

ORIGINAL ARTICLE

Combined Analysis of Phase I and Phase II Data to Enhance the Power of Pharmacogenetic Tests

A Tessier^{1,2,3*}, J Bertrand⁴, M Chenel³ and E Comets^{1,2,5}

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 with six genetic markers randomly sampled in each simulated dataset. We compared penalized 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 if sparse, increases the power to identify the associations between genetics and PK due to the larger sample size. Design optimization brings a further improvement, and we highlight a direct relationship between η -shrinkage and loss of genetic signal.

CPT Pharmacometrics Syst. Pharmacol. (2016) 5, 123–131; doi:10.1002/psp4.12054; published online 14 March 2016.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC? Most pharmacogenetic analyses in pharmacokinetic studies recently published included a limited number of subjects (fewer than 50). Previous simulations showed that such sample sizes result in a low probability to detect polymorphisms. But with large numbers of subjects, extensive pharmacokinetic information is difficult to obtain in drug development. • WHAT QUESTION DID THIS STUDY ADDRESS? This simulation study explored realistic ways to increase the amount of information by combining rich phase I data and sparse phase II data, and optimizing such sparse designs. • WHAT THIS STUDY ADDS TO OUR KNOWLEDGE This study shows that even sparse data from phase II allow a marked improvement in the probability to detect genetic variants when combined with rich data from phase I, even more when sparse designs are optimized. • HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS The pharmacogenetic analyses should be planned later in drug development to take advantage of larger sample sizes by combining data that would increase the power to detect genetic effects.

Studying the sources of the variability observed in drug response facilitates individualization of prescription. One of the sources of variability in drugs' pharmacokinetics (PK)¹ is the variation in activity of enzymes and transporters involved in the drug absorption, distribution, metabolism, or elimination. Pharmacogenetics² studies the genetic component of interindividual variability (IIV) observed in PK to identify populations at risk of treatment inefficacy or adverse effects.³ Single nucleotide polymorphisms (SNPs) are the genetic variants most frequently studied in pharmacogenetics and screened more and more often in clinical studies.

Genetic data offer some unique challenges, in particular because they may lead to a very unbalanced number of subjects, which impacts the power of tests in pharmacogenetic analyses.^{4,5} In a previous simulation work, we showed that typical phase I studies have low power to detect genetic effects because of the limited sample size.⁶ On the other hand, phase I studies generally provide good quality PK information, allowing characterization of the PK profile of the drug. We showed that from the different approaches used at this stage to estimate PK parameters, nonlinear

mixed effects models (NLMEM)⁷ could be considerably more powerful than noncompartmental analyses (NCA)⁸ for complex PK models.⁶ Our simulations also showed that increasing the sample size, as in phase II studies, would improve the power to detect genetic variants. However, sparse designs typically used in phase II may result in biased estimations for empirical Bayes estimates (EBE)⁹ used in a generalized additive model (GAM) covariate analysis procedure.¹⁰

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

In the present work we propose practical designs involving phase I and phase II data, and we quantify through simulations their ability to detect genetic associations with PK. A motivating example was provided by IRIS (Institut de Recherches Servier), a pharmaceutical industry, to generate realistic genetic and PK data. We compared two association methods, a penalized regression method and a stepwise procedure.⁶

¹INSERM, IAME, UMR 1137, Paris, France; ²Université Paris Diderot, IAME, UMR 1137, Sorbonne Paris Cité, Paris, France; ³Division of Clinical Pharmacokinetics and Pharmacometrics, Institut de Recherches Internationales Servier, Suresnes, France; ⁴University College London, Genetics Institute, London, UK; ⁵INSERM CIC 1414, Université Rennes 1, Rennes, France. *Correspondence: A Tessier (adrien.tessier@inserm.fr)

Received 10 August 2015; accepted 11 December 2015; published online on 14 March 2016. doi:10.1002/psp4.12054

MATERIALS AND METHODS

Simulation study

Figure 1 presents the framework of the simulation study, which was designed based on PK data from drug S (IRIS) collected in 78 subjects from three phase I clinical studies.¹¹ All subjects were genotyped at baseline using a DNA microarray developed by IRIS of 176 SNPs known for being involved in the PK of drugs. These 176 polymorphisms were matched to a reference Hapmap panel (Hapmap 3, release 2) for a Caucasian population¹² and we used the Hapgen2 software¹³ to simulate genetic variants retaining their frequencies and the correlations between polymorphisms found in the human genome (see details in **Supplementary Material, Supplementary Figures S1–S3**).

PK profiles were simulated with a two-compartment model with dose-dependent double absorption (**Supplementary Figure S4**), with the parameters in **Table 1**, under two conditions: 1) no gene effect (H_0); 2) gene effect on clearance CL (H_1). Under H_1 , six SNPs were drawn randomly without physiological assumptions or prior knowledge, and assumed to explain in total 30% of the IIV on CL through the following additive genetic model on the log-transformed CL:

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

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

$$\beta_k = \sqrt{\frac{R_{GCK} \times \omega^2_{CL}}{2p_k(1-p_k) - R_{GCK} \times 2p_k(1-p_k)}} \quad (2)$$

where ω^2_{CL} is the variance of interindividual random effects on CL due to nongenetic sources. R_{GCK} was respectively equal to 1, 2, 3, 5, 7 and 12% for the six causal variants¹⁴ to mimic a multifactorial genetic effect. Then under H_0 , $\omega^2_{CL} = 0.06$ (as in **Table 1**), while under H_1 30% of the variance is explained by the genetics so that $\omega^2_{CL} = 0.04$ (example on the magnitude of simulated effect sizes is available in **Supplementary Table S1**).

