

Comparisons Between Hamiltonian Monte Carlo and Maximum A Posteriori For A Bayesian Model For Apixaban Induction Dose & Dose Personalization

A. Demetri Pananos

APANANOS@UWO.CA

*Department of Epidemiology & Biostatistics
Western University
London, Ontario, Canada*

Daniel J. Lizotte

DLIZOTTE@UWO.CA

*Department of Computer Science
Department of Epidemiology & Biostatistics
Western University
London, Ontario, Canada*

Abstract

Precision medicine’s slogan is “right drug – right patient – right time.” Implicit in the slogan is “right dose”; however, determining the right dose for any one patient can be challenging when dose-response data are limited. Bayesian methods, with their ability to explicitly incorporate prior information to supplement limited data, have been proposed as a solution to this problem. Although Hamiltonian Monte Carlo (HMC) is a leading methodology for inference in Bayesian models because of its ability to capture posterior distributions with high fidelity, dose personalization studies commonly use simpler Maximum A Posteriori (MAP) inference methods. The impact of the choice of inference engine on dose decision-making has not been explored. To better understand this issue, we perform a simulation study characterizing the differences between inferences made via MAP and HMC for personalized dosing strategies. The simulation study uses a new Bayesian pharmacokinetic model for apixaban pharmacokinetics written in an open source Bayesian language; the model code and posterior summaries of all parameters will be publicly available. We demonstrate that the differences between HMC and MAP are non-trivial and can greatly affect the choices surrounding dose selection for personalized medicine.

1. Introduction

Precision medicine’s slogan is “right drug – right patient – right time”. Implicit in the slogan is “right dose”; however, determining the right dose for any one patient can be challenging. The anticoagulant Warfarin offers a good example of these challenges; physicians choose an initial dose based on guidelines and their own experience. They then closely monitor the patient’s International Normalised Ratio (INR), which measures how long it takes blood to clot, and in response they adjust the dose over time.

Pharmacokinetic and statistical models of how drugs behave within an individual can alleviate some of these challenges by predicting the effects of different doses based on patient covariates. In some studies ([Schwarz et al., 2008](#); [Sohrabi and Tajik, 2017](#); [Caldwell](#)

et al., 2007) a cohort of patients will have an appropriate maintenance doses determined empirically and these are then regressed onto patient covariates. In others (Ohara et al., 2019; Zhu et al., 2017; Xue et al., 2017) patient pharmacokinetics are directly modeled and can be simulated under different dosing regimens to find an appropriate dose. In both cases, uncertainty in the models can be assessed and can help guide clinical decisions as to what dose is best or what dose to try next.

Both types of models can provide guidance for individual patients, but only when there is enough data so that the models are accurate and reliable. For personalized pharmacokinetics-based dosing, this amount of data is rarely available in practice. Obtaining sufficient data to learn a patient’s pharmacokinetic parameters requires a lengthy observation period which few patients are willing or capable of committing to. Population pharmacokinetic models could be used in place of a patient’s pharmacokinetics, but treating the patient as “average” is precisely what precision medicine seeks to improve upon.

In many contexts where limited data are available, Bayesian methods with informative priors have been proposed. Model priors allow analysts to specify their beliefs about model parameters prior to seeing a new patient’s data, and to combine those beliefs with new observations to form personalized predictive models. This allows models to “hit the ground running” so to speak, and makes use of all available data to support decision-making. To use all but the simplest of Bayesian models for decision-making requires computational approximation techniques to obtain model estimates and predictions. Several approaches exist for generating approximate samples from the posterior distribution, with Hamiltonian Monte Carlo (HMC) being considered the gold standard (Neal, 1996; Hoffman and Gelman, 2014; Carpenter et al., 2017; Tripuraneni et al., 2017). Despite HMC being the preferred method by theorists and applied Bayesians alike, methods like Maximum A Posteriori (MAP), in which the posterior mode is computed via optimization and then a Laplace approximation is performed, continue to be used in population Bayesian pharmacokinetic studies as late as 2020 (Brooks et al., 2016; Nguyen et al., 2016; Preijers et al., 2019; Stiff et al., 2020). HMC and MAP are two different approaches with different strengths and different theoretical motivations. Naturally, this raises questions regarding how decisions in personalized medicine may be affected by the use of different methods for performing inference, even using the same model and data. We seek to answer these questions by developing a new, high-fidelity Bayesian pharmacokinetic model and then investigating the impact of the choice of inference method on precision medicine decisions.

Generalizable Insights about Machine Learning in the Context of Healthcare

The main methodological insight we gained was that *although predictions made by HMC and MAP may appear to be very similar according to common error metrics, they can lead to very different personalized dosing decisions*. The main contributions of this paper are as follows:

1. A new Bayesian model for apixaban pharmacokinetics written in an open source Bayesian language. We make the model code and posterior summaries of all parameters publicly available at <https://github.com/Dpananos/PKBayes>.

2. A simulation study demonstrating that inferences made via MAP and HMC lead to very different dosing strategies.
3. An induction dosing model for apixaban based on desired trough concentration level after a first dose.

2. Background

Pharmacokinetic Modelling

Pharmacokinetics is the study of the dynamics of a mass of drug in the body and is concerned with the absorption, distribution, metabolism, and excretion of that drug. Differential equations (equations which relate the derivative of an unknown function to itself) are often used to describe how these dynamics evolve over time. The differential equation models in pharmacokinetics are called “compartmental models” as they idealize different parts of the body as compartments from which drug can flow in and out at rates proportional to how much drug is presently in that compartment. If the differential equation is not too complex, the solution can be written in terms of analytic functions. In the case where the differential equation cannot be solved in terms of analytic functions, a rich literature of numerical techniques exist to approximate the solution to within quantifiable precision. In either case, estimation of model parameters is of interest as they represent pharmacokinetic measures, such as the volume of distribution or rate constants for which the drug is absorbed into/excreted out of a compartment. If the parameters for such a model are known, we can use the model to make predictions about drug concentration as a function of time and dose. This in turn can be used to select a dose that meets given criteria about what the concentration function should look like.

Parameter estimation for these models can be done in both frequentist and Bayesian frameworks. In a Bayesian framework, parameter estimation begins by specifying a prior distribution which reflects the knowledge of parameters before seeing data. Once data are observed, Bayes’ rule can be used to get the posterior distribution. This distribution provides information about what parameter values have most plausibly generated the observed data. By virtue of being a probability distribution, the posterior can be summarized by expectations to get point estimates of model parameters. Shown in figure 1 is a visual summary of how Bayes’ rule and Bayesian modelling of pharmacokinetics works using pseudodata. The leftmost panel is our prior distribution. Each concentration curve results from specific combinations of parameters for the model which are believed to be plausible before seeing data. Once data are observed (the middle panel), application of Bayes’ rule yields the rightmost panel. Concentration curves in this panel correspond to combinations of parameters which have most plausibly generated the data, resulting in concentration curves which have most plausibly generated the data. Note that in this setting, because we have many measurements, the pharmacokinetic model is well-determined and the posterior uncertainty is small. Except in very simple cases, the integrals required to evaluate the posterior become intractable; thus, computational approximations are required to fit Bayesian models.

