Developing and Evaluating Pharmacokinetics-Driven Dynamic Personalized Medicine: A Framework and Case Study

A. Demetri Pananos, M.Math[1], Rommel G. Tirona Ph.D[2], Simon Bonner Ph.D[3], and Daniel J, Lizotte, Ph.D[1,4]

[1]Department of Epidemiology and Biostatsitics, Western University

[2]Department of Physiology and Pharmacology, Western University

[3]Department of Statistics and Actuarial Sciences, Western University

[4]Department of Computer Science, Western University

Developing and Evaluating Pharmacokinetics-Driven Dynamic Personalized Medicine: A Framework and Case Study

1 Introduction

Personalized medicine has been characterized by four goals: 1) to identify drugs for which between-subject variability in effectiveness or toxicity is a key issue for effective treatment, 2) to identify predictors which may explain this variability, 3) to decide on the right dose of the right drug by considering these factors, and 4) to prevent adverse reactions to drugs [1]. Progress in all four goals has accelerated within the last decade: For example, recent studies on DPYD genotype testing prior to starting fluoropyrimidine-based chemotherapy showed promise in preventing adverse events, making good arguments for integration of DPYD genotype testing into standard of care practices [2].

With regard to the third goal—personalized dosing—the intent of most efforts has been what we call *static* personalization. Such approaches inform dose at one point in time (usually induction) with the goal of eliminating the need for "trial-and-error" adjustments (titration) where the dose is adapted to the patient over time in response to its effects, both therapeutic and adverse [1]. Although significant progress has been made, for example in warfarin dosing [3], the need for titration has been reduced but not eliminated. Thus, there is an opportunity to personalize not only the initial doses but also the titration process to achieve the best result—we call this *dynamic* personalization. This approach has been used in other contexts by applying techniques from disciplines such as control theory, operations research, machine learning, and biostatistics to define and apply models for optimal sequential decision-making for patient care [4, 5].

Despite its potential to improve care, dynamic personalization imposes additional burden on the patient and provider, because it requires ongoing monitoring, for example by gathering lab results and returning for additional clinic visits. It is therefore natural to ask whether dynamic personalization is "worth it." Is the additional control over dose worth the additional burden? To help answer this question, we present a unified framework for the development and simulation-based evaluation of static and dynamic personalization based on pharmacokientic (PK) modelling. The knowledge created by our framework can be integrated into a system-level decision-making framework like Know4Go, for example, which can be used to evaluate whether such a personalized medicine program should be implemented into a particular health care system [6]. Having established our framework, we investigate the static and dynamic personalization of apixaban dosing as a case study.

We begin in Section 2 with an overview of dynamic treatment regimes, which underpin our models for dynamic personalization, and we review Bayesian PK modelling, which allow us to predict drug concentrations and to generate simulated patient data. In Section 3, we present our framework, which describes how to estimate optimal dynamic treatment regimes for personalization by combining Bayesian PK modelling with Q-learning, and describes a simulation-based approach for assessing the potential benefits of different modes of static and dynamic personalization. We then present our case study of personalized apixaban dosing in Section 4. Finally in Section 6 we discuss the results of the case study, and we identify broader issues relevant to the further development and implementation of PK-driven static and dynamic personalization.

2 Background

We briefly review dynamic treatment regimes, which are used to develop optimal decision-making models, and Bayesian PK models, which are used to capture relationships among patient characteristics, measurements, pharmacokinetics, and dose so that optimal dosing decisions can be derived using the dynamic treatment regime framework.

2.1 Dynamic Treatment Regimes

A *dynamic treatment regime* (DTR) is a mathematical formalism intended to model the practise of evaluating a patient, choosing a treatment, and observing a response. A DTR is defined as a sequence of decision rules $d = (d_1, \cdots, d_K)$, each of which is a function that takes information about a patient produces an *action*, like a dose initiation or change, that is intended to affect the status of the patient, like their plasma concentration [7, 8, 9]. The application of each decision rule is called a *stage* in the DTR, and applying a DTR generates a *trajectory* of data $O_1, A_1, O_2, A_2, ..., O_K, A_K, O_{K+1}$; these are are represented in upper case to emphasize that they are random variables which represent potentially noisy observations of a patient and actions which depend on the observations. We define the *history* of the patient at stage j to be $H_j = (O_1, A_1, O_2, A_2, \cdots, O_{j-1}, A_{j-1}, O_j)$; this encompasses all information available for decision-making at stage j.

2.1.1 Defining and Estimating Optimal DTRs

To define the performance of a decision rule (and of a DTR) we also define a *reward* $Y_j = Y_j(H_j, A_j, O_{j+1})$ which is a quantitative measure of success of the outcome that follows the stage j action, coded so that higher values are preferable. The sum of the rewards achieved over a single trajectory is called the *return*. Given this definition, the *value* of a DTR is given by

$$V^d = \mathbb{E}\left[\sum_{k=1}^{K} Y_k\right], \tag{1}$$

which is the expectation of the return if we follow DTR d. Typically, a DTR is defined to be optimal if it achieves the highest possible value among those under consideration; this corresponds to the concept of maximizing utility or minimizing expected loss in statistical decision theory [10].

There are different ways of estimating an optimal DTR [9]. One way, called "Q-learning" relies on estimating the optimal Q function. We give an overview of Q-learning for DTRs here, and refer the reader to other sources for more detail [7]. The optimal Q function at stage $j < K$ is a function of the observed history h_j and a proposed action a_j given by

$$Q_j^{\mathsf{opt}}(h_j, a_j) = \mathbb{E}\left[Y_j(h_j, a_j, O_{j+1}) + \max_a Q_{j+1}^{\mathsf{opt}}(H_{j+1}, a) | H_j = h_j, A_j = a_j\right]. \tag{2}$$

and $Q_K^{\mathsf{opt}}(h_K, a_K) = \mathbb{E}[Y_j(h_K, a_K, O_{K+1}) | H_K = h_K, A_K = a_K]$. The function Q_j^{opt} represents the expected return if we choose action a_j when history is h_j and subsequently always choose actions that are optimal,

that is, give highest expected return. Given the optimal Q function, an optimal DTR is given by choosing the action that maximizes it:

$$d_j^{\mathsf{opt}}(h_j) = \arg\max_{a \in \mathcal{A}} \left\{ Q_j^{\mathsf{opt}}(h_j, a) \right\} . \tag{3}$$

The Q-learning algorithm proceeds by first estimating Q_K^{opt}, often by acquiring a dataset of tuples of the form (h_K, a_K, y_K) and regressing the y_K on the h_K and a_K. The resulting $\hat{Q}_K^{\mathsf{opt}}$ can estimate the expected reward for any choice of h_K and a_K. It is then used to estimate Q_{K-1}, which is in turn used to estimate Q_{K-2}, and so on. This "backward induction" approach emphasizes that the optimal decision rule at earlier stages depends on the decision rules at later stages, and that they cannot in general be optimized independently.

