



Local Identifiability Analysis, Parameter Subset Selection and Verification for a Minimal Brain PBPK Model

Kamala Dadashova¹ · Ralph C. Smith¹  · Mansoor A. Haider¹ 

Received: 18 May 2023 / Accepted: 3 November 2023 / Published online: 3 January 2024

© The Author(s), under exclusive licence to Society for Mathematical Biology 2023

Abstract

Physiologically-based pharmacokinetic (PBPK) modeling is important for studying drug delivery in the central nervous system, including determining antibody exposure, predicting chemical concentrations at target locations, and ensuring accurate dosages. The complexity of PBPK models, involving many variables and parameters, requires a consideration of parameter identifiability; i.e., which parameters can be uniquely determined from data for a specified set of concentrations. We introduce the use of a local sensitivity-based parameter subset selection algorithm in the context of a minimal PBPK (mPBPK) model of the brain for antibody therapeutics. This algorithm is augmented by verification techniques, based on response distributions and energy statistics, to provide a systematic and robust technique to determine identifiable parameter subsets in a PBPK model across a specified time domain of interest. The accuracy of our approach is evaluated for three key concentrations in the mPBPK model for plasma, brain interstitial fluid and brain cerebrospinal fluid. The determination of accurate identifiable parameter subsets is important for model reduction and uncertainty quantification for PBPK models.

Keywords Parameter subset selection · Local sensitivity analysis · Identifiability · PBPK modeling · Energy statistics

Ralph C. Smith and Mansoor A. Haider have contributed equally to this work.

✉ Mansoor A. Haider
mahaider@ncsu.edu

Kamala Dadashova
kdadash@ncsu.edu

Ralph C. Smith
rsmith@ncsu.edu

¹ Department of Mathematics, North Carolina State University, Box 8205, Raleigh, NC 27695, USA

1 Introduction

Physiologically-based pharmacokinetic (PBPK) models commonly include a large number of parameters and several model responses (dependent variables) of interest. Calibrating model parameters can be challenging and depends on the number of responses of physiological interest, the model's inherent structure and the scope of available data. Typically, many parameters cannot be directly measured and instead must be specified at a set of nominal values obtained from or motivated by the available literature. Sensitivities of individual model responses near the set of nominal or estimated parameter values are also of interest in many applications. However, these sensitivities can be different depending on the particular response of interest and the time at which it is evaluated. Overall, it is crucial to understand how variation or uncertainty of the parameters in PBPK models affect individual responses across the time domain of interest. Techniques that quantify such relationships enable one to fix values of a subset of the parameters and lead to more robust optimization, statistical inference or uncertainty quantification focused on these responses. Such techniques can also inform the design of experiments when choosing one (or a few) data sources, among several choices of observed responses, is advantageous.

Local sensitivity analysis has been widely used to study how PBPK model responses depend on perturbations in parameter values relative to a set of baseline or nominal values (Clewell et al. 1994; Evans et al. 2020; Evans and Andersen 1995; Matthews et al. 2009). When parameter ranges across the scope of application of a PBPK model are known, global sensitivity analysis methods can also be applied (Hsieh et al. 2018). Identifiability refers to the ability to determine a unique set of parameter values given a model response or a set of responses. Local identifiability analysis is carried out in the context of data from experiments or for model responses of interest at a set of nominal parameter values. Furthermore, the accuracy of a parameter subset deemed to be locally identifiable for a particular response (based on its values across the full time domain) can vary depending upon the time at which this response is evaluated.

Structural identifiability analysis has the objective of determining parameter dependencies which are global in the sense that they persist over the full time domain of the model. Several software packages exist for structural identifiability analysis of model parameters in ODE systems (Bellu et al. 2007; Hong et al. 2019). Structural identifiability analysis has been widely used for noise-free PBPK models that are small enough to be analytically manipulated or that can be linearized without loss of model accuracy (Brown et al. 2022; Calvier et al. 2018; Carter et al. 2020; Kendrick et al. 2019; Slob et al. 1997; Kendrick et al. 2017; Yates 2006). Techniques for scaling up to larger dynamical systems have also been developed, but may be challenging to execute in practice, due to the large number of nonlinear algebraic equations that arise (Bellu et al. 2007). Such approaches require assumptions that the functional dependence in the system involves only polynomials or rational functions (Raue et al. 2014). Overall, parameters deemed to be non-identifiable via structural identifiability analysis may not capture variations in parameter identifiability that occur more locally in time, i.e. when a particular response is evaluated within a subdomain of the full time domain, or for different sets of nominal parameter values.

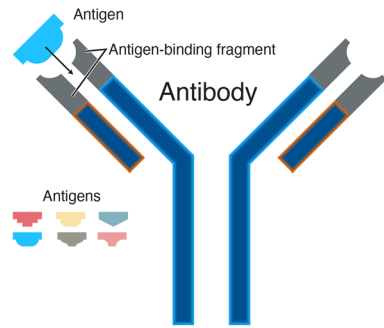
Local identifiability analysis is well-suited to PBPK models since its partitioning of parameters into identifiable and non-identifiable subsets can change when studying different time regimes, for different model variables or quantities of interest, and for differing nominal parameter values. This is due, in part, to the diverse range of underlying time scales that are inherent to PBPK models. These approaches are based on local sensitivities (Clewell et al. 1994; Evans et al. 2020; Evans and Andersen 1995; Matthews et al. 2009; Pearce et al. 2021), and can also involve properties of the Fisher information matrix (Koyama et al. 2021; Lavezzi et al. 2018). Properties of this matrix are also employed in the model sloppiness literature, using both standard and Bayesian approaches (Apgar et al. 2010; Gutenkunst et al. 2007; Monsalve-Bravo et al. 2022; Transtrum et al. 2015). However, this approach is strongly tied to the data when sensitivity derivatives are applied to residuals for model responses relative to data for the same responses. By contrast, our approach considers specific model responses across all times in the domain of interest, and in the absence of data from experiments.

Due to this variability in local identifiability across both model responses and times, techniques for evaluating the accuracy of identifiable PBPK parameter subsets are needed. In this paper, we develop and evaluate such techniques and approaches in the context of a minimal PBPK (mPBPK) model of the brain for antibody therapeutics (Bloomingdale et al. 2021). This mPBPK model is a reduced form of a previous multi-species platform brain PBPK model comprising 100 differential equations (Chang et al. 2019); the mPBPK model contains 16 differential equations and 31 parameters.

We construct a local sensitivity-based parameter subset selection (PSS) algorithm to determine subsets of identifiable parameters for three responses (concentrations) of interest in the mPBPK model, relative to a set of nominal parameter values. In our PSS approach, the user specifies a single time domain of interest for the identifiability analysis. For the mPBPK model (Bloomingdale et al. 2021), this time domain is 0–1,000 h (at 1 h intervals), and captures the underlying dynamics for the three responses of interest (see also Fig. 4). We also systematically evaluate the performance of this PSS algorithm, comparing it to a more standard PSS approach using qualitative and quantitative (energy statistics) methods. This includes considering both the accuracy of the identifiable parameter subsets and the calibration of threshold values in our PSS algorithm for model predictions of key concentrations across the full time domain of the simulations. In particular, for each of these three responses of interest we evaluate the accuracy of the parameter subsets at each time point across the full time domain.

We first present a summary of the mPBPK model in Sect. 2, including its physiological structure. In Sect. 3, we outline the methods and algorithms for parameter subset selection, including two approaches for verifying accuracy of the subsets of identifiable parameters. Our results are presented in Sect. 4, for three key responses of interest in the model which are the plasma concentration, the brain interstitial fluid concentration, and the brain cerebrospinal fluid concentration. We end with a brief discussion and conclusions in Sect. 5.

Fig. 1 Antibodies and antigens
(National Human Genome
Research Institute 2023)



2 Model

An antibody (Fig. 1) is an immune system protein that circulates in the blood, identifies foreign molecules known as antigens (e.g. bacteria, viruses), and neutralizes them. Therapeutic antibodies are used to treat a wide range of diseases including those affecting the central nervous system (CNS), as well as cancer, autoimmune diseases, cardiovascular diseases, inflammation, and allergy (Dwek 2009). The Food and Drug Administration (FDA) authorized the first antibody in 1986 (Todd et al. 1989), called muromonab-CD3, whose goal was to reduce acute rejection in recipients of organ transplants. The development of novel therapeutic antibodies remains an active area of research and development with many undergoing clinical trials and others still in preclinical development.

Antibody therapies for CNS disorders are lacking, thus motivating the need for drug discovery. One challenge in CNS drug development is that the blood-brain barrier (BBB) prevents sufficient quantities of large molecules and small compounds from entering the brain to activate the desired target (Paul 2011). Furthermore, measurement of antibody exposure is difficult due to the lack of knowledge concerning antibody location in the brain. Thus, mathematical models that can predict antibody concentrations in CNS compartments, and exchange of antibodies across compartments are important components of CNS research and development for novel antibody therapies.

Physiologically-based pharmacokinetic (PBPK) models are widely used for this purpose (Jones and Rowland-Yeo 2013). PBPK models represent physiological structure and functions that influence a chemical's disposition. Pharmacokinetics accounts for chemical absorption, distribution, metabolism, and excretion. PBPK models also use parameters and equations that incorporate body weight, blood flow rate, and metabolic rate.

2.1 Model Structure

To develop our methods, we employ the minimal PBPK (mPBPK) model introduced in Bloomingdale et al. (2021), which was constructed by reducing the full PBPK model in Chang et al. (2019). As detailed in Bloomingdale et al. (2021), this reduction decreases the complexity of the full model while retaining its primary physiological

features. The mPBPK compartments are shown in Fig. 2, and many of the model simplifications are obtained by combining compartments from the original model. For example, a tissue compartment in the mPBPK model was formed by combining fourteen non-brain organ compartments in the full model. In the reduced model, the tissue compartment comprises three subcompartments: vascular, endosomal, and interstitium. Furthermore, a cerebrospinal fluid (CSF) compartment was created by merging four CSF compartments in the full model (lateral ventricle, third-fourth ventricle, cisterna magna, and subarachnoid space), while retaining the endosomal, interstitial and brain vascular compartments. Nominal values of the model parameters for humans are given in Table 5 in the Appendix. Note that three nominal parameter values in the full PBPK model were obtained using maximum likelihood estimation based on published and in-house pharmacokinetic data; the remaining values were fixed based on values reported in prior literature (see Chang et al. 2019). We refer readers to Bloomingdale et al. (2021) for further information about the structural reduction of the full PBPK model to the mPBPK model.

