Méthodes du maximum de vraisemblance et alternatives Bayésiennes pour le criblage à haut débit de marqueurs génétiques en modélisation pharmacocinétique

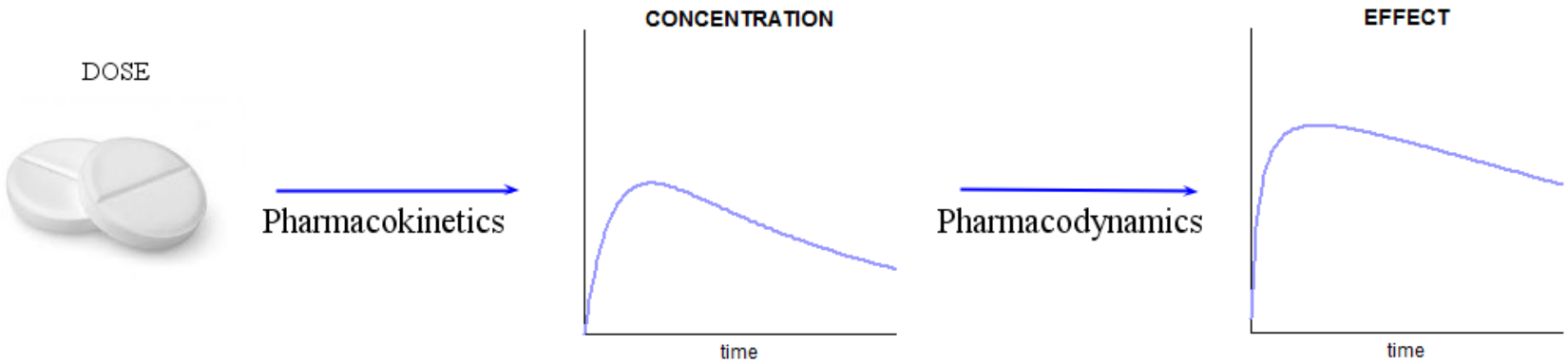
Julie Bertrand

Genetics Institute, University College London, London, UK

27 novembre 2014

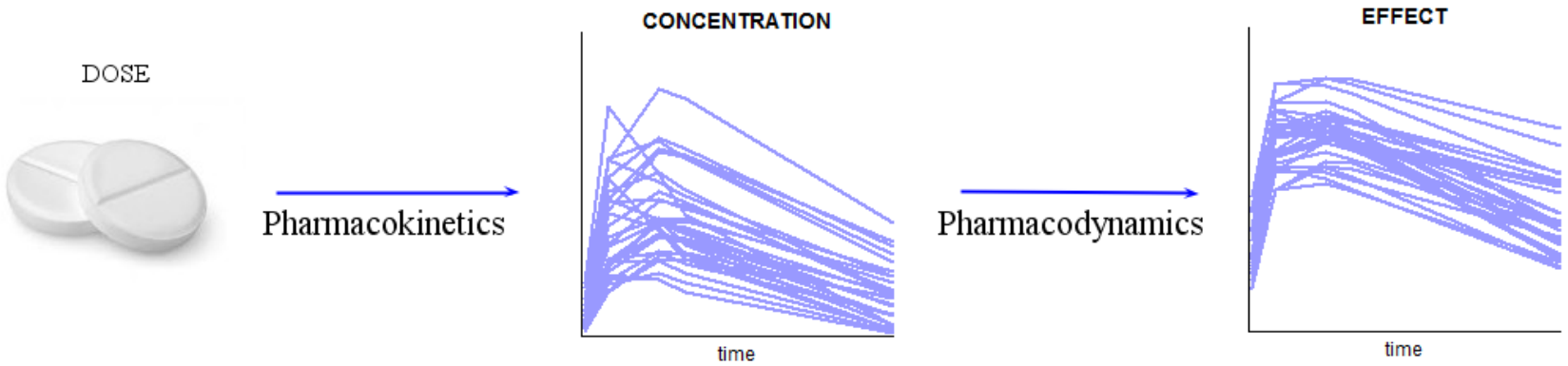
Pharmacological and genetic variability

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



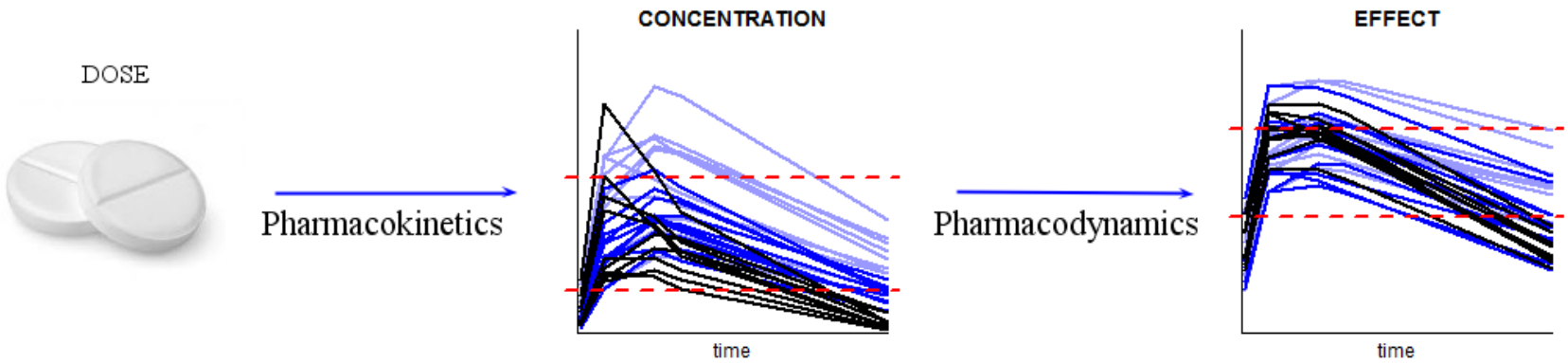
Pharmacological and genetic variability

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



Pharmacological and genetic variability

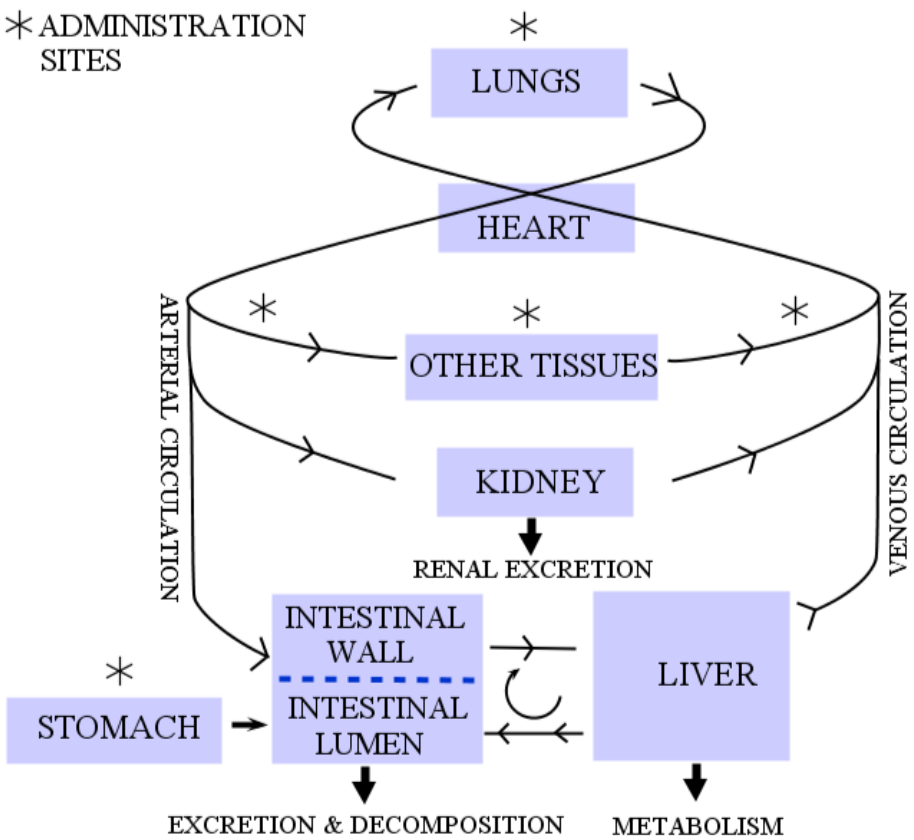
- Clinical pharmacology: study the interaction between the organism and the drug and its variability
  - pharmacokinetics (PK) and pharmacodynamics (PD)

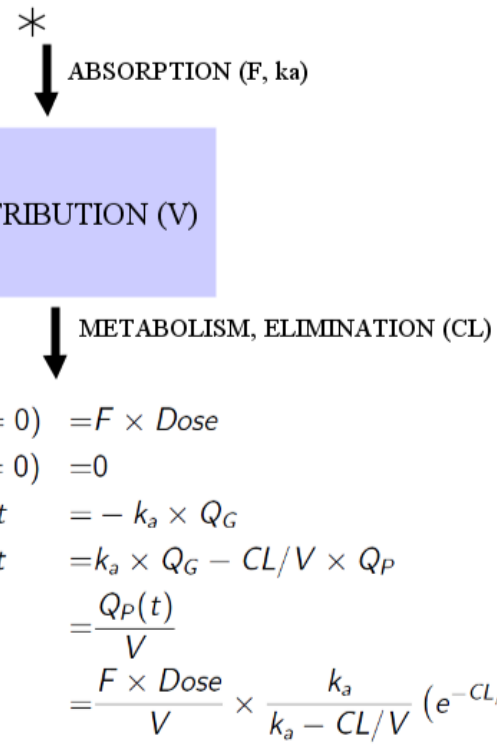


- Pharmacogenetics (PG): genetic part of the variability
  - stratified medicine
- Genes coding for proteins involved in PK/PD processes
