Bayesian approaches for pharmacogenetic models with JAGS and Stan

Julie Bertrand, Maria De Iorio, David J. Balding

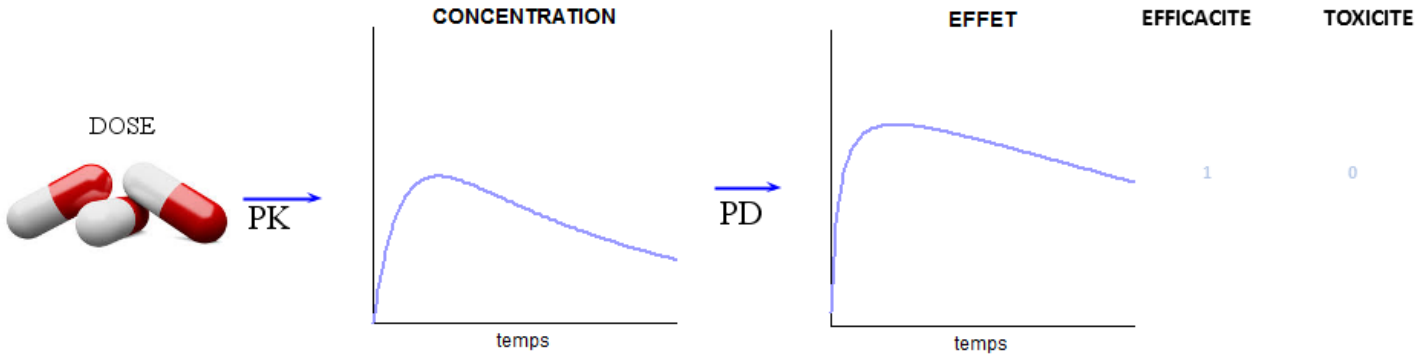
Inserm
Institut national
de la santé et de la recherche médicale

lame
UMR 1137
Infection • Antimicrobials • Modelling • Evolution

UCL

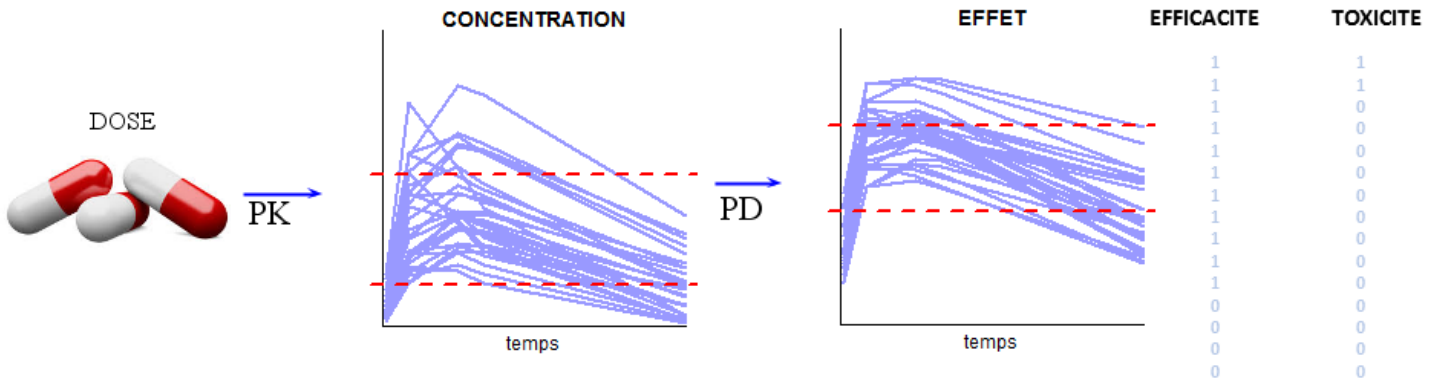
Pharmacological and genetic variability

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



Pharmacological and genetic variability

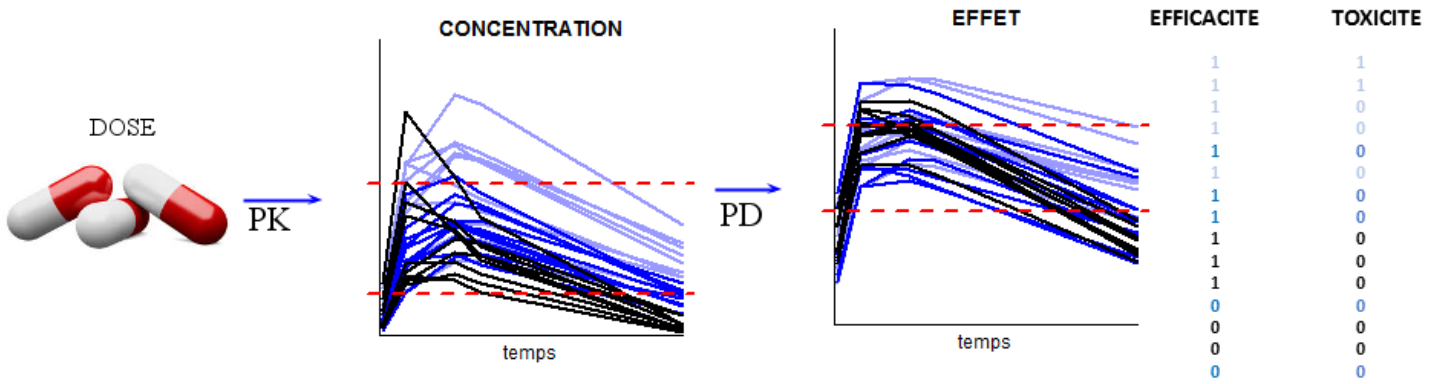
- Clinical pharmacology: study the interaction between the organism and the drug and its variability
  - pharmacokinetics (PK) and pharmacodynamics (PD)



