Pharmacokinetic parameters estimation using adaptive Bayesian P-splines models

Astrid Jullion¹,*†, Philippe Lambert², Benoit Beck³, and F. Vandenhende¹,⁴

¹Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
²Institut des Sciences Humaines et Sociales, Université de Liège, Belgium
³Axiosis, Belgium
⁴ClinBAY, Belgium

In preclinical and clinical experiments, pharmacokinetic (PK) studies are designed to analyse the evolution of drug concentration in plasma over time i.e. the PK profile. Some PK parameters are estimated in order to summarize the complete drug’s kinetic profile: area under the curve (AUC), maximal concentration (Cmax), time at which the maximal concentration occurs (tmax) and half-life time (t1/2).

Several methods have been proposed to estimate these PK parameters. A first method relies on interpolating between observed concentrations. The interpolation method is often chosen linear. This method is simple and fast. Another method relies on compartmental modelling. In this case, nonlinear methods are used to estimate parameters of a chosen compartmental model. This method provides generally good results. However, if the data are sparse and noisy, two difficulties can arise with this method. The first one is related to the choice of the suitable compartmental model given the small number of data available in preclinical experiment for instance. Second, nonlinear methods can fail to converge. Much work has been done recently to circumvent these problems (J. Pharmacokinet. Pharmacodyn. 2007; 34:229–249, Stat. Comput., to appear, Biometrical J., to appear, ESAIM P&S 2004; 8:115–131).

In this paper, we propose a Bayesian nonparametric model based on P-splines. This method provides good PK parameters estimation, whatever be the number of available observations and the level of noise in the data. Simulations show that the proposed method provides better PK parameters estimations than the interpolation method, both in terms of bias and precision. The Bayesian
nonparametric method provides also better AUC and $t_{1/2}$ estimations than a correctly specified compartmental model, whereas this last method performs better in $t_{max}$ and $C_{max}$ estimations.

We extend the basic model to a hierarchical one that treats the case where we have concentrations from different subjects. We are then able to get individual PK parameter estimations. Finally, with Bayesian methods, we can get easily some uncertainty measures by obtaining credibility sets for each PK parameter. Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: pharmacokinetic parameters estimation; preclinical experiments; pharmacokinetic profiles; Bayesian P-splines; adaptive penalties

1. INTRODUCTION

In drug discovery experiments, pharmacokinetic (PK) studies are designed to assess the systemic exposure of subjects to a compound under investigation. This kind of studies attempts to analyse the evolution of drug concentration in plasma over time i.e. the PK profile. To do so, blood samples are collected at several time points after drug administration and some PK parameters are estimated in order to summarize the complete drug’s kinetic profile. The usual parameters are: area under the curve (AUC), maximal concentration ($C_{max}$), time at which the maximal concentration occurs ($t_{max}$) and half-life time ($t_{1/2}$). These estimated PK parameters are notably used for drug screening to rapidly take the decision of dropping a compound or of keeping on working with promising ones.

There are several types of design in PK studies. In clinical studies or in preclinical studies with large animals, a series of blood samples may be taken from each individual such that the whole PK profile can be characterized on each subject. In this case, we talk about complete design. For small animals, due to ethical considerations, only a limited volume of blood can be collected from each subject. A classical approach is to reduce the sampling frequency per animal and to collect blood from different blocks of animals across time points [1]. The design is then said to be incomplete. In the extreme case where we only have one observation per animal, we talk about destructive design [2]. The scope of this paper includes both complete and incomplete designs.

Several methods exist to estimate PK parameters. A first method, named below the ‘traditional method’, is based on interpolating between observed concentrations. Typically, the interpolation is linear in the ascending phase and log-linear in the descending phase although other interpolation methods have been proposed in the literature such as, for example, splines interpolation [3]. In an incomplete design, the traditional method consists of interpolating at each time point, the means of the observed concentrations and, thus, it does not allow to have individual estimations of PK parameters for each animal.

Compartmental modelling approach is an alternative. With this method, some time and care should be devoted to the choice of a suitable compartmental model. Once the model is chosen, nonlinear estimation methods are used to estimate the parameters of the model. These methods may fail if the data are sparse and noisy. However, some efforts have been recently made to improve the estimation algorithms [4–7].

In this paper, we propose a Bayesian nonparametric method. The idea is to fit individual PK profile using penalized splines. With this method, we get quick results and we do not face the problem of model choice. The proposed method can be easily handled by non-statisticians since it does not require prior guesses for the parameters as all the prior elicitation processes have been automated. Furthermore, some comparisons with
the traditional method will show that our approach provides better estimations of the PK parameters, i.e. smaller bias and higher precision. In comparison with the compartmental models, it provides better estimation when the level of noise is high. Finally, with Bayesian methods, one can easily obtain measures of uncertainty for the estimated PK parameters using credibility sets.

The plan of the paper is as follows. In Section 2, we present the PK experiments as well as a brief reminder of the traditional method. Our proposal is presented in Section 3. Section 4 explains how to estimate PK parameters with the Bayesian non-parametric method and how to get measures of uncertainty. Section 5 gives the results of some simulations the aim of which is to compare the performances of our approach with those of traditional and compartmental methods. Some applications on real data are shown in Section 6. We end this paper with a discussion in Section 7.

2. PK EXPERIMENT

PK studies are aimed at studying the absorption, distribution, metabolism and elimination of a pharmaceutical product. To do so, blood samples are collected at multiple times after dosing.

### 2.1. Incomplete and complete designs

In an incomplete design, subjects are sampled at one of the possible subsets of predefined time points. Typically, the time points are assigned to blocks of subjects in the following way. The population of subjects \( G \) is divided into \( G \) groups \( (g = 1, \ldots, G) \), containing each \( n_g \) individuals. The \( K \) sampling times \( \{t_j : j = 1, \ldots, K\} \) are distributed into these groups so that \( k_g \) blood samples are taken from subjects in group \( g \) at time points \( \{t_i : i \in I_g \subset \{1, \ldots, K\}\} \) with \( \#I_g = k_g \). Sampling times are different between groups. The complete design is a particular case when \( G = 1 \) and \( \#I_g = K \), i.e. each subject is sampled at every time point. In this situation, a PK profile may be straightforwardly estimated for each subject.

