Author/s:
Bertrand, J;De Iorio, M;Balding, DJ

Title:
Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics

Date:
2015-04-22

Citation:
Bertrand, J., De Iorio, M. & Balding, D. J. (2015). Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics. Pharmacogenetics and Genomics, 25 (5), pp.231-238. https://doi.org/10.1097/FPC.0000000000000127.

Persistent Link:
https://hdl.handle.net/11343/262040

Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics

Julie Bertrand[a], Maria De Iorio[b] and David J. Balding[a]

Context **In a previous work, we have shown that penalized regression approaches can allow many genetic variants to be incorporated into sophisticated pharmacokinetic (PK) models in a way that is both computationally and statistically efficient. The phenotypes were the individual model parameter estimates, obtained a posteriori of the model fit and known to be sensitive to the study design.**

Objective **The aim of this study was to propose an integrated approach in which genetic effect sizes are estimated simultaneously with the PK model parameters, which should improve the estimate precision and reduce sensitivity to study design.**

Methods **A total of 200 data sets were simulated under the null and each of the following three alternative scenarios: (i) a phase II study with $N = 300$ participants and $n = 6$ sampling times, wherein six unobserved causal variants affect the drug elimination clearance; (ii) the addition of participants with a residual concentration collected in clinical routine ($N = 300$, $n = 6$ plus $N = 700$, $n = 1$); and (iii) a phase II study ($N = 300$, $n = 6$) in which four unobserved causal variants affect two different model parameters.**

Results **In all scenarios the integrated approach detected fewer false positives. In scenario (i), true-positive rates were low and the stepwise procedure outperformed the integrated approach. In scenario (ii), approaches performed similarly and rates were higher. In scenario (iii), the integrated approach outperformed the stepwise procedure.**

Conclusion **A PK phase II study with $N = 300$ lacks the power to detect genetic effects on PK using genetic arrays. Our approach can simultaneously analyse phase II and clinical routine data and identify when genetic variants affect multiple PK parameters.** ***Pharmacogenetics and Genomics*** **25:231–238**

Pharmacogenetics and Genomics 2015, **25**:231–238

Keywords: high throughput analysis, nonlinear mixed-effect models, penalized regression, pharmacogenetics

[a]University College London Genetics Institute and [b]University College London, Statistical Science Department, London, UK

Correspondence to Julie Bertrand, PhD, University College London, Darwin Building, room 208, Gower St, WC1E 6BT London
Tel: +44 (0)20 7679 2189; e-mail: j.bertrand@ucl.ac.uk

Received 4 November 2014 **Accepted** 27 January 2015

Introduction

Some established drugs can show high interindividual variability. Adverse drug reactions are an extreme consequence of this variability, which account for 6.5% of hospital admissions [1]. Ingelman-Sundberg and Gomez [2] estimated that 10–20% of such drug reactions could be due to genetic factors. Moreover, *CYP2C9* and *VKORC1* polymorphisms have been shown to explain up to 40% of the variability in the response to warfarin [3]. Furthermore, regulation authorities now recommend to perform genetic association studies when a polymorphic transporter has been shown to play a major role in the pharmacokinetics (PK) of a drug and/or there is marked interindividual variability or inexplicable outliers reported in phase I or subsequent studies [4]. Likewise, several initiatives have been created to identify and control for genetic sources of variability in drug response, such as the recent P4Medecine [5].

However, large randomized control trials such as the European EU-PACT [6] and the American COAG [7] can raise inconsistent results, partly because of genetic heterogeneity but also because they were not using the same genetic-based dosing algorithm. Indeed, COAG used no loading dose, so decisions were made based on concentrations not yet at steady-state, whereas EU-PACT used a dosing algorithm based on a nonlinear mixed-effect (NLME) model of warfarin PK and pharmacodynamics. One possible explanation is that doses predicted from such a model were better tailored to patients, which explains the EU-PACT success in assessing the benefit of genetic-guided dosing when COAG showed no significant differences between the genetic-guided and standard care treatment arms.

This example provides an incentive to use NLME models in the exploration of further genetic associations with drug response.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.pharmacogeneticsandgenomics.com*).

1744-6872

DOI: 10.1097/FPC.0000000000000127

Indeed a few years ago, Lehr *et al.* [8] proposed a simple but computationally intensive stepwise regression approach to perform high throughput genetic associations using NLME models. In a previous work, we compared their procedure to an approach combining penalized regression and NLME models. We showed through a realistic simulation study that our approach was computationally and statistically efficient to analyse a large set of single nucleotide polymorphisms (SNPs) [9]. The phenotypes, in both our approach and the stepwise procedure, are the empirical Bayes estimates of the subject-specific model parameters. However, the latter are obtained a posteriori of the model fit and are known to be sensitive to the study design.