2.2 Model Equation

Each response (dependent variable) in the model represents an evolving antibody concentration, and the corresponding differential equation describes its rate of change in the respective compartment (Fig. 2). To illustrate, the differential equation for the change in plasma antibody concentration (C_P) is

$$V_P \cdot \frac{dC_P}{dt} = (Q_T - L_T) \cdot C_{T_V} + (Q_B - L_B) \cdot C_{B_V} + (L_T + L_B) \cdot C_L - Q_T \cdot C_P - Q_B \cdot C_P. \quad (1)$$

Each term on the right-hand side of (1) models the mass transport of plasma antibodies into (postive terms) and out of (negative terms) the plasma compartment, based on the rates specified in Fig. 2. Here, C_P , C_{T_V} , C_{B_V} and C_L , denote concentrations in the plasma, tissue vascular, brain vascular, and lymph compartments, respectively. Q_T , L_T , Q_B , and L_B are parameters governing the rates of transport, and V_P is the (fixed) volume of the plasma. The full mPBPK model is summarized in the Appendix.

The mPBPK model can be expressed in the following general form:

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(t, \mathbf{u}(t), \boldsymbol{\beta}), \quad \mathbf{u}(t_0) = \mathbf{u}_0(\boldsymbol{\beta}). \quad (2)$$

Here $\mathbf{u}(t) = [u_1, \dots, u_m]^T$ ($m = 16$) are the model variables (or responses), $\boldsymbol{\beta} = [\beta_1, \dots, \beta_r]^T$ ($r = 31$), and $\boldsymbol{\theta} = [\boldsymbol{\beta}, \mathbf{u}_0]$ comprises the full set of p model parameters. Nominal values for these parameters, based on Bloomingdale et al. (2021), are given in the Appendix. Note that, for this model, initial conditions for only three responses are assumed to be nonzero, and sensitivities with respect to the initial conditions are not incorporated in our identifiability analysis.

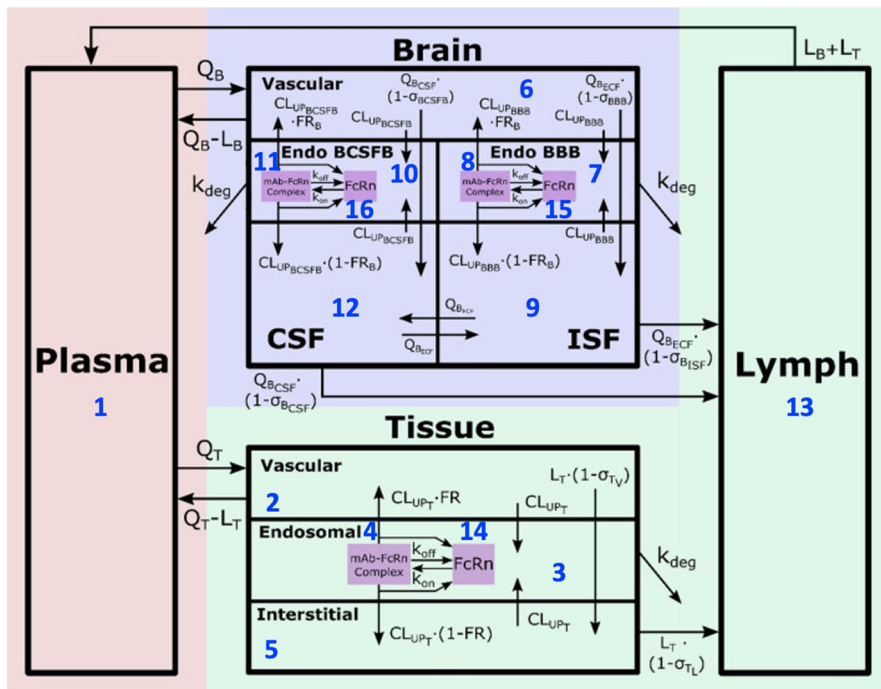


Fig. 2 Brain mPBPK model structure (Bloomingdale et al. 2021). Compartments included in the model represent antibody concentrations in (1) Plasma, (2) Tissue vascular, (3) Tissue endosome (unbound), (4) Tissue endosome (FcRnbound), (5) Tissue interstitium, (6) Brain vascular, (7) BBB endosome (unbound), (8) BBB endosome (FcRnbound), (9) Brain interstitium, (10) BCSFB endosome (unbound), (11) BCSFB endosome (FcRnbound), (12) Brain CSF, (13) Lymph, (14) FcRn concentration in tissue endosome, (15) FcRn BBB endosome, and (16) FcRn BCSFB endosome. This image is adapted from Bloomingdale et al 2021

3 Methods

In this section, we summarize techniques for parameter identifiability analysis, including verification methods, based on local sensitivities about a set of nominal parameter values and a specified time domain of interest. We outline two parameter subset selection (PSS) algorithms to determine identifiable and non-identifiable subsets of mPBPK model parameters for a chosen model response. We then summarize verification techniques for evaluating the accuracy of identifiable parameter subsets for the model response evaluated at each time across the prescribed time domain of interest (0–1000 h).

3.1 Sensitivity Analysis

Sensitivity analysis is frequently used to quantify how changes in model parameters impact the response(s). In this investigation, we quantify the impact of model parameters on antibody concentrations using local (sensitivity-based) identifiability analysis

(Burth et al. 1999). For models with a known set of nominal values and weak-to-moderate levels of nonlinearity, local sensitivity analysis is a reliable and efficient approach. In such problems, sensitivity analysis can achieve a variety of objectives, including model reduction (by fixing non-influential model parameters), better strategies for data collection in experiments, and identifying possible limitations of the model (Brown et al. 2022). Note that our analysis of local sensitivities only considers the influence of one parameter at a time. By contrast, global sensitivity analysis accounts for uncertainties in a response that depend on uncertainties in parameters.

3.1.1 Local Sensitivity Analysis

Local sensitivities quantify variability of the model responses $\mathbf{u}(t_i)$ when input parameters $\boldsymbol{\theta}$ are perturbed about nominal parameter values $\boldsymbol{\theta}^*$ (Smith 2014). Note that the concentrations in the model represent $\mathbf{u}(t)$ and $\boldsymbol{\theta} = \boldsymbol{\beta}$. Local sensitivities are computed at n discrete time points using the partial derivatives

$$s_j(t_i) = \frac{\partial \mathbf{u}(t_i)}{\partial \theta_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, p, \quad (3)$$

of the model responses $\mathbf{u}(t_i)$ with respect to the individual parameters. Here, s_j is the local sensitivity with respect to the j^{th} parameter θ_j , t_i is the i^{th} time point, p is the number of parameters in the model, and n is the number of time steps. In this investigation, sensitivity analysis is applied only to parameters which are coefficients in (2), not the initial states; henceforth, $p = r$ and $\boldsymbol{\theta} = \boldsymbol{\beta}$. Local sensitivities can be computed by either solution of the sensitivity equations or via a numerical approximation. The former approach can be tedious for moderate to large size models usually requiring the numerical solution of a system of ODEs, whereas the latter approach is extensively utilized in practice.

We first demonstrate the process of deriving the sensitivity equations for the ODE model (2). We differentiate equation (2) with respect to β_k , and interchange the order of differentiation to obtain

$$\frac{ds}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{u}} \mathbf{s}(t) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}, \quad \mathbf{s}(t_0) = \frac{\partial \mathbf{u}_0}{\partial \boldsymbol{\beta}}, \quad (4)$$

where $\mathbf{s}(t) = [\mathbf{s}_1(t), \mathbf{s}_2(t), \dots, \mathbf{s}_p(t)]_{n \times p}$ is constructed from the sensitivity vectors $\mathbf{s}_k(t) = \frac{\partial \mathbf{u}}{\partial \beta_k}$, $k = 1, \dots, p$. For brevity, the dependence of the sensitivities on the model parameters is suppressed. Here, $\frac{\partial \mathbf{f}}{\partial \mathbf{u}}$ is an $(n \cdot p) \times (n \cdot p)$ block diagonal matrix having p blocks consisting of the $n \times n$ Jacobians on each block, and $\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}$ is a column vector with np components. To illustrate, consider a case with $n = 2$ responses C_P and C_{TV} , with $p = 2$ parameters $\boldsymbol{\beta} = [Q_T, FR]^T$. We obtain

$$\mathbf{s}(t) = \left[\frac{\partial C_P}{\partial Q_T}, \frac{\partial C_{TV}}{\partial Q_T}, \frac{\partial C_P}{\partial FR}, \frac{\partial C_{TV}}{\partial FR} \right]^T,$$

$$\frac{\partial f}{\partial \mathbf{u}} = \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} \frac{-(Q_T + Q_B)}{V_P} & \frac{(Q_T - L_T)}{V_P} \\ \frac{Q_T}{V_{TV}} & -\frac{((Q_T - L_T) + (1 - \sigma_{TV}) \cdot L_T + CLU P_T)}{V_{TV}} \end{bmatrix},$$

$$\frac{\partial f}{\partial \beta} = [C_{TV} - C_P, C_P - C_{TV}, 0, CLU P_T \cdot C_{T_{E,B}}]^T.$$

Alternatively, one can employ numerical techniques, such as finite-difference or complex-step approximations (Lyneess and Moler 1967), or automatic-differentiation (Gautschi 1997) to estimate local sensitivities. For a quantity of interest $f(t_j, \theta^*)$ evaluated about its nominal parameter values θ^* , finite-difference approximations are evaluated by using

$$\frac{\partial f(t_j, \theta^*)}{\partial \theta_i} = \frac{f(t_j, \theta^* + \mathbf{e}_i h_i) - f(t_j, \theta^*)}{h_i} + \mathcal{O}(h_i), \quad (5)$$

where $h_i (\ll 1)$ is a scalar step-size and $\mathbf{e}_i$ is i^{th} unit vector. This numerical technique has certain limitations. One potential issue is roundoff error due to subtractive cancellation, which occurs when both the denominator and the numerator are sufficiently small. Another limitation is that the selection of h_k must be determined while taking the magnitude of the parameters into account.

The complex-step approximation (Banks et al. 2015; Martins et al. 2003) is a more efficient and straightforward approach compared to deriving sensitivity equations for complicated models. It is defined as

$$\frac{\partial f(t_j, \theta^*)}{\partial \theta_i} = \frac{Im(f(t_j, \theta^* + I \mathbf{e}_i h_i))}{h_i} + \mathcal{O}(h_i^2), \quad (6)$$

where $I \equiv \sqrt{-1}$. Unlike the finite-difference method, it is nearly insensitive to the step-sizes since it avoids subtractive cancellation. Furthermore, it is numerically demonstrated to be accurate up to points of discontinuity.