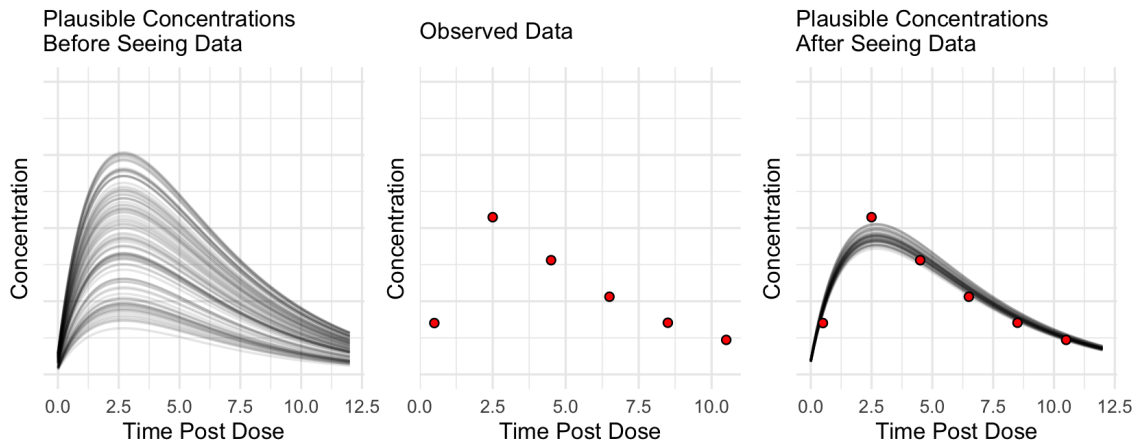


Figure 1: A demonstration of a Bayesian workflow for pharmacokinetic models. The left-most panel represents the prior. Each curve corresponds to a unique set of model parameters which induce each concentration function. In the center panel is the data observed from a single patient. Conditioning on this data yields the right-most panel. Each curve corresponds to a unique set of model parameters drawn from the posterior distribution.

Dosing Decisions

Vitamin K antagonists, such as the popular oral anticoagulant Warfarin, are known to have narrow therapeutic windows as well as drug and food interaction. Determination of a maintenance dose is consequently a procedure with frequent monitoring and followup, with some sources recommending monitoring daily or every other day until the INR stabilizes for two days. The narrow therapeutic window forces investigators to also consider the pharmacodynamics (the study of the onset, intensity, and duration of the drug response and how these are related to the concentration of the drug at its site of action) of Warfarin in addition to the pharmacokinetics when determining dose size as concentration of the drug alone is not sufficient to infer the antithrombotic effect in patients. The introduction of factor Xa inhibitors like apixaban has alleviated some of the difficulties in prescribing anticoagulants. Factor Xa inhibitors have been shown to have lower risk for bleeding than Warfarin in patients with atrial fibrillation (Vinogradova et al., 2018) and also allow for fixed dosing without frequent monitoring of INR. Furthermore, unlike Warfarin, the pharmacodynamic effect of apixaban on clotting is closely correlated with the concentration in the plasma (Byon et al., 2019), making pharmacokinetic modelling more informative on antithrombotic effect as compared to Warfarin. However, as of writing this paper there is little information on the therapeutic window, making selecting dose sizes large enough to avoid thromboembolism difficult. Furthermore, studies have demonstrated that inter-patient variability of apixaban plasma concentrations is much higher than was initially believed (Gulilat et al., 2020). In this work,

we develop personalized dosing whose goal is to find the minimum dose that avoids plasma concentrations that are too low.

3. Methods

Bayesian Model

To achieve our three objectives of 1) developing a new Bayesian pharmacokinetic model for apixaban, 2) investigating the impact of MAP versus HMC inference on dosing decisions and 3) developing an induction dosing model for apixaban, we fit a hierarchical mixed effects model of apixaban pharmacokinetics using data from [Beaton et al. \(2018\)](#). Thirty-six participants were given 5 mg of apixaban and 100 ml of water in a fasted state. Blood plasma concentrations of apixaban were recorded over the course of 12 hours.

Table 1: Summary of data from [Beaton et al. \(2018\)](#).

	Female (N=23)	Male (N=13)	Overall (N=36)
Age			
Mean (SD)	48.8 (11.6)	51.7 (11.6)	49.8 (11.5)
Median [Min, Max]	49.0 [26.0, 67.0]	51.0 [31.0, 70.0]	50.0 [26.0, 70.0]
Weight			
Mean (SD)	84.6 (23.5)	94.0 (25.6)	88.0 (24.4)
Median [Min, Max]	82.8 [54.7, 136]	87.2 [62.0, 137]	83.5 [54.7, 137]
Creatinine			
Mean (SD)	68.5 (12.1)	67.2 (13.5)	68.0 (12.5)
Median [Min, Max]	66.0 [50.0, 95.0]	65.0 [50.0, 95.0]	65.0 [50.0, 95.0]
BMI			
Mean (SD)	29.9 (6.74)	31.5 (5.62)	30.5 (6.33)
Median [Min, Max]	29.5 [18.3, 42.3]	31.8 [23.3, 40.7]	31.3 [18.3, 42.3]

Since participants were given a single dose of apixaban in a fasted state, we use a single-compartment pharmacokinetic model with first order absorption and elimination. The solution to the differential equation describing mass transit, and consequently the concentration function, is then

$$y(t) = \begin{cases} \frac{F \cdot D}{Cl} \frac{k_e \cdot k_a}{k_e - k_a} \left(e^{-k_a(t-\delta)} - e^{-k_e(t-\delta)} \right) & \delta \leq t \\ 0 & \text{else} \end{cases} \quad (1)$$

Here, D is the size of the dose in mg, F is the bioavailability (fixed to 0.5 for apixaban [Byon et al. \(2019\)](#)), Cl is the clearance rate in units litres per hour, k_a is the rate constant of absorption into the volume of distribution in units 1/hours, and k_e is the elimination rate constant in units 1/hours. We include a time delay, δ , to relax the assumption that

absorption begins immediately after ingestion. Parameters are considered as random effects, with some population mean and variance (which is estimated from the data).

