Deliverable 3: A simulation study protocol to evaluate the performance characteristics of uncertainty quantification methods of complex models

CONFIRMS Consortium

Date: November 7, 2025

Contents

1 Introduction

Mechanistic models such as physiologically based pharmacokinetic (PBPK) models have become important tools in drug development, particularly within the framework of model-informed drug development (MIDD)[1]. These models enable predictions of drug–drug interactions (DDIs), first-in-human (FIH) dose estimations, etc. In the regulatory aspect, PBPK modeling is commonly used to support decisions related to DDIs, food effects, and dose extrapolation to special populations. One challenge in its application is that PBPK models are characterized by a large number of parameters based on previous studies and literature[2]. The uncertainties of those parameters and their impact on the output metrics are difficult to study systematically. A common approach is sensitivity analysis[3], which only investigates how the change of certain parameters affects the relevant output metrics.

Usually, PBPK modeling is performed on modeling platforms, such as Simcyp, GastroPlus, and PK-Sim, which collectively implement mechanistic models based on previous literature and studies for model structure and system-related parameters. In 2018, the European Medicines Agency (EMA) published a Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation[4], which outlines requirements for qualification of PBPK platforms for high-impact regulatory applications, such as waiving clinical DDI studies or informing dosing decisions in specific populations (e.g., pediatric subgroups) based on limited or no clinical data. Certara conducted a Bayesian hierarchical meta analysis (BHMA) and received EMA's regulatory qualification opinion for Simcyp in 2025 for predicting CYP-mediated DDIs with low to moderate inhibition[5]. The BHMA qualification was based on 220 DDI validation datasets (Certara Drug Interaction Database, formerly known as the University of Washington database), evaluating the predictive performance of SimCyp in terms of bias and imprecision of geometric mean ratio (GMR) and between-subject variances (BSV) of AUC or Cmax. However, the methodology and performance of BHMA for mechanistic models in drug development have not been systematically assessed and compared with other meta analysis approaches. It is unclear how its performance and conclusion are impacted by the design characteristics of the validation datasets, operational features, and model misspecification.

Uncertainty can be understood from two different perspectives. The first concerns the general predictive performance of a complex model platform (such as SimCyp) within a defined range of contexts of use. In this case, uncertainty refers to the imprecision in predicting output metrics of the model platform (e.g., GMR of AUC), which partly reflects variations in data quality and model development practices among previous PBPK models. BHMA method is an example of this approach, as it estimates the prediction bias and imprecision based on the comparison between observed and predicted values of previous DDI studies The second perspective focuses on the uncertainty at the level of a drug-specific model developed using a model platform, e.g., a pbpk model developed for a specific drug by setting drug-specific parameters based on available research results. In this context, the main interest lies in the uncertainty of model input parameters (e.g., drug-specific parameters) and how these uncertainties affect the prediction results for a specific case. To study the impact of uncertainty in input parameters, the current most common approach is sensitivity analysis, which has several drawbacks, including a lack of probabilistic interpretation, neglect of correlation between parameters, and challenges in directly informing the predictive performance. Probabilistic sensitivity analysis (PSA) is a method used to assess how the impact of uncertainty in model input parameters affects the output metrics by assigning probability distributions based on prior knowledge[6]. Unlike deterministic sensitivity analysis, PSA offers a probabilistic framework for understanding the influence of input uncertainty and facilitates more informed decision-making.

2 Objectives

The objectives of this project (Figure 1) are to explore and/or evaluate the uncertainty analysis methods or approaches for regulatory evaluations of complex models:

- To explore and evaluate methods for qualifying PBPK platforms in terms of prediction performance (uncertainty in the output metrics), with a focus on BHMA:
  - To investigate the needed design characteristics of validation datasets for BHMA
  - To investigate the optimal operational characteristics of BHMA
  - To evaluate the performance of BHMA under a series of scenarios, including model misspecification
  - To explore other methods, such as traditional meta-analysis and the frequentist prior method
- To explore PSA for analyzing the uncertainty in model input parameters
  - To implement PSA for PBPK modeling with a suggested workflow for its application
  - To investigate the impact of selected prior uncertainty distributions on PSA results.
- To explore approaches for combining the two sources of uncertainty (uncertainty in output metrics and uncertainty in model input parameters)
  - To get a better understanding of the different types of uncertainty
  - To propose how the investigated approaches can quantitatively inform regulatory decision-making.