2.2 Bayesian Models of Pharmacokinetics

In order to estimate the optimal Q functions, we need to be able to predict how a patient's concentration is likely to evolve over time in response to a hypothetical dose (action). Our approach is to build a Bayesian model of patient pharmacokinetics that can use baseline clinical information, as well as any available concentration measurements, to make tailored predictions of future concentrations that are as accurate as possible given the model structure and available data. The model is flexible in that it can condition on whatever information is available—for example, if previous dose and measurement information is not available for a specific patient, the model will rely on baseline information alone. If it is available, the model will use it to (hopefully) make improved predictions. This allows us to optimize both initial doses and later dose adjustments after additional information about concentration is acquired.

Bayesian models have another key property that we use in our framework. Once they are fit to data, and assuming the model is fit well, they are able to simulate the trajectories of patients drawn from a distribution that is similar to the distribution of the data that the models were trained on, but in the simulated data, *all* variables—including normally-hidden PK parameters—are fully observed. This allows us to conduct a form of internal validation where we use the simulated patients to assess the relative benefits of different modes of static and dynamic personalization, because we can know for each simulated patient exactly what the effect of any dose would be. This process is described in detail in the next section, where we present our framework, and the details of the Bayesian model itself are provided in Appendix A.

3 Framework

In this section, we present the components of our framework for assessing static and dynamic personalization, including details for fitting a hierarchical Bayesian PK model to concentration data from a cohort of patients, assessing the behaviour of Markov chains via diagnostics, and using the Bayesian model to generate simulated data for evaluation. We then outline several modes of static and dynamic personalization ranging from no personalization (every patient gets the same dose) to a complex dynamic mode of personalization (estimation

of the optimal Dynamic Treatment Regime for dosing). Finally, we outline steps for assessing the benefits of each mode of personalization.

3.1 Bayesian Modelling

The first step in our framework is to fit a Bayesian model that relates patient covariates and dose to drug concentration as a function of time. For example, previous work [11] describes a hierarchical Bayesian model of apixaban pharmacokinetics, in which the clearance Cl (L/hour), time to maximum concentration t_{max} (hours), absorption time delay δ (hours), and ratio between the elimination and absorption rate constants ($\alpha = k_e/k_a$, a unitless parameter) are hierarchically modelled. In our case study, we extend that model by regressing the latent pharmacokinetic parameters on baseline clinical variables (age, sex, weight, and creatinine) to permit personalization. The model could equally well be extended with pharmacokinetic or biomarker information if the relevant theory and data were available for a particular use case. We details for our Bayesian hierarchical pharmacokinetic model and information on interpretation of sampler diagnostics in appendix A.

3.2 Modes of Personalization & Assessment of Personalization

The second step in our framework is to identify modes of personalization that we wish to evaluate. We classify these modes of personalization into two types: static and dynamic personalization.

Static modes of personalization seek to inform the dose at one point in time (usually treatment initiation) with the goal of eliminating the need for "trial-and-error" adjustments. We consider two modes of static personalization in our case study:

1. **One size fits all**. This mode of personalization is not very personal at all. All patients receive the same dose size at the onset of treatment ($\approx 8.5mg$ taken twice daily). This dose was selected so that the average value across patients was maximized.

2. **Dose based on clinical variables**. In this mode of personalization, the patient's covariates, for example age, sex, weight, creatinine (a measure of kidney function)measurements, and possibly genetic or biomarker information, are provided to the pharmacokinetic model. A dose size is then selected using the model to maximize the value function conditional on these measurements.

Dynamic modes of personalization seek to personalize the initial doses but also the titration process. We consider four modes of dynamic personalization for our case study:

1. **One size fits all initial dose *and* one dose adjustment**. This mode of personalization provides patients the same dose to start, but requires a concentration measurement to be made sometime in the future, which is then used to adjust the dose. For example, in our case study, patients take their initial dose once every 12 hours with perfect adherence for five days, and, a sample is taken randomly in the

second 12 hour period of the fourth day. Our pharmacokinetic model conditions on this measurement, and the dose is adjusted in order to maximize the reward for another five days by updating our Bayesian model with the measurement.

2. **Initial dose based on clinical variables *and* one dose adjustment**. Here, the initial dose provided to the patient is determined by the patient's clinical measurements. For example, in our case study, in the second half of the fifth day, a concentration measurement is made at a random time. The model is conditioned on this concentration and the dose is adjusted to optimize the reward.

3. **Initial dose based on clinical variables *and* optimally-timed observation**. Similar to the previous mode of personalization, but the time at which the measurement is made is under our control and tuned to maximize reward. The time at which the sample is taken can yield more or less information about particular parameters in the model, but increases the burden by necessitating an additional constraint on when the observation should be obtained. For example, measuring much later after the dose is taken yields more information about the elimination rate constant k_e than it does about the absorption rate constant k_a because later in time, the majority of the dose has been absorbed and is now being eliminated by the body. In this mode of personalization, the initial dose and the timing of the adjustment are optimized independently.

4. **Optimal sequential dosing**. The approach of this mode is the same as the previous mode, except that the initial dose and the timing of the adjustment are *jointly optimized* using Q-learning to maximize the expected reward.

We stress that these are just examples of some modes of personalization, and that we do not mean that these modes should always be candidate modes for personalization, nor that they are the only modes of interest. These modes may not be appropriate for all drugs across all indications, and were selected in order to illustrate natural extensions and combinations of static personalization with additional information collection. A strength of our approach is that many possible modes of personalization may be considered depending on what is appropriate for the use-case at hand.

Each simulated patient has their dose(s) selected under each mode of personalization. Since the patients are simulated, we can compute what the return under the proposed dose(s) obtained from each mode of personalization and compare the return achieved to the largest return theoretically possible (i.e. the return achieved were we to know the pharmacokinetic parameters exactly when providing the initial dose, which is not possible in practice). The difference between this optimal return and the actual return is called the *regret*. Comparing the regrets achieved by different modes allows us to assess their relative performance, and helps us decide which are most appropriate for a given personalized medicine problem.

