	GT	GT	GT
TT	TT	CC	AA	AA	TT	GG	TT	TT	TT
TT	TT	CT	AG	AG	CT	GG	GT	GT	GT
TT	TT	TT	GG	GG	CC	GG	GG	GG	GG
TT	TT	CT	AG	AG	CT	GG	GT	GT	GT
TT	TT	CT	AG	AG	CT	GG	GT	GT	GT
TT	TT	TT	GG	GG	CC	GG	GG	GG	GG
TT	TT	CC	AA	AA	TT	GG	TT	TT	TT
TT	TT	TT	GG	GG	CC	GG	GG	GG	GG
CT	TT	CT	AG	AG	CT	GG	GT	GT	GT
CT	TT	CC	AA	AA	TT	GG	GT	GT	TT
CT	TT	TT	GG	GG	CC	GG	GG	GG	GG

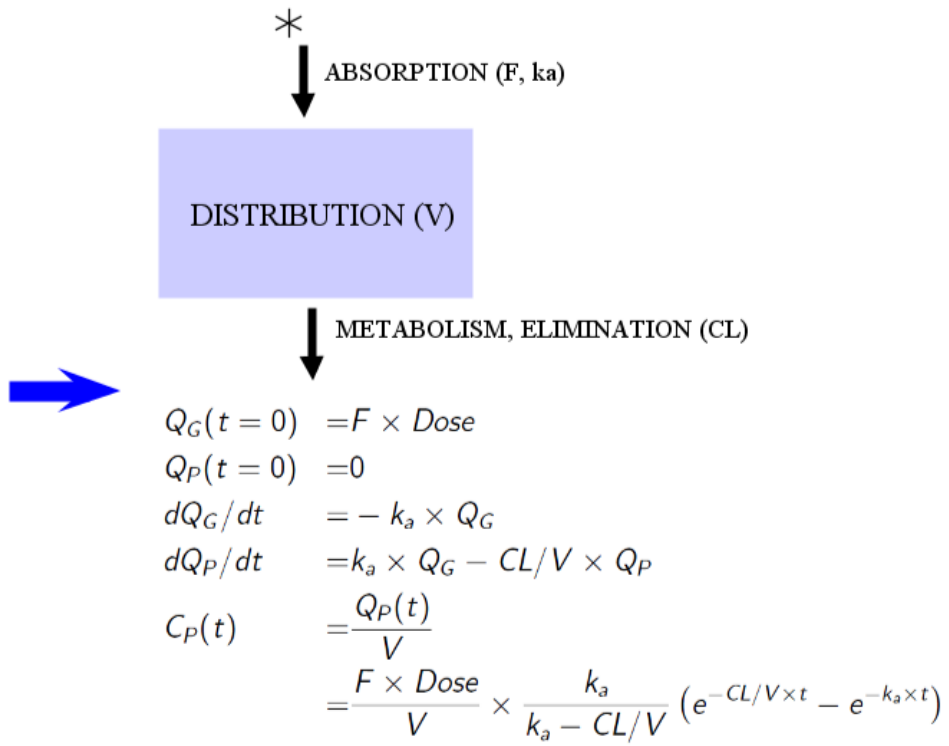


- Pharmacogenetics (PG): genetic part of the variability
  - stratified medicine
- Genes coding for proteins involved in PK/PD processes
  - metabolism enzymes (CYP450, NAT)
  - single nucleotide polymorphism (SNP)

Modelling in pharmacology

- Semi-physiological models integrating the *a priori* knowledge on the drug
  - parameters characterizing each physiological processes
  - model nonlinear in its parameters





Modelling in pharmacology

- Semi-physiological models integrating the *a priori* knowledge on the drug
  - parameters characterizing each physiological processes
  - model nonlinear in its parameters
- Mixed effect models
  - all patients analyzed simultaneously

Modelling in pharmacology

- Semi-physiological models integrating the *a priori* knowledge on the drug
  - parameters characterizing each physiological processes
  - model nonlinear in its parameters
- Mixed effect models
  - all patients analyzed simultaneously
  - parameter decomposed in fixed and random effects

  ↪ covariates identification

Example : $CL_i = CL + \beta \times SNP_i + \eta_i$ with $SNP = \{0, 1, 2\}$

Modelling in pharmacology

- Semi-physiological models integrating the *a priori* knowledge on the drug
  - parameters characterizing each physiological processes
  - model nonlinear in its parameters
- Mixed effect models
  - all patients analyzed simultaneously
  - parameter decomposed in fixed and random effects

  ↪ covariates identification

Example : $CL_i = CL + \beta \times SNP_i + \eta_i$ with $SNP = \{0, 1, 2\}$

- Estimation methods
  - Maximum likelihood: model linearization, Gaussian quadrature, SAEM
  - Bayesian inference (Lunn et al. 2002)

Methodological challenges in PGx

- PK/PD phenotype $\rightarrow$ not observed
  - data : plasma or insulin concentrations, ...

  $\hookrightarrow$ dynamical models
- Variable informativeness of genetic markers
  - uneven distribution, small sample size of some genotypes

  $\hookrightarrow$ mixed effect models
- Increased size of the genetic data sets toward high throughput screening
  - dimensionality curse $N << p$
  - structural correlation along the genome (linkage disequilibrium)

  $\hookrightarrow$ statistical genetics

Genetic association analyses in PGx

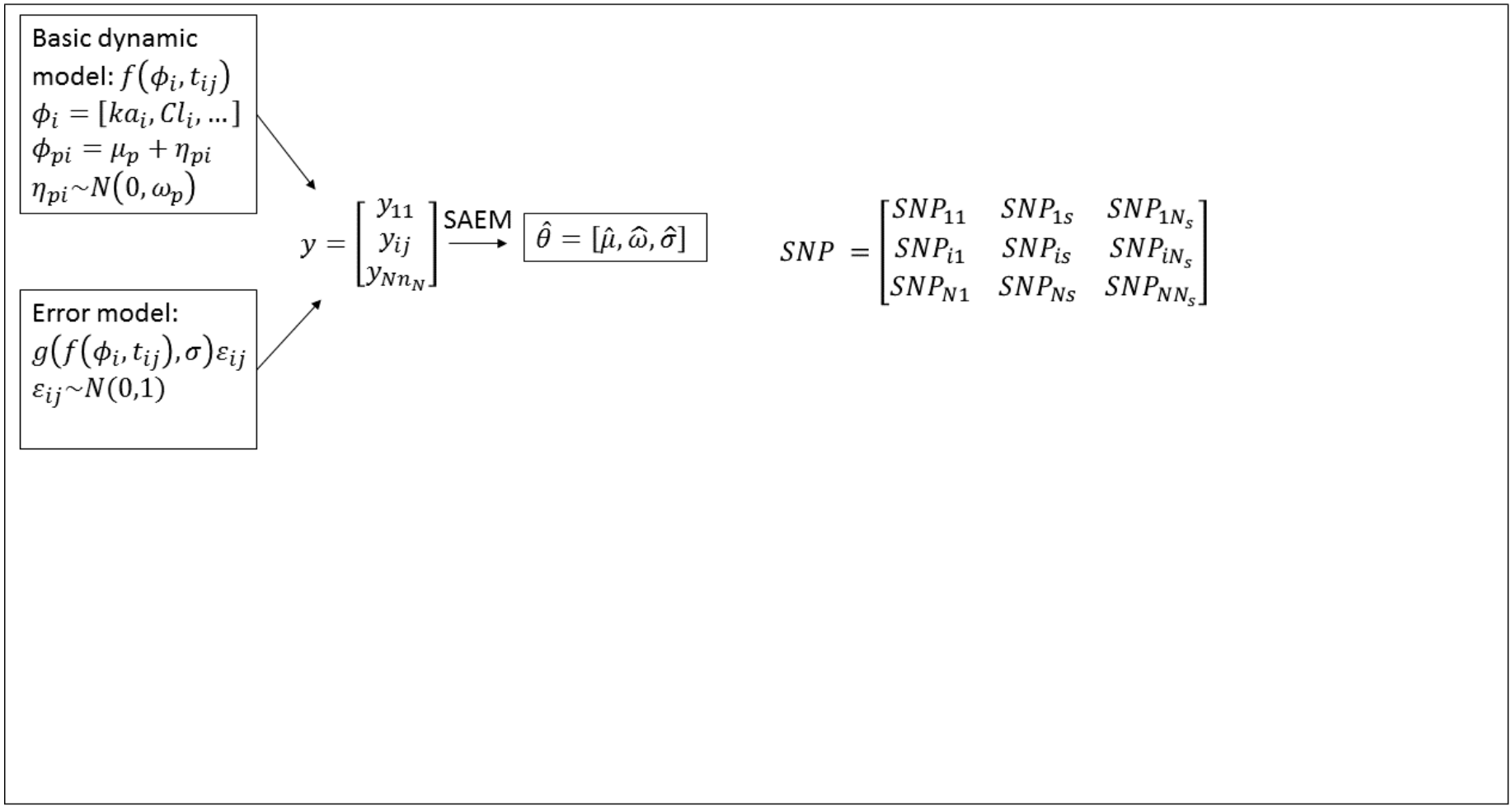
- Stepwise procedure
  - commonly used for covariate model building
  - Lehr et al. (2010) adaptation for high throughput screening
- Integrated approach based on penalized regression
  - penalized regression established in animal and plant genetics
  - Lasso (Tibshirani. 1996)
  - HLasso (Hoggart et al. 2008)
    - developed for genome-wise association studies
    - higher effect size once included in the model
- Bayesian variable selection (O'Hara & Sillanpaa 2009)