3.1.2 Scaling Techniques for Parameters

Many models require scaling of local sensitivities prior to their use in identifiability analysis. Two examples occur when one has a significant difference in the magnitude of parameters or when local sensitivities do not have the same units. For this scaling, since we have positive parameter values, we use a log transformation $\hat{\theta}_j = \log(\theta_j)$ (Keene 1995). The log-scaled sensitivities about the nominal values θ^* are

$$\hat{s}_j(t) \equiv \frac{\partial \mathbf{u}(t)}{\partial \hat{\theta}_j} = \frac{\partial \mathbf{u}(t)}{\partial \theta_j} \theta_j^* = s_j(t) \theta_j^*, \quad j = 1, \dots, p. \quad (7)$$

3.1.3 Identifiability of the Parameters

PBPK models consist of ordinary differential equations having many unknown parameters. Some of these parameters are often practically non-identifiable, implying that

their values cannot be uniquely determined from experimental data or synthetic data. An example of the latter is a simulated response based on perturbations of a subset of parameters about their nominal values. There are several reasons for the lack of influence on the measured outputs, such as algebraic dependencies among parameters or poor data availability or quality.

3.2 Parameter Subset Selection

Once we obtain local sensitivities for specific model responses of interest using techniques outlined in Sect. 3.1.1, we can use them to determine identifiable parameters for each response. Fixing non-identifiable parameters at nominal values is useful for subsequent optimization, frequentist or Bayesian inference, or uncertainty quantification. Nominal values are often a starting point for such techniques. Fixing non-identifiable parameters at their nominal values often leads to more accurate estimates of values for the identifiable parameters. This is due to more robust performance of the associated algorithms since the computational cost of these procedures decreases as fewer parameters need to be explored.

In this section, we present parameter subset selection (PSS) algorithms for determining identifiability of the model parameters for a particular response on a specified time domain of interest. We also outline two techniques to verify accuracy of results obtained from our PSS algorithms when simulating the same response at each time in the domain of interest.

3.2.1 Parameter Subset Selection Algorithm (PSS)

We focus on the statistical observation model

$$y_i = f(t_i, \boldsymbol{\theta}) + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n, \quad (8)$$

where y_i is experimental data or synthetic data, t_i is time, $\boldsymbol{\theta}$ is the parameter vector, and ε_i is the observation error.

The mean squared error of the response relative to the data is

$$J(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n [y_i - f(t_i, \boldsymbol{\theta})]^2 \quad (9)$$

Now consider the Taylor series expansion of the response about the nominal parameter values; i.e., $\boldsymbol{\theta} = \boldsymbol{\theta}^*$

$$f(t_i, \boldsymbol{\theta}^* + \Delta\boldsymbol{\theta}) \approx f(t_i, \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} f(t_i, \boldsymbol{\theta}^*) \cdot \Delta\boldsymbol{\theta}$$

Since $y_i \approx f(t_i, \theta^*)$, substituting into (9) yields

$$J(\theta^* + \Delta\theta) \approx \frac{1}{n} \sum_{i=1}^n (\nabla_{\theta} f(t_i, \theta^*) \cdot \Delta\theta)^2 = \frac{1}{n} (S\Delta\theta)^T (S\Delta\theta) = \frac{1}{n} \Delta\theta^T S^T S \Delta\theta, \quad (10)$$

where the sensitivity matrix S is

$$S(\theta^*) = \begin{bmatrix} \frac{\partial f}{\partial \theta_1}(t_1, \theta^*) & \dots & \frac{\partial f}{\partial \theta_p}(t_1, \theta^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial \theta_1}(t_n, \theta^*) & \dots & \frac{\partial f}{\partial \theta_p}(t_n, \theta^*) \end{bmatrix}. \quad (11)$$

Denoting an eigenpair of $S^T S$ by $(\lambda, \Delta\theta)$, we obtain

$$J(\theta^* + \Delta\theta) \approx \frac{\lambda}{n} \|\Delta\theta\|_2^2, \quad (12)$$

where $F = S^T S$ is the scaled Fisher information matrix. This provides a way to quantify information about parameter identifiability and correlations among parameters. The eigenvalues of F are real and nonnegative since F is symmetric and nonnegative definite. This implies that if we let $\lambda \approx 0$ in (12), then $J(\theta^* + \Delta\theta) \approx 0$ and respective parameters are non-identifiable at θ^* . We note that there are various theoretical parameter subset selection algorithms (Burth et al. 1999; Cintrón-Arias et al. 2009; Friswell et al. 1997; Kim and Lee 2019; Pearce et al. 2021; Quaizer and Mönningmann 2009) which employ the eigenvalues of the Fisher information matrix. However, using these methods may result in a loss of accuracy when computing the small eigenvalues of $F = S^T S$ for certain matrices S (Ipsen 2009). We address this issue by computing the singular values of S instead, as has been done in several other studies on parameter identifiability (Monsalve-Bravo et al. 2022; Stigter and Molenaar 2015).

Consequently, we use the singular value decomposition (SVD) of the scaled sensitivity matrix, based on (7), as $\hat{S} = U \Sigma V^T$, where the diagonal matrix Σ contains the singular values of $\hat{S}$, and the orthogonal matrix V^T contains its right singular vectors. The intuition behind the resulting algorithms is to sequentially remove the least identifiable parameter and determine a subset of the uniquely identifiable parameters.

We consider two approaches termed *all-at-once* (AAO, Algorithm 1) and *one-at-a-time* (OAT, Algorithm 2). In the former case, a threshold η is used once to partition the singular values and use information in the singular vectors to flag parameters as non-identifiable. In contrast, in the latter case the dimension of $\hat{S}$ is reduced by one at each iteration, corresponding to labeling a parameter as unidentifiable and removing it from the sensitivity matrix, also based on a prescribed threshold η . The SVD is recomputed at each iteration. Each algorithm produces a subset θ_{id} of identifiable parameters; the complementary (non-identifiable) subset is denoted θ_{nid} .

In both algorithms, the SVD is used to determine which parameters have the strongest effects on the outputs in the system. We compute the ratio of the largest

to the smallest singular value, and if its square is greater than the threshold value, all parameters are considered identifiable (line 5). If this ratio is less than the threshold, some parameters are non-identifiable.

We choose the smallest singular value and its corresponding singular vector, for which we find the position of the largest component in the singular vector; this position corresponds to the least identifiable parameter. The algorithm then removes this parameter from the set of identifiable parameters and repeats the process, iteratively, for the remaining parameters. We carry out this procedure either all-at-once by computing the SVD one time (Algorithm 1), or one-at-a-time (Algorithm 2) by recomputing the SVD at each iteration, with the number of columns in $\mathbf{S}$ decreasing by one at each iteration.

There is no precise rule for choosing the threshold value η . Its value needs to be calibrated in a problem-dependent manner. Therefore, the resulting subset of non-identifiable parameters should be verified using the techniques presented in the next section, ideally over a range of η values spanning a few orders of magnitude.

Algorithm 1 PSS AAO (all-at-once)

Input: Nominal $p \times 1$ parameter vector θ^* , $t > 0$, and $n > p$ values t
Output: Subset $\theta_{id} \subseteq \theta$ of identifiable parameters

- 1: Set threshold $0 < \eta < 1$ and specify a nominal input vector θ^* .
- 2: Construct $n \times p$ scaled sensitivity matrix $\hat{\mathbf{S}}(\theta^*)$ having entries $\hat{S}_{ij} = \theta_j^* \frac{\partial f}{\partial \theta_j}(t_i, \theta)$
- 3: Compute the scaled singular value decomposition $\hat{\mathbf{S}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$, where the $p \times p$ diagonal matrix $\mathbf{\Sigma}$ contains the singular values $\sigma_1 \geq \dots \geq \sigma_p \geq 0$ of $\hat{\mathbf{S}}$, and $p \times p$ orthogonal matrix $\mathbf{V}^T$ contains the right singular vectors.
- 4: **for** $k \rightarrow p$ to 1 **do**:
- 5: **if** $\left(\frac{\sigma_k}{\sigma_1}\right)^2 > \eta$ **then**
- 6: Return $\theta_{id} = \theta$, i.e. all parameters currently in θ are identifiable.
- 7: **else**
- 8: i) For the singular (column) vector $\tilde{v}_k \in \mathbf{V}^T$ associated with σ_k , determine the position l of component with the largest magnitude, which corresponds to the least identifiable parameter.
- 9: ii) Remove the l^{th} column in $\mathbf{V}^T$, and the corresponding parameter (θ_l) from the identifiable subset.
- 10: **end if**
- 11: **end for**

3.3 Verification Techniques

In this section, we provide two techniques to evaluate accuracy of the PSS algorithm outcomes and select appropriate η values. Once we have identifiable parameter subsets, we can verify their accuracy using quantitative or qualitative methods across the time domain of interest. Both approaches ensure the validity of the algorithm and the selected threshold value.

3.3.1 Qualitative Method with Relative Error

To verify the PSS algorithm's outcomes, we compare response distributions with all parameters randomly sampled to those when only identifiable parameters are randomly

Algorithm 2 PSS OAT (one-at-a-time)

Input: Nominal $p \times 1$ parameter vector θ^* , $t > 0$, and $n > p$ values t
Output: Subset $\theta_{id} \subseteq \theta$ of identifiable parameters

- 1: Set threshold $0 < \eta < 1$ and specify a nominal input vector θ^* .
- 2: Construct $n \times p$ scaled sensitivity matrix $\hat{S}(\theta^*)$ having entries $\hat{S}_{ij} = \theta_j^* \frac{\partial f}{\partial \theta_j}(t_i, \theta)$
- 3: **for** $k \rightarrow 1$ to p **do**:
- 4: Compute the scaled singular value decomposition $\hat{S} = U \Sigma V^T$, where the $r \times r$ diagonal matrix Σ contains the singular values $\sigma_1 \geq \dots \geq \sigma_r \geq 0$ of $\hat{S}$, and $r \times r$ orthogonal matrix V^T contains the right singular vectors.
 Here $r = p - k + 1$.
- 5: **if** $(\frac{\sigma_r}{\sigma_1})^2 > \eta$ **then**
- 6: Return $\theta_{id} = \theta$, i.e. all parameters currently in θ are identifiable.
- 7: **else**
- 8: *i)* For the singular (column) vector $\tilde{v}_r \in V^T$ associated with σ_r ,
 determine the position l of component with the largest magnitude,
 which corresponds to the least identifiable parameter.
- 9: *ii)* Remove the l^{th} column in $\hat{S}$ and the corresponding parameter (θ_l)
 from the identifiable subset.
- 10: **end if**
- 11: **end for**

sampled about nominal values. In the latter case, the non-identifiable parameters are fixed at their nominal values.

Plotting estimates of kernel density for these response distributions provides a qualitative comparison using this approach (Smith 2014). Randomly sampled parameters are uniformly perturbed $\pm 10\%$ about their nominal values for each parameter. We compute the responses by solving the model at times t between 1–1,000 h, every hour.