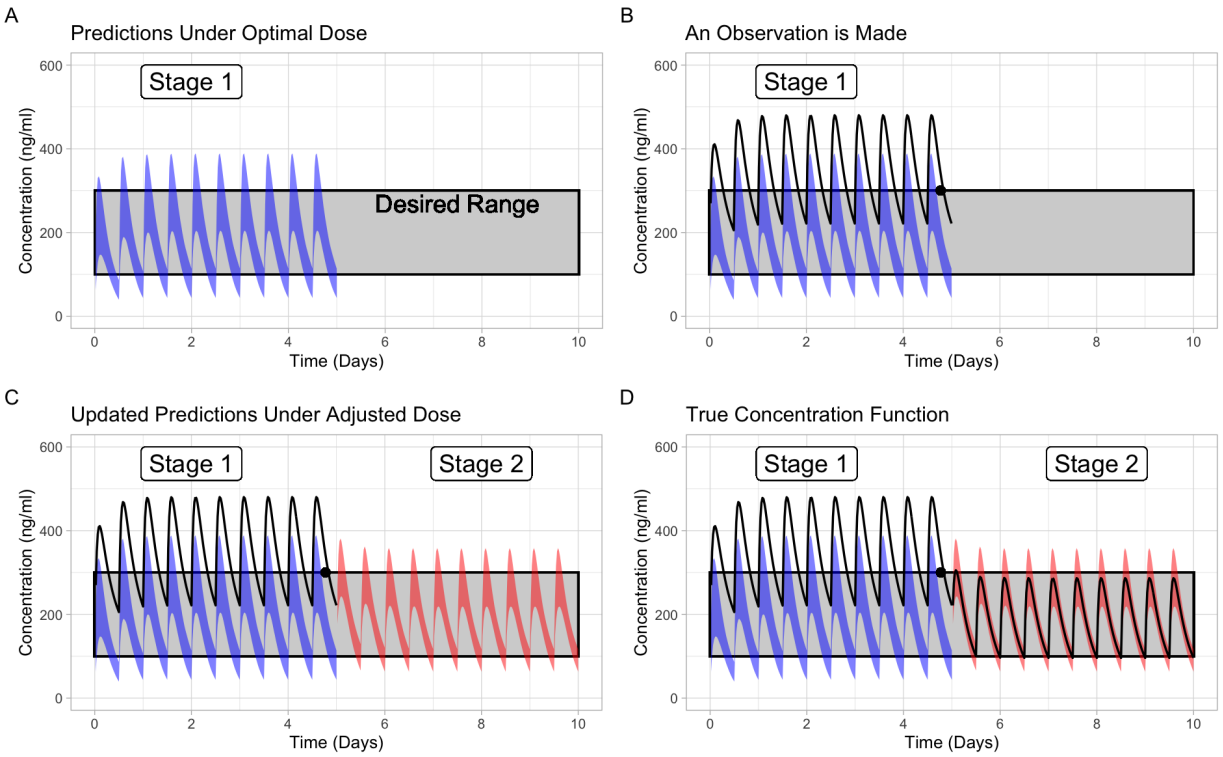

Figure 1: Visual representation of the steps in our framework. **A** Using only clinical information, the model is used to select a dose that is expected to keep the patient in the desired range for as long as possible. The blue ribbon indicates 90% credible interval for the latent concentration. **B** Some time later, the patient's blood serum concentration is measured (black dot). **C** The observation is incorporated into the model and a new dose is selected to keep the patient in range for as long as possible. The red ribbon indicates 90% credible interval for the latent concentration after adjusting the dose. **D** The black line indicates the true latent concentration under each dose. Note the observation (black dot) is not on the black line (true concentration) because there is measurement error, which the model accounts for.

4 Case Study

We present a case study of applying our framework to investigate the potential benefits of static and dynamic personalization of apixaban dosing. Apixaban is a direct acting anticoagulant medication used to treat active blood clots occurring with deep venous thrombosis or pulmonary embolism, or to prevent stroke in patients with atrial fibrillation. Prescribing an apixaban dose that achieves blood concentrations within an optimal range is expected to provide optimal treatment benefits while minimizing harms (e.g., serious bleeding). Clinical variables measuring age, weight, and kidney function are routinely used for dosing, and female sex, co-medications and genetic factors are known to contribute to higher circulating apixaban concentrations [12]. However, these variables only explain 35% of the pharmacokinetic variability in apixaban, which serves as rationale for considering dynamic dose optimization supported by post-initiation blood concentration monitoring.

4.1 Bayesian Modelling

To create the necessary model for apixaban personalization, we extend a previously proposed one-compartment Bayesian pharmacokinetic model [11] to include fixed effects of covariates on pharmacokinetic parameters in order to incorporate baseline clinical information (age, sex, weight, and creatinine.) Full details of the model structure, fitting, and diagnostic checks are provided in Appendix A. We fit the model to previously-collected data on apixaban concentration [13] and then use the fitted model to simulate patients with known "ground truth" pharmacokinetic parameters as described previously. We then use this population of simulated patients in our experiments to explore different modes of dose personalization and their relative benefits.

4.2 Modes of Personalization

We consider the 6 modes of personalization as outlined in section 3. To evaluate these modes of personalization, we generate 1000 simulated patients taking a dose of apixaban once every 12 hours with perfect adherence for a total of 10 days. The goal is to maximize the time spent with blood concentration level between between 100 ng/ml and 300 ng/ml. We choose this range as it is not so narrow that even optimal doses perform poorly, but not so wide that any dose can achieve high reward. For static modes of personalization, the selected initial dose is fixed over the 10 day period. For dynamic modes of personalization, some time in the second 12 hour period on the fourth day (between 108 and 120 hours after the initial dose), the simulated patient's blood concentration is measured, and then at the start of the fifth day, the dose is adjusted based on all the pre-dose clinical measurements plus the observed concentration by incorporating the new information into the Bayesian model.

4.2.1 Defining the Dynamic Treatment Regimes

To implement the two dynamic modes of personalization, we estimate DTRs with two stages (the first five days, and the latter five days). For the dynamic personalization policies our experiments, we develop a DTR for selecting the best dose for keeping a patient's blood plasma concentration within a desired range. In terms of the DTR, the system is the patient for whom a dose is selected, the actions correspond to selection of dose sizes (and a time in the future to sample the patient, should the DTR require that), and the reward is the proportion of time spent within the desired concentration range. The trajectories we will use to estimate the optimal Q functions are of the form

$$O_1, A_1, Y_1, O_2, A_2, Y_2 \tag{4}$$

The interpretation of a given trajectory is:

- O_1 is any pre-dose clinical measurements of the patient. In our experiments, we consider age in years, renal function (as measured by serum creatinine in mMol/L), weight in kilograms, and dichotomous biological sex (dummy coded so that male=1 and female=0). We choose these variables as they are known to affect the pharmacokinetics of apixaban [14].
- A_1 is the initial dose to provide the patient. If the DTR allows us to specify a time in the future at which to measure the patient's blood serum concentration, then A_1 is the dual action of initial dose plus a time in the future at which to measure.
- Y_1 is the proportion of time spent within the concentration range in the first five days.
- O_2 is the pre clinical measurements of the patient plus the observed concentration made on the fourth day.
- A_2 is the dose adjustment
- Y_2 is the proportion of time spent within the concentration range in the final five days after the dose adjustment.

