

# Contribution of nonlinear mixed effects models and penalised regression approaches in pharmacogenetic population studies

Tessier A<sup>1,2</sup>, Bertrand J<sup>3</sup>, Chenel M<sup>2</sup>, Comets E<sup>1,4</sup>

1. INSERM, IAME, UMR 1137, F-75018 Paris, France; Université Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France  
2. Clinical Pharmacokinetics and Pharmacometrics, Institut de Recherches Internationales Servier, Suresnes, France  
3. University College London, Genetics Institute, London, UK  
4. INSERM CIC 1414, Université Rennes 1, Rennes, France

email: adrien.tessier@inserm.fr

## CONTEXT

**Pharmacogenetics:** study of the genetic variation in relation to interindividual variability in drug response

**Drug response:** pharmacokinetics (PK)

- No consensus on approaches to obtain the phenotype in population PK:
  - two-stage approaches (noncompartmental analyses, NCA) vs nonlinear mixed effects models (NLMEM)

**Genetic variation:** single nucleotide polymorphisms (SNPs)

- No consensus on methods adapted to handle hurdles of pharmacogenetic data:
  - number of polymorphisms >> number of subjects
  - local correlation between markers (linkage disequilibrium)

## OBJECTIVES

Investigate, through a simulation study based on a motivating example:

- four association tests (a stepwise procedure, three penalised regression methods)
- applied to four phenotypes (2 observed, estimated by NCA, model-based approach (NLMEM))

## METHODS

**Genetic association model**

Additive effect of rare alleles:

$$Phenotype_i = \beta_0 + \sum_{k=1}^P \beta_k \cdot SNP_{ik} + \varepsilon_i; SNP_{ik} = \{0, 1, 2\}$$

**Phenotypes**

Observed concentration at 24h (C24h) and 192h (C192h)

AUC estimated by noncompartmental analysis

CL estimated by nonlinear mixed effects model

**Association methods**

**Stepwise procedure:** multi-step iterative process based on the Wald test (fig. 1) [1]

**Penalised regression:** one-step method to estimate model coefficients using different *a priori* distributions

- Ridge regression [2]: Wald test used to select significant coefficients
- Lasso [3] and HyperLasso[4]: most coefficients driven to 0
  - SNPs with coefficients different from 0 are selected