Stepwise procedure (Lehr et al. 2010)

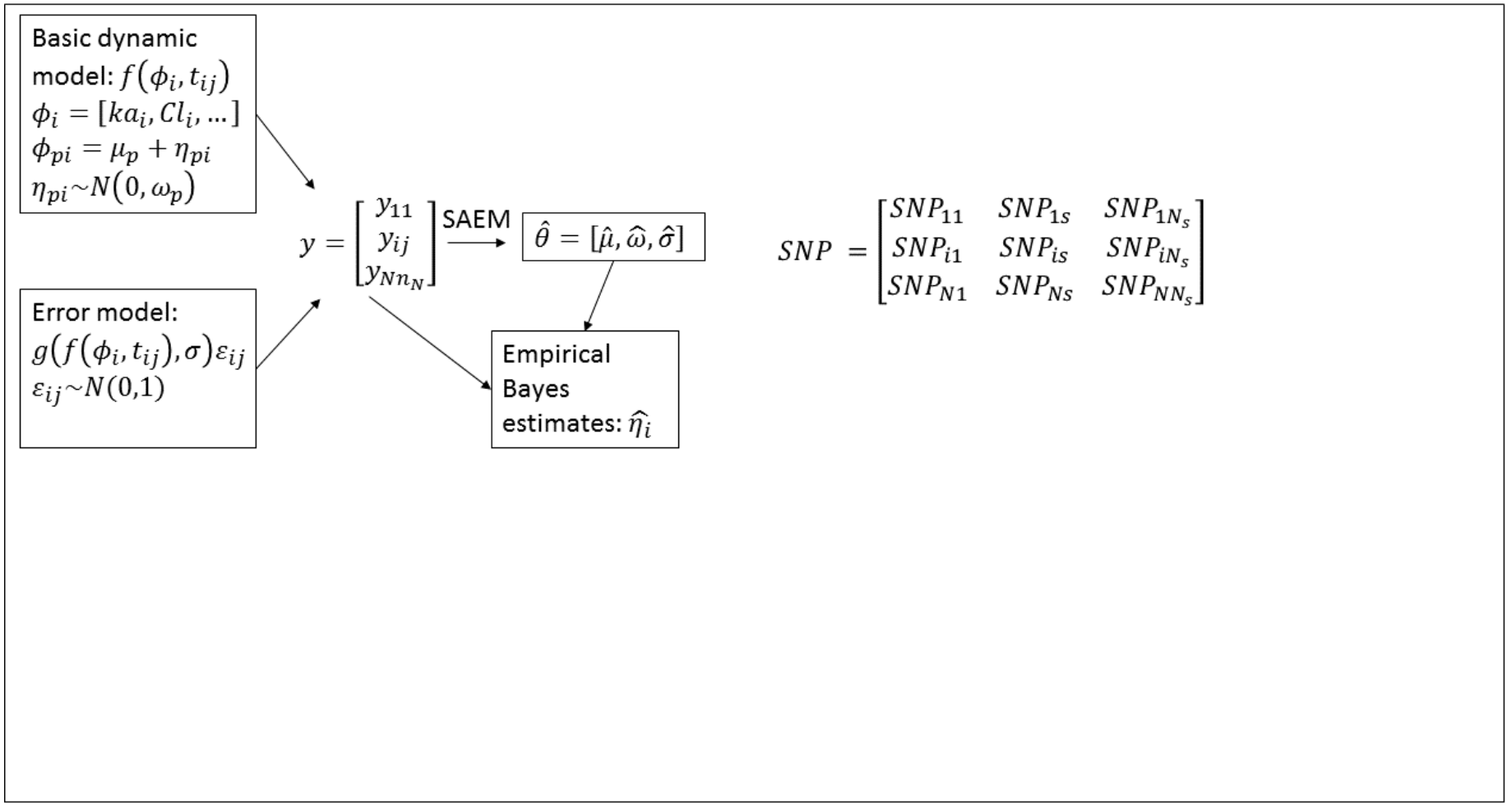
$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix}$$

$$SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

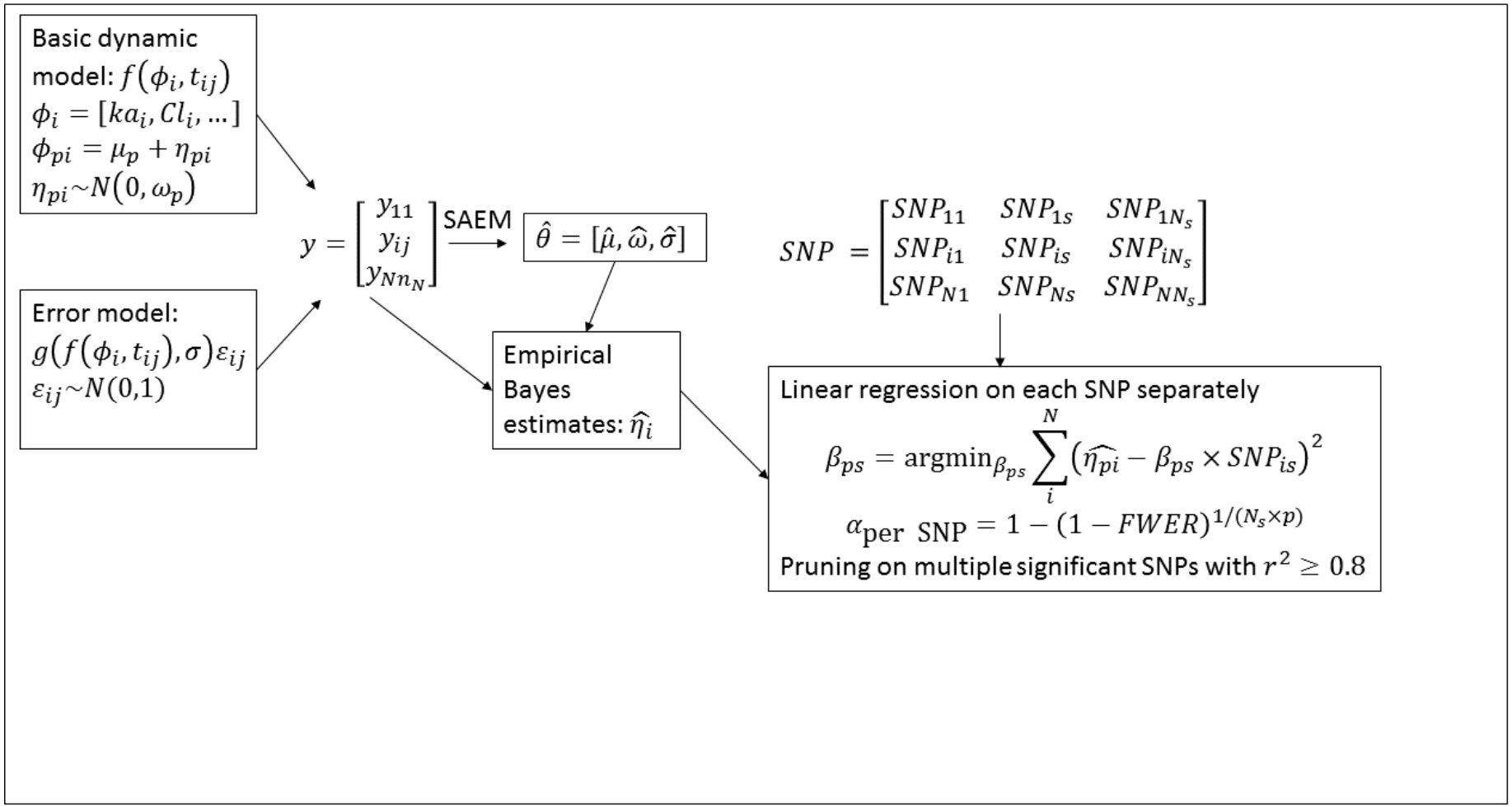
Stepwise procedure (Lehr et al. 2010)



