

# Combined analysis of phase I and phase II data to enhance the power of pharmacogenetic tests

Adrien Tessier<sup>1,2,3</sup>, Julie Bertrand<sup>4</sup>, Marylore Chenel<sup>3</sup> and Emmanuelle Comets<sup>1,2,5</sup>

1 INSERM, IAME, UMR 1137, F-75018 Paris, France

2 University Paris Diderot, IAME, UMR 1137, Sorbonne Paris Cité, F-75018 Paris, France

3 Division of Clinical Pharmacokinetics and Pharmacometrics, Institut de Recherches Internationales Servier, Suresnes, France

4 University College London, Genetics Institute, London, UK

5 INSERM CIC 1414, Université Rennes 1, Rennes, France

## Running head: Combined Analyses In Pharmacogenetic Studies

### ABSTRACT

We show through a simulation study how the joint analysis of data from phase I and phase II studies enhances the power of pharmacogenetic tests in pharmacokinetic (PK) studies. PK profiles were simulated under different designs along with 176 genetic markers. The null scenarios assumed no genetic effect, while under the alternative scenarios, drug clearance was associated to 6 genetic markers randomly sampled in each simulated dataset. We compared penalised regression Lasso and stepwise procedures to detect the associations between empirical Bayes estimates of clearance, estimated by nonlinear mixed effects models, and genetic variants. Combining data from phase I and phase II studies, even sparse, increases the power to identify the associations between genetics and PK due to the larger sample size. Design optimisation brings a further improvement, and we highlight a direct relationship between  $\eta$ -shrinkage and loss of genetic signal.

### KEYWORDS

Pharmacogenetics, Pharmacokinetics, Nonlinear mixed effects models, Design, Clinical drug development.

## 1 ABBREVIATIONS

2	EBE	Empirical Bayes estimates
3	FWER	Family wise error rate
4	FIM	Fisher information matrix
5	GAM	Generalised additive model
6	$H_0$	Null scenarios
7	$H_1$	Alternative scenarios
8	IRIS	Institut de Recherches Servier
9	IIV	Interindividual variability
10	LD	Linkage disequilibrium
11	LRT	Likelihood ratio test
12	NCA	Noncompartmental analyses
13	NLMEM	Nonlinear mixed effects models
14	PK	Pharmacokinetics
15	RD	Relative deviation
16	REE	Relative estimation error
17	$R_{GC}$	Genetic component of the interindividual variability
18	SE	Standard error
19	Shr	$\eta$ -shrinkage
20	SNPs	Single nucleotide polymorphisms
21	TP	True positive
22	TPR	True positives rate
23	FP	False positive
24	FPR	False positives rate
25	$\alpha$	Type I error per test
26	$\beta$	Effect size coefficient
27	$\xi$	Tuning parameter for Lasso method
28	$\rho$	Correlation coefficient for stepwise procedure
29	$N_t$	Number of tests
30	$p_k$	Frequency of the minor allele

## 31 INTRODUCTION

32 Studying the sources of the variability observed in drug response facilitates individualisation  
33 of prescription. One of the sources of variability in drugs' pharmacokinetics (PK)<sup>1</sup> is the  
34 variation in activity of enzymes and transporters involved in the drug absorption,  
35 distribution, metabolism or elimination. Pharmacogenetics<sup>2</sup> studies the genetic component  
36 of interindividual variability (IIV) observed in PK to identify populations at risk of treatment  
37 inefficacy or adverse effects<sup>3</sup>. Single nucleotide polymorphisms (SNPs) are the genetic  
38 variants most frequently studied in pharmacogenetics and screened more and more often in  
39 clinical studies.

40 Genetic data offer some unique challenges, in particular because they may lead to very  
41 unbalanced number of subjects, which impacts the power of tests in pharmacogenetic  
42 analyses<sup>4,5</sup>. In a previous simulation work, we showed that typical phase I studies have low  
43 power to detect genetic effects because of the limited sample size<sup>6</sup>. On the other hand,  
44 phase I studies generally provide good quality PK information allowing to characterise the PK  
45 profile of the drug. We showed that from the different approaches used at this stage to  
46 estimate PK parameters, nonlinear mixed effects models (NLMEM)<sup>7</sup> could be considerably  
47 more powerful than noncompartmental analyses (NCA)<sup>8</sup> for complex PK models<sup>6</sup>. Our  
48 simulations also showed that increasing the sample size, as in phase II studies, would  
49 improve the power to detect genetic variants. However sparse designs typically used in  
50 phase II may result in biased estimations for empirical Bayes estimates (EBE)<sup>9</sup> used in GAM  
51 covariate analysis procedure (generalised additive models)<sup>10</sup>.

52 To increase the detection of genetic covariates, a way could be to combine for the analysis  
53 data from a study collected with a rich design, as expected in phase I, with sparser, but still  
54 informative, data from a phase II study.

55 In the present work we propose practical designs involving phase I and phase II data, and we  
56 quantify through simulations their ability to detect genetic associations with PK. A  
57 motivating example was provided by IRIS, a pharmaceutical industry, to generate realistic  
58 genetic and PK data. We compared two association methods, a penalised regression method  
59 and a stepwise procedure<sup>6</sup>.

## 60 MATERIALS AND METHODS

### 61 Simulation study

62 Figure 1 presents the framework of the simulation study, which was designed based on PK  
63 data from drug S (IRIS) collected in 78 subjects from 3 phase I clinical studies<sup>11</sup>. All subjects  
64 were genotyped at baseline using a DNA microarray developed by IRIS of 176 SNPs known  
65 for being involved in the PK of drugs. These 176 polymorphisms were matched to a  
66 reference Hapmap panel (Hapmap 3 release 2) for a Caucasian population<sup>12</sup> and we used the  
67 Hapgen2 software<sup>13</sup> to simulate genetic variants retaining their frequencies and the  
68 correlations between polymorphisms found in the human genome (see details in  
69 **Supplementary Material, Supplementary Figures S1-S3**).

70 PK profiles were simulated with a two-compartment model with dose-dependent double  
71 absorption (**Supplementary Figure S4**), with the parameters in **Table 1**, under two  
72 conditions: (i) no gene effect ( $H_0$ ); (ii) gene effect on clearance CL ( $H_1$ ). Under  $H_1$ , 6 SNPs  
73 were drawn randomly without physiological assumptions or prior knowledge, and assumed  
74 to explain in total 30% of the IIV on CL through the following additive genetic model on the  
75 log-transformed CL:

$$\log(CL_{sim_i}) = \log(\mu_{CL}) + \sum_{k=1}^6 \beta_k \times SNP_{ik} + \eta_{i_{CL}} \quad (1)$$