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

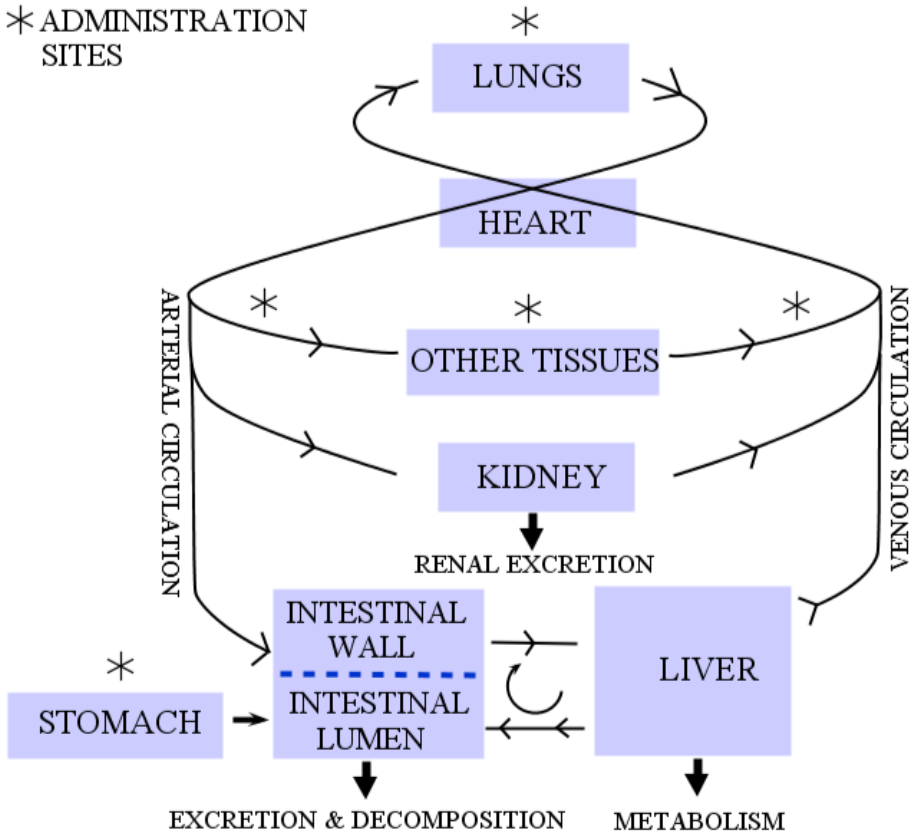


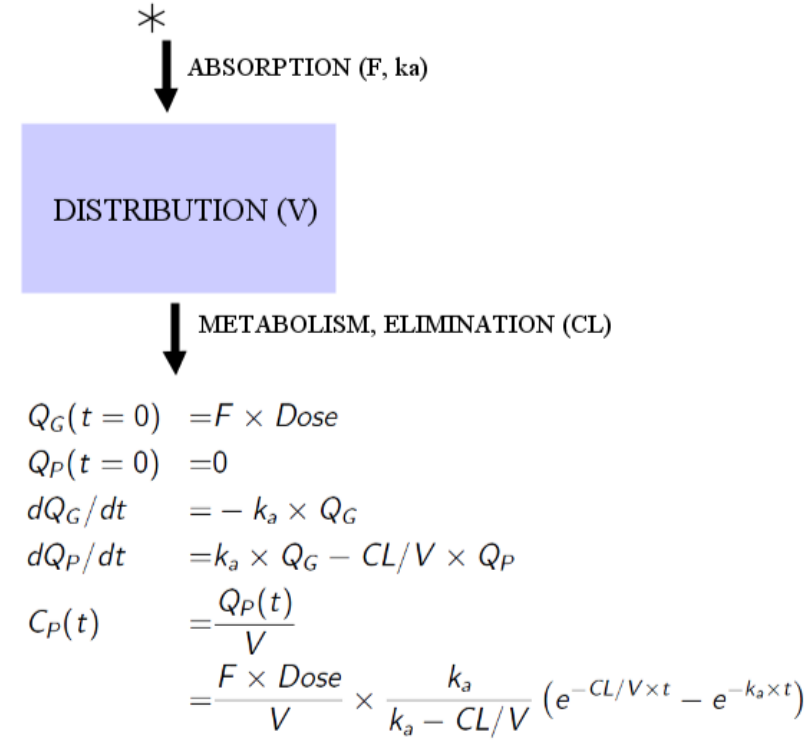
- Pharmacogenetics (PG): genetic part of the variability
  - stratified/precision medicine
  - genes coding for proteins involved in PK/PD processes
    - metabolism enzymes (CYP450, NAT)
    - single nucleotide polymorphism (SNP)

Methodological challenges in PGx

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

$\hookrightarrow$ dynamical model (nonlinear in its parameters)





$$
\begin{aligned}
Q_G(t=0) &= F \times Dose \\
Q_P(t=0) &= 0 \\
dQ_G/dt &= -k_a \times Q_G \\
dQ_P/dt &= k_a \times Q_G - CL/V \times Q_P \\
C_P(t) &= \frac{Q_P(t)}{V} \\
&= \frac{F \times Dose}{V} \times \frac{k_a}{k_a - CL/V}\left(e^{-CL/V \times t} - e^{-k_a \times t}\right)
\end{aligned}
$$

Methodological challenges in PGx

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

  $\hookrightarrow$ dynamical model (nonlinear in its parameters)

  $\hookrightarrow$ mixed effect models
    - all patients analyzed simultaneously
    - parameter decomposed in fixed and random effects

Example : $CL_i = \mu_{CL} + \eta_i$ with $\eta_i \sim N(0, \omega^2_{CL})$

Methodological challenges in PGx

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

  $\hookrightarrow$ dynamical model (nonlinear in its parameters)

  $\hookrightarrow$ mixed effect models
    - all patients analyzed simultaneously
    - parameter decomposed in fixed and random effects
    - covariates identification:

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

Methodological challenges in PGx

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

  $\hookrightarrow$ dynamical model (nonlinear in its parameters)

  $\hookrightarrow$ mixed effect models
- Increased size of genetic data sets up to high throughput screening
  - dimensionality curse $N << p$

  $\hookrightarrow$ Stepwise procedure

  $\hookrightarrow$ Penalized regression

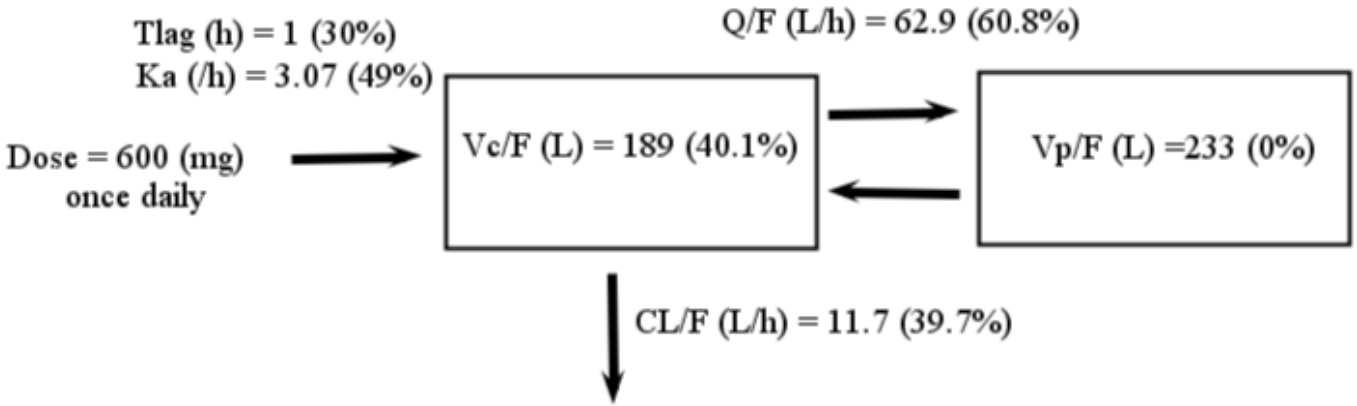
  $\hookrightarrow$ Bayesian variable selection

Objectives

- To evaluate, through a realistic simulation study, the performance to detect a pharmacogenetic association
  - stepwise procedure (Monolix software and lm R function)
  - penalized regression (saemix R package and hlasso program)
  - bayesian variable selection (R2jags)

Genetic and PK data

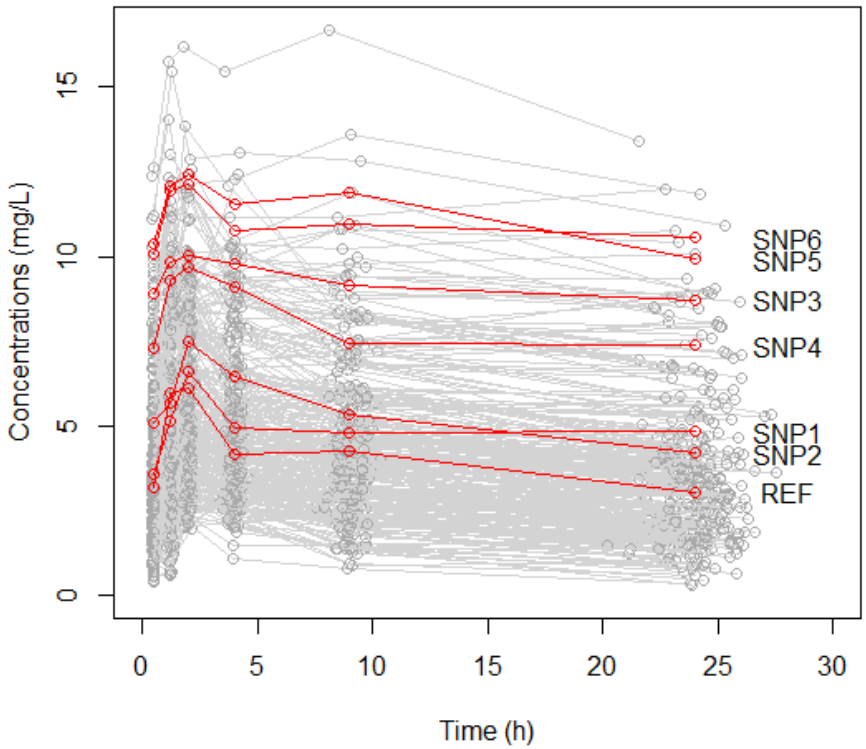
- Generation of genotypes using HAPGEN (Su et al. 2011)
  - N_s=1227 snps on 171 genes from the DMET Chip (Daly et al. 2007)
  - 6 [1-56] snps per gene
  - HAPMAP caucasian reference haplotypes
- Pharmacokinetic profiles inspired from real study (Kappelhoff et al. 2005)