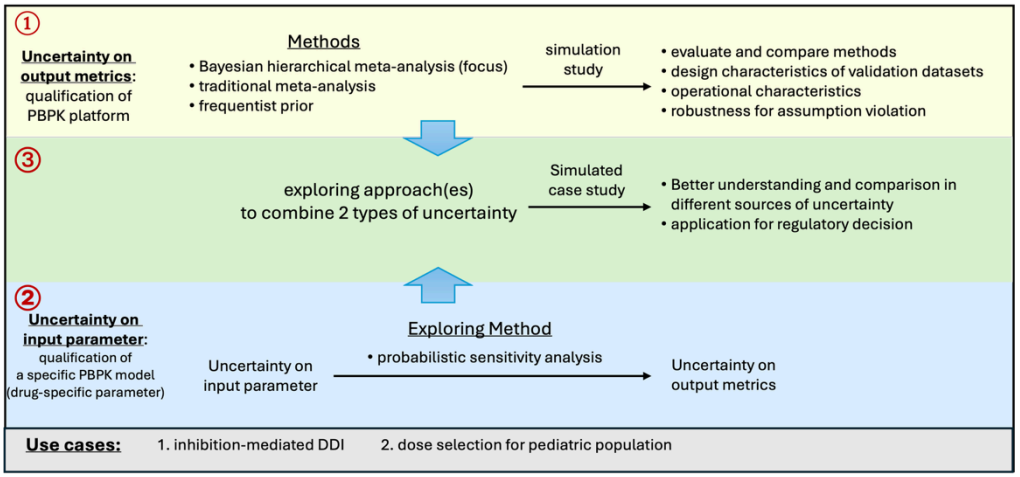

Figure 1: Overview of the simulation study plan

3 Simulation plan

3.1 BHMA evaluation

The model structure of the BHMA method evaluated here is based on the final model reported in Qualification Opinion for Simcyp Simulator[5]. Briefly, the input data for the analysis includes statistics of previous studies (GMR, geometric BSV, and sample size) and the corresponding predicted values from a PBPK platform (predicted GMR and predicted BSV) for those studies. The estimated parameters in the model characterize the uncertainty of the PBPK platform performance in terms of bias and imprecision in predicting GMR and BSV, respectively. The simulation will include a step of applying the BHMA results to inform regulatory decision-making for a hypothetical drug, which compares an uncertainty interval (the 90% prediction interval) calculated from the model with an assumed safety limit (e.g., 2-fold for GMR in the case of inhibition-mediated DDI). If the uncertainty

interval falls within the safety limit, the conclusion based on the PBPK platform is that the DDI is likely to pose no risk. Otherwise, there is a risk of toxicity (overdose) or lack of efficacy (underdose) depending on the selected context of use.

3.1.1 Simulation Workflow

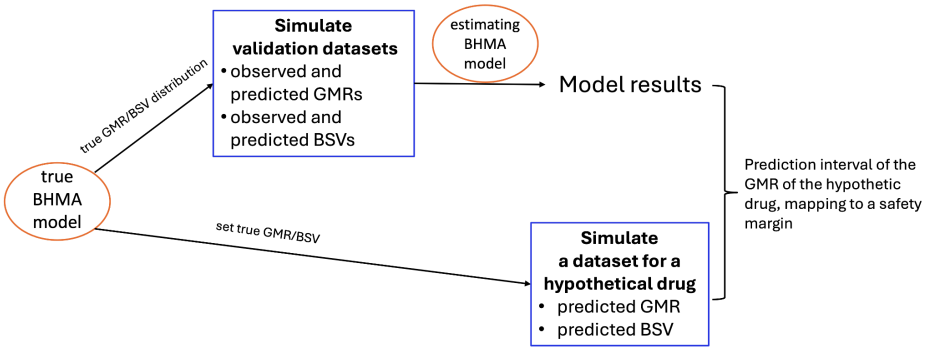

Figure 2: Simulation workflow for Bayesian Hierarchical meta analysis

The simulation workflow for evaluating BHMA (Figure 2) will include the following steps:

- **Simulating datasets:** An assumed true BHMA model will be used to simulate: (1) validation datasets with pre-specified design features (simulated data are composed of observed and predicted GMR and BSV for each dataset); (2) a dataset for a hypothetical drug (predicted GMR and predicted BSV) representing a simulation using a PBPK platform. It should be noted that a distribution of true GMRs is needed for simulating validation datasets, since the BHMA quantifies the differences between observed and predicted GMRs without information about the true GMR distribution. The used distribution for true GMRs should meet the EMA requirement that the validation datasets include inhibitors of different potencies (mild, moderate, strong)[4]. A true GMR value will be set for the hypothetical drug, depending on whether the simulation is to calculate type I error or power. A similar procedure will be used for generating BSV values from an assumed true distribution. It should be noted that while the simulation setup will focus on GMR of AUC, the evaluated method can be readily extended to GMR of Cmax.
- **BHMA analysis:** BHMA analysis is performed on the simulated validation datasets. It should be noted that the BHMA model used for estimation may differ from the model used for data simulation (the true model) when evaluating the impact of model misspecification, such as different prior assumptions.
- **Conclusion for the hypothetical drug:** The resultant BHMA model will be used to create a 90% prediction interval for the hypothetical drug based on the simulated data. The conclusion regarding the risk of DDI will then be obtained by comparing the prediction interval to the pre-set safety limit.
- **Simulation summary:** The simulation study will be conducted with at least 1,000 simulation runs. Based on this, the percentage of cases concluding no DDI risk (i.e., 90% prediction intervals within the safety limit) will be calculated, which corresponds to either the type I error or power, depending on the design scenario (true GMR value for the hypothetical drug).

3.1.2 Use Case 1

Use Case 1 will focus on quantifying the impact of competitive, inhibition-mediated drug interactions on investigational compounds. For example, the investigational drug is a victim of CYP3A4, and the predicted drug interaction is caused by a co-administered drug that is a mild or moderate inhibitor of CYP3A4. This use case will be the main case for evaluating the BHMA approach and investigating operational characteristics. The safety limit will be set as a 2-fold increase in GMR, i.e., a clinically relevant DDI is associated with a risk of toxicity if the GMR (with vs. without the inhibitor) is larger than 2. It should be noted that we only focus on the upper limit of the GMR since the inhibition-mediated DDI is expected to have a higher GMR.

3.1.2.1 Design characteristics of validation datasets

Different numbers of validation datasets will be evaluated in the simulation study, with a planned range from 4 to 200. If encountering identifiability/converge issues, the dataset number will be adjusted. In addition, the subject number of each dataset will be another design characteristic to investigate, which will be between 5 and 30, a common range for a DDI study. The simulation study will cover two scenarios to investigate the minimal size for validation datasets: (1) all datasets have the same subject number, and (2) they have different subject numbers.

3.1.2.2 Operational characteristics

In the final model for the Simcyp qualification[5], the priors used in the BHMA are standard normal distributions for bias and truncated Cauchy distributions for imprecision (i.e., variances). We will investigate other prior settings, such as distributions with different spreads and locations.

The method's robustness against a violation of the assumption of constant bias and imprecision regardless of DDI magnitude will be investigated. To achieve this, in a simulation study, we will use a BHMA model with varying magnitudes of bias and/or imprecision for different GMRs (i.e., different DDI potencies) and apply a BHMA with constant bias/imprecision as an estimation model.

3.1.2.3 Evaluation scenarios

For evaluating type I error, the true GMR for the hypothetical drug will be set at 2. For evaluating power, the true GMR will be set at 0.8, 1, and 1.5, respectively.

In addition, the BHMA approach will be evaluated under different PBPK platform performance:

- High bias, high imprecision
- Low bias, high imprecision
- High bias, low imprecision
- Low bias, low imprecision

The high and low bias may be set at 50% and 5%, respectively. The high and low imprecision may be set at 100% and 20%, respectively.

3.1.3 Use Case 2

In Use Case 2, PBPK platforms are used to support dose selection for the pediatric population (e.g., a certain age group) based on simulated exposure in the case of limited or no clinical data. The simulation study of this case is to evaluate BHMA performance for exposure matching compared to adults based on GMR of AUC with both lower and upper boundaries, which may be set to 0.6 and 1.67, respectively (modified based on the FDA reference[7]). The GMR between pediatric and adult populations under selected dose needs to be within 0.6-1.67 to claim comparable exposure.

Under this newly defined safety zone, the BHMA method will be evaluated across various scenarios, similar to those in Use Case 1 (Section 3.1.2), including:

- Different dataset sizes;
- Different prior settings;
- Different PBPK platform performance levels (combinations of high/low bias and imprecision).