76 where  $CL_{sim_i}$  is the simulated individual clearance,  $\mu_{CL}$  the typical clearance,  $\beta_k$  the effect  
77 size associated to the variant allele of  $SNP_{ik}$  and  $\eta_{i_{CL}}$  the interindividual random effect for  
78 clearance of subject  $i$ . Causal SNPs were different from one dataset to another. Assuming an  
79 additive genetic model, genotypes take values 0, 1 or 2, reflecting the number of mutated  
80 alleles. We chose this model to simplify the simulations but dominant or recessive genetic  
81 models could be easily simulated by changing genotype values.  $\beta_k$  was computed as a  
82 function of the coefficient of genetic component ( $R_{GC_k}$ , the percentage of the interindividual  
83 variability in CL explained by the SNP) and the minor allele fraction ( $p_k$ ), as follows:

$$\beta_k = \sqrt{\frac{R_{GC_k} \times \omega_{CL}^2}{2p_k(1-p_k) - R_{GC_k} \times 2p_k(1-p_k)}} \quad (2)$$

84 where  $\omega_{CL}^2$  is the variance of interindividual random effects on CL due to non-genetic  
85 sources.  $R_{GC_k}$  was respectively equal to 1, 2, 3, 5, 7 and 12% for the 6 causal variants<sup>14</sup> to

86 mimic a multifactorial genetic effect. Then under  $H_0$ ,  $\omega^2_{CL} = 0.06$  (as in Table I), while  
87 under  $H_1$  30% of the variance is explained by the genetics so that  $\omega^2_{CL} = 0.04$  (example on  
88 the magnitude of simulated effect sizes is available in **Supplementary Table S1**).

89 The simulated datasets were then fitted with the base model without genetic covariates.  
90 Individual clearance estimates ( $EBE_{CL_i}$ ) were estimated and all associations with the 176  
91 simulated polymorphisms were tested assuming a linear relation without re-estimating  
92 model parameters, as in a GAM analysis<sup>10</sup>.

93 We compared two association methods to detect gene effects. Lasso<sup>15</sup> is a multivariate  
94 penalised regression which simultaneously estimates effect size coefficients and selects  
95 variants by setting a large number of coefficients to 0. The penalty is set by a tuning  
96 parameter ( $\xi$ ) which depends on  $\alpha$ , the type I error per test, and the number of subjects<sup>16,17</sup>  
97 (**Figure 1**). Alternatively in practice the penalty can be determined through permutation or  
98 cross-validation methods, which are more time consuming. Stepwise procedure includes  
99 relationships one by one depending on the significance of a Wald test compared to a  
100 threshold  $\alpha$ . The correlation between two significant SNPs, due to linkage disequilibrium, is  
101 computed through the Pearson's correlation coefficient  $r$  and if two significant SNPs are  
102 strongly correlated ( $|r| > 0.89$ ), only the most significant is kept. Finally the most significant  
103 variant among selected SNPs is kept in the final model and steps are repeated until no more  
104 association is significant<sup>6</sup> (**Figure 1**).

105 In both approaches, we control the Family Wise Error Rate (FWER), representing the  
106 percentage of datasets where at least one variant is selected under  $H_0$ , by correcting the  
107 nominal  $\alpha$  by the number of tests performed (Sidak correction) corresponding to the  
108 number of polymorphic SNPs considered  $N_t$  (**Figure 1**). The FWER was set to 20% (with a  
109 prediction interval for 200 datasets equal to [14.5-25.5]) for an exploratory analysis. The  
110 prediction interval determined when to adjust  $\alpha$  to control the FWER under  $H_0$ .

### 111 **Simulated designs and analysis scenarios**

112 We simulated a phase I study corresponding to the motivating example, including 78  
113 subjects (N1) receiving 8 different single doses (5, 10, 20, 50, 100, 200, 400 or 800 units, for  
114 respectively 6, 6, 24, 12, 12, 6, 6 and 6 subjects per dose) and sampled at 16 times. Three  
115 designs of phase II study were simulated. They included 306 subjects (N2), receiving 3 doses  
116 (20, 50 or 100 units, 102 subjects per dose), sampled at steady state. Two phase II studies

117 included three samples per subject, optimised using the PFIM software<sup>18</sup> to ensure a  
 118 reasonable precision of CL estimates. The last sampling time was limited to 24h in one, while  
 119 a late sample was allowed after the last dose administration in the other. The third study  
 120 included only one trough concentration (24h). We considered 4 analysis scenarios (Figure 1),  
 121 three combining the phase I and one of the phase II study (respectively SPI/II<sub>3s,24h</sub>, SPI/II<sub>3s,96h</sub>  
 122 and SPI/II<sub>1s,24h</sub>), and one, for comparison, with only the phase I subjects (SPI).  
 123 We also investigated the impact of a higher variability on phenotype on the results. For this,  
 124 we simulated the same four scenarios increasing the IIV on CL to 60% (instead of 25% in  
 125 previous settings).

## 126 Evaluation

127 For each analysis scenario, 200 datasets were simulated under  $H_0$  and  $H_1$ .

128 The ability of the designs to estimate the population and individual parameters under  $H_0$  was  
 129 first evaluated through estimation bias and  $\eta$ -shrinkage (see details in **Supplementary**  
 130 **Material, Supplementary Figures S5-S6**).

131 Under  $H_1$  we evaluated the performance of each scenario in terms of true and false positives  
 132 counts (TP and FP) and rates (TPR, the proportion of TP detected among the causal variants;  
 133 and FPR, the proportion of FP detected among all potential false associations) for parameter  
 134 CL, as well as the probability to detect genetic variants. Assuming that SNPs located on genes  
 135 coded for metabolism enzymes and transporters affect mostly the drug distribution and  
 136 elimination, we also applied association tests on Q, the intercompartmental clearance and  
 137 V2, the peripheral volume, separately. Any variants associated to Q and V2 were counted as  
 138 false positives. The central volume V1 was not considered because it had no random effects.  
 139 We also evaluated the loss of genetic signal between simulated and estimated individual  
 140 clearances, comparing slopes  $b$  of univariate linear regressions on  $\log(CL_{sim_i})$  or  
 141  $\log(EBE_{CL_i})$  for each causal variant. A relative deviation of the genetic signal  $RD_{signal}$  was  
 142 computed as follows:

$$RD_{signal}(\%) = \frac{b_{EBE_{CL_i}} - b_{CL_{sim_i}}}{b_{CL_{sim_i}}} \times 100 \quad (3)$$

143  $RD_{signal}$  quantifies the departure of the estimated genetic signal ( $b_{EBE_{CL_i}}$ ) from the one  
 144 simulated ( $b_{CL_{sim_i}}$ , see details in **Supplementary Material**).

## 145 RESULTS

### 146 Control of FWER under $H_0$

147 Lasso and stepwise procedure both tended to be too conservative, as the FWER was lower  
148 than expected in some scenarios (**Table 2**). After an empirical correction by increasing the  
149 type I error per SNP  $\alpha$ , FWER was properly controlled around 20%. This correction was  
150 applied in the corresponding simulation under  $H_1$ . Previous simulations suggested that this  
151 decrease in FWER is influenced by correlations between polymorphisms<sup>6</sup>.