Therefore, we propose, here, integrating penalized regression and NLME models, in which the selection of SNPs is simultaneous with the estimation of the model parameters. We assessed this integrated approach through a simulation study based on the drug metabolizing enzymes and transporters (DMET) Chip [10] for the SNP set and a moderately complex dose-concentration model with six parameters for the drug response. We compared its performance with the stepwise procedure on three scenarios differing in terms of SNP effect on PK model parameter or study design.

Methods

Nonlinear regression

To describe the time course of the drug concentration in each participant after intake, we used a compartmental representation of the human body. Each compartment represented a biological unit (such as a group of organs with similar physicochemical properties) where the drug distributes in an homogenous fashion. The absorption and elimination are described as first-order input and output functions from a central unit (e.g. the systemic circulation), and distribution to less perfused organs is represented by additional compartments communicating with the central unit through first-order influx and efflux. This dynamic system can be translated in a mathematical function nonlinear in its parameters, the absorption (k_a) and elimination (k) rates, the volumes of distribution of the central (V_c) and peripheral (V_p) compartments and the influx and efflux rates to a peripheral compartment (k_{cp} and k_{pc}). The volumes, the elimination rate from the central unit and the influx and efflux rates define the time required to clear the body from the drug, as well as the time to reach an equilibrium for a given dosage regimen. Pharmacologists often combine those in the elimination ($Cl = k/V_c$) and intercompartmental ($Q = k_{cp}/V_c = k_{pc}/V_p$) clearances, which express the volumes of drug cleared per time unit.

Mixed-effect framework

Each of the participants had been sampled following the drug intake, at a number of timepoints that can differ across individuals. The mixed-effect framework enables a simultaneous model fit, borrowing information from participants with several sampling times for the analysis of participants with fewer profiles.

Let y_i be the vector of concentrations of participant $i = 1, ..., N$, sampled at the n_i timepoints in vector t_i after intake of dose D_i. f is a function that may be nonlinear in ϕ_i, its parameter vector, and which specifies the concentration of subject i:

$$\begin{aligned} y_i &= f(\phi_i;\, D_i,\, t_i) + \varepsilon_i \\ \phi_i &= \mu + \eta_i \\ \eta_i &\sim N(0,\, \Omega) \\ \varepsilon_i &\sim N(0,\, I_{ni}\sigma^2), \end{aligned} \tag{1}$$

where η_i and ε_i are the vectors of random effects and residual errors for participant i, and $\theta = (\mu, \Omega, \sigma)$ denotes population-level model parameters for the vector of fixed effects μ, the matrix of interindividual variance-covariance Ω and the residual error variance σ, respectively. To ensure positive concentrations, concentrations can be modelled on the log-scale or have a proportional residual variance at the likelihood level.

In this work, we considered the stochastic approximation of the expectation maximization (SAEM) algorithm [11] to obtain the model parameter estimates (see Appendix 1, Supplemental digital content 1, *http://links.lww.com/FPC/A810*).

Genetic association analyses

The influence of SNPs on the PK of each individual is incorporated in the NLME model through the addition of a SNP-PK parameter relationship matrix C_i to Eq. (1):

$$\phi_i = C_i\mu + \eta_i, \tag{2}$$

where C_i is a block-diagonal matrix specific to each individual i that contains a block of SNPs under study per PK model parameter. μ is then a stacked vector of the PK parameter-specific regression coefficients. It contains successively for each PK parameter (such as Cl and V_c), its intercept and as many β as there are SNPs for that parameter. η_i captures for each PK parameter the departure of the individual i from the population value. Therefore, mixed effects enable to explore SNP effect on separate PK parameters and moreover to quantify the part of the interindividual variability explained. However, when the number of SNPs is large or even exceeds the number of participants, maximum likelihood estimation can no longer be performed.

The most common method for a large number of SNPs is the stepwise procedure. First, an NLME PK model is fitted to the data to obtain estimates of $\hat{\theta}$. Thereafter, for each participant i, an empirical Bayes estimate of each PK parameter is derived using these $\hat{\theta}$ and for each PK model

parameter d the vector $\hat{\phi}_d$ of size N is regressed on each candidate SNP in a separate linear model. In a second step, for each SNP that passes the screening step, an NLME PK model including the SNP-PK parameter relationship as in (2) is fitted to the data and the best model is selected using a likelihood ratio test (LRT). From the latter, new vectors $\hat{\eta}_d$ are derived and regressed on each candidate SNP as in step 1.