The actions A_j affect the reward Y_j mediated by their effects on concentration. For example, a larger dose will elicit larger concentrations which may put the patient in range for longer (more reward) or take them out of range for some time (less reward). Thus, our reward function can be thought of as a composition of the reward function and the concentration function. In our experiments, we create a mesh of $2K$ times at which we can evaluate the latent concentration and compute the reward function. Each stage in our DTR consists of $K = 240$ times (equivalent to evaluating the latent concentration function every 30 minutes after ingestion). Let $c_i\ , i = 1...2K$, be the i^{th} latent concentration value at time t_i. The reward function in the first stage is

$$Y_1(H_1, A_1) = Y_1(c_1(A_1), \ldots, c_K(A_1)) = \frac{1}{K}\sum_{i=1}^{K} \mathbb{I}(0.1 < c_i(A_1) < 0.3) \tag{5}$$

Here, $\mathbb{I}$ is an indicator function returning 1 if c_i is between 100 ng/ml and 300 ng/ml and 0 else. The reward function in the second stage is

$$Y_2(H_2, A_2) = Y_1(c_{K+1}(A_2), \ldots, c_{2K}(A_2)) = \frac{1}{K}\sum_{i=1}^{K} \mathbb{I}(0.1 < c_{K+i}(A_2) < 0.3) \tag{6}$$

Our stage 2 optimal Q function is then

$$Q_2^{\mathsf{opt}}(H_2, A_2) = E\left[Y_2\left(c_{K+1}(A_2), \cdots, c_{2K}(A_2)\right) \middle| H_2, A_2 \right], \tag{7}$$

and our stage 1 optimal Q function is

$$Q_1^{\mathsf{opt}}(H_1, A_1) = E\left[Y_1\left(c_1(A_1), \cdots, c_K(A_1)\right) + \max_{a_2} Q_2^{\mathsf{opt}}(H_2, a_2) \middle| H_1, A_1 \right] \tag{8}$$

We seek to maximize the stage 1 optimal Q function to learn the optimal DTR for dosing patients under the constraint we can measure them at most once and are limited to the aforementioned pre-dose clinical variables. The interpretation of stage 1 optimal Q function is as follows: *Given the pre-dose clinical variables of the patient and a proposed initial dose and measurement time, the stage 1 optimal Q function gives the expected proportion of time the patient's blood serum concentration is between 100 ng/ml and 300 ng/ml assuming that we provide the patient with the best dose possible at the start of the* 5^{th} *day.* The decision rules which choose A_1 and A_2 to maximize these functions constitutes the estimated optimal DTR.

4.3 Evaluating Modes of Personalization

We measure the performance of different modes in terms of *regret*, the difference between theoretically largest possible return if the individual's PK parameters were precisely known and the achieved return by each mode of personalization. The results are shown in figure 2, ordered from least amount of information and burden (top) to most amount of information and burden (bottom) and colored by their personalization strategy (static or dynamic).

Modes of personalization which use less information have larger regret. The One Size Fits All approach (which uses no information about the patient) performs worst with a median regret of 0.145. The distribution of regrets for this mode is right skewed with some exceeding 0.95, meaning the patient could have been in range for nearly the entire time if the correct PK parameters were known, but the mode selected a dose which failed to put the patient in range.

The Clinical Variables mode nearly cuts the regret in half, achieving a median regret of 0.086 with smaller right skew. Modes which use observed concentration information (Clinical Variables + One Sample, Optimal

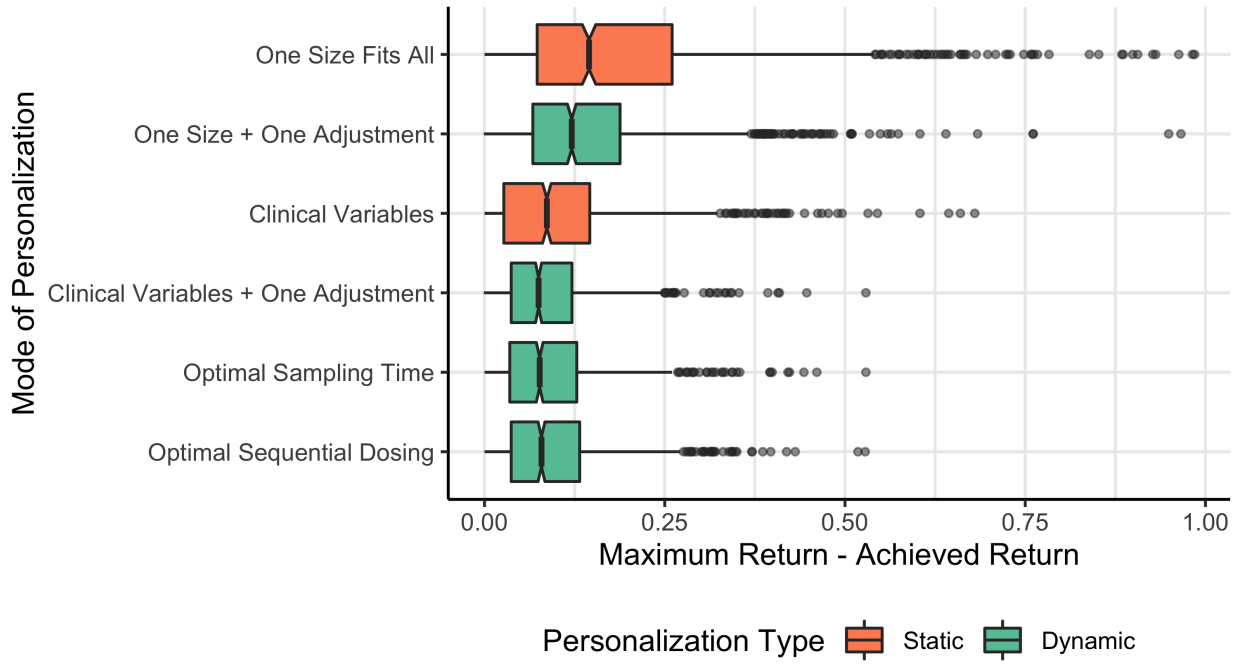

Figure 2: Boxplots of the *regret* – the difference between the largest possible return and the achieved return for each of the 1000 simulated patients. patients who achieve a return close to their maximum possible return have a regret near 0, while patients who achieve a return less than their maximum possible have larger regrets, with the largest possible regret being 1.

Sampling Time, and Optimal Sequential Dosing) lead to slightly lower median regrets (0.075, 0.076, 0.079 respectively) as compared to the Clinical Variables mode.

5 Limitations

Our framework relies on the quality of the Bayesian model used to predict plasma concentrations. The data required to construct a high quality Bayesian model of pharmacokinetics require multiple observations of a single patient over an extended time, preferably over multiple well timed doses with near perfect adherence. Obtaining such data requires well organized efforts and is high burden for both investigators and participating subjects. This makes acquiring a robust Bayesian model for use in dose personalization difficult. For this reason, our framework is not intended to replace prospective evaluation of personalized medicine programs; rather it is intended to estimate relative performance given existing evidence so that the most promising modes of personalization can be identified for potential implementation and further study.

6 Conclusions

As expected, modes of personalization that use more information result in lower regret (larger achieved rewards.) The static Clinical Variables mode balances relatively low implementation burden with high reward. However, there is right skew in the distribution of regret, with some 'outliers' who are at risk of obtaining less than half of their best possible return. If avoiding this risk is important, the Clinical Variables + One Adjustment mode may be preferable, even though it imposes additional burden.

