



**HAL**  
open science

# Hierarchical Bayesian pharmacometrics analysis of Baclofen for alcohol use disorder

Nina Baldy, Nicolas Simon, Viktor K Jirsa, Meysam Hashemi

► **To cite this version:**

Nina Baldy, Nicolas Simon, Viktor K Jirsa, Meysam Hashemi. Hierarchical Bayesian pharmacometrics analysis of Baclofen for alcohol use disorder. *Machine Learning: Science and Technology*, 2023, 4, pp.035048. 10.1088/2632-2153/acf6aa . inserm-04544203

**HAL Id: inserm-04544203**

**<https://inserm.hal.science/inserm-04544203>**

Submitted on 12 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PAPER • OPEN ACCESS

## Hierarchical Bayesian pharmacometrics analysis of Baclofen for alcohol use disorder

To cite this article: Nina Baldy *et al* 2023 *Mach. Learn.: Sci. Technol.* 4 035048

View the [article online](#) for updates and enhancements.

You may also like

- [Indoor location-aware medical systems for smart homecare and telehealth monitoring: state-of-the-art](#)  
Fendy Santoso and Stephen J Redmond
- [Multifractal analysis of financial markets: a review](#)  
Zhi-Qiang Jiang, Wen-Jie Xie, Wei-Xing Zhou et al.
- [An evaluation and comparison of commercial driver sleepiness detection technology: a rapid review](#)  
Jennifer M Cori, Jessica E Manousakis, Sjaan Koppel et al.



## PAPER

## OPEN ACCESS

RECEIVED  
9 March 2023REVISED  
4 August 2023ACCEPTED FOR PUBLICATION  
4 September 2023PUBLISHED  
14 September 2023

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



# Hierarchical Bayesian pharmacometrics analysis of Baclofen for alcohol use disorder

Nina Baldy<sup>1</sup>, Nicolas Simon<sup>2,3</sup>, Viktor K Jirsa<sup>1,\*</sup> and Meysam Hashemi<sup>1,\*</sup> <sup>1</sup> Aix Marseille Univ, INSERM U1106, INS, Inst Neurosci Syst, Marseille, France<sup>2</sup> Aix Marseille Univ, INSERM U912, SESSTIM, Marseille, France<sup>3</sup> Consultation d'addictologie UF6525, Hôpital de la Timone, APHM, Marseille, France

\* Authors to whom any correspondence should be addressed.

E-mail: [viktor.jirsa@univ-amu.fr](mailto:viktor.jirsa@univ-amu.fr) and [meysam.hashemi@univ-amu.fr](mailto:meysam.hashemi@univ-amu.fr)**Keywords:** Hierarchical Bayesian inference, adaptive Monte Carlo sampling, convergence diagnostics, prior information, model selection and validation, population pharmacokinetic analysis, alcohol use disorder

## Abstract

Alcohol use disorder (AUD), also called alcohol dependence, is a major public health problem, affecting almost 10% of the world's population. Baclofen, as a selective GABA<sub>B</sub> receptor agonist, has emerged as a promising drug for the treatment of AUD. However, the inter-trial, inter-individual and residual variability in drug concentration over time in a population of patients with AUD is unknown. In this study, we use a hierarchical Bayesian workflow to estimate the parameters of a pharmacokinetic (PK) population model from Baclofen administration to patients with AUD. By monitoring various convergence diagnostics, the probabilistic methodology is first validated on synthetic longitudinal datasets and then applied to infer the PK model parameters based on the clinical data that were retrospectively collected from outpatients treated with oral Baclofen. We show that state-of-the-art advances in automatic Bayesian inference using self-tuning Hamiltonian Monte Carlo (HMC) algorithms provide accurate and decisive predictions on Baclofen plasma concentration at both individual and group levels. Importantly, leveraging the information in prior provides faster computation, better convergence diagnostics, and substantially higher out-of-sample prediction accuracy. Moreover, the root mean squared error as a measure of within-sample predictive accuracy can be misleading for model evaluation, whereas the fully Bayesian information criteria correctly select the true data generating parameters. This study points out the capability of non-parametric Bayesian estimation using adaptive HMC sampling methods for easy and reliable estimation in clinical settings to optimize dosing regimens and efficiently treat AUD.

## 1. Introduction

Alcohol use disorder (AUD) is a major public health problem, affecting almost 10% of the world's population (Schuckit 2009). The harmful use of alcohol is responsible for 5.1% of the global burden of disease, resulting in more than 200 diseases, injuries and other health conditions, and contributing to 3 million deaths globally each year, i.e. 5.3% of all deaths according to World Health Organization (2022). Beyond health consequences, this brain disorder brings significant mental and behavioral symptoms, as well as social and economic losses to individuals and society at large.

Baclofen is a selective agonist of the Gamma-Aminobutyric Acid type B receptors (GABA<sub>B</sub>), which may exert an inhibitory effect on dopaminergic neurons (Doherty and Gratton 2007). It was originally marketed as a muscle relaxant for the treatment of neurological-induced spasticity. Various experiments have shown a positive effect of Baclofen on AUD in rodents and non-human primates (see Colombo *et al* 2018 for a review). Although there is no clear evidence on the dosing, efficacy, safety, and ideal duration of Baclofen treatment for AUD, clinical studies have shown its promise as a medication for patients with moderate to severe AUD (Agabio *et al* 2002, Garbutt *et al* 2021). Preclinical experiments have also shown its efficacy in

reducing alcohol withdrawal syndrome (Brennan *et al* 2013) and voluntary alcohol consumption (Colombo *et al* 2004). The optimal dosage of Baclofen varies according to individuals and has not been well established with a general agreement. It is mainly prescribed to patients for whom the existing treatments have not been effective, or in those with liver disease due to its minimal damage to the liver. See de Beurepaire *et al* (2019), for a review of the use of Baclofen as a treatment in human AUD. Although in recent years, Baclofen has been used to reduce craving, voluntary alcohol intake and withdrawal syndrome of alcoholic patients, a wide inter-individual variability has been observed, and the potential high risk of sedation is unknown (Imbert *et al* 2015, Simon *et al* 2018). Thus, the precise efficacy of dosing and the ideal duration of treatment in different AUD patient groups need to be evaluated.

Mathematical modeling is widely used in many areas of science to learn about the data generation process, make predictions on outcomes (unseen data), and to justify hypotheses. In the modeling framework, differential equations provide us with the natural evolution of the system under study at any given time point. They are often used to describe the principles that govern the dynamics of the system, allowing us to adequately describe the processes involved and make quantitative predictions, such as the rate of change of drug concentration and clinical efficacy. In this study, we use a population pharmacokinetic (PK) model of the Baclofen effect on patients with AUD to estimate the variability in drug concentration over time in a population of patients. The PK analysis has been widely used in clinical development, with several applications, for instance, to improve the understanding of the *in vivo* behavior of complex delivery systems, as it allows for the separation of drug-, carrier-, and pharmacological system-specific parameters (Sheiner and Ludden 1992, Meibohm and Derendorf 1997, Bonate 2011, Upton and Mould 2014). Analysis of PK model is currently an indispensable component of drug discovery, by ensuring that robust evidence from preclinical models closely shapes the design of clinical studies (Zou *et al* 2020). Although PK modeling has facilitated the drug development process, the accurate and reliable estimation of its parameters from noisy data is a major challenge in conducting clinical research to determine the safety and efficacy of a drug within a particular disease or specific patient population. To estimate unknown quantities, optimization methods within the frequentist approach are often used in practice, in which an objective (or cost) function is defined to score the performance of the model by comparing the observed and predicted values. However, such a parametric approach results in only a point estimation, and the optimization algorithms may easily get stuck in a local maximum, requiring multi-start strategies to address the potential multi-modalities. Moreover, the estimation depends critically on the form of the objective function defined for optimization, and the models involving differential equations often have non-identifiable parameters (Hashemi *et al* 2018).

In this study, we use the Bayesian approach to address these challenges in the estimation of a population PK model's parameters from synthetic data (generated by known values for validation purposes), and then from routine clinical data that were retrospectively collected from 67 adult outpatients treated with oral Baclofen. The Bayesian framework is a principled method for inference and prediction with a broad range of applications, wherein the uncertainty in parameter estimation is naturally quantified through probability distribution placed on the parameters, which are updated with the information provided by data (Bishop 2006, Gelman *et al* 2014). Such a probabilistic technique provides the full posterior distribution of unknown quantities in the underlying data generating process, given only observed responses and the existing information about uncertain quantities expressed as a prior probability distribution (Ferreira *et al* 2020, Hashemi *et al* 2020, 2021). In other words, Bayesian inference provides all plausible parameter ranges that are consistent with observation by integrating information from both domain expertise and experimental data.

In the context of clinical trials, the frequentist approach utilizes prior information (based on evidence from previous trials) only in the design of a trial, not in the analysis of the data (Gupta 2012, Jack and Chu 2012). In contrast, the Bayesian approach provides a formal mathematical framework to combine prior information with available information at the design, conduct, and analysis stages of the experiment (Spiegelhalter *et al* 1999, Berry 2006, Yarnell *et al* 2021). To conduct Bayesian data analysis, Markov chain Monte Carlo (MCMC) methods have often been used to sample from and, hence, approximate the exact posterior distributions. However, MCMC sampling in high-dimensional parameter spaces, which converge to the desired target distribution is non-trivial and computationally expensive (Betancourt 2014, 2017). In particular, the use of differential equations, such as PK population models, together with noise in data raises many convergence issues (Hashemi *et al* 2020, Grinsztajn *et al* 2021, Jha *et al* 2022). Designing an efficient MCMC sampler to perform principled Bayesian inference on high-dimensional and correlated parameters remains a challenging task. Although Bayesian inference requires painstaking model-specific derivations and hyper-parameter tuning, probabilistic programming languages such as Stan (Carpenter *et al* 2017) provide high-level tools to reliably solve complex parameter estimation problems. Stan (see <https://mc-stan.org>) is a state-of-the-art platform for high-performance statistical computation and automatic Bayesian data analysis (Gelman *et al* 2020), which provides advanced algorithms (Hoffman and Gelman 2014), efficient gradient

computation (Margossian 2018), and numerous diagnostics to check the convergence of sampling (Vehtari *et al* 2021), hence, the reliability of inference and prediction.

In the present work, to estimate the full posterior distribution of a population PK model's parameters, we use a self-tuning variant of Hamiltonian Monte Carlo (HMC) in Stan. This algorithm adapts the parameters of the HMC, making the sampling strategy more efficient and automated (Hoffman and Gelman 2014). The Bayesian setting in this study, which is first validated on synthetic data according to various diagnostics, enables us to efficiently and accurately estimate the effect of Baclofen on retrospective patients with AUD. Importantly, we demonstrate that leveraging the level of information in priors provide faster computation, better convergence diagnostics, and substantially higher out-of-sample prediction accuracy. Moreover, we show that the root mean squared error as a measure of within-sample predictive accuracy can be misleading for model evaluation and selection. In constrats, the fully Bayesian information criteria (such as Watanabe–Akaike information criterion; Gelman *et al* 2014) correctly select the true data generating parameters. This work may pave the way to reliably predicting the efficacy of treatment drugs from longitudinal patient data in a fully Bayesian setting, thus optimizing strategies for clinical decision-making, especially in brain disorders.

## 2. Materials and methods

### 2.1. Clinical data

In this study, we used the routine clinical data retrospectively collected from 67 adult outpatients (43 men and 24 women, with weights ranging from 42 kg to 128 kg and ages ranging from 29 to 68 years old), with AUD treated with oral Baclofen, in the Department of Addictology at Sainte Marguerite Hospital in Marseille, France. Ethics committee approval and patient consent were not compulsory in France for the use of retrospective therapeutic drug monitoring data, before *Loi Jardé* (n° 2012-300 du 5 Mars 2012 relative aux recherches impliquant la personne humaine). Our data were collected before 2012, thus they are retroactively considered as *Hors Loi Jardé*. In October 2018, France became the first country to officially approve Baclofen for AUD (Rolland *et al* 2020). The French Medicines Agency (ANSM) has granted marketing authorization for Baclofen to support drinking reduction in AUD, which is a major and insufficiently addressed public health issue, with more than 60 000 patients treated with Baclofen in France (Rolland *et al* 2020). The data has been thoroughly described and published in a previous study by Imbert *et al* (2015).

In brief, Baclofen was administered orally in conventional tablet form and the dosing schedule consisted of a tailored and variable dose for each patient. All patients met the DSM-5 criteria for AUD (American Psychiatric Association, 2013). Therapeutic drug monitoring enabled the measurement of plasma concentrations. The treatment was conducted similarly for all patients, that is, doses were progressively increased (subject to a good tolerance to the treatment) from an initial dose of 15 mg/d (5 mg three times per day [t.i.d]) for 1 week to reach the target dose of 180 mg at 3 months (representing a Baclofen increase of 15 mg every week). The clinical event schedule (dosing and measurements) of each subject is specified according to the Nonlinear Mixed Effects Modeling software (NONMEM, Beal *et al* 2009) conventions. The number of observations is subject-dependant, with a total of 427 recordings. For each patient, the number of measurements is comprised of between 1 and 16 observations (median = 6). The total follow-up time ranges between 1 and 613 days (average is 185 days). The following data were collected for each patient: drug administration data with the patient's dosage regimen of Baclofen, serum concentration of Baclofen with the date and time of each measurement, demographic data (age, body weight, height, gender), biological data (creatinine, urea, hepatic transaminases aspartate aminotransferase/alanine aminotransferase, albumin, prothrombin ratio, fibrinogen, mean corpuscular volume, bilirubin, alkaline phosphatase, carbohydrate-deficient transferrin, gamma-glutamyl transferase) and alcohol and tobacco consumption. The planned visits were every week the first month then at months 2 and 3 following the treatment initiation. At each visit, routine psychological support counseling was provided by the same professional staff and craving level for alcohol was assessed using the obsessive compulsive drinking scale (OCDS). According to a one-compartment PK population model with first-order elimination (ADVAN2 TRANS2), none of the covariates tested was able to improve the fit using an exponential model, or decrease the intersubject variability and the base model, and previous analysis showed no formal bias (Imbert *et al* 2015).

### 2.2. Synthetic data

Using synthetic data for fitting allows us to validate the inference process as we know the ground-truth of the parameters being estimated. Therefore, we can use standard error metrics to measure the similarity between the inferred parameters and those used for data generation. Synthetic data were generated following the same event schedule as empirical clinical data and using a one-compartment population model with first-order absorption (see equation (1)). The synthetic data was generated using the R package *mrgsolve* (Baron 2022),

which enables simulation from ODE-based hierarchical models, including random effects and covariates. The `mrgsolve` package (Elmokadem *et al* 2019) uses Livermore Solver for Ordinary Differential Equations (LSODE; Radhakrishnan *et al* 1993) of the Fortran ODEPACK library (Hindmarsh 1992) integrated in R through the `Rcpp` (Eddelbuettel and Francois 2011) package.