We extend this qualitative approach to the full time interval by also computing a relative error norm for a concentration of interest u_k ($k = 1, \dots, m$). This error norm is based on a difference between model simulations of u_k at randomly chosen values of all parameters ($\theta^*(1 + \epsilon X)$), and at randomly chosen values of identifiable parameters ($\theta_{id}^*(\eta)(1 + \epsilon X)$), where X is a random variable such that $X \sim U(-1, 1)$. In the second term of the numerator, non-identifiable parameters are fixed at their nominal values ($\theta_{nid}^*(\eta)$). We define this relative error norm by,

$$R(t; \eta) = \frac{\|u_k(t, \theta^*(1 + \epsilon X)) - u_k(t, \theta_{nid}^*(\eta) \cup \theta_{id}^*(\eta)(1 + \epsilon X))\|_2}{\|u_k(t, \theta^*(1 + \epsilon X))\|_2}. \quad (13)$$

3.3.2 Quantitative Method Based on Energy Statistics

A quantitative technique, based on energy statistics, determines if two sets of samples are drawn from the same distribution. At a specified confidence level, we use this technique to accept or reject the null hypothesis H_0 that two samples come from the same distribution. Let X (size n_1) and Z (size n_2) be random independent samples drawn from distributions G_X and G_Y , respectively. Here, we associate X with randomly sampling all parameters $\theta(\eta)$, whereas we associate Z with randomly sampling the identifiable parameters $\theta_{id}(\eta)$ with the non-identifiable parameters are fixed at their nominal values.

The null hypothesis is as follows:

$H_0 : G_X = G_Y$, which identifies if the two sets of samples are drawn from the same distribution

Using the strategy detailed in Székely and Rizzo (2013, 2004), we evaluate the energy distance

$$\epsilon_{n_1, n_2}(X, Z; \eta) = \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} |X_i - Z_m| - \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} |X_i - X_j|$$

and the test statistic

$$T_{n_1, n_2} = \frac{n_1 n_2}{n_1 + n_2} \epsilon_{n_1, n_2}.$$

We reject the null hypothesis if $T_{n_1, n_2} > T_c$, where T_c is defined as

$$T_c = p = \frac{1}{M} \sum_{k=1}^M I(T_{n_1, n_2}^k \leq T_c). \quad (14)$$

The critical or p -value at the α significance level is the value of T_c for which $P(T_{n_1, n_2}^k \leq T_c) = 1 - \alpha$. In (14), $I(\cdot)$ is an indicator function, and this method is used to approximate critical values via bootstrapping (Efron and Tibshirani 1993).

4 Results

We solve the initial value problem for the mPBPK model numerically using the Matlab (ver. 2022a) routine “ode15s”, which utilizes a variable-step, variable-order ODE solver. We employed this solver as the system of differential equations is stiff due, in part, to the variety of time scales inherent to the model. Both the absolute tolerance and the relative tolerance for this routine were set at the value 10^{-10} . We calibrated these two values to ensure that the final choice yielded accurate results, as determined by the verification techniques outlined in Sect. 3.3 (see also the Discussion).

In Sect. 4.1, we illustrate our methodological approach for the concentration in plasma (C_P) based on the nominal parameter values θ^* for humans discussed in Table 5 in the Appendix. This approach is then applied for concentrations in the brain CSF and interstitial fluid (ISF) in Sects. 4.2 and 4.3, respectively. We plot the response curves for these dependent variables over a time interval from zero to 1000 h in Fig. 3.

4.1 Concentration in Plasma

Prior to constructing a sensitivity matrix using nominal values θ^* , we compare the numerical evaluation of scaled sensitivities (7) for the variable C_P using both the sensitivity equations and derivative approximations. Specifically, we first verify qualitative

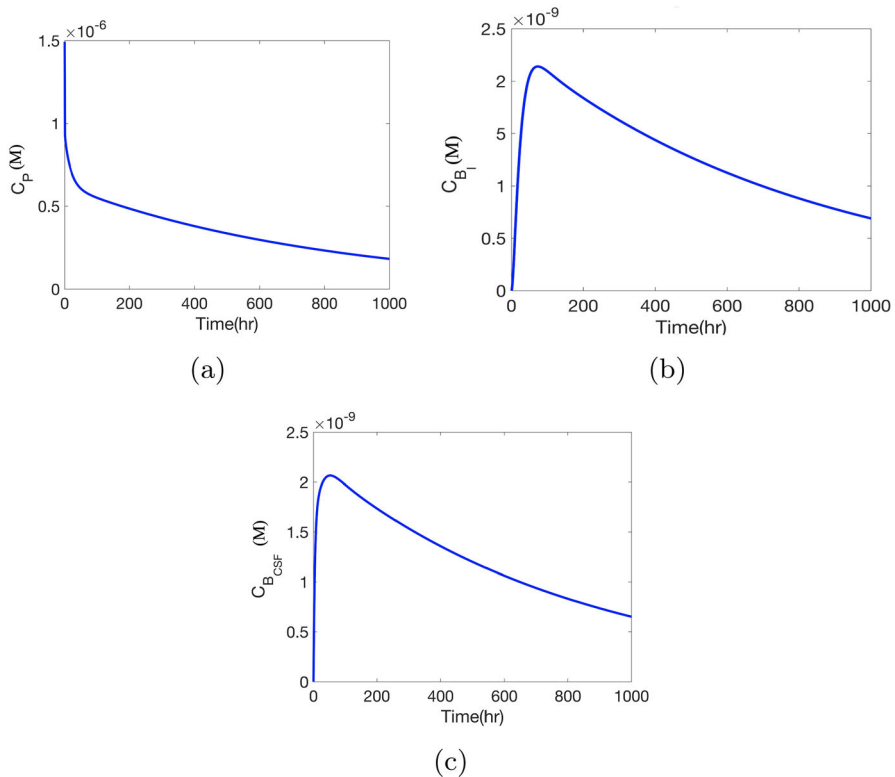


Fig. 3 Response plots for concentration in **a** plasma (C_P), **b** brain ISF (C_{B_I}), and **c** brain CSF ($C_{B_{CSF}}$) at the set of nominal parameter values shown in Table 5

agreement between a complex-step derivative approximation (6) and solutions of the sensitivity Eq. (4) for the scaled sensitivities with larger magnitudes in Fig. 4a and b. We plot all 31 sensitivities of concentration in the plasma, grouped by magnitude, in Fig. 4. We then create a sensitivity matrix $\mathbf{S}$ (11) using these sensitivities which we use as input for the algorithms presented in Sect. 3.2.

4.1.1 Comparison of PSS Approaches

We implemented Algorithm 1 and Algorithm 2 for PSS, detailed in Sect. 3.2, in Matlab. For a fixed value of η , typical run times to generate a plot of the relative error norm (see Fig. 6) for both algorithms were under 30 s on a MacBook Pro laptop (1.7 GHz Quad-Core Intel i7). We compare results for the sets of the identifiable parameters determined by these two algorithms in Table 1. Specifically, we compare four values of the threshold in the range $\eta = 10^{-12}$ to $\eta = 10^{-6}$. In the case $\eta = 10^{-6}$, we observe that the two subsets of identifiable parameters differ by two elements. We also observe that the number of identifiable parameters increases as we decrease the threshold. However, elements of the identifiable parameter subsets deviate further

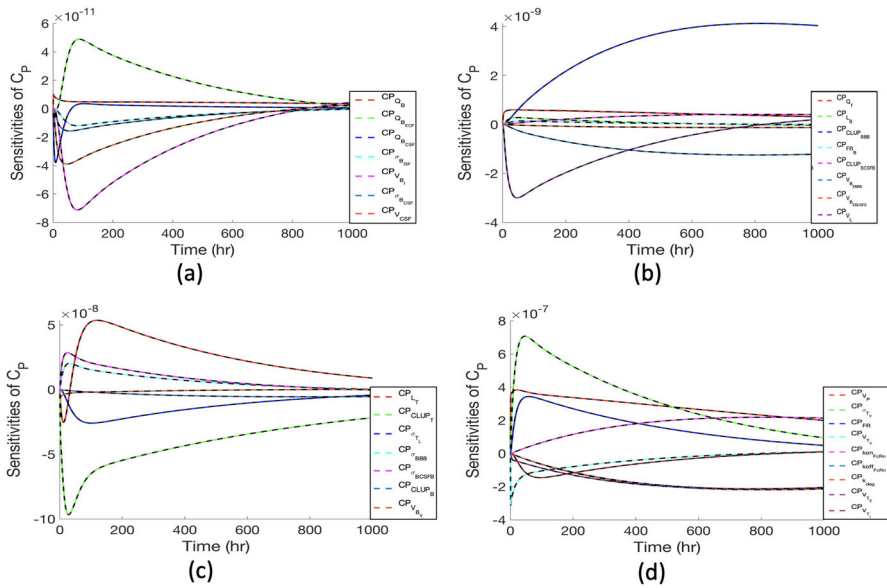


Fig. 4 Scaled local sensitivities for the concentration of plasma (C_P) with respect to the 31 parameters, grouped by magnitude. Sensitivity equations (solid) vs complex-step derivative approximation (dashed) in (a–d)

as the threshold value is reduced to 10^{-12} . For this case ($\eta = 10^{-12}$), the one-at-a-time approach (Algorithm 2) includes L_T , FR , k_{deg} in the identifiable parameter subset, whereas the all-at-once approach (Algorithm 1) includes $CLUP_T$, $koff_{FcRn}$, $CLUP_{BBB}$ in the identifiable parameter subset. Note that both identifiable subsets contain the following parameters: V_P , σ_{TV} , V_{TV} , V_{TE} , V_{TI} , σ_{BBB} , and σ_{BCSFB} .

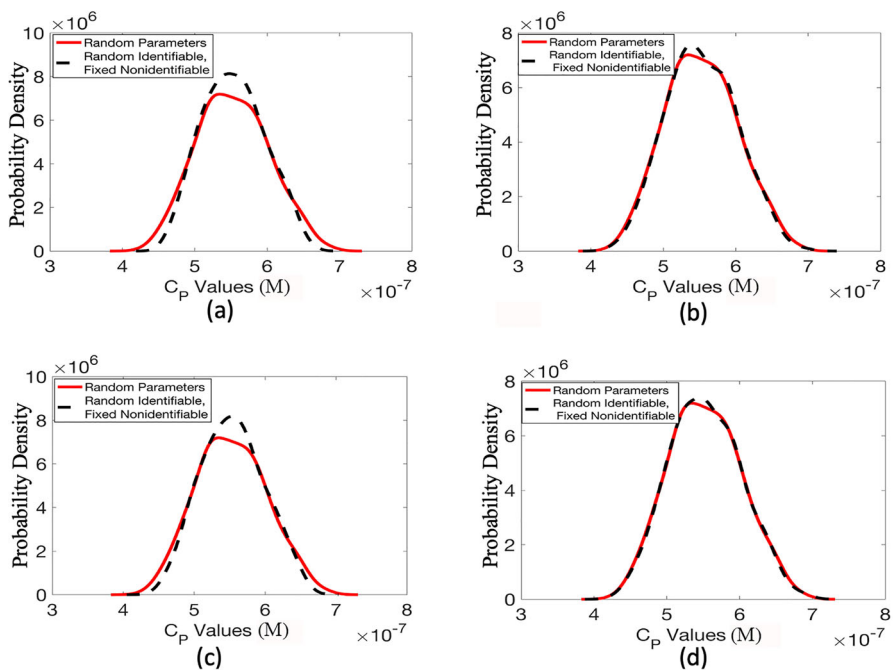
We then evaluate our results using the qualitative method discussed in Sect. 3.3.1. We illustrate an underlying distribution plot at $t = 100$ hr in Fig. 5. These plots demonstrate that, for the values $\eta = 10^{-8}$ and $\eta = 10^{-12}$ at $t = 100$, Algorithm 2 outperforms Algorithm 1 based on the fact that the distributions match more closely for Algorithm 2. Table 1 also verifies this outcome since for all values of η , we achieve the lowest average relative error norm with Algorithm 2. Note that we compute average relative error norms of concentration in plasma using Eq. (13).