3.1.4 Exploring other meta-analysis methods

In addition to BHMA, we will explore alternative meta-analysis approaches, such as traditional meta-analysis (where bias and imprecision are treated as fixed values) and the frequentist prior approach[8,9] (which incorporates prior information as a penalty term for the maximized likelihood function to account for discrepancy between the prior and the available data). We will first attempt to implement these methods, and if successful, they will be included in the simulation study and compared with BHMA.

3.2 Investigation of PSA for uncertainty in model input parameters

Our objective is to implement PSA to quantify the impact of uncertainty in model input parameters, which include two steps:

Step 1: Construction of Parameter Uncertainty Distributions

In practice, uncertainty in model input parameters may arise from expert opinion, preclinical and clinical data, and/or literature reports. In this study, we will consider the following elements of prior information: the most probable value (mode), the uncertainty variance (or standard error), and plausible parameter bounds. Using these inputs, we will construct parameter uncertainty distributions with potential options such as (truncated) normal, (truncated) t distribution, logit-normal distribution, uniform distribution, and beta distributions[6].

Step 2: Uncertainty Propagation

Once prior distributions for parameter uncertainty are established, propagation of this parameter uncertainty will be performed, using, e.g., random sampling (Monte Carlo simulation) to generate a set of parameter values. For each sampled parameter set, the corresponding output metric (GMR of AUC) will be obtained through the PBPK model. The result of PSA is an uncertainty distribution for the model output.

The investigation will focus on Use Case 1, i.e., applying PBPK models to predict CYP competitive inhibition-mediated DDI. Our implementation of PSA will focus on a single parameter. If results are promising and time allows, the analysis may be extended to two parameters, with and without correlation.

We will also investigate how the choice of prior uncertainty distribution affects PSA outcomes. Briefly, resulting output distributions will be compared across different prior distributions constructed using the same information (i.e., mode and variance). Furthermore, we will explore effective ways to visualize, summarize, and interpret PSA results to better support regulatory decision-making. With all these efforts, we aim to propose a workflow for implementing probabilistic sensitivity analysis to characterize parameter uncertainty in complex models and inform regulatory decision-making in a transparent and quantitative manner.

3.3 Exploring approaches to combine uncertainties from different sources

In Section 3.1, the bias and uncertainty analyzed through meta-analysis reflect the general performance of a PBPK platform under a specific context of use. The imprecision or uncertainty in the output metrics partly reflects variations in data quality and model development practices among previous PBPK models. In contrast, Section 3.2 focuses on uncertainty related to the input parameter uncertainty of a specific PBPK model, largely in drug-specific parameters in the case of DDIs. When making regulatory decisions based on complex modeling, it is important to consider the uncertainties associated with both the model platform and the specific model implementation. The key research question for this section is whether it is possible to combine these two sources of uncertainty to better inform regulatory decision-making.

This work will be exploratory in nature. We will explore potential approaches using a simulated case study (Use Case 1: PBPK model for DDI prediction). A potential simulated case study is illustrated in Figure 3. Briefly, one or more datasets of clinical DDI studies will be simulated based on published models (either PBPK or population PK models). Using these simulated datasets, key parameters (e.g., fraction metabolized by a specific CYP, fm) will be estimated with uncertainty (such as SE). A selected UQ method (e.g., PSA) may be carried out based on the resultant updated PBPK model with the uncertainty on the estimated parameter(s) to obtain predicted output metrics (such as GMR of AUC) with uncertainty and corresponding BSV. In parallel, a BHMA model may be estimated using a collection of DDI datasets across different drugs under a certain context of use (e.g., the BHMA model reported in the SimCYP qualification study).

Comparing the uncertainties obtained from these two paths will provide insights into the relative magnitudes. Furthermore, different strategies may be explored to integrate these two types of uncertainties—those arising from the PBPK model input parameters and those captured by the BHMA model—to support regulatory decision-making. One potential approach is to incorporate the propagated uncertainty directly into the BHMA framework.