  - metabolism enzymes (CYP450, NAT)
  - single nucleotide polymorphism (SNP)

Modelling in pharmacology

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





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

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

$\hookrightarrow$ 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

$\hookrightarrow$ 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 approaches

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 model-based PK

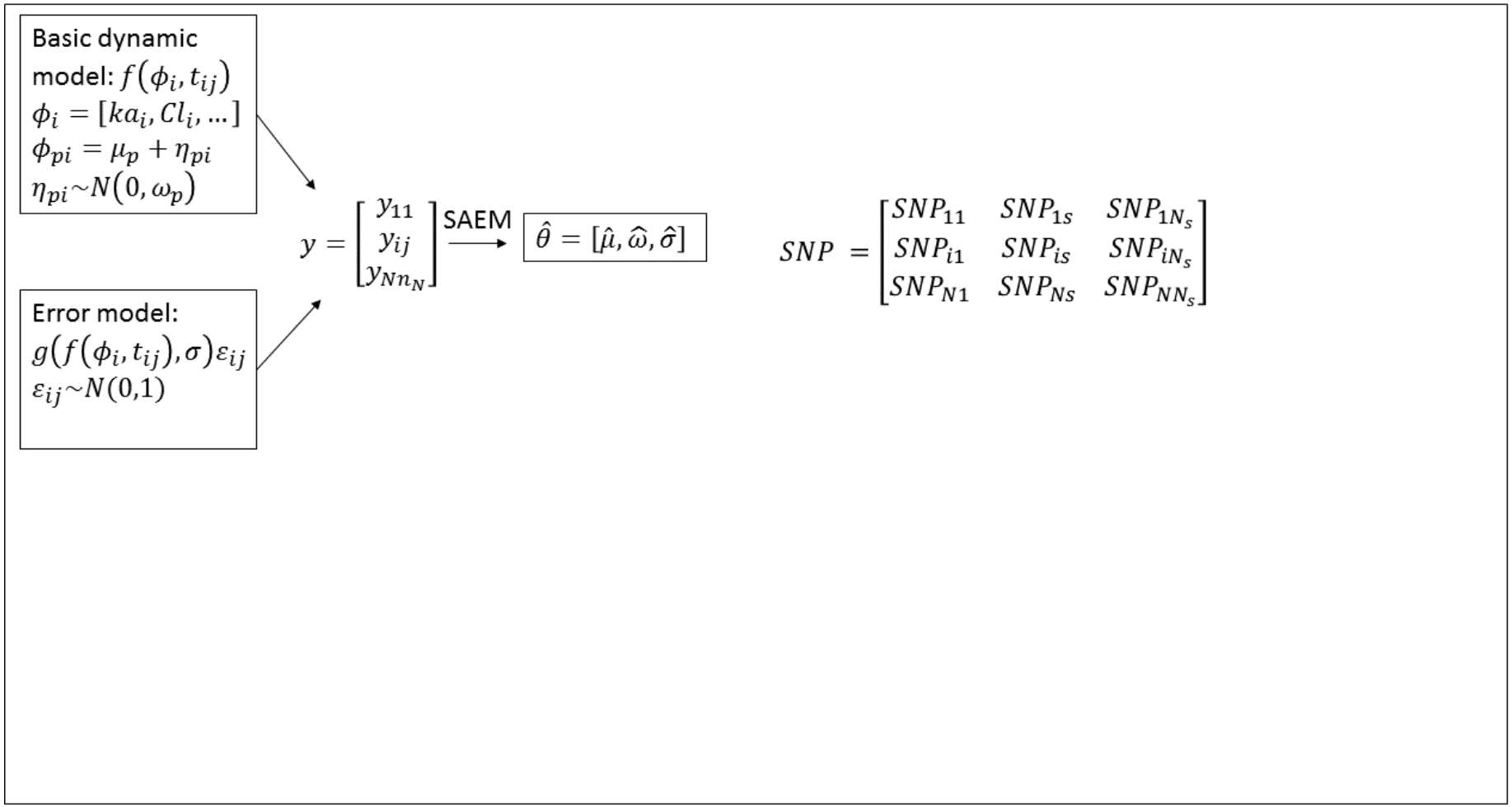
- Stepwise procedure
  - commonly used for covariate model building
  - Lehr et al. (2010) adaptation for high throughput screening
- Penalized regression-based approach
  - established in animal and plant genetics
  - Lasso (Tibshirani. 1996)
  - HLasso (Hoggart et al. 2008)
    - developed for genome-wise association studies
    - higher effect size once included in the model
- Bayesian variable selection
- KWII parsimonious interaction metric (Knights et al. 2013)