Figure 6 shows the relative error norm values of C_P for the complete time interval $[0, 1000]$. This more comprehensive comparison demonstrates that Algorithm 2 outperforms Algorithm 1. Although both algorithms yield nested subsets of identifiable parameters as η is decreased, Algorithm 2 has magnitudes of R that are significantly lower over the full time interval. Furthermore, Algorithm 2 performs significantly better for the cases $\eta = \{10^{-8}, 10^{-10}, 10^{-12}\}$ versus the case $\eta = 10^{-6}$. However, error values (R) increase substantially for times beyond $t \approx 300$ h.

We also verify our results with the quantitative energy statistic, summarized in Sect. 3.3.2. In this case, G_X is the distribution constructed by randomly sampling all parameters, and G_Y is the distribution generated by randomly sampling identifiable parameters with fixed unidentifiable parameters. We reject the null hypothesis that

Table 1 Parameter subset selection result for C_P , Id-number of identifiable parameters, R-relative error norm

η values	Id	Identifiable parameters (θ_{id})	$\bar{R}_1$	$\bar{R}_2$
10^{-6}	5	Alg 1: $V_P, \sigma_{TV}, koff_{FcRn}, V_{TE}, CLUP_{BBB}$ Alg 2: $V_P, \sigma_{TV}, k_{deg}, V_{TE}, V_{TI}$	0.17755	0.17642
10^{-8}	7	Alg 1: $V_P, \sigma_{TV}, CLUP_T, koff_{FcRn}, V_{TE}, \sigma_{BCSFB}, CLUP_{BBB}$ Alg 2: $V_P, \sigma_{TV}, FR, k_{deg}, V_{TE}, V_{TI}, \sigma_{BCSFB}$	0.18418	0.098272
10^{-10}	9	Alg 1: $V_P, \sigma_{TV}, CLUP_T, koff_{FcRn}, V_{TE}, V_{TI}, \sigma_{BBB}, \sigma_{BCSFB}, CLUP_{BBB}$ Alg 2: $LT, V_P, \sigma_{TV}, FR, k_{deg}, V_{TE}, V_{TI}, \sigma_{BBB}, \sigma_{BCSFB}$	0.18013	0.10277
10^{-12}	10	Alg 1: $V_P, \sigma_{TV}, CLUP_T, V_{TV}, koff_{FcRn}, V_{TE}, V_{TI}, \sigma_{BBB}, \sigma_{BCSFB}, CLUP_{BBB}$ Alg 2: $LT, V_P, \sigma_{TV}, FR, V_{TV}, k_{deg}, V_{TE}, V_{TI}, \sigma_{BBB}, \sigma_{BCSFB}$	0.11989	0.023192

**Fig. 5** Response distribution plots for C_P at $t=100$; **a** Algorithm 1 ($R_1 = 0.2238$) versus **b** Algorithm 2 ($R_2 = 0.09364$) for $\eta = 10^{-8}$, **c**) Algorithm 1 ($R_1 = 0.16041$) versus **d**) Algorithm 2 ($R_2 = 0.05795$) for $\eta = 10^{-12}$

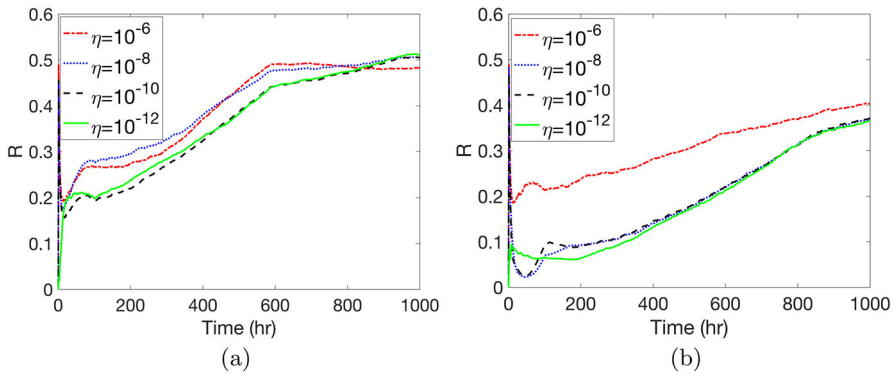


Fig. 6 Concentration in plasma: **a** Algorithm 1 and **b** Algorithm 2

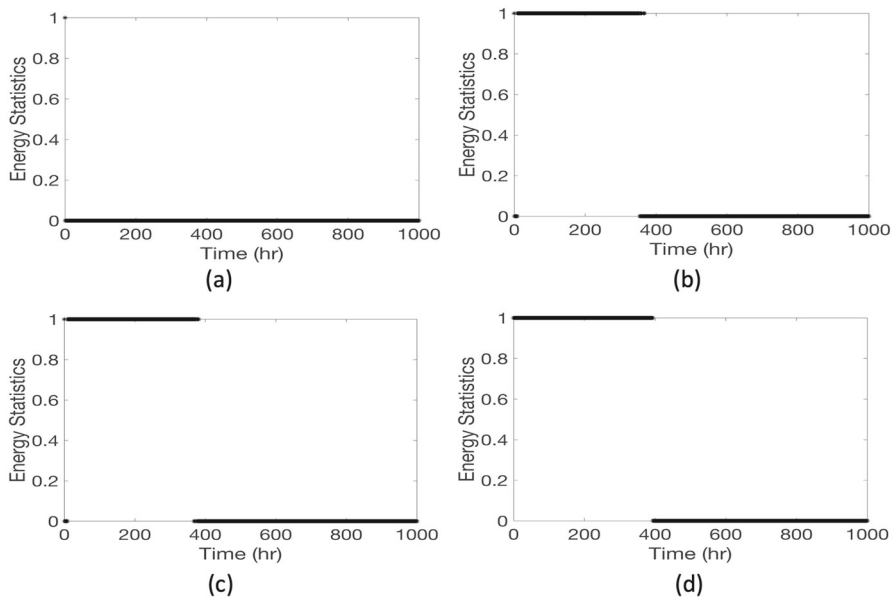


Fig. 7 Quantitative verification with energy statistics using Algorithm 2: **a** $\eta = 10^{-6}$, **b** $\eta = 10^{-8}$, **c** $\eta = 10^{-10}$, **d** $\eta = 10^{-12}$

these two distributions are the same if test statistic is greater than the critical value after calculating energy distance, test statistics, and critical values. Throughout the time interval, we continue this process, marking each rejection with a 0 and each acceptance with a 1. In Fig. 7, we observe that we reject the null hypothesis for $\eta = 10^{-6}$ for almost the entire time interval, whereas we accept the null hypothesis for $\{10^{-8}, 10^{-10}, 10^{-12}\}$ for the first 400 hr and reject it for the remainder of the time interval. Hence, the quantitative and qualitative techniques are in agreement.

4.2 Concentration in Brain ISF

In this section, we consider the concentration in the brain ISF (C_{B_I}) and repeat the analysis procedure in Sect. 4.1. The outcomes of the identifiability analysis for this concentration are compiled in Table 2 for four values of the threshold in the range $\eta = 10^{-12}$ to $\eta = 10^{-6}$. Based on these results, the concentration in ISF yields fewer identifiable parameters than the concentration in plasma for different threshold levels. Furthermore, as the values of η decrease, elements in the identifiable parameter subset increase by one or two parameters.

Figure 8 shows the relative error norm values of C_{B_I} for the complete time interval $[0, 1000]$, for both algorithms. Overall, we observe that both algorithms have significantly lower relative error values for C_{B_I} as compared to C_P over the time domain of interest. We also observe that the error values in Algorithm 2 have magnitudes of R that are lower than those for Algorithm 1 until $t \approx 300$ hr for the cases $\eta = \{10^{-8}, 10^{-10}, 10^{-12}\}$. Furthermore the error curves become less sensitive to the threshold values as the magnitude of η decreases.

We also verify our results with the quantitative energy statistic (Fig. 8c). We observe that all four threshold values yield an identical plot, and we can conclude that we always accept our hypothesis for the full time interval. Hence, the quantitative and qualitative techniques are in agreement.

4.3 Concentration in Brain CSF

Lastly, we consider the concentration in the brain CSF ($C_{B_{CSF}}$) and repeat the analysis procedure in Sects. 4.1 and 4.2. The outcomes of the identifiability analysis for this concentration are compiled in Table 3 for four values of the threshold in the range $\eta = 10^{-12}$ to $\eta = 10^{-6}$. We note that the results in Table 3 are very similar to those in Table 2. In particular, the identifiable parameter subset for the case $\eta = 10^{-12}$ is identical to that corresponding entry in Table 2 for C_{ISF} . This suggests that, in terms of identifiability, the concentrations in ISF and CSF yield similar information regarding the identifiable parameter subset.

Figure 9a and b demonstrate the relative error norm values for Algorithm 1 and 2, respectively. Similar to concentration in brain ISF, the relative error norm for concentration in brain CSF is also significantly smaller than for plasma across the full time domain. In this case, these error values are also slightly smaller across time for Algorithm 2 as compared to Algorithm 1. We present a quantitative validation in Fig. 9c. We observe that the null hypothesis is accepted over the full time domain of interest.