### 2.3. Population model for pharmacometrics

PK models are currently used to predict clinical response in humans through understanding the mechanism of action and disease behavior for a given drug. Inference on the PK model from data can optimize the design of clinical trials, provide an appropriate dose range, anticipate the effect in certain subpopulations, and better predict drug–drug interactions. The PK population models invoke nonlinear mixed-effects models that allow for the quantification of exposure–response relationships on both a global population and individual level (Bonate 2011, Owen and Fiedler-Kelly 2014, Keizer *et al* 2018). Such a modeling approach provides a flexible tool for hypothesis testing and refining assumptions, e.g. the random phenomena underlying the exposure–response mechanism by including group-specific (in this case, subject-specific) effects within a population. This model also allows for the inclusion of covariates whose influence on inter-group variations can be inferred.

In the present work, drug concentration in the plasma is modeled by a one-compartment model with first-order absorption

$$\begin{aligned}\frac{dy_{\text{dosing}}}{dt} &= -k_a y_{\text{dosing}}, \\ \frac{dy_{\text{central}}}{dt} &= k_a y_{\text{dosing}} - \frac{CL}{V} y_{\text{central}}.\end{aligned}\quad (1)$$

The above linear ODE system describes the evolution of the amount  $y$  of the drug in both dosing compartment<sup>4</sup> and the central compartment (plasma). The first-order kinetics parameter  $k_a$  ( $\text{h}^{-1}$ ) is the absorption constant of the drug. The parameter  $V$  (L) is the volume of distribution, which represents the hypothetical volume in which the drug would need to be diluted to the same level as in plasma. The clearance  $CL$  (L/h) controls the elimination of the drug from the organism, and such a linear ODE system has an analytical solution, as we used in this study. Let  $y = (y_{\text{dosing}}, y_{\text{central}})^T$ , the linear system given by equation (1) can be expressed in matrix form

$$\frac{dy}{dt} = \mathbf{A}y, \quad \text{with } \mathbf{A} = \begin{pmatrix} -k_a & 0 \\ k_a & -\frac{CL}{V} \end{pmatrix}.\quad (2)$$

Solutions of this homogeneous linear system at time  $t$  are given by

$$y(t) = e^{t\mathbf{A}}y_0,\quad (3)$$

where  $y_0 = y(0)$  is the initial condition and  $e^{t\mathbf{A}}$  denotes the matrix exponential<sup>5</sup>.

### 2.4. Bayesian modeling

In Bayesian modeling, all model parameters are treated as random variables and the values vary based on the underlying probability distribution (Bishop 2006). That is, in the case of a one-compartment PK model, kinetic parameters  $CL$ ,  $V$  and  $k_a$  will be interpreted as random variables, and we aim to infer their probability distributions based on *prior* knowledge (e.g. derived from physiological information or previous evidence), updated with available information in the observed data through the so-called *likelihood* function, i.e. the probability of some observed outcomes given a set of parameters. Although the likelihood function can provide the best-fit points (maximum likelihood estimators), we are interested in the whole posterior distribution over unknown model parameters, as it contains all relevant information about parameters after observing data to perform a reliable model selection (Gelman *et al* 2014, Hashemi *et al* 2021).

Formally, given data  $y$  and model parameters  $\theta$ , Bayes rule gives the posterior distribution as

$$p(\theta | y) = \frac{p(\theta)p(y | \theta)}{p(y)}\quad (4)$$

that combines and actualizes prior knowledge on parameters (before seeing data) with knowledge acquired from observed data (through likelihood function). The denominator  $p(y) = \int p(y | \theta)p(\theta)d\theta$  represents the

<sup>4</sup> For oral dosing, doses are added in the dosing compartment; infusion and IV bolus injections must be added directly in the central compartment.

<sup>5</sup> Matrix exponential  $e^{t\mathbf{A}}$  is defined as  $e^{t\mathbf{A}} = \sum_{k=0}^{\infty} \frac{(t\mathbf{A})^k}{k!}$ . Stan features a matrix exponential function `matrix_exp`.

probability of the data and it is known as evidence or marginal likelihood. In the context of inference, this term amounts to simply a normalization factor, thus equation (4) is reduced to a proportionality relation

$$p(\theta | y) \propto p(\theta)p(y | \theta). \quad (5)$$

The Bayesian approach applied to PK analysis provides a fully probabilistic description of unknown quantities, allowing not only a straightforward interpretation of the inferred parameters and outcomes, but also the modeling of uncertainty about the inferred values of these quantities. Moreover, it provides us with a principled method for model comparison, selection, and decision-making by measuring the out-of-sample model predictive accuracy (i.e. the measure of the model's ability in new data prediction). This metric evaluates the generalizability of a model, as it provides an estimate of how well the model is likely to perform when used to make predictions on unseen data. To assess the out-of-sample predictive accuracy, we used Watanabe–Akaike (or widely available) information criterion (WAIC; Gelman *et al* 2014). This metric is a fully Bayesian information criterion that utilizes the entire posterior distribution, thus allowing us to incorporate our prior knowledge into the model selection. Following Gelman *et al* (2014), WAIC is given by:

$$\text{WAIC} = -2(\text{lppd} - p_{\text{eff}}), \quad (6)$$

where *lppd* is the log pointwise predictive density for a new data point (as the accuracy term<sup>6</sup>), and  $p_{\text{eff}}$  is the effective number of parameters (as penalty term to adjust for overfitting)<sup>7</sup>. In practice, we can replace the expectations with the average over the draws from the full posterior to calculate WAIC (for more details see Gelman *et al* 2014, Hashemi *et al* 2021). Note that WAIC uses the whole posterior distribution rather than the point estimation used in the classical information criteria such as AIC and BIC. Finally, the relative difference in WAIC is used to measure the level or the strength of evidence for each candidate model under consideration. The lower value of WAIC indicates a better model fit. Following Burnham and Anderson (2002), Hashemi *et al* (2021), a relative difference larger than 10 between the best model (with minimum WAIC) and other candidate models indicates that an alternative model is very unlikely; that is, an alternative model has essentially no support.

## 2.5. MCMC sampling

Monte Carlo (MC) sampling is a family of computational algorithms used for uncertainty quantification by drawing random samples from distributions, in which the sampling process does not require knowledge of the entire distribution. Markov chain Monte Carlo (MCMC) methods construct Markov chains (sequences of probabilistically linked states, with the probability of transitioning to a given state depending only on the current state) that have a desired stationary probability distribution, to extract samples from this distribution (Gelman *et al* 2014). Using MCMC, the samples are obtained by retrieving the history of states visited by the chain; if the chain is run for long enough, these samples will originate from the stationary distribution (independent of the initial states). From these samples, we can construct MC estimates describing the target probability distribution, such as mean and quantiles; the sample average approximates an expectation with respect to the stationary distribution of the chain. In this context, Monte Carlo standard errors (MCSE) provide an indication of the quality of reported estimates, that is, a quantification of the estimation noise. The MCSE of an estimator  $\hat{\theta}_n$  is given by the posterior standard deviation divided by the square root of the effective sample size (ESS). In addition to sample mean and sample standard deviation, MCMC methods provide the estimated error (standard deviation) in the posterior mean estimation, on the scale of the parameter value (Carpenter *et al* 2017).

## 2.6. Adaptive HMC sampling using Stan

Stan (<https://mc-stan.org>) is an open-source statistical tool that implements automatic gradient-based algorithms for Bayesian modeling and probabilistic machine learning (Carpenter *et al* 2017). Hamiltonian Monte Carlo (HMC) is a powerful MCMC method that uses the derivatives of the density function being sampled to generate efficient transitions that explore the whole posterior distributions in complex probabilistic models (Duane *et al* 1987, Neal 2011). However, the performance of HMC is highly sensitive to the algorithmic parameters, such as the step size and the number of steps in the leapfrog integrator. Stan provides a self-tuning variant of the HMC, the so-called No U-Turn Sampler (Hoffman and Gelman 2014), which uses a recursive algorithm to eliminate the need for setting the hyperparameters, i.e. it adaptively tunes

<sup>6</sup> Log predictive density or log-likelihood is proportional to the mean squared error if the model is normal with constant variance.

<sup>7</sup> The effective number of parameters is calculated by summing over all the posterior variance of the log predictive density for each data point.

the HMC algorithm without requiring the user intervention (Hoffman and Gelman 2014, Betancourt 2017). The NUTS algorithm avoids random walk behavior and sensitivity to correlated parameters to construct efficient Markov chain exploration of the distribution's *typical set*, that is, the set that concentrates the majority of the volume-density trade-off (Betancourt 2017).

Starting from an initial value (random or defined by the user), the NUTS algorithm updates parameters  $\theta$  through a series of iterations by evolving them according to Hamiltonian dynamics and submitting them to a Metropolis proposal. More precisely, HMC transforms the problem of sampling from a target distribution into the problem of simulating Hamiltonian dynamics; it artificially introduces an auxiliary momentum  $\rho$  to suppress random walk behavior, which extends the representation of the target parameter space into a phase space of joint parameters  $(\rho, \theta)$ . The auxiliary momentum is sampled from the conditional distribution  $\pi(\rho|\theta)$ . In Stan implementation, it is sampled from a multivariate normal distribution  $\mathcal{N}$  that is independent of  $\theta$ :

$$\rho \sim \mathcal{N}(0, M), \quad (7)$$

with covariance matrix  $M$  estimated as the inverse of posterior covariance during a *warm-up* phase. The joint phase space parameters  $(\rho, \theta)$  are evolved through Hamiltonian dynamics (the expression of the Hamiltonian is given in appendix A.1) by solving the Hamiltonian equations of motion (see appendix A.2). To do so, Stan uses the leapfrog integrator (see appendix A.3). The output of the integrator  $(\rho^*, \theta^*)$  is then submitted as a move proposal inside the Metropolis algorithm and is either accepted or rejected with a given probability and the system is evolved accordingly. The process is repeated iteratively starting with the momentum (re)sampling step (equation (7)).

## 2.7. Convergence of MCMC sampling

After running a class of MCMC algorithms, it is necessary to monitor the convergence of the samples. This can be carried out in different ways, including traceplots (showing the evolution of parameter estimates from MCMC draws over the iterations), pair plots (to identify collinearity between variables), and autocorrelation plots (to measure the degree of correlation between draws of MCMC samples). More quantitative metrics are also used in this study to assess the MCMC convergence, as described in the following.

### 2.7.1. Split- $\hat{R}$

Sampling from a target distribution to estimate expectations can only be reliable if the chains have converged to the stationary distribution. The primary step to ensure the reliability of MCMC posterior chains is to check the convergence of chains with different random initializations. MCMC is asymptotically exact in the limit of infinite runs ( $N \rightarrow \infty$ ), however, in our finite-time settings, MCMC convergence cannot be guaranteed with a limited number of samples, thus, we have to rely on diagnostic quantities to assure convergence. Potential scale reduction statistic  $\hat{R}$  (Gelman and Rubin 1992) is widely used in statistics to assess the convergence of multiple Markov chains involved in posterior sampling. The  $\hat{R}$  diagnostic provides an estimate of how much variance could be reduced by running the chains longer. The  $\hat{R}$  close to 1 (in practice, lower than 1.1) indicates good mixing of the chains; otherwise, the chains have not converged to the same stationary distribution.

Computation of  $\hat{R}$  relies on between- ( $B$ ) and within- ( $W$ ) chain variances<sup>8</sup>. In particular, the marginal posterior variance is (over-) estimated as a combination of these two quantities

$$\widehat{\text{var}}^+(\theta | y) = \frac{N-1}{N}W + \frac{1}{N}B, \quad (8)$$

where  $N$  is the number of draws in each of the  $M$  chains. The potential scale reduction  $\hat{R}$  is then defined as

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\theta | y)}{W}}. \quad (9)$$

Stan reports summary statistics including  $\hat{R}$  computed on the split-half chains, an additional precaution to detect non-stationarity in individual chains<sup>9</sup>.

<sup>8</sup>  $W$  is defined as the averaged sum of the squares of the within-chain variance of each chain (Gelman et al 2014).

<sup>9</sup> A simple example: two chains with complementary increasing and decreasing values will yield  $\hat{R}$  close to 1, even though they are not mixing well.

### 2.7.2. Divergent transitions

The trajectories generated by numerical (leapfrog) integrator in the Hamiltonian equations may diverge, typically due to highly varying posterior curvature (such as funnel-shaped distributions; Betancourt 2017), then, the samples cannot be trusted. With symplectic integrators<sup>10</sup>, divergences build up very quickly, making them easy to detect. We have carefully monitored the Stan summary statistics for divergent transitions encountered during sampling.

### 2.7.3. ESS

A complication with MCMC methods is the autocorrelation within the generated samples of a chain, which is often caused by strong correlations among variables. Higher autocorrelation in chains increases the uncertainty of the estimation of posterior quantities of interest. Highly correlated MCMC samplers affect the efficiency of the sampler, as they require more samples to produce the same level of MC error for an estimate (Gelman et al 2014). The ESS provides an estimation of the independency of draws from posterior distributions that generated Markov chain would be equivalent to. Thus, a large enough ESS is an important factor for the reliability of the inferred quantities from posterior draws. We use the bulk-ESS and tail-ESS computed on rank-normalized posterior draws (Vehtari et al 2021) to assess the efficiency of sampling in the bulk (mean estimates efficiency), and tail (quantiles estimates efficiency) of the sampled posterior distribution, respectively. In the case of sampling with four individual chains (each with  $iter = 2000$  and  $warmup = 1000$ ), we consider a minimum ESS of 400 (Vehtari et al 2021).

### 2.7.4. Posterior predictive checks

Validating model assumptions and assessing the reliability of the model performance requires both domain expertise and the evaluation of the model's predictive performance. A posterior predictive check (PPC) is often used for model validation, i.e. generating data from the model using parameters drawn from the estimated posterior, and then comparing it with observed data. PPCs address such questions as does the fitted model and its estimated parameters generate data similar to those observed experimentally? Are the individual and population variability, such as the influence of covariates, consistently modeled? The principle of PPCs is to simulate multiple sets of data according to the predictive distribution  $y_i \sim y^{rep}, i = 1, 2, \dots, N$ , and compare them to the observed data  $y$ , visually or quantified by summary statistics. Visual predictive checks usually depict the predicted mean and 95% confidence/credibility intervals (2.5% and 97.5% empirical quantiles of the posterior predictive draws) confronted with observed data. Note that PPCs are not predictive per se because we are not comparing predicted data to new observations, but rather we are comparing it to the same data that the model was fitted with. Such a process involves the double use of the data and should not be confused with validation prediction, which is evaluated on a separate validation dataset or unseen data (e.g. using WAIC). Systematic discrepancies between model predictions and data would indicate an inadequate model predictive power, and in this case, could be investigated to identify model failures.