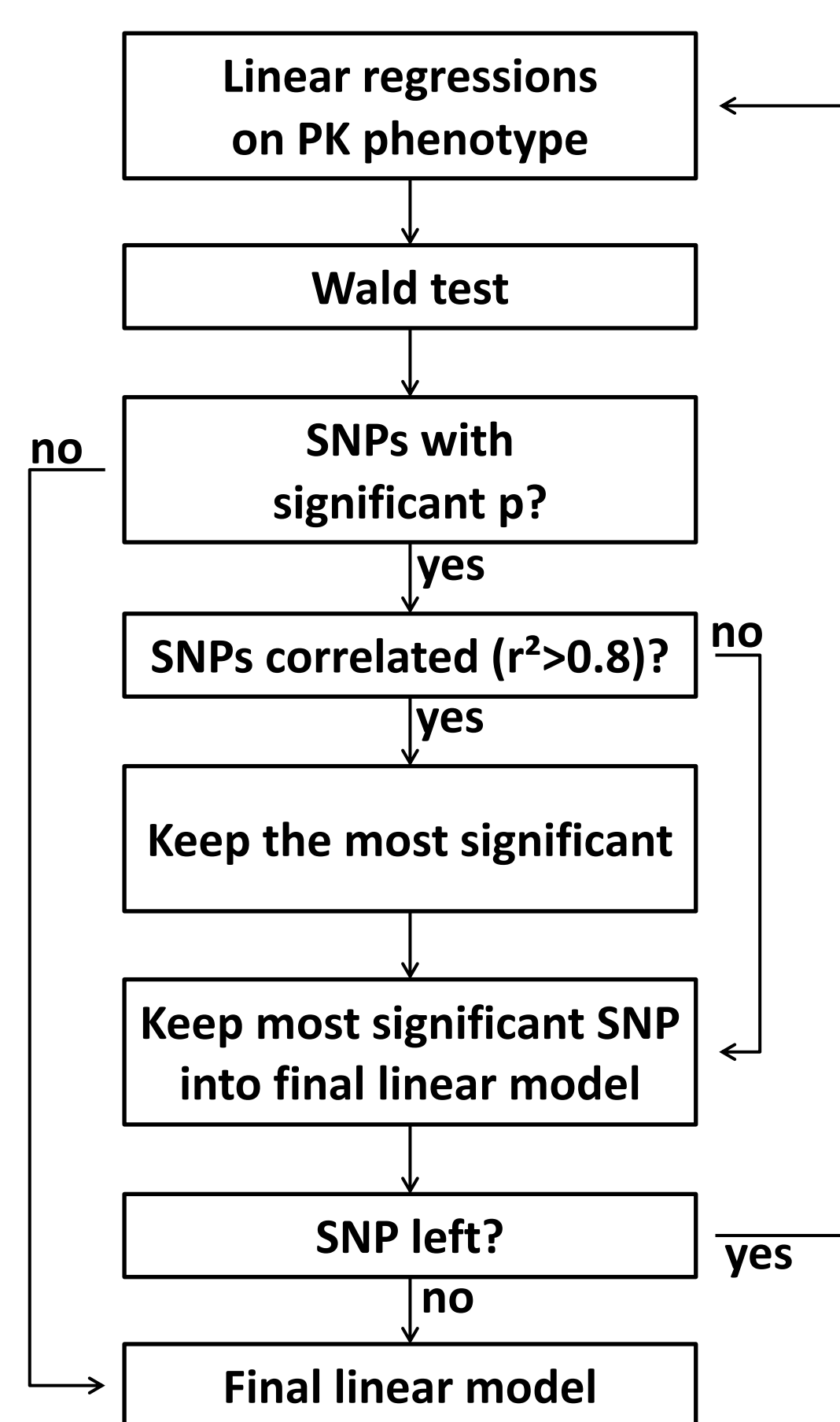


Figure1. Stepwise procedure algorithm

**Simulation study**

**Motivating example**

Inspired by real data sets from phase I clinical studies [5]

- limited number of subjects (N = 78) frequently sampled (n = 16)
- extensive genetic data: 176 SNPs from genes involved in PK of drugs
- complex PK profiles
  - nonlinear absorption (fig. 2)

**PK settings**

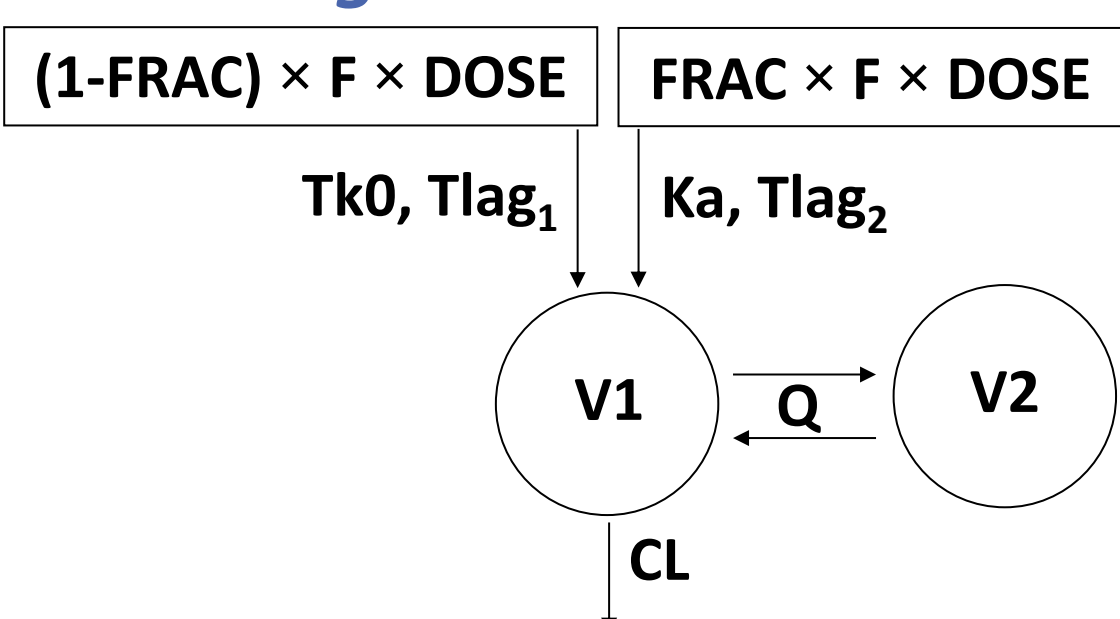


Figure 2. Structural PK model for simulations

Phenotypes are obtained (concentrations) or estimated (AUC, CL) from simulated PK profiles