The Optimal Sampling Time and Optimal Sequential Dosing modes use the same information as Clinical Variables + One Adjustment, but impose more burden on the patient and clinic. Both require samples be taken at specified times, and the second uses a more complex optimization procedure. Neither of these improved performance beyond the simpler modes in our study. There are two characteristics of our study that may explain this result. First, the clinical variables used are known to affect the pharmacokinetics of apixiban and are useful for predicting concentrations for static personalization, hence their use alone makes it possible to choose good doses. In other settings where clinical variables are not as predictive, we would expect dynamic personalization to have a bigger advantage. Second, the elimination rates k_e in our study were relatively high. This means that the effect of an initial dose on levels at the time subsequent doses are taken (i.e., a day later) is relatively small, so that doses can be optimized largely independently. If this were not the case, optimization using Q-learning would be expected to be more important to ensure that initial doses were not too large to successfully adapt later.

7 Implications

Any decision to implement personalized medicine must assess the costs and benefits of doing so. Despite the potential for personalized medicine to reduce health care costs [15, 16], the cost of patient testing and monitoring, personnel, and training required to operate a personalized medicine clinic carry a burden, and it is not yet clear in what circumstances personalized medicine is cost effective. In their 2019 scoping review of personalized medicine cost effectiveness, Kasztura et al [17] found that willingness-to-pay thresholds vary wildly from country to country (citing that cost per quality adjusted life year (QALY) for some modes of personalized medicine range from $20,000 USD per QALY in Europe and the United Kingdom to $200,000 USD per QALY in the United States). This high variability means the burden required for start up may result in a positive return on investment in some areas but not others. This variability should prompt would be adopters to more closely examine whether taking on the initial burden is worth the result.

Others have noted that much work on personalized medicine has not centred the needs, constraints, and utilities of the patient [18, 19]. Patients can be burdened by frequent followup for clinical measurement (as in the case with warfarin), be burdened by costly expenses related to obtaining care, or may be more risk adverse/tolerant than the "typical" patient. As an example, transportation has been found to be a large financial burden for patients receiving cancer treatment [20], and continues to burden patients, with a 2020

study finding that the cost of parking alone can climb as high as $1600 over the course of treatment in the United States [21]. Additional visits to a clinic have the potential to further burden patients by requiring them to miss a day of work, and find means of childcare during their absence (if necessary). Incorporating patient preferences and reducing the burden of personalization on the patient can result in sustained adherence [22], thereby increasing effectiveness and further preventing adverse events.

The complexity of balancing these burdens means that Implementation decisions made at the organizational level need to attend to a broad array of evidence and contextual factors. Know4Go [23] is one framework for explicitly considering factors from expanded domains of influence surrounding adoption of new technologies/interventions in a healthcare setting (like a clinic or hospital). These expanded domains of influence include: social, legal, ethical, environmental/institutional, political, entrepreneurial/innovative, research opportunity, and reversibility factors in conjunction with objective evidence of benefits versus risks, systematic review, and costs. Broadly, once evidence has been synthesized through systematic review and/or meta-analysis, the evidence is contextualized to local healthcare system perspective. Evidence is converted onto a benefit scale, derived from the number of patients likely to benefit from adoption of the technology/intervention. Budget impact of the adoption is estimated using costing data from the hospital/clinic, and new technologies can be triaged according to their impact and cost.

Policy decisions around personalization are complex. They carry burden for health systems and patients that vary widely by context. While there are frameworks like Know4Go for navigating these decisions, applying them requires quality evidence for the potential benefits of different kinds of personalization, even for deciding on potential pilot studies. Our framework can produce evidence for the potential effectiveness of a range of modes of personalization to inform organization-level decisions surrounding the investigation and implementation of personalized medicine programs that reduce cost, respect burden, and improve outcomes.

References

[1] Bridget L Morse and Richard B Kim. Is personalized medicine a dream or a reality? *Critical reviews in clinical laboratory sciences*, 52(1):1–11, 2015.

[2] Theodore John Wigle, Brandi Povitz, Wendy Teft, Robin Legan, John Gordon Lenehan, Markus Gulilat, Stephanie Nevison, Justin Kritzinger, Veera Punaganty, Denise Keller, et al. Prospective cohort study of the impact of hospital-wide dihydropyrimidine dehydrogenase (dpyd) genotype testing for fluoropyrimidine-based chemotherapy on adverse events and hospital costs. *American Society of Clinical Oncology*, 2019.

[3] Inna Y Gong, Rommel G Tirona, Ute I Schwarz, Natalie Crown, George K Dresser, Samantha LaRue, Nicole Langlois, Alejandro Lazo-Langner, Guangyong Zou, Dan M Roden, et al. Prospective evaluation of a pharmacogenetics-guided warfarin loading and maintenance dose regimen for initiation of therapy. *Blood, The Journal of the American Society of Hematology*, 118(11):3163–3171, 2011.

[4] Kristine Zhang, Yuanheng Wang, Jianzhun Du, Brian Chu, Leo Anthony Celi, Ryan Kindle, and Finale Doshi-Velez. Identifying decision points for safe and interpretable reinforcement learning in hypotension treatment, 2021.

[5] Barbara E Engelhardt, Niranjani Prasad, Li-Fang Cheng, Corey Chivers, Michael Draugelis, Kai Li, and Finale Doshi-Velez. The importance of modeling patient state in reinforcement learning for precision medicine. 2021.

[6] Janet Martin, Avtar Lal, Jessica Moodie, Fang Zhu, and Davy Cheng. *Hospital-Based HTA and Know4Go at MEDICI in London, Ontario, Canada*, pages 127–152. Springer International Publishing, Cham, 2016. ISBN 978-3-319-39205-9.

[7] Bibhas Chakraborty. *Statistical methods for dynamic treatment regimes*. Springer, 2013.

[8] Marie Davidian, Brian Everitt, Ron S. Kenett, Geert Molenberghs, Walter Piegorsch, and Fabrizio Ruggeri, editors. *Wiley StatsRef*, chapter Reinforcement Learning. Wiley, 2017. 3000 words.

[9] A.A. Tsiatis, M. Davidian, S.T. Holloway, and E.B. Laber. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2019. ISBN 9781498769785. URL `https://books.google.ca/books?id=FD0PEAAAQBAJ`.

[10] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[11] A Demetri Pananos and Daniel J Lizotte. Comparisons between hamiltonian monte carlo and maximum a posteriori for a bayesian model for apixaban induction dose & dose personalization. In *Machine Learning for Healthcare Conference*, pages 397–417. PMLR, 2020.

[12] 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.

[13] Rommel G Tirona, Zahra Kassam, Ruth Strapp, Mala Ramu, Catherine Zhu, Melissa Liu, Ute I Schwarz, Richard B Kim, Bandar Al-Judaibi, and Melanie D Beaton. Apixaban and rosuvastatin pharmacokinetics in nonalcoholic fatty liver disease. *Drug Metabolism and Disposition*, 46(5):485–492, 2018.

[14] Wonkyung Byon, Samira Garonzik, Rebecca A Boyd, and Charles E Frost. Apixaban: a clinical pharmacokinetic and pharmacodynamic review. *Clinical pharmacokinetics*, 58(10):1265–1279, 2019.

[15] Margot de Looff, Bob Wilffert, Cornelis Boersma, Lieven Annemans, Stefan Vegter, Job FM van Boven, and Maarten J Postma. Economic evaluations of pharmacogenetic and pharmacogenomic screening tests: a systematic review. second update of the literature. *PloS one*, 11(1):e0146262, 2016.