The procedure will continue until no more SNP effect is found on PK parameters in steps 1 or 2. In genetic association analyses, a certain amount of correlation due to linkage disequilibrium is expected among the SNPs.

Therefore, Lehr *et al.* [8] proposed that, if after the first step there are multiple significant SNPs with r^2 of 0.8 or more, only the most significant is advanced to the next step of the analysis.

The integrated approach, that we propose, is an automatic variable selection method in which we use a penalized regression in the maximization step of μ at each iteration k of the SAEM algorithm, considering each PK parameter d independently:

$$\mu_{dk+1} = \text{Argmin}_{\mu_d} \sum_{i=1}^{N} (S_{d,i,k} - C_{di}\mu_{dk})^2 + P(\mu_{dk}), \qquad (3)$$

where $S_{d,i,k}$ is the sufficient statistic derived from the stochastic approximation of the model likelihood for updating μ_d (see details in Appendix 1, Supplemental digital content 1, *http://links.lww.com/FPC/A810*), C_{di} is the block of SNPs under study for PK model parameter d and P is the penalty term. Penalized regression shrinks effect size coefficients of SNPs towards zero. We considered two types of penalized regression: the Lasso and HyperLasso (HLasso), a generalization of the Lasso. The Lasso requires only one penalty term ξ; the higher this term, the stronger the penalty and the fewer the SNPs that enter the model:

$$P(\mu_{dk}) = \xi \sum_{v} |\beta_{dvk}|,$$

where ξ is a regularization parameter and β_{dvk} is the effect size of SNP v on PK parameter d at iteration k. In contrast, HLasso requires two penalty terms: the shape γ and the scale λ:

$$P(\mu_{dk}) = -\sum_{v} \left(\frac{\beta_{dvk}^2}{4\gamma^2} + \log D_{(-2\lambda-1)} \left(\frac{|\beta_{dvk}|}{\gamma} \right) \right),$$

where D is the parabolic cylinder function. When both γ and λ tend towards infinity, HLasso converges to the Lasso. Here, we set γ to 1, as it has been found to perform well [12].

For the screening step of the stepwise procedure, the per-test type I error α was set to achieve a target family wise error rate (FWER) using the Sidak correction: $\text{FWER} = 1-(1-\alpha)^{N_{\text{par}}}$, where N_{par} is the total number of SNP tested on all PK model parameters. For the ensuing LRT step, the per-test type I error was set to 1%.

For the integrated approach, we set ξ and λ using an asymptotic approximation [13], which ensures the same target α as the screening step of the stepwise procedure:

$$g'(\beta_{dvk} = 0^+) = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{N}{\sigma^{*2}}}, \qquad (4)$$

where $g'(\beta_{dvk} = 0^+)$ is the first derivative of the log-prior distribution specific to the penalized regression, Φ^{-1} is the inverse normal distribution function and σ^* the standard deviation of $s_{d,i,k}$ (see details in Appendix 2 Supplemental digital content 2, *http://links.lww.com/FPC/A811*).

Simulation study

The genotypic data for each patient were simulated using Hapgen (Wellcome Trust Centre for Human Genetics, Oxford, UK) [14] for 1227 genetic variants from the DMET Chip [10] located on 171 genes involved in drug metabolism spanning the 22 autosomes and chromosome X. The median (range) interval covered by the SNPs is 29 (0–804) kb per gene, with 6 (1–56) SNPs per gene. Hapgen resamples known haplotypes and can thereby produce samples with patterns of linkage disequilibrium mimicking those in real data. The reference haplotype set used in our simulations comes from HapMap release 21 for the White population. Haplotypes were paired randomly, thus imposing Hardy–Weinberg equilibrium. SNPs were coded 0, 1, or 2 expressing the allele dosage.

The concentration data were simulated from a two-compartment model. The parameter values were inspired from a real case study with non-negligible interindividual coefficients of variation from 30 to 70% and heteroscedastic variance for the residual errors [9].

We simulated 200 data sets per hypothesis (null hypothesis of no SNP effect on PK model parameters, H_0 and an alternative hypothesis of existing SNP effects on PK model parameters, H_1) for three different scenarios with different study designs or SNP effects under H_1.