  - 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

- 6 unobserved causal variants explaining 30% of the PK



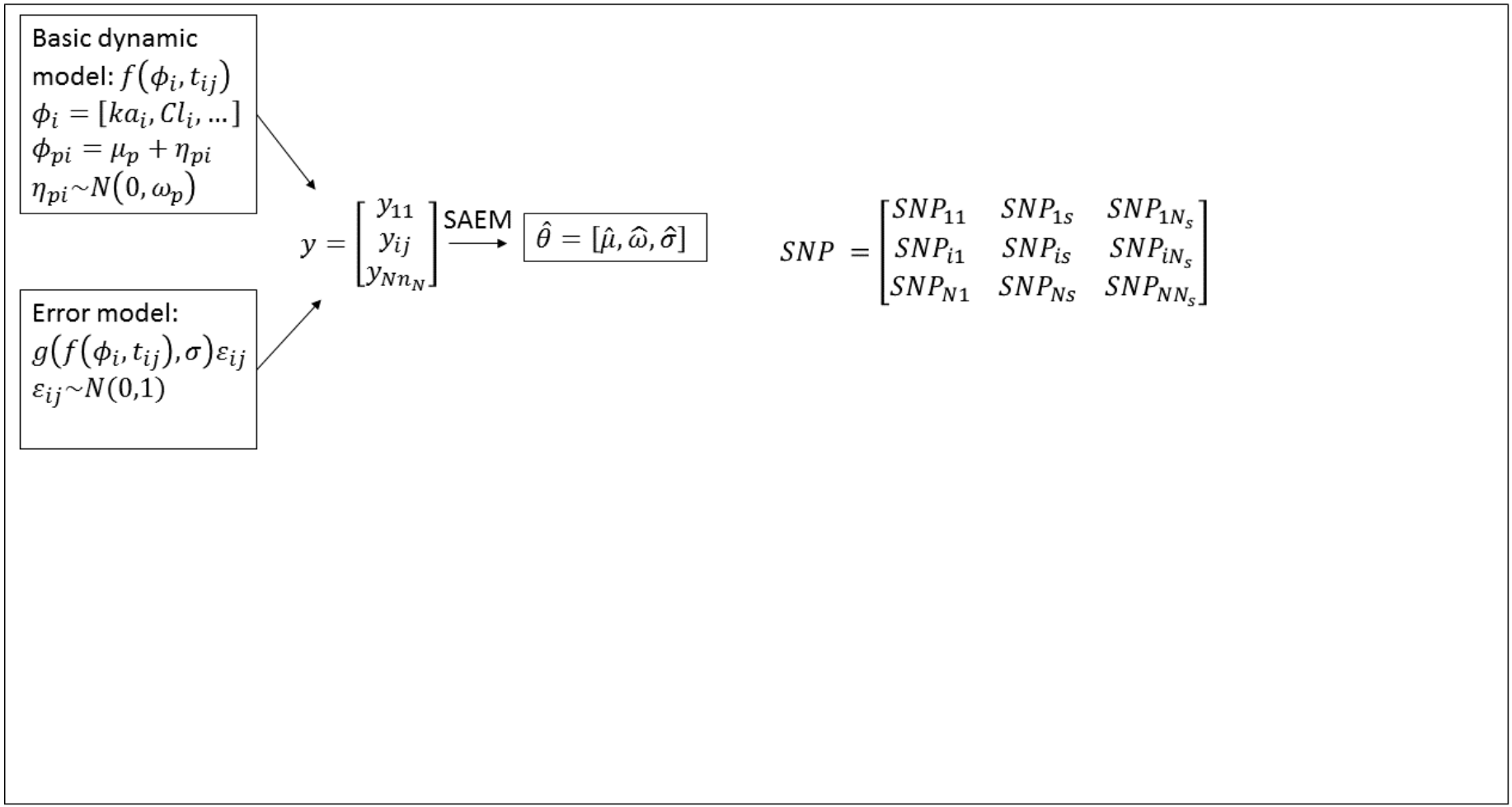
- $\log\ Cl_i = \log\ Cl + \sum_{s=1}^{6} \beta_{Cl_s} SNP_{s_i} + \eta_{Cl_i}$ with $SNP_{s_i} = 0, 1, 2$

Stepwise procedure (Lehr et al. 2010)

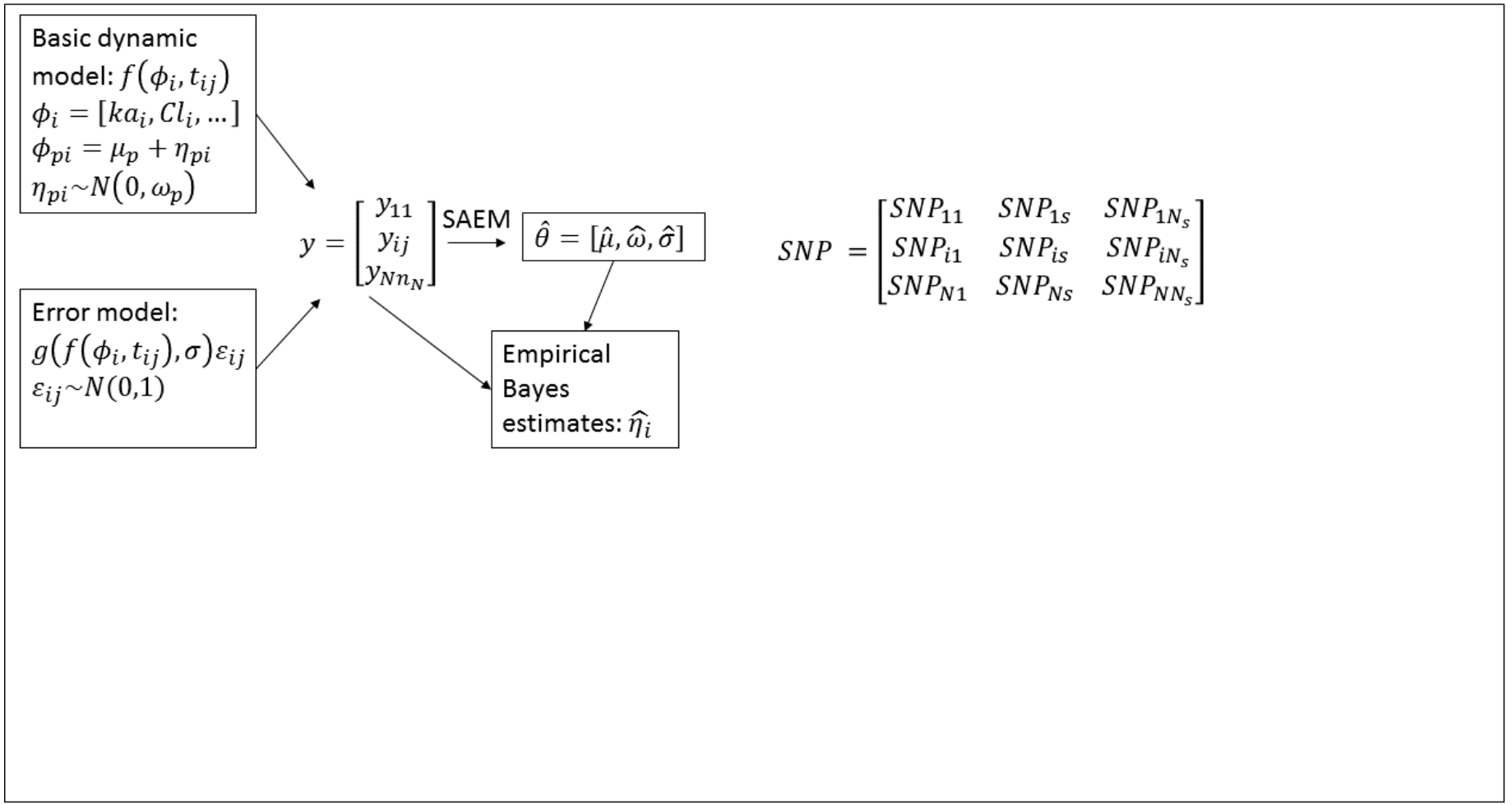
$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix}$$