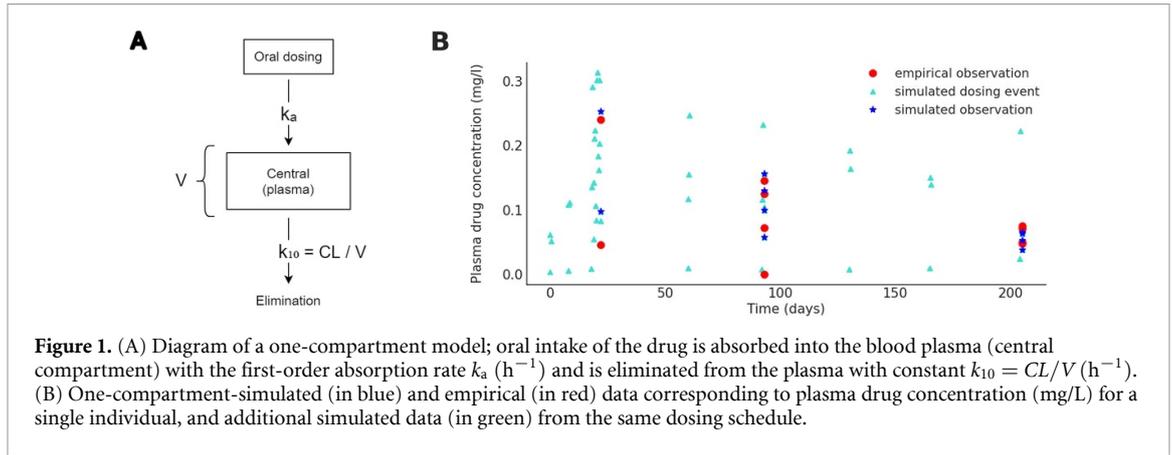
## 2.8. Solving ODE systems in Stan

Linear ODE systems have the virtue of being analytically solvable, which not only provides an exact solution but is also very computationally efficient. The one-compartment linear ODE system (equation (1)) was solved analytically as provided by the Torsten library ready-to-use solver. Torsten package (Zhang et al 2021) is based on Stan software (v2.27.0) and provides functions that facilitate the analysis of pharmacometrics data (Margossian et al 2021). The library handles clinical event schedule data written in NONMEM (Beal et al 2009) conventions and the computation of steady-state dosing<sup>11</sup>. A steady-state is reached when the quantity of drug eliminated matches the quantity of drug that reaches the systemic circulation. In the context of a repeat-dose regimen, steady-state concentrations of drugs in the blood will rise and fall according to a repeating pattern as long as the dosing schedule remains unchanged. Torsten also offers specific functions for one- and two-compartment models for which it implements analytical solutions using matrix exponential in the Stan language.

The model simulation and parameter estimation were performed on a Linux machine with 3.60 GHz Intel Core i7-7700 and 8 GB of memory.

<sup>10</sup> A numerical integration scheme for Hamiltonian systems to nearly conserve the total energy and is particularly useful when treating long times.

<sup>11</sup> Computation of steady-state dosing in Torsten is detailed in appendix D in the case of one-compartment modeling.



### 3. Results

#### 3.1. Bayesian inference on synthetic data

In order to validate the Bayesian workflow, a synthetic dataset is generated from a one-compartment population model with first-order absorption, based on the same dosing schedule as the clinical setting, and with population PK parameters as  $CL_{\text{pop}} = 7$  (L/h),  $V_{\text{pop}} = 65$  (L),  $k_{\text{a pop}} = 2.5$  ( $\text{h}^{-1}$ ). Inter-individual variations  $\omega = (\omega_{CL}, \omega_V, \omega_{k_a})$  are equally set to 0.3, and residual intra-individual variation  $\sigma$  is set to 0.1. An example of simulated drug concentration data compared to empirical observation for a single individual is shown in figure 1.

Bayesian inference of the unknown parameters of the underlying data generating process depends on both the observed responses and prior information about the underlying generative process. For the generated dataset, we aim to explore the actual sensitivity of the inference process to the level of information encoded in the prior. To do so, we consider two different prior specifications on population parameters  $CL_{\text{pop}}$ ,  $V_{\text{pop}}$  and  $k_{\text{a pop}}$ , while the prior distributions placed on other parameters remain identical. Inter-individual variations are modeled by a variance-covariance matrix  $\Omega$ , in which for computational efficiency in MC sampling, it is decomposed as

$$\Omega = \text{diag}(\omega) \mathbf{L} \mathbf{L}' \text{diag}(\omega) \tag{10}$$

with  $\mathbf{L}$  (resp.  $\mathbf{L}'$ ) as the Cholesky factor (resp. its transpose) of a correlation matrix  $R = \mathbf{L} \mathbf{L}'$ , and  $\omega = (\omega_{CL}, \omega_V, \omega_{k_a})$  as the vector of standard deviation of PK parameters. Here,  $\mathbf{L}$  is assigned with a Lewandowski–Kurowicka–Joe (LKJ; Lewandowski et al 2009) prior distribution (a distribution over the set of correlation matrices or their Cholesky factor) with uniform density over all Cholesky factors of correlation matrices, whereas  $\omega$  is assigned with a half-normal prior with zero mean and a standard deviation of 0.1:

$$\mathbf{L} \sim \text{lkj\_corr\_cholesky}(\eta = 1), \tag{11}$$

$$\omega \sim \text{Half-}\mathcal{N}(0, 0.1). \tag{12}$$

Individual parameters (normalised without covariate influence) denoted by  $\theta_i = \{CL_i, V_i, k_{a i}\}_{i=1}^N$  are modeled by a Gaussian distribution centered on population parameters  $\theta_{\text{pop}} = \{CL_{\text{pop}}, V_{\text{pop}}, k_{\text{a pop}}\}$ , and

$$\log(\theta_i) \sim \mathcal{N}(\log(\theta_{\text{pop}}), \Omega) \quad \forall i. \tag{13}$$

Given the parameters  $\theta_i$  for the  $i$ th individual, the virtual amount of drug in both absorption and plasma compartments is computed by analytically solving the one-compartment ODE system given by equation (1). Here, we employed a ready-to-use one-compartment model solver provided by the Torsten package for event schedule specified data (Zhang et al 2021). Estimation of drug concentration in the blood for individual  $i$ , with unknown  $C_i$  underlying observations  $C_i[\text{obs}]$ , is given by  $\hat{C}_i(t) = \frac{\hat{y}_{\text{central}_i}(t)}{V_i}$ , with the likelihood of

$$\log(C_i[\text{obs}]) \sim \mathcal{N}(\log(\hat{C}_i[\text{obs}]), \sigma). \tag{14}$$

Given the above formulation, we compare two classes of *a priori* information on the population PK parameters. In the first set-up (denoted by **model I**), we place a weakly informative prior distribution centered around the true value of the parameters:

**Table 1.** Summary statistics and convergence diagnostics for Bayesian inference on synthetic data (figure 1), using weakly informative priors (given by model I) and diffuse priors (given by model II), averaged over four HMC chains. Overall, the Bayesian model I with a higher level of information in prior, yields faster computation, better convergence diagnostics, and substantially higher prediction accuracy.

		model I	model II
Step size	average	0.046	0.033
	SE	0.00	0.01
Integrator steps	average	121	159
	SE	19	56
Effective Sample Size (ESS)	min	65	27
	average	5876	3631
	max	13 367	13 138
$\hat{R} > 1.1$		0%	0%
$\hat{R} > 1.01$		2.1%	8.9%
Divergences		0	0
Sampling time		17 min	24 min
Gradient evaluation		0.009 sec	0.012 sec
Maximum correlation		0.20 ( $k_{a\text{pop}}, \omega_{ka}$ )	0.71 ( $V_{\text{pop}}, k_{a\text{pop}}$ )
Number of parameters		418 <sup>a</sup>	418
Root Mean Square Error (RMSE)		0.349	0.331
Effective number of parameters ( $p_{\text{eff}}$ )		40 669.36	56 303.03
Watanabe-Akaike information criterion (WAIC)		82 401.03	113 679.84

<sup>a</sup> Three population parameters ( $CL_{\text{pop}}, V_{\text{pop}}, k_{a\text{pop}}$ ) and their respective variability ( $\omega_{CL}, \omega_V, \omega_{ka}$ );  $3 \times 67$  individual parameters  $\{CL_i, V_i, k_{ai}\}_{i=1}^{N=67}$ ; elements of  $3 \times 3$  Cholesky matrix  $L$ ;  $3 \times 67$  standard normal variables for non-centered parametrization of equation (13); residual variability  $\sigma$ .

$$\begin{aligned}
 CL_{\text{pop}} &\sim \mathcal{N}(7, 5) \\
 V_{\text{pop}} &\sim \mathcal{N}(70, 5) \\
 k_{a\text{pop}} &\sim \text{Half-}\mathcal{N}(3, 0.5).
 \end{aligned}
 \tag{model I}$$

In the second set-up (denoted by model II), we place more diffuse (uninformative) priors, but still they are centered on true value for parameters  $V_{\text{pop}}$  and  $k_{a\text{pop}}$ :

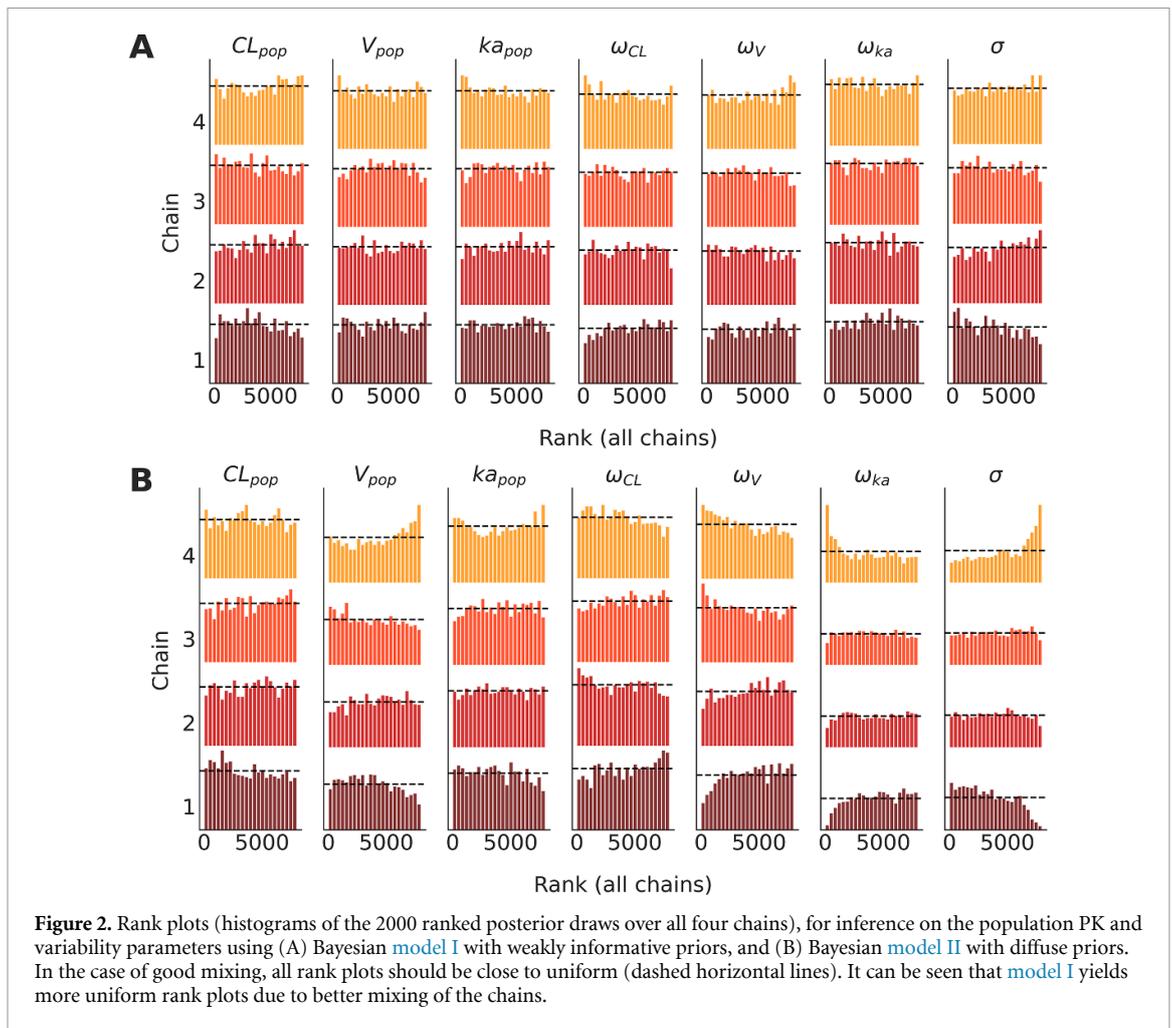
$$\begin{aligned}
 CL_{\text{pop}} &\sim \mathcal{N}(7, 5) \\
 V_{\text{pop}} &\sim \mathcal{N}(70, 10) \\
 k_{a\text{pop}} &\sim \text{Half-}\mathcal{N}(3, 2).
 \end{aligned}
 \tag{model II}$$

To perform inference, prediction, and model comparison, we ran four individual chains with NUTS default setting in Stan (e.g. 2000 discarded warm-up iterations and 2000 saved sampling iterations, an expected acceptance probability of 0.85, and a maximum tree-depth of 10). In the following, we first carefully monitor the convergence of Markov chains, then we show the estimated posterior distributions and PPCs for population PK parameters. Finally, we demonstrate model comparison using WAIC to reliably select the model that best explains the data.

### 3.1.1. Sampling diagnostics

Sampling properties of both Bayesian models with weakly informative and diffuse priors (given by model I and model II, respectively) are overall satisfactory (see table 1). Neither of the models exhibited sampling pathologies such as divergent transitions or reaching maximum tree-depth. The values of the potential scale reduction factor are  $\hat{R} < 1.1$  for all parameters in both Bayesian models, indicating that the Markov chains have converged to a stationary distribution. When considering a more exigent threshold, then  $\hat{R} < 1.01$  for all parameters of model I, except those relative to the 53th individual. More precisely, the individual parameter which partly controls the individual variation for the 53th individual and as a result biases the PK parameters and quantities (estimated and predicted concentration) for this individual. The Bayesian model II encounters more convergence issues with more individuals (16th, 44th, 53th, 64th). This model exhibits around 9% of its total 418 parameters that do not fall under the threshold, compared to 2.1% for the other model (see table 1).

The Bayesian model I also exhibits a better convergence according to HMC specific diagnostics compared to model II (see table 1): the self-tuning step size of the leapfrog integrator is larger, and has fewer integration steps, resulting in faster computational time (around 1.5 order of magnitude), and more efficient sampling. Towards a principled Bayesian workflow (Gelman et al 2020), we investigate more closely the convergence of



the two fitted Bayesian models using the methods and quantities proposed by Vehtari *et al* (2021) to address possible flaws in  $\hat{R}$ -based convergence diagnostics.