### 2.2. Parameter definitions

The most familiar non-compartmental PK parameters are the AUC, \( C_{\text{max}} \), \( t_{\text{max}} \) and \( t_{1/2} \). The \( C_{\text{max}} \) represents the systemic concentration and \( t_{\text{max}} \) is the time needed to achieve \( C_{\text{max}} \). AUC is the area under the PK curve. It measures the extent of systemic exposure. The terminal half-life time, \( t_{1/2} \), is the half-life time associated with the terminal phase, which is the final log-linear portion of the concentration versus time curve, for multicompartmental PK.

### 2.3. Traditional methods

We give here a brief reminder of the most simple and popular way to estimate the different PK parameters non-compartmentally [8,9]. Denote by \( C(t) \) the concentration at the observation time \( t \) and assume that the concentrations are observed at \( K \) different times. To estimate \( C_{\text{max}} \), we use

\[
\hat{C}_{\text{max}} = \max\{C(t) : t \in \{t_1, \ldots, t_K\}\}
\]

\( t_{\text{max}} \) is estimated by

\[
\hat{t}_{\text{max}} = \arg \max\{C(t) : t \in \{t_1, \ldots, t_K\}\}
\]

Several investigators have considered different rules to estimate the AUC [10–13]. The simplest one is the trapezoidal rule given by:

\[
\hat{\text{AUC}} = \frac{1}{5}(C(t_i) + 4C(t_{i+1}) + C(t_i))t_{i+1} - t_i
\]

Methods that use the log-trapezoidal rule instead of the trapezoidal one on the descending portion of the curve were found to improve results with respect to accuracy while not losing out on statistical precision [11–13]. With this method, we estimate the AUC as

\[
\hat{\text{AUC}} = \sum_{i=1}^{K-1} 0.5(C(t_{i+1}) + C(t_i))(t_{i+1} - t_i)
\]

\[
+ \sum_{i=j}^{K-1} (C(t_{i+1}) - C(t_i))(t_{i+1} - t_i)
\]

\[
/\log(C(t_i)/C(t_{i+1}))
\]

where \( t_j \) is the first observation time in the descending portion of the curve.
To estimate $t_{1/2}$, there exist several methods, more or less sophisticated [14–16]. In this paper, we shall consider a simple one where we first estimate the slope $\lambda$ given by the last two observations on the time-log (concentration) scale and, then, we compute

$$i_{1/2} = \log(2)/\lambda$$

In the case of an incomplete design, everything stays the same except that the first step consists in computing the means of the concentrations at each time point. It means that, while we can get individual estimations of PK parameters in a complete design, only mean PK parameters estimations are available in an incomplete one. With an incomplete design, we can get uncertainty measures for the AUC. Indeed, several efforts have been made to get confidence intervals for this PK parameter [6,7,17–20]. For the other parameters, obtaining uncertainty measures is still an issue. However, methods based on bootstrap may be a solution to circumvent this problem [21].

### 3. BAYESIAN P-SPLINES MODEL

In this section, we present Bayesian models to obtain a nonparametric estimation of the individual PK profiles for complete and incomplete design. These models are based on P-splines techniques. However, other techniques exist such as the cubic smoothing splines, for instance. We refer to Hastie and Tibshirani [22] for a complete review of the different smoothing techniques.

#### 3.1. P-splines definition

To obtain a nonparametric fit to a curve, we use penalized B-splines, also named P-splines by Eilers and Marx [23]. A B-spline of degree $q$ consists of $q + 1$ polynomial pieces, each of degree $q$. These polynomial pieces join at $q$ inner knots of the experimental domain. Each B-spline is positive on a domain spanned by $q + 2$ knots and it is zero everywhere else. Figure 1(a) presents a B-splines basis of degree 2 with 20 equidistant knots between 0 and 1.

Let $b(x)$ denote the B-spline basis at $x$ for a given equidistant grid of knots. A fitted curve $\hat{y}$ to data $\{(x_i, y_i)\}$ is a linear combination $\hat{y}(x) = b(x)'\hat{\theta}$ where $\hat{\theta}$ is the estimated vector of B-splines coefficients. Figure 1(b) presents an example of a fitted curve obtained with a linear combination of the B-splines basis presented in Figure 1(a).

When $m$ data points $(x_i, y_i)$ are available, the least-squares estimator $\hat{\theta}$ minimizes the function

$$S = \sum_{i=1}^{m} \{y_i - b(x_i)'\hat{\theta}\}^2$$

The parameters estimates are highly dependent on the number of knots and their location. The fitted curve will show more variation than is justified by the data if we let the number of knots to be relatively large. To make the estimates less sensitive, Eilers and Marx propose to consider a large set of equidistant knots and to introduce a penalty term in the objective function

$$S = \sum_{i=1}^{m} \{y_i - b(x_i)'\theta\}^2 + \lambda \theta' P \theta$$

where $P = D'D$ is the penalty matrix and $D$ the $r$th-order difference matrix, yielding $\theta' P \theta = \sum_k$
where \( \Delta \) is the first-order difference operator. Thus, for \( r = 2 \), we have

\[ D = \begin{bmatrix}
1 & -2 & 1 & 0 & \ldots & 0 \\
0 & 1 & -2 & 1 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 1 & -2 & 1 \\
\end{bmatrix} \]

By adding the term \( \lambda \sum_k (\Delta k \theta_k)^2 \), we add a penalty on finite differences of the coefficients of adjacent B-splines. Parameter \( \lambda \) expresses the weight that we give to this penalty. Large values for \( \lambda \) force the coefficients of adjacent B-splines to be close to each other. It will result in a very smooth fit. In contrast, if \( \lambda \) is small, the penalty plays a small role and the coefficients of adjacent B-splines are allowed to be highly different from each other. It will yield a wiggly fit. \( \lambda \) is usually selected using cross-validation [24] or information criteria [25].