Priors for k_e and k_a are not defined explicitly. Rather, our model puts priors on the time to max concentration, which can be expressed as a function of the parameters in equation (1)

$$t_{max} = \frac{\ln(k_a) - \ln(k_e)}{k_a - k_e} \quad (2)$$

and on the ratio between k_e and k_a , which we call α

$$\alpha = \frac{k_e}{k_a} . \quad (3)$$

We choose to place a prior on the quantity α because it arises when non-dimensionalizing (Lin and Segel, 1988) the differential equation governing mass transit of the drug in and out of the volume of distribution. The plasma concentration function is a version of the “flip-flop” model (Wakefield, 1996; Salway and Wakefield, 2008), since different parameterizations of this model can yield the same curve, leading to model un-identifiability. To ensure the model is identifiable, we require $k_e < k_a$ as has been done in previous Bayesian analyses of this model (Wakefield, 1996; Salway and Wakefield, 2008). This requirement bounds α to the unit interval. In principle, information on the elimination rate constant could be obtained by performing a linear regression on the log concentration values in the latter half of the concentration profiles where the drug is being eliminated from the body. To preserve as much data as possible for model fitting, we forgo this approach. These two sets of equations are used to parameterize the absorption and elimination rate constants as follows

$$k_a = \frac{1}{t_{max}} \frac{\ln(\alpha)}{\alpha - 1} \quad (4)$$

$$k_e = k_a \alpha \quad (5)$$

Details on the prior distributions for each parameter and the model likelihood can be found in the supplementary material. We also include a summary of the model’s posterior as well as details on simulating data from the model’s posterior predictive distribution to generate pseudopatients.

Model Fitting and Diagnostics

For HMC, prior/posterior predictive checks and model fitting was performed using Stan (Carpenter et al., 2017) to draw from the prior/posterior. Twelve chains were initialized and run for 4000 iterations each (1000 for warmup allowing the Markov chain the opportunity to find the correct target distribution and 3000 to use as samples from the posterior). Stan monitors several diagnostics none of which detected problematic HMC behavior¹.

We use Stan’s optimization capabilities to compute the MAP estimates. The L-BFGS optimizer was used to find the posterior mode. The optimizer terminated when either 10,000 iterations had been performed or when the value of the objective function stopped

1. 0 divergences, all Gelman-Rubin diagnostics < 1.01 , smallest effective sample size ratio 16%.

changing within a tolerance of 1×10^{-10} . Once the mode was located, 10,000 samples from the Laplace approximation to the posterior were obtained. Constrained parameters were transformed to the appropriate space before sampling.

Determination of a Personalized Dose

The results from both HMC and MAP yield samples from the approximate posterior of Cl_j , $k_{e,j}$, $k_{a,j}$, and δ_j for each of the j patients. For any given posterior sample, these parameters can be combined to compute a predicted concentration for patient j at time t by using equation (1). We determine a personalized dose size by evaluating the pseudopatients' concentration function under different dose sizes D and then computing posterior probabilities of failing to surpass concentration thresholds. When we have posterior samples, equation (1) turns into a function of the dose size and time conditioned on patient.

We perform two experiments to compare HMC and MAP. In our first experiment, we determine the posterior probability of failing to exceed a concentration of 20 ng/ml 12 hours post dose for each pseudopatient across a variety of dose sizes. We choose 20 ng/ml as our threshold for this experiment because the median concentration at 12 hours post dose in the data from [Beaton et al. \(2018\)](#) is approximately 20 ng/ml. In our second experiment, we determine the posterior probability that the max concentration fails to exceed 80 ng/ml. We choose 80 ng/ml as our threshold for this experiment because the median max concentration in the data is 79 ng/ml (though it is important to note that it is unlikely that these patients were observed exactly at the time which the peak concentration was achieved). These quantities represent two different ways of assessing a patient's risk of being below a given threshold. The chosen threshold is arbitrary, but our method generalizes to any threshold. For each experiment, the risks are computed across a grid of dose sizes of 0 mg to 60 mg, yielding risk as a function of dose size. For each pseudopatient, we interpolate these estimates using a monotone Hermite spline and then invert the risk curve; the inverted risk curve maps risk to dose. This allows us to determine a dose size which produces a specified risk level.

4. Results

Bayesian Model

Diagnostic plots for Bayesian model fit to the real apixaban data is shown in figure 2. Posterior population prediction intervals (that is, the result of integrating out the random effect of each patient) of the observed concentration are realistic and to the eye appear similar to the observed data (top left of figure 2). Residual plots (observed minus posterior mean) indicate homogeneity of variance on the log scale, which is consistent both with expert knowledge on the measurement process and the likelihood we choose (bottom left of figure 2). Predicted concentrations tend to agree with observed concentrations (top right of figure 2), and posterior predictive draws have similar empirical cumulative distribution functions as the observed data (bottom right of figure 2). In figure 3, we show concentration functions obtained from draws from our prior distribution as well as two patients with best and worst fit as measured by mean absolute percentage error (best: 3.29%, worst: 26.4%). Because our HMC diagnostics do not indicate problematic behaviour in the Markov chains,