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

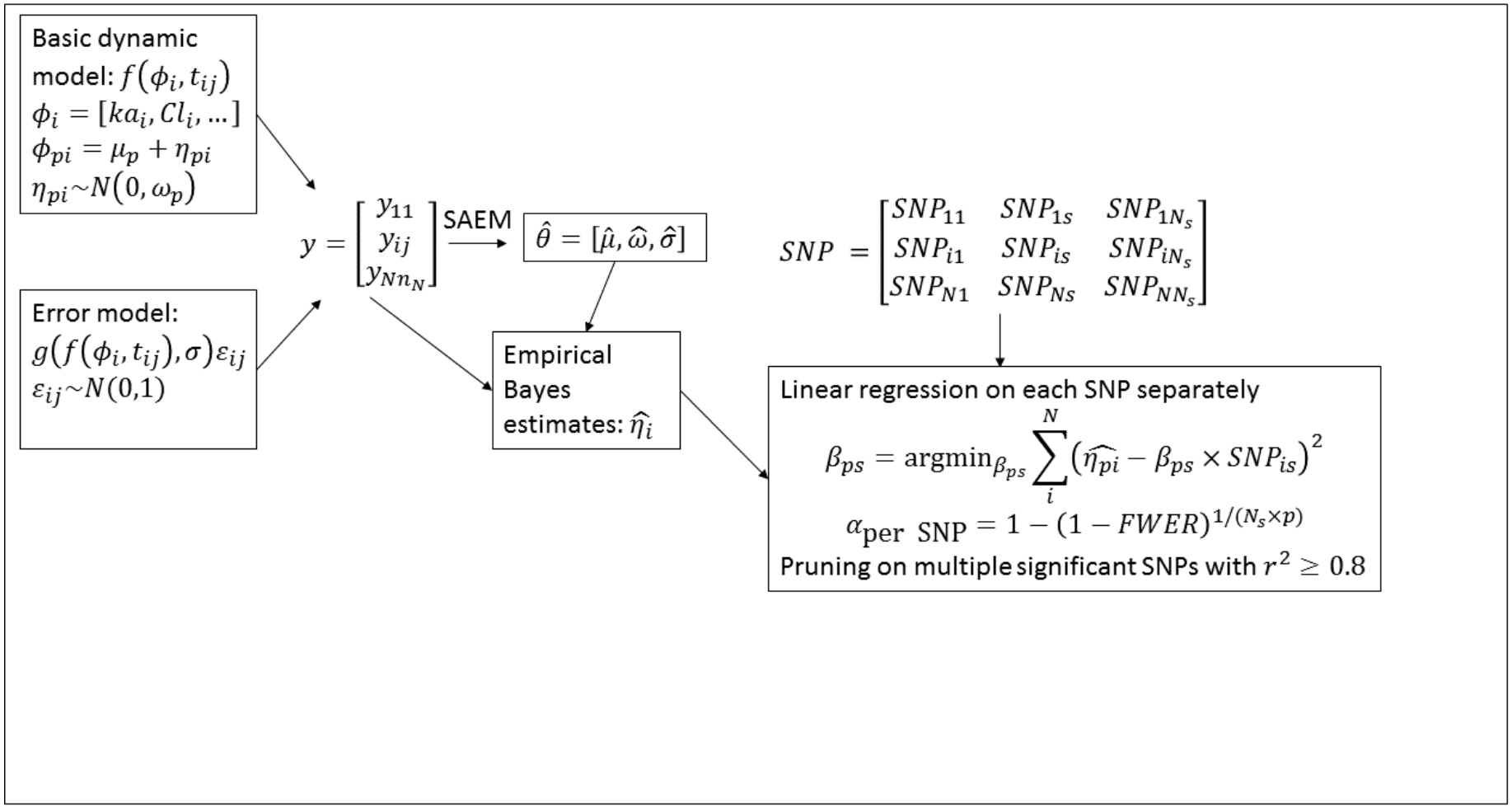
Stepwise procedure (Lehr et al. 2010)



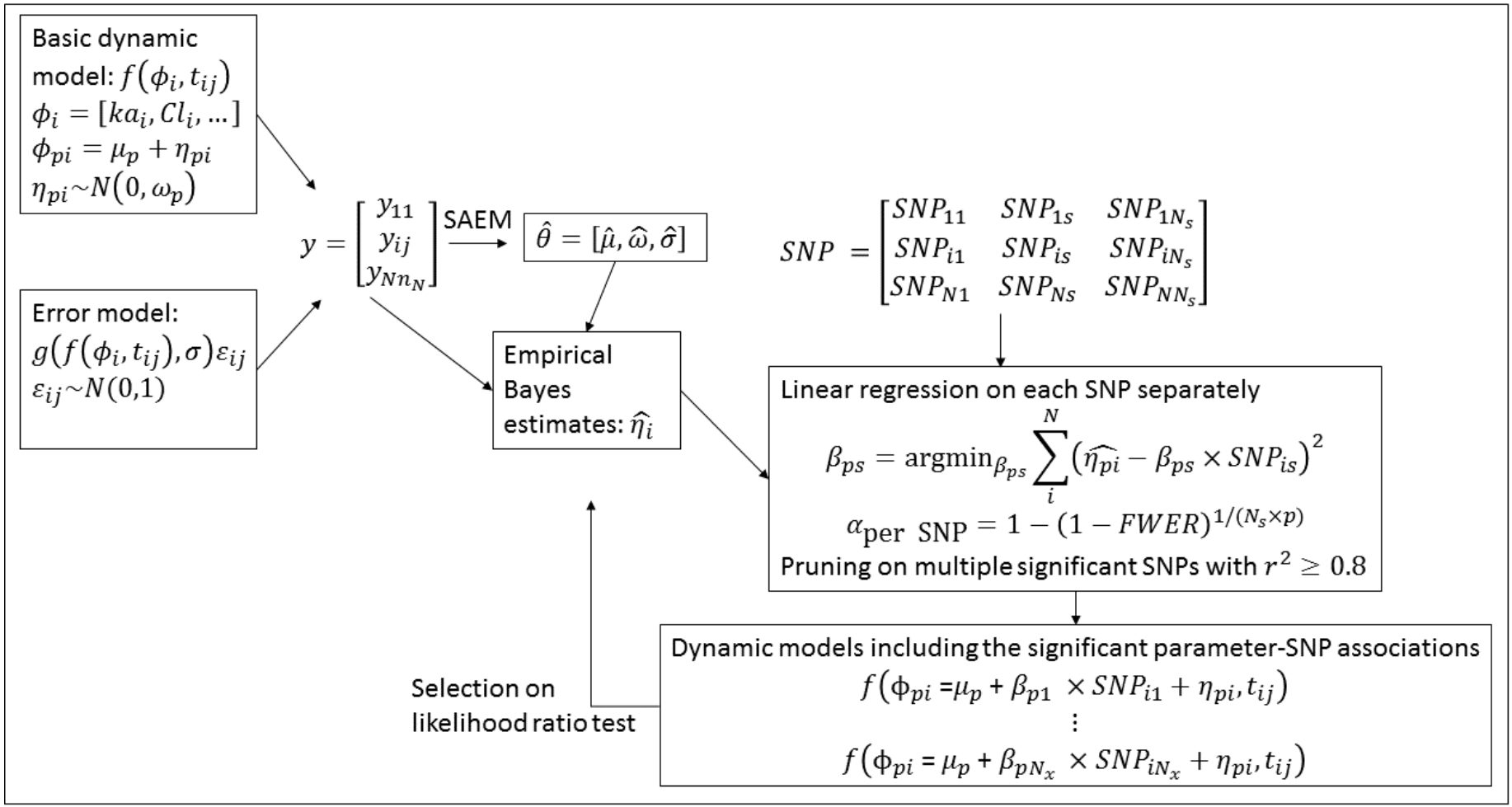
Stepwise procedure (Lehr et al. 2010)