### 152 Detection of genetic effects

153 Under  $H_1$  the TPR (**Figure 2, top left**) was higher in scenarios including phase II data (from 22  
154 to 32%) compared to scenario with only phase I data (SPI, 4%) and was the highest in  
155 scenario SPI/II<sub>3s,96h</sub>. The FPR was lowest (0.2%) in scenario SPI where a limited number of  
156 SNPs was selected, and only slightly higher in scenarios including phase II data, ranging from  
157 0.6 to 0.8% for both methods. Very few TP were effectively detected in scenario SPI (around  
158 44 for both methods) where the number of subjects was limited ( $N_1 = 78$ ) (**Supplementary**  
159 **Tables S2-S3**). By adding more subjects ( $N_2 = 306$ ) to the analysis, the number of TP  
160 increased sharply. Scenario SPI/II<sub>3s,96h</sub> allowed detecting the largest number of TP (380 or  
161 more), while in SPI/II<sub>3s,24h</sub> around 326 TP were detected. In SPI/II<sub>1s,24h</sub> the number of TP was  
162 lower (around 270 TP), but remained much higher than scenario SPI with only phase I data.  
163 In the same way, the number of FP increased when including phase II data to the analysis,  
164 but to a much lesser extent.

165 With only phase I data, the probability to detect at least one genetic variant on CL was low  
166 (**Figure 2, bottom left**), around 20% (SPI). This probability decreased quickly when trying to  
167 detect more polymorphisms and reached 0 for 3 variants or more. Adding phase II data to  
168 the analysis increased the probability to detect at least one variant about 85% in scenario  
169 SPI/II<sub>1s,24h</sub>, and up to 95% in scenario SPI/II<sub>3s,96h</sub>. Scenarios including phase II data showed  
170 good detection of 1 to 3 SNPs and SPI/II<sub>3s,96h</sub> had always the higher detection. This shows  
171 that the major determinant of power is the number of subjects, and that optimising the  
172 design for more informativeness can bring a smaller further improvement. The low  
173 probability to detect 4 SNPs or more ( $\leq 4\%$ ) in scenarios combining phase I and phase II data  
174 can be explained by those variants having a very weak impact; polymorphisms only  
175 explaining 1, 2 or 3% of the variability of CL.

176 In **Supplementary Table S4**, the TPR was computed separately for each causal SNP. The  
177 variants associated to the lowest  $R_{GC}$  had low TPR, close to the FPR. Thus the signal  
178 associated to these variants was close to the noise created by the non-causal variants

### 179 **Shrinkage**

180 Two  $\eta$ -shrinkage estimates were computed using metric proposed by Bertrand *et al.*<sup>4</sup> based  
181 on estimated variances, with respect to the estimate of  $\hat{\omega}^2$  in the dataset; one over the  $\hat{\eta}_i$   
182 from phase I subjects and one over the  $\hat{\eta}_i$  from phase II subjects (**Figure 3**). The  $\eta$ -shrinkage  
183 for phase I subjects was low (median = 23%) thanks to the large number of observations per  
184 subject. A large range of  $\eta$ -shrinkage estimates for phase II data was observed across  
185 analysis scenarios, but was below 50% in scenario SPI/II<sub>3s,96h</sub>.

### 186 **Loss of signal**

187  $RD_{signal}$  was always negative for the 6 SNPs, indicating that part of the signal was lost  
188 during the estimation step (**Figure 4, top**). This loss was smaller in the scenario with phase I  
189 data alone (SPI) than in scenarios combining phase I and phase II data. In each scenario the  
190 signal loss was of the same magnitude for the 6 SNPs, regardless of the value of associated  
191  $R_{GC}$ . For phase I data (**Figure 4, bottom**), the loss was of a constant magnitude across  
192 scenarios (median = -30%). For phase II data, in the most informative scenario (SPI/II<sub>3s,96h</sub>)  
193 the loss was of a similar magnitude (median = -41%) than the loss in phase I data, where  
194 subjects were extensively sampled. The loss was higher in scenario SPI/II<sub>3s,24h</sub> (median = -  
195 56%), and even more when only one time was sampled (SPI/II<sub>1s,24h</sub>, median = -70%).  
196 The signal loss and  $\eta$ -shrinkage values changed accordingly across phase II scenarios, while  
197 the probability of detection changed in the opposite direction.

### 198 **Influence of the phenotype variance**

199 Increasing IIV for the CL parameter to 60% led to a sharp increase in the number of TP  
200 (**Supplementary Tables S5-S6**), resulting in higher TPR and higher probabilities to detect the  
201 causal variants (**Figure 2, right**), compared to when individual CL were simulated with a  
202 moderate IIV. This higher number of TP is explained first and foremost by the increase in  
203 simulated effect sizes which depended on the variance of interindividual random effects on  
204 CL due to non-genetic sources (equation 2). A second consequence of the larger IIV was that  
205 the estimated  $\eta$ -shrinkages became much smaller. Lower  $\eta$ -shrinkages resulted in lower



206 signal losses in all scenarios for phase I and phase II data (**Supplementary Figure S7-S8**),  
207 which again favoured a higher probability to detect the genetic effects.

## 208 **DISCUSSION**

209 In this work, we show and evaluate practical designs to combine data from studies occurring  
210 in phase I and II of a drug development. We assess through a simulation study, inspired by a  
211 real example, the probability to detect genetic variants and the influence of the phase II  
212 study design. We considered phenotypes estimated by NLMEM, which can handle the  
213 analysis of heterogeneous data involving sparsely sampled subjects.

214 Genetic variants are unbalanced and so the amount of information they provide is directly  
215 related to the variant allele frequency and the study sample size. On the other hand, PK  
216 information depends also on the number and times of sampling which drives the precision of  
217 the PK model parameter estimates. A limited number of samples, as in phase II studies, may  
218 lead to missing a true association when EBE are used as phenotypes<sup>9</sup>. Savic and Karlsson  
219 suggested a more extensive use of the likelihood ratio test (LRT) for covariate selection when  
220  $\eta$ -shrinkage is large, but Combes et al. showed that the power to detect a covariate effect is  
221 the same with a LRT or a simple correlation test on EBE<sup>19</sup>.

222 The effect of sample size can be distinctly observed in our simulations. In the context of  
223 phase I studies, where the number of subjects is limited, the probability to detect the  
224 genetic effects was low, in line with our previous results<sup>6</sup>. The combined analysis of phase I  
225 and phase II data allowed a marked improvement in this detection probability, irrespective  
226 of the phase II study design. By modifying the design of the phase II data, we highlighted a  
227 direct link between  $\eta$ -shrinkage, loss of genetic signal and probability to detect genetic  
228 variants. Our results showed that poor PK information due to the phase II study design  
229 results in higher  $\eta$ -shrinkage, which increases the loss of genetic signal at the estimation step  
230 and translates to a lower probability to detect genetic variant. The dilution of the individual  
231 information by adding subjects with sparse designs to subjects with rich designs increases as  
232 expected the loss of genetic signal. But this is accompanied with a sharp increase in  
233 detection power thanks to a larger sample size.  $\eta$ -shrinkage may also modify the EBE-EBE  
234 relationship, falsely inducing or masking correlations between model parameters<sup>9</sup>. This  
235 could result in an increased number of false positives associated to other parameters than

236 CL, although in our simulations the number of FP on CL, V2 and Q remained of a similar  
237 magnitude across scenarios (**Supplementary Table S3**), showing no systematic effect.