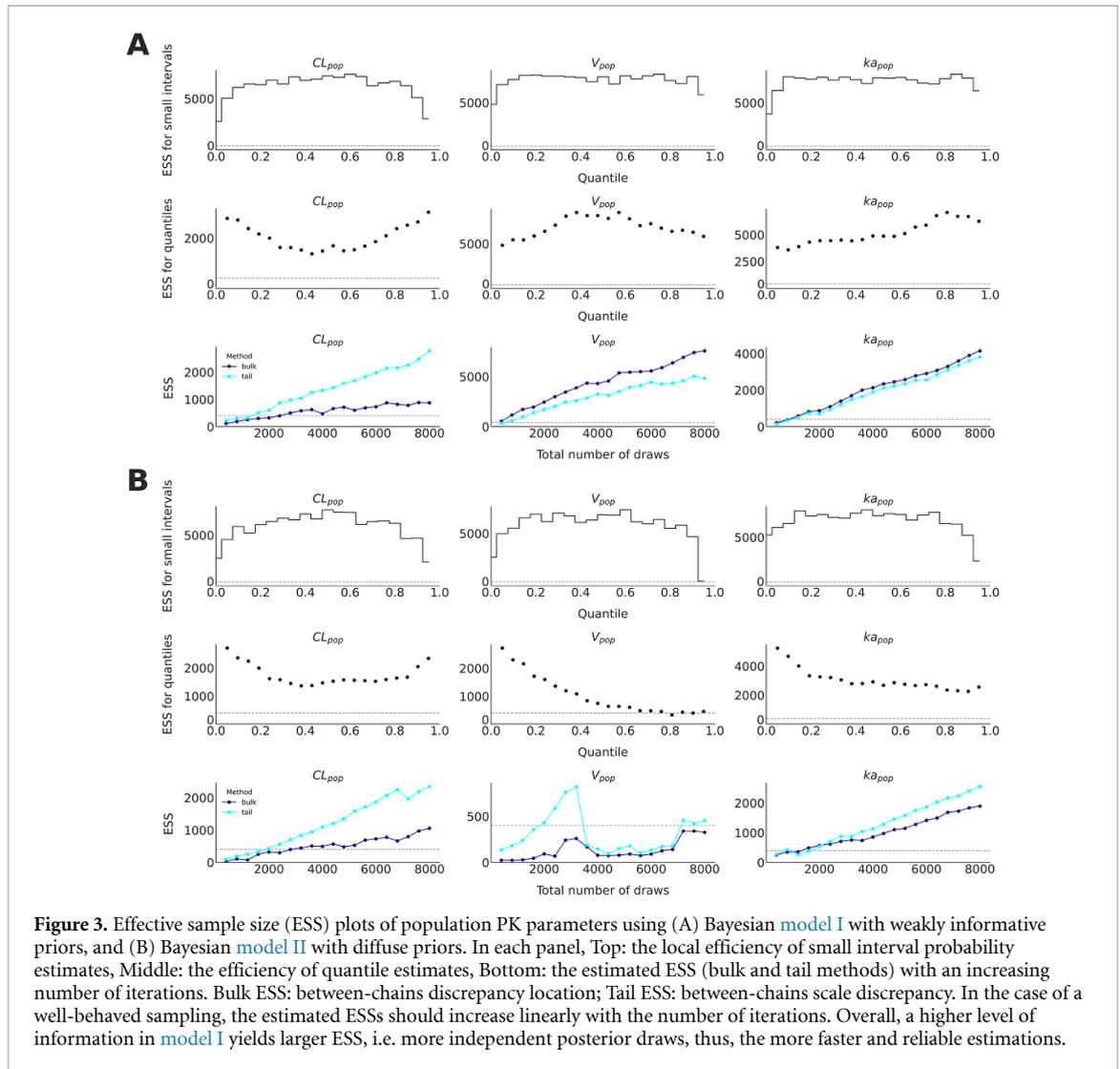
### 3.1.2. Mixing of the chains

Rank plots are an alternative approach to trace plots for visual sanity checking of convergence in posterior chains (Vehtari *et al* 2021) as shown in figure 2. For comparison to raw trace plots of the chains for the population PK parameters, see appendix figure B.1. Rank plots are histograms of posterior draws, ranked over all four chains, and plotted for each chain separately. If all chains are targeting the same posterior distribution, the rank histograms should not deviate significantly from uniformity, which is expected with a higher ESS. Figure 2 illustrates rank plots for population PK parameters using both Bayesian **model I** and **model II**. From this figure, we can see that a higher level of information in **model I** yields more uniform rank plots due to better mixing of the chains.

### 3.1.3. ESS

We have carefully monitored the ESS as an indicator of the number of independent draws, which affects the uncertainty of the estimation. We investigate the ESS for the three population PK parameters using the computation of split- $\hat{R}$  and ESS on rank-normalized (normal scores) samples, and the corresponding diagnostics proposed by Vehtari *et al* (2021). They have observed that the convergence of the chains is not necessarily uniform over the entire distribution of a parameter of interest. In addition to rank-normalized based ESS (or bulk-ESS), they proposed a tail-ESS to evaluate convergence in extreme quantiles.

In figure 3, we illustrate the checks of efficiency of quantiles and small-interval probabilities, as well as the efficiency evolution with the number of iterations. The Bayesian **model I** yields an overall larger ESS for the three population PK parameters, especially for parameter  $V_{pop}$ . The Bayesian **model I** quantile estimates efficiency is comparatively improved compared to **model II** for parameter  $V_{pop}$ , and high quantiles of  $k_{a_{pop}}$ , the first being higher and uniform and the latter being increased in the right tail of the distribution. The



parameter  $V_{pop}$  in **model II** has noticeable small and decreasing efficiency in high quantiles.  $CL_{pop}$  quantile estimates efficiency is similar in both models. In the case of a well-behaved sampling, the estimated ESSs should increase linearly with the number of iterations. This is the case for the population PK parameters in **model I**. However, in **model II**, the parameter  $V_{pop}$  exhibits a problematic drop at half of the total iterations in tail-ESS, which indicates inter-chain scale discrepancy. Population PK parameters of both models reach an acceptable ESS over the recommended threshold of 400 when the number of draws exceeds 2000, except for pathological parameter  $V_{pop}$  in **model II**. Together, these diagnostics validate that the samples in **model I** have better convergence to the target distribution.

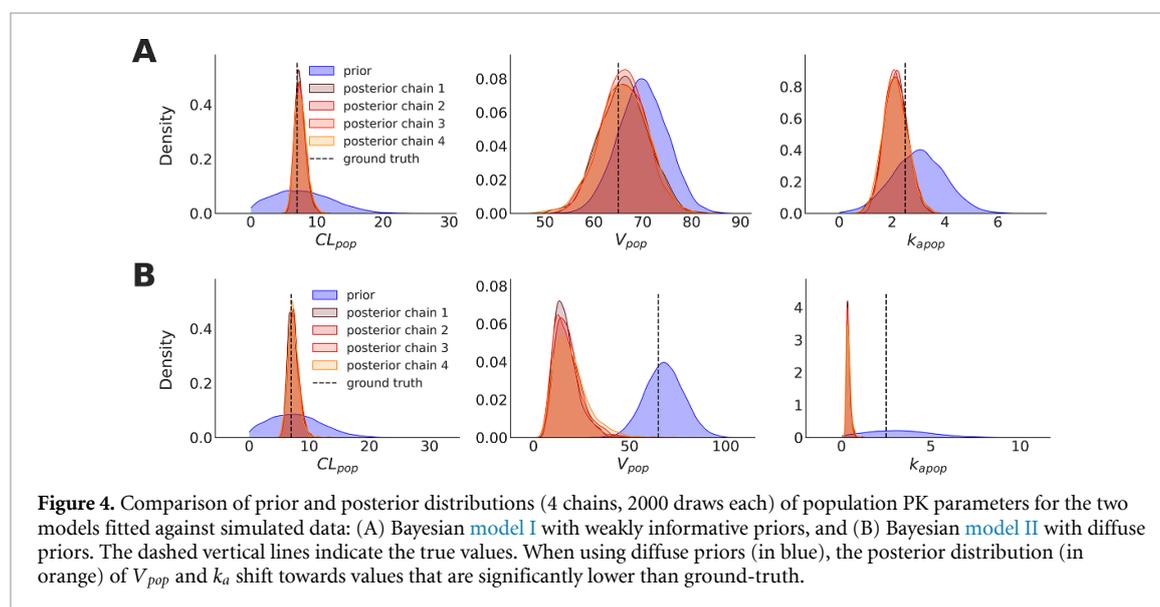
### 3.1.4. Posterior behaviour

The prior and posterior distributions of **model I** and **model II** are illustrated in figure 4. It can be seen that, when using **model I**, the true values of all PK parameters (dashed vertical lines) are well within the support of the posterior densities, indicating that the Bayesian parameter recovery was successful. In contrast, the setting in **model II** failed to recover the parameters  $V_{pop}$  and  $k_{a_{pop}}$ . When using a wider set of priors (**model II**), the posterior distribution of  $V_{pop}$  shifts towards values that are significantly lower than ground-truth, under over-influence of the data (over-fitting). A similar phenomenon is observed for the absorption parameter  $k_a$ . The uncertainty of the estimation of posterior quantities (mean, MCSE of the mean, standard deviation, 95% credible intervals, and ESS) for the population PK parameters in both Bayesian models are summarised in table 2.

Visualizing the joint posterior samples in pair plots is especially useful for identifying collinearity between parameters as well as the presence of non-identifiability (banana-shaped distributions). As shown in

**Table 2.** Comparison of sampled posterior distributions of population PK parameters using Bayesian [model I](#) and [model II](#); sample mean, MCSE, sample standard deviation (SD), posterior 95% credible intervals and effective sample size ( $ESS_{bulk}$ ). Parameters marked with a \* have ESS below the limit of 400 recommended by Vehtari et al (2021).

Population parameters	model I					model II				
	Mean	MCSE	SD	95% CI	$ESS_{bulk}$	Mean	MCSE	SD	95% CI	$ESS_{bulk}$
One-compartment PK										
$CL_{pop}$ (L/h)	7.5	0.028	0.86	[6.05; 9.45]	873	7.4	0.028	0.9	[5.87; 9.39]	1058
$V_{pop}$ (L)	66	0.056	4.8	[56.68; 75.42]	7582	18	0.5	7.5	[7.94; 36.37]	326*
$k_{apop}$ ( $h^{-1}$ )	2.1	0.007	0.45	[1.29; 3.05]	4148	0.39	0.003	0.12	[0.22; 0.68]	1893
Intersubject variability $\omega$										
$\omega_{CL}$	0.52	0.002	0.05	[0.43; 0.62]	446	0.55	0.001	0.05	[0.46; 0.65]	1272
$\omega_V$	0.24	0.002	0.08	[0.11; 0.41]	1404	0.35	0.004	0.086	[0.19; 0.52]	450
$\omega_{k_a}$	0.41	0.001	0.07	[0.28; 0.54]	4682	0.32	0.008	0.084	[0.28; 0.54]	142*
Residual variability										
$\sigma$ (exp)	0.08	0.0002	0.003	[0.07; 0.09]	496	0.072	0.0005	0.0034	[0.07; 0.08]	40*



**Figure 4.** Comparison of prior and posterior distributions (4 chains, 2000 draws each) of population PK parameters for the two models fitted against simulated data: (A) Bayesian [model I](#) with weakly informative priors, and (B) Bayesian [model II](#) with diffuse priors. The dashed vertical lines indicate the true values. When using diffuse priors (in blue), the posterior distribution (in orange) of  $V_{pop}$  and  $k_a$  shift towards values that are significantly lower than ground-truth.

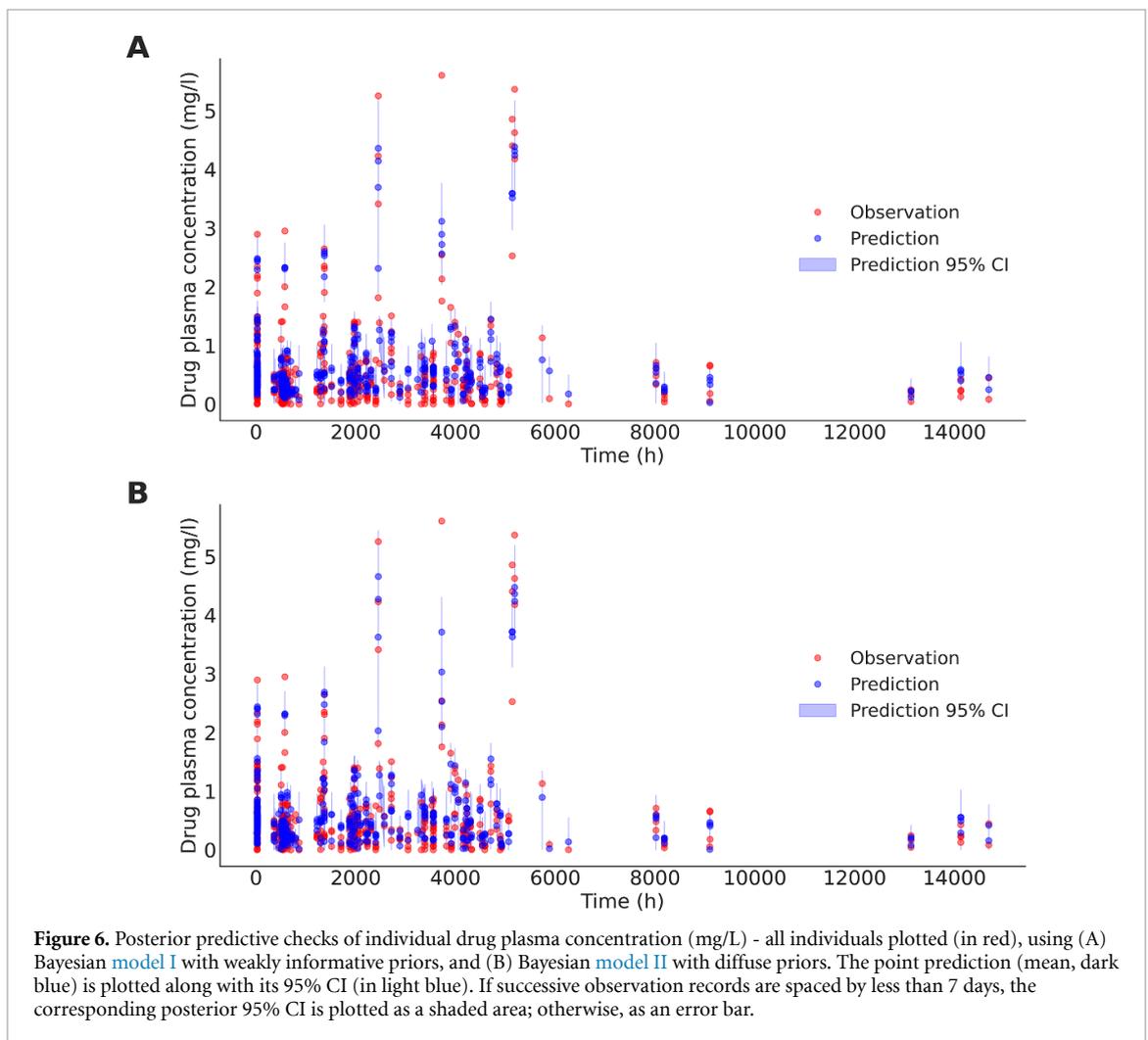
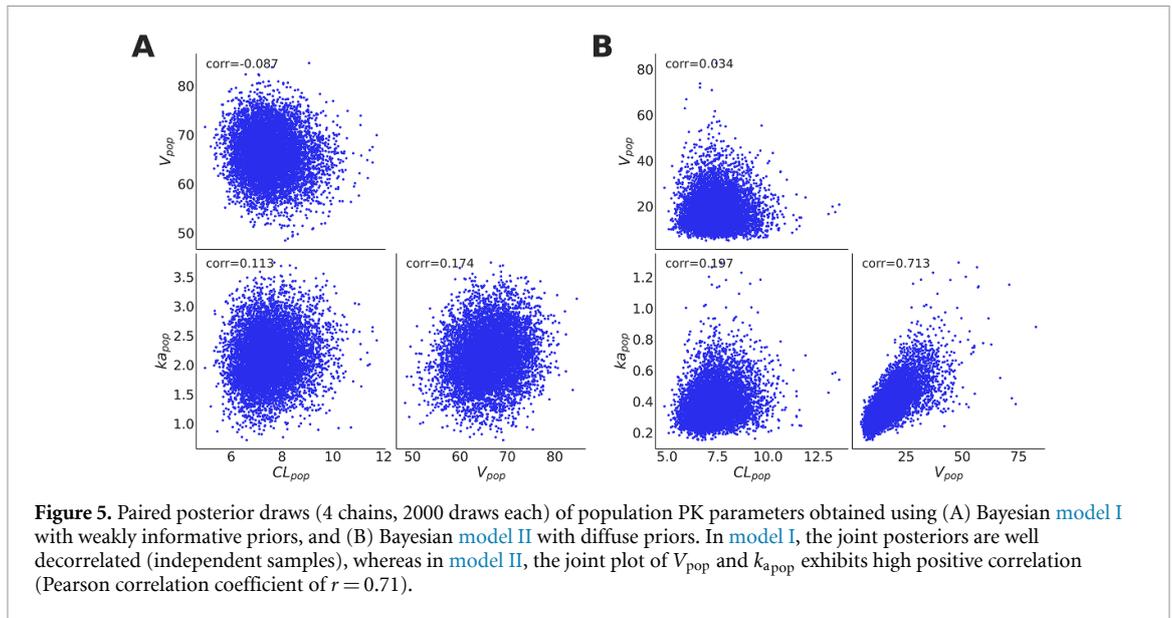
figure 5, [model II](#) exhibits a larger degenerate behavior in the form of a positive correlation ( $r > 0.7$ ) between  $V_{pop}$  and  $k_{apop}$  posterior samples that is not seen in the more informative [model I](#) ( $r < 0.2$ ). Such a high collinearity leads to an inefficient exploration of the posterior, which can be quantifiably observed in decreased numbers of effective samples and increased  $\hat{R}$  values. Moreover, by randomly initializing the chains, the estimation of posterior density of  $CL_{pop}$  is more robust than that of the other two parameters. The robustness of the sampling of  $CL_{pop}$  with respect to other parameters was also observed in the ESS plots (figure 3). According to these results, the Bayesian set-up used in [model I](#) substantially improved the convergence of the recovery for population PK parameters.