### 3.2. Basic Bayesian P-splines model

In terms of likelihood, the penalty appears as a term that we subtract from the log-likelihood \( l(y; \theta) \). The penalized likelihood function has the following form:

\[
l_{\text{pen}} = l(y; \theta) - \frac{\lambda}{2} \theta'P\theta
\]

We get the same log-posterior in a Bayesian setting with the following model specification [26,27]:

\[
(Y_{i|x}; \theta) \sim \mathcal{N}(\mathbf{b}(x)' \theta, \tau^{-1})
\]

\[
p(\theta|\tau) \propto \exp[-0.5 \tau \theta'P\theta]
\]

\( \tau \) is the roughness penalty parameter and plays the same role as \( \lambda \) in the frequentist setting. The penalty from the frequentist penalized likelihood approach translates, in a Bayesian setting, into a prior distribution for the \( r \)th-order differences of successive B-splines parameters, \( \theta_j \).

### 3.3. Prior specification

There are some hyperparameters for which we have to propose prior distributions. For \( \tau \), the conditional precision of the vector response \( Y_{i|x} \), it is common to take a noninformative prior:

\[
p(\tau) \propto \tau^{-1}
\]

The prior of the roughness penalty parameter \( \tau \) can be conveniently chosen to be the conditional conjugate prior

\[
\tau \sim \mathcal{G}(a, b)
\]

where \( \mathcal{G}(a, b) \) denotes a gamma distribution with mean \( a/b \) and variance \( a/b^2 \). Lang and Brezger [26] have recommended using a large variance by setting \( a \) equal to 1 and \( b \) equal to a small quantity, or \( a = b \) equal to a small quantity.

### 3.4. Three extensions to the basic Bayesian P-splines model

In this part, we propose some extensions of the basic Bayesian P-splines model to improve its fit to PK data [27].

#### 3.4.1. Preliminary remark

In preclinical PK studies, sampling schedules are often specified to have the last two times widely separated [19]. This suggests to specify the Bayesian model for log-transformed data, \( \log(1 + \text{time}) \) and \( \log(1 + \text{concentration}) \). This ensures positive values for the fitted concentrations and better handles the large interval between the last two observations.

#### 3.4.2. First extension: robust prior

Jullion and Lambert [27] have emphasized the sensitivity of the Bayesian fit to the choice of the prior for \( \tau \) in some specific circumstances. In Figure 2, we have fitted the model of Section 3.2 on simulated sparse PK data with several gamma priors for \( \tau \) corresponding to different values for \( a \) and \( b \). We can see the influence of these hyperparameters on the fit in such a setting. When \( a = 1, b = 0.001 \), we even get a flat curve.

To deal with this issue, Jullion and Lambert [27] propose to consider as prior distribution for \( \tau \) a
weighted sum of $M$ gamma distributions with different values for $b$. In this case, $a$ is fixed to 1. This gives the prior

\[
(\tau_j | \mathbf{p}) \sim \sum_{m=1}^{M} \rho_m \mathcal{G}(1, b_m)
\]

\[
\mathbf{p} \sim \mathcal{D}(\mathbf{u})
\]

where \{\(b_1, \ldots, b_M\}\} is a set of prespecified values. For instance, we may consider a grid of 33 values, logarithmically equally spaced between \(10^{-5}\) and \(10^3\). \(\mathcal{D}\) stands for the Dirichlet distribution, and \(\mathbf{u}' = \{u_1, \ldots, u_M\}\) is a set of (small and equal) hyperprior parameters expressing our likely prior ignorance about the optimal choice for $b$. We use, for instance, the value 0.01 for each $u_i$.

### 3.4.3. Second extension: adaptive penalties

We provide even more flexibility to the model by allowing the roughness penalty to change in a progressive way along the $x$-axis. Adapting penalties can be integrated into the previous model as follows:

\[ p(\theta | \tau_j, A) \propto \exp[-0.5 \tau_j \theta' D' A D \theta] \]

\[ \lambda_k \sim \mathcal{G}(\omega, \omega) \quad \text{when } k > r + 1, \lambda_{r+1} = 1 \]

where

\[
\lambda^{(k)} = \prod_{l=r+1}^{k} \lambda_l
\]

\[ A = \text{diag}(\lambda^{(r+1)}, \ldots, \lambda^{(K)}) \]

Instead of having a single penalty $\tau_\lambda$, we now have a penalty parameter $\tau_{\lambda}^{(k)}$ for each $r$th-order difference between successive components of $\theta$. The penalty parameters are obtained sequentially by multiplying the previous one by a gamma random variable with mean 1 and an (arbitrarily large) variance $\omega^{-1}$. That construction yields a progressive evolution of the penalty parameters with $x$ (see Appendix A). For more details, we refer to [27].

### 3.4.4. Third extension: concavity condition

If found necessary, a third extension may be added to the model. Indeed, one could constrain $\theta$ through its prior. A general information about a PK profile with oral dosing is its global shape. As Gibadli expresses in [28], kinetic profiles after oral administration of a drug first show a continuous increase in drug concentration in the blood stream; then, after having reached a peak, drug concentration slowly decreases over time, following a negative exponential elimination curve. We can constrain the estimated profile to have this global shape by imposing a concavity condition on the fitted curve on the log scale (see Figure 3).

If we want to add this extension, we simply have to reject, in the Gibbs sampler (see Section 4), a $\theta$ generated in the unconstrained specification if it does not meet the concavity condition.

In Figure 4, the thick solid line is the true PK profile. The dashed line is estimated with basic Bayesian P-splines model combined with the robust prior and a fixed penalty, whereas the thin solid line is estimated with the model also having the adaptive penalties and the concavity condition. The fit is markedly improved.

The Matlab code used to estimate these curves can be obtained from the authors. Just a few seconds are required to run 1000 iterations using...
this Matlab code on a Pentium IV. To use this model, we advise to take as many knots as the number of observations and to place them in an equidistant way. The method can be automated in order to be easily applied by the end-user.