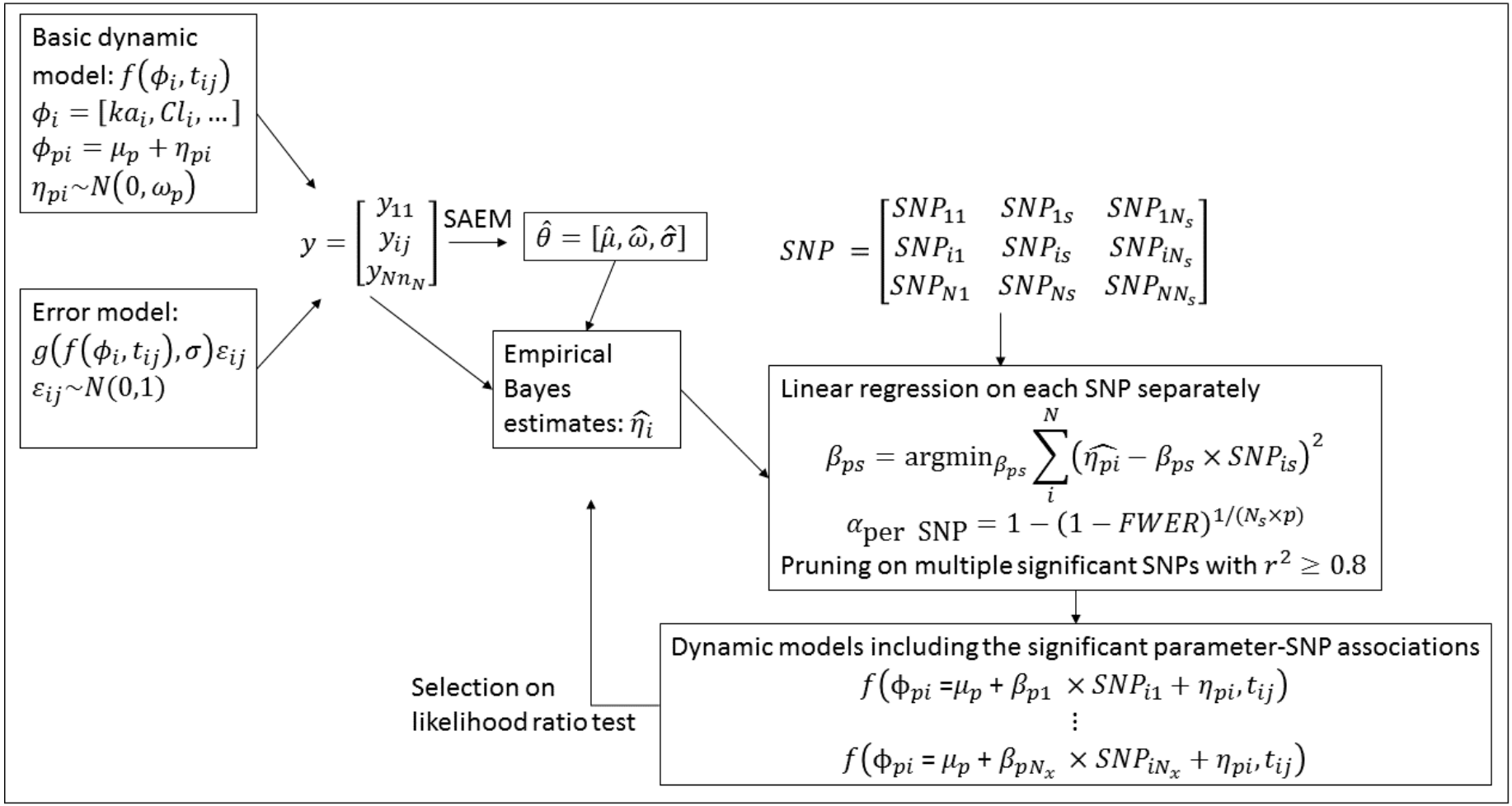
Stepwise procedure (Lehr et al. 2010)



Stepwise procedure (Lehr et al. 2010)



Stepwise procedure (Lehr et al. 2010)



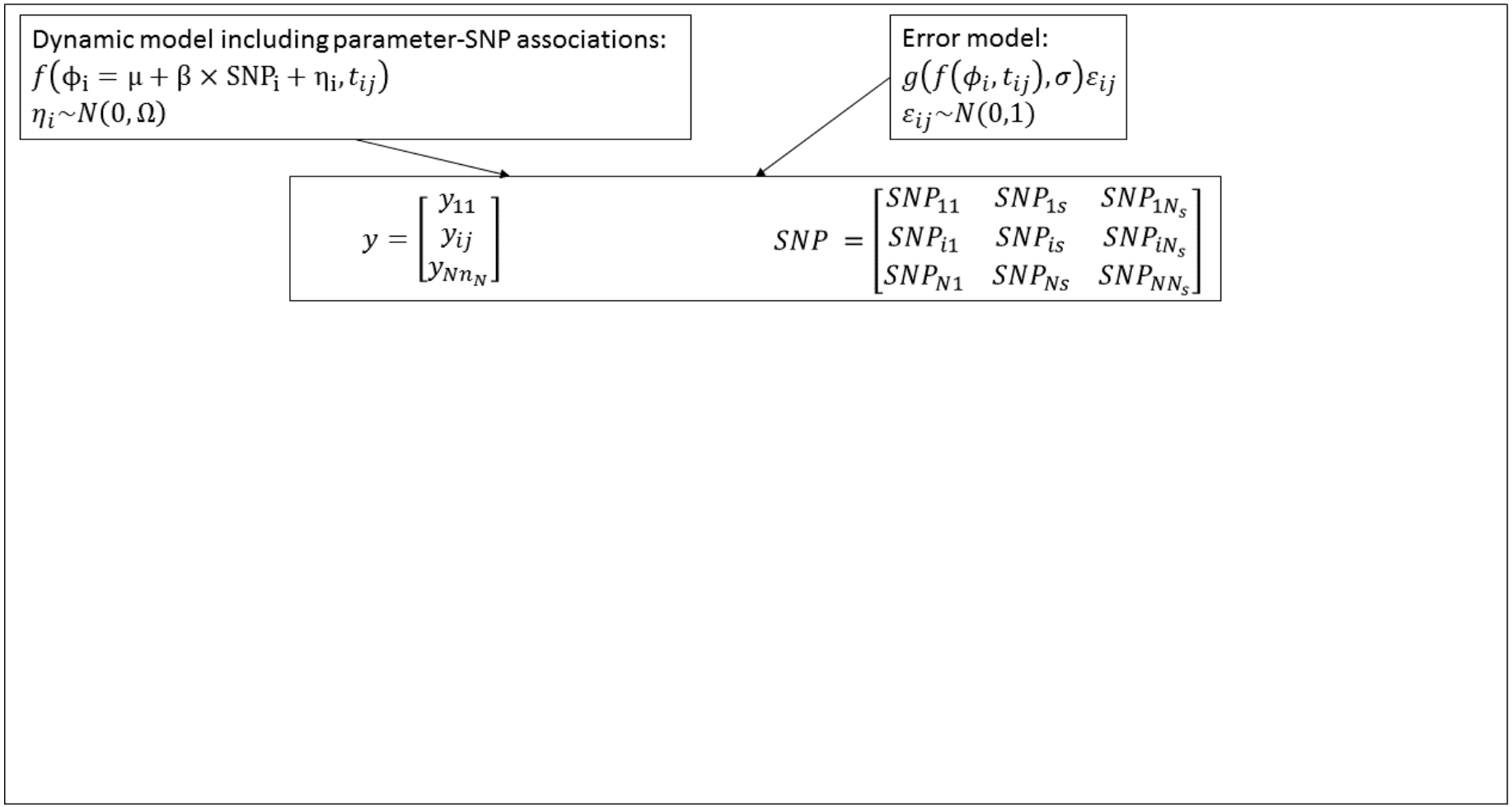
↪ SNP selection after estimation of model parameters

↪ SNP considered independently

Integrated appr. (Bertrand et al. 2015)

$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix} \qquad SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

Integrated appr. (Bertrand et al. 2015)



Integrated appr. (Bertrand et al. 2015)

Dynamic model including parameter-SNP associations:

$f(\phi_i = \mu + \beta \times \mathrm{SNP_i} + \eta_i, t_{ij})$

$\eta_i \sim N(0, \Omega)$

Error model:

$g(f(\phi_i, t_{ij}), \sigma)\varepsilon_{ij}$

$\varepsilon_{ij} \sim N(0,1)$

$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix} \qquad SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

SAEM

Stochastic Approximation Expectation-step at iteration k

ϕ'_{ik} drawn from $p(.|y; \theta_k)$, Metropolis Hasting algorithm

$s_{\mathrm{ik}} = s_{\mathrm{ik}} + \tau_k(\phi'_{\mathrm{ik}} - s_{\mathrm{ik-1}})$, τ_k sequence of decreasing positive numbers

Maximization-step of μ and β at iteration k

$(\widehat{\mu, \beta}) = \mathrm{argmin}_{\mu,\beta} \sum_{i=1}^{N} (s_{\mathrm{ik}} - \mu - \beta \times \mathrm{SNP_i}) \Omega^{-1} (s_{\mathrm{ik}} - \mu - \beta \times \mathrm{SNP_i})$

$\hat{\theta} = [\hat{\mu}, \hat{\beta}, \hat{\omega}, \hat{\sigma}]$

Integrated appr. (Bertrand et al. 2015)

Dynamic model including parameter-SNP associations:
$f(\phi_i = \mu + \beta \times SNP_i + \eta_i, t_{ij})$
$\eta_i \sim N(0, \Omega)$

Error model:
$g(f(\phi_i, t_{ij}), \sigma)\varepsilon_{ij}$
$\varepsilon_{ij} \sim N(0,1)$

$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix} \qquad SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

SAEM modified

Stochastic Approximation Expectation-step at iteration k

ϕ'_{ik} drawn from $p(.|y; \theta_k)$, Metropolis Hasting algorithm

$s_{ik} = s_{ik} + \tau_k(\phi'_{ik} - s_{ik-1})$, τ_k sequence of decreasing positive numbers

Maximization-step of μ and β at iteration k

$(\widehat{\mu, \beta}) = \text{argmin}_{\mu,\beta} \sum_{i=1}^{N} (s_{ik} - \mu - \beta \times SNP_i)\Omega^{-1}(s_{ik} - \mu - \beta \times SNP_i) + P(\beta)$

Lasso: $P_\xi(\beta)$ approx. a double exponential prior on β

Hlasso: $P_{\gamma,\lambda}(\beta)$ approx. a normal exponential gamma prior on β

ξ and γ set using an asymptotic approximation

$\hat{\theta} = [\hat{\mu}, \hat{\beta}, \hat{\omega}, \hat{\sigma}]$

↪ Simultaneous SNP selection and estimation of model parameters

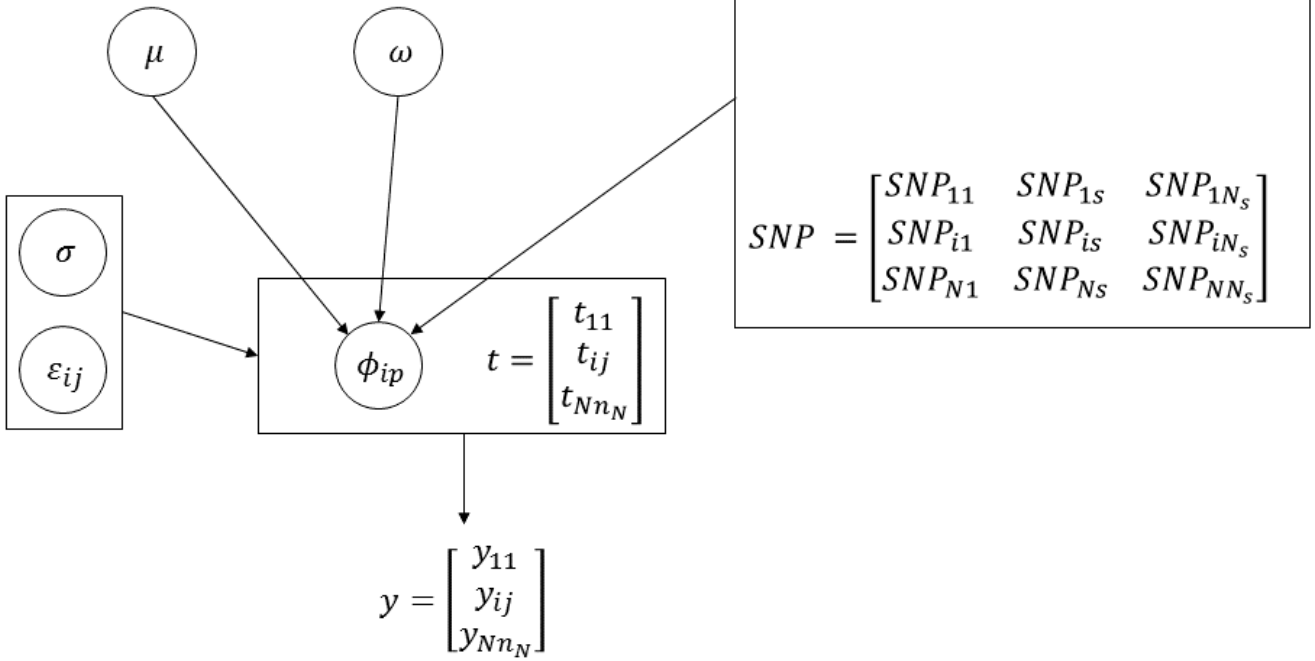
↪ All parameter-SNP associations considered simultaneously

Bayesian approach (Kuo & Mallick 1998)

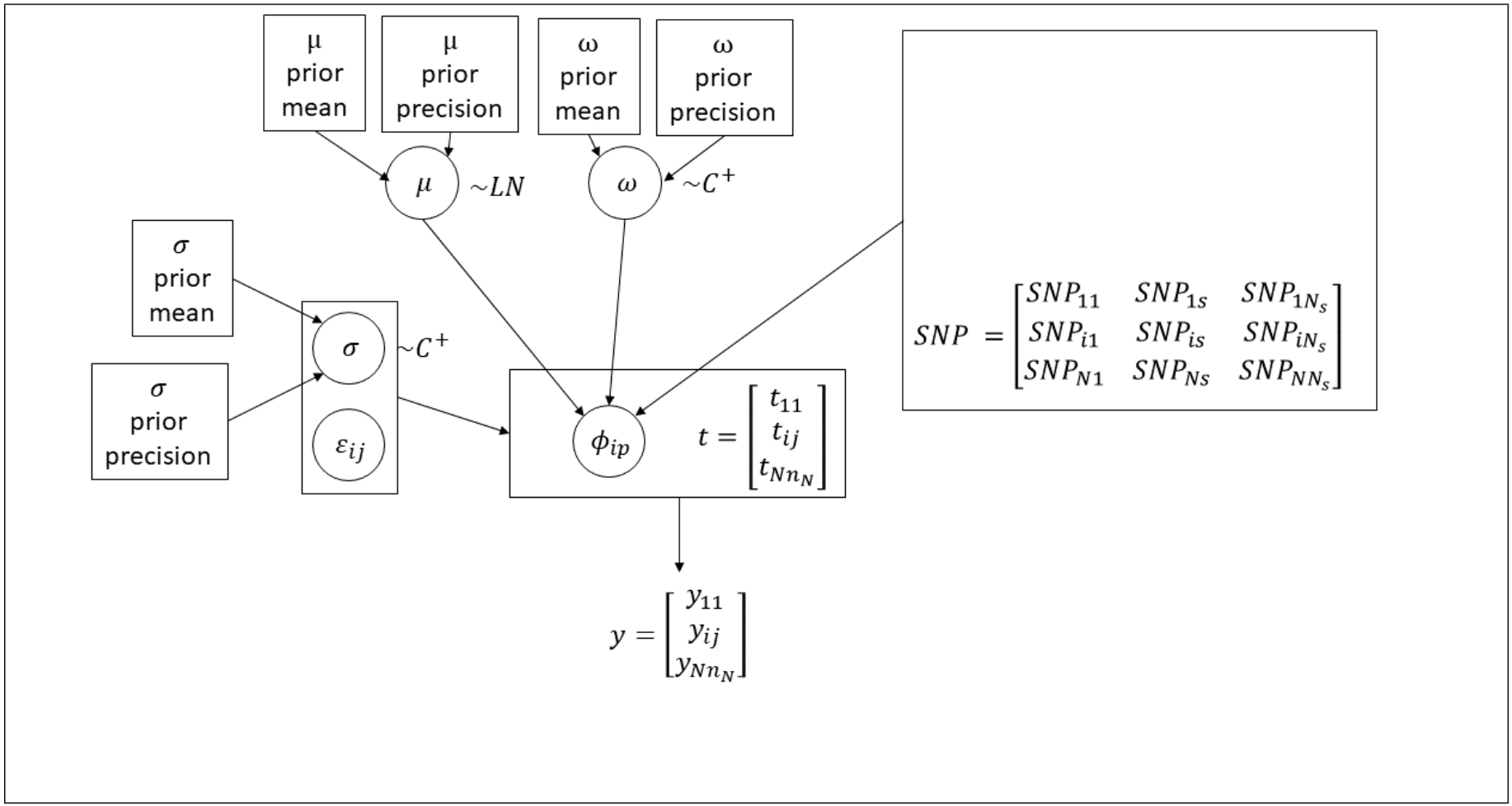
$$SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix}$$