The simulated datasets were then fitted with the base model without genetic covariates. Individual clearance estimates (EBE_{CL_i}) were estimated and all associations with the 176 simulated polymorphisms were tested assuming a linear relation without reestimating model parameters, as in a GAM analysis.¹⁰

We compared two association methods to detect gene effects. Lasso¹⁵ is a multivariate penalized regression which simultaneously estimates effect size coefficients and selects variants by setting a large number of coefficients to 0. The penalty is set by a tuning parameter (ξ) which depends on α , the type I error per test, and the number of subjects^{16,17} (**Figure 1**). Alternatively, in practice the penalty can be determined through permutation or cross-validation methods, which are more time-consuming. A stepwise procedure includes relationships one-by-one depending on the significance of a Wald test compared to a threshold α . The correlation between two significant SNPs, due to linkage disequilibrium, is computed through the Pearson's correlation coefficient r and if two significant SNPs are strongly correlated ($|r| > 0.89$), only the most significant is kept. Finally, the most significant variant among selected SNPs is kept in the final model and steps are repeated until no association is significant⁶ (**Figure 1**).

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

Simulated designs and analysis scenarios

We simulated a phase I study corresponding to the motivating example, including 78 subjects (N1) receiving eight different single doses (5, 10, 20, 50, 100, 200, 400, or 800 units, for respectively 6, 6, 24, 12, 12, 6, 6, and 6 subjects per dose) and sampled 16 times. Three designs of phase II study were simulated. They included 306 subjects (N2), receiving three doses (20, 50, or 100 units, 102 subjects per dose), sampled at steady state. Two phase II studies included three samples per subject, optimized using the PFIM software¹⁸ to ensure a reasonable precision of CL estimates. The last sampling time was limited to 24 hours in one, while a late sample was allowed after the last dose administration in the other. The third study included only one trough concentration (24 hours). We considered four analysis scenarios (**Figure 1**), three combining the phase I and one of the phase II study (respectively, SPI/II_{3s,24h}, SPI/II_{3s,96h}, and SPI/II_{1s,24h}), and one, for comparison, with only the phase I subjects (SPI).

We also investigated the impact of a higher variability on phenotype on the results. For this, we simulated the same four scenarios increasing the IIV on CL to 60% (instead of 25% in previous settings).