238 We assume homogeneity of the PK between subjects simulated for the phase I and the  
239 phase II study. In practice healthy volunteers are often included in phase I while phase II  
240 studies focus on patients. A difference in typical values, for example of CL, between the two  
241 populations should not impact the detection power by combination of data, as the  
242 association tests use the phenotype variance, provided that the genetic effect is the same  
243 and that the model accounts for the systematic difference between clearances. It is more  
244 difficult to predict what would happen if the variability of clearance is different in the two  
245 populations, as the magnitude of the shrinkage in each subpopulation could affect the signal  
246 detection. When the assumption that the two populations are similar breaks down, we  
247 would suggest instead to combine rich and sparse data within the phase II study.

248 Pharmacogenetic studies including a large number of subjects combining sparse and rich  
249 designs have already been published<sup>20,21</sup>, showing that the combination of different  
250 sampling designs is feasible within the same study to assure more homogeneity.

251 Situations where pharmacogenetic analyses in PK studies are recommended are described  
252 by health authorities<sup>22</sup>. In our work, we simulated a blinded pharmacogenetic analysis,  
253 exploring a large number of genetic markers. In real applications, other considerations than  
254 the statistical significance of genetic variants such as their physiological and clinical  
255 relevance could be factored in the analysis and its interpretation. Lehr et al.<sup>23</sup> proposed in  
256 their stepwise procedure to select only significant polymorphisms having a physiologic  
257 relevance in the final model, and the same constraint could be integrated in penalised  
258 regression approaches. The probability to detect genetic variants could also be increased  
259 through the targeted inclusion of subjects for a few polymorphisms of interest, but this  
260 approach requires hypotheses on which polymorphisms to test, with a risk to miss important  
261 associations. We focused in this work on PK variability, which is a part of the variability in  
262 drug response. But the conclusions from the simulation study could be extended to  
263 pharmacodynamics. A previous survey indicated that most pharmacogenetic analyses in  
264 clinical PK studies used a phenotype estimated by NCA and furthermore included a limited  
265 number of subjects (lower than 50 subjects in two thirds)<sup>6</sup>. Authorities in fact recommend to  
266 study pharmacogenetics in phase I<sup>22</sup>, where the number of subjects is limited. Our work  
267 shows that such analyses do not have the power to detect polymorphisms efficiently but can

268 generate hypotheses to assess in later studies. A recent simulation work<sup>24</sup> studied the  
269 sample size required to detect a binary covariate. They conclude that around 60 subjects  
270 combining rich or sparse designs was sufficient to detect the covariate with at least 80%  
271 power. Again our simulations showed that genetic covariates require higher sample sizes  
272 because they are highly unbalanced.

273 In the first series of simulations a moderate IIV on CL (25%) was used, resulting in a low  
274 impact of the genetics on PK, since overall 30% of the moderate CL variability was explained  
275 by genetic variants. This setting represented a realistic case to challenge the detection of  
276 genetic variants through modelling. We also evaluated the same scenarios with a higher IIV  
277 for CL, set to 60%. The  $\eta$ -shrinkage was much lower, as a higher IIV downweighs the  
278 population priors in the combined criterion used to compute EBE. This decrease in CL  $\eta$ -  
279 shrinkages resulted in lower signal loss because of the direct relationship between the two.  
280 Associated with larger simulated effect sizes, the number of TP and the probability to detect  
281 genetic variants increased in these scenarios. The effect of  $\eta$ -shrinkage on the probability to  
282 detect genetic effects was in these simulations higher than the one we observed with the  
283 main settings, because the decrease of  $\eta$ -shrinkage was associated with a sharp increase in  
284 the number of TP. This shows that our conclusions don't depend on the level of IIV.

285 This simulation study also confirms the results of our previous work concerning the relative  
286 performance of the different association methods<sup>6</sup>. The penalised regression method Lasso  
287 and the stepwise procedure showed a similar probability to detect genetic variants in all  
288 scenarios. However, the Lasso is a slightly more complex method which requires computing  
289 the penalty in a first step before testing the associations. In this work we assessed methods  
290 to detect genetic effects on EBE, after an initial fit. An algorithm proposed by Lehr et al.<sup>23</sup>  
291 uses univariate regressions to select variants to test in the PK model through LRT. This  
292 approach is easy to implement but run-times depend on the number of iterations leading to  
293 the full covariates model. An alternative is to use an integrated approach where effect sizes  
294 are estimated and significant variants selected using a penalised regression in the same  
295 step<sup>17</sup>; it showed similar performance than the stepwise procedure proposed by Lehr et al.,  
296 but with longer computing times<sup>17</sup>. The results for the two other penalised regression  
297 methods tested in the previous work, ridge regression and HyperLasso, were similar  
298 **(Supplementary Tables S7-S9, Figures S9-S10)**. None of the methods detected the 6 SNPs  
299 simultaneously, as 3 of the polymorphisms only explained 1 to 3% of the clearance

300 variability, making them difficult to detect. Because association methods relate the  
301 polymorphisms to the phenotype variance, we fixed the variance explained by the causal  
302 variants (through the parameter  $R_{GC}$ ) and computed the effect sizes as a function of their  
303 allelic frequencies. For a given  $R_{GC}$  an infrequent polymorphism was therefore associated  
304 with higher effect sizes. This reflects that a clinically relevant polymorphism (with a high  
305 impact on PK), present in few subjects because of its low frequency, will explain a limited  
306 proportion of the phenotype variance. Detecting such polymorphisms is crucial to identify  
307 subpopulation at risk but required much larger sample sizes, as in genome-wide studies<sup>25</sup>. As  
308 an example, the rs3918290 polymorphism from gene DPYD has a frequency lower than 1%,  
309 but results in a deficient dihydropyrimidine dehydrogenase activity associated with a 40%  
310 decrease of the maximum conversion capacity of the chemotherapeutic drug 5-  
311 fluorouracil<sup>26</sup>, resulting in severe toxicities.

312 The power to detect polymorphisms is also closely related to the type I error chosen for the  
313 analysis. In a context of exploratory analyses, we fixed the global type I error to 20%. But  
314 using the Sidak correction the significance thresholds were finally lower than 0.1% for each  
315 test, so that only strong effects of causal variants will be detected, and our simulations show  
316 that polymorphisms explaining a limited part of the phenotype variance are not detected.  
317 Approaches based on FWER and corrections as Bonferroni or Sidak are easy to implement  
318 but are conservative and may reduce the power of analyses, but limit the number of  
319 polymorphisms to test in later confirmatory trials. In practice other corrections for type I  
320 error could be considered, as permutation methods which are more time consuming but less  
321 conservative.

322 Although this correction was conservative and was calibrated under  $H_0$  to control the FWER,  
323 the proportion of FP under  $H_1$  amongst selected variants was higher than the expected 20%.  
324 This could reflect the correlations between polymorphisms we simulated.