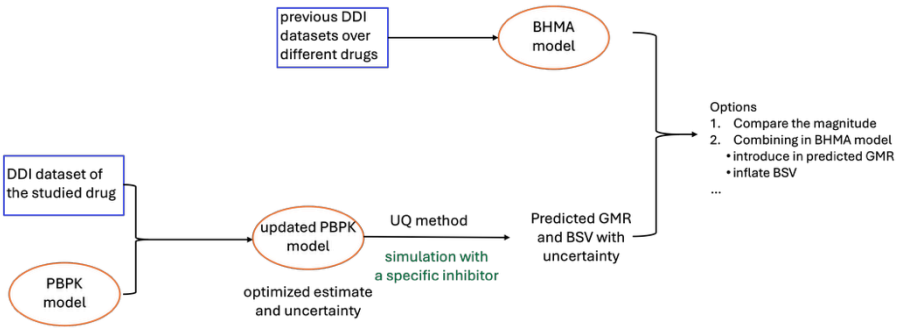

Figure 3: a potential simulated case study to explore approaches of combining uncertainties of different sources

4 Software

The Simulations will be performed in R[11], making use of the SimDesign package[12] to initialize random number generators, save random seeds, dispatch the computations, and collect the results. Functions to generate the data, apply the selected methods, and aggregate the results will be written by the consortium. The different scenarios will be implemented in functions that call the data-generating functions from other packages with the respective parameters or their own code.

Implementations of the data analysis methods applied to the generated data will be used from different packages where they are already implemented. For example, the R package Rstan[13] may be used for Bayesian analysis. NONMEM may be used to implement the frequentist prior method. Wrapper functions compatible with the SimDesign framework will be provided. All R functions and documentation will be published in an R package as a GitHub repository; the code to reproduce the simulation study, tables, and graphs from the report will be published as one or more vignettes to the R package. Code used to generate data from other scenarios will also be made available.

In the simulated case study (Section 3.3) and possibly implementing sensitivity analysis (Section 3.2), the PBPK model used may be selected from the Open Systems Pharmacology PBPK model library[14]. PK-sim, an open access software, will likely be used for PBPK modeling.

5 Discussion

Choice of methods to explore

The methods to explore proposed in this simulation plan were carefully selected based on an overall assessment of several factors.

- The proposed methods were identified as suitable for evaluating uncertainty of mechanistic models for model-informed drug development during the literature review performed by the consortium (Deliverable 2).
- The proposed methods were selected based on the relevance of the method for regulatory decision-making according to the EMA PBPK guideline[4].
- Practical considerations were deemed necessary considering the available resources for the research project (including project timelines).

In addition to the above, the proposed methods were also considered aligned with the use cases and planned evaluation of operational characteristics as outlined in the preliminary study plan written at the beginning of the project (Deliverable 1).

Two distinct types of methods were identified in the literature review including methods mainly concerning uncertainty on model output (e.g. uncertainty in predicted GMR of AUC) and methods mainly concerning uncertainty on model input parameters. Thus, we propose to focus on one method mainly concerning uncertainty in model output and to explore one method mainly concerning uncertainty in model input.

The qualification of PBPK platforms is required for applying PBPK models in high-impact regulatory decisions according to the EMA PBPK guideline[4], which suggests comparisons of predicted output vs observed data (i.e. validation datasets). This emphasizes the regulatory relevance of exploring a method which mainly concerns uncertainty on model output. BHMA was identified as a suitable method according to the literature review. BHMA was used in a recent important regulatory precedent where the SimCYP simulator got a positive EMA qualification[5]. Furthermore, the EFPIA Pharma industry PBPK expert team shared a position statement at an EMA workshop on mechanistic models which commended the use of BHMA for PKPK platform qualification. Taken together, there is a high regulatory relevance of exploring and evaluating BHMA. Therefore, our simulation plan focuses on evaluating the BHMA method and investigating its operational characteristics (Section 3.1). Notably, BHMA does not directly involve PBPK models and could potentially be extended to other complex modeling frameworks.

Several UQ methods identified in the literature review (Deliverable 2) concentrate on uncertainties in model input parameters. These methods are more relevant for assessing the quality of individual PBPK models rather than qualifying modeling platforms *per se*. The EMA PBPK guideline[4] describes the role of sensitivity analyses when developing and applying mechanistic models. While such methods are also important, further exploratory work is needed to implement them and to develop strategies for interpreting their results in a regulatory context. This is the focus of Section 3.2. PSA was identified as a suitable method in the literature review[6,15], mainly concerning uncertainty on model input parameters. We propose to explore PSA due to its relative simplicity, ease of implementation, and interpretability.