Evaluation

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

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

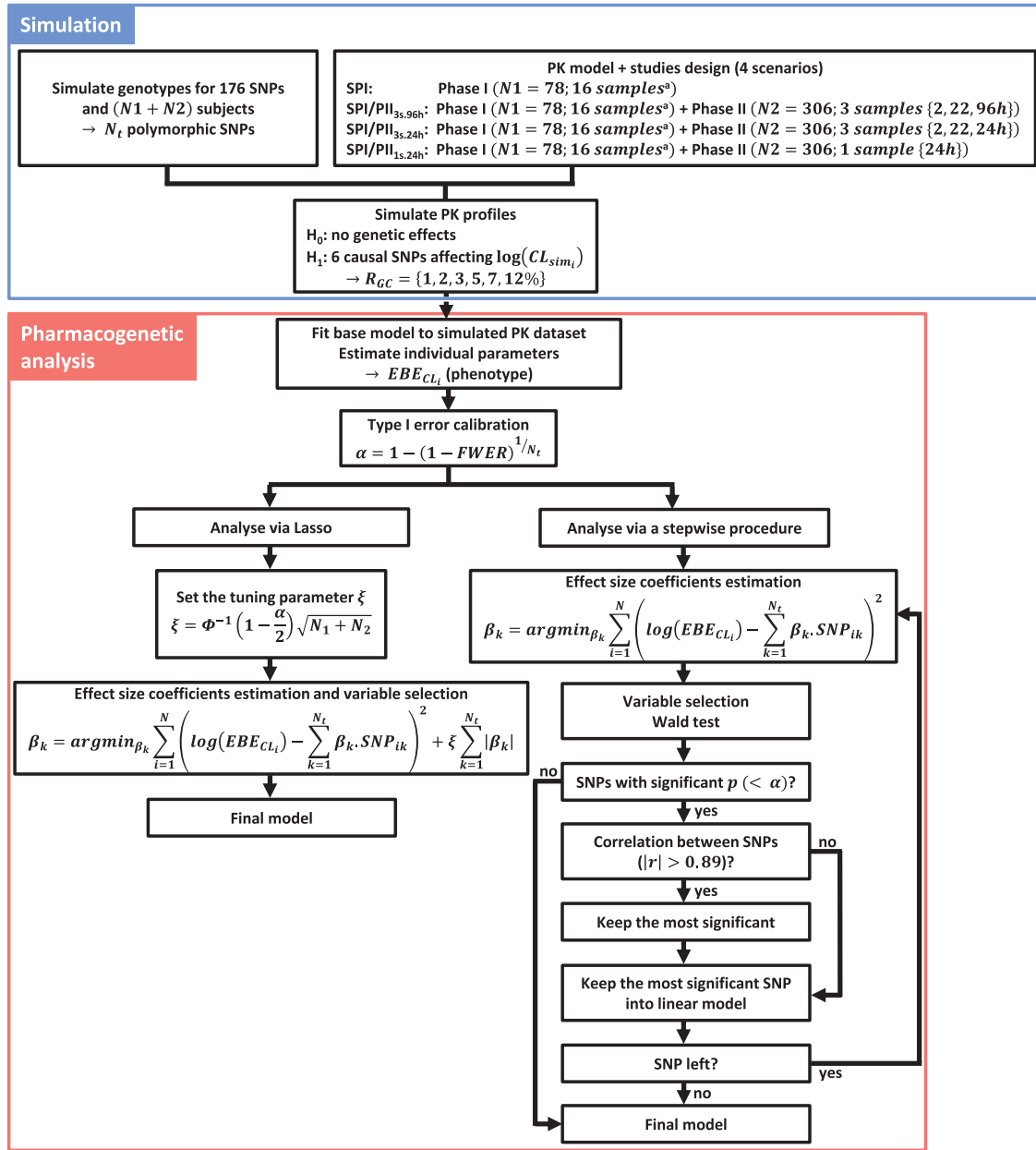


Figure 1 Workflow of the simulation study divided in the simulation (blue box) and analysis part (red box). At 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 16, 24, 48, 72, 96, 120, and 192 hours. CL_{sim_i} , simulated individual clearance ($i=1, \dots, N=N_1+N_2$); EBE_{CL_i} , empirical Bayes estimate of clearance; 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 analyze; p , P value; r , correlation coefficient between variants; R_{GC} , genetic component of the interindividual variability; SNP_{ik} , single nucleotide polymorphism ($k=1, \dots, N_t$); α , type I error per test; β_k , effect size coefficient; ξ , Lasso tuning parameter.

Under H_1 we evaluated the performance of each scenario in terms of true and false positive counts (TP and FP) and rates (TPR, the proportion of TP detected among the causal variants; and FPR, the proportion of FP detected among all potential false associations) for parameter CL, as well as the probability to detect genetic variants. Assuming that SNPs located on genes coded for metabolism enzymes and transporters affect mostly the drug distribu-

tion and elimination, we also applied association tests on Q, the intercompartmental clearance, and V2, the peripheral volume, separately. Any variants associated to Q and V2 were counted as false positives. The central volume V1 was not considered because it had no random effects.

We also evaluated the loss of genetic signal between simulated and estimated individual clearances, comparing slopes b of univariate linear regressions on $\log(CL_{sim_i})$ or

Table 1 Population values (μ) and interindividual variability (ω) for the model parameters of drug S used in the simulation study

Parameters	μ	ω (%)
F ^a	Imax _F	0.8
	D50 _F	41.7
FRAC ^b	E _{max} _{FRAC}	0.45
	D50 _{FRAC}	18.6
Tlag ₁	0.401	35.1
Tk0	1.59	31.6
Tlag ₂	22.7	–
Ka	0.203	–
V1	1520	–
Q	147	89.9
V2	2130	44.2
CL	94.9	25.1
σ_{slope} (%)	20	–

F, bioavailability; FRAC, fraction of dose; Tk0, zero-order absorption duration; Tlag₁, lag time of zero order absorption; Ka, first order absorption constant rate; Tlag₂, lag time of first order absorption; V1, central compartment volume; V2, peripheral compartment volume; Q, intercompartmental clearance; CL, linear elimination clearance.

^aFor doses < 20 units $F=1$, for doses ≥ 20 units $F=1 - \frac{\text{Imax}_F (\text{dose} - 20)}{\text{D50}_F + \text{dose} - 20}$, where dose is the amount administered.

^b $\text{FRAC} = \frac{\text{E}_{\text{maxFRAC}} \text{dose}}{\text{D50}_{\text{FRAC}} + \text{dose}}$

$\log(\text{EBE}_{CL_i})$ for each causal variant. A relative deviation of the genetic signal RD_{signal} was computed as follows:

$$RD_{\text{signal}}(\%) = \frac{b_{\text{EBE}_{CL_i}} - b_{\text{CL}_{\text{sim}_i}}}{b_{\text{CL}_{\text{sim}_i}}} \times 100 \quad (3)$$

RD_{signal} quantifies the departure of the estimated genetic signal ($b_{\text{EBE}_{CL_i}}$) from the one simulated ($b_{\text{CL}_{\text{sim}_i}}$, see details in **Supplementary Material**).