325 To make more specific recommendations for study designs is difficult because it is closely  
326 related to the developed drug. In our simulations a late sample allowed larger information  
327 on the elimination phase to estimate CL. This result can be generalised to pharmacogenetic  
328 studies involving clearance and drugs with a long half-life. Taking a late sample requires to  
329 suspend treatment long enough to observe a decrease in concentrations, which may not be  
330 possible in patients from phase II trials.

331 In any case, it is essential that the sampling protocol, although limited, is as informative as  
332 possible to minimise the estimation error and shrinkage in individual parameters estimation.  
333 The detection of genetic polymorphisms could highly benefit from the use of larger sample  
334 sizes through combined analysis and optimised design<sup>18,27</sup>.  
335 In conclusion, this work confirmed the very limited likelihood that weak genetic effects can  
336 be detected in a typical phase I study, due to the small sample size. Such studies have to be  
337 considered only as hypothesis generating<sup>28</sup>. On the basis of our results in term of detection  
338 probability when analysing together data from phase I and phase II study, we claim that  
339 phase II is the best moment to identify the impact of genetic variants on drug response. It  
340 would be less efficient to start the study of the pharmacogenetics of a new drug in phase III  
341 trials or in post-marketing, because these take place too late in drug development<sup>29</sup> and the  
342 new treatment could be administered in non-responders or expose subjects to high  
343 toxicities. Furthermore genetic subpopulations can be better targeted and potentially some  
344 subjects excluded from the study to increase the efficacy and reduce the risk of toxicity of  
345 the drug in these phase III studies.

## 346 **STUDY HIGHLIGHTS**

### 347 ***What is the current knowledge on the topic?***

348 Most pharmacogenetic analyses in pharmacokinetic studies recently published included a  
349 limited number of subjects (less than 50). Previous simulations showed that such sample  
350 sizes result in low probability to detect polymorphisms. But with large number of subjects,  
351 extensive pharmacokinetic information is difficult to obtain in drug development.

### 352 ***What question did this study address?***

353 This simulation study explored realistic ways to increase the amount of information by  
354 combining rich phase I data and sparse phase II data, and optimising such sparse designs.

### 355 ***What this study adds to our knowledge?***

356 This study shows that even sparse data from phase II allow a marked improvement in the  
357 probability to detect genetic variants when combined with rich data from phase I, even more  
358 when sparse designs are optimised.

### 359 ***How this might change clinical pharmacology and therapeutics?***

360 The pharmacogenetic analyses should be planned later in the drug development to take  
361 advantage of larger sample sizes by combining data which would increase the power to  
362 detect genetic effects.

### 363 **Acknowledgments**

364 Adrien Tessier received funding from Institut de Recherches Internationales Servier. The  
365 authors thank Laurent Ripoll and Bernard Walther from Institut de Recherches  
366 Internationales Servier for their advices in pharmacogenetics. The authors would also like to  
367 thank Hervé Le Nagard for the use of the computer cluster services hosted on the “Centre de  
368 Biomodélisation UMR1137”.

### 369 **Conflict of Interest**

370 Adrien Tessier has a research grant from Institut de Recherches Internationales Servier and  
371 the French government. Marylore Chenel works for Institut de Recherches Internationales  
372 Servier, heading the department of Clinical Pharmacokinetics and Pharmacometrics.

### 373 **Author Contributions**

374 Adrien Tessier, Julie Bertrand, Marylore Chenel and Emmanuelle Comets designed the  
375 research. Adrien Tessier performed the research. Adrien Tessier, Julie Bertrand, Marylore  
376 Chenel and Emmanuelle Comets analyzed the results. Adrien Tessier, Julie Bertrand,  
377 Marylore Chenel and Emmanuelle Comets wrote the manuscript.

## 378 **REFERENCES**

- 379 1. Aarons, L. Population pharmacokinetics: theory and practice. *Br. J. Clin. Pharmacol.* **32**,  
380 669–670 (1991).
- 381 2. Motulsky, A. G. Drugs and genes. *Ann. Intern. Med.* **70**, 1269–1272 (1969).
- 382 3. Guo, Y., Shafer, S., Weller, P., Usuka, J. & Peltz, G. Pharmacogenomics and drug  
383 development. *Pharmacogenomics* **6**, 857–864 (2005).
- 384 4. Bertrand, J., Comets, E., Laffont, C. M., Chenel, M. & Mentré, F. Pharmacogenetics and  
385 population pharmacokinetics: impact of the design on three tests using the SAEM  
386 algorithm. *J. Pharmacokinet. Pharmacodyn.* **36**, 317–339 (2009).
- 387 5. Bertrand, J., Comets, E., Chenel, M. & Mentré, F. Some alternatives to asymptotic tests  
388 for the analysis of pharmacogenetic data using nonlinear mixed effects models.  
389 *Biometrics* **68**, 146–155 (2012).
- 390 6. Tessier, A., Bertrand, J., Chenel, M. & Comets, E. Comparison of nonlinear mixed effects  
391 models and noncompartmental approaches in detecting pharmacogenetic covariates.  
392 *AAPS J.* **17**, 597–608 (2015).
- 393 7. Sheiner, L. B., Rosenberg, B. & Melmon, K. L. Modelling of individual pharmacokinetics  
394 for computer-aided drug dosage. *Comput. Biomed. Res. Int. J.* **5**, 411–459 (1972).