Stepwise procedure (Lehr et al. 2010)



Stepwise procedure (Lehr et al. 2010)



↪ SNP selection after estimation of model parameters

↪ SNP considered independently

Penalized regression (Bertrand et al. 2015)

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

Penalized regression (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:

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

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

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

Penalized regression (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

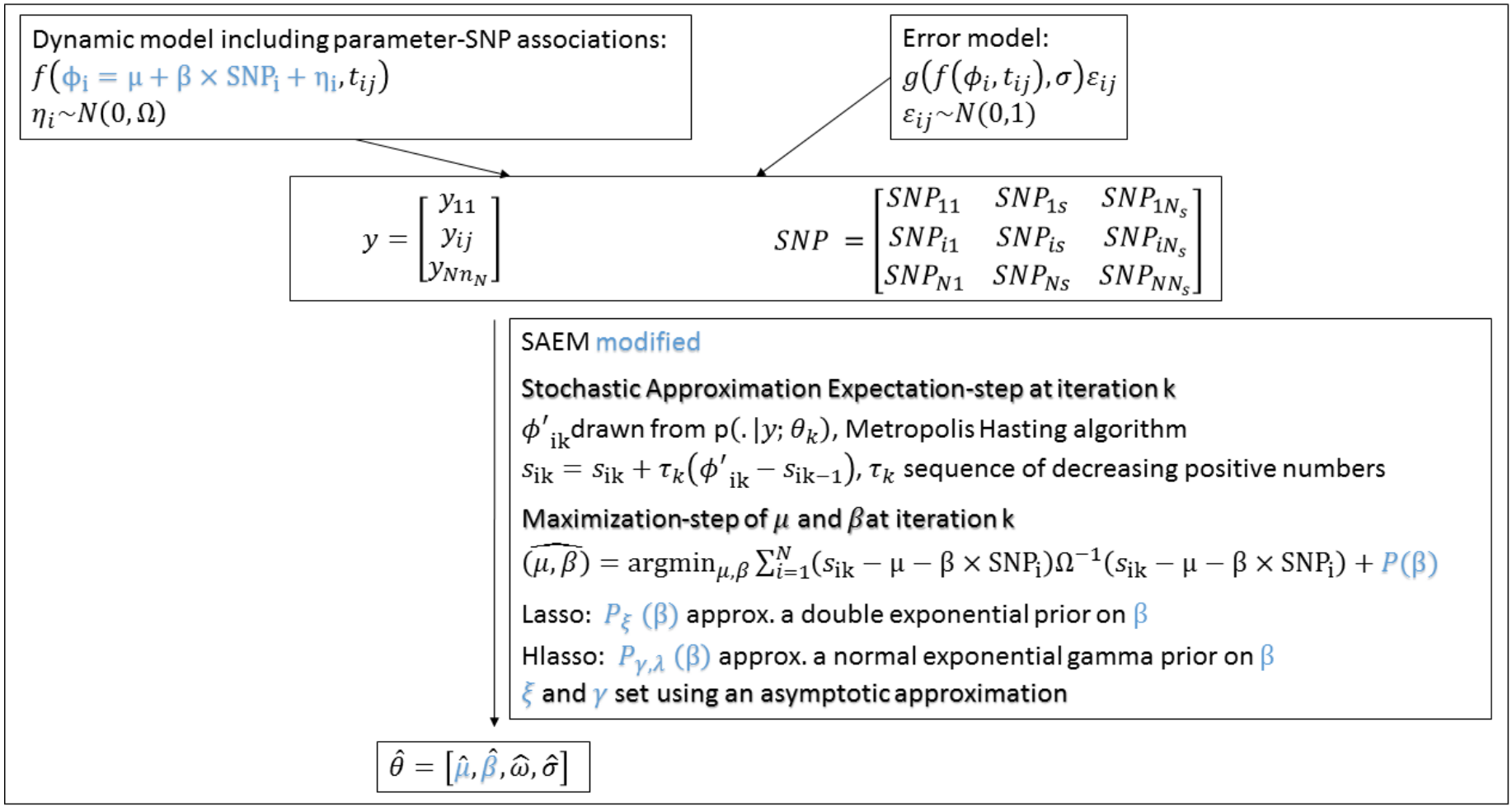
$s_{\mathrm{ik}} = s_{\mathrm{ik}} + \tau_k(\phi'_{\mathrm{ik}} - s_{\mathrm{ik-1}})$, τ_k sequence of decreasing positive numbers

Maximization-step of μ and β at iteration k

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

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

Penalized regression (Bertrand et al. 2015)



↪ Simultaneous SNP selection and estimation of model parameters

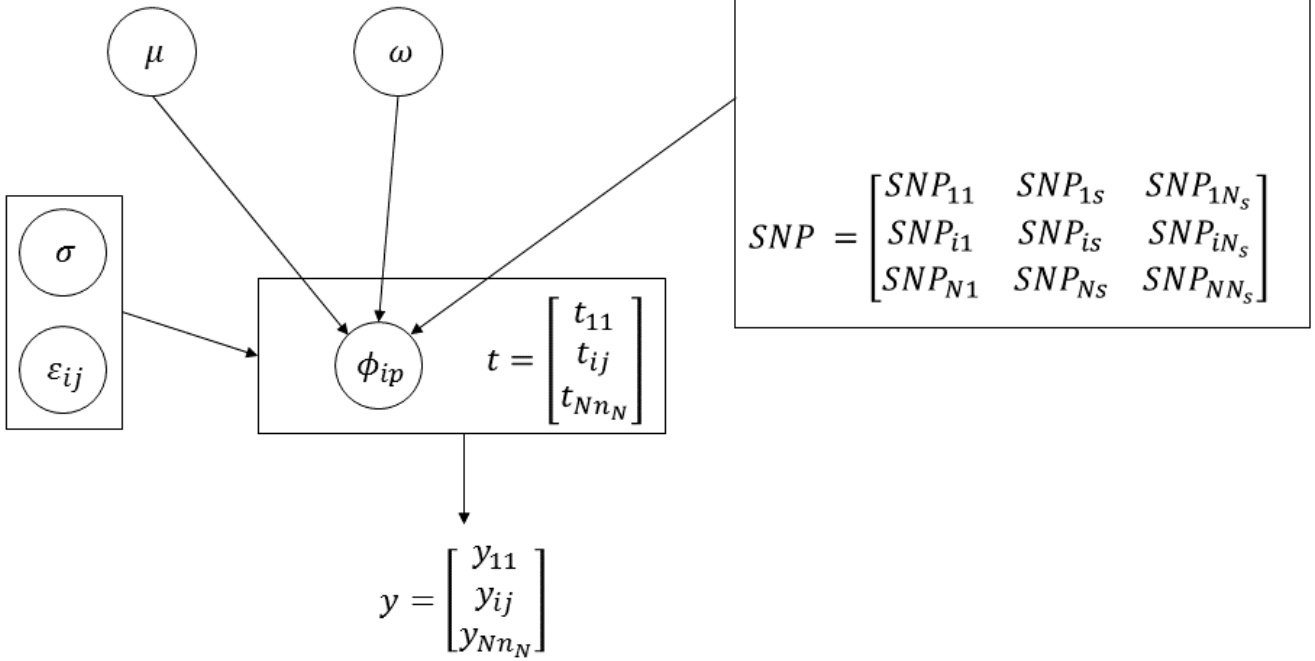
↪ All parameter-SNP associations considered simultaneously