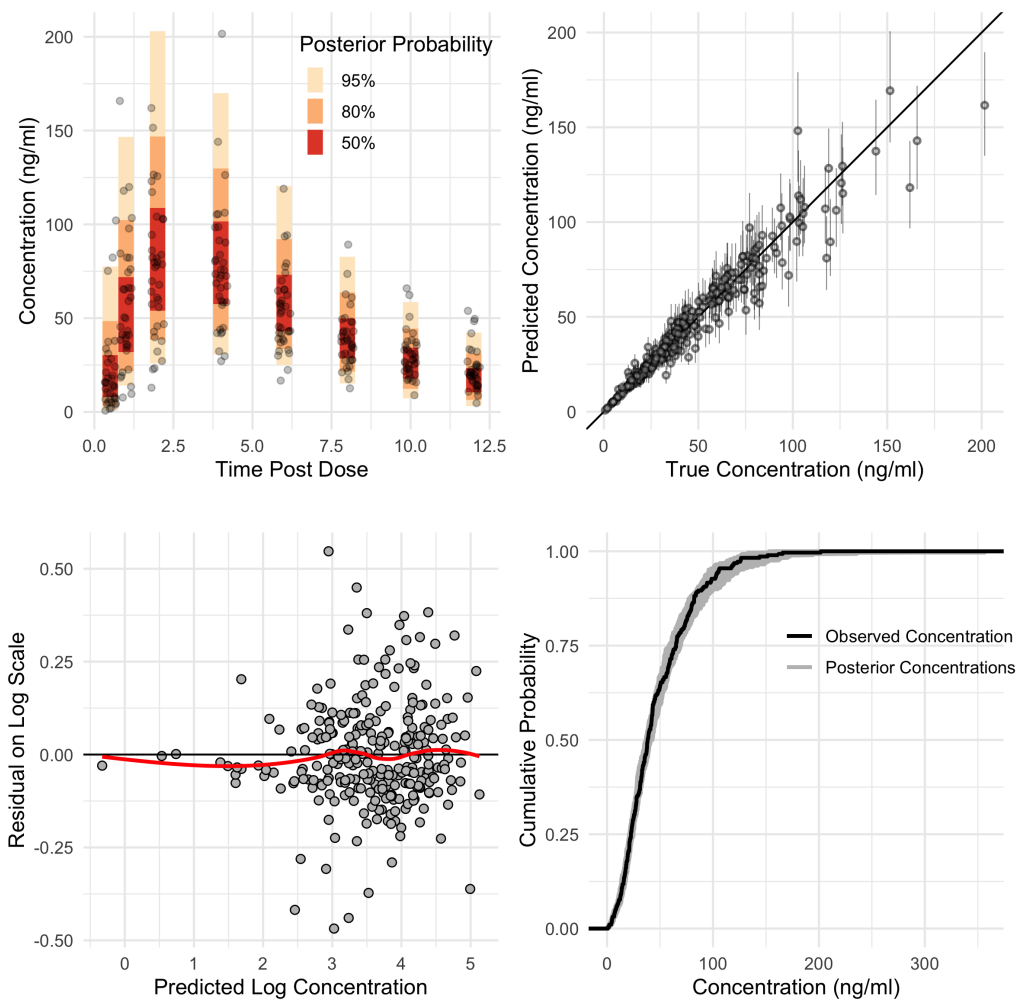


Figure 2: Diagnostic plots for our Bayesian model. Top left shows the posterior predictive distribution plus observed data. Data points have been perturbed to prevent overlapping. Top right shows the predicted values along with accompanying 95% equal-tailed posterior credible interval. Bottom left shows the residuals (on the log scale) between the observed concentrations and the posterior mean concentration, bottom right shows the cumulative density function for the observed data (black) as well as draws from the posterior predictive distribution (gray).

and because the model diagnostics indicate adequate fit, we believe the obtained model's posterior predictive distribution is adequate for simulation of pseudopatients.

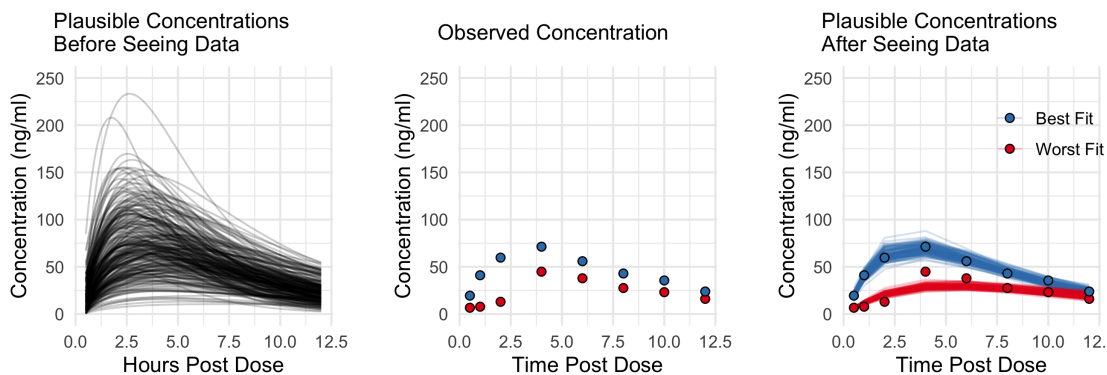


Figure 3: The leftmost panel shows 250 draws from the prior defined in the previous section. The center panel shows data from two patients who achieved the best (blue) and worst (red) model fit as measured through mean absolute percent error. The rightmost panel shows 250 draws from the posterior for these patients. Not shown here are the other 34 patients in our data, for which the model is also capable of performing predictions for.

	HMC	MAP
MSE (SD)	6.67 (15.93)	8.57 (19.93)
MAE (SD)	1.71 (1.94)	1.97 (2.17)
MAPE (SD)	0.04 (0.03)	0.05 (0.03)

Table 2: Comparison of HMC and MAP on three loss functions common in pharmacokinetics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The loss was computed on samples not seen by our model. Included in parentheses are the standard deviations of the loss values.

Fit on Simulated Patients Using HMC and MAP

Initial comparisons of predicted values indicate that both HMC and MAP yield similar predictions to one another, and similar predictions to actual values of unseen data. Examining predictions alone, it would seem that HMC and MAP are equivalent, or at the very least similar enough so as to not have strong preference for one over the other. When using posterior means, HMC results in lower prediction error on unseen data (see table 2) as measured with Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), but these are not stark differences. Estimates of posterior uncertainty between MAP and HMC can however vary a great deal. Shown in figure 4 are 19 of the 100 simulated patients which have a MAP equal tailed posterior interval at least 50% larger as compared to their HMC equal tail posterior interval at the widest point. We note that while not shown explicitly, unobserved concentrations lie entirely within the HMC and MAP posterior intervals.