[16] Fatiha H Shabaruddin, Nigel D Fleeman, and Katherine Payne. Economic evaluations of personalized medicine: existing challenges and current developments. *Pharmacogenomics and personalized medicine*, 8:115, 2015.

[17] Miriam Kasztura, Aude Richard, Nefti-Eboni Bempong, Dejan Loncar, and Antoine Flahault. Cost-effectiveness of precision medicine: a scoping review. *International journal of public health*, 64(9): 1261–1271, 2019.

[18] Wolf Rogowski, Katherine Payne, Petra Schnell-Inderst, Andrea Manca, Ursula Rochau, Beate Jahn, Oguzhan Alagoz, Reiner Leidl, and Uwe Siebert. Concepts of 'personalization'in personalized medicine: implications for economic evaluation. *Pharmacoeconomics*, 33(1):49–59, 2015.

[19] Antonello Di Paolo, François Sarkozy, Bettina Ryll, and Uwe Siebert. Personalized medicine in europe: not yet personal enough? *BMC health services research*, 17(1):1–9, 2017.

[20] Peter S Houts, Allan Lipton, Harold A Harvey, Barbara Martin, Mary A Simmonds, Richard H Dixon, Santo Longo, Thomas Andrews, Robert A Gordon, John Meloy, et al. Nonmedical costs to patients and their families associated with outpatient chemotherapy. *Cancer*, 53(11):2388–2392, 1984.

[21] Anna Lee, Kanan Shah, and Fumiko Chino. Assessment of parking fees at national cancer institute–designated cancer treatment centers. *JAMA oncology*, 6(8):1295–1297, 2020.

[22] Rachel A Elliott, Judith A Shinogle, Pamela Peele, Monali Bhosle, and Dyfrig A Hughes. Understanding medication compliance and persistence from an economics perspective. *Value in health*, 11(4):600–610, 2008.

[23] Janet Martin, Avtar Lal, Jessica Moodie, Fang Zhu, and Davy Cheng. Hospital-based hta and know4go at medici in london, ontario, canada. In *Hospital-Based Health Technology Assessment*, pages 127–152. Springer, 2016.

[24] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo, 2018.

[25] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[26] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Burkner. Rank-normalization, folding, and localization: An improved r-hat for assessing convergence of mcmc. *arXiv preprint arXiv:1903.08008*, 2019.

[27] Andrew Gelman, Daniel Lee, and Jiqiang Guo. Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543, 2015.

A Appendix

A.1 Bayesian PK Model Details

The Bayesian model used to predict personalized concentration in response to dose, which we refer to as $\mathcal{M}_1$, is

$$y_{i,j} \sim \text{Lognormal}\left(C_i(t_j), \sigma_y^2\right) \tag{9}$$

$$\sigma^2 \sim \text{Lognormal}\left(0.1, 0.2\right) \tag{10}$$

$$C_i(t_j) = \begin{cases} \frac{D_i \cdot F}{Cl_i} \cdot \frac{k_{e,i} \cdot k_{a,i}}{k_{e,i} - k_{a,i}} \left(e^{-k_{a,i}(t_j - \delta_i)} - e^{-k_{e,i}(t_j - \delta_i)}\right) & t_j > \delta_i \\ 0 & \text{else} \end{cases} \tag{11}$$

$$k_{e,i} = \alpha_i \cdot k_{a,i} \tag{12}$$

$$k_{a,i} = \frac{\log(\alpha_i)}{t_{max,i} \cdot (\alpha_i - 1)} \tag{13}$$

$$\delta_i \sim \text{Beta}(\phi, \kappa) \tag{14}$$

$$\text{logit}(\alpha_i) | \beta_\alpha, \sigma_\alpha^2 \sim \text{Normal}(\mu_\alpha + \mathbf{x}_i^T \beta_\alpha, \sigma_\alpha^2) \tag{15}$$

$$\log(t_{max,i}) | \beta_{t_{max}}, \sigma_{t_{max}} \sim \text{Normal}(\mu_{t_{max}} + \mathbf{x}_i^T \beta_{t_{max}}, \sigma_{t_{max}}^2) \tag{16}$$

$$\log(Cl_i) | \beta_{Cl}, \sigma_{Cl} \sim \text{Normal}(\mu_{Cl} + \mathbf{x}_i^T \beta_{Cl}, \sigma_{Cl}^2) \tag{17}$$

$$p(\phi) \sim \text{Beta}(20, 20) \tag{18}$$

$$p(\kappa) \sim \text{Beta}(20, 20) \tag{19}$$

$$p(\mu_{Cl}) \sim \text{Normal}(\log(3.3), 0.15^2) \tag{20}$$

$$p(\mu_{t_{max}}) \sim \text{Normal}(\log(3.3), 0.1^2) \tag{21}$$

$$p(\mu_\alpha) \sim \text{Normal}(-0.25, 0.5^2) \tag{22}$$

$$p(\sigma_y) \sim \text{Lognormal}(\log(0.1), 0.2^2) \tag{23}$$

$$p(\sigma_{CL}) \sim \text{Gamma}(15, 100) \tag{24}$$

$$p(\sigma_{t_{max}}) \sim \text{Gamma}(5, 100) \tag{25}$$

$$p(\sigma_\alpha) \sim \text{Gamma}(10, 100) \tag{26}$$

$$p(\beta_{Cl,k}) \sim \text{Normal}(0, 0.25^2) \quad k = 1...4 \tag{27}$$

$$p(\beta_{t_{max},k}) \sim \text{Normal}(0, 0.25^2) \quad k = 1...4 \tag{28}$$

$$p(\beta_{\alpha,k}) \sim \text{Normal}(0, 0.25^2) \quad k = 1...4 \tag{29}$$

Here, normal distributions are parameterized by their mean and variance, lognormal distributions are parameterized by the mean and variance of the random variable on the log scale, and gamma distributions

	β_α	β_{Cl}	$\beta_{t_{max}}$
Age	-0.08 (-0.27,0.1)	0.01 (-0.06,0.08)	-0.01 (-0.1,0.08)
Creatinine	-0.06 (-0.25,0.14)	0.02 (-0.05,0.09)	-0.05 (-0.14,0.04)
Sex	-0.2 (-0.53,0.15)	0.39 (0.23,0.54)	-0.01 (-0.18,0.15)
Weight	0.32 (0.11,0.55)	0.2 (0.12,0.27)	0.09 (0.01,0.18)

Table 1: Posterior means for coefficients for each covariate in our pharmacokinetic model. In parantheses are 95% credible interval estimates.

are parameterized by their shape and rate. The μ in the model above represent population means on either the log or logit scale, the β are regression coefficients for the indicated pharmacokinetic parameter, the sigmas are the population level standard deviations on the log or logit scale, δ is aparameter which relaxes the assumption that the dose is absorbed into the blood immeditately upon ingestion, F is the bioavailability of apixiban (which we fix to 0.5 [14]) and D is the size of the dose in milligrams. All continuous variables were standardized using the sample mean and standard deviation prior to being passed to the model.