Bayesian variable selection (Kuo & Mallick 1998)

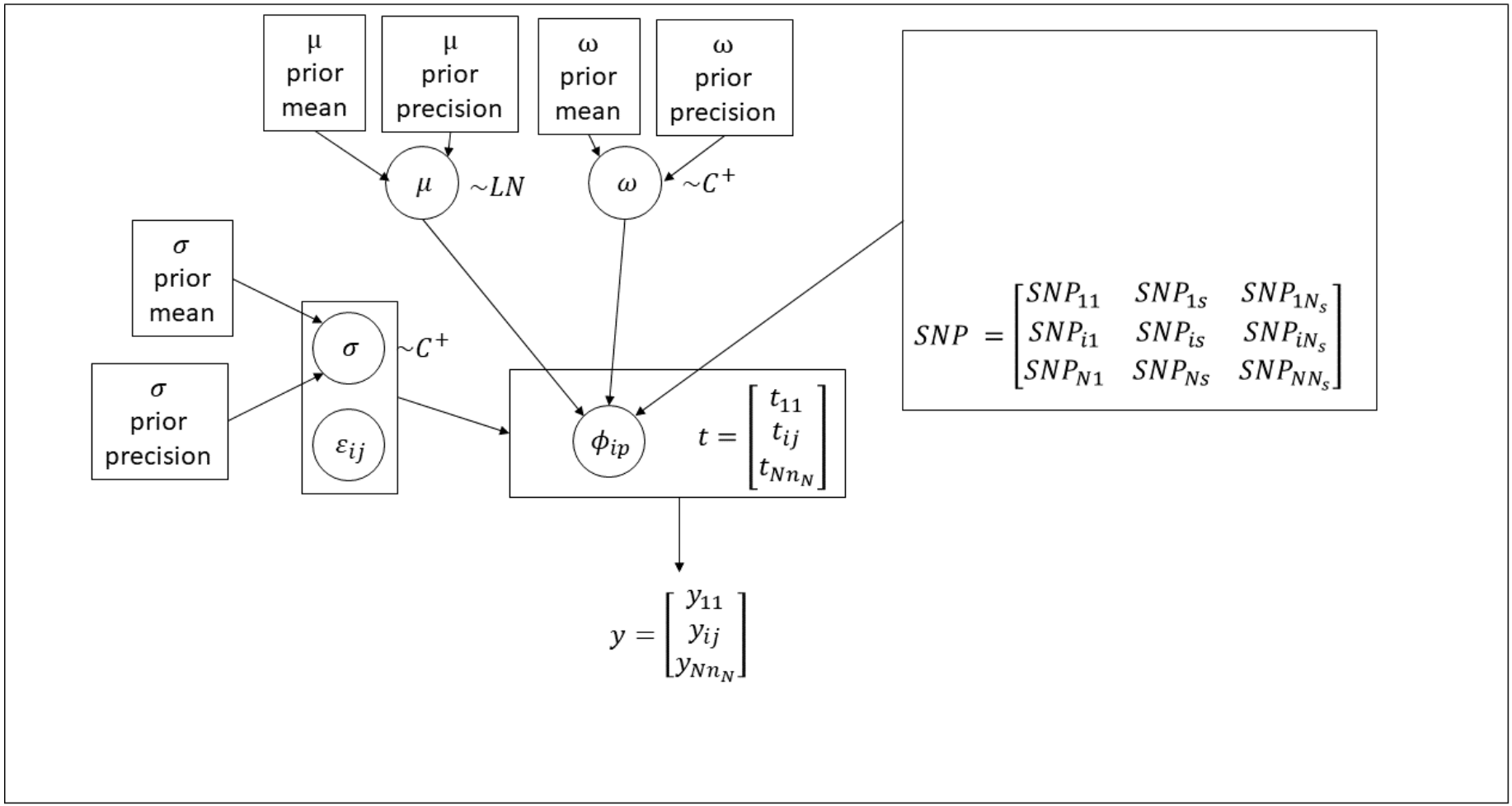
$$SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

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

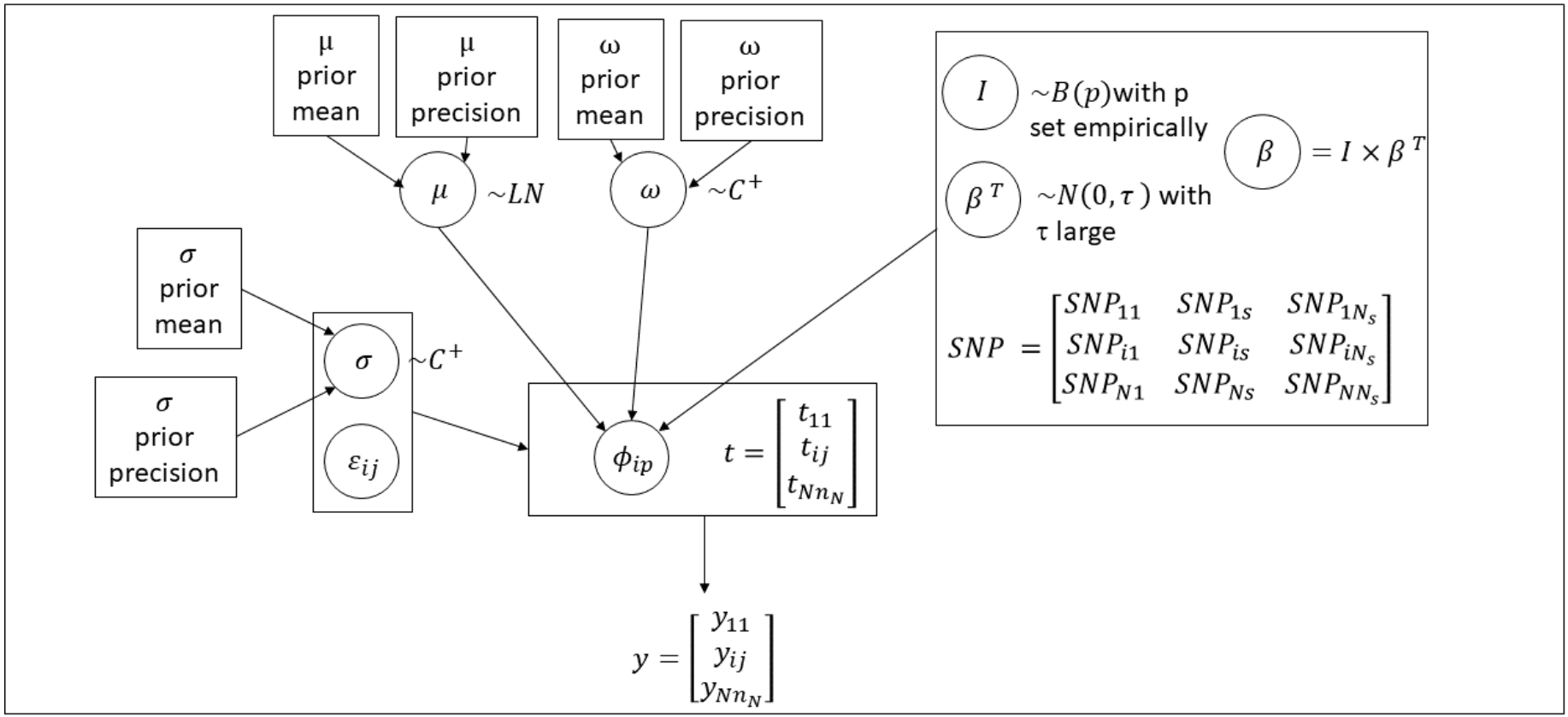
Bayesian variable selection (Kuo & Mallick 1998)