RESULTS

Control of FWER under H₀

The Lasso and stepwise procedure methods both tended to be too conservative, as the FWER was lower than expected in some scenarios (**Table 2**). After an empirical correction by increasing the type I error per SNP α FWER was properly controlled around 20%. This correction was applied in the corresponding simulation under H₁. Previous simulations suggested that this decrease in FWER is influenced by correlations between polymorphisms.⁶

Detection of genetic effects

Under H₁ the TPR (**Figure 2**, top left) was higher in scenarios including phase II data (from 22 to 32%) compared to scenario with only phase I data (SPI, 4%) and was the highest in scenario SPI/II_{3s,96h}. The FPR was lowest (0.2%) in scenario SPI, where a limited number of SNPs was selected, and only slightly higher in scenarios including phase II data, ranging from 0.6 to 0.8% for both methods. Very few TP were effectively detected in scenario SPI (around 44 for both methods) where the number of subjects was limited (N1 = 78) (**Supplementary Tables S2–S3**). By adding more subjects (N2 = 306) to the analysis, the number of TP increased sharply. Scenario SPI/II_{3s,96h} allowed detecting the largest number of TP (380 or more), while in SPI/II_{3s,24h} around 326 TP were detected. In SPI/II_{1s,24h} the number of TP was lower (around 270 TP), but remained much higher than scenario SPI with only phase I data. In the same way, the number of FP increased when including phase II data in the analysis, but to a much lesser extent.

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

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

Shrinkage

Two η -shrinkage estimates were computed using a metric proposed by Bertrand *et al.*⁴ based on estimated variances, with respect to the estimate of $\hat{\omega}^2$ in the dataset; one over the $\hat{\eta}_I$ from phase I subjects and one over the $\hat{\eta}_{II}$ from phase II subjects (**Figure 3**). The η -shrinkage for phase I

Table 2 Empirical estimates of family-wise error rate under H₀ for both association tests

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

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

^aSet of empirical family-wise error rates (FWER) obtained without correction.

^bSet of empirical FWER obtained after correction of type I error per tests.

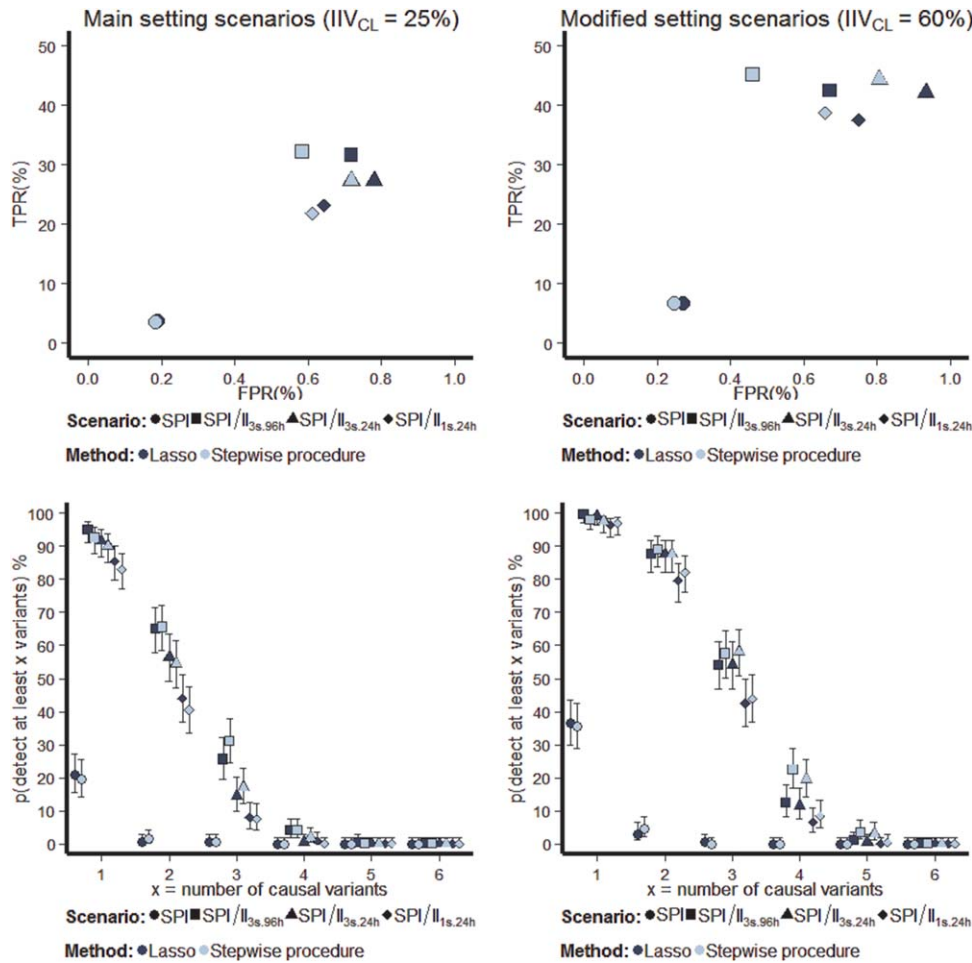


Figure 2 True positive rate (TPR) vs. false positive rate (FPR) under H_1 (top) and probability estimates (points) and 95% confidence interval (bars) to detect at least x variants explaining the interindividual variability of CL ($x=1, \dots, 6$) under H_1 (bottom) for main scenarios simulated with $IIV_{CL}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

subjects was low (median = 23%) thanks to the large number of observations per subject. A large range of η -shrinkage estimates for phase II data was observed across analysis scenarios, but was below 50% in scenario SPI/II_{3s,96h}.