The first scenario corresponds to a large phase II study with $N = 300$ participants and $n = 6$ sampling times allocated to ensure a reasonable precision of parameter estimates for the basic model using the optimization algorithm PFIM [15]. Under H_1, six unobserved causal variants decreased *Cl*, explaining 1, 2, 3, 5, 7 and 12% of its interindividual variability, respectively (see Appendix 3, Supplemental digital content 3, *http://links.lww.com/FPC/A812*).

The second scenario corresponds to the combination of this phase II study ($N=300$, $n=6$) with residual concentrations collected from several participants ($N=700$, $n=1$).

The last scenario considers four unobserved causal variants affecting two different model parameters under H_1. The first and the third causal variants decrease *Cl*, whereas the second and the fourth decrease V_c. Moreover, the first causal variant is correlated with the second one and the third causal variant with the fourth ($r^2 \geq 0.5$). Each causal variant explains 15% of the inter-individual variability of its associated parameter (see Appendix 3, Supplemental digital content 3, *http://links.lww.com/FPC/A812*). The study design is the same as in scenario 1 (i.e. $N=300$, $n=6$).

Evaluation

The DMET Chip mainly contains genes coding for proteins involved in distribution or elimination processes. Thus, in our model we explored SNP effect on *Cl*, *Q* and V_c, which describe these processes. *Vp* is not considered because it has no interindividual variability.

We specified an overall FWER of 20%, more liberal than usual, because this allows better power comparisons for lower-effect SNPs [9].

To be realistic, the causal variants are not observed (i.e. not among the tested SNPs); therefore, a true positive (TP) was defined as any significant SNP with an r^2 of 0.05 or more with a causal variant. For both TP and false positive (FP) evaluation, sets of SNPs, with each pair in the set having an r^2 of 0.8 or more, were considered as one signal. Scenarios in which the six causal variants affect *Cl*, any signal for *Q* or V_c is a FP, whereas in scenario 3 in which two causal variants affect *Cl* and two causal variants affect V_c, any signal on *Q* is a FP.

Counts of TP per data set were compared across methods using the Friedman test [16] and pairwise Wilcoxon signed-rank tests using a Holm correction [17] for multiple testing.

Furthermore, to explore the impact of the study design on the performances of the different methods, we calculated the median (range) shrinkage over the 200 data sets (see Appendix 3, Supplemental digital content 3, *http://links.lww.com/FPC/A812*). This metric quantifies how much empirical Bayes estimates of individuals with little information (i.e. few sampling times) are moved towards the fixed effect (i.e. typical value). A spurious covariate association can result from a large shrinkage – for example, greater than 0.5 [18].

Moreover, we assessed how our modification of the SAEM algorithm impacts the accuracy and precision of its estimates under H_0 by calculating relative estimation errors, relative and relative root mean square errors for each model parameter (see Appendix 3, Supplemental digital content 3, *http://links.lww.com/FPC/A812*).

We used the SAEM algorithm implementation in the SAEMIX R package [19]. Initial conditions were set to the simulated values, as our purpose was not to challenge the estimation algorithm itself. In the expectation phase, the number of iterations was set to 300 for the stochastic phase and to 100 for the cooling phase. The number of Monte Carlo Markov chains for simulation of the individual parameters in the E-step was set to 5. The marginal log likelihood to be used for the LRT in the stepwise procedure was calculated using a first-order linearization of the model around the posterior mode of the random effects.

For the penalized regression, we used the HLasso software (C program written by Dr Clive Hoggart, London, UK) [13] to run both the Lasso and HLasso models, keeping the highest mode of 10 iterations, where the order in which the SNPs were updated was permuted to account for the potential multimodality of the posterior density.

All analyses were run on the UCL Legion High Performance Computing Facility on cores with 2GB RAM. Central processing unit times are given as median (range) on the 200 data sets.