### 3.1.5. PPCs

Posterior predictive (or retrodictive) checks are a widely-used tool for verifying the reliability of the fitted model by monitoring its predictive performance on the fitted dataset. Synthetic datasets (67 individuals, total of 427 observations) of plasma drug concentration are repeatedly drawn from the predictive distribution given by the fitted model (4 chains, each with 2000 draws). Punctual predictions (mean) and the corresponding 95% credible intervals are extracted from the draws and plotted along with the actual observations (see figure 6). For both Bayesian [model I](#) and [model II](#), these visual checks suggest no systematic bias in the predictive distribution compared to the fitted data.

### 3.1.6. Residuals

Non-normalised residuals are calculated as differences between observed plasma concentration  $C_{obs}$  and the estimated concentration computed by solving the one-compartment ODE system (see equation (1)), over the fitted individual PK parameters ( $CL_i, V_i, k_{ai}$ ) and evaluated at observation points  $\hat{C}_{obs}$ . Monitoring the



residuals can help in pointing out issues with the estimation of PK parameters, an inadequate compartmental modeling, or ODE-solver inaccuracies. Histograms of non-normalised residuals for both **model I** and **model II** are given in appendix figure B.2. They are reasonably zero-centered (not biased), and there is no significant difference in the distributions of the non-normalised residuals of **model I** and **model II**. Note that the root mean squared error (RMSE, figure B.2, and table 1) as a within-sample

predictive accuracy is slightly smaller in [model II](#); this is consistent with the marginal posterior densities (figure 4) where the interaction of the likelihood with more diffuse priors fails to recover the true parameters; hence, the resulting model fits the observed data even better, but at the cost of a poor representation of the true data generating process (Gelman *et al* 2017, Betancourt 2018).

### 3.1.7. Model selection

To select the best Bayesian model considered in this study for the synthetic dataset, we use the widely applicable information criterion (WAIC) as a measure of model evidence. WAIC is an indicator for the comparison of point-wise out-of-sample predictive accuracy of Bayesian models, based on the whole estimated posterior distribution. WAIC is here reported (see table 1) in the  $-2$  log-score scale (deviance): a model with a smaller WAIC value suggests higher predictive accuracy, i.e. a better balance between accuracy and complexity, even though the number of parameters is equal for both candidate models. Our model comparison based on WAIC substantially favors [model I](#) (since  $\Delta\text{WAIC} \gg 10$ ). Taken together, the Bayesian set-up used in [model I](#) substantially improved the convergence and model predictive power for estimation over population PK parameters.

### 3.1.8. Effect of noise on estimation

Lastly, we investigate the effect of the observation noise level on model parameter estimation. For varying levels of exponential residual noise, we examine the robustness of inference by monitoring HMC sampling diagnostics, PK parameter estimation, the within-sample accuracy of the fit (RMSE), and the out-of-sample predictive accuracy (WAIC). A sweep over noise values is depicted in figure B.3 for the best selected model (i.e. [model I](#)). We observed that lower residual noise levels ( $\sigma_{exp} < 2$ ) yield better estimation of posterior distributions of model parameters, a better RMSE between observed and fitted data, and enhanced model evidence according to WAIC measures. However, a large value of residual noise ( $\sigma_{exp} \geq 2$ ) results in a significantly small predictive accuracy (i.e. the larger RMSE and WAIC values). This indicates that HMC sampling with the Bayesian set-up used in [model I](#) is robust when dealing with high levels of noise in observed data.

### 3.1.9. Out-of-distribution (OOD) PK parameters

The OOD parameters refer to situations where the observed data falls outside the range of typical or expected values. We generate biologically implausible data from unrealistic PK parameters (e.g. a volume of distribution  $V$  of 1 L), and we challenge the model's ability to infer from these data, which are unrealistic. Figure B.4 shows the predictive ability of [model I](#) measured using WAIC, when marginally changing the values of the population parameters in the data generation process. We observed a significantly lower WAIC value for plausible values of the parameters, resulting in better prediction accuracy than OOD parameters. This result indicates that the WAIC is able to effectively capture the underlying patterns and relationships in the data when the parameter values fall within a plausible range, thus establishing bounds or ranges in which the current model performs well.

## 3.2. Bayesian inference on clinical data

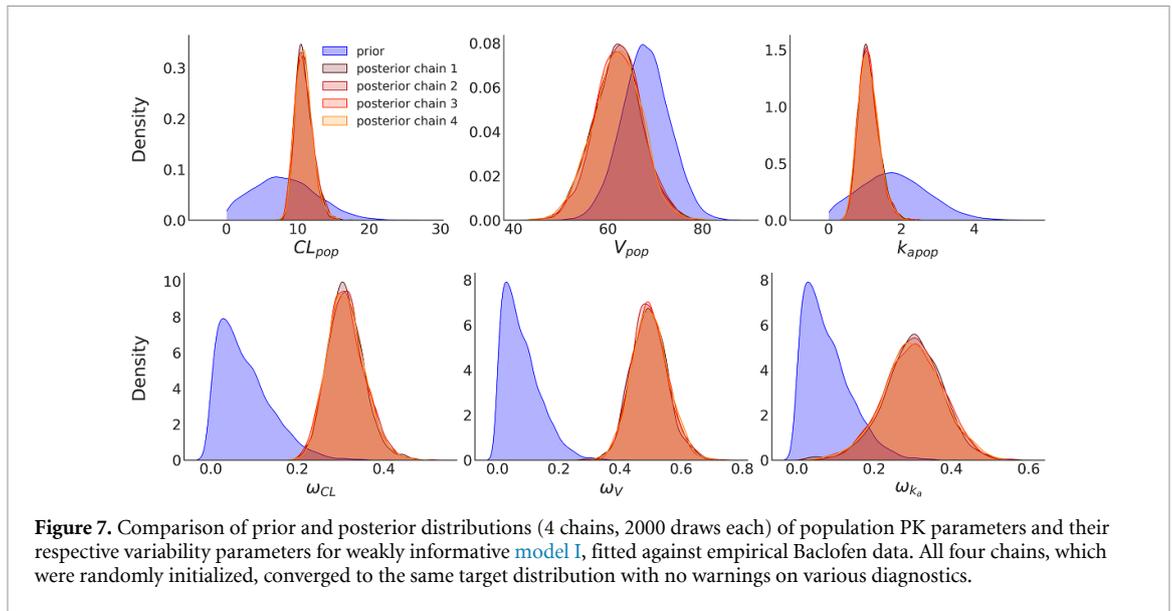
After calibrating models on simulated data, we fit [model I](#) on empirical Baclofen data. In particular, we use the set of population PK priors defined in [model I](#). As in the previous section, we ran 4 chains of 2000 iterations each, with default settings in Stan (e.g. 2000 warm-up iterations, a target Metropolis acceptance rate of 0.85, and a maximum tree-depth of 10).

### 3.2.1. Sampling diagnostics

Stan diagnostics reported no warning on convergence, in particular, there is no  $\hat{R}$  value above the threshold of 1.01, no diverging transition, and no quantity with insufficient ESS. We also rely on convergence diagnostics as shown in the investigation of models on simulated data: trace plots in figure C.1 show that the chains are mixing well; rank plots in figure C.2 are close to uniform; ESS (bulk and tail methods) in figure C.3 is large for the main three population PK parameters; paired samples of population PK parameters in figure C.4 exhibit no significant correlation.

### 3.2.2. Posterior distributions

The sampled posterior distributions of the three main population PK parameters and their respective inter-individual variability parameters are shown and compared to their prior distributions in figure 7. Statistical summary of posterior distributions of the population PK parameters is reported in table 3. We report the mean, MCSE of the mean, standard deviation, 95% credible intervals, and ESS. Interpretation of Bayesian posterior credible intervals is straightforward: a 95% CI is the central portion of the posterior that



**Figure 7.** Comparison of prior and posterior distributions (4 chains, 2000 draws each) of population PK parameters and their respective variability parameters for weakly informative model I, fitted against empirical Baclofen data. All four chains, which were randomly initialized, converged to the same target distribution with no warnings on various diagnostics.

**Table 3.** Statistical summary of posterior distributions of population PK parameters sampled from empirical Baclofen data; sample mean, Monte Carlo standard error (MCSE), sample standard deviation (SD), and posterior 95% credible interval.

Population parameters	Posterior sampled distributions				
	Mean	MCSE	SD	95% C.I.	$ESS_{bulk}$
<b>One-compartment PK</b>					
$CL_{pop}$ (L/h)	11	0.023	1.2	[8.80 ; 13.68]	2861
$V_{pop}$ (L)	62	0.05	5.1	[52 ; 72.07]	10 419
$k_{apop}$ ( $h^{-1}$ )	1.1	0.0041	0.27	[0.64 ; 1.72]	4307
<b>Intersubject variability <math>\omega</math></b>					
$\omega_{CL}$	0.31	0.0006	0.042	[0.24 ; 0.40]	4421
$\omega_V$	0.5	0.0008	0.058	[0.39 ; 0.62]	5400
$\omega_{ka}$	0.3	0.0013	0.076	[0.15 ; 0.45]	3354
<b>Residual variability</b>					
$\sigma$ (exp)	0.084	0.00 003	0.003	[0.08 ; 0.09]	10 805

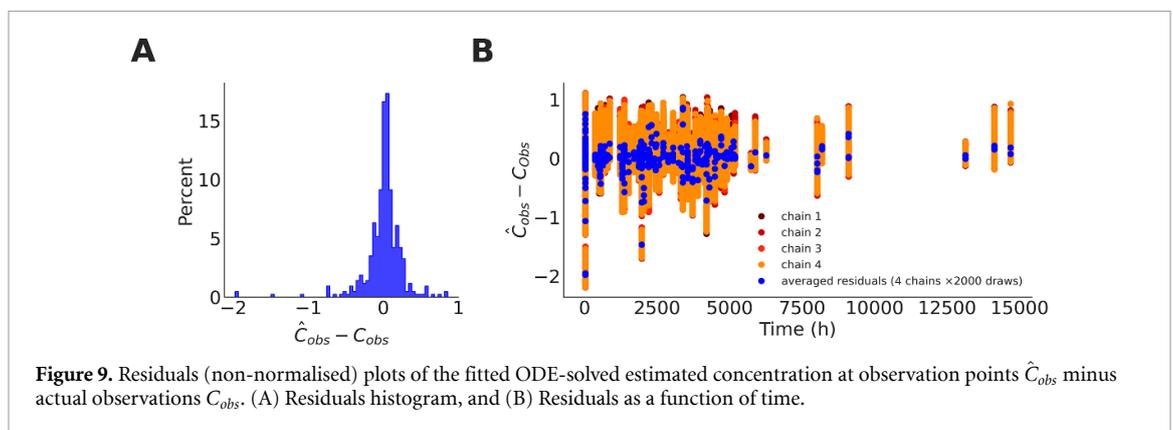
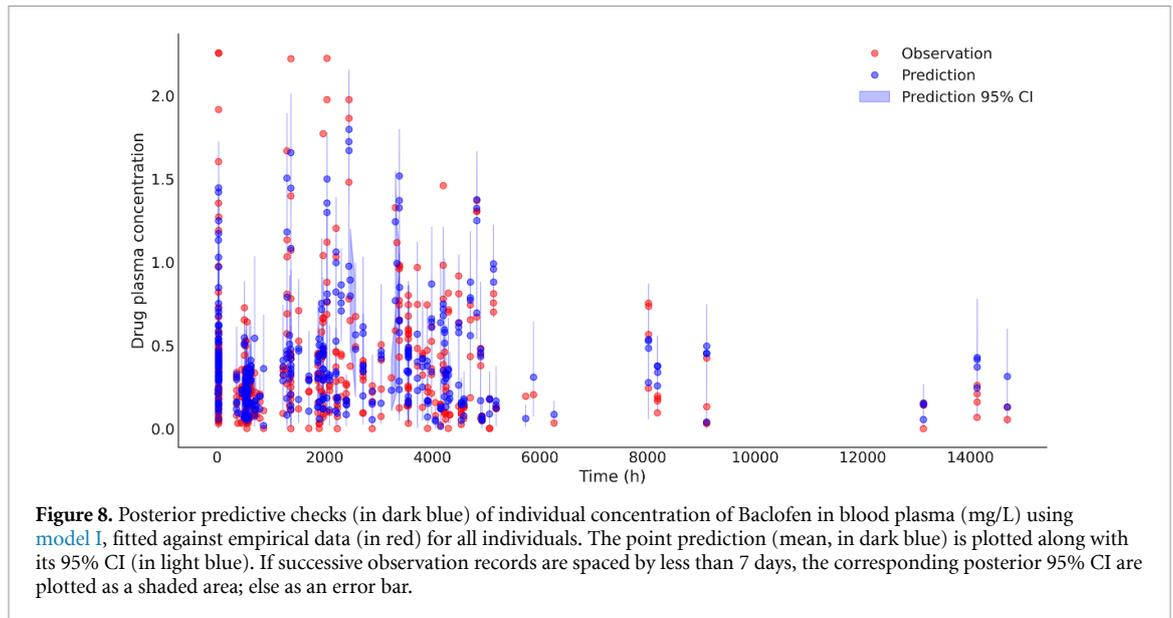
concentrates 95% of sampled values; given the observed data, the parameter falls into the interval with probability of 95%. Distributions of sampled individual PK parameters  $CL_i$ ,  $V_i$  and  $k_{ai}$  are plotted for every individual in figure C.5.