Loss of signal

RD_{signal} was always negative for the six SNPs, indicating that part of the signal was lost during the estimation step (Figure 4, top). This loss was smaller in the scenario with phase I data alone (SPI) than in scenarios combining phase I and phase II data. In each scenario the signal loss was of the same magnitude for the six SNPs, regardless of the value of associated R_{GC} . For phase I data (Figure 4, bottom), the loss was of a constant magnitude across scenarios (median = -30%). For phase II data, in the most informative scenario (SPI/II_{3s,96h}), the loss was of a similar magnitude (median = -41%) than the loss in phase I data, where subjects were extensively sampled. The loss was higher in scenario SPI/II_{3s,24h} (median = -56%), and even

more when only one time was sampled (SPI/II_{1s,24h}, median = -70%).

The signal loss and η -shrinkage values changed accordingly across phase II scenarios, while the probability of detection changed in the opposite direction.

Influence of the phenotype variance

Increasing IIV for the CL parameter to 60% led to a sharp increase in the number of TP (Supplementary Tables S5–S6), resulting in higher TPR and higher probabilities to detect the causal variants (Figure 2, right), compared to when individual CLs were simulated with a moderate IIV. This higher number of TP is explained first and foremost by the increase in simulated effect sizes, which depended on the variance of interindividual random effects on CL due to nongenetic sources (Eq. 2). A second consequence of the larger IIV was that the estimated η -shrinkages became much smaller. Lower η -shrinkages resulted in lower signal losses in all scenarios for phase I and phase II data (Supplementary Figures S7–S8), which again favored a higher probability to detect the genetic effects.

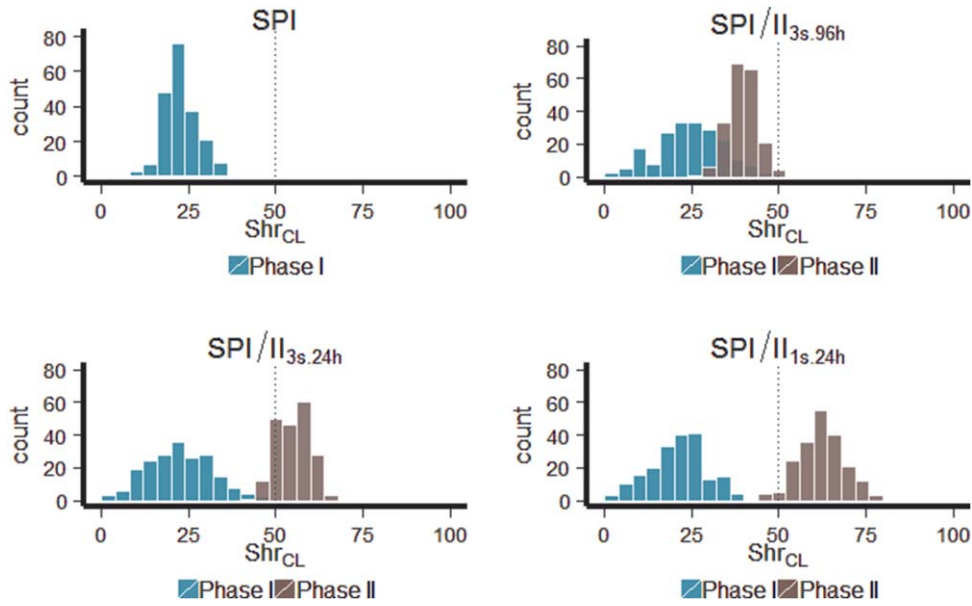


Figure 3 Distribution of the η -shrinkages on clearance for subjects in the phase I dataset (blue) and for subjects in the phase II dataset (brown), for each main scenario simulated under H_0 with $IIV_{CL} = 25\%$.