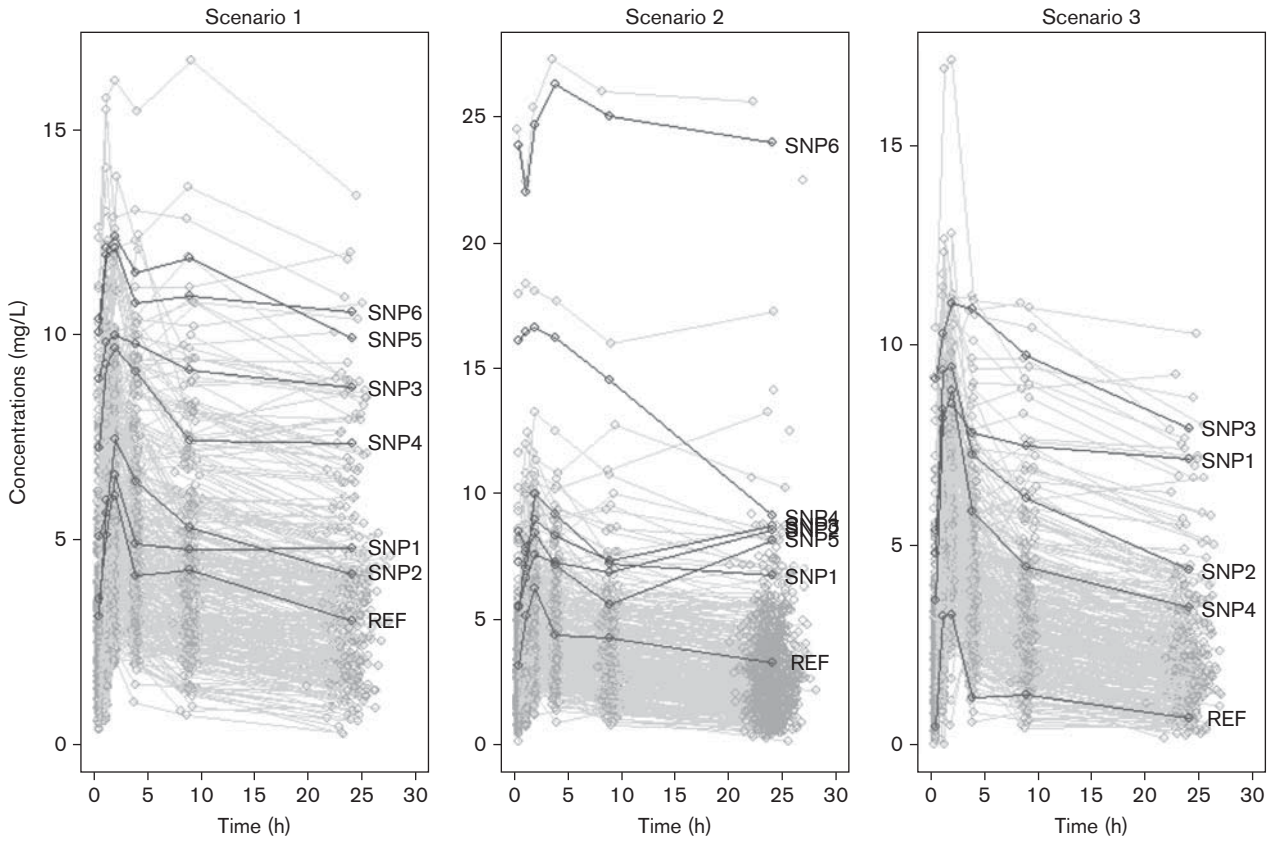
Results

As shown in Fig. 1, we simulated profiles with large interindividual variability. In this figure, we have highlighted one homozygote individual per causal variant. One could be surprised that the PK profiles are not superposed according to the percentage of inter-individual variability explained. This arises both because variance explained is a function of both effect size and population allele frequency, and because a subject PK profile results from the effect of multiple genetic variants.

Table 1 contains the FWER estimates of each method for scenario 1 and scenario 2. Scenarios 1 and 3 share the same study design ($N=300/n=6$) and only differ in the presence of causal variants. All FWER estimates are within the interval 0.145–0.255, which is a 95% prediction interval under H_0.

Figure 2 highlights the trade-off of the three methods on each scenario in terms of TP and FP (see Table 3 in supplementary material for details, Supplemental digital content 4, *http://links.lww.com/FPC/A813*). In scenario 1, the TP rate was 30.5% for the stepwise approach, which was significantly higher than the integrated approach with Lasso (28.3%, $P=0.003$) or HLasso (27.9%, $P=0.002$), but the FP count for the stepwise approach was also significantly higher ($P=0.002$ and 0.001, respectively). In scenario 2, the increase in the sample size despite most of the participants having only one concentration led to a TP rate of 60% with no significant difference across methods, but still a significantly higher FP count for the stepwise approach compared with the integrated approach with Lasso ($P \leq 0.001$ from pairwise

Fig. 1



Spaghetti plots of the simulated concentration versus time profiles for scenario 1: $N = 300/n = 6$, scenario 2: $N = 300/n = 6$ and $N = 700/n = 1$ with six causal variants and scenario 3: $N = 300/n = 6$ with four causal variants. On each plot for each causal variant x, one homozygote individual has been highlighted (SNPx), as well as an individual with zero causal variant allele (REF). SNP, single nucleotide polymorphism.