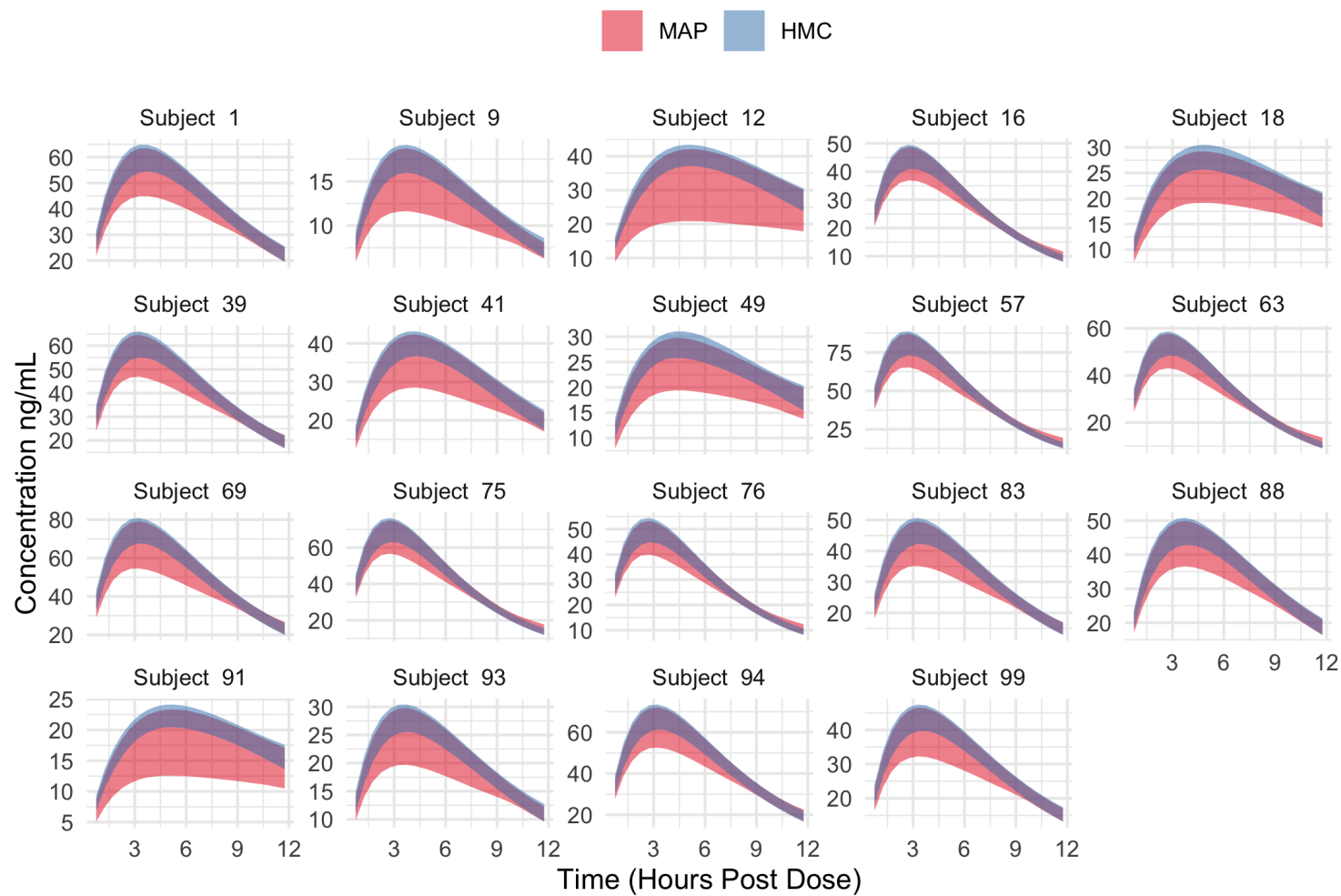


Figure 4: Comparisons of equal tail posterior intervals from MAP and HMC. Note that the concentration scales differ from subplot to subplot. Selected patients are those which have a MAP posterior interval at least 50% as wide or wider than their HMC interval.

Difference in Estimated Dose To Achieve Target Risk

By inverting the risk curve, we can obtain dose size as a function of risk. Shown in figure 5 are the differences between doses computed from HMC and MAP posteriors to achieve the indicated level of risk. The left panel shows the difference in doses in order to achieve a concentration of at least 20 ng/ml 12 hours post dose. For a majority of patients, MAP and HMC agree to within 1 mg though some pseudopatients see a much larger dose recommendation by HMC than by MAP. The right panel shows the difference in doses in order to achieve a max concentration of 80 ng/ml. MAP tends to always recommend larger doses than HMC for this scenario, with the difference between recommended doses becoming larger as the desired risk becomes smaller.

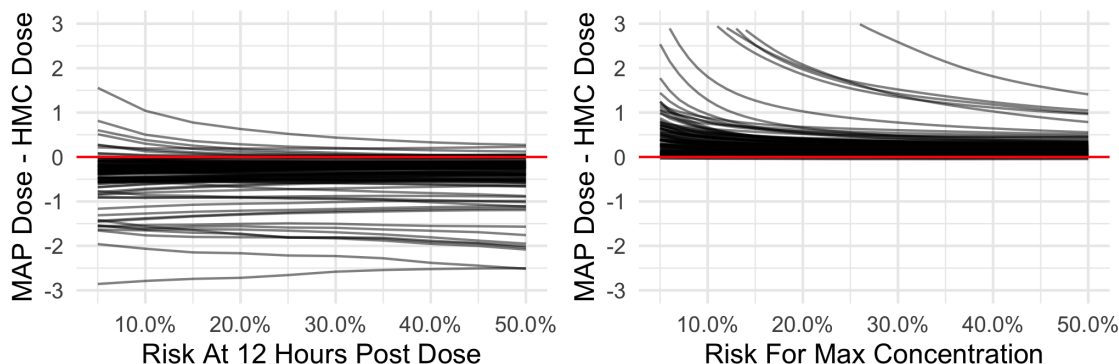


Figure 5: Left: Differences between the estimated doses from MAP and HMC to achieve the indicated risk of having a concentration of apixiban smaller than 20 ng/ml at 12 hours post dose. Each line corresponds to one of the 100 pseudopatients. HMC tends to recommend larger doses than MAP to achieve a desired risk of having the patient’s concentration at 12 hours post dose be below 20 ng/ml. This tendency to recommend larger doses is consistent across desired risk levels. Right: Differences between estimated doses from MAP and HMC to achieve the indicated risk of having a max concentration smaller than 80 ng/ml. Some pseudopatients see dose recommendation differences as large as 10 mg and are thus cut off by the y axis limits. Red lines indicate where the two methods would perfectly agree.

Calibration for Dosing Decisions

Since all pseudopatients were simulated, the true concentration function as a function of the dose size (equation (1)) was known. To further compare HMC and MAP for decision making, we took estimated dose size to achieve a desired risk and computed what the concentration curve under the recommended dose size. We could then compute the number of pseudopatients which actually exceeded the threshold, thus allowing us to examine the cali-

bration of HMC and MAP. The calibration curves for HMC and MAP for both experiments are shown in figure 6.

In the first experiment (left of figure 6), HMC is better calibrated than MAP. This means that when a dose is selected, in order to achieve a risk of being below 20 ng/ml at 12 hours post dose of r , approximately $r \times 100$ pseudopatients have a true concentration function which is smaller than 20 ng/ml at 12 hours post dose. In contrast, MAP is poorly calibrated, and sees more pseudopatients failing to exceed the 20 ng/ml threshold than was desired. Calibration in our second experiment again shows that HMC is better calibrated than MAP, but calibration seems to become worse as the desired risk becomes larger. When we use dose sizes recommended by MAP to achieve a 50% probability of exceeding the 80 ng/ml threshold, only 26% of pseudopatients actually have a max concentration which exceeds the threshold.