### 3.2.3. PPCs

Predicted Baclofen plasma concentration datasets ( $N = 4 \text{ chains} \times 2000 \text{ draws}$ ) are generated from the estimated predictive distribution (see figure 8). Each shown data-point corresponds to an actual individual observation (in red). The PPCs are shown for all individual observations (in blue). For every individual observation, the corresponding punctual prediction (mean) and 95% confidence interval over the repeated simulations are also plotted (in light blue). No systematic deviation of predictive simulations from reality is visible, suggesting that the model’s predictive distribution accurately captured the plasma concentration.

### 3.2.4. Residuals

Checking the residuals between the fitted ODE-solved estimated concentration at observation points  $\widehat{C}_{obs}$  and actual observations  $C_{obs}$  allows us to ensure that there is no bias in the estimation of the PK parameters and issues in the resolution of the underlying ODE system. The histogram of non-normalized residuals and the residuals along the time-axis are shown in figure 9, indicating that there is no biased estimation.



### 3.2.5. Covariates

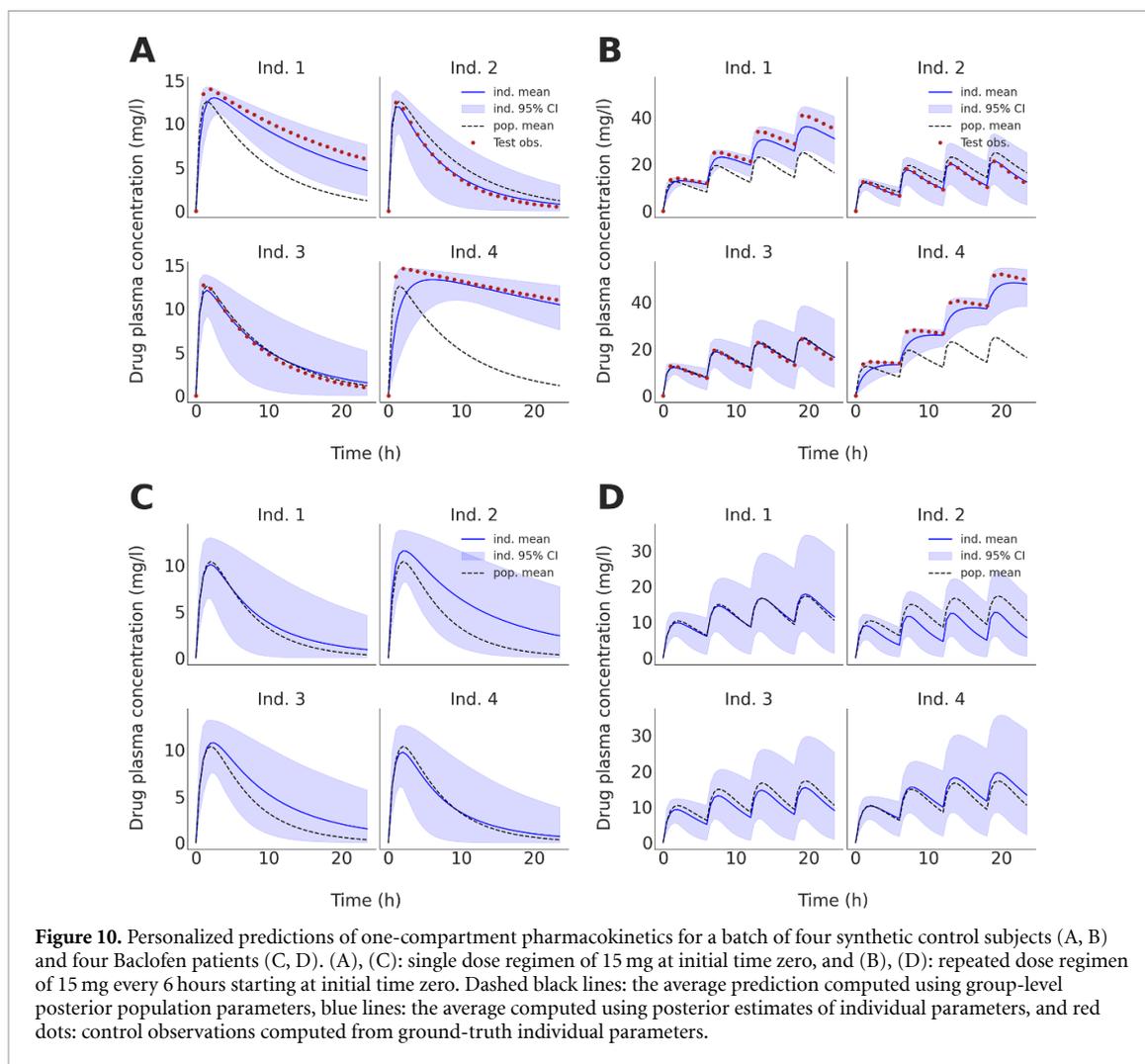
Clinical data includes 18 biological covariates, which are detailed in Imbert *et al* (2015). The influence of a continuous covariate (denoted by COVAR) on individual parameter  $\theta_i$  is modeled by

$$\theta_{i,COVAR} = \theta_i \left[ \frac{COVAR_i}{\text{median}(COVAR)} \right]^p. \quad (15)$$

Searching for significant covariates on PK parameters was performed by evaluating repeated regressions (in linear and nonlinear forms of equation (15)) on HMC-fitted individual PK parameters. The statistical significance ( $p < 0.01$ ) of regression coefficients was tested for covariates such as gender, body weight, age, height, creatinine serum, creatinine clearance, urea, alanine aminotransferase, aspartate aminotransferase, albumin, mean corpuscular volume, prothrombin ratio, fibrinogen, gamma-glutamyl transferase, alkaline phosphatase, carbohydrate-deficient transferrin, tobacco, and fagerstrom score. None exhibited significance, and as a result, no covariate was retained in the final model. This is consistent with the covariate selection that was performed with a NONMEM (Beal *et al* 2009) routine in Imbert *et al* (2015) and resulted in no selected covariate.

### 3.2.6. Prediction on individual dose regimen

The individualized estimation of PK model parameters allows for personalized prediction in customized dosing and temporal contexts. By placing the estimated posterior samples of the individual parameters into equation (3), we are able to predict the drug concentration in the blood under any time frame and different dose regimens, thus deriving reliable projections for optimizing the doses for individuals. Figure 10 shows examples of short-term posterior simulations for control subjects and Baclofen patients, under two different dose regimens. These simulations were obtained following a one-compartment modelling and using each individual's set of posterior samples of kinetic parameters. Note that in control cases, individual predictions



based on posterior estimates of individual parameters are considerably less accurate predictors compared to the population mean when individual settings deviate from the average.

#### 4. Discussion

Bayesian inference is a principled method for estimating the posterior distribution of unknown quantities given only observed responses and prior beliefs about unobserved hidden states (or latent variables). An advantage of using the Bayesian framework in the context of inference/prediction is the ability to generate not only a single point estimate (e.g. in the frequentist approach), but also full probability distributions for the quantities of interest (uncertainty quantification for decision-making process). From the latter, one can directly extract quantiles, with the possibility to answer questions such as ‘what is the probability that the parameter of interest is greater/smaller than a specific value?’, with the confidence intervals in estimation. In addition, the propagation of uncertainty in the Bayesian framework provides a more robust and reliable predictive capability for the model under study, rather than point estimation with optimization methods. Importantly, the out-of-sample prediction accuracy (i.e. the measure of the model’s ability in new data prediction e.g. using WAIC) enables reliable and efficient evaluation of potential hypotheses, as performed in this study. Several previous studies have used a scoring function (such as root mean square error or correlation) to measure the similarity between empirical and fitted data (Imbert *et al* 2015). The choice of scoring function can dramatically affect the ranking of model candidates, and, ultimately, the decision-making processes (see RMSE in table 1). Rather, we used non-parametric probabilistic mythology to analyze data, while various convergence diagnostics were carefully monitored to assess when the sampling procedure has converged to sampling from the target distribution.

Another advantage of the Bayesian approach used in this study is to integrate our prior information (domain expertise before seeing the data) into the inference process to better improve the model's ability to predict new data. For instance, it has been shown that measuring the prediction accuracy of a whole-brain network model with a higher level of information in the prior provides decisive evidence in favor of the true hypothesis regarding the degree of epileptogenicity, across different brain regions (Hashemi *et al* 2020, 2021, Vattikonda *et al* 2021). Accuracy of the predictive models is critical in determining the quality of their predictions for evidence-informed decision-making. It is important to note that the prior information is relevant in estimating the parameters rather than in calculating predictive accuracy (Gelman *et al* 2014). However, a substantial change in priors will affect the computations of the marginal likelihood, thus, the accuracy of the predictive models. An inappropriate choice of prior can lead to weak inferences and, consequently, poor predictions (see figure 4). On the other hand, a sufficient amount of information encoded through the prior distribution can provide decisive evidence in favor of the correct hypothesis, as shown in this study (see WAIC values in table 1). In addition to achieving a model with higher performance prediction accuracy, the appropriate priors can dramatically improve the exploration of the search space in terms of computational cost and algorithmic diagnostics, such as effective numbers of samples, rendering a more efficient inference process (see table 1, figures 2, 3 and 5).

MC methods allow us to sample from and, thus, approximate the exact posterior densities, without requiring knowledge of the whole distribution. HMC uses the gradient information of the posterior to avoid the undesired random walk of traditional sampling algorithms, thereby samples efficiently from posterior distributions with correlated parameters, particularly in high dimensional settings. Thanks to the open-source Stan programs (Carpenter *et al* 2017), adaptive HMC sampling refined the inference on population PK parameters, with the uncertainty in estimated parameters compared to point estimation in the previous work of Imbert *et al* (2015). In this study, we integrated domain expertise into inference process through the prior distribution to obtain the full posterior distribution that is well consistent with the data and domain expertise. For instance, prior information refined inference for the clearance parameter  $CL_{pop}$  and its variability  $\omega_{CL}$  that are outside the 95% bootstrap confidence intervals compared to previous analysis. The posterior distributions of other inferred population PK parameters exhibit mean values that are within the previously proposed confidence intervals, and the credibility intervals extracted from their posterior distributions have been narrowed in comparison. Moreover, here the residual variance has been reduced by a factor of 10.

In this study, we estimated the effect of Baclofen on retrospective patients with AUD through Bayesian estimation of the parameters of a population PK model. Note that, retrospective studies do not involve the manipulation of variables, making it challenging to establish a cause-and-effect relationship between the exposure and the outcome. Although, they can provide important information about the potential associations between exposures and outcomes, but they cannot prove causation. This is because retrospective studies are subject to various sources of bias, including selection bias, confounding, and measurement bias. Despite these limitations, retrospective studies can still provide valuable information for generating hypotheses, guiding further research, and informing clinical decision-making. Using the prospective data, to establish a causal relationship between Baclofen's dose and exposure, and measuring the probability of cause and effect remains to be investigated in future studies.

The Bayesian approach, relying on automatic MCMC sampling algorithms used in this study provided accurate and reliable estimation of Baclofen effect, validated by posterior behavior analysis and various convergence diagnostics. This principled and probabilistic methodology enabled us to integrate the prior information in the explanation of observations to maximize the model prediction power for a given patient with AUD. This work offers proper guidance for the prediction of drug efficacy in clinical practice, bringing personalized medicine closer to reality in the treatment of brain disorders.

### Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

## Acknowledgments

This work was funded by Human Brain Project SGA3, and the Fondation pour la Recherche Médicale (DIC20161236442). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

Conceptualization: M H and V K J. Methodology: N B, and M H. Data curation: N S. Software: N B and M H. Investigation: N B, and M H. Visualization: N B. Supervision: M H and V K J. Funding acquisition: V K J. Writing—original draft: N B and M H. Writing—review & editing: N B, N S, V K J and M H.

## Information sharing statement

The patient datasets cannot be made publicly available due to the data protection concerns regarding potentially identifying and sensitive patient information. The data that support the findings of this study are available upon request from Nicolas Simon ([nicolas.simon@univ-amu.fr](mailto:nicolas.simon@univ-amu.fr)), at Aix-Marseille Université. The main source codes needed to reproduce the presented results are available on GitHub ([https://github.com/ins-amu/Stan\\_PKPD\\_examples](https://github.com/ins-amu/Stan_PKPD_examples)).

## Appendix A. Hamiltonian equations of motion

### A.1. Expression of the Hamiltonian

The Hamiltonian function is expressed in terms of parameter  $\theta$  and sampled auxiliary momentum  $\rho$  by developing the joint probability formula,

$$\begin{aligned} H(\theta, \rho) &\equiv -\log \pi(\theta, \rho) \\ &\equiv -\log \pi(\rho | \theta) - \log \pi(\theta) \\ &\equiv K(\rho, \theta) + V(\theta) \end{aligned} \quad (\text{A.1})$$

where  $K(\rho, \theta)$  corresponds to the kinetic energy and  $V(\theta)$  is the potential energy.