5 Discussion and Conclusions

In this investigation, we presented a systematic approach for identifiability analysis of PBPK models based on local sensitivities in the absence of experimental data, but when a time domain of interest and a set of responses of interest are prescribed. Techniques were also developed for evaluating the accuracy of identifiable parameter

Table 2 Parameter subset selection result for C_{B_I} , Id-number of identifiable parameters, R-relative error norm

η values	Id	Identifiable parameters (θ_{Id})	$\bar{R}_1$	$\bar{R}_2$
10^{-6}	3	Alg 1: $\sigma_{BBB}, \sigma_{BCSFB}, V_{BeBCSFB}$ Alg 2: $V_{TE}, \sigma_{BBB}, \sigma_{BCSFB}$	0.08051	0.07028
10^{-8}	4	Alg 1: $FR, \sigma_{BBB}, \sigma_{BCSFB}, FR_B, V_{BeBCSFB}$ Alg 2: $V_{TE}, \sigma_{BBB}, \sigma_{BCSFB}, \sigma_{TV}$	0.06042	0.06108
10^{-10}	6	Alg 1: $FR, \sigma_{BBB}, \sigma_{BCSFB}, CLUP_{BBB}, FR_B, V_{BeBCSFB}$ Alg 2: $V_{TE}, \sigma_{BBB}, V_{T_I}, \sigma_{BCSFB}, \sigma_{TV}, FR_B$	0.05833	0.06272
10^{-12}	7	Alg 1: $FR, \sigma_{BBB}, \sigma_{BCSFB}, CLUP_{BBB}, FR_B, V_{BeBCSFB}, VCSF$ Alg 2: $V_{TE}, \sigma_{BBB}, V_{T_I}, \sigma_{BCSFB}, \sigma_{TV}, FR_B, VCSF$	0.05856	0.06280

Table 3 Parameter subset selection result for C_{BCSF} , Id-number of identifiable parameters, R-relative error norm

η values	Id	Identifiable parameters (θ_{Id})	$\bar{R}_1$	$\bar{R}_2$
10^{-6}	2	Alg 1: $FR, \sigma_{BBB}, \sigma_{BCSFB}$ Alg 2: $\sigma_{BBB}, \sigma_{BCSFB}$	0.04128	0.05738
10^{-8}	4	Alg 1: $\sigma_{TV}, FR, \sigma_{BBB}, \sigma_{BCSFB}$ Alg 2: $V_{TE}, \sigma_{BBB}, \sigma_{BCSFB}, \sigma_{TV}$	0.05815	0.03027
10^{-10}	6	Alg 1: $\sigma_{TV}, FR, V_{T1}, \sigma_{BBB}, \sigma_{BCSFB}, VCSF$ Alg 2: $V_{TE}, \sigma_{BBB}, V_{T1}, \sigma_{BCSFB}, \sigma_{TV}, VCSF$	0.05929	0.02873
10^{-12}	7	Alg 1: $\sigma_{TV}, FR, V_{T1}, \sigma_{BBB}, \sigma_{BCSFB}, CLUP_{BBB}, VCSF$ Alg 2: $V_{TE}, \sigma_{BBB}, V_{T1}, \sigma_{BCSFB}, \sigma_{TV}, FR_B, VCSF$	0.05909	0.028552

Table 4 Average sensitivities with respect to identifiable parameters for each concentration of interest

C_P	C_{BI}				C_{BCSF}				
	Param	θ^*	Sens	Param	θ^*	Sens	Param	θ^*	Sens
C_P	σ_{TV}	0.92	3.277e-07	σ_{BBB}	1	- 2.462e-07	σ_{BCSFB}	0.99	- 3.619e-07
	V_P	3.12	2.847e-07	σ_{BCSFB}	0.99	- 1.978e-07	σ_{BBB}	1	- 8.677e-08
	V_{TE}	0.33	- 1.728e-07	FR_B	0.71	- 1.977e-09	σ_{TV}	0.92	- 1.141e-09
	k_{deg}	26.6	- 1.687e-07	σ_{TV}	0.92	1.181e-09	FR_B	0.71	- 7.959e-10
	FR	0.71	1.664e-07	V_{TE}	0.33	- 6.056e-10	V_{TE}	0.33	- 5.926e-10
	V_{TI}	11.09	- 5.178e-08	V_{TI}	11.09	- 1.893e-10	V_{TI}	11.09	- 1.816e-10
	V_{TV}	1.68	- 4.098e-08	$VCSF$	0.14	- 2.237e-12	$VCSF$	0.14	- 3.910e-12
	LT	0.32	2.623e-08						
	σ_{BCSFB}	0.99	8.602e-09						
	σ_{BBB}	1	6.654e-09						

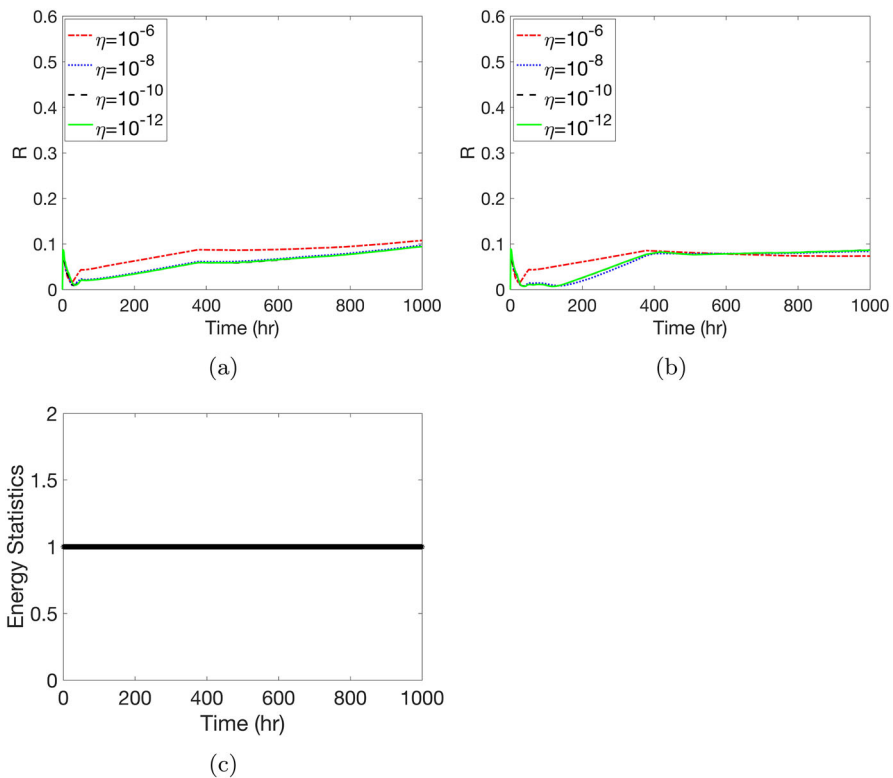


Fig. 8 Qualitative verification for concentration in brain ISF with **a** Algorithm 1 **b** Algorithm 2 **c** quantitative verification with Algorithm 1

subsets with respect to random perturbations about a set of nominal parameter values. These techniques were developed and evaluated in the context of a minimal brain PBPK model for investigating antibody therapeutics in the CNS (Bloomingdale et al. 2021). We considered three key antibody concentrations: the antibody concentrations in plasma (C_P), brain BSF (C_{BSF}) and brain ISF (C_{ISF}), over a time domain of interest spanning up to 1000 h. Our approach includes a highly accurate complex-step approximation of local sensitivities. The verification component in our approach is used to evaluate accuracy of the set of identifiable parameters using both a relative error norm across the full time domain and energy statistics.

Our findings reveal that accuracy of our parameter identifiability analysis for C_P is much higher when using a one-at-a-time algorithm (Algorithm 2) versus an all-at-once approach (Algorithm 1). However, for both approaches, parameter identifiability becomes less accurate for C_P for times beyond $t \approx 300$ h. By contrast, identifiability analysis for C_{BSF} and C_{BI} demonstrated greater accuracy for all times and was also less dependent on the choice of algorithm.

To investigate this difference further, we calculate average sensitivities across the time interval for the set of all identifiable parameters (Algorithm 2, $\eta = 10^{-12}$) in Table 4. In particular, for each of the three concentrations of interest, we list the iden-

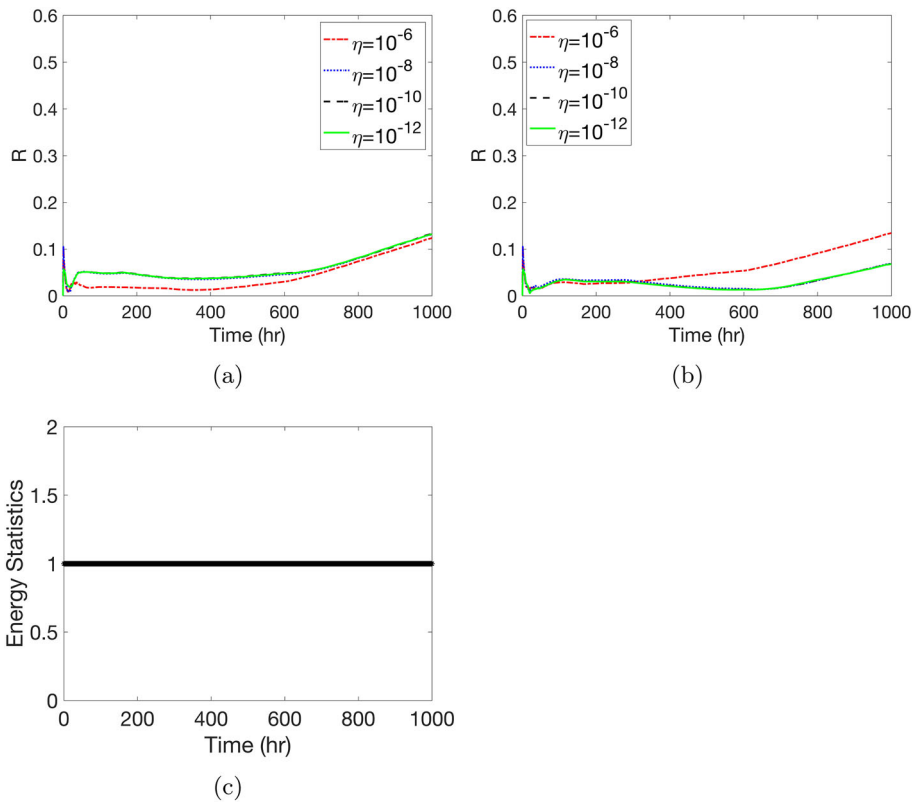


Fig. 9 Qualitative verification for concentration in brain CSF with **a** Algorithm 1 **b** Algorithm 2 **c** quantitative verification with Algorithm 1

tifiable parameters in decreasing order of the magnitude of their average sensitivities. With this effective ranking and a screening threshold of 10^{-7} , we observe that C_P retains 5 parameters (σ_{TV} , V_P , V_{TE} , k_{deg} , FR). In contrast, C_{B_I} retains two parameters (σ_{BBB} , σ_{BCSFB}) and C_{BCSF} retains one parameter (σ_{BCSFB}). Furthermore, the parameters retained for both C_{B_I} and C_{BCSF} have an upper bound of 1 and nominal values of 0.99 and 1. Taken together, these observations suggest that achieving low relative error norms for C_P is more challenging than for C_{B_I} and C_{BCSF} . We believe that this is due to the wider exploration of parameter space in the random perturbations of the most sensitive parameters about the nominal values (Sect. 4.1); the most identifiable parameters for C_P are greater in number and less constrained.