Bayesian variable selection (Kuo & Mallick 1998)



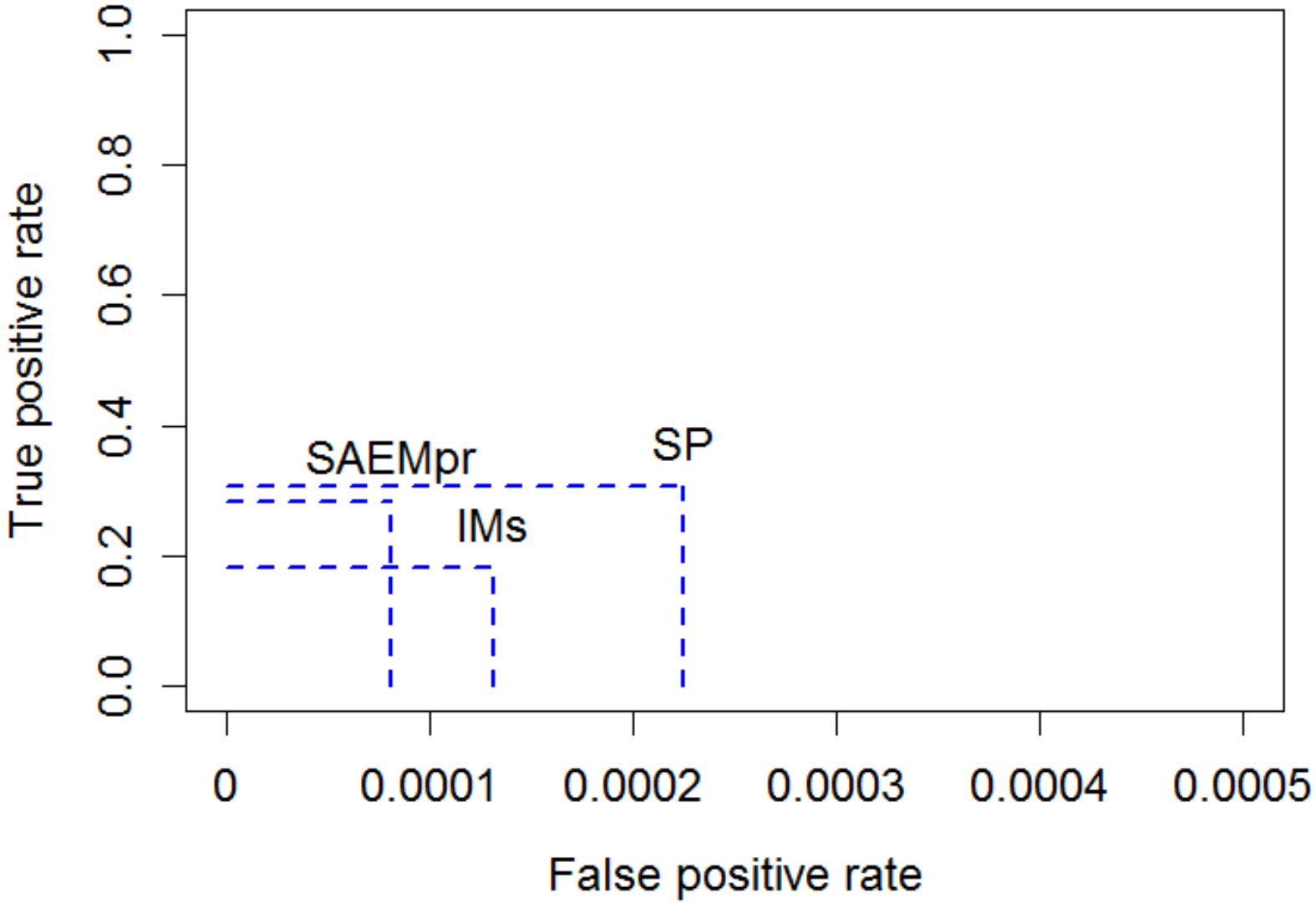
Bayesian variable selection (Kuo & Mallick 1998)



↪ Simultaneous SNP selection and estimation of model parameters

↪ All parameter-SNP associations considered simultaneously

True and false positive rates



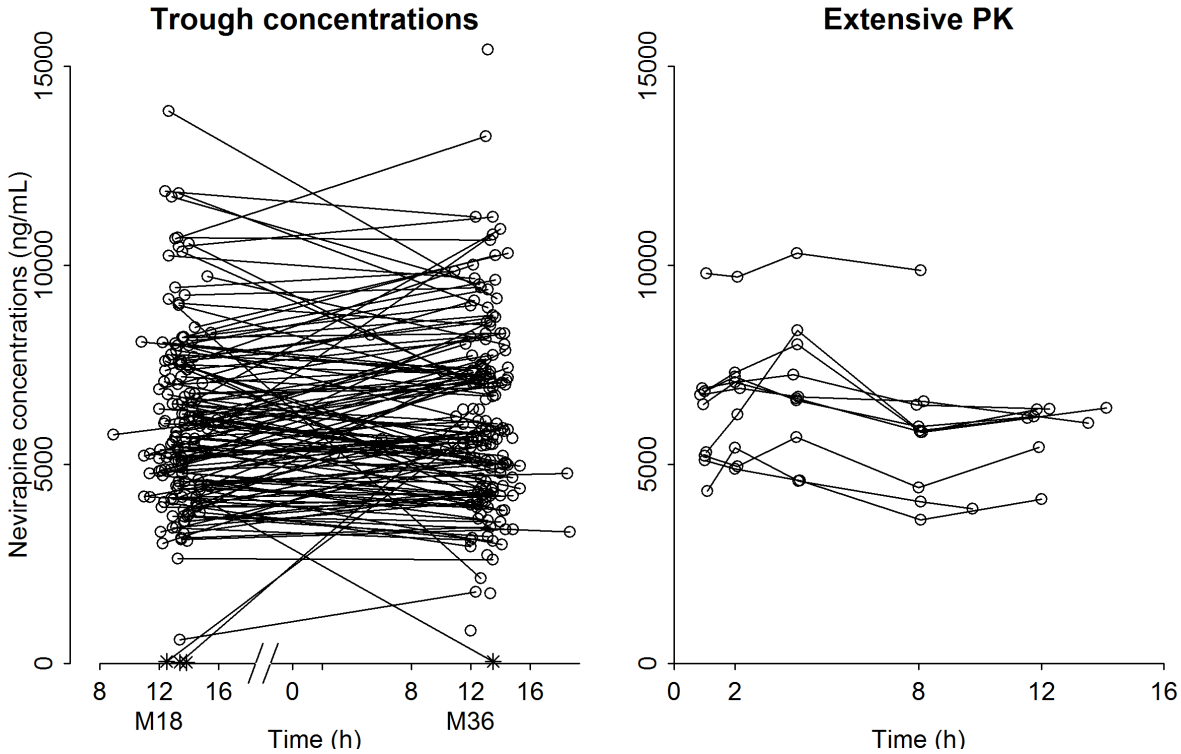
Computing times in hours - mean [range]

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 penalized regression 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 model selection (Gibbs variable selection, Stochastic search variable selection)
  - shrinkage priors (Laplacian, Horseshoe)
  - other MCMC algorithms (HMC in the 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 SNP with a maximum of 7 per subject

Analyses

- One compartment with 1st-order absorption and elimination
- Inter-individual and -occasion variability on CL/F
  - adjustment on CYP2B6 G516T (rs3745274)