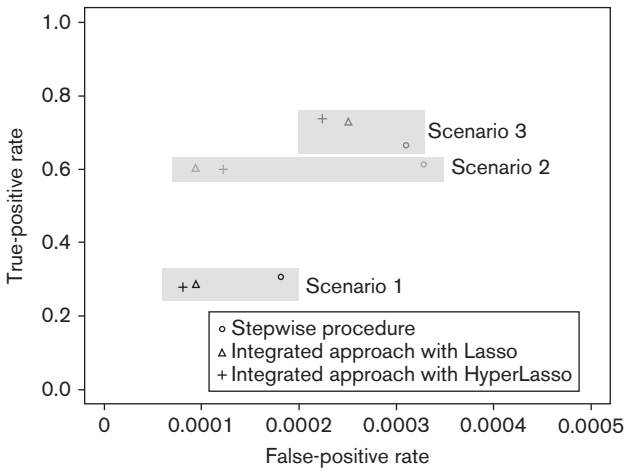
Table 1 Family wise error rate (FWER in %) estimates and their confidence intervals for the three methods on scenario 1 and/or 3 ($N = 300/n = 6$) and scenario 2 ($N = 300/n = 6$ and $N = 700/n = 1$); FWER expected value of 20% with prediction interval on 200 data sets of 14.5–25.5%

Method	Scenario 1 and/or 3	Scenario 2
Stepwise procedure	19.0 [14.2–25.0]	22.5 [17.3–28.7]
Lasso	20.0 [15.0–26.1]	21.0 [15.9–27.2]
HyperLasso	22.5 [17.3–28.7]	21.5 [16.4–27.7]

Wilcoxon signed-rank test) or HLasso ($P \leq 0.001$) was observed. In scenario 3, the stepwise procedure had a TP rate of 66% versus 73 and 74% for the integrated approach with Lasso and HLasso ($P \leq 0.001$ and 0.001, respectively); there were no significant differences in FP count.

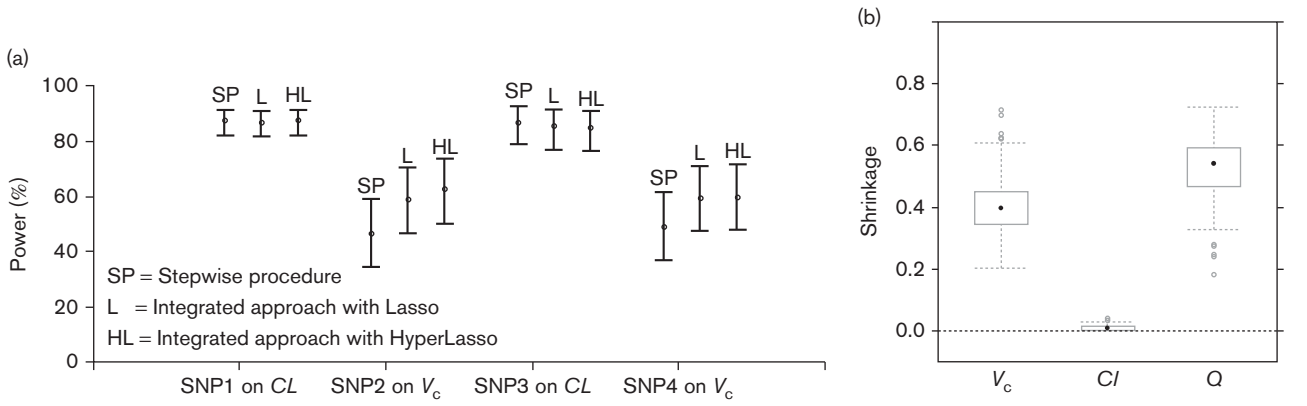
Figure 3 displays the power estimates of all methods to detect each of the four causal variants in scenario 2, as well as the mean and range of shrinkage estimates for V_c, Cl and Q. All methods were about 50% less powerful for

Fig. 2



The mean true versus false-positive rate estimates over the 200 data sets under H_1 for the three approaches on each scenario.

Fig. 3



(a) Mean power estimates to detect each of the four causal variants, with 90% confidence intervals, for the three approaches on scenario 3 and (b) boxplots of the shrinkage estimates from the basic model (i.e. without covariate) for the 200 data sets under H_1 on scenario 3. SNP, single nucleotide polymorphism.

causal variants underlying V_c compared with *Cl*, the drop being greater for the stepwise approach. When considering the shrinkage on V_c and *Cl*, we observed that it was notably larger on V_c, about 40%, than on *Cl*, less than 5%. These results suggest that the shrinkage lowers the power of the methods to detect a genetic effect on a model parameter, the stepwise procedure being more sensitive to its influence than the integrated approaches.