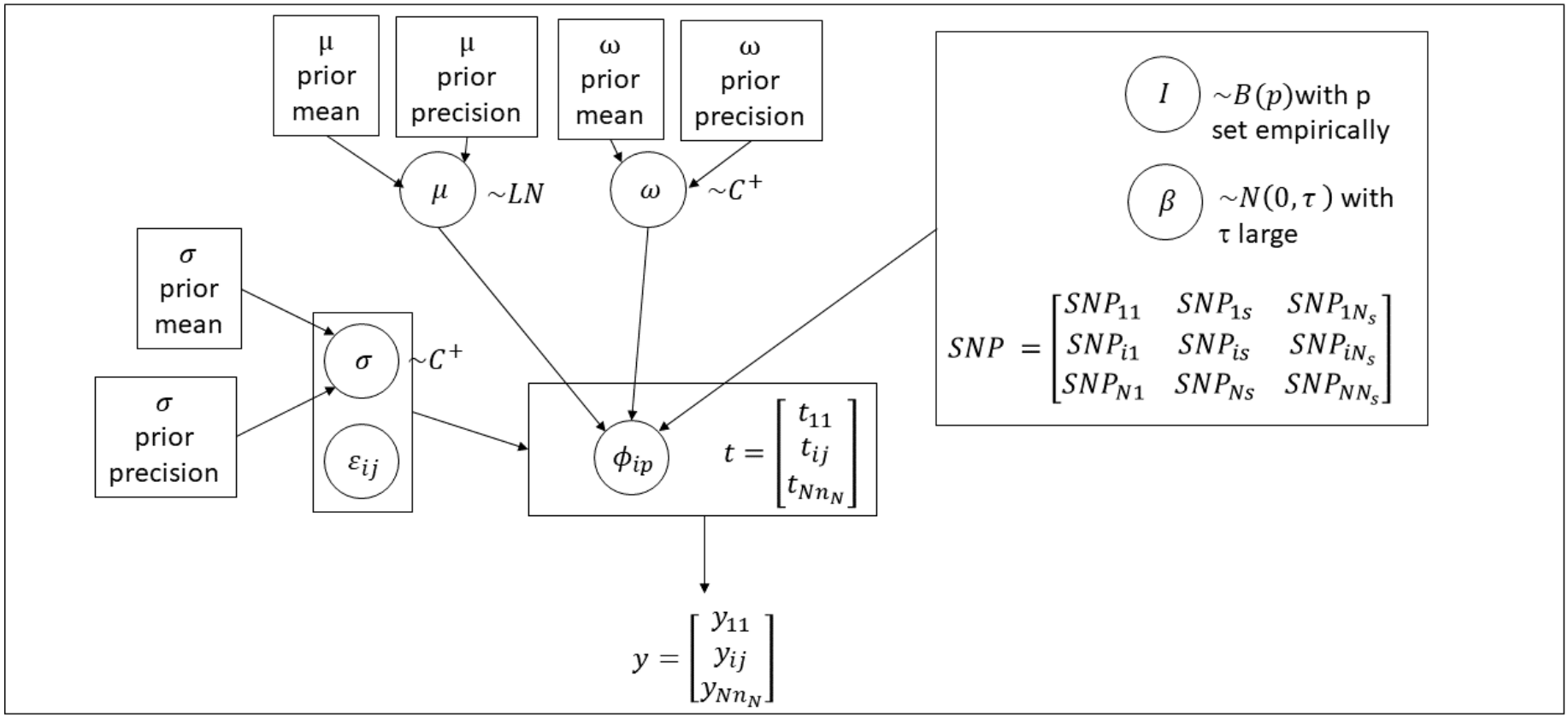
Bayesian approach (Kuo & Mallick 1998)



Bayesian approach (Kuo & Mallick 1998)



Bayesian approach (Kuo & Mallick 1998)



↪ Simultaneous SNP selection and estimation of model parameters

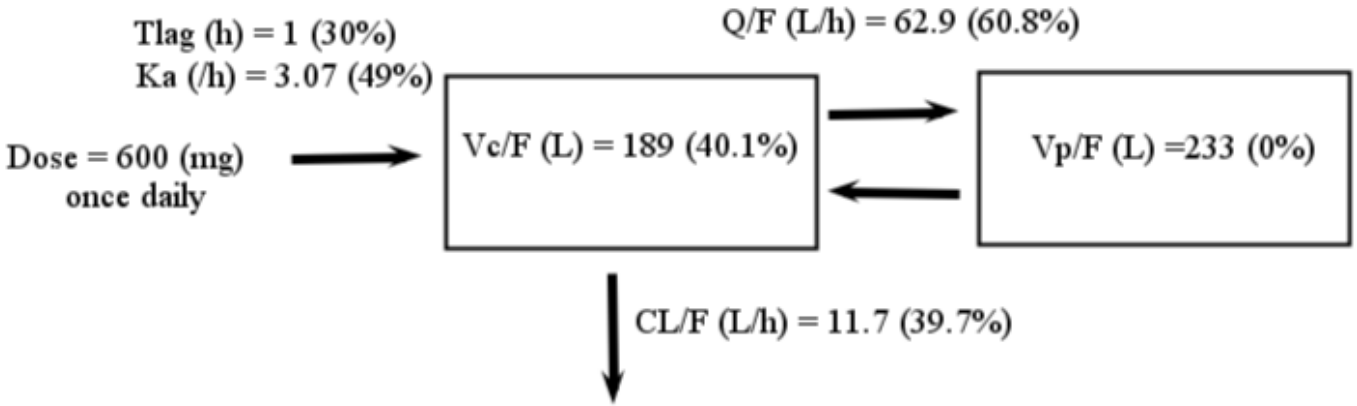
↪ All parameter-SNP associations considered simultaneously

Objectives

- To evaluate, through a realistic simulation study, the performance to detect a pharmacogenetic association
  - stepwise procedure
  - integrated approach with penalized regression
  - bayesian approach

Genetic and PK data

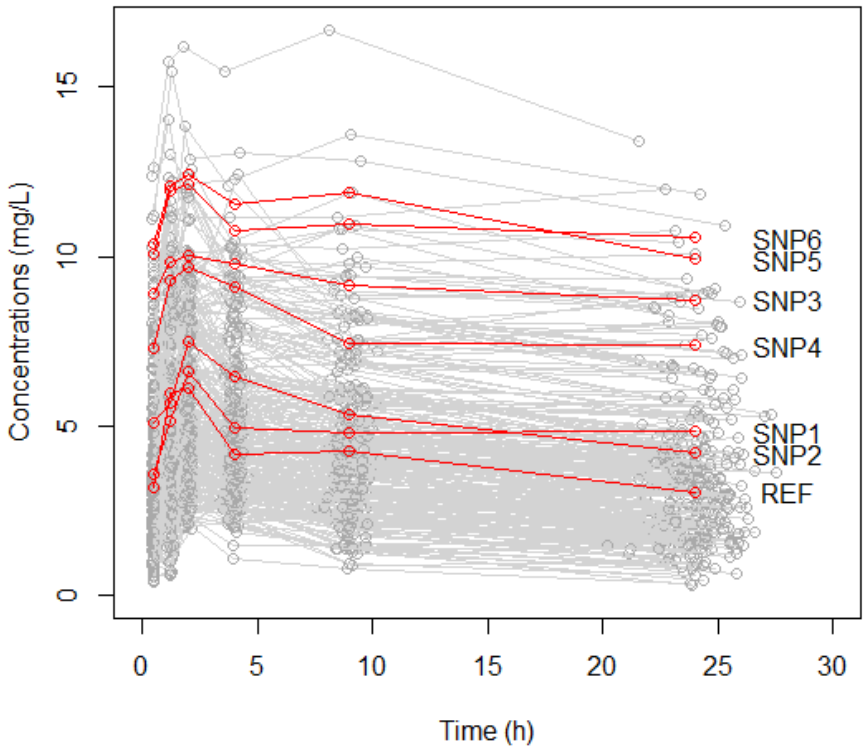
- Generation of genotypes using HAPGEN (Su et al. 2011)
  - N_s=1227 snps on 171 genes from the DMET Chip (Daly et al. 2007)
  - 6 [1-56] snps per gene
  - HAPMAP caucasian reference haplotypes
- Pharmacokinetic profiles inspired from real study (Kappelhoff et al. 2005)