- 395 8. Rowland, M. & Tozer, T. N. *Clinical pharmacokinetics and pharmacodynamics: concepts*  
396 *and applications*. (Wolters Kluwer Health/Lippincott William & Wilkins, Philadelphia,  
397 2011).
- 398 9. Savic, R. M. & Karlsson, M. O. Importance of shrinkage in empirical bayes estimates for  
399 diagnostics: problems and solutions. *AAPS J.* **11**, 558–569 (2009).
- 400 10. Mandema, J. W., Verotta, D. & Sheiner, L. B. Building population pharmacokinetic--  
401 pharmacodynamic models. I. Models for covariate effects. *J. Pharmacokinet. Biopharm.*  
402 **20**, 511–528 (1992).
- 403 11. Tessier, A., Bertrand, J., Fouliard, S., Comets, E. & Chenel, M. High-throughput genetic  
404 screening and pharmacokinetic population modeling in drug development. (2013).  
405 Abstract #2836 at <[www.page-meeting.org/?abstract=2836](http://www.page-meeting.org/?abstract=2836)>
- 406 12. International HapMap Consortium The International HapMap Project. *Nature* **426**, 789–  
407 796 (2003).
- 408 13. Su, Z., Marchini, J. & Donnelly, P. HAPGEN2: simulation of multiple disease SNPs.  
409 *Bioinformatics* **27**, 2304–2305 (2011).
- 410 14. Bertrand, J. & Balding, D. J. Multiple single nucleotide polymorphism analysis using  
411 penalized regression in nonlinear mixed-effect pharmacokinetic models.  
412 *Pharmacogenet. Genomics* **23**, 167–174 (2013).
- 413 15. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **58**,  
414 267–288 (1994).
- 415 16. Hoggart, C. J., Whittaker, J. C., De Iorio, M. & Balding, D. J. Simultaneous analysis of all  
416 SNPs in genome-wide and re-sequencing association studies. *PLoS Genet.* **4**, e1000130  
417 (2008).
- 418 17. Bertrand, J., De Iorio, M. & Balding, D. J. Integrating dynamic mixed-effect modelling and  
419 penalized regression to explore genetic association with pharmacokinetics.  
420 *Pharmacogenet. Genomics* **25**, 231–238 (2015).
- 421 18. Bazzoli, C., Retout, S. & Mentré, F. Design evaluation and optimisation in multiple  
422 response nonlinear mixed effect models: PFIM 3.0. *Comput. Methods Programs Biomed.*  
423 **98**, 55–65 (2010).
- 424 19. Combes, F., Retout, S., Frey, N. & Mentré, F. Powers of the Likelihood Ratio Test and the  
425 Correlation Test Using Empirical Bayes Estimates for Various Shrinkages in Population  
426 Pharmacokinetics. *CPT Pharmacomet. Syst. Pharmacol.* **3**, 1–9 (2014).
- 427 20. Chou, M. *et al.* Population pharmacokinetic-pharmacogenetic study of nevirapine in HIV-  
428 infected Cambodian patients. *Antimicrob. Agents Chemother.* **54**, 4432–4439 (2010).
- 429 21. Bertrand, J. *et al.* Multiple genetic variants predict steady-state nevirapine clearance in  
430 HIV-infected Cambodians. *Pharmacogenet. Genomics* **22**, 868–876 (2012).

- 431 22. EMA *Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic*  
432 *evaluation of medicinal products*. (2012).
- 433 23. Lehr, T., Schaefer, H.-G. & Staab, A. Integration of high-throughput genotyping data into  
434 pharmacometric analyses using nonlinear mixed effects modeling. *Pharmacogenet.*  
435 *Genomics* **20**, 442–450 (2010).
- 436 24. Klopogge, F., Simpson, J. A., Day, N. P. J., White, N. J. & Tarning, J. Statistical Power  
437 Calculations for Mixed Pharmacokinetic Study Designs Using a Population Approach.  
438 *AAPS J.* **16**, 1110–1118 (2014).
- 439 25. Takeuchi, F. *et al.* A genome-wide association study confirms VKORC1, CYP2C9, and  
440 CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433  
441 (2009).
- 442 26. Kuilenburg, A. B. P. van *et al.* Evaluation of 5-fluorouracil pharmacokinetics in cancer  
443 patients with a c.1905+1G>A mutation in DPYD by means of a Bayesian limited sampling  
444 strategy. *Clin. Pharmacokinet.* **51**, 163–174 (2012).
- 445 27. Combes, F. P., Retout, S., Frey, N. & Mentré, F. Prediction of shrinkage of individual  
446 parameters using the bayesian information matrix in non-linear mixed effect models  
447 with evaluation in pharmacokinetics. *Pharm. Res.* **30**, 2355–2367 (2013).
- 448 28. Bromley, C. M. *et al.* Designing pharmacogenetic projects in industry: practical design  
449 perspectives from the Industry Pharmacogenomics Working Group. *Pharmacogenomics*  
450 *J.* **9**, 14–22 (2009).
- 451 29. O'Donnell, P. H. & Stadler, W. M. Pharmacogenomics in early-phase oncology clinical  
452 trials: is there a sweet spot in phase II? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**,  
453 2809–2816 (2012).
- 454



## TABLES

**Table 1.** Population values ( $\mu$ ) and interindividual variability ( $\omega$ ) for the model parameters of drug S used in the simulation study.

Parameters		$\mu$	$\omega$ (%)
F <sup>a</sup>	Imax <sub>F</sub>	0.8	32.9
	D50 <sub>F</sub>	41.7	
FRAC <sup>b</sup>	E <sub>max</sub> <sub>FRAC</sub>	0.45	-
	D50 <sub>FRAC</sub>	18.6	
Tlag <sub>1</sub>		0.401	35.1
Tk0		1.59	31.6
Tlag <sub>2</sub>		22.7	-
Ka		0.203	-
V1		1520	-
Q		147	89.9
V2		2130	44.2
CL		94.9	25.1
$\sigma_{\text{slope}}$ (%)		20	-

a. For doses  $< 20$  units  $F = 1$ , for doses  $\geq 20$  units  $F = 1 - \frac{Imax_F (dose - 20)}{D50_F + dose - 20}$ , where dose is the amount administered.

b.  $FRAC = \frac{E_{max_{FRAC}} dose}{D50_{FRAC} + dose}$

F: bioavailability ; FRAC: fraction of dose ; Tk0: zero order absorption duration ; Tlag<sub>1</sub>: lag time of zero order absorption ; Ka: first order absorption constant rate ; Tlag<sub>2</sub>: lag time of first order absorption ; V1: central compartment volume ; V2: peripheral compartment volume ; Q: intercompartmental clearance ; CL: linear elimination clearance