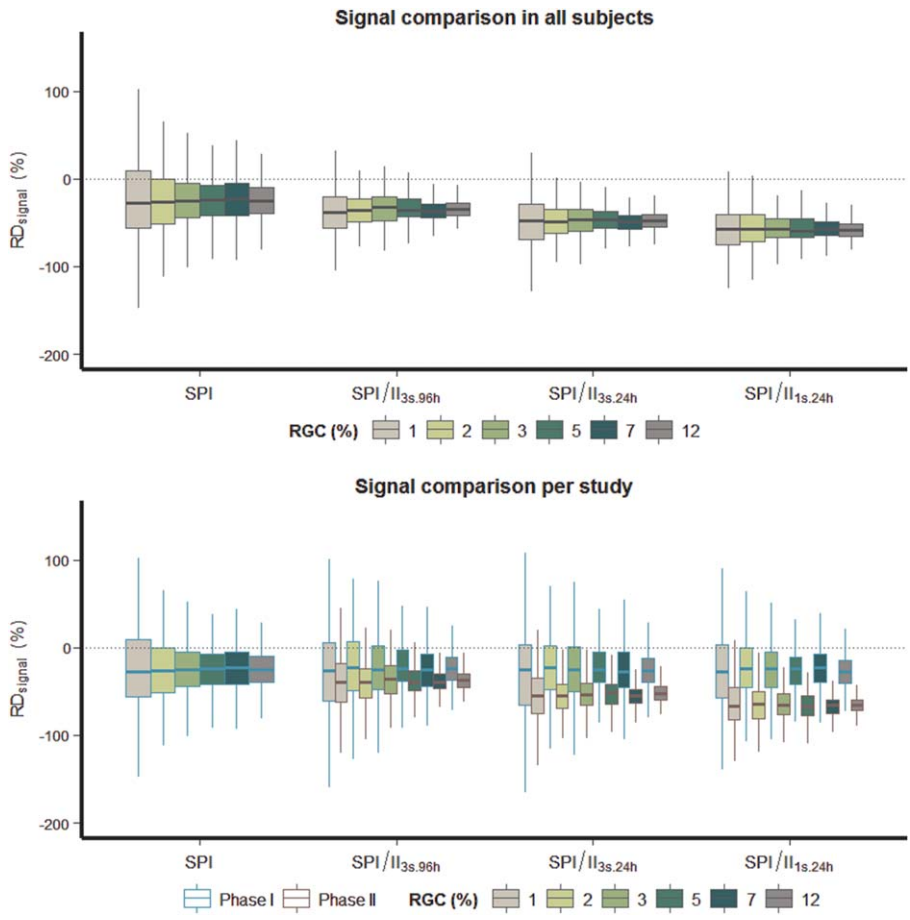


Figure 4 Boxplots showing the loss of the signal for genetic effect in the overall population (top), as well as separately for the phase I data (blue borders) and for the phase II data (brown borders) (bottom). A boxplot is shown separately for each main scenario simulated under H_1 with $IIV_{CL} =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

DISCUSSION

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

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

The effect of sample size can be distinctly observed in our simulations. In the context of phase I studies, where the number of subjects is limited, the probability to detect the genetic effects was low, in line with our previous results.⁶ The combined analysis of phase I and phase II data allowed a marked improvement in this detection probability, irrespective of the phase II study design. By modifying the design of the phase II data, we highlighted a direct link between η -shrinkage, loss of genetic signal, and probability to detect genetic variants. Our results showed that poor PK information due to the phase II study design results in higher η -shrinkage, which increases the loss of genetic signal at the estimation step and translates to a lower probability to detect genetic variants. The dilution of the individual information by adding subjects with sparse designs to subjects with rich designs increases, as expected, the loss of genetic signal. But this is accompanied with a sharp increase in detection power thanks to a larger sample size. η -shrinkage may also modify the EBE–EBE relationship, falsely inducing or masking correlations between model parameters.⁹ This could result in an increased number of false positives associated with other parameters than CL, although in our simulations the number of FP on CL, V2, and Q remained of a similar magnitude across scenarios (**Supplementary Table S3**), showing no systematic effect.

We assume homogeneity of the PK between subjects simulated for the phase I and the phase II study. In practice, healthy volunteers are often included in phase I, while phase II studies focus on patients. A difference in typical values, for example of CL, between the two populations should not impact the detection power by combination of data, as the association tests use the phenotype variance, provided that the genetic effect is the same and that the model accounts for the systematic difference between clearances. It is more difficult to predict what would happen if the variability of clearance is different in the two popula-

tions, as the magnitude of the shrinkage in each subpopulation could affect the signal detection. When the assumption that the two populations are similar breaks down, we would suggest instead to combine rich and sparse data within the phase II study. Pharmacogenetic studies including a large number of subjects combining sparse and rich designs have already been published,^{20,21} showing that the combination of different sampling designs is feasible within the same study to assure more homogeneity.

Situations where pharmacogenetic analyses in PK studies are recommended are described by health authorities.²² In our work, we simulated a blinded pharmacogenetic analysis, exploring a large number of genetic markers. In real applications, other considerations than the statistical significance of genetic variants such as their physiological and clinical relevance could be factored in the analysis and its interpretation. Lehr *et al.*²³ proposed in their stepwise procedure to select only significant polymorphisms having a physiologic relevance in the final model, and the same constraint could be integrated in penalized regression approaches. The probability to detect genetic variants could also be increased through the targeted inclusion of subjects for a few polymorphisms of interest, but this approach requires hypotheses on which polymorphisms to test, with a risk to miss important associations. We focused in this work on PK variability, which is a part of the variability in drug response. But the conclusions from the simulation study could be extended to pharmacodynamics. A previous survey indicated that most pharmacogenetic analyses in clinical PK studies used a phenotype estimated by NCA and furthermore included a limited number of subjects (fewer than 50 subjects in two-thirds).⁶ Authorities in fact recommend studying pharmacogenetics in phase I,²² where the number of subjects is limited. Our work shows that such analyses do not have the power to detect polymorphisms efficiently, but can generate hypotheses to assess in later studies. A recent simulation work²⁴ studied the sample size required to detect a binary covariate. They concluded that around 60 subjects combining rich or sparse designs was sufficient to detect the covariate with at least 80% power. Again, our simulations showed that genetic covariates require higher sample sizes because they are highly unbalanced.