  - phase II study: N=300/t=0.5, 1.25, 2, 4, 9, 24h
  - diagonal variance matrix of random effects
  - combined residual error model $g = \sigma_{inter} + \sigma_{prop} f(\phi_i, t_{ij})$

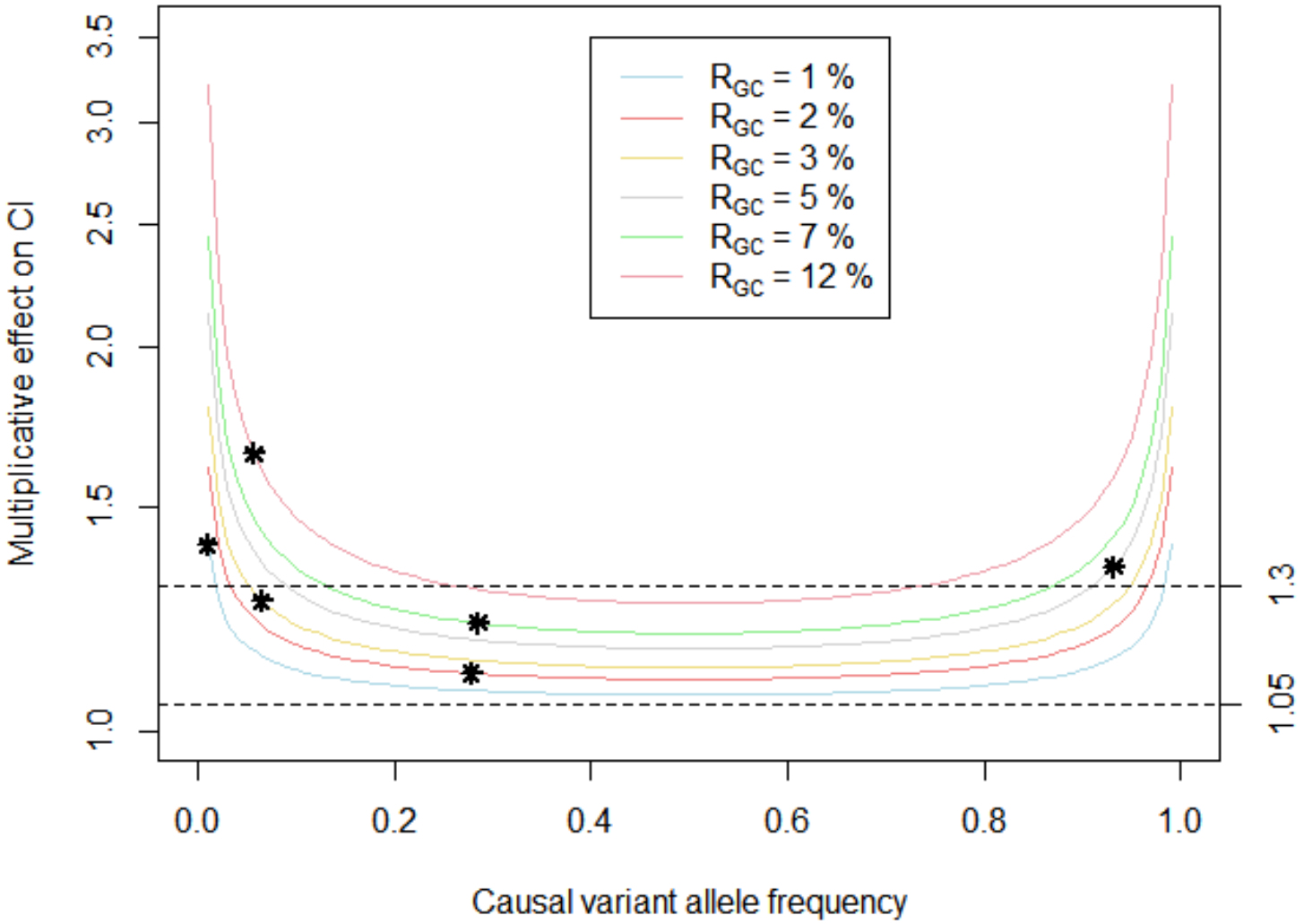
Pharmacogenetic effect - 1/2

- 6 unobserved causal variants explaining 30% of the PK

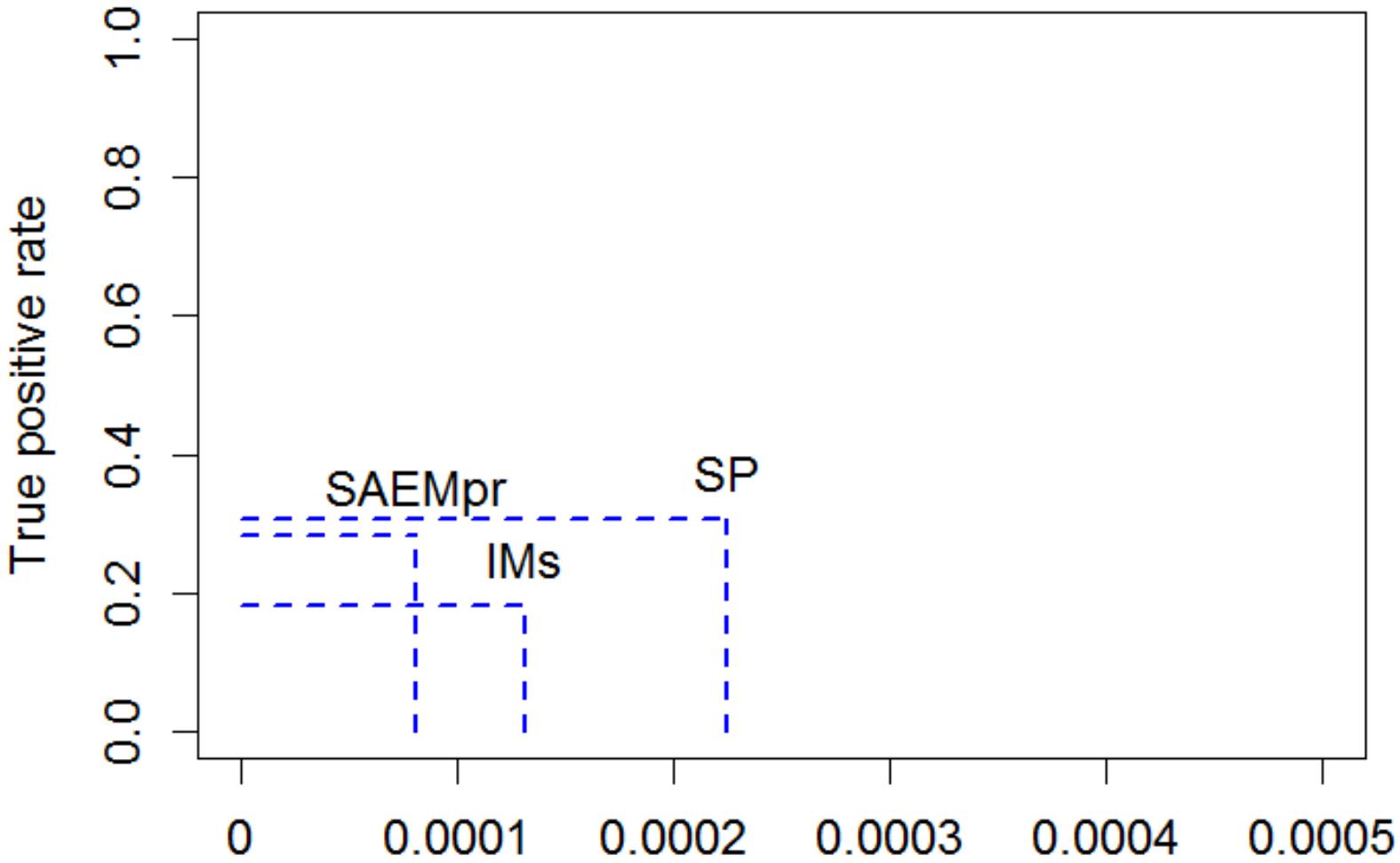


- $\log\ Cl_i = \log\ Cl + \sum_{s=1}^{6} \beta_{Cl_s} SNP_{s_i} + \eta_{Cl_i}$ with $SNP_{s_i} = 0, 1, 2$ i.e. allele dosage model