Stepwise procedure

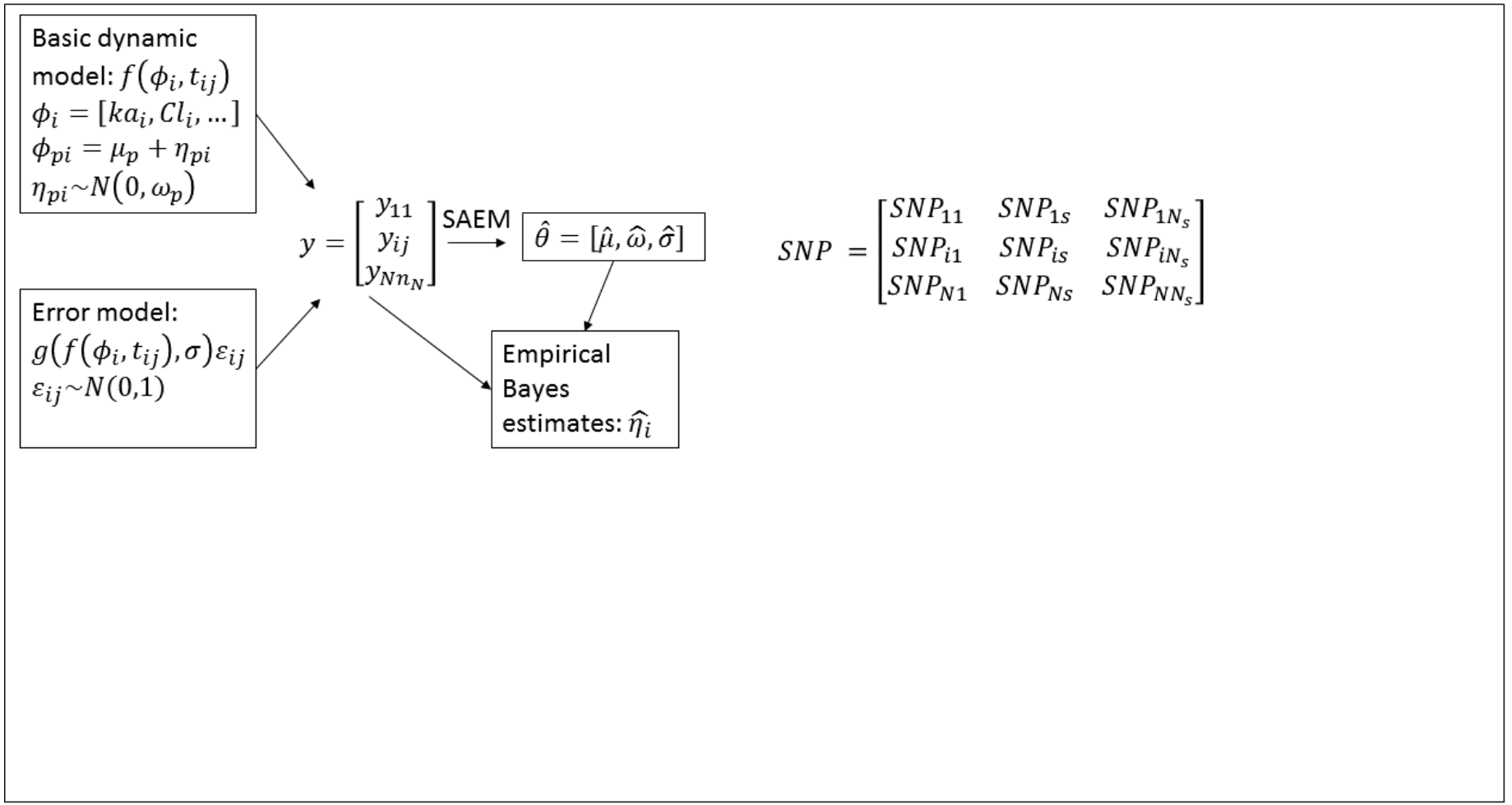
$$y = \begin{bmatrix} y_{11} \\ y_{ij} \\ y_{Nn_N} \end{bmatrix}$$