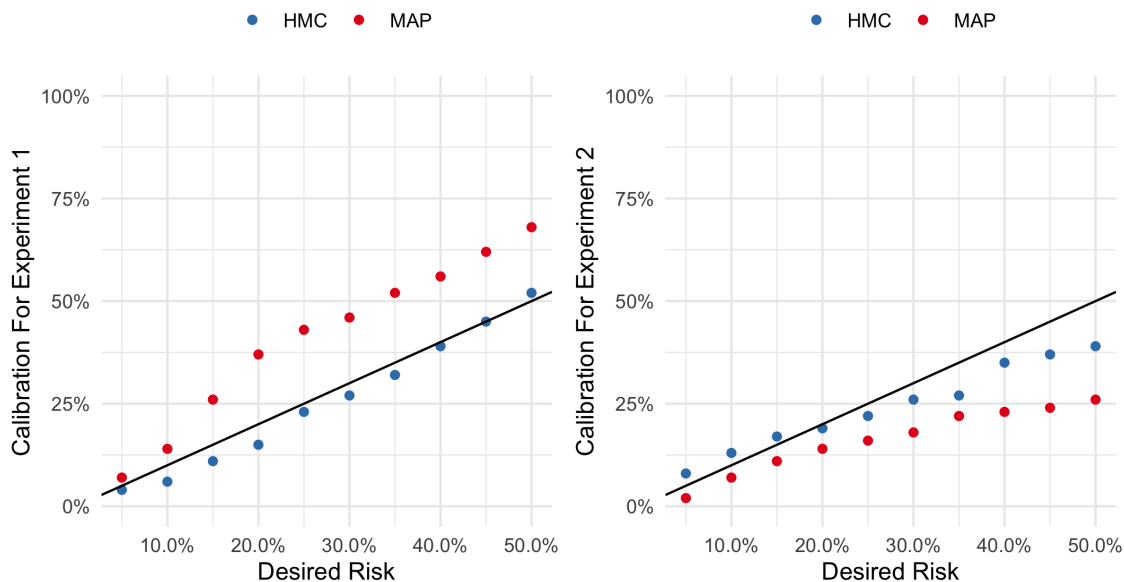


Figure 6: Left: Calibration curves for assessing risk of being below 20 ng/ml. Each dot represents the proportion of the 100 pseudopatients which fail to exceed the 20 ng/ml threshold. When doses are chosen from samples obtained via HMC, then probabilities are well calibrated. When doses are chosen from samples obtained via MAP, more pseudopatients fail to exceed the threshold than were specified. Right: Calibration curves for assessing risk of the max concentration being below 80 ng/ml. HMC appears to be better calibrated than MAP, though the calibration could stand to improve.

5. Discussion

While prediction errors of the point estimates produced by MAP and HMC are very similar as measured by 3 losses common to pharmacokinetic research, they each produce very different estimates of uncertainty. Such estimates are necessary for decision making under uncertainty, where an expected loss is computed over a posterior distribution. In this study, for a given dose, MAP assigns a significant amount of probability mass to concentrations which are far lower than the concentrations considered plausible by HMC. The extent to which this discrepancy would change decisions depends on the loss function, but we see substantial differences in our two example experiments for personalized dosing. The difference in uncertainty between MAP and HMC results in small disagreement for dose size for the majority of patients, but very large disagreement for a sizable minority of about 20%. The observed difference in uncertainty for the concentration function between MAP and HMC are likely due to discrepancies in uncertainty over the individual PK parameters: Each of the 19 pseudopatients in figure 4 see MAP and HMC disagree strongly on posterior uncertainty for the parameters k_a and k_e . MAP lends credence to higher values of k_a and lower values of k_e as compared to HMC. The differences in uncertainty in these parameters is likely the cause of the observed difference in uncertainty in concentration levels. This translates into much wider equal-tailed posterior intervals for concentration using MAP, with 19 out of 100 patients having an MAP equal tailed posterior credible interval at least 50% as wide or wider at their widest point than their HMC equal tailed posterior interval. For each of these 19 patients, MAP appears to have a lower interval estimate far below that of HMC, making it appear as if lower predicted concentrations are probable. This in can make a given proposed dose appear risky in terms of allowing concentration to fall too low; which in turn leads to an increase in recommended dose when the model is asked for a dose that bounds this risk.

While the posterior distribution for this model is too complex to be analyzed analytically, there are good theoretical reasons to prefer HMC over MAP when analysts seek the posterior expectation of some function of parameters. These reasons are nicely summarized by [Betancourt \(2017\)](#), but can be explained by the fact that expectations are computed over volumes, and in high dimensional space there exists more volume away from the mode than in a neighbourhood around it. Because the volume near the mode is so small, these regions of parameter space contribute negligibly to expectations. Instead, regions of parameter space where the product of probability density and volume is large (i.e. *the typical set*) should contribute more to expectations, and this is where our chosen method should be focusing its computational power. Hamiltonian Monte Carlo does exactly this, and thus we prefer it to Maximum A Posteriori.

Neither HMC nor MAP provide perfect representations of the posterior, and discrepancies between the two methods are expected. However, the degree of discrepancy observed in this study and its impact on dosing decisions reveals that these techniques are not interchangeable. One reason for the observed difference might be an insufficient number of observations. However, with 24 equally-spaced observations per each of the 100 simulated patients, this simulation study represents an extremely optimistic (and likely unrealistic) best case scenario. Even specialized studies of pharmacokinetics would collect fewer samples from fewer patients, and even less data collection is practical in clinical practice. Hence,

even if MAP and HMC were to converge to each other with enough data, this amount of data is not available in practice. Another possible reason could be the chosen priors and/or the likelihood, but the model used was identical for both inference methods and had strong priors informed by existing pharmacokinetic data. We note that our proposed model does not account for patient covariates (e.g. weight, BMI, creatinine, etc). Our model could be extended to include covariates as input to the model to further personalize dose. Although our models did not account for patient covariates, a secondary analysis was performed in which patient level pharmacokinetic parameters were regressed onto patient covariates. In this analysis, we observed the same disagreement in model uncertainty as shown in figure 4. Because the uncertainty directly affects dosing decisions, we believe a model sampled using MAP which included covariates would also show similar calibration relative to a model sampled using HMC.

6. Conclusion

We have presented a new Bayesian model for apixaban pharmacokinetics and an induction dosing model for apixaban based on desired trough concentration level after a first dose. We have also presented a simulation study demonstrating that inferences made via MAP and HMC lead to very different dosing strategies; from this simulation study, we derive some general conclusions and guidelines for Bayesian pharmacokinetic modelling as applied to precision medicine.