1 Stepwise procedure
  - missing data removed

2 Indicator Model selection using `R2jags`
  - missing data imputed using observed allele frequency

2 Horseshoe shrinkage prior using `Rstan` (Piironen et al. 2017)

$$\beta_j|\lambda_j,\tau \sim N(0,\lambda_j^2\tau^2)$$
$$\lambda_j \sim C^+(0,1) \text{ and } \tau \sim C^+(0, p0/[(N_{SNP}-p0)\sqrt{N}])$$

  - missing data imputed to most common genotype

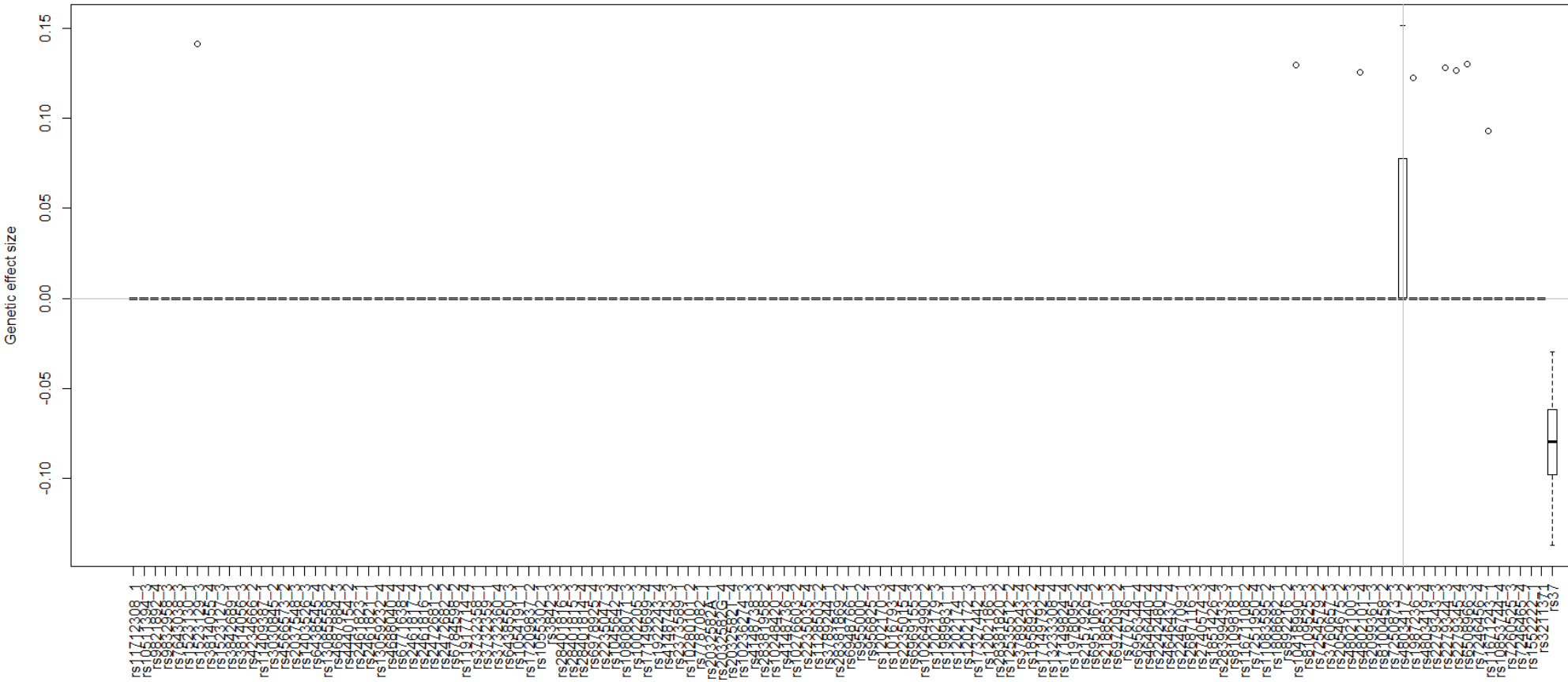
Stepwise procedure

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}

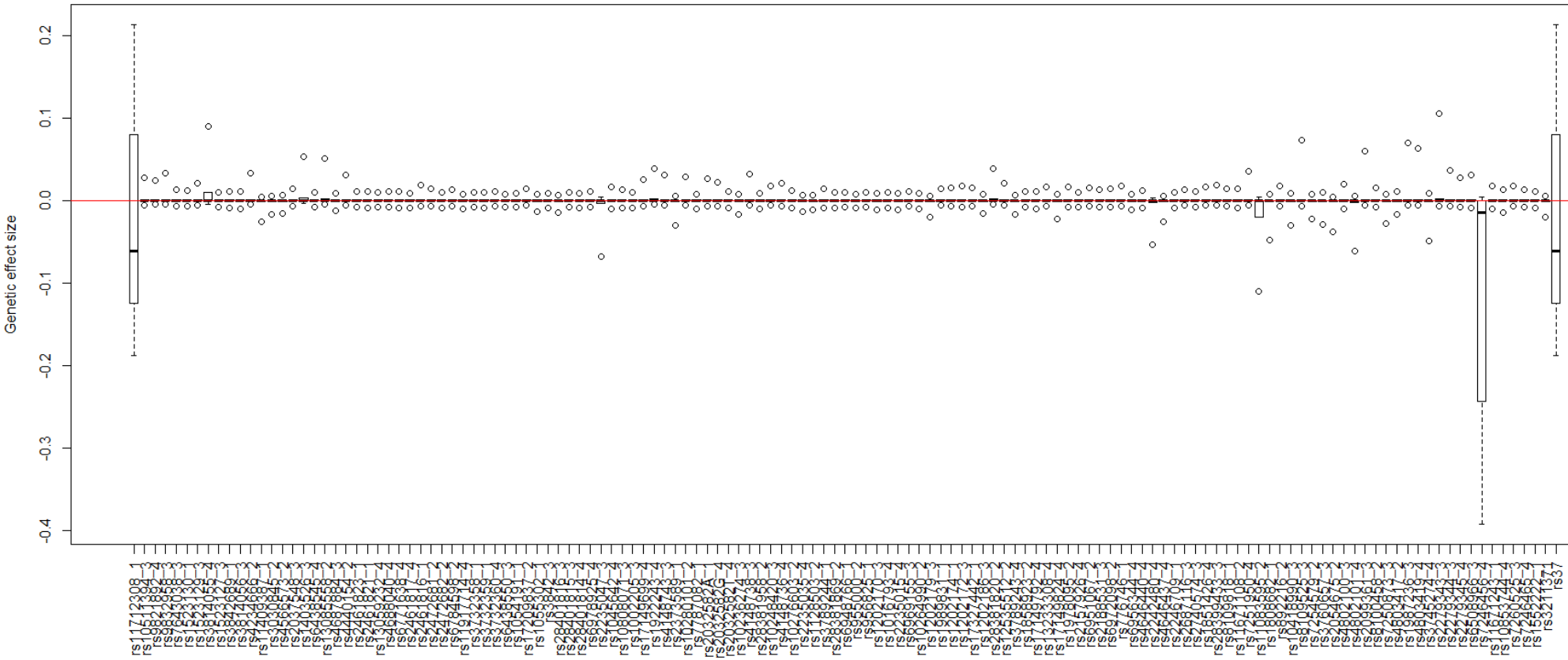
[a]Data from one individual outlier (extremely high) nevirapine clearance value were censored from this analysis.

Indicator model selection in R2jags



- n.chains=4, n.iter=5000, n.burnin=1000, n.thin=10
  $\text{Neff}_{CL/F} = 120$ / approx. 7 h

Horseshoe prior in Rstan



- n.chains=4, n.iter=1600, n.burnin=1600, n.thin=4
  Neff$_{CL/F}$ =1300 / approx. 1.5 h

Conclusions

- Similar model PK parameter estimates across methods
- No other snp associations, besides CYP2B6 G516T on CL/F, identified with Bayesian selection methods
- Rstan sampling marginally more efficient than R2jags
- Perspective
  - evaluate adaptive shrinkage in simulation study
  - 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