In the first series of simulations a moderate IIV on CL (25%) was used, resulting in a low impact of the genetics on PK, since overall 30% of the moderate CL variability was explained by genetic variants. This setting represented a realistic case to challenge the detection of genetic variants through modeling. We also evaluated the same scenarios with a higher IIV for CL, set to 60%. The η -shrinkage was much lower, as a higher IIV downweights the population prior in the combined criterion used to compute EBE. This decrease in CL η -shrinkages resulted in lower signal loss because of the direct relationship between the two. Associated with larger simulated effect sizes, the number of TP and the probability to detect genetic variants increased in these scenarios. The effect of η -shrinkage on the probability to detect genetic effects in these simulations was higher than the one we observed with the main settings,

because the decrease of η -shrinkage was associated with a sharp increase in the number of TP. This shows that our conclusions do not depend on the level of IIV.

This simulation study also confirms the results of our previous work concerning the relative performance of the different association methods.⁶ The penalized regression method Lasso and the stepwise procedure showed a similar probability to detect genetic variants in all scenarios. However, Lasso is a slightly more complex method that requires computing the penalty in a first step before testing the associations. In this work we assessed methods to detect genetic effects on EBE, after an initial fit. An algorithm proposed by Lehr *et al.*²³ uses univariate regressions to select variants to test in the PK model through LRT. This approach is easy to implement but runtimes depend on the number of iterations leading to the full covariates model. An alternative is to use an integrated approach where effect sizes are estimated and significant variants selected using a penalized regression in the same step¹⁷; this showed similar performance as the stepwise procedure proposed by Lehr *et al.*, but with longer computing times.¹⁷ The results for the two other penalized regression methods tested in the previous work, ridge regression and Hyper-Lasso, were similar (**Supplementary Tables S7–S9, Figure S9**). None of the methods detected the six SNPs simultaneously, as three of the polymorphisms only explained 1 to 3% of the clearance variability, making them difficult to detect. Because association methods relate the polymorphisms to the phenotype variance, we fixed the variance explained by the causal variants (through the parameter R_{GC}) and computed the effect sizes as a function of their allelic frequencies. For a given R_{GC} an infrequent polymorphism was therefore associated with higher effect sizes. This reflects that a clinically relevant polymorphism (with a high impact on PK), present in few subjects because of its low frequency, will explain a limited proportion of the phenotype variance. Detecting such polymorphisms is crucial to identify subpopulations at risk but require much larger sample sizes, as in genome-wide studies.²⁵ As an example, the rs3918290 polymorphism from gene *DPYD* has a frequency lower than 1%, but results in a deficient dihydropyrimidine dehydrogenase activity associated with a 40% decrease of the maximum conversion capacity of the chemotherapeutic drug 5-fluorouracil,²⁶ resulting in severe toxicities.

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

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

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

In any case, it is essential that the sampling protocol, although limited, is as informative as possible to minimize the estimation error and shrinkage in individual parameters estimation. The detection of genetic polymorphisms could highly benefit from the use of larger sample sizes through combined analysis and optimized design.^{18,27}

In conclusion, this work confirmed the very limited likelihood that weak genetic effects can be detected in a typical phase I study, due to the small sample size. Such studies have to be considered only as hypothesis-generating.²⁸ On the basis of our results in terms of detection probability when analyzing together data from phase I and phase II studies, we claim that phase II is the best moment to identify the impact of genetic variants on drug response. It would be less efficient to start the study of the pharmacogenetics of a new drug in phase III trials or in postmarketing, because these take place too late in drug development²⁹ and the new treatment could be administered in nonresponders or expose subjects to high toxicities. Furthermore, genetic subpopulations can be better targeted and potentially some subjects excluded from the study to increase the efficacy and reduce the risk of toxicity of the drug in these phase III studies.

Acknowledgments. Adrien Tessier received funding from Institut de Recherches Internationales Servier. The authors thank Laurent Ripoll and Bernard Walther from Institut de Recherches Internationales Servier for their advice in pharmacogenetics. The authors also thank Hervé Le Nagard for the use of the computer cluster services hosted on the “Centre de Biomodélisation UMR1137.” This work received the Lewis Sheiner Student Award from the Population Approach Group in Europe (PAGE) committee and was presented as oral communication in the Lewis Sheiner Student Session at the 24th annual PAGE meeting: Tessier, A. *et al.* Modelling pharmacogenetic data in population studies during drug development. *PAGE Abstract* #3333 <<http://www.page-meeting.org/?abstract=3333>> (2015).

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

Author Contributions. A.T., J.B., M.C., and E.C. wrote the article; A.T., J.B., M.C., and E.C. designed the research; A.T., J.B., M.C., and E.C. performed the research; A.T., J.B., M.C., and E.C. analyzed the data.