Bayesian modelling using informative priors provides a practical approach for developing personalized dosing strategies when data are limited. However, the evaluation of Bayesian models, particularly with informative priors, typically focuses on the model itself - are the priors plausible? Do posterior predictive checks look appropriate? In this work, we have demonstrated that the inference technique can have an impact on decision making that is as important as model fidelity, even when the impact on point prediction quality is minimal. Specifically, we have shown that MAP-based inference, which is very commonly used in pharmacokinetics, can lead to very different personalized dosing decisions than HMC-based inference, even in a well-validated model.

Studies using MAP for Bayesian inference in pharmacokinetic models have been published as recently as 2020. The speed and similarity to maximum likelihood makes MAP an attractive and familiar approach as compared to HMC, which can take several minutes to return samples and can use quite complex mechanisms to draw from the posterior. The aforementioned studies have largely focused on point predictions of latent concentrations where, as we have shown, MAP and HMC yield similar results. However, when uncertainty information is used for decision-making, MAP and HMC can lead to very different outcomes.

We recommend that if practitioners do use MAP, that they also compare model results with HMC. Libraries exist to perform HMC in a variety of languages including R, python, and Julia, making HMC widely accessible. Use of these libraries has the added benefit of making analysis more transparent and reproducible for the community at large.

Acknowledgement

We would like to thank Drs. Melanie Beaton and Rommel Tirona for feedback on the pharmacokinetic models and for providing the data to fit our model. This work was supported by the Ontario Ministry of Research, Innovation and Science (Ontario Research Fund—Research Excellence [RE08-063], Principal Investigator: Kim RB) and by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- Melanie D Beaton, Rommel G Tirona, Ruth Strapp, Mala Ramu, Ute Schwarz, Richard Kim, Bandar Aljudaibi, and Zahra Kassam. Su1495 - apixaban and rosuvastatin pharmacokinetics in nonalcoholic fatty liver disease. *Gastroenterology*, 154(6):S-1158, 2018.
- Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo, 2017. arXiv:1701.02434 [stat.ME].
- Emily Brooks, Susan E Tett, Nicole M Isbel, and Christine E Staatz. Population pharmacokinetic modelling and bayesian estimation of tacrolimus exposure: Is this clinically useful for dosage prediction yet? *Clin. Pharmacokinet.*, 55(11):1295–1335, November 2016.
- Wonkyung Byon, Samira Garonzik, Rebecca A Boyd, and Charles E Frost. Apixaban: A clinical pharmacokinetic and pharmacodynamic review. *Clin. Pharmacokinet.*, 58(10):1265–1279, October 2019.
- Michael D Caldwell, Richard L Berg, Kai Qi Zhang, Ingrid Glurich, John R Schmelzer, Steven H Yale, Humberto J Vidaillet, and James K Burmester. Evaluation of genetic factors for warfarin dose prediction. *Clin. Med. Res.*, 5(1):8–16, March 2007.
- Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.
- Markus Gulilat, Denise Keller, Bradley Linton, A Demetri Pananos, Daniel Lizotte, George K Dresser, Jeffrey Alfonsi, Rommel G Tirona, Richard B Kim, and Ute I Schwarz. Drug interactions and pharmacogenetic factors contribute to variation in apixaban concentration in atrial fibrillation patients in routine care. *Journal of Thrombosis and Thrombolysis*, 49(2):294–303, 2020.
- Mathew Hoffman and Andrew Gelman. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.
- C C Lin and L A Segel. *Mathematics Applied to Deterministic Problems in the Natural Sciences*. SIAM, Philadelphia, 1988.
- Radford M Neal. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer, New York, 1996.

- Thu Thuy Nguyen, Henri Bénech, Alain Pruvost, and Natacha Lenuzza. A limited sampling strategy based on maximum a posteriori bayesian estimation for a five-probe phenotyping cocktail. *Eur. J. Clin. Pharmacol.*, 72(1):39–51, January 2016.
- Minami Ohara, Yasuhiko Suzuki, Saki Shinohara, Inna Y Gong, Crystal L Schmerk, Rommel G Tirona, Ute I Schwarz, Ming-Shien Wen, Ming Ta Michael Lee, Kiyoshi Mihara, et al. Differences in warfarin pharmacodynamics and predictors of response among three racial populations. *Clinical pharmacokinetics*, 58(8):1077–1089, 2019.
- Tim Preijers, Britta A P Laros-vanGorkom, Ron A A Mathôt, and Marjon H Cnossen. Pharmacokinetic-guided dosing of factor VIII concentrate in a morbidly obese severe haemophilia a patient undergoing orthopaedic surgery. *BMJ Case Reports*, 12(1):bcr-2018, 2019.
- Ruth Salway and Jon Wakefield. Gamma generalized linear models for pharmacokinetic data. *Biometrics*, 64(2):620–626, June 2008.
- Ute I Schwarz, Marylyn D Ritchie, Yuki Bradford, Chun Li, Scott M Dudek, Amy Frye-Anderson, Richard B Kim, Dan M Roden, and C Michael Stein. Genetic determinants of response to warfarin during initial anticoagulation. *New England Journal of Medicine*, 358(10):999–1008, 2008.
- Mohammad Karim Sohrabi and Alireza Tajik. Multi-objective feature selection for warfarin dose prediction. *Comput. Biol. Chem.*, 69:126–133, August 2017.
- Frank Stiff, Franciscus Vandermeer, Cees Neef, Sander van Kuijk, and Maarten H L Christiaans. A limited sampling strategy to estimate exposure of once-daily modified release tacrolimus in renal transplant recipients using linear regression analysis and comparison with bayesian population pharmacokinetics in different cohorts. *Eur. J. Clin. Pharmacol.*, 76(5):685–693, May 2020.
- Nilesh Tripuraneni, Mark Rowland, Zoubin Ghahramani, and Richard Turner. Magnetic Hamiltonian Monte Carlo. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML’17, pages 3453–3461. JMLR.org, August 2017.
- Yana Vinogradova, Carol Coupland, Trevor Hill, and Julia Hippisley-Cox. Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care. *BMJ*, 362:k2505, 2018.
- Jon Wakefield. The bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association*, 91(433):62–75, 1996.
- Ling Xue, Nick Holford, Xiao-liang Ding, Zhen-ya Shen, Chen-rong Huang, Hua Zhang, Jing-jing Zhang, Zhe-ning Guo, Cheng Xie, Ling Zhou, Zhi-yao Chen, Lin-sheng Liu, and Li-yan Miao. Theory-based pharmacokinetics and pharmacodynamics of S- and R-warfarin and effects on international normalized ratio: influence of body size, composition and genotype in cardiac surgery patients. *British Journal of Clinical Pharmacology*, 83(4):823–835, 2017.