Several other promising UQ methods—such as Bayesian calibration[16] and global sensitivity[17,18] analysis—also hold potential for application to complex models including PBPK, QSP, and machine learning models in drug development and evaluation. These methods merit investigation in future projects. Importantly, beyond technical implementation, it is essential to understand how the results from these UQ methods can be linked to and inform regulatory decision-making.

Terminology considerations

In this simulation plan, type I error and statistical power are used as evaluation metrics for BHMA. It should be noted that these metrics are borrowed from the concepts of frequentist hypothesis testing to assess the chances of making correct or incorrect decisions. However, since BHMA is a Bayesian approach, it does not involve predefined significance levels. Therefore, we do not expect type I error to align with conventional thresholds (e.g., 5%) used in frequentist frameworks.

Reference

1. Jones, H. M. & Rowland-Yeo, K. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacomet. Syst. Pharmacol.* **2**, (2013).
2. Kuepfer, L. *et al.* Applied Concepts in PBPK Modeling: How to Build a PBPK/PD Model. *CPT Pharmacomet. Syst. Pharmacol.* **5**, 516 (2016).
3. Tsamandouras, N., Rostami-Hodjegan, A. & Aarons, L. Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. *Br. J. Clin. Pharmacol.* **79**, 48–55 (2015).
4. Committee for Medicinal Products for Human Use (CHMP) Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. *Eur. Med. Agency* (2018).
5. European Medicines Agency (EMA) *Qualification opinion for Simcyp Simulator*. (2025).
6. Lash, T. L. *et al.* Good practices for quantitative bias analysis. *Int. J. Epidemiol.* **43**, 1969–1985 (2014).
7. Wang, Y., Jadhav, P. R., Lala, M. & Gobburu, J. V. Clarification on precision criteria to derive sample size when designing pediatric pharmacokinetic studies. *J. Clin. Pharmacol.* **52**, 1601–1606 (2012).
8. Langdon, G., Gueorguieva, I., Aarons, L. & Karlsson, M. Linking preclinical and clinical whole-body physiologically based pharmacokinetic models with prior distributions in NONMEM. *Eur. J. Clin. Pharmacol.* **63**, 485–498 (2007).
9. Gisleskog, P. O., Karlsson, M. O. & Beal, S. L. Use of prior information to stabilize a population data analysis. *J. Pharmacokinet. Pharmacodyn.* **29**, 473–505 (2002).
10. Posada, M. M. *et al.* Predicting Clinical Effects of CYP3A4 Modulators on Abemaciclib and Active Metabolites Exposure Using Physiologically Based Pharmacokinetic Modeling. *J. Clin. Pharmacol.* **60**, 915–930 (2020).

11. R Core Team *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2021).at <https://www.r-project.org/>
12. Chalmers, R. P. & Adkins, M. C. Writing Effective and Reliable Monte Carlo Simulations with the SimDesign Package. *Quant. Methods Psychol.* **16**, 248–280 (2020).
13. Stan Development Team. RStan: the R interface to Stan. (2020).at <http://mc-stan.org/>
14. Lippert, J. *et al.* Open Systems Pharmacology Community—An Open Access, Open Source, Open Science Approach to Modeling and Simulation in Pharmaceutical Sciences. *CPT Pharmacomet. Syst. Pharmacol.* **8**, 878–882 (2019).
15. Pelekis, M., Nicolich, M. J. & Gauthier, J. S. Probabilistic Framework for the Estimation of the Adult and Child Toxicokinetic Intraspecies Uncertainty Factors. *Risk Anal.* **23**, 1239–1255 (2003).
16. Nong, A. *et al.* Bayesian Calibration of a Physiologically Based Pharmacokinetic/Pharmacodynamic Model of Carbaryl Cholinesterase Inhibition. *J. Toxicol. Environ. Health A* **71**, 1363–1381 (2008).
17. Scherholz, M. L., Forder, J. & Androulakis, I. P. A framework for 2-stage global sensitivity analysis of GastroPlus™ compartmental models. *J. Pharmacokinet. Pharmacodyn.* **45**, 309–327 (2018).
18. Najjar, A., Hamadeh, A., Krause, S., Schepky, A. & Edginton, A. Global sensitivity analysis of Open Systems Pharmacology Suite physiologically based pharmacokinetic models. *CPT Pharmacomet. Syst. Pharmacol.* **13**, 2052–2067 (2024).