Once fit, $\mathcal{M}_1$ can be used to predict the pharmacokinetics of new patients, using the patient's covariates as predictors. To do so, the marginal posterior distributions for μ_{Cl}, $\mu_{t_{max}}$, μ_α, β_{Cl}, $\beta_{t_{max}}$, β_α, σ_{Cl}, $\sigma_{t_{max}}$, σ_α, and σ_y must be summarized. We use maximum likelihood on the posterior samples to summarize the marginal posterior distributions. We model the population means and regression coefficients as normal, and the standard deviations as gamma. The maximum likelihood estimates are used to construct priors for a new model, which we call $\mathcal{M}_2$. We construct $\mathcal{M}_2$ so as to be able to predict plasma concentration after multiple doses (of potentially different sizes) administered over time, and remove the time delay (δ) to simplify our simulations. Model priors for $\mathcal{M}_2$ are then

$$p(\mu_{Cl}) \sim \text{Normal}(0.5, 0.04) \tag{30}$$

$$p(\mu_{t_{max}}) \sim \text{Normal}(0.93, 0.05) \tag{31}$$

$$p(\mu_{\alpha}) \sim \text{Normal}(-1.35, 0.13) \tag{32}$$

$$p(\sigma_{Cl}) \sim \text{Gamma}(69.15, 338.31) \tag{33}$$

$$p(\sigma_{t_{max}}) \sim \text{Gamma}(74.96, 349.56) \tag{34}$$

$$p(\sigma_{\alpha}) \sim \text{Gamma}(10.1, 102.07) \tag{35}$$

$$p(\beta_{Cl,1}) \sim \text{Normal}(0.39, 0.08^2) \tag{36}$$

$$p(\beta_{Cl,2}) \sim \text{Normal}(0.19, 0.04^2) \tag{37}$$

$$p(\beta_{Cl,3}) \sim \text{Normal}(0.02, 0.04^2) \tag{38}$$

$$p(\beta_{Cl,4}) \sim \text{Normal}(0.01, 0.04^2) \tag{39}$$

$$p(\beta_{t_{max},1}) \sim \text{Normal}(-0.01, 0.08^2) \tag{40}$$

$$p(\beta_{t_{max},2}) \sim \text{Normal}(0.09, 0.05^2) \tag{41}$$

$$p(\beta_{t_{max},3}) \sim \text{Normal}(-0.05, 0.04^2) \tag{42}$$

$$p(\beta_{t_{max},4}) \sim \text{Normal}(-0.01, 0.04^2) \tag{43}$$

$$p(\beta_{\alpha,1}) \sim \text{Normal}(-0.19, 0.17^2) \tag{44}$$

$$p(\beta_{\alpha,2}) \sim \text{Normal}(0.33, 0.11^2) \tag{45}$$

$$p(\beta_{\alpha,3}) \sim \text{Normal}(-0.06, 0.1^2) \tag{46}$$

$$p(\beta_{\alpha,4}) \sim \text{Normal}(-0.09, 0.1^2) \tag{47}$$

$$\tag{48}$$

For our experiments, we generate the pharmacokinetic parameters of 1000 simulated patients from the prior predictive model of $\mathcal{M}_2$. Bayesian models are generative models, meaning they can generate pseudo-data by drawing random variables according to the model specification going from top (model priors) to bottom (model likelihood). To do so, we begin by resampling 1000 tuples of age, sex, weight, and creatinine from the dataset used to fit $\mathcal{M}_{\infty}$. We sample one draw of r μ_{Cl}, $\mu_{t_{max}}$, μ_{α}, β_{Cl}, $\beta_{t_{max}}$, and β_{α} from their respective prior distributions in $\mathcal{M}_2$. The values of these parameters remained fixed for all 1000 patients. Conditioned on the values of these mus and betas, we compute the expectation of the population distribution for each pharmacokinetic parameter by computing $\mu_{Cl} + \mathbf{x}^T\beta_{Cl}$, $\mu_{t_{\max}} + \mathbf{x}^T\beta_{t_{max}}$, $\mu_{\alpha} + \mathbf{x}^T\beta_{\alpha}$, where $\mathbf{x}^T$ is

the resampled tuple. From the prior distribution of M2, we sample one draw ofσ_{Cl}, $\sigma_{t_{max}}$, σ_α, and σ_y. These remained fixed for all 1000 patients. Using the previously computed expectations and σ, we sample 1000 tuples of pharmacokinetic parameters, one for each of the simulated patients. The clearance rate and time to max concentration were sampled assuming a lognormal distribution. Alpha was sampled using a logitnormal distribution. The pharmacokinetics can then be determined conditional on the pharmacokinetic parameters. Each of simulated patients' pharmacokinetic parameters remained fixed through the experiments. We simulate the latent concentration using $C(t)$ as written in $\mathcal{M}_2$, and can simulate observed concentrations by drawing a sample from a lognormal distribution with mean $\ln(C(t))$ and standard deviation σ_y

We use Stan, an open source probabilistic programming language, for fitting our Bayesian models via Hamiltonian Monte Carlo (a Markov Chain Monte Carlo technique) and computing markov chain diagnostics. Twelve chains are initialized and run for 2000 iterations each (1000 for warmup allowing the Markov chain the opportunity to find the correct target distribution and 1000 to use as samples from the posterior).

A.2 Diagnostics For Bayesian Models Fit Via MCMC

Once the form of the model is specified, creating simulated patients or estimating the PK parameters of a real patient requires computation of or sampling from the posterior distribution of the relevant variables given the relevant data. However, exact computation of the posterior distribution is intractable for all but very simple models, so Markov chain Monte Carlo (MCMC) techniques are often used to approximate the expectations with respect to the posterior distribution. Presently, the gold standard for generating samples from the posterior is Hamiltonian Monte Carlo (HMC), which works by generating a sequence of samples that "explores" the posterior distribution by solving a system of ordinary differential equations which describe the motion of an imaginary particle as it rolls along the surface of the log posterior density. Many implementations of HMC come with diagnostics which monitor the behaviour of the Markov chains that are used to generate samples and help to ensure that they are representative of the posterior distribution. That these Markov chains behave well is crucial, as any inferences about or from the model are obtained from samples generated by the chains. To assess the quality of the Markov chains, several diagnostics are commonly used including: number of divergences, the Gelman-Rubin convergence diagnostic, and effective sample size [24].

In practice, several Markov chains are used simultaneously to generate samples from the posterior. The chains are assessed with within-chain and between-chain diagnostics. First, individual chains may sometimes *diverge*. A divergence in a Markov chain indicates that the HMC Markov chain has encountered a region of high curvature in the posterior distribution which cannot be adequately explored. Consequently, Monte Carlo estimators of any expectations can be biased due to incomplete exploration of the posterior distribution. It is important that none of the Markov chains generated by HMC display a divergence, and that many chains (typically 4 or more) are initialized and are allowed to explore the posterior distribution.