- observed concentrations and AUC are normalised by dose for PK nonlinearity
- PK phenotypes are log transformed

**Genotypes simulations**

176 SNPs simulated from a Hapmap data set to retain correlations [6]

6 SNPs drawn randomly and set to affect the drug elimination:

$$\log(CL_i) = \log(\mu_{CL}) + \sum_{k=1}^6 \beta_k \times SNP_{ik} + \eta_{i_{CL}}; SNP_{ik} = \{0, 1, 2\}$$

Associated  $\beta_k$  is computed according to the proportion of CL variability explained by the 6 SNPs

| SNP <sub>1</sub> | SNP <sub>2</sub> | SNP <sub>3</sub> | SNP <sub>4</sub> | SNP <sub>5</sub> | SNP <sub>6</sub> |
|------------------|------------------|------------------|------------------|------------------|------------------|
| 1 %              | 2 %              | 3 %              | 5 %              | 7 %              | 12 %             |

The 6 SNPs explain overall 30% of CL variability

**Scenarios**

200 data sets simulated per scenario

**S<sub>real</sub>**: N = 78 subjects, n = 16 samples

**S<sub>large</sub>**: N = 384 subjects, n = 16 samples

- influence of the structural PK model:
  - S<sub>large, no IIV<sub>F</sub></sub>**: no variability on F
  - S<sub>linearPK</sub>**: linear absorption, FRAC = 0.5, F = 0.8, IIV(F) = 32.9%

## RESULTS

**Example of simulated data set**

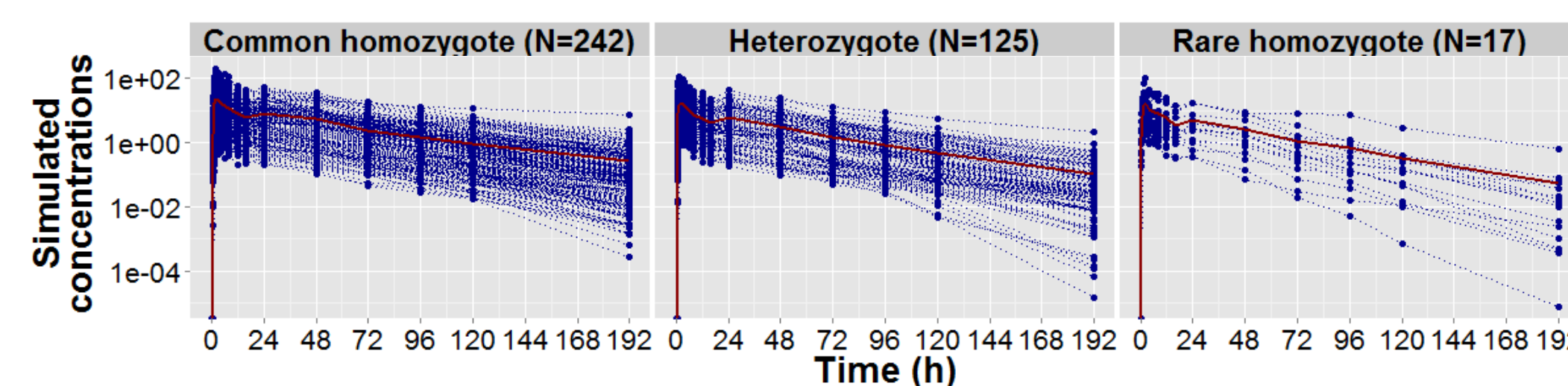


Figure 3. Individual and mean concentration versus time profiles per genotype simulated for one data set

Lower trough concentrations in heterozygotes and rare homozygotes (fig. 3)

**Phenotypes and association methods comparison**

True positive (TP): a SNP which was indeed associated to CL in the simulation

False positive (FP): a SNP selected in the model but not present in the simulation

$$\text{True Positive Rate: } TPR = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$\text{False Positive Rate: } FPR = \frac{\text{number of false positives}}{\text{number of false positives} + \text{number of true negatives}}$$