Table 2 contains the mean and range runtime estimates of all methods on all scenarios under both hypotheses. The stepwise procedure runs up to 20 times faster than the integrated approach under H_0, but the gap is reduced in the presence of causal variants and even reverses on large sample size. A reduction of the runtime gain is also observed if more than one parameter is affected by a causal variant.

All methods performed similarly in terms of accuracy and precision of estimation, with mean (range) relative bias of 0.01 and (−0.05 to 0.05) for the fixed effects and −0.001 (−0.110 to 0.140) for the variances, and relative root mean square errors 0.10 (0.01–0.23) for the fixed effects and 0.21 (0.05–0.56) for the variances (see Supplementary Figure 4, Supplemental digital content 5, *http://links.lww.com/FPC/A814*).

Discussion

In this work, we propose an integrated approach to simultaneously test for SNP effect on PK model parameters, and estimate effect sizes, as opposed to a stepwise procedure, which test for SNP effect a posteriori of the NLMEM estimation step.

Because we used Lasso-type penalization for the selection process, the integrated approach tends to be more conservative than the stepwise procedure. Indeed, as seen in scenario 1, the integrated approach detects significantly less FP for a small reduction in power. However, the integrated approach appears more powerful when genetic variants affect multiple PK parameters. Indeed, by accounting for the effect of SNPs on *Cl* during the estimation process instead of regressing each SNP on the empirical Bayes estimates, the integrated approach was less sensitive to the shrinkage on V_c in scenario 3. Of note, we see little difference using Lasso or HLasso probably because of the SNP set not being on the genome-wide scale. Our approach could be extended to genome-wide genotyping data, especially using HLasso. However, in scenario 2 (1200 SNPs and 1000 participants) on call to HLasso requires ∼17 s, whereas with 650k SNPs and 1000 participants, Hoffman *et al.* [20] reported an average HLasso runtime of 52 min. Moreover, this additional runtime would be multiplied

Table 2 Median [range] computing times in hours for the three methods on each scenario under H_0 and H_1

	Scenario 1		Scenario 2		Scenario 3
Method	H_0	H_1	H_0	H_1	H_1
Stepwise procedure	0.05 [0.05–0.24]	0.24 [0.06–1.09]	0.16 [0.11–0.93]	2.31 [0.31–8.53]	0.84 [0.06–2.56]
Lasso	1.04 [0.81–1.48]	1.14 [0.83–1.61]	1.79 [1.40–2.48]	1.90 [1.47–3.73]	1.51 [1.23–9.05]
HyperLasso	1.08 [0.81–1.57]	1.19 [0.84–1.60]	1.91 [1.44–3.14]	1.94 [1.58–4.19]	1.62 [0.79–1.83]

by the number of iterations of the modified-SAEM algorithm.

Therefore, we expect runtimes to increase dramatically.

Phase II study designs are highly variable and the sample size can vary from tens to hundreds [21,22]. Here, $N = 300$ was the largest value we considered and nevertheless found a lack of power to detect multiple realistic and clinically relevant genetic effects on PK. With our approach or using the stepwise procedure, the power to detect at least three causal variants (in scenario 1 and 2) is less than 20%. However, NLMEM can handle few and unbalanced data and thereby can analyse combinations of different clinical studies to increase the power to detect covariate effects [23]; indeed we show, here, that combining a large phase II study with single subject measurements permits a large increase in power for both methods equally. Being capable of the simultaneous analysis of multiple studies with different designs, the approach we propose can make the most out of projects such as tranSMART [24], combining clinical and genomic data collected on open-access and open-source platforms.

One main feature of the integrated approach was that the FWER depends on the penalty term only. For the stepwise procedure, in contrast, it depends on the thresholds at the screening step and the model inclusion step, which have to be chosen based on an asymptotic distribution (as performed here) or using permutations [25]. Moreover, the stepwise procedure can have many variations; the screening step can be performed using an analysis of variance or a linear regression, whereas the model inclusion step can be performed using criteria such as the AIC or the BIC, a LRT (as performed here) or a Wald test. The use of an integrated approach comes nevertheless at a non-negligible cost in terms of runtime. However, the runtime difference with using a stepwise procedure diminishes with increasing sample size.