3.5. Hierarchical model

The above model allows the fit of an individual time curve. In this section, we adapt it to the case where we have several subjects such that we can get an estimate not only of the mean PK profile but also of individual PK profiles for each subject. The model below is then suitable to incomplete designs. Considering that we have $n_s$ subjects, we define the vector of parameters

$$\mathbf{\theta} = [\mathbf{\theta}_0', \mathbf{\theta}_1', \ldots, \mathbf{\theta}_{n_s}']'$$

where the coefficient vector $\mathbf{\theta}_0$ yields the reference PK profile and the coefficient vectors $\mathbf{\theta}_j$ ($j = 1, \ldots, n_s$) enable to correct $\mathbf{\theta}_0$ to get the spline parameters for subject $j$. We denote by $K$ the number of B-splines in the basis. In order to force the above interpretation for $\mathbf{\theta}_0$, we specify for the first and last B-splines coefficients $y_j$ and $y_{jK}$, $j = 1, \ldots, n_s$, a normal prior distribution with mean 0 and variance $\eta^{-1}$:

$$y_j | \eta \sim \mathcal{N}(0, \eta^{-1}) \quad \forall j = 1, \ldots, n_s$$

$$y_{jK} | \eta \sim \mathcal{N}(0, \eta^{-1}) \quad \forall j = 1, \ldots, n_s$$

By combining the previous equations with the usual (improper) smoothness prior:

$$p(\mathbf{\theta}_j | \tau_j) \propto \exp[-0.5 \tau_j \mathbf{\theta}_j'D_j D_j \mathbf{\theta}_j] \quad \forall j = 1, \ldots, n_s$$

we get the following proper prior distribution for $\mathbf{\theta}_j$:

$$\mathbf{\theta}_j | \eta, \tau_j \sim \mathcal{N}(\mathbf{0}, \Sigma_j) \quad \forall j = 1, \ldots, n_s$$

with

$$\Sigma_j^{-1} = \tau_j D_j D_j + \eta \text{diag}(1, 0, \ldots, 0, 1)$$

The detailed model specification is given in Appendix B.

4. POSTERIOR AND PARAMETER ESTIMATION

In this section, we shall explain how to explore the posterior distribution using Markov chain Monte Carlo (MCMC) techniques and, from this, how to derive estimates and credibility sets for the PK parameters in the Bayesian P-splines model.
4.1. Exploring the posterior using MCMC

MCMC technique is a powerful method to generate samples from posterior distributions in a Bayesian framework [29]. The Gibbs sampler [30] is an MCMC sampler by which each component is updated conditionally on the last available updates for the other components.

In our case, for each presented model, all the conditional posterior distributions can be identified. Thus, we can use the Gibbs sampler to generate random samples from the posterior distribution. We give here the conditional posterior distributions for the basic Bayesian P-splines model of Section 3.2:

\[(\theta; \tau_0; y) \sim N(\tau \Sigma \theta; \Sigma \theta)\]

\[(\tau|\text{other}; y) \equiv (\tau|\theta; y) \sim G(0.5n, 0.5(y - B\theta)^\top(y - B\theta))\]

\[(\tau_z|\text{other}; y) \equiv (\tau_z|\theta; y) \sim G(a + 0.5\rho(P), b + 0.5\rho^\top P)\]

where \(\rho(P)\) is the rank of \(P\) and \(B = \{b(x_1), \ldots, b(x_n)\}\) and \(\Sigma^{-1} = \tau B^\top B + \tau_z P\) and ‘other’ generically denotes all the other parameters from the joint distribution.

The conditional posterior distributions for the extended P-splines model and for the hierarchical one are given in Appendix C.

4.2. Estimation of PK parameters

At each iteration of the MCMC sampler, a vector \(\theta\) is generated yielding a chain of vectors \(\{\theta^{(1)}, \ldots, \theta^{(M)}\}\) (see Section 4.1). For each \(\theta^{(m)}\), the predicted concentrations are calculated over a detailed grid of time points and PK parameters are derived as follows. The estimate for \(C_{\text{max}}\) is the maximal predicted concentration on the detailed grid and the estimate for \(t_{\text{max}}\) is the time at which that maximum occurs. An estimate for the AUC can be computed using trapezoid integration based on this grid. To obtain \(t_{1/2}\), even if more sophisticated methods could be used, such as, for instance, by using linear regression [14,15], we simply compute \(t_{1/2}\) as \(\log(2)/\lambda\) where \(\lambda\) is the slope formed by the predicted values at the last two time points.

PK parameters are so obtained for each generated \(\theta^{(m)}\), yielding an MCMC sample of size \(M\) for each PK parameter. The posterior median of these chains is used to estimate the PK parameters and 95% credibility sets are obtained by taking the 2.5 and 97.5 percentiles of the chains.

5. SIMULATIONS

In this part, we perform some simulations to compare the performances of the extended Bayesian P-splines model with the traditional and compartmental methods for complete and incomplete designs. We also compare this method with another smoothing technique: the robust smoothing splines method [22].

For the complete design, we consider that we have one subject sampled at the following time points: 0, 0.25, 0.5, 0.75, 1, 2, 4, 6, 7, 24 h and postdosing. The choice of the design is important when using traditional methods. Indeed, the estimated values for \(C_{\text{max}}\) and \(t_{\text{max}}\) highly depend on the choice of the observed time points.

At each time point \(t_i\), we generate concentration \(y_i\) using a one-compartment model with oral dosing and first-order absorption and elimination:

\[y(t_i) = \mu(t_i)(1 + \sigma c_i)\]

\[\mu(t_i) = \text{Dose exp}(\log(K_e) + \log(K_d)) - \log(Cl)/(K_d - K_e)\times(\exp(-K_e t_i) - \exp(-K_d t_i))\]

where \(\text{Dose} = 10\), \(\log(K_e) = -1\), \(\log(K_d) = -0.45\), \(\log(Cl) = -5.186\) and \(c \sim N(0, 1)\). We consider three different values for \(\sigma\): \(\sigma = 0.1, 0.2\) and 0.4, which correspond to a low, medium and high level of noise in the data. We have generated 500 simulations in each case. An example of generated data sets is shown in Figure 5 where the circles are generated with \(\sigma = 0.1\), the plus with \(\sigma = 0.2\) and the triangles with \(\sigma = 0.4\).