### A.2. Hamiltonian equations of motion

Parameters  $(\theta, \rho)$  in phase space are evolved through Hamiltonian dynamics by solving the Hamiltonian equations of motion:

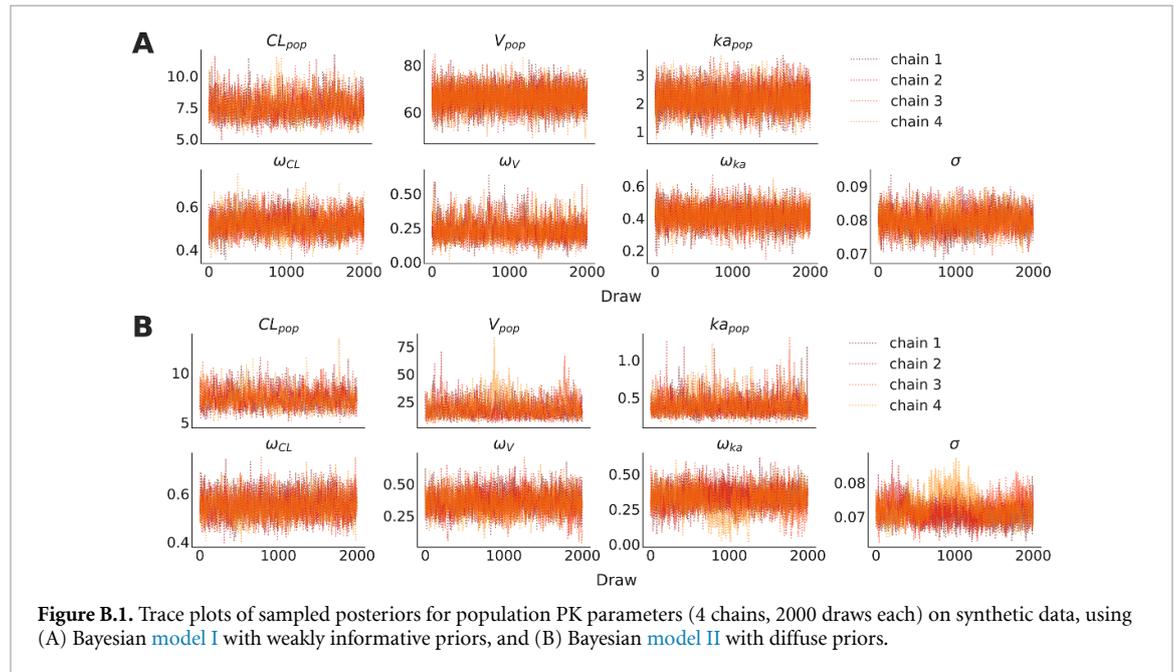
$$\begin{cases} \frac{d\theta}{dt} = \frac{\partial H}{\partial \rho} = \frac{\partial K}{\partial \rho}, \\ \frac{d\rho}{dt} = -\frac{\partial H}{\partial \theta} = -\frac{\partial K}{\partial \theta} - \frac{\partial V}{\partial \theta}. \end{cases} \quad (\text{A.2})$$

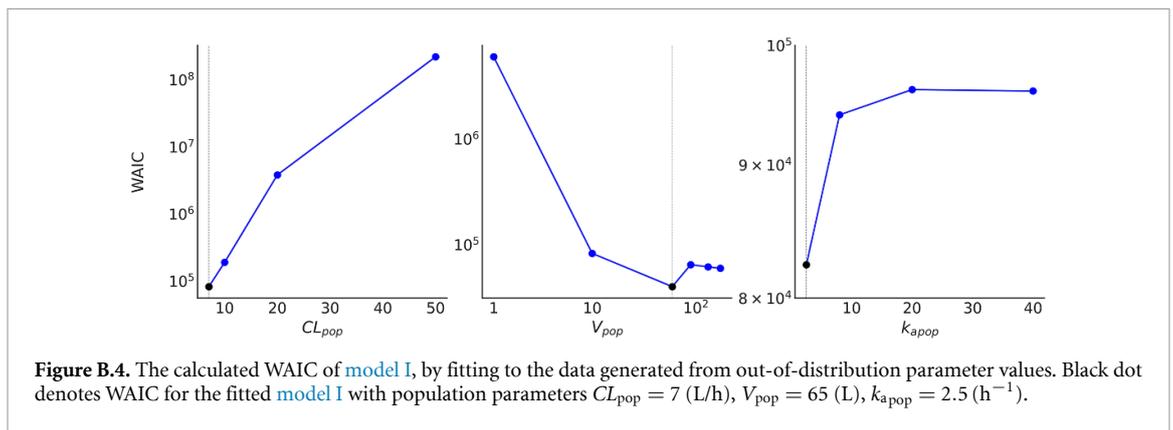
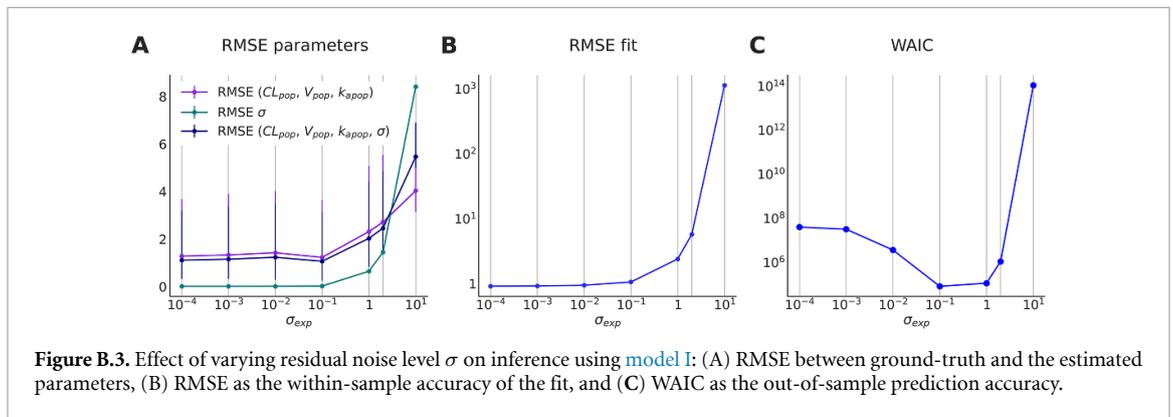
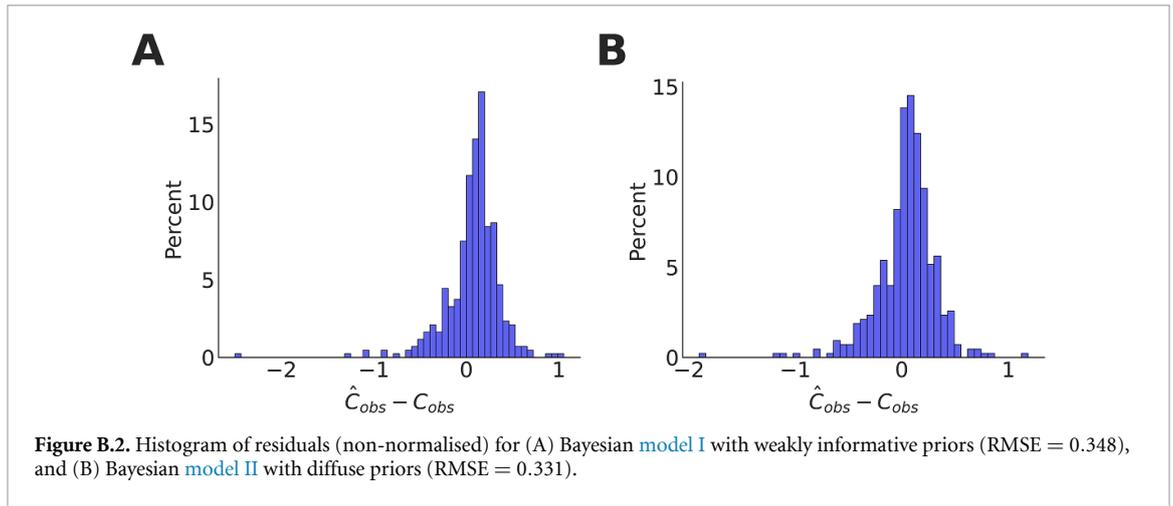
### A.3. Leapfrog integrator

For a number of iterations  $L$ , the leapfrog integrator successively updates the momentum of half a  $\epsilon$  step then parameters of a complete step and then updates the momentum of half a step again.

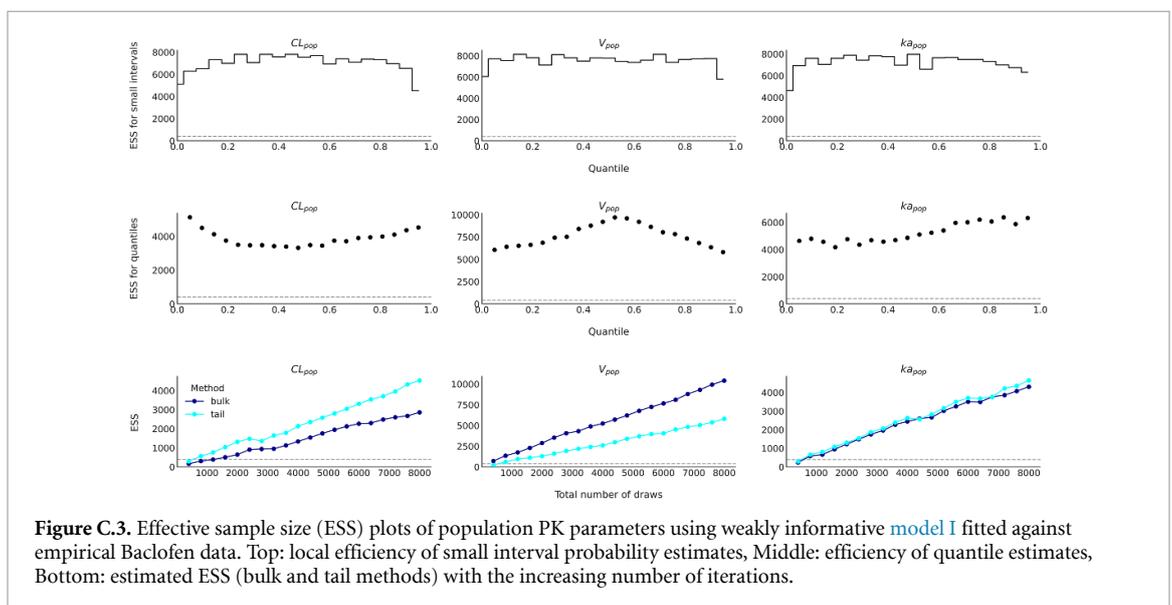
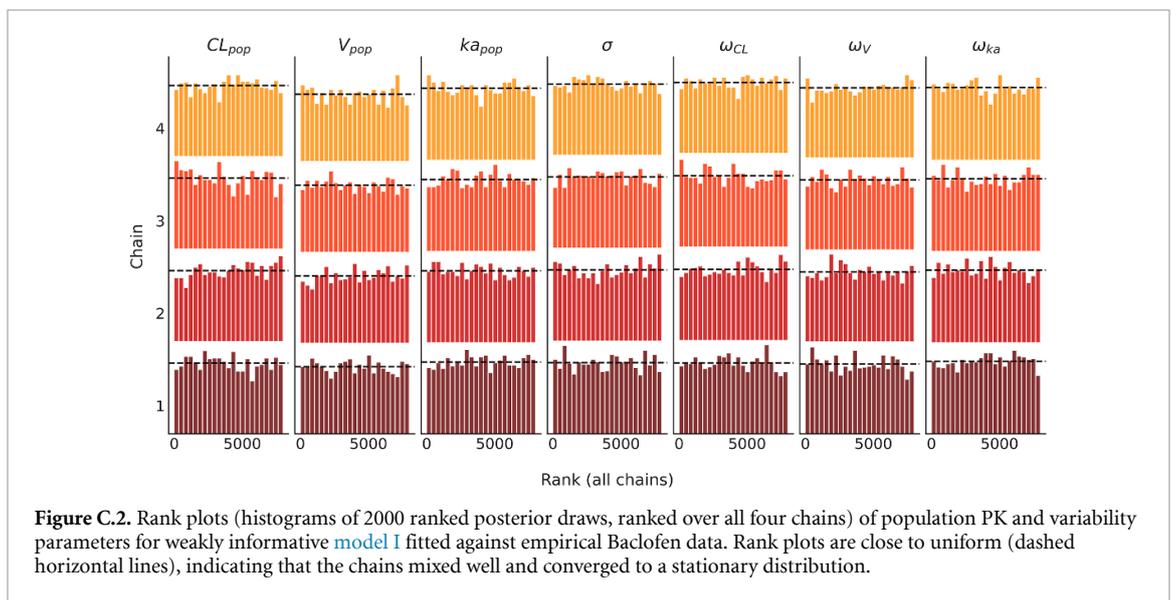
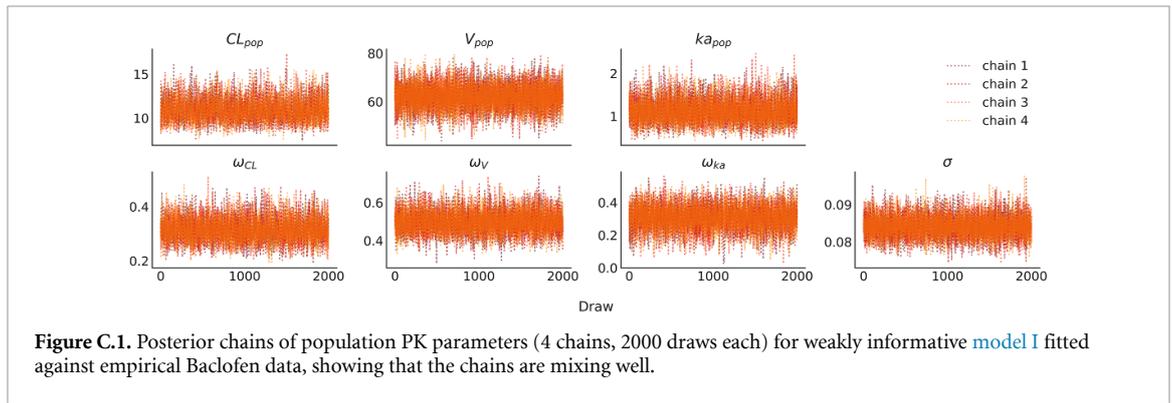
$$\begin{cases} \rho \leftarrow \rho - \frac{\epsilon}{2} \frac{\partial V}{\partial \theta}, \\ \theta \leftarrow \theta + \epsilon M^{-1} \rho, \\ \rho \leftarrow \rho - \frac{\epsilon}{2} \frac{\partial V}{\partial \theta}. \end{cases} \quad (\text{A.3})$$

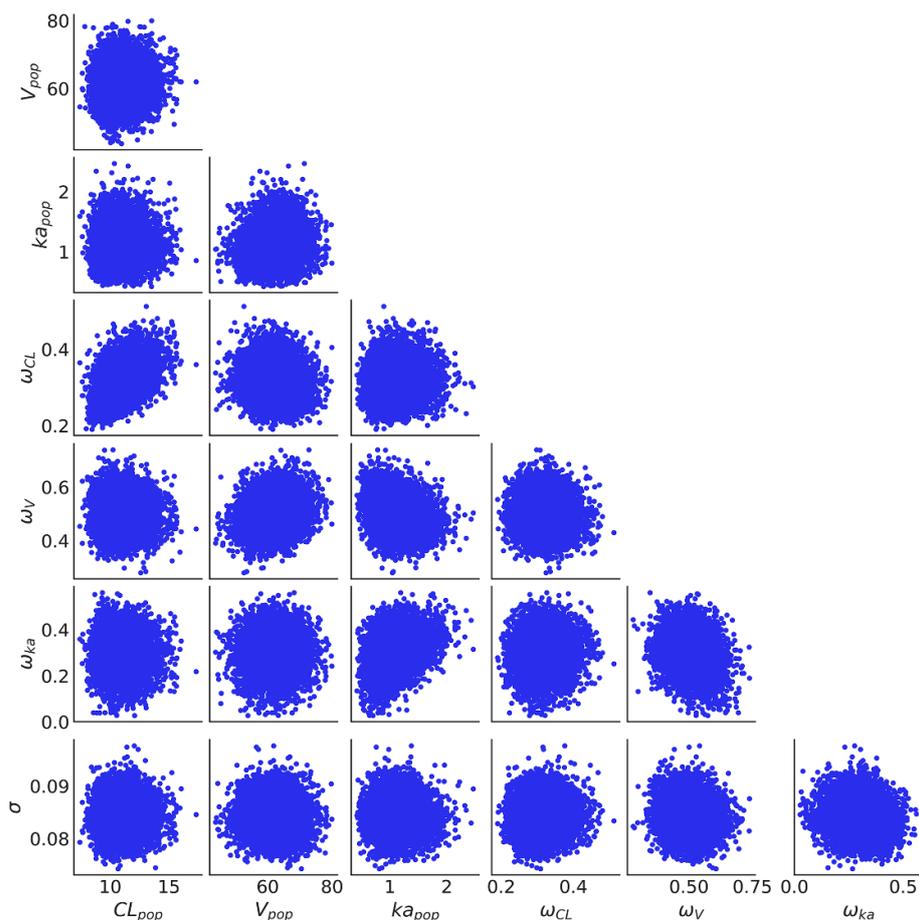
## Appendix B. Bayesian inference on synthetic data



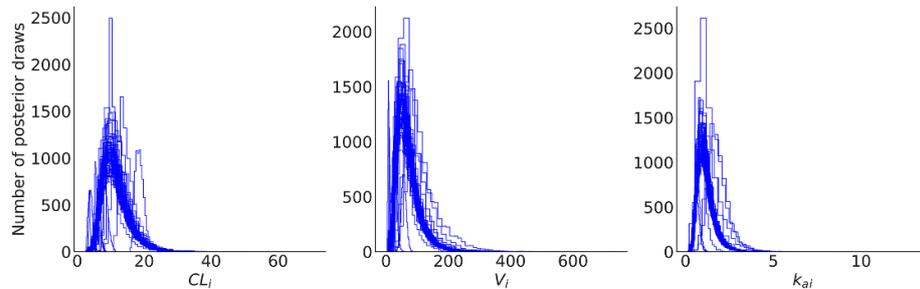


### Appendix C. Bayesian inference on empirical data





**Figure C.4.** Paired posterior draws (4 chains, 2000 draws each) of population PK parameters using weakly informative model I fitted against empirical Baclofen data.