**Table 2.** Empirical estimates of Family-Wise Error Rate under H<sub>0</sub> for both association tests.

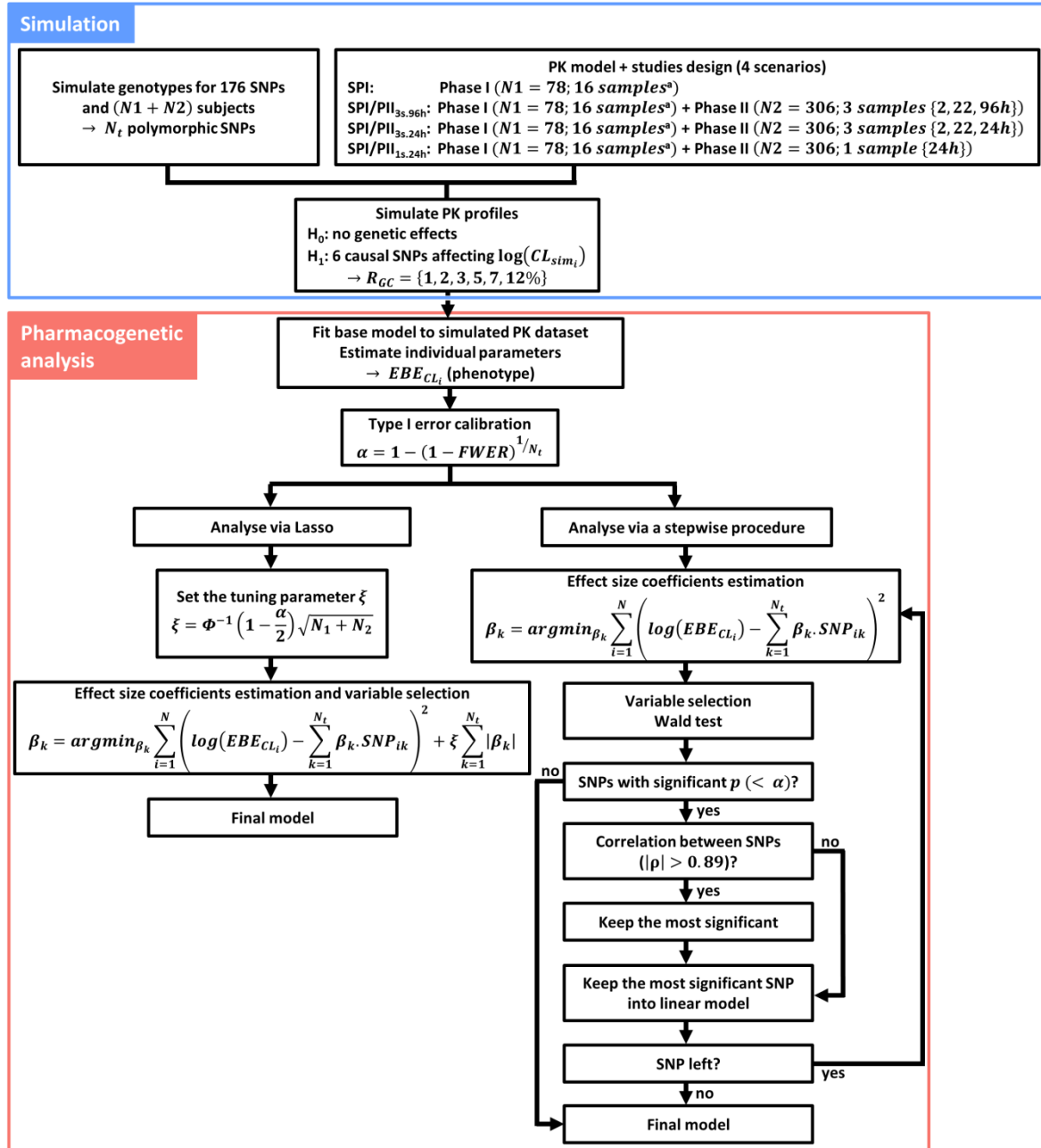
Method		FWER (%)			
		SPI	SPI/II <sub>3s,96h</sub>	SPI/II <sub>3s,24h</sub>	SPI/II <sub>1s,24h</sub>
Lasso	Without correction <sup>a</sup>	14	17.5	21.5	13.5
Stepwise procedure	Without correction <sup>a</sup>	20	18.5	22.5	15.5
Lasso	After empirical correction <sup>b</sup>	20	19.5	21.5	19.5
Stepwise procedure	After empirical correction <sup>b</sup>	20	20.5	22.5	20.5

a. Set of empirical family wise error rates (FWER) obtained without correction.

b. Set of empirical FWER obtained after correction of type I error per tests.

The 95% prediction interval around 20 for 200 simulated datasets is [14.5-25.5].