Yu-Bin Zhu, Xian-Hua Hong, Meng Wei, Jing Hu, Xin Chen, Shu-Kui Wang, Jun-Rong Zhu, Feng Yu, and Jian-Guo Sun. Development of a novel individualized warfarin dose algorithm based on a population pharmacokinetic model with improved prediction accuracy for Chinese patients after heart valve replacement. *Acta Pharmacol. Sin.*, 38(3): 434–442, March 2017.

Appendix A. Model Priors

Time to max concentration values for patient j are drawn from a log normal distribution

$$t_{max,j} | \mu_t, \sigma_t \sim \text{LogNormal}(\mu_t, \sigma_t) \quad (6)$$

and α is drawn from a weakly informative beta prior to prevent degenerate cases when α is 0 or 1

$$\alpha_j \sim \text{Beta}(2, 2) . \quad (7)$$

The rate constants for patient j , $k_{e,j}$ and $k_{a,j}$, are determined from equations (4) and (5). The clearance rate is modelled hierarchically

$$Cl_j | \mu_{Cl}, \sigma_{Cl} \sim \text{LogNormal}(\mu_{Cl}, \sigma_{Cl}) . \quad (8)$$

Each patient is observed to have a non-zero concentration at time 0.5, so the time delay for each patient is no larger than 0.5 hours. We place a beta prior on the delay

$$\delta_j | \phi, \kappa \sim \text{Beta}(\phi/\kappa, (1 - \phi)/\kappa) \quad (9)$$

and multiply delta by 0.5 in our model to ensure the maximum delay is 0.5 hours. Here, ϕ is the mean of this beta distribution and κ determines the precision of the distribution. Shown in figure 7 is a Bayes net to exposit model structure at a high level.

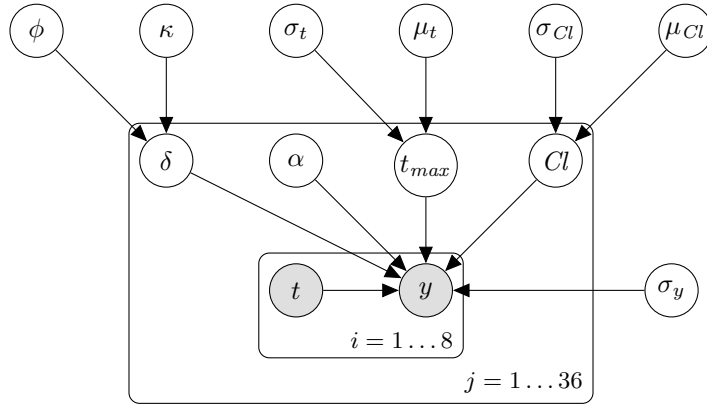


Figure 7: Graphical description of the data generating process for our model. The data consist of 36 patients, indexed by j . Each of the j patients are observed a total of 8 times, with each observation index by i . The data are generated by drawing random variables from their appropriate distribution at the top level and then drawing child random variables directly there after. As an example, ϕ and κ are drawn, which are then used to draw the δ_j , which are then used to draw each of the 8 concentration values, y_i for each of the j patients.

Priors for Model Hyperparameters

Estimates of the time to max concentration for apixaban place the population median t_{max} near 3.3 hours post dose (Byon et al., 2019). Assuming the median and the mean are similar, this provides information for μ_t and so we use specify

$$p(\mu_t) = \text{Normal}(\log(3.3), 0.25) \quad (10)$$

The standard deviation of the prior for μ_t was selected via prior predictive checks in which profiles are drawn and priors are assessed as realistic or not. We choose to err on the side of caution and inflate the uncertainty in this estimate to account for population differences between the measured patients in the data and the patients used in studies to determine the estimates of t_{max} . The population variability of t_{max} was modeled as

$$p(\sigma_t) = \text{Gamma}(10, 100) \quad (11)$$

Using these priors, we recover similar median, min, and max t_{max} values as reported by Byon et al. (2019). Similarly, we model the population mean and variability for the clearance rate as

$$p(\mu_{Cl}) = \text{Normal}(\log(3.3), 0.15) \quad (12)$$

$$p(\sigma_{Cl}) = \text{Gamma}(15, 100) \quad (13)$$

so that population estimates of the mean clearance rate are near 3.3 litres per hour with inflated uncertainty to account for possible population differences. We use weakly informative priors for ϕ and κ which induces an approximately uniform prior on δ .

$$p(\phi) = \text{Beta}(20, 20) \quad (14)$$

$$p(\kappa) = \text{Beta}(20, 20) \quad (15)$$

The tools used to measure the concentration of apixaban are believed to be within 10% of the real concentration. This implies that the observational model is heteroskedastic. We use a log-normal likelihood so that positivity of observed concentrations and heteroskedasticity are respected. We place a lognormal prior on the likelihood's variability with

$$p(\sigma_y) = \text{LogNormal}(\ln(0.1), 0.2) \quad (16)$$

$$C_j(t) | Cl_j, k_{a,j}, k_{e,j}, \delta_j \sim \text{LogNormal}(\ln(y(t)), \sigma_y) \quad (17)$$

Posterior Summarization and Generating New Data

Once our model was fit on the pharmacokinetic data, the marginal posteriors were summarized to create priors for the new model. Parameters for these priors were determined by using maximum likelihood on the posterior samples. The priors for the new model are as follows:

$$\mu_{Cl} \sim \text{Normal}(1.64, 0.09) \tag{18}$$

$$\sigma_{Cl} \sim \text{LogNormal}(-0.94, 0.11) \tag{19}$$

$$\mu_t \sim \text{Normal}(0.97, 0.05) \tag{20}$$

$$\sigma_t \sim \text{LogNormal}(-1.40, 0.12) \tag{21}$$

$$\alpha_j \sim \text{Beta}(2, 2) \tag{22}$$

$$\sigma_y \sim \text{LogNormal}(-1.76, 0.12) \tag{23}$$

Lognormal distributions were used to respect positivity of some parameters. The posterior predictive distribution of the model fit to the data from [Beaton et al. \(2018\)](#) was then used to simulate 100 pseudopatients. The time delay, δ was not used to generate these data as δ does not affect the overall shape of the concentration function, it merely shifts it right. The model with the priors defined by equations (18) to (23) was then refit on the 100 pseudopatients in order to examine differences between HMC and MAP in a “best case” scenario. The pseudopatients were sampled between 0.5 and 12.0 hours after ingestion in increments of 0.5. Draws from the posterior were used to predict latent concentration for each patient at times 0.75 to 11.75 in increments of 0.5.