In conclusion, the integrated approach we propose seems well suited to the analysis of large SNP sets with PK collected across multiple clinical studies. We only considered a hypothetical two-compartment PK model in this simulation study. However, we expect our approach to be effective on more complex drug response models, as the SAEM algorithm has been successfully used on HIV viral dynamics [26] and in the hidden Markov model of daily seizures data [27]. Our results suggest implementation of the integrated approach as an option in the next version of SAEMIX in future studies.

At present, it does not handle nonsolvable ordinary differential equation system, missing SNPs and interoccasion variability, as those features are not yet covered by the saemix R package. In parallel of these developments, we would like to consider alternatives such as Bayesian variable selection methods. Indeed, the flexibility of Bayesian modelling enables fitting complex physiological-based models to heterogenous data [28] but their application in pharmacogenetics, representing an additional layer of complexity, remains to be further explored.

Acknowledgements

The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

Conflicts of interest

There are no conflicts of interest.

References

1 Pirmohamed M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annu Rev Genomics Hum Genet* 2014; **15**:349–370.
2 Ingelman-Sundberg M, Gomez A. The past, present and future of pharmacoepigenomics. *Pharmacogenomics* 2010; **11**:625–627.
3 Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin Pharmacol Ther* 2011; **90**:625–629.
4 EMEA. *EMEA: reflection paper on the use of pharmacogenetics in the pharmacokinetic evaluation of medicinal products.* London 2007.
5 Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *J Intern Med* 2012; **271**:111–121.
6 Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, *et al.* A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med* 2013; **369**:2294–2303.
7 Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, *et al.* A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med* 2013; **369**:2283–2293.
8 Lehr T, Schaefer HG, Staab A. Integration of high-throughput genotyping data into pharmacometric analyses using nonlinear mixed effects modeling. *Pharmacogenet Genomics* 2010; **20**:442–450.
9 Bertrand J, Balding DJ. Multiple single nucleotide polymorphism analysis using penalized regression in nonlinear mixed-effect pharmacokinetic models. *Pharmacogenet Genomics* 2013; **23**:167–174.
10 Daly TM, Dumaual CM, Miao X, Farmen MW, Njau RK, Fu DJ, *et al.* Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. *Clin Chem* 2007; **53**:1222–1230.
11 Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat* 1999; **27**:94–128.
12 Vignal CM, Bansal AT, Balding DJ. Using penalised logistic regression to fine map HLA variants for rheumatoid arthritis. *Ann Hum Genet* 2011; **75**:655–664.
13 Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* 2008; **4**:e1000130.
14 Su Z, Marchini J, Donnelly P. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics* 2011; **27**:2304–2305.
15 Bazzoli C, Retout S, Mentre F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Comput Methods Programs Biomed* 2010; **98**:55–65.
16 Friedman M. A comparison of alternative tests of significance for the problem of m rankings. *Ann Math Stat* 1940; **11**:86–92.
17 Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979; **6**:65–70.
18 Savic RM, Karlsson MO. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *AAPS J* 2009; **11**:558–569.
19 Comets E, Lavenu A, Lavielle M. *SAEMIX, an R version of the SAEM algorithm.* Athens, Greece: Population Approach Group in Europe; 2011.
20 Hoffman GE, Logsdon BA, Mezey JG. PUMA: a unified framework for penalized multiple regression analysis of GWAS data. *PLoS Comput Biol* 2013; **9**:e1003101.

21 Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005; **23**:4450–4457.
22 Brown SR, Gregory WM, Twelves CJ, Buyse M, Collinson F, Parmar M*, et al*. Designing phase II trials in cancer: a systematic review and guidance. *Br J Cancer* 2011; **105**:194–199.
23 Sheiner LB, Steimer JL. Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu Rev Pharmacol Toxicol* 2000; **40**:67–95.
24 Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Jt Summits Transl Sci Proc* 2013; **2013**:6–8.
25 Bertrand J, Comets E, Chenel M, Mentre F. Some alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models. *Biometrics* 2012; **68**:146–155.
26 Chan PL, Jacqmin P, Lavielle M, McFadyen L, Weatherley B. The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *J Pharmacokinet Pharmacodyn* 2011; **38**:41–61.
27 Delattre M, Radojka S, Miller R, Karlsson MO, Lavielle M. Estimation of mixed hidden markov model with SAEM. Application to daily seizures data. American Conference of Pharmacometrics; 2009. Mashantucket, CT, October 4–7.
28 Leil TA, Kasichayanula S, Boulton DW, LaCreta F. Evaluation of 4β-hydroxycholesterol as a clinical biomarker of CYP3A4 drug interactions using a Bayesian mechanism-based pharmacometric model. *CPT Pharmacometrics Syst Pharmacol* 2014; **3**:e120.