Having ensured that no chains are diverging, a group-level diagnostic is used to assess whether all chains

have converged to the same limiting distribution. The *Gelman-Rubin (sometimes called $\hat{R}$) convergence diagnostic* is designed to detect if the Markov chains have converged to the same distribution by measuring the within-chain variance to the between chain-variance. In practice, $1.05 < \hat{R}$ indicates that there is poor mixing of the Markov chains and inference from the samples should not be performed lest the Monte Carlo estimators are biased by this poor mixing.

Even if the chains do not exhibit divergences and arrive at the same limiting distribution, the Markov chains could still exhibit high within-chain correlation, thereby increasing the uncertainty of estimation of key posterior quantities such as means, variances, or quantiles [25]. The *effective sample size* is a measure of how much the within chain autocorrelation increases uncertainty estimates. Presently, the guidance is that the effective sample size ratio should be larger than $100 \times$ (number of chains) [26].

In addition to monitoring divergences, Gelman-Rubin convergence diagnostics, and effective sample sizes, the model should be evaluated against existing domain knowledge. Evaluating that the model has learned appropriate behaviour (e.g. that as one quantity increases, another should decrease) can be performed by plotting model predictions. Additionally, *posterior predictive checks* – generating synthetic data from the model's posterior distribution and comparing against the real data – can be performed to ensure the model is not generating data which are physically impossible or completely unrealistic. Once the model is fit, important diagnostics indicate no pathological behaviour, and the model is deemed to fit the data sufficiently well, the model can then used to generate synthetic pharmacokinetic data for use in experiments to compare different forms of personalization. Each generated data point may be thought of as one synthetic patient, with observed covariates and observed pharmacokinetic parameters. These parameters, which are never observed in real data, allow us to compute the effects of any dosing decisions (which are made *without* direct knowledge of the parameters), and thus allow us to evaluate the performance of different modes of personalized dosing on the sampled population.

B Bayesian Model Diagnostics for Case Study

We fit our model to real pharmacokinetic data using the open source probabalistic programming language, Stan [27]. Stan monitors several Markov chain diagnostics, none of which detected problematic Markov chain behavior, which indicates that Stan's sampling algorithm was able to converge (0 divergences, all all Gelman-Rubin diagnostics$<$1.01, all effective sample sizes $>$ 2600).

The inclusion of covariates in the model results in a better fit than excluding them. Shown in figure 3 are the estimated random effects for the clearance pharmacokinetic parameter of each patient as a function of weight. Patient sex is indicated by color, the overall trend is shown in the black dashed line. Failing to include patient sex and weight results in males having on average a larger random effect than females of the same weight, and heavier patients having a larger random effect than lighter patients. When covariates are added into the model, the variation in the random effects attenuates, resulting in closer alignment to model

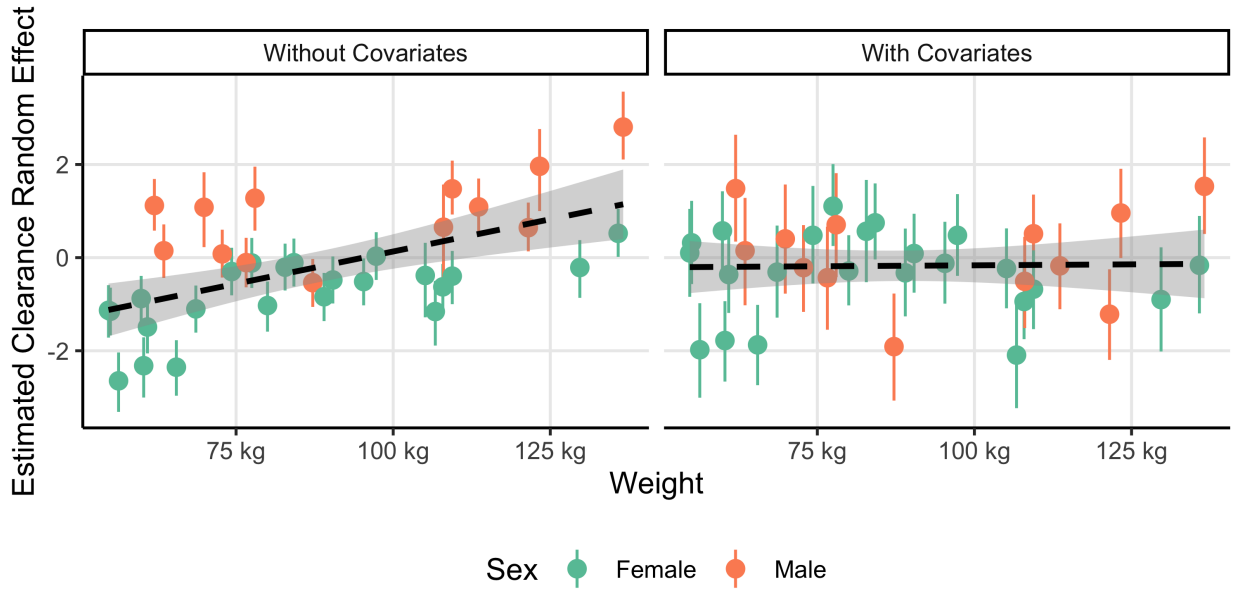

Figure 3: Random effects estimates for clearance CL_i and 95% credible intervals (left). Random effects estimates are colored by patient sex. Prior to adjusting for covariates, a general trend in weight can be seen in the random effects. Patients who are heavier tend to have larger random effect, and males tend to have larger random effects than females of the same weight. Patterns such as these indicate that weight and sex can be used to explain variation in the random effects. After adjusting for sex and weight (right), the random effects have no discernable pattern.

assumptions. A better fit to the data means data generated from the model may be closer aligned with the true data generating process.

Examining the posterior distributions of the regression coefficients provides further insights into the relationships between covariates and pharmacokinetics. Greater patient weight is associated with an increase in the expected value of alpha (which is used to compute the elimination and absorption rates in the first order one compartment PK model. The parameter α is the ratio of how fast the drug exits the central compartment how fast the drug enters the central compartment) which impacts the time to maximum concentration after each dose. There is an estimated effect of sex on α (males have smaller alpha than females, meaning the drug leaves their central compartment slower or enters the central compartment quicker), however the uncertainty is large (estimated effect -0.2 on the logit scale, 95% credible interval -0.53 to 0.15). See appendix A.1 in the Appendix for a full summary of the regression coefficients.

Model training error is comparable between the two models; the model without covariates achieves an average error of 8.31 ng/ml as measured by root mean squared error. The model with covariates achieves a root mean squared error of 8.36 ng/ml. Estimates of concentration uncertainty remain similar between the two models as well. We conclude the inclusion of covariates in the model improves model inferences but does not substantially improve the fit of the model in this case.

While prediction error and concentration uncertainty are comparable between the two models, the most important differences are between inter-individual uncertainty. The inclusion of the covariates explains variation between individual pharmacokinetic parameters, hence the between patient variability $\sigma_{Cl}, \sigma_{t_{max}}$ and σ_α are smaller in the covariate model as opposed to the no covariate model. This uncertainty effects decision making, as the no covariate model is more uncertain about the pharmacokinetics of new patients.