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

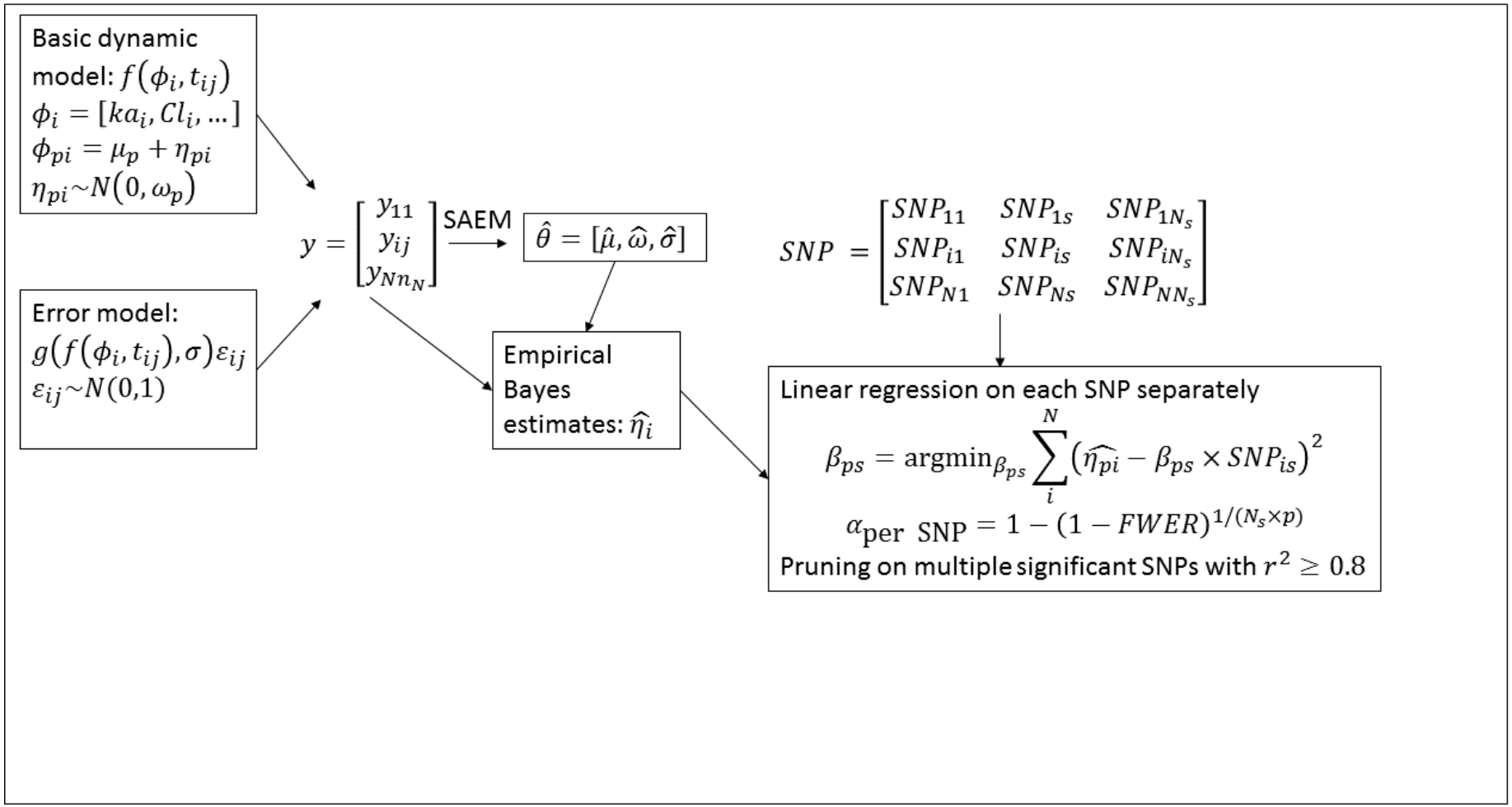
Stepwise procedure



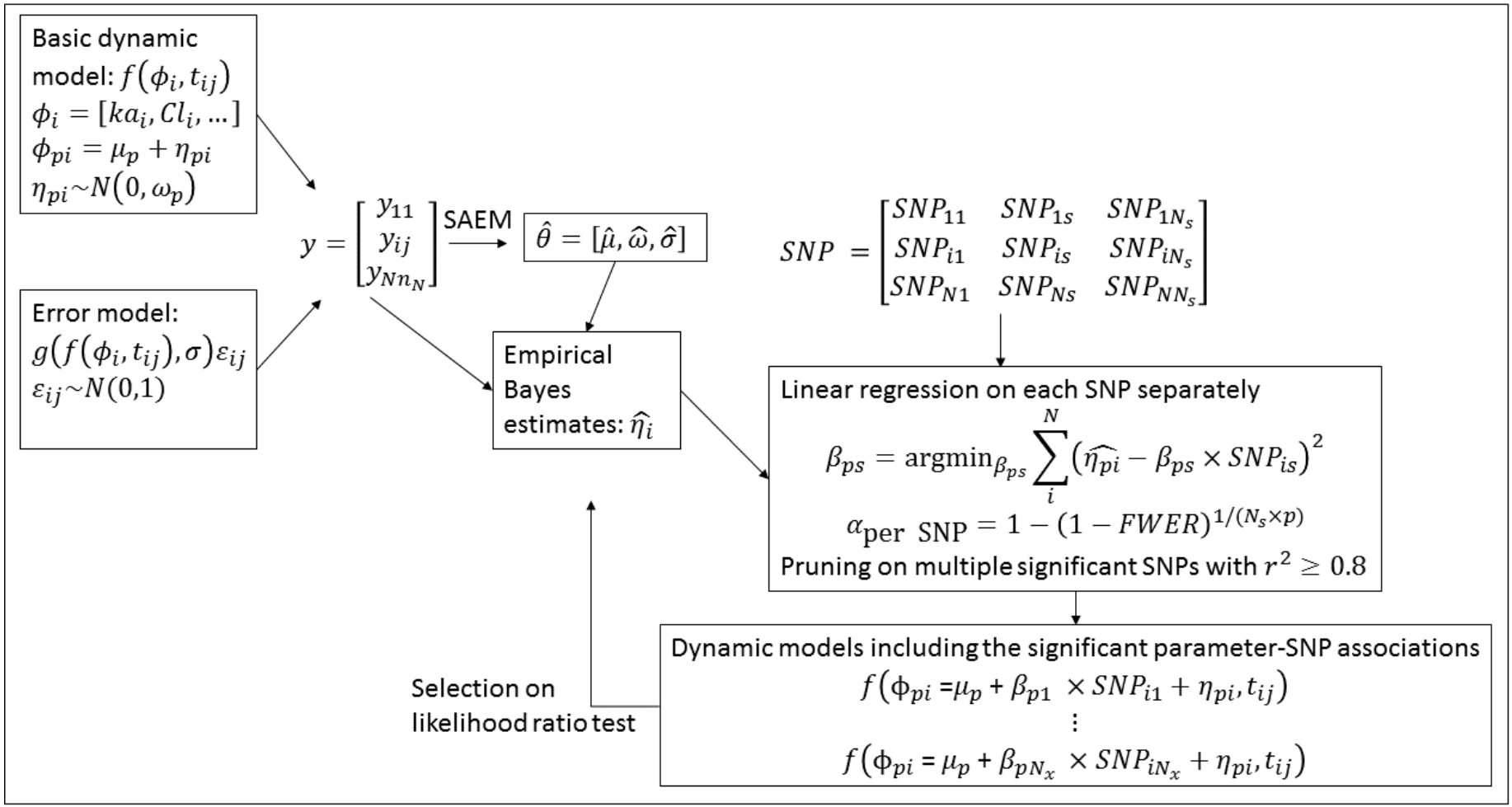
Stepwise procedure



Stepwise procedure



Stepwise procedure



↪ SNP selection after estimation of model parameters

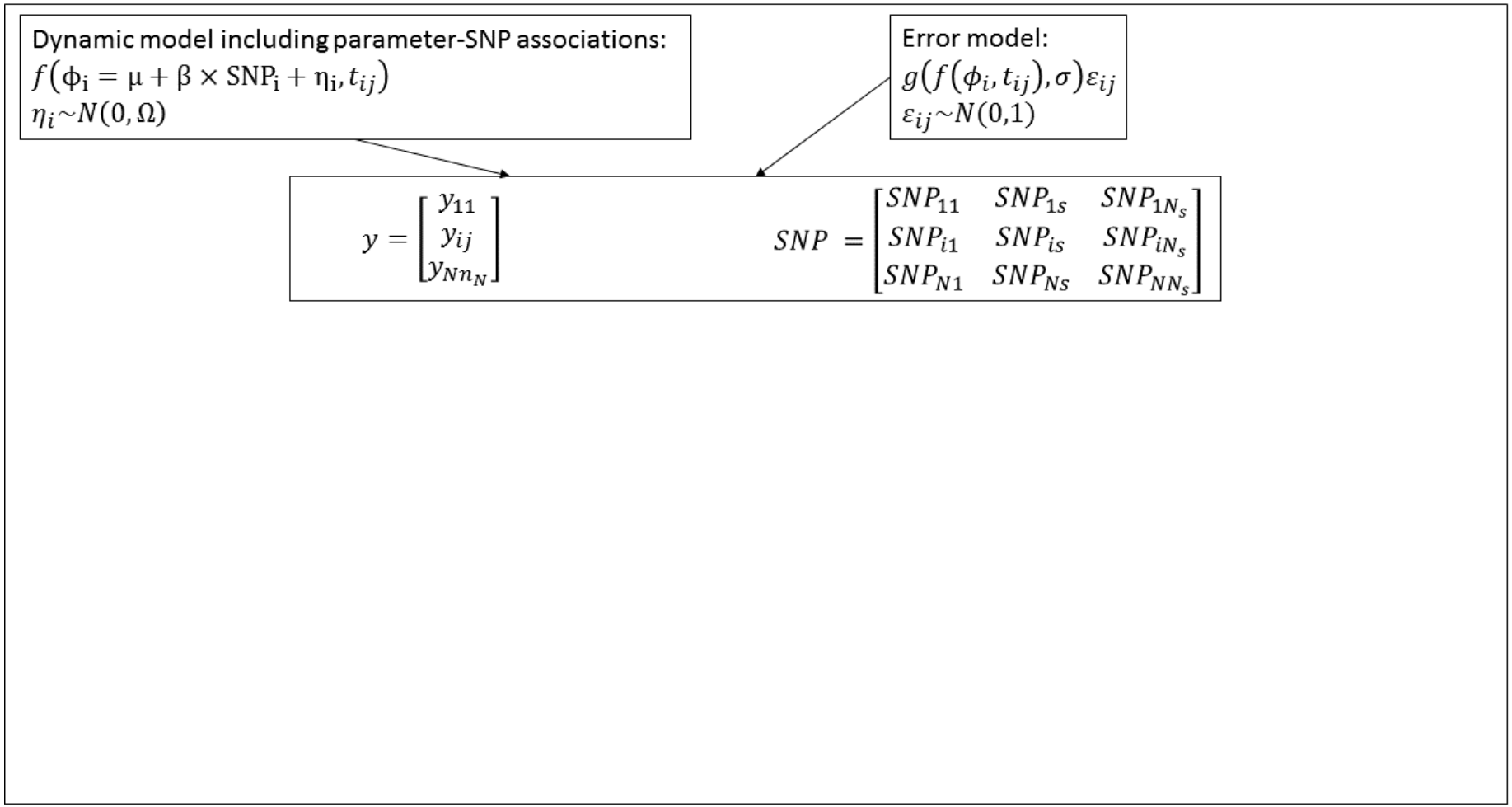
↪ SNP considered independently

Integrated appr. with penalized regression

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

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

Integrated appr. with penalized regression



Integrated appr. with penalized regression