1. Aarons, L. Population pharmacokinetics: theory and practice. *Br. J. Clin. Pharmacol.* **32**, 669–670 (1991).
2. Motulsky, A.G. Drugs and genes. *Ann. Intern. Med.* **70**, 1269–1272 (1969).
3. Guo, Y., Shafer, S., Weller, P., Usuka, J. & Peltz, G. Pharmacogenomics and drug development. *Pharmacogenomics* **6**, 857–864 (2005).
4. Bertrand, J., Comets, E., Laffont, C.M., Chenel, M. & Mentré, F. Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm. *J. Pharmacokinet. Pharmacodyn.* **36**, 317–339 (2009).
5. Bertrand, J., Comets, E., Chenel, M. & Mentré, F. Some alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models. *Biometrics* **68**, 146–155 (2012).
6. Tessier, A., Bertrand, J., Chenel, M. & Comets, E. Comparison of nonlinear mixed effects models and noncompartmental approaches in detecting pharmacogenetic covariates. *AAPS J.* **17**, 597–608 (2015).
7. Sheiner, L.B., Rosenberg, B. & Melmon, K.L. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput. Biomed. Res. Int. J.* **5**, 411–459 (1972).
8. Rowland, M. & Tozer, T.N. *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications* (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2011).
9. Savic, R.M. & Karlsson, M.O. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *AAPS J.* **11**, 558–569 (2009).
10. Mandema, J.W., Verotta, D. & Sheiner, L.B. Building population pharmacokinetic—pharmacodynamic models. I. Models for covariate effects. *J. Pharmacokinet. Biopharm.* **20**, 511–528 (1992).
11. Tessier, A., Bertrand, J., Fouliard, S., Comets, E. & Chenel, M. High-throughput genetic screening and pharmacokinetic population modeling in drug development. (2013). Abstract 2836. <www.page-meeting.org/?abstract=2836>.
12. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
13. Su, Z., Marchini, J. & Donnelly, P. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics* **27**, 2304–2305 (2011).
14. Bertrand, J. & Balding, D.J. Multiple single nucleotide polymorphism analysis using penalized regression in nonlinear mixed-effect pharmacokinetic models. *Pharmacogenet. Genomics* **23**, 167–174 (2013).
15. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1994).
16. Hoggart, C.J., Whittaker, J.C., De Iorio, M. & Balding, D.J. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet.* **4**, e1000130 (2008).
17. Bertrand, J., De Iorio, M. & Balding, D.J. Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics. *Pharmacogenet. Genomics* **25**, 231–238 (2015).
18. Bazzoli, C., Retout, S. & Mentré, F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Comput. Methods Programs Biomed.* **98**, 55–65 (2010).
19. Combes, F., Retout, S., Frey, N. & Mentré, F. Powers of the likelihood ratio test and the correlation test using empirical Bayes estimates for various shrinkages in population pharmacokinetics. *CPT Pharmacomet. Syst. Pharmacol.* **3**, 1–9 (2014).
20. Chou, M. *et al.* Population pharmacokinetic-pharmacogenetic study of nevirapine in HIV-infected Cambodian patients. *Antimicrob. Agents Chemother.* **54**, 4432–4439 (2010).
21. Bertrand, J. *et al.* Multiple genetic variants predict steady-state nevirapine clearance in HIV-infected Cambodians. *Pharmacogenet. Genomics* **22**, 868–876 (2012).
22. EMA *Guideline on the Use of Pharmacogenetic Methodologies in the Pharmacokinetic Evaluation of Medicinal Products*. (2012).
23. Lehr, T., Schaefer, H.-G. & Staab, A. Integration of high-throughput genotyping data into pharmacometric analyses using nonlinear mixed effects modeling. *Pharmacogenet. Genomics* **20**, 442–450 (2010).
24. Klopogge, F., Simpson, J.A., Day, N.P.J., White, N.J. & Tarning, J. Statistical power calculations for mixed pharmacokinetic study designs using a population approach. *AAPS J.* **16**, 1110–1118 (2014).
25. Takeuchi, F. *et al.* A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433 (2009).
26. van Kuilenburg, A. B. P. *et al.* Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c.1905 + 1G>A mutation in DPYD by means of a Bayesian limited sampling strategy. *Clin. Pharmacokinet.* **51**, 163–174 (2012).
27. Combes, F.P., Retout, S., Frey, N. & Mentré, F. Prediction of shrinkage of individual parameters using the bayesian information matrix in non-linear mixed effect models with evaluation in pharmacokinetics. *Pharm. Res.* **30**, 2355–2367 (2013).
28. Bromley, C.M. *et al.* Designing pharmacogenetic projects in industry: practical design perspectives from the Industry Pharmacogenomics Working Group. *Pharmacogenomics J.* **9**, 14–22 (2009).
29. O'Donnell, P.H. & Stadler, W.M. Pharmacogenomics in early-phase oncology clinical trials: is there a sweet spot in phase II? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 2809–2816 (2012).

© 2016 The Authors CPT: Pharmacometrics & Systems Pharmacology published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (<http://www.wileyonlinelibrary.com/psp4>)