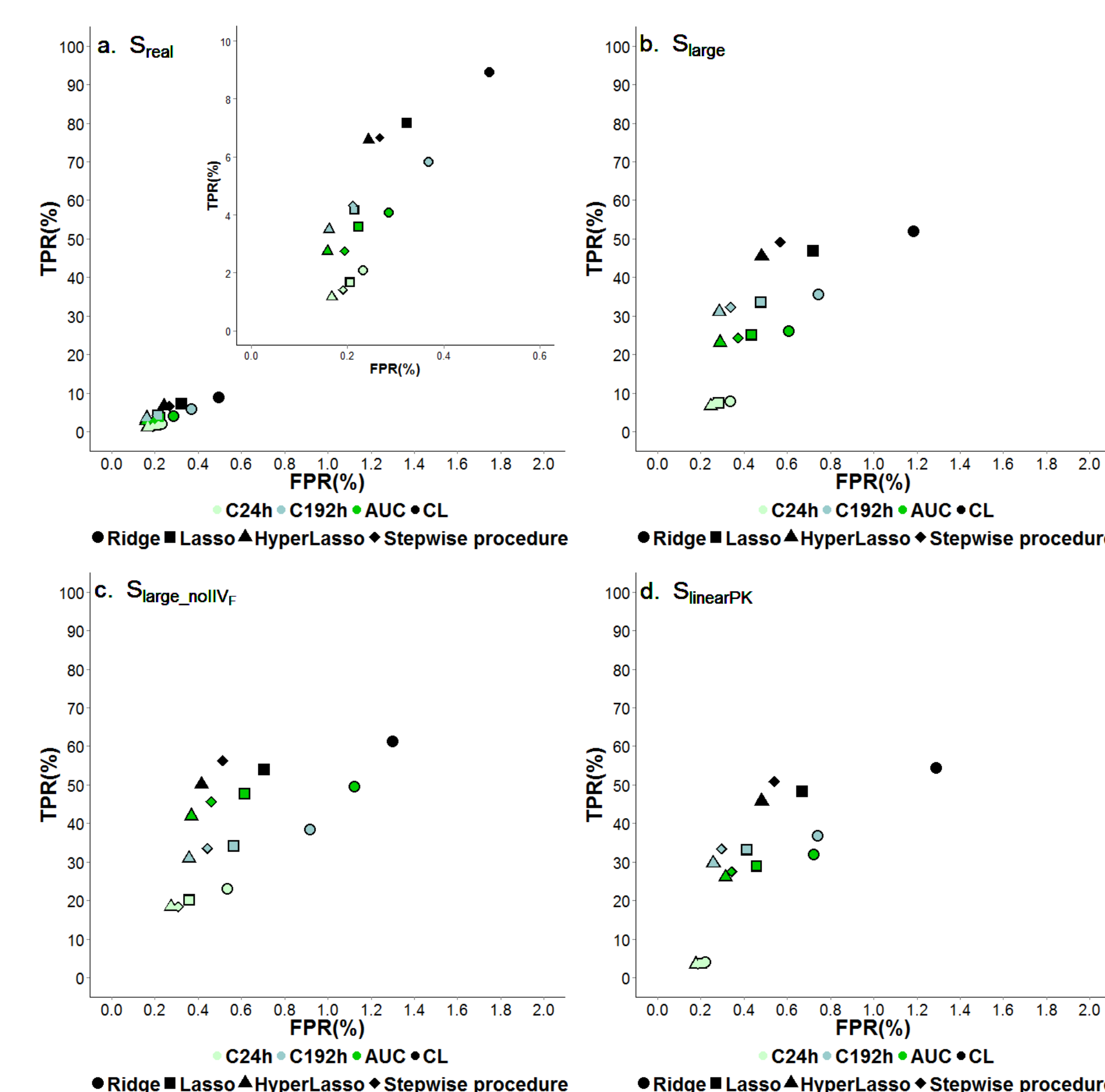


Figure 4. Percentage of True Positive Rate (TPR) versus False Positive Rate

Similar FPR between phenotypes in all scenarios (fig. 4)

Highest TPR for CL in all scenarios

- despite the fact that AUC is well correlated with CL in rich designs
- lower TPR for AUC due to both the nonlinearity in the absorption parameter F and the interindividual variability in F
  - additional simulation with linear absorption and no interindividual variability on F (data not shown) yields similar TPR and FPR for CL and AUC

TPR always lower for observed concentrations

Ridge regression: highest TPR and FPR in all scenarios

**Impact of number of subjects on probability of detection on CL**

Probabilities to detect the genetic effect: p(detect at least x variants) = % of data sets where at least x of the 6 SNPs associated with CL are selected [x=1,...,6]