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

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

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

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

Integrated appr. with penalized regression

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 (e.g. $\xi = \Phi^{-1}(1-\alpha/2)\sqrt{\frac{N}{\sigma^{*2}}}$)

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

↪ Simultaneous SNP selection and estimation of model parameters

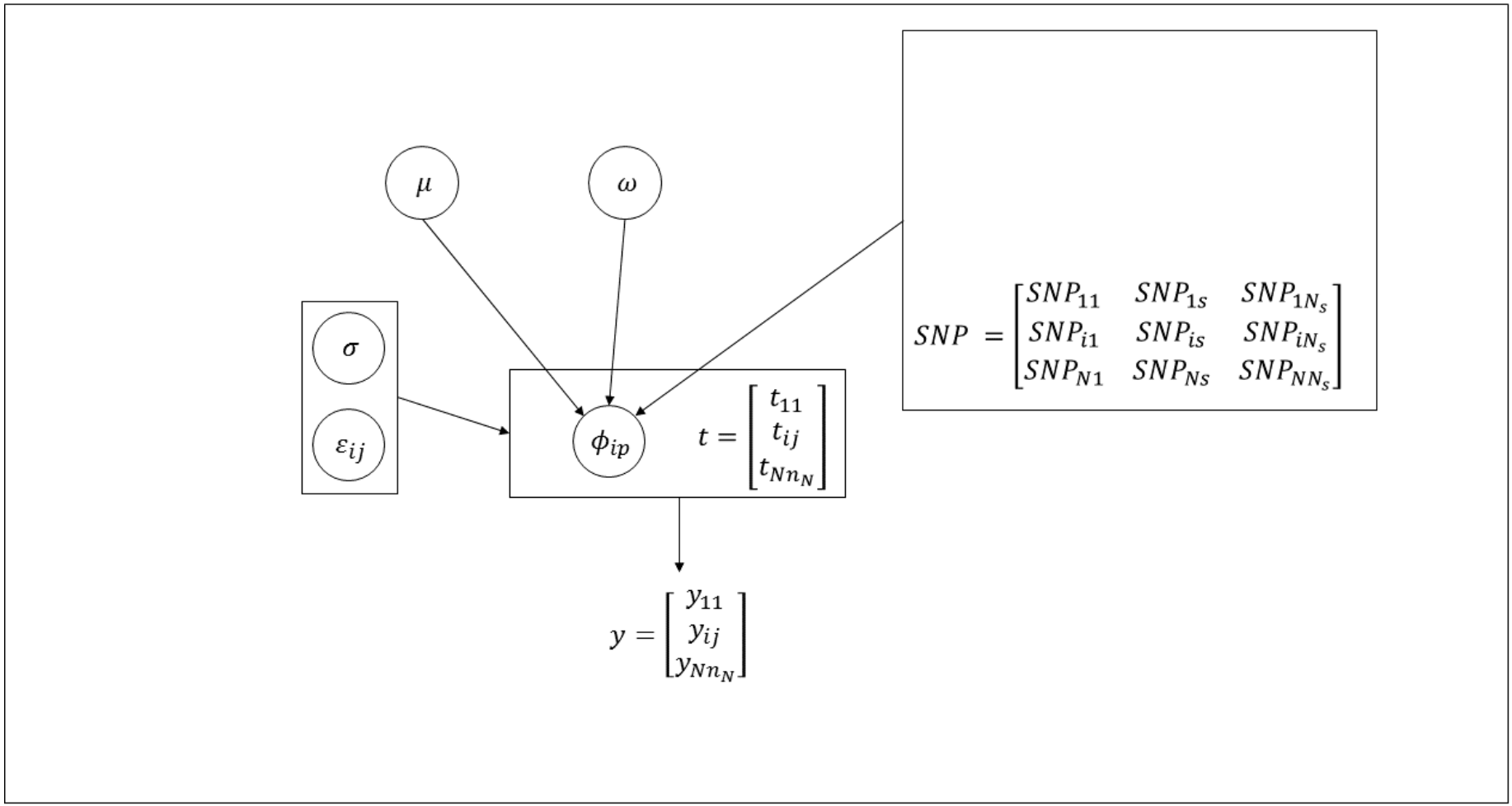
↪ All parameter-SNP associations considered simultaneously