Since PBPK models have a diverse range of inherent and interacting time scales, determining an appropriate balance between analyzing quantities of interest on physically meaningful time domains, and robustly determining parameter identifiability is an important consideration in applying this methodology. Our overall approach not only permits the user to specify the time domain of interest for the PSS, but also enables verification of identifiable parameter subset accuracy at an arbitrary time in the domain. This is in contrast to structural identifiability analysis, which may not

capture variations in parameter identifiability with respect to time and will not flag parameters that are locally unidentifiable about a particular set of nominal values, but not among the subset of parameters deemed to be (globally) unidentifiable. We note that, for simpler dynamic models, the ability to identify a small number of time scales can be used to devise a more tailored PSS approach that generates multiple sets of identifiable parameter candidates (Pearce et al. 2021).

For all values of η considered, C_P has a greater number of identifiable parameters than C_{B_I} and $C_{B_{CSF}}$; in the case $\eta = 10^{-12}$, C_P has 10 identifiable parameters while C_{B_I} and $C_{B_{CSF}}$ both have 7 identifiable parameters. This outcome is not unexpected since the mPBPK model has more internal connections among compartments in the brain compartment of the model as compared to the plasma component of the model (Fig. 2).

Overall, our results demonstrate that our PSS algorithm based on local identifiability analysis, combined with verification techniques, provides a systematic and robust assessment of identifiable parameter subsets. Furthermore, the accuracy of these subsets depends on both the time at which predictions for key responses are evaluated as well as the nominal parameter values. By contrast, structural identifiability analysis (Slob et al. 1997; Yates 2006) determines identifiable subsets that are likely smaller, since they would necessarily need to be valid across the entire time domain of interest. Future studies could combine the two approaches, but this may be challenging since our attempt to perform structural identifiability analysis on our model using the SIAN package (Hong et al. 2019) failed to yield results after running for over 20 h (MAPLE 2021, MacMini, Apple M1 Chip). Nominal parameter values are often a starting point for parameter estimation in developing extended models; accurately identifying subsets of non-identifiable parameters can lead to more robust optimization during parameter estimation. Robust and accurate identifiability analysis for PBPK models is also a crucial step prior to carrying out Bayesian methods for parameter estimation or uncertainty quantification. The ability to accurately and reliably fix a subset of parameters at their nominal values can improve and accelerate performance of associated algorithms like Markov chain Monte Carlo (MCMC).

Funding This study was funded in part by grant DMS-1638521 from the National Science Foundation.

Availability of data and materials Data sharing is not applicable to this article as no datasets were analyzed or generated during the current study.

Appendix A

The brain minimal PBPK (mPBPK) model (Bloomingdale et al. (2021)) comprises the following system of ordinary differential equations for the 16 antibody concentrations:

1. Plasma (C_P)

$$V_P \cdot \frac{dC_P}{dt} = (Q_T - L_T) \cdot C_{T_V} + (Q_B - L_B) \cdot C_{B_V} + (L_T + L_B) \cdot C_L - Q_T \cdot C_P - Q_B \cdot C_P \quad (15)$$

Table 5 Parameter values for human in the model of Bloomingdale et al. 2021

Parameter	Human	Unit
FR	0.715	–
FR_B	0.715	–
k_{deg}	26.6	1/h
kon_{FcRn}	5.59E+08	1/M/h
$koff_{FcRn}$	23.85	1/h
V_P	3.1259	L
V_{TV}	1.6814	L
V_{TE}	0.3352	L
V_{TI}	11.0988	L
V_{BV}	0.0319	L
V_{BBBB}	0.0066	L
$V_{BEC SFB}$	6.5904e–04	L
V_{BI}	0.2610	L
V_{BCSF}	0.1425	L
V_L	0.2743	L
Q_T	160.4592	L/h
Q_B	21.4533	L/h
L_T	0.3209	L/h
L_B	0.0345	L/h
Q_{BECF}	0.0105	L/h
Q_{BCSF}	0.024	L/h
σ_{TV}	0.9233	–
σ_{TL}	0.2	–
σ_{BBB}	1	–
σ_{BCSFB}	0.9974	–
σ_{BISF}	0.2	–
σ_{BCSF}	0.2	–
$CLUP_T$	0.1844	L/h
$CLUP_{BCSFB}$	2.0081e–05	L/h
$CLUP_{BBB}$	2.0081e–04	L/h
$CLUP_B$	2.2089e–04	L/h
$FcRn$	4.9820e–05	M
k_{CLUPT}	0.55	1/h
k_{CLUPB}	0.03047	1/h
SA_{BBB}	17	m ²
SA_{BCSFB}	1.7	m ²
f_{BBB}	0.9091	–

2. Tissue vascular (C_{TV})

$$V_{TV} \cdot \frac{dC_{TV}}{dt} = Q_T \cdot C_P - (Q_T - L_T) \cdot C_{TV} - ((1 - \sigma_{TV}) \cdot L_T \cdot C_{TV}) \\ - CLUP_T \cdot C_{TV} + CLUP_T \cdot FR \cdot C_{TE,B} \quad (16)$$

3. Tissue endosome (unbound) ($C_{TE,U}$)

$$V_{TE} \cdot \frac{dC_{TE,U}}{dt} = CLUP_T \cdot (C_{TV} + C_{TI}) - V_{TE} \cdot (kon_{FcRn} \cdot C_{TE,U} \cdot C_{TE,B} \\ - koff_{FcRn} \cdot C_{TE,B} + k_{deg} \cdot C_{TE,U}) \quad (17)$$

4. Tissue endosome (FcRnbound) ($C_{TE,B}$)

$$V_{TE} \cdot \frac{dC_{TE,B}}{dt} = V_{TE} \cdot (kon_{FcRn} \cdot C_{TE,U} \cdot C_{TE,B} - koff_{FcRn} \cdot C_{TE,B}) \\ - CLUP_T \cdot C_{TE,B} \quad (18)$$

5. Tissue interstitium (C_{TI})

$$V_{TI} \cdot \frac{dC_{TI}}{dt} = (1 - \sigma_{TV}) \cdot L_T \cdot C_{TV} - (1 - \sigma_{TI}) \cdot L_T \cdot C_{TI} \\ + CLUP_T \cdot (1 - FR) \cdot C_{TE,B} - CLUP_T \cdot C_{TI} \quad (19)$$

6. Brain vascular (C_{BV})

$$V_{BV} \cdot \frac{dC_{BV}}{dt} = Q_B \cdot C_P - (Q_B - L_B) \cdot C_{BV} - (1 - \sigma_{BBB}) \cdot Q_{BECF} \cdot C_{BV} \\ - (1 - \sigma_{BCSFB}) \cdot Q_{BCSF} \cdot C_{BV} - CLUP_B \cdot C_{BV} \\ + CLUP_{BBB} \cdot FR_B \cdot C_{BEBB,B} \\ + CLUP_{BCSFB} \cdot FR_B \cdot C_{BEBCSFB,B} \quad (20)$$

7. BBB endosome (unbound) ($C_{BEBB,U}$)

$$V_{BEBB} \cdot \frac{dC_{BEBB,U}}{dt} = CLUP_{BBB} \cdot (C_{BV} + C_{BI}) + V_{BEBB} \cdot (-kon_{FcRn} \cdot \\ C_{BEBB,U} \cdot C_{BEBB,B} + koff_{FcRn} \cdot C_{BEBB,B} \\ - k_{deg} \cdot C_{BEBB,U}) \quad (21)$$

8. BBB endosome (FcRnbound) ($C_{BEBB,B}$)

$$V_{BEBB} \cdot \frac{dC_{BEBB,B}}{dt} = V_{BEBB} \cdot (kon_{FcRn} \cdot C_{BEBB,U} \cdot C_{BEBB,B} \\ - koff_{FcRn} \cdot C_{BEBB,B}) \\ - CLUP_{BBB} \cdot C_{BEBB,B} \quad (22)$$

9. Brain interstitium (C_{B_I})

$$\begin{aligned}
 V_{B_I} \cdot \frac{C_{B_I}}{dt} = & (1 - \sigma_{BBB}) \cdot Q_{B_{ECF}} \cdot C_{B_V} - (1 - \sigma_{B_{ISF}}) \cdot Q_{B_{ECF}} \cdot C_{B_I} \\
 & + CLUP_{BBB} \cdot (1 - FR_B) \cdot C_{B_{EBBB,B}} - CLUP_{BBB} \cdot C_{B_I} \\
 & - Q_{B_{ECF}} \cdot C_{B_I} + Q_{B_{ECF}} \cdot C_{B_{CSF}}
 \end{aligned} \quad (23)$$

10. BCSFB endosome (unbound) ($C_{B_{EBCSFB,U}}$)

$$\begin{aligned}
 V_{B_{EBCSFB}} \cdot \frac{C_{B_{EBCSFB,U}}}{dt} = & CLUP_{BCSFB} \cdot C_{B_V} + CLUP_{BCSFB} \cdot C_{B_{CSF}} \\
 & + V_{B_{EBCSFB}} \cdot (-kon_{FcRn} \cdot C_{B_{EBCSFB,U}} \\
 & \cdot C_{B_{BCSFBFcRn,U}} + koff_{FcRn} \cdot C_{B_{EBCSFB,B}} \\
 & - k_{deg} \cdot C_{B_{EBCSFB,U}})
 \end{aligned} \quad (24)$$

11. BCSFB endosome (FcRnbound) ($C_{B_{EBCSFB,B}}$)

$$\begin{aligned}
 V_{B_{EBCSFB}} \cdot \frac{C_{B_{EBCSFB,B}}}{dt} = & V_{B_{EBCSFB}} \cdot (kon_{FcRn} \cdot C_{B_{EBCSFB,U}} \\
 & \cdot C_{B_{BCSFBFcRn,U}} - koff_{FcRn} \cdot C_{B_{EBCSFB,B}}) \\
 & - CLUP_{BCSFB} \cdot C_{B_{EBCSFB,B}}
 \end{aligned} \quad (25)$$

12. Brain CSF ($C_{B_{CSF}}$)

$$\begin{aligned}
 V_{B_{CSF}} \cdot \frac{C_{B_{CSF}}}{dt} = & (1 - \sigma_{BCSFB}) \cdot Q_{B_{CSF}} \cdot C_{B_V} - CLUP_{BCSFB} \cdot C_{B_{CSF}} \\
 & + CLUP_{BCSFB} \cdot (1 - FR_B) \cdot C_{B_{EBCSFB,B}} + Q_{B_{ECF}} \cdot C_{B_I} \\
 & - (1 - \sigma_{B_{CSF}}) \cdot Q_{B_{CSF}} \cdot C_{B_{CSF}} - Q_{B_{ECF}} \cdot C_{B_{CSF}}
 \end{aligned} \quad (26)$$