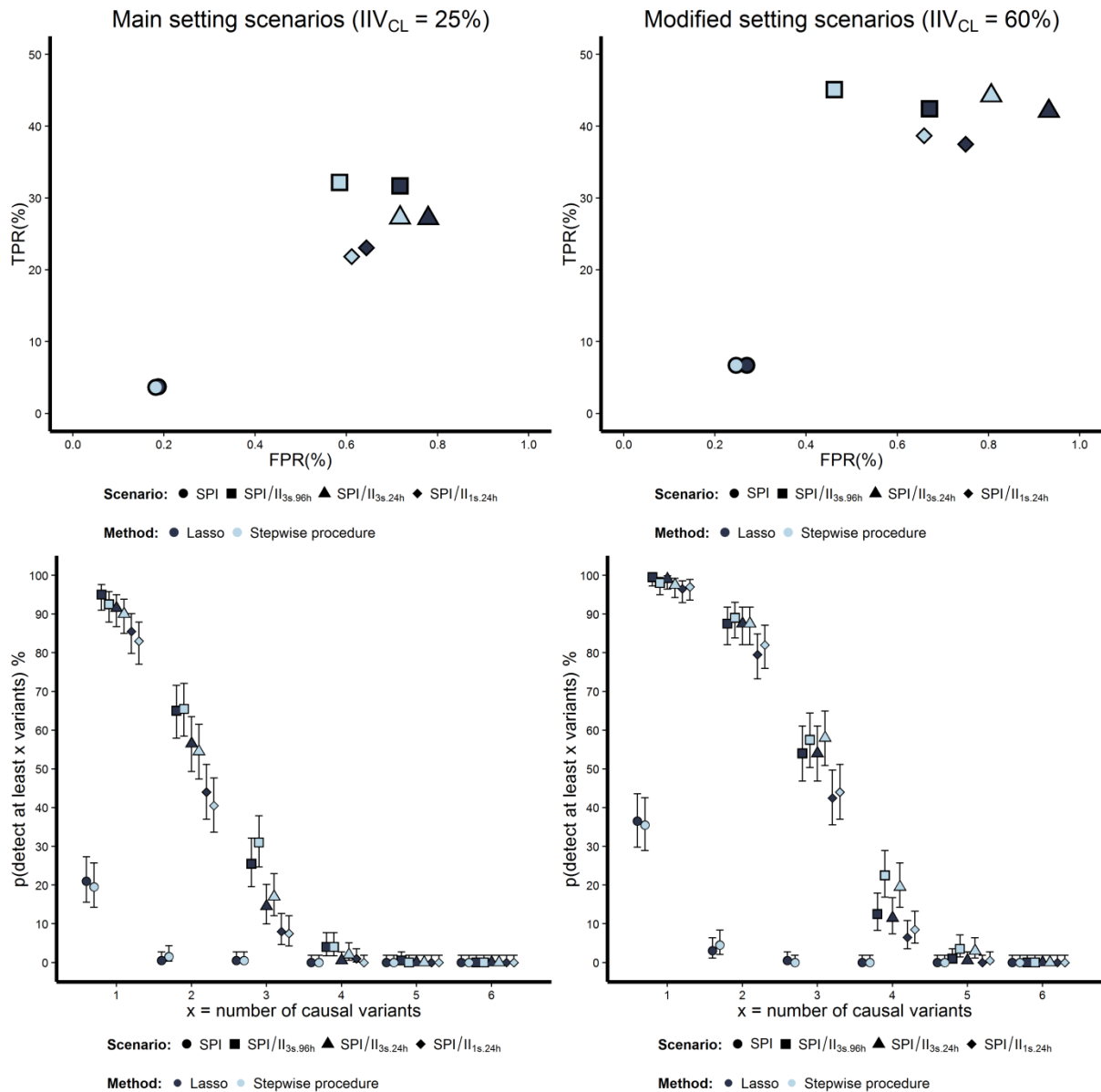
## FIGURES



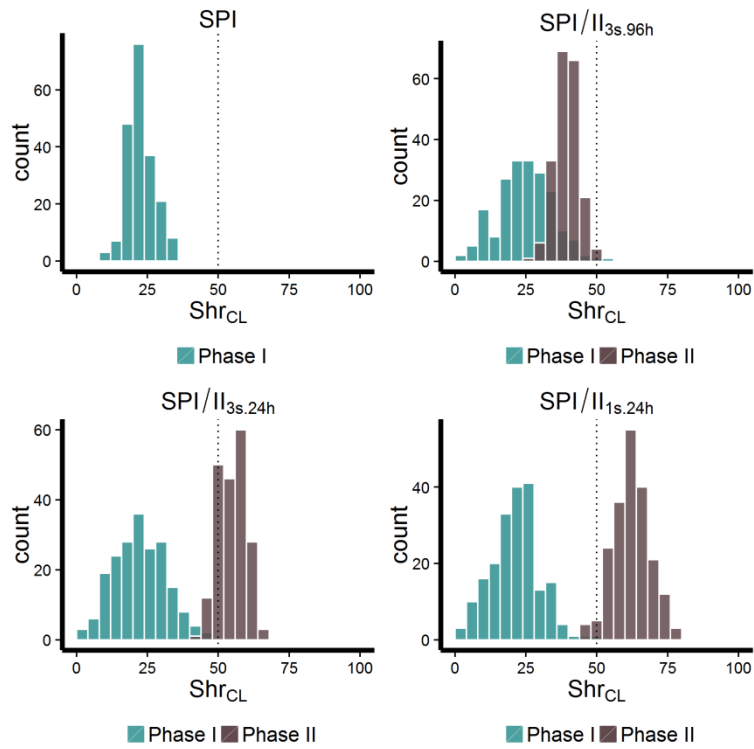
**Figure 1.** Workflow of the simulation study divided in the simulation (blue box) and analysis part (red box).

a. at 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 16, 24, 48, 72, 96, 120 and 192h.

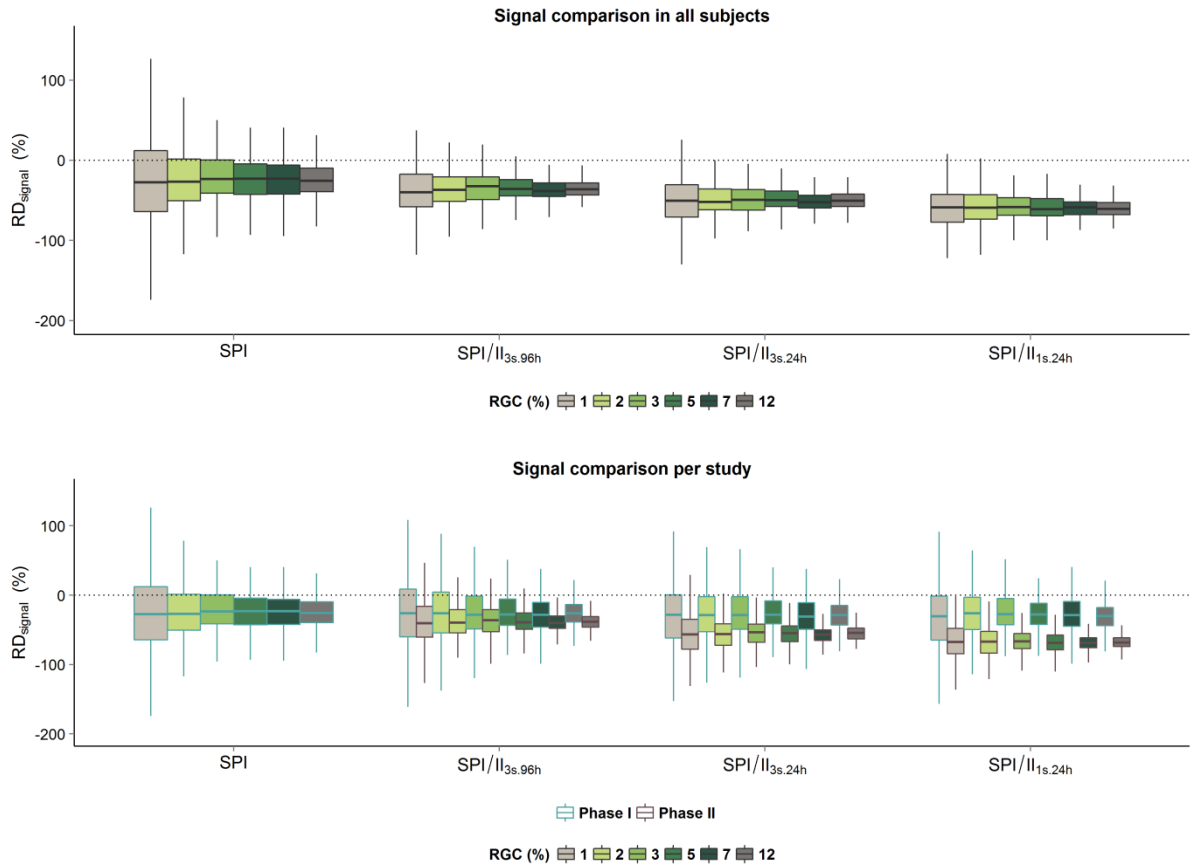
$CL_{sim_i}$ : simulated individual clearance ( $i = 1, \dots, N = N_1 + N_2$ );  $EBE_{CL_i}$ : empirical Bayes estimate of clearance;  $H_0$ : null scenarios;  $H_1$ : alternative scenarios;  $FWER$ : family wise error rate;  $N_1$ : number of subjects from the phase I study;  $N_2$ : number of subjects from the phase II study;  $N_t$ : number of polymorphic SNP to analyse;  $R_{GC}$ : genetic component of the interindividual variability;  $SNP_{ik}$ : single nucleotide polymorphism ( $k = 1, \dots, N_t$ );  $p$ : p value;  $\rho$ : correlation coefficient between variants;  $\alpha$ : type I error per test;  $\beta_k$ : effect size coefficient;  $\xi$  Lasso tuning parameter.



**Figure 2.** True Positive Rate (TPR) *versus* False Positive Rate (FPR) under  $H_1$  (*top*) and probability estimates (points) and 95% confidence interval (bars) to detect at least  $x$  causal variants explaining the interindividual variability of CL ( $x = 1, \dots, 6$ ) under  $H_1$  (*bottom*) for main scenarios simulated with  $IIV_{CL} = 25\%$  (*left*) or modified scenarios simulated with  $IIV_{CL} = 60\%$  (*right*). Different symbols are used for each scenario, and colours denote the Lasso (grey) and the stepwise procedure (light blue).



**Figure 3.** Distribution of the  $\eta$ -shrinkages on clearance for subjects in the phase I dataset (blue) and for subjects in the phase II dataset (brown), for each main scenario simulated under  $H_0$  with  $IIV_{CL} = 25\%$ .



**Figure 4.** Boxplot showing the loss of the signal for genetic effect in the overall population (*top*), as well as separately for the phase I data (blue borders) and for the phase II data (brown borders) (*bottom*). A boxplot is shown separately for each main scenario simulated under  $H_1$  with  $IIV_{CL} = 25\%$  as a function of increasing  $R_{GC}$  (boxplots colour).