For the incomplete case, we consider a design with two subjects sampled at times (0.25, 1, 5 h),...
two at times (0.5, 2, 7 h) and two at times (0.75, 3, 24 h). We use the same PK model as the one used in the complete design but with an extra inter-subject variability: we generate a different curve $\mu_j(t) (j = 1, \ldots, 6)$ for each subject by adding to $\log(K_e)$ and $\log(Cl)$, a normal perturbation with standard deviation equal to 0.1. For each of the generated curves, we generate an observation at the three time points selected by the design as done previously.

For the compartmental method, the fitted model is a one-compartment model with oral dosing: this is a favourable situation for that modelling strategy as the assumed model corresponds to the data-generating mechanism. To estimate its parameters, we have used the function `nls` implemented in the R software [31], which is based on least-squares estimations. The `gnls` function, which allows to model the variance as a function of the expected response, could also be used. However, we would then get a perfect adequation between the fitted model and the data-generating one, a situation not realistic in practice. A (nonreported) simulation study has been performed to compare the Bayesian method with the `nls` and the `gnls` ones: the same conclusions are obtained using either the `nls` or the `gnls` methods.

For the incomplete case, we have used the function `nlme` in R, where we add a random effect on $\log(K_e)$ and $\log(Cl)$ of the compartmental model. Finally, to estimate the robust smoothing splines model, we have used the function `qsreg` proposed in R (fields package).

Table I. Simulation results for the complete cases: RMSE, ML1 and standard deviation for the Bayesian method (ab), the traditional method ($t$), the compartmental method (nls) and the cubic smoothing splines (sr).

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma = 0.1$</th>
<th></th>
<th>$\sigma = 0.2$</th>
<th></th>
<th>$\sigma = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>ML1</td>
<td>St. dev.</td>
<td>RMSE</td>
<td>ML1</td>
</tr>
<tr>
<td>AUC</td>
<td>ab</td>
<td>5.6</td>
<td>3.8</td>
<td>4.4</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>5.6</td>
<td>4.3</td>
<td>4.0</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>nls</td>
<td>5.5</td>
<td>3.4</td>
<td>4.9</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>sr</td>
<td>5.4</td>
<td>3.6</td>
<td>5.1</td>
<td>10.6</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>ab</td>
<td>8.2</td>
<td>5.9</td>
<td>6.2</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>8.2</td>
<td>5.9</td>
<td>7.6</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>nls</td>
<td>6.1</td>
<td>4.0</td>
<td>5.4</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>sr</td>
<td>7.5</td>
<td>5.5</td>
<td>7.5</td>
<td>12.8</td>
</tr>
<tr>
<td>$t_{\text{max}}$</td>
<td>ab</td>
<td>14.8</td>
<td>2.2</td>
<td>14.1</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>24.3</td>
<td>2.0</td>
<td>23.6</td>
<td>45.5</td>
</tr>
<tr>
<td></td>
<td>nls</td>
<td>7.2</td>
<td>4.6</td>
<td>5.8</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>sr</td>
<td>21.1</td>
<td>10.3</td>
<td>21.0</td>
<td>36.4</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>ab</td>
<td>2.8</td>
<td>2.0</td>
<td>2</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>3.0</td>
<td>2.0</td>
<td>2.3</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>nls</td>
<td>35.5</td>
<td>20.4</td>
<td>26.7</td>
<td>55.7</td>
</tr>
<tr>
<td></td>
<td>sr</td>
<td>42.1</td>
<td>41.8</td>
<td>3</td>
<td>43.9</td>
</tr>
</tbody>
</table>
Table II. Simulation results for the mean fits of the incomplete cases: RMSE, ML1 and standard deviation for the Bayesian method (ab), the traditional method (t), the compartmental method (nls), the cubic smoothing splines (sr) and the compartmental method with random effect (nlme).

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>ML1</th>
<th>St. dev.</th>
<th>RMSE</th>
<th>ML1</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>7.4</td>
<td>4.3</td>
<td>6.9</td>
<td>10.7</td>
<td>9.0</td>
<td>7.0</td>
</tr>
<tr>
<td>t</td>
<td>7.9</td>
<td>5.8</td>
<td>7.1</td>
<td>8.9</td>
<td>6.0</td>
<td>8.6</td>
</tr>
<tr>
<td>nls</td>
<td>12.4</td>
<td>8.2</td>
<td>10.7</td>
<td>60.0</td>
<td>43.7</td>
<td>37.7</td>
</tr>
<tr>
<td>sr</td>
<td>12.5</td>
<td>9.0</td>
<td>8.8</td>
<td>15.4</td>
<td>14.6</td>
<td>4.1</td>
</tr>
<tr>
<td>nlme</td>
<td>12.7</td>
<td>7.5</td>
<td>12.0</td>
<td>32.4</td>
<td>17.7</td>
<td>29.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>ML1</th>
<th>St. dev.</th>
<th>RMSE</th>
<th>ML1</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>10.3</td>
<td>7.4</td>
<td>8.1</td>
<td>22.7</td>
<td>16.1</td>
<td>19.7</td>
</tr>
<tr>
<td>t</td>
<td>11.1</td>
<td>7.4</td>
<td>11.1</td>
<td>32.2</td>
<td>2.0</td>
<td>32.1</td>
</tr>
<tr>
<td>nls</td>
<td>9.6</td>
<td>6.7</td>
<td>8.3</td>
<td>11.9</td>
<td>8.9</td>
<td>9.0</td>
</tr>
<tr>
<td>sr</td>
<td>11.7</td>
<td>7.7</td>
<td>12.7</td>
<td>36.3</td>
<td>36.4</td>
<td>8.4</td>
</tr>
<tr>
<td>nlme</td>
<td>9.1</td>
<td>6.5</td>
<td>8.2</td>
<td>10.3</td>
<td>7.2</td>
<td>8.5</td>
</tr>
</tbody>
</table>

For each PK parameter, Tables I and II give, for the complete and incomplete cases (based on the mean fit for the latter case), the root mean-squared error (RMSE), the median $L_1$ error (ML1) and the standard deviation of the errors, expressed as a percentage of the true value. For instance, RMSE and ML1 of AUC are computed as

$$\text{RMSE(AUC)} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (AUC^m - AUC^{\text{true}})^2} \times 100\%$$

$$\text{ML1(AUC)} = \text{median} \left( \frac{|AUC - AUC^{\text{true}}|}{AUC^{\text{true}}} \right) \times 100\%$$

where $AUC$ is the $M$-length vector of the estimated AUC with $M$, the number of simulations.