Pharmacogenetic effect - 2/2



True and false positive rates



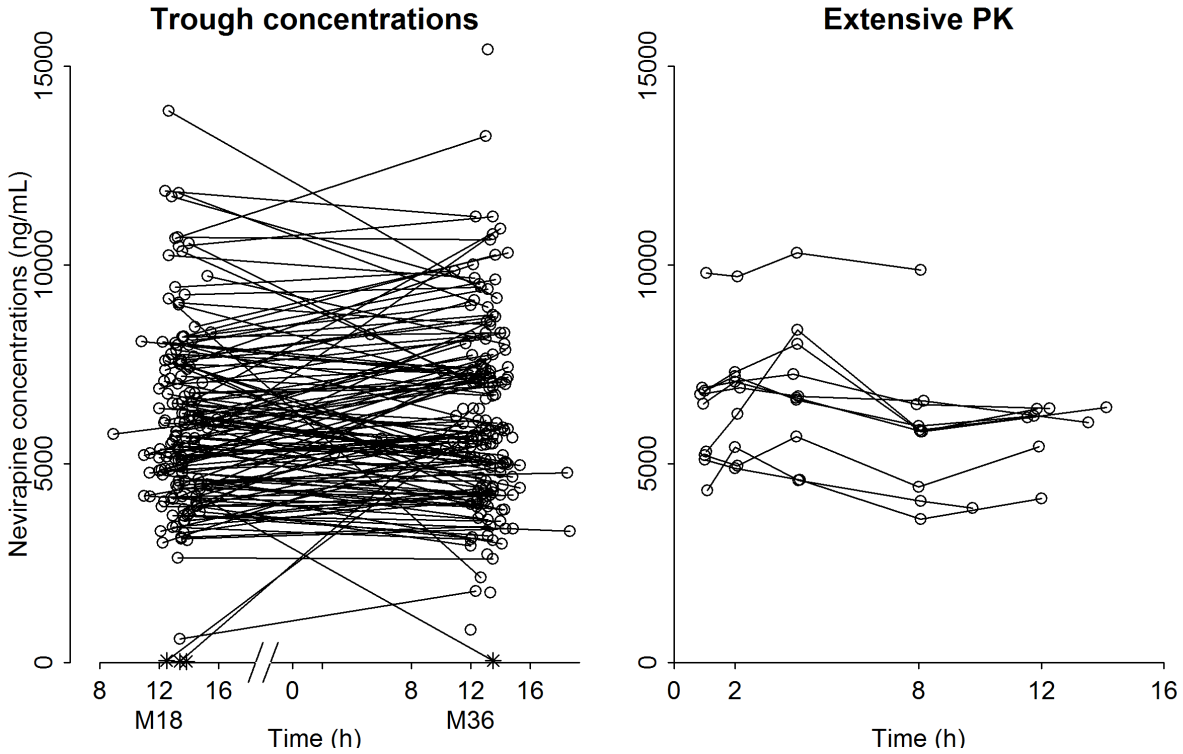
Computing times in hours - mean [range]

Stepwise procedure	**0.24 [0.06 - 1.09]**
Integrative appr.	**1.14[0.83 - 1.61]**
Bayesian appr.	**19.58 [11.51 - 23.12]**

Conclusions

- Feasibility of model-based PGx analyses on a real-case scenario
  - real need of increased sample size compared to classical drug development study designs
- Similar TPR between Integrative appr. and stepwise procedure
  - cost in computing time non-negligible
  - less so on larger data sets (not shown today)
  - better performance when multiple parameter-SNP associations (not shown today)
- Bayesian appr. performance not yet competitive
  - other indicator-based selection and/or shrinkage priors
  - use of Hamiltonian Monte Carlo (rstan package)

PECAN ANRS 12154 (Dr AM Taburet and Prof D Haas)



- 129 patients on up to 3 occasions with 196 markers

Chromosome	3	7		19	
Gene	*NR1I2* (PXR)	ABCB1 (P-gp)	*CYP3A5*	*CYP3A4*	*CYP2A6*	*CYP2B6*
Number of markers	49	63	1	36	1	47

- 218 missing polymorphisms with a maximum of 7 per subject

Analyses

- One compartment model with 1st-order absorption and elimination
- Inter-individual and inter-occasion variability on CL

1 Stepwise procedure on empirical Bayes estimates with a priori adjustment for rs3745274
   - missing data removed

2 Indicator Model selection in `JAGS`
   - missing data imputed from Binomial with empirical allele frequency

2 Shrinkage prior in `rstan`
   - missing data imputed beforehand to most common genotype

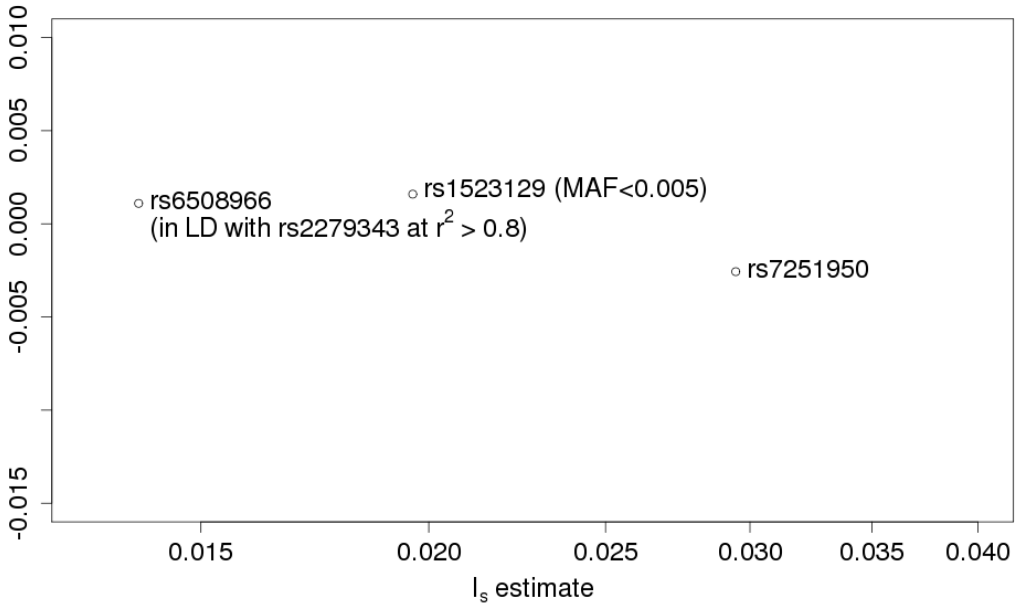
Stepwise procedure (Bertrand et al. 2012)

Table 3 Multivariate linear regression for association between genetic variants and nevirapine clearance estimate

Variants	Gene	β	Statistic	P value
$n=129$ (all participants)				
rs3745274	*CYP2B6*	−1.606	−6.66	1.01×10^{-9}
rs2687116	*CYP3A4*	−1.913	−4.68	7.95×10^{-6}
rs2279343	*CYP2B6*	0.904	4.13	7.07×10^{-5}
rs7251950	*CYP2B6*	−0.499	−4.23	4.77×10^{-5}
$n=128$ (without outlier)[a]				
rs3745274	*CYP2B6*	−1.568	−7.66	6.78×10^{-12}
rs2279343	*CYP2B6*	0.835	4.50	1.66×10^{-5}
rs7251950	*CYP2B6*	−0.404	−4.01	1.09×10^{-4}
rs2032582A	*ABCB1*	0.456	3.60	4.73×10^{-4}

[a]Data from one individual outlier (extremely high) nevirapine clearance value were censored from this analysis.