Bayesian variable selection

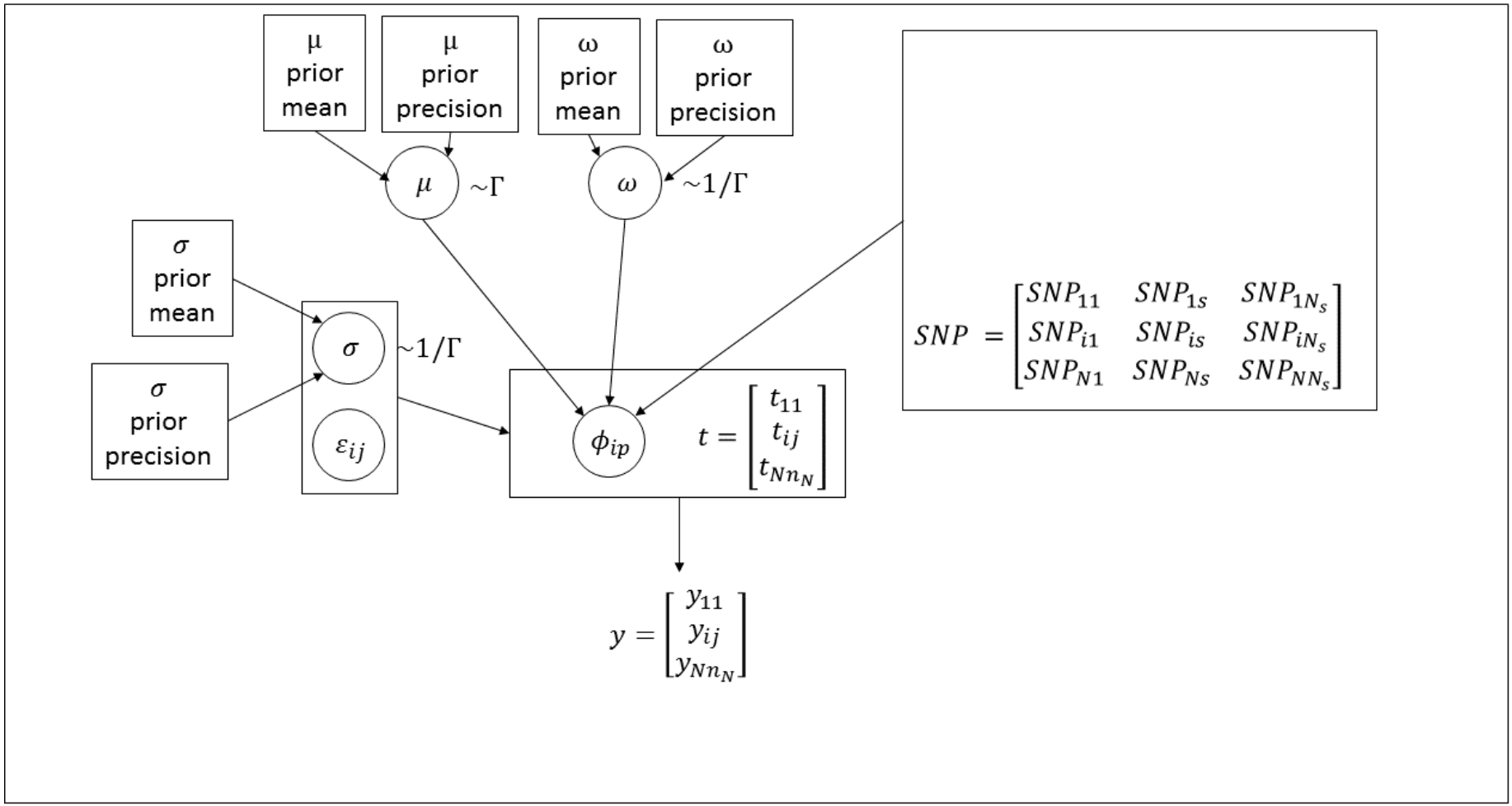
$$SNP = \begin{bmatrix} SNP_{11} & SNP_{1s} & SNP_{1N_s} \\ SNP_{i1} & SNP_{is} & SNP_{iN_s} \\ SNP_{N1} & SNP_{Ns} & SNP_{NN_s} \end{bmatrix}$$