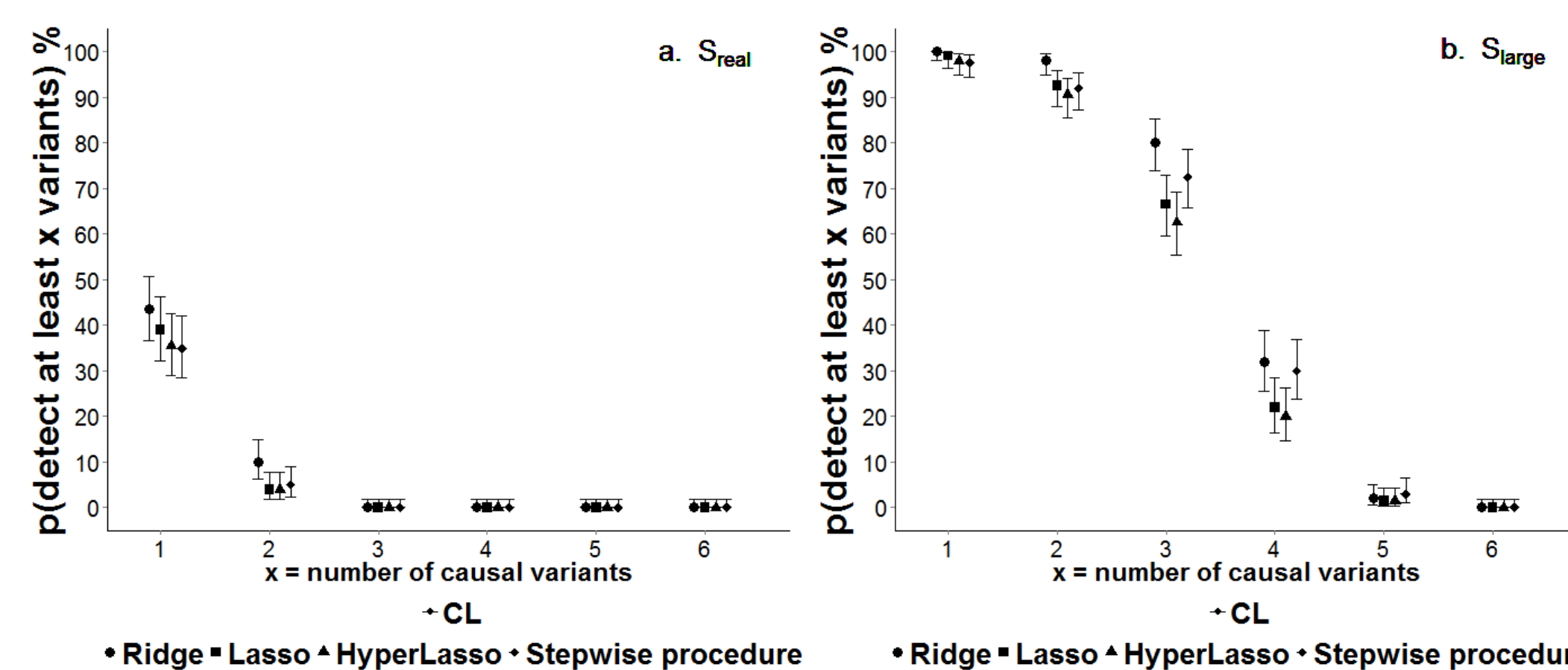


Figure 5. Probability Estimates and 95% confidence interval with each method to detect at least x variants explaining interindividual variability of CL

Similar probability of detection across methods in all scenarios (fig. 5)

**S<sub>real</sub>**: low probability to detect at least one or two genetic variant on CL. Probability reaches 0 for 3 variants or more

**S<sub>large</sub>**: probability of all methods reaches almost 100% to detect at least one genetic variant. The 6 SNPs are never detected

## DISCUSSION

Recent methodological work comparing the association methods applied on a single PK phenotype (model-based CL) by a simulation study [7].

With nonlinear PK, the power to detect an effect of genetic covariates was much higher when using the model-based phenotype CL than when using AUC or observations.

The probability of detection was similar in the different association methods.

However, all approaches needed a large number of subjects (approx. 400) to detect a clinically relevant effect, especially with infrequent variants.