Abbreviation ab stands for the adaptive Bayesian model, t for the traditional method, nls for the compartmental model, sr for the robust smoothing splines and nlme for the compartmental model with a random effect.

Let us first compare the different estimation methods for the complete design (see Table I). One can see that, for AUC, all the methods perform similarly, but with a poor performance of the compartmental model at a high level of noise. Comparable results are obtained by the different approaches in the estimation of $C_{\text{max}}$, except for relatively bad results of the traditional and the cubic smoothing splines methods when the level of noise is high ($\sigma = 0.4$). For the estimation of $t_{\text{max}}$, the compartmental method is the best choice whatever be the level of noise, whereas the traditional and the cubic smoothing splines methods should be avoided; the Bayesian method lies in between. Finally, for the estimation of $t_{1/2}$, the Bayesian and the traditional methods are, by far, the best choices for all levels of noise.

In the incomplete design case (see Table II), one can see that all the methods perform similarly in the estimation of $C_{\text{max}}$. However, the Bayesian and the traditional methods are clearly the best choices to estimate AUC and $t_{1/2}$. Finally, as in the complete design case, a compartmental model should be preferred to estimate $t_{\text{max}}$.

When several PK profiles are observed in the incomplete design case, a separate curve can be fitted using the hierarchical Bayesian model. The performances of this method are compared with those obtained with a compartmental model including a random effect for $\log(C_l)$ and for $\log(K_e)$. This last model was fitted using the function nlme in the R software. The RMSE, the ML1 and the standard deviation of the errors were computed from the individual relative errors obtained for each subject at each simulation (see Table III). For instance, the individual relative error of AUC for subject $j$ at simulation $m$ is computed as

$$\text{AUC}_{\text{estimated}} - \text{AUC}_{\text{true}} \div \text{AUC}_{\text{true}}$$

One can see that the two methods perform similarly in the estimation of $C_{\text{max}}$. However, the Bayesian method is preferable to estimate the AUC and $t_{1/2}$. Finally, as in the non-hierarchical case, a compartmental method should be preferred to estimate $t_{\text{max}}$.

Note that the results of the Bayesian and of the traditional methods with regard to $t_{\text{max}}$ and $C_{\text{max}}$ do not surprisingly depend on the position of the design (time) points with respect to (the unknown) $t_{\text{max}}$. For example, a (nonreported) simulation...
with observations added at 1.5 and 3-h postdosing ($\sigma = 0.1$) revealed a 30% decrease in the RMSE of $t_{\text{max}}$ and $C_{\text{max}}$. Then the Bayesian method becomes a valuable tool for estimating all aspects of PK profiles even when the level of noise is large.

### 6. APPLICATION

The Bayesian models were applied to data coming from a real PK study. The parametric compartmental models could not be fitted to these data because of convergence problems. In such situations where parametric models do not provide satisfactory results, Bayesian models are a worthwhile alternative to the simple interpolation methods.

#### 6.1. Complete design

We present an application of the Bayesian model (see Section 3.4) on real data observed on a single subject (see Table IV).

Figure 6 shows the estimated PK profile with 90% credibility sets for it. To obtain these curves, we estimate at each MCMC generation, the concentrations over a detailed grid of time. Then, we take the quantiles 5%, 50% and 95% of the MCMC chain of the concentration estimated at each time point. Table V reports the estimations of each PK parameter with 90% credibility set.

#### 6.2. Incomplete design

Table VI gives PK data for six rats observed in an incomplete design.

Figure 7 shows the fitted mean PK profile (thick solid line) and the individual PK profiles. The estimated PK parameters are given in Table VII.

### Table III. Simulation results for the individual fits of the incomplete cases: RMSE, ML1 and standard deviation for the Bayesian method (ab) and the compartmental method with random effect (nlme).

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>ML1</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>ab</td>
<td>20.7</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>nlme</td>
<td>27.6</td>
<td>14.9</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>ab</td>
<td>11.7</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>nlme</td>
<td>12.7</td>
<td>8.6</td>
</tr>
<tr>
<td>$t_{\text{max}}$</td>
<td>ab</td>
<td>24.1</td>
<td>15.9</td>
</tr>
<tr>
<td></td>
<td>nlme</td>
<td>14.0</td>
<td>8.3</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>ab</td>
<td>11.7</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>nlme</td>
<td>60.2</td>
<td>34.6</td>
</tr>
</tbody>
</table>

### Table IV. Data observed on a single subject.

<table>
<thead>
<tr>
<th>Times (in hours)</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrations</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td>35</td>
<td>84</td>
<td>102</td>
<td>38</td>
<td>16</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PK parameter</th>
<th>Median</th>
<th>90% Credibility set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{max}}$</td>
<td>2.01</td>
<td>[1.69, 2.44]</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>80.2</td>
<td>[53.3, 108.5]</td>
</tr>
<tr>
<td>AUC</td>
<td>392</td>
<td>[304, 558]</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>4.29</td>
<td>[2.98, 5.76]</td>
</tr>
</tbody>
</table>

### Table V. Estimated PK parameters with credibility sets.

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat 1</td>
<td>5.4</td>
<td>6.76</td>
<td>8.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat 2</td>
<td>3.56</td>
<td>8.54</td>
<td>8.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat 3</td>
<td>3.59</td>
<td>13.1</td>
<td>3.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat 4</td>
<td>21</td>
<td>18.2</td>
<td>4.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat 5</td>
<td>14</td>
<td>28.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat 6</td>
<td>7.79</td>
<td>12.3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table VI. Observed concentrations (in mg/l) in six rats under an incomplete design.
7. DISCUSSION

A Bayesian method based on P-splines was developed to estimate PK profiles and PK parameters. Like the traditional non-compartmental method, Bayesian estimation is fast and does not require the assumptions of a compartmental model.