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

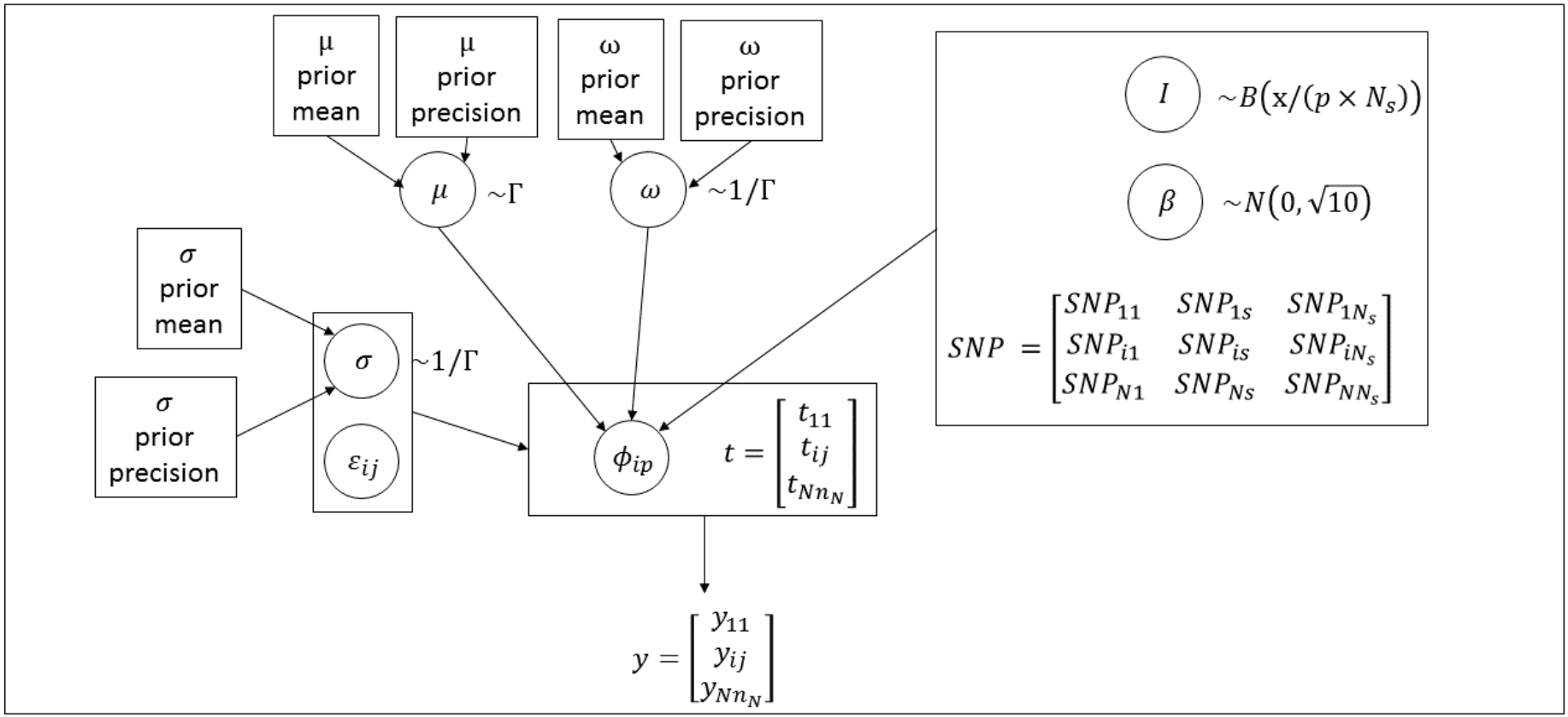
Bayesian variable selection



Bayesian variable selection



Bayesian variable selection



↪ Simultaneous SNP selection and estimation of model parameters

↪ All parameter-SNP associations considered simultaneously

Implementation

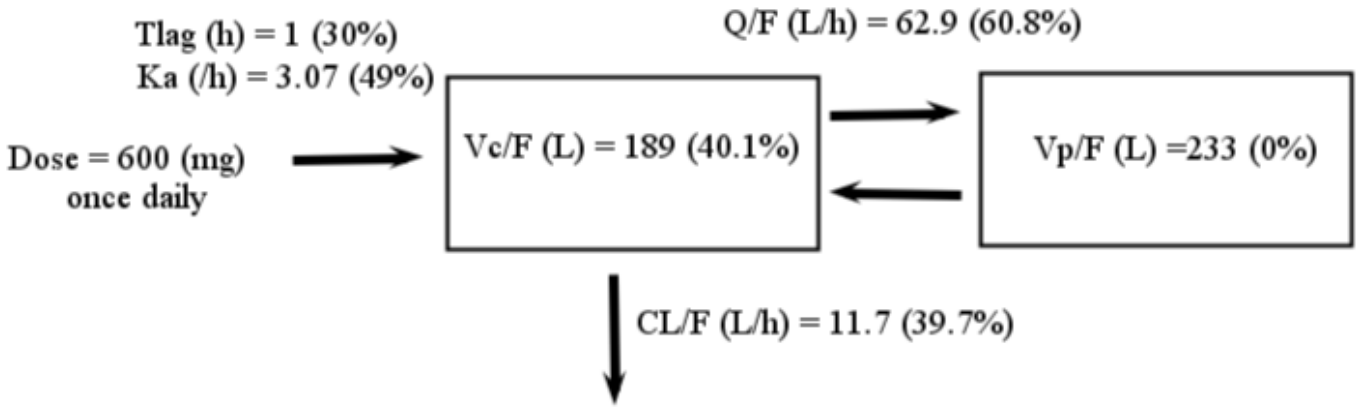
- Stepwise procedure
  - R and `saemix` R package
- Integrated appr. with penalized regression
  - extension of the `saemix` R package
- Bayesian variable selection
  - `R2jags`

Objectives

- To evaluate and compare through a realistic simulation study (200 data sets under H_0 and H_1)
  - stepwise procedure
  - integrated approach using
    - Lasso
    - HLasso
  - Bayesian variable selection

Simulation settings - 1/2

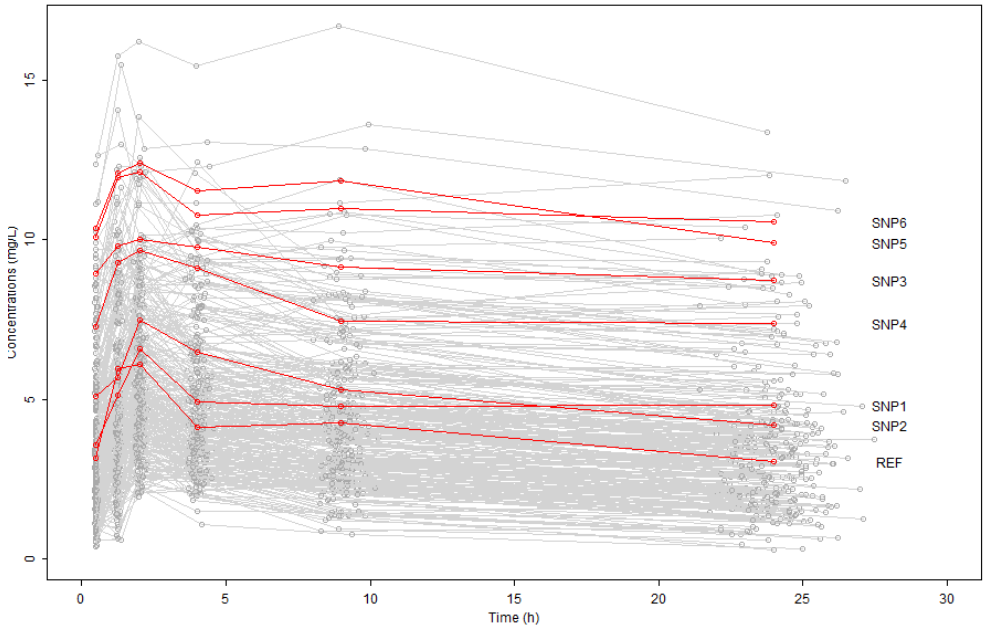
- Generation of genotypes using HAPGEN
  - N_s=1227 snps on 171 genes from the DMET Chip
  - 6 [1-56] snps per gene
  - HAPMAP caucasian reference haplotypes
- Pharmacokinetic profiles inspired from real study



  - diagonal variance matrix of random effects
  - combined residual error model $g = \sigma_{inter} + \sigma_{prop} f(\phi_i, t_{ij})$

↪ Genetic association explored on V_c, Cl and Q

Simulation settings - 2/2



- Phase II study
  - N=300/t=0.5, 1.25, 2, 4, 9, 24h
- Six unobserved and uncorrelated causal variants
  - decreasing *Cl*
  - each explaining 1, 2, 3, 5, 7 and 12% of the inter-individual variability

FWER and True/False positives (T/FP)

Method	FWER	TP	FP_{CL}	FP_V	FP_Q
Step. Proc.	19.0	366	15	7	5
Integr. appr. with Lasso	20.0	340	7	6	0
Integr. appr. with HLasso	22.5	335	4	7	1
Bayesian Variable Selection*	-	176	35	0	3

Family wise error rate (FWER), expected value of 20[14.5–25.5]%
TP = count of SNP in $r^2 \geq 0.05$ with a causal with a maximum of 1200
*188 data sets

FWER and True/False positives (T/FP)

Method	FWER	TP	FP_{CL}	FP_V	FP_Q
Step. Proc.	19.0	366	15	7	5
Integr. appr. with Lasso	20.0	340	7	6	0
Integr. appr. with HLasso	22.5	335	4	7	1
Bayesian Variable Selection*	-	176	35	0	3

Family wise error rate (FWER), expected value of 20[14.5–25.5]%
TP = count of SNP in $r^2 \geq 0.05$ with a causal with a maximum of 1200
*188 data sets

FWER and True/False positives (T/FP)

Method	FWER	TP	FP_{CL}	FP_V	FP_Q
Step. Proc.	19.0	366	15	7	5
Integr. appr. with Lasso	20.0	340	7	6	0
Integr. appr. with HLasso	22.5	335	4	7	1
Bayesian Variable Selection*	-	176	35	0	3