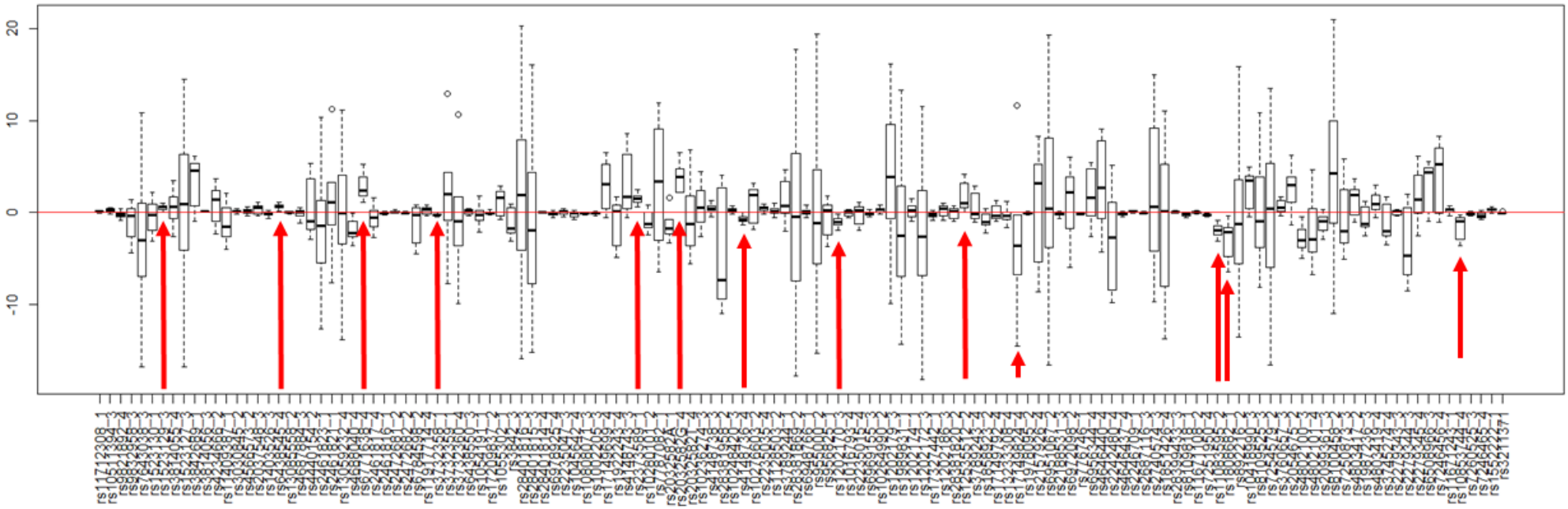
Indicator Model selection in JAGS (Kuo & Mallick 1998)



- $\beta = I \times \beta_s$

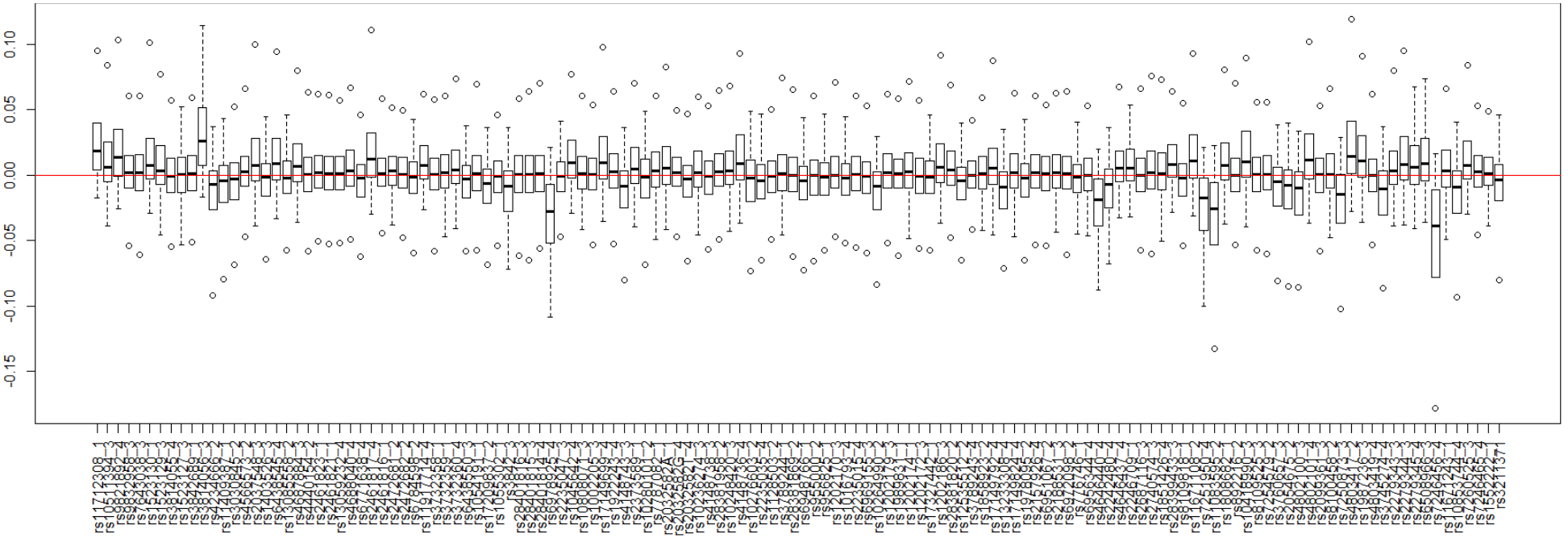
  with $\beta_s \sim N(0, 1)$, $I \sim B(p_{ind})$ and $p_{ind} \sim Beta(.5, .5)$
- 3 chains / nburn=200 / niter=2000 / Neff = [15 -750] / approx. 5 min

Shrinkage prior in `rstan`



- normal prior (Hoerl & Kennard 1970)
  $\beta \sim N(0, 10)$
- 3 chains / nburn=300 / niter=900 /Neff = [81 - 663] / approx. 20 min

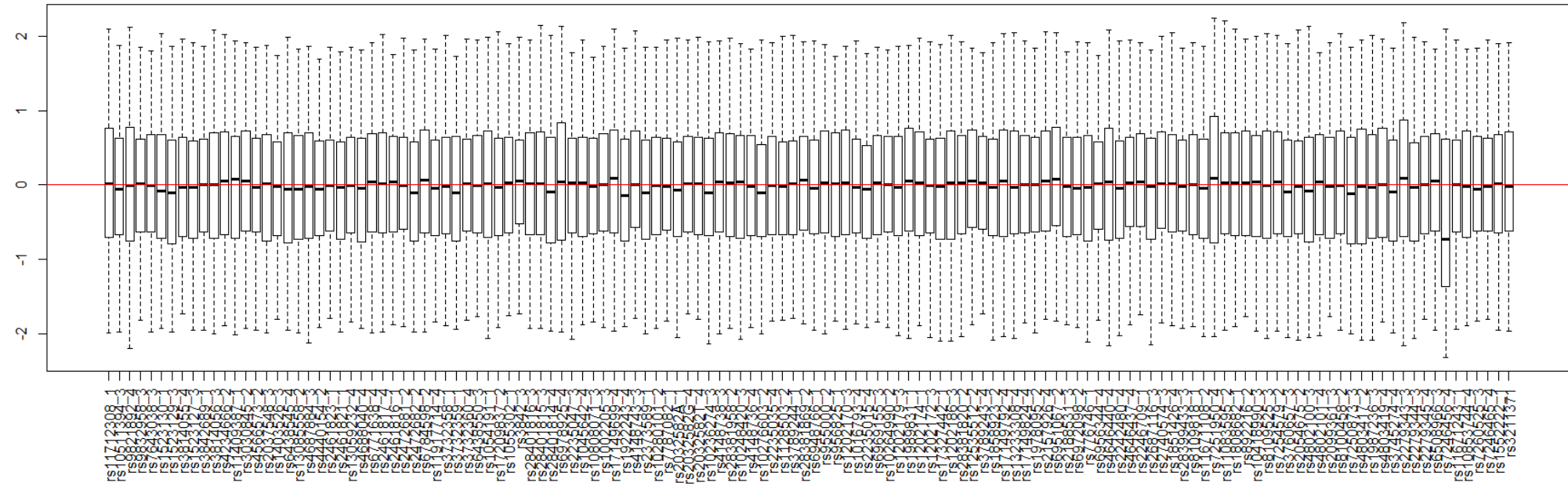
Shrinkage prior in `rstan`



- Lasso prior (Tibshirani 1996)

  $\beta \sim N(0, \tau)$ with $\tau \sim C^{+}(0, .01)$
- 3 chains / nburn=300 / niter=900 /Neff = [81 - 663] / approx. 20 min

Shrinkage prior in `rstan`



- Horseshoe prior (Peltola et al. 2014)

  $\beta \sim N(0, \lambda_i \tau)$ with $\lambda_i \sim C^+(0,1)$ and $\tau \sim U(0,1)$
- 3 chains / nburn=300 / niter=900 /Neff = [81 - 663] / approx. 20 min

Conclusion

- Stepwise approach and Bayesian variable selection in JAGs detect similar signals
- Rstan sampling marginally more efficient than JAGs
  - too strong shrinkage of Lasso and Horseshoe ?
- Work in progress with yet a lot to do...
  - run rstan on simulation study
  - investigate how to set prior parameters
  - application to HIV Swiss cohort and EARSII trial on diabetes data

Acknowledgements

Prof. David J Balding, UCL Genetics Institute, UK

Dr. Maria de Iorio, UCL Statistical Science Dept., UK

Dr Emmanuelle Comets, INSERM, France