**Figure C.5.** Individual posterior distributions (histograms with 40 bins) of individual one-compartment PK parameters fitted against empirical Baclofen data.

## Appendix D. Workflow of the steady state calculation for one-compartment models with Torsten

Translated to algorithm text from the `PredSS_oneCpt.hpp` file of Torsten library<sup>12</sup>.

**Case of interest:** bolus dose in the gut compartment (`cmt = 1`), with `ss = True`.

Set  $\alpha := (\frac{CL}{V}, k_a)^T$

- Computation of amount in gut  $y_{gut}$ :  
 $\mathbf{a} := (0; 1)^T$

$$\text{pred}(y_{gut}) = \text{PolyExp}(t=ii, \text{amt}, \text{rate}=0, t_{inj}=0, \tau=ii, \text{ss}=\text{True}, \mathbf{a}, \alpha, \text{nCmt}=2)$$

→ Calls `PolyExp.hpp`:

If only one dose (`ii=0`):

For  $i = 0, i+ = 1, i < \text{nCmt}$ :  $\text{result} += \mathbf{a}[i] \exp(-\alpha[i] * t)$

i.e.  $\text{result} = \exp(-k_a * ii)$

Else:

If `SS=True` (case of interest):

For  $i = 0, i+ = 1, i < \text{nCmt}$ :  $\text{result} += \frac{\mathbf{a}[i] \exp(-\alpha[i] * t)}{1 - \exp(-\alpha[i] * \text{tau})}$

That is,  $\text{result} = \frac{1}{\exp(k_a * ii) - 1}$

Return  $\text{result} * \text{amt}$

- Computation of amount in central compartment  $y_{cent}$ :

$\mathbf{a} := (\frac{k_a}{(k_a - (CL/V))}; -\frac{k_a}{(k_a - (CL/V))})^T$

$$\text{pred}(y_{cent}) = \text{PolyExp}(t=ii, \text{amt}, \text{rate}=0, t_{inj}=0, \tau=ii, \text{ss}=\text{True}, \mathbf{a}, \alpha, \text{nCmt}=2)$$

→ Calls `PolyExp.hpp`:

If only one dose (`ii=0`):

For  $i = 0, i+ = 1, i < \text{nCmt}$ :  $\text{result} += \mathbf{a}[i] \exp(-\alpha[i] * t)$

That is,  $\text{result} = 0$

Else:

If `SS=True` (case of interest):

For  $i = 0, i+ = 1, i < \text{nCmt}$ :  $\text{result} += \frac{\mathbf{a}[i] \exp(-\alpha[i] * t)}{1 - \exp(-\alpha[i] * \text{tau})}$

That is,  $\text{result} = \frac{k_a}{(k_a - (CL/V))} \left[ \frac{1}{\exp((CL/V) * ii) - 1} - \frac{1}{\exp(k_a * ii) - 1} \right]$

Return  $\text{result} * \text{amt}$

Return  $\text{result} * \text{amt}$

## ORCID iDs

Viktor K Jirsa  <https://orcid.org/0000-0002-8251-8860>

Meysam Hashemi  <https://orcid.org/0000-0001-5289-9837>

## References

- Addolorato G, Caputo F, Capristo E, Domenicali M, Bernardi M, Janiri L, Agabio R, Colombo G, Luigi Gessa G and Gasbarrini G 2002 Baclofen efficacy in reducing alcohol craving and intake: a preliminary double-blind randomized controlled study *Alcohol Alcohol* **37** 504–8
- Baron K T 2022 `mrgsolve`: simulate from ODE-Based Models R package version 1.0.4 (available at: <https://github.com/metrumresearchgroup/mrgsolve>)
- Beal S, Boeckmann L, Bauer R J and Sheiner L B 2009 Nonmem user's guides (1989-2009)
- Berry D A 2006 Bayesian clinical trials *Nat. Rev. Drug Discovery* **5** 27–36
- Betancourt M 2017 A conceptual introduction to Hamiltonian Monte Carlo (arXiv:1701.02434)
- Betancourt M 2018 Towards a principled Bayesian workflow
- Betancourt M, Byrne S, Livingstone S and Girolami M 2014 The geometric foundations of Hamiltonian Monte Carlo (arXiv:1410.5110)

<sup>12</sup> (<https://github.com/metrumresearchgroup/Torsten>).

- Bishop C M 2006 *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer)
- Bonate P L 2011 *Pharmacokinetic-Pharmacodynamic Modeling And Simulation* vol 20 (Springer)
- Brennan J L, Leung J G, Gagliardi J P, Rivelli S K and Muzyk A J 2013 Clinical effectiveness of baclofen for the treatment of alcohol dependence: a review *Clin. Pharmacol.* **5** 99–107
- Burnham K P and Anderson D 2002 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 2nd edn (Springer)
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P and Riddell A 2017 Stan: a probabilistic programming language *J. Stat. Softw.* **76** 1
- Colombo G, Addolorato G, Agabio R, Carai M A, Pibiri F, Serra S, Vacca G and Gessa G L 2004 Role of GABA(B) receptor in alcohol dependence: reducing effect of Baclofen on alcohol intake and alcohol motivational properties in rats and amelioration of alcohol withdrawal syndrome and alcohol craving in human alcoholics *Neurotox. Res.* **6** 403–14
- Colombo G and Gessa G L 2018 Suppressing effect of Baclofen on multiple alcohol-related behaviors in laboratory animals *Front. Psychiatry* **9** 475
- de Beaurepaire R et al 2019 The use of Baclofen as a treatment for alcohol use disorder: a clinical practice perspective *Front. Psychiatry* **9** 708
- Doherty M and Gratton A 2007 Differential involvement of ventral tegmental GABA<sub>A</sub> and GABA<sub>B</sub> receptors in the regulation of the nucleus accumbens dopamine response to stress *Brain Res.* **1150** 62–68
- Duane S, Kennedy A D, Pendleton B J and Roweth D 1987 Hybrid Monte Carlo *Phys. Lett. B* **195** 216–22
- Eddelbuettel D and Francois R 2011 Rcpp: Seamless r and c++ integration *J. Stat. Softw.* **40** 1–18
- Elmokadem A, Riggs M M and Baron K T 2019 Quantitative systems pharmacology and physiologically-based pharmacokinetic modeling with mrgsolve: a hands-on tutorial *CPT Pharmacometrics Syst. Pharmacol.* **8** 883–93
- Ferreira D, Barthoulot M, Pottecher J, Torp K D, Diemunsch P and Meyer N 2020 Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review *Br. J. Anaesthesia* **125** 201–7
- Garbutt J C, Kampov-Polevoy A B, Pedersen C, Stansbury M, Jordan R, Willing L and Gallop R J 2021 Efficacy and tolerability of Baclofen in a us community population with alcohol use disorder: a dose-response, randomized, controlled trial *Neuropsychopharmacology* **46** 2250–6
- Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A and Rubin D B 2014 *Bayesian Data Analysis* 3rd edn (Chapman & Hall/CRC Texts in Statistical Science)
- Gelman A, Hwang J and Vehtari A 2014 Understanding predictive information criteria for Bayesian models *Stat. Comput.* **24** 997–1016
- Gelman A and Rubin D B 1992 Inference from iterative simulation using multiple sequences *Stat. Sci.* **7** 457–72
- Gelman A, Simpson D and Betancourt M 2017 The prior can often only be understood in the context of the likelihood *Entropy* **19** 555
- Gelman A, Vehtari A, Simpson D, Margossian C C, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P C, and Modrák M 2020 Bayesian workflow
- Grinsztajn L, Semanova E, Margossian C C and Riou J 2021 Bayesian workflow for disease transmission modeling in Stan *Stat. Med.* **40** 6209–34
- Gupta S K 2012 Use of Bayesian statistics in drug development: Advantages and challenges *Int. J. Appl. Basic Med. Res.* **2** 3
- Hashemi M, Hutt A, Buhry L and Sleigh J 2018 Optimal model parameter estimation from eeg power spectrum features observed during general anesthesia *Neuroinformatics* **16** 231–51
- Hashemi M, Vattikonda A N, Sip V, Diaz-Pier S, Peyser A, Wang H, Guye M, Bartolomei F, Woodman M M and Jirsa V K 2021 On the influence of prior information evaluated by fully Bayesian criteria in a personalized whole-brain model of epilepsy spread *PLOS Comput. Biol.* **17** 1–34
- Hashemi M, Vattikonda A N, Sip V, Guye M, Bartolomei F, Woodman M M and Jirsa V K 2020 The Bayesian virtual epileptic patient: a probabilistic framework designed to infer the spatial map of epileptogenicity in a personalized large-scale brain model of epilepsy spread *NeuroImage* **217** 116839
- Hindmarsh A 1992 Odepack. a collection of ODE system solvers vol 1 (available at: [www.osti.gov/biblio/145724](http://www.osti.gov/biblio/145724))
- Hoffman M D and Gelman A 2014 The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo *J. Mach. Learn. Res.* **15** 1593–623
- Imbert B, Alvarez J and Simon N 2015 Anticraving effect of Baclofen in alcohol-dependent patients *Alcohol. Clin. Exper. Res.* **39** 1602–8
- Jack Lee J and Chu C T 2012 Bayesian clinical trials in action *Stat. Med.* **31** 2955–72
- Jha J, Hashemi M, Nihalani Vattikonda A, Wang H and Jirsa V 2022 Fully Bayesian estimation of virtual brain parameters with self-tuning Hamiltonian Monte Carlo *Mach. Learn.: Sci. Technol.* **3** 035016
- Keizer R J, Ter Heine R, Frymoyer A, Lesko L J, Mangat R and Goswami S 2018 Model-informed precision dosing at the bedside: scientific challenges and opportunities *CPT Pharmacomet. Syst. Pharmacol.* **7** 785–7
- Lewandowski D, Kurowicka D and Joe H 2009 Generating random correlation matrices based on vines and extended onion method *J. Multivariate Anal.* **100** 1989–2001
- Margossian C C 2018 A review of automatic differentiation and its efficient implementation *CoRR* (arXiv:1811.05031)
- Margossian C C, Zhang Y and Gillespie W R 2021 Flexible and efficient Bayesian pharmacometrics modeling using stan and torsten, part I (arXiv:2109.10184)
- Meibohm B and Derendorf H 1997 Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling *Int. J. Clin. Pharmacol. Ther.* **35** 401–13
- Neal R M 2011 MCMC using Hamiltonian dynamics *Handbook of Markov Chain Monte Carlo* vol 54 (CRC Press) pp 113–62
- Owen J and Fiedler-Kelly J S 2014 Introduction to population pharmacokinetic/pharmacodynamic analysis with nonlinear mixed effects models (<https://doi.org/10.1002/9781118784860>)
- Radhakrishnan K and Hindmarsh A C 1993 Description and use of LSODE, the livermore solver for ordinary differential equations *Technical Report*
- Rolland B, Simon N, Franchitto N and Aubin H-J 2020 France grants an approval to Baclofen for alcohol dependence *Alcohol Alcohol.* **55** 44–45
- Schuckit M 2009 Alcohol use disorders *Lancet* **373** 492–501
- Sheiner L B and Ludden T M 1992 Population pharmacokinetics/dynamics *Annu. Rev. Pharmacol. Toxicol.* **32** 185–209
- Simon N, Moirand R, Dematteis M, Bordet R, Deplanque D and Rolland B 2018 Full-profile pharmacokinetic study of high dose Baclofen in subjects with alcohol use disorder *Front. Psychiatry* **9** 385
- Spiegelhalter D J, Myles J B, Jones D R and Abrams K R 1999 An introduction to Bayesian methods in health technology assessment *BMJ* **319** 508–12

- Upton R N and Mould D R 2014 Basic concepts in population modeling, simulation and model-based drug development: part 3-introduction to pharmacodynamic modeling methods *CPT: Pharmacomet. Syst. Pharmacol.* **3** 1–16
- Vattikonda A N, Hashemi M, Sip V, Woodman M M, Bartolomei F and Jirsa V K 2021 Identifying spatio-temporal seizure propagation patterns in epilepsy using Bayesian inference *Commun. Biol.* **4** 1–13
- Vehtari A, Gelman A, Simpson D, Carpenter B and Bürkner P-C 2021 Rank-normalization, folding and localization: an improved  $\hat{r}$  for assessing convergence of MCMC (with discussion) *Bayesian Anal.* **16** 667–718
- World Health Organization 2022 Alcohol fact sheet (available at: [www.who.int/news-room/fact-sheets/detail/alcohol](http://www.who.int/news-room/fact-sheets/detail/alcohol)) (Accessed 12 September 2022)
- Yarnell C J et al 2021 Clinical trials in critical care: can a Bayesian approach enhance clinical and scientific decision making? *Lancet Respirat. Med.* **9** 207–16
- Zhang Y, Gillespie W R and Margossian C 2021 Torsten user guide (available at: <https://metrumresearchgroup.github.io/Torsten/>)
- Zou H, Banerjee P, Leung S S Y and Yan X 2020 Application of pharmacokinetic-pharmacodynamic Modeling in drug delivery: development and challenges *Front. Pharmacol.* **11** 997