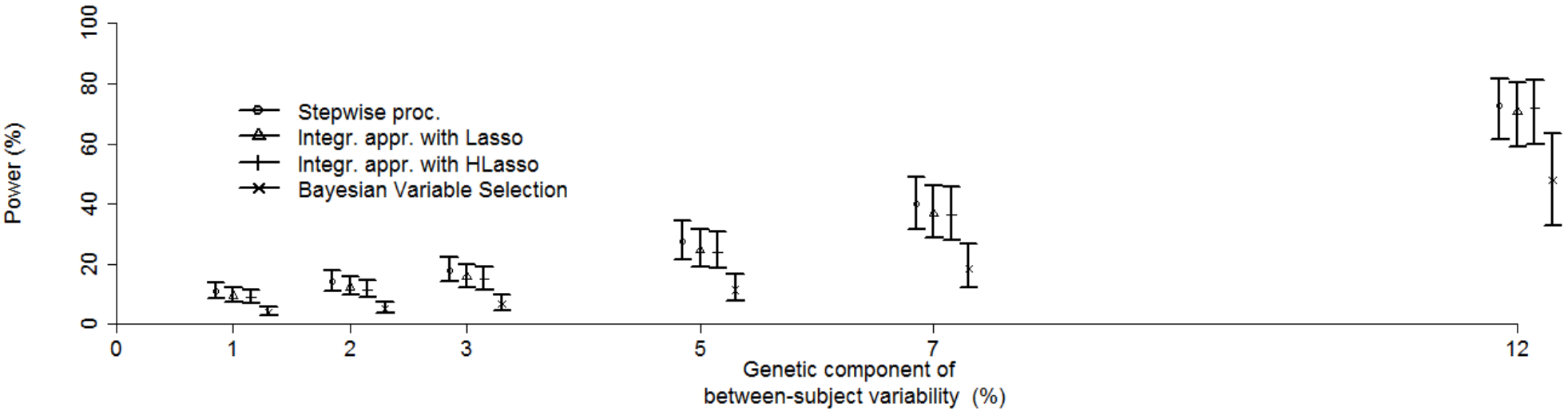
Family wise error rate (FWER), expected value of 20[14.5–25.5]%
TP = count of SNP in $r^2 \geq 0.05$ with a causal with a maximum of 1200
*188 data sets

FWER and True/False positives (T/FP)

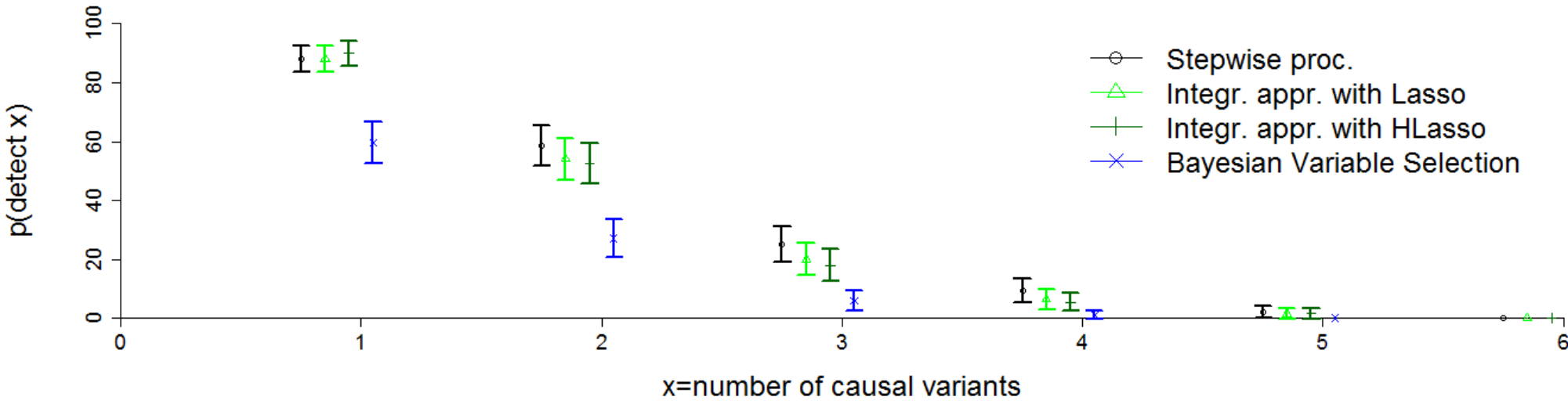
Method	FWER	TP	FP_{CL}	FP_V	FP_Q
Step. Proc.	19.0	366	15	7	5
Integr. appr. with Lasso	20.0	340	7	6	0
Integr. appr. with HLasso	22.5	335	4	7	1
Bayesian Variable Selection*	-	176	35	0	3

Family wise error rate (FWER), expected value of 20[14.5–25.5]%
TP = count of SNP in $r^2 \geq 0.05$ with a causal with a maximum of 1200
*188 data sets

Power


Power (%)
0
20
40
60
80
100
Stepwise proc.
Integr. appr. with Lasso
Integr. appr. with HLasso
Bayesian Variable Selection
0
1
2
3
5
7
12
Genetic component of
between-subject variability (%)

Power to detect multiple SNPs



Computing times (hours)

Method	H_0	H_1
Step. Proc.	0.05 [0.05-0.24]	0.24 [0.06-1.09]
Integr. appr. with Lasso	1.04 [0.81-1.48]	1.14 [0.83-1.61]
Integr. appr. with HLasso	1.08 [0.81-1.57]	1.19 [0.84-1.60]
bayesian Variable Selection*	-	32 [23-66]

median [range]
*188 data sets

Computing times (hours)

Method	H_0	H_1
Step. Proc.	0.05 [0.05-0.24]	0.24 [0.06-1.09]
Integr. appr. with Lasso	1.04 [0.81-1.48]	1.14 [0.83-1.61]
Integr. appr. with HLasso	1.08 [0.81-1.57]	1.19 [0.84-1.60]
bayesian Variable Selection*	-	32 [23-66]

median [range]
*188 data sets

Computing times (hours)

Method	H_0	H_1
Step. Proc.	0.05 [0.05-0.24]	0.24 [0.06-1.09]
Integr. appr. with Lasso	1.04 [0.81-1.48]	1.14 [0.83-1.61]
Integr. appr. with HLasso	1.08 [0.81-1.57]	1.19 [0.84-1.60]
bayesian Variable Selection*	-	32 [23-66]

median [range]
*188 data sets

Maximum likelihood approaches

- Feasibility of model-based PGx analyses on real-case scenarios
- Real need of increased sample size compared to classical drug development study designs
- Integrated approach
  - much less FP for a slightly inferior power
  - cost in computing time non-negligible
  - better capture multiple parameter-SNP associations (data not shown)
- Current implementation in the `saemix` R package limited
  - no inter-occasion variability
  - remove SNP with missing individuals

Bayesian Variable Selection

- Early results and need to improve on JAGs settings
  - prior distribution / initial condition on PK parameters
  - thinning, stepsize
- Sensibility analyses
  - standardize the genotypes
  - not estimating random effect variances on some PK parameters
- Alternative approaches
  - slab and spike selection
  - prior on the effect size and inclusion probability

Acknowledgements

Prof. David J Balding, UCL Genetics Institute, UK

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

Dr Emmanuelle Comets, INSERM, France

All my colleagues at the UCL Genetics Institute