The advantages of the Bayesian method over the traditional one are shown with the simulations where more accurate and more precise estimates are generally obtained for the PK parameters, especially when the noise is large. Furthermore, the Bayesian method offers the advantage to provide uncertainty measures for all PK parameters via the credibility sets obtained by MCMC, even in an incomplete design.

The Bayesian method is always a good choice to estimate the AUC, $C_{\text{max}}$ and $t_{1/2}$. Its superiority over the (correctly specified) compartmental model arises in the estimation of the AUC and $t_{1/2}$ when the level of noise is large in the complete design case, and already with a low level of noise in the incomplete design case.

Better results are obtained in the estimation of $t_{\text{max}}$ with a correctly specified compartmental model. However, a careful choice of the design points markedly improves the performances of the Bayesian method in that respect. Then, the Bayesian method is a valuable tool for estimating all aspects of PK profiles.

The presented hierarchical model enables to fit an individual profile for each subject even when the design is incomplete. In comparison with the results obtained with the correctly specified population compartmental model, the Bayesian method provides better estimates for AUC and $t_{1/2}$, comparable estimates for $C_{\text{max}}$ and inferior results for $t_{\text{max}}$. These conclusions are in agreement with those in the non-hierarchical setting.

The Bayesian method is particularly advisable when the level of noise is high or when the suitable

![Figure 7. Application to real data for the incomplete case. The thick solid line is the mean profile, the thin dashed one is the profile for rat 1 (circles), the thin dotted one for rat 2 (plus), the thin dashed-dotted one for rat 3 (diamonds), the thick dashed-dotted one for rat 4 (stars), the thick dashed one for rat 5 (triangles) and the thin solid one for rat 6 (full circles).](image)

<table>
<thead>
<tr>
<th></th>
<th>$t_{\text{max}}$</th>
<th>$C_{\text{max}}$</th>
<th>AUC</th>
<th>$t_{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>2.15 [1.19; 3.35]</td>
<td>11.75 [8.57; 17.48]</td>
<td>106.32 [80.17; 154.29]</td>
<td>5.95 [4.09; 8.53]</td>
</tr>
<tr>
<td>Rat 2</td>
<td>2.19 [1.14; 3.35]</td>
<td>12.16 [8.94; 17.87]</td>
<td>111.15 [81.22; 155.60]</td>
<td>5.91 [4.06; 8.35]</td>
</tr>
<tr>
<td>Rat 3</td>
<td>2.02 [1.05; 2.95]</td>
<td>10.39 [7.33; 15.09]</td>
<td>85.57 [60.49; 120.01]</td>
<td>6.49 [3.95; 10.55]</td>
</tr>
<tr>
<td>Rat 4</td>
<td>1.38 [0.90; 2.36]</td>
<td>21.89 [13.22; 33.06]</td>
<td>125.07 [92.18; 173.95]</td>
<td>6.08 [3.61; 9.95]</td>
</tr>
<tr>
<td>Rat 6</td>
<td>2.47 [1.59; 4.05]</td>
<td>12.22 [8.92; 17.18]</td>
<td>109.80 [81.04; 162.60]</td>
<td>5.90 [4.02; 8.37]</td>
</tr>
</tbody>
</table>
compartmental model is uncertain. It can easily be automated to be used by non-statisticians. For our data sets, we do not notice any influence of the choice for the hyperparameters in the prior distribution on the fitted curves. However, sensitivity of the results to such a choice has been reported in other contexts with some hierarchical models [32]. If necessary, a mixture prior (for \( \tau_j \), see Section 3.4.2) can be used [27].

The hierarchical model can be extended to analyse repeated intravenous/oral dosing, dose proportionality studies or to perform group comparisons, for instance.

ACKNOWLEDGEMENTS

Astrid Jullion thanks Eli Lilly for the financial support through a patronage research grant and the UCL for an FSR research grant. Financial support from the IAP research network nr. P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is also gratefully acknowledged by Philippe Lambert. No relevant conflict of interest is to be declared.

REFERENCES

15. Wright J, Boddy A. Half-lifes are wrong but some half-lifes are useful. *Clinical Pharmacokinetics* 2001; 40:237–244.
APPENDIX A

Model specification with the two extensions:

\[ (Y, \theta, \tau) \sim \mathcal{N}(\beta(x)\theta, \tau^{-1}) \]

\[ p(\tau) \propto \tau^{-1} \]

\[ p(\theta | \tau, A) \propto \exp[-0.5 \tau J^{-1} \theta^T D^T A \theta] \]

\[ \lambda_k \sim \mathcal{G}(\omega, \omega) \quad \text{when } k > r + 1, \lambda_{r+1} = 1 \]

\[ (\tau_{2j} | p) \sim \sum_{m=1}^{M} p_m \mathcal{G}(1, b_m) \]

\[ p \sim \mathcal{G}(u) \]