13. Lymph (C_L)

$$\begin{aligned}
 V_L \cdot \frac{C_L}{dt} = & (1 - \sigma_{T_L}) \cdot L_T \cdot C_{T_I} + (1 - \sigma_{B_{CSF}}) \cdot Q_{B_{CSF}} \cdot C_{B_{CSF}} \\
 & + (1 - \sigma_{B_{ISF}}) \cdot Q_{B_{ECF}} \cdot C_{B_I} - (L_T + L_B) \cdot C_L
 \end{aligned} \quad (27)$$

14. FcRn concentration in tissue endosome ($C_{T_{FcRn,U}}$)

$$\begin{aligned}
 V_{T_E} \cdot \frac{C_{T_{FcRn,U}}}{dt} = & -V_{T_E} \cdot (kon_{FcRn} \cdot C_{T_{E,U}} \cdot C_{T_{FcRn,U}} - koff_{FcRn} \cdot C_{T_{E,B}}) \\
 & + CLUP_T \cdot C_{T_{E,B}}
 \end{aligned} \quad (28)$$

15. FcRn BBB endosome ($C_{B_{BBBFcRn,U}}$)

$$V_{B_{EBBB}} \cdot \frac{C_{B_{BBBFcRn,U}}}{dt} = V_{B_{EBBB}} \cdot (-kon_{FcRn} \cdot C_{B_{EBBB,U}} \cdot C_{B_{BBBFcRn,U}} + koff_{FcRn} \cdot C_{B_{EBBB,B}}) + CLUP_{BBB} \cdot C_{B_{EBBB,B}} \quad (29)$$

16. FcRn BCSFB endosome ($C_{B_{BCSFBFcRn,U}}$)

$$V_{B_{EBCSFB}} \cdot \frac{C_{B_{BCSFBFcRn,U}}}{dt} = V_{B_{EBCSFB}} \cdot (-kon_{FcRn} \cdot C_{B_{EBCSFB,U}} \cdot C_{B_{BCSFBFcRn,U}} + koff_{FcRn} \cdot C_{B_{EBCSFB,B}}) + CLUP_{BCSFB} \cdot C_{B_{EBCSFB,B}} \quad (30)$$

This system of equations is augmented with initial conditions. The first concentration is prescribed an initial value $\frac{Dose_0}{V_P}$, where $Dose_0 = \frac{Bodyweight \times Dose_{amount}}{Molecular_{weight}} \times 1000$ with $Bodyweight = 70kg$, $Molecular_{weight} = 150kDa$, $Dose_{amount} = 10mg/kg$. The subsequent 12 concentrations are prescribed initial values of zero, and the remaining three concentrations ($C_{T_{FcRn,U}}$, $C_{B_{BBBFcRn,U}}$, $C_{B_{BCSFBFcRn,U}}$) are all prescribed a nonzero initial value of 0.4982×10^{-4} . In this study, a 10mg/kg dosage was given over 1000h to a 70kg individual using a 1-hour step size.

References

- Apgar JF, Witmer DK, White FM, Tidor B (2010) Sloppy models, parameter uncertainty, and the role of experimental design. *Mol BioSyst* 6:1890–1900
- Banks HT, Bekele-Maxwell K, Bociu L, Noorman M, Tillman K (2015) The complex-step method for sensitivity analysis of non-smooth problems arising in biology. *Eurasian J Math Comput Appl* 3:15–68
- Bloomingdale P, Bakshi S, Maass C, van Maanen E, Pichardo-Almarza C, Yadav DB, van der Graaf P, Mehrotra N (2021) Minimal brain PBPK model to support the preclinical and clinical development of antibody therapeutics for CNS diseases. *J Pharmacokinet Pharmacodyn* 48(6):861–871
- Bellu G, Saccomani MP, Audoly S, D'Angiò L (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput Methods Programs Biomed* 88(1):52–61
- Brown LV, Coles MC, McConnell M, Ratushny AV, Gaffney EA (2022) Analysis of cellular kinetic models suggest that physiologically based model parameters may be inherently, practically unidentifiable. *J Pharmacokinet Pharmacodyn* 49(5):539–556
- Burth M, Verghese GC, Vélez-Reyes M (1999) Subset selection for improved parameter estimates in on-line identification of a synchronous generator. *IEEE Trans Power Syst* 14(1):218–225
- Calvier EAM, Nguyen TT, Johnson TN, et al (2018) Can population modelling principles be used to identify key PBPK parameters for paediatric clearance predictions? An innovative application of optimal design theory. *Pharm Res* 209
- Carter SJ, Chouhan B, Sharma P, Chappell MJ (2020) Prediction of clinical transporter-mediated drug-drug interactions via comeasurement of pitavastatin and eltrombopag in human hepatocyte models. *CPT Pharmacometrics Syst Pharmacol* 9(4):211–221
- Chang HY, Wu S, Meno-Tetang G, Shah DK (2019) A translational platform PBPK model for antibody disposition in the brain. *J Pharmacokinet Pharmacodyn* 46(4):319–338
- Cintrón-Arias A, Banks HT, Capaldi A, Lloyd AL (2009) A sensitivity matrix based methodology for inverse problem formulation. *J Inverse Ill Posed Probl* 17(6):545–564

- Clewell HJ III, Lee TS, Carpenter RL (1994) Sensitivity of physiologically based pharmacokinetic models to variation in model parameters: methylene chloride. *Risk Anal* 1994(14):521–531
- Dwek R (2009) Antibodies and antigens: it's all about the numbers game. *Proc Natl Acad Sci USA* 106(7):2087–2088
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton
- Evans MV, Eklund CR, Williams DN, Sey YM, Simmons JE (2020) Global optimization of the Michaelis-Menten parameters using physiologically based pharmacokinetic (PBPK) modeling and chloroform vapor uptake data in F344 rats. *Inhalation Toxicol* 32(3):97–109
- Evans MV, Andersen ME (1995) Sensitivity analysis and the design of gas uptake inhalation studies. *Inhalation Toxicol* 7(7):1075–1094
- Friswell MI, Penny JET, Garvey SD (1997) Parameter subset selection in damage location. *Inverse Probl Eng* 5:189–215
- Gautschi W (1997) Numerical analysis. Birkhäuser, Basel
- Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3(10):1871–78
- Hong H, Ovchinnikov A, Pogudin G, Yap C (2019) SIAN: software for structural identifiability analysis of ODE models. *Bioinformatics* 35(16):2873–2874
- Hsieh NH, Reisfeld B, Bois FY, Chiu WA (2018) Applying a global sensitivity analysis workflow to improve the computational efficiencies in physiologically-based pharmacokinetic modeling. *Front Pharmacol* 9:588
- Ipsen ICF (2009) Numerical matrix analysis. SIAM, Philadelphia
- Jones H, Rowland-Yeo K (2013) Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacometrics Syst Pharmacol* 2(8):e63
- Keene ON (1995) The log transformation is special. *Stat Med* 14(8):811–819
- Kendrick F, Evans ND, Arnulf B, Avet-Loiseau H, Decaux O, Dejoie T, Fouquet G, Guidez S, Harel S, Hebraud B, Javauge V, Richez V, Schraen S, Touzeau C, Moreau P, Leleu X, Harding S, Chappell MJ (2017) Analysis of a compartmental model of endogenous immunoglobulin G metabolism with application to multiple myeloma. *Front Physiol* 8:147
- Kendrick F, Evans ND, Berlanga O, Harding SJ, Chappell MJ (2019) Parameter identification for a model of neonatal Fc receptor-mediated recycling of endogenous immunoglobulin G in humans. *Front Immunol* 10:674
- Kim B, Lee JH (2019) Parameter subset selection and biased estimation for a class of ill-conditioned estimation problems. *J Process Control* 81:65–75
- Koyama S, Toshimoto K, Lee W, Aoki Y, Sugiyama Y (2021) Revisiting nonlinear Bosentan pharmacokinetics by physiologically based pharmacokinetic modeling: target binding, albeit not a major contributor to nonlinearity, can offer prediction of target occupancy. *Drug Metab Dispos* 49(4):298–304
- Lavezzi SM, Mezzalana E, Zamuner S, De Nicolao G, Ma P, Simeoni M (2018) MPBPk-TMDD models for mAbs: alternative models, comparison, and identifiability issues. *J Pharmacokinet Pharmacodyn* 45(6):787–802
- Lyness J, Moler C (1967) Numerical differentiation of analytic functions. *SIAM J Numer Anal* 4:202–210
- Martins JRRA, Sturdza P, Alonso JJ (2003) The complex-step derivative approximation. *ACM Transactions on Mathematical Software* 29(3):245–262
- Matthews JL, Schultz IR, Easterling MR, Melnick RL (2009) Physiologically based pharmacokinetic modeling of dibromoacetic acid in F344 rats. *Toxicol Appl Pharmacol* 244(2010):196–207
- Monsalve-Bravo GM, Lawson BAJ, Drovandi C, Burrage K, Brown KS, Baker CM, Vollert SA, Mengersen K, McDonald-Madden E, Adams MP (2022) Analysis of sloppiness in model simulations: Unveiling parameter uncertainty when mathematical models are fitted to data. *Sci Adv* 8(38):eabm595
- National Human Genome Research Institute (2023). <https://www.genome.gov/genetics-glossary/Antibody>
- Paul S (2011) Therapeutic antibodies for brain disorders. *Sci Transl Med* 3(84):1–5
- Pearce KJ, Nellenbach K, Smith RC, Brown AC, Haider MA (2021) Modeling and parameter subset selection for fibrin polymerization kinetics with applications to wound healing. *Bull Mathem Biol* 83(5):1–22
- Quaiser T, Mönningmann M (2009) System identifiability testing for unambiguous mechanistic modeling—application to JAK-STAT, MAP kinase, and NF- κ B signaling pathway models. *BMC Syst Biol* 3(1):50
- Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J (2014) Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* 30(10):1440–1448
- Slob W, Janssen PH, van den Hof JM (1997) Structural identifiability of PBPK models: practical consequences for modeling strategies and study designs. *Crit Rev Toxicol* 27(3):261–272

- Smith RC (2014) Uncertainty quantification: theory, implementation, and applications. SIAM, Philadelphia
- Stigter JD, Molenaar J (2015) A fast algorithm to assess local structural identifiability. *Automatica* 58:118–124
- Székely GJ, Rizzo ML (2004) Testing for equal distributions in high dimensions. *InterStat* 5:1249–1272
- Székely GJ, Rizzo ML (2013) Energy statistics: a class of statistics based on distances. *J Stat Plan Inference* 143(8):1249–1272
- Todd P, Brogden R, Muromonab C (1989) A review of its pharmacology and therapeutic potential. *Drugs* 37(6):871–899
- Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP (2015) Perspective: sloppiness and emergent theories in physics, biology, and beyond. *J Chem Phys* 143(1):010901
- Yates JW (2006) Structural identifiability of physiologically based pharmacokinetic models. *J Pharmacokinet Pharmacodyn* 33(4):421–439

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.