APPENDIX B

The model specification for the hierarchical model is the following:

\[ (Y, \theta, \tau) \sim \mathcal{N}(\mathbf{M}_0, \tau^{-1}) \]

\[ p(\theta_0 | \tau_0, A) \propto \exp[-0.5 \tau_0 J^{-1} \theta_0^T D^T A \theta_0] \]

\[ (\theta | \eta, \tau_j) \sim \mathcal{N}(0, \Sigma_j) \quad \forall j = 1, \ldots, n_s \]

\[ \lambda_k \sim \mathcal{G}(\omega, \omega) \quad \text{when } k > r + 1, \lambda_{r+1} = 1 \]

\[ p(\tau) \propto \tau^{-1} \]

\[ (\tau_0 | p_0) \sim \sum_{m=1}^{M} p_{0m} \mathcal{G}(1, b_m) \]

\[ p_0 \sim \mathcal{G}(u) \]

\[ (\tau_j | p_j) \sim \sum_{m=1}^{M} p_{jm} \mathcal{G}(1, b_m) \quad \forall j = 1, \ldots, n_s \]

\[ p_j \sim \mathcal{G}(u) \quad \forall j = 1, \ldots, n_s \]

\[ \eta \sim \mathcal{G}(a_n, b_n) \]

with

\[ \Sigma_j^{-1} = \tau_j J^{-1} + \eta \text{diag}(1, 0, \ldots, 0, 1) \quad \forall j = 1, \ldots, n_s \]

We give here an example of matrix \( \mathbf{M} \) for \( n_s = 6 \) subjects with four observations per subject. \( B \) is the B-splines basis for the 24 observations with 9 knots. The matrix \( \mathbf{M} \) is given by

\[
\mathbf{M} = \begin{bmatrix}
B(1:4,:) & B(1:4,:) & 0 & 0 & 0 & 0 & 0 \\
B(5:8,:) & 0 & B(5:8,:) & 0 & 0 & 0 & 0 \\
B(9:12,:) & 0 & 0 & B(9:12,:) & 0 & 0 & 0 \\
B(13:16,:) & 0 & 0 & 0 & B(13:16,:) & 0 & 0 \\
B(17:20,:) & 0 & 0 & 0 & 0 & B(17:20,:) & 0 \\
B(21:24,:) & 0 & 0 & 0 & 0 & 0 & B(21:24,:) 
\end{bmatrix}
\]
where \( B(i_1 : i_2, :) \) stands for the submatrix of \( B \) corresponding to its rows \( i_1 \) to \( i_2 \). We summarize the matrix \( M \) with this notation:

\[
M = [M_0, M_1, M_2, \ldots, M_n,]
\]

**APPENDIX C**

**C.1. Extended P-splines model**

The conditional posterior distributions are

\[
(\theta | \tau, \lambda ; y) \sim \mathcal{N}(\tau \Sigma_0 B' y, \Sigma_0) \\
(\tau | y, \lambda) \sim \mathcal{N}(0.5y, 0.5(y - B \theta)'(y - B \theta)) \\
(\lambda | \theta, \lambda, y) \sim \mathcal{N}(0, 0.5y) \\
(\theta | y, \lambda) \sim \mathcal{N}(\mu, \Sigma_0^{-1}) \\
(\mu, \Sigma_0^{-1}) \sim \mathcal{N}(\mu_0, \Sigma_0) \\
(\lambda | y) \sim \mathcal{N}(0, 0.5) \\
(\tau | y) \sim \mathcal{N}(a, 0.5) \\
(\theta | y, \lambda) \sim \mathcal{N}(\mu_0, \Sigma_0)
\]

where

\[
\begin{align*}
\theta &= \tau \Sigma_0 B' y \\
\Sigma_0^{-1} &= \tau B' B + \tau_j D' AD \\
\mu_0 &= \tau \Sigma_0 B' W \\
W &= y - \sum_{i=1}^n M_i \theta_i \\
\mu_j &= \tau \Sigma_{p_j} B' W_j, \quad \Sigma_{p_j}^{-1} = \tau B' B + \Sigma_j^{-1} \\
W_j &= y - M_0 \theta_0 - \sum_{i=1, i \neq j}^n M_i \theta_i \\
\Sigma_j^{-1} &= \tau_j D' D + \eta \text{diag}(1, 0, \ldots, 0, 1) \\
c_{j,m} &= \exp(-\tau_j b_m) b_m^T \sum_{k=1}^M \frac{u_k}{u_m} \\
c_{0,m} &= \exp(-\tau_0 b_m) b_m^T \sum_{k=1}^M \frac{u_k}{u_m}
\end{align*}
\]

**C.2. Hierarchical model**

The conditional distributions are

\[
(\theta | y) \sim \mathcal{N}(0.5y, 0.5(y - M_0 \mu_0)') (y - M_0 \mu_0) \\
(\theta | y, \lambda) \sim \mathcal{N}(\mu, \Sigma_0^{-1}) \\
(\tau | \lambda, \lambda, y) \sim \mathcal{N}(0, 0.5y) \\
(\lambda | \theta, \lambda, y) \sim \mathcal{N}(0, 0.5y) \\
(\theta | \lambda, y) \sim \mathcal{N}(\mu, \Sigma_0^{-1}) \\
(\lambda | y) \sim \mathcal{N}(0, 0.5) \\
(\tau | y) \sim \mathcal{N}(a, 0.5) \\
(\theta | y, \lambda) \sim \mathcal{N}(\mu_0, \Sigma_0) \\
(\lambda | y) \sim \mathcal{N}(0, 0.5)
\]

where

\[
\begin{align*}
\theta &= \tau \Sigma_0 B' y \\
\Sigma_0^{-1} &= \tau B' B + \tau_j D' AD \\
\mu_0 &= \tau \Sigma_0 B' W \\
W &= y - \sum_{i=1}^n M_i \theta_i \\
\mu_j &= \tau \Sigma_{p_j} B' W_j, \quad \Sigma_{p_j}^{-1} = \tau B' B + \Sigma_j^{-1} \\
W_j &= y - M_0 \theta_0 - \sum_{i=1, i \neq j}^n M_i \theta_i \\
\Sigma_j^{-1} &= \tau_j D' D + \eta \text{diag}(1, 0, \ldots, 0, 1) \\
c_{j,m} &= \exp(-\tau_j b_m) b_m^T \sum_{k=1}^M \frac{u_k}{u_m} \\
c_{0,m} &= \exp(-\tau_0 b_m) b_m^T \sum_{k=1}^M \frac{u_k}{u_m}
\end{align*}
\]