



A latent variable approach to account for correlated inputs in global sensitivity analysis

Nicola Melillo¹ · Adam S. Darwich²

Received: 23 February 2021 / Accepted: 6 May 2021 / Published online: 25 May 2021
© The Author(s) 2021

Abstract

In drug development decision-making is often supported through model-based methods, such as physiologically-based pharmacokinetics (PBPK). Global sensitivity analysis (GSA) is gaining use for quality assessment of model-informed inference. However, the inclusion and interpretation of correlated factors in GSA has proven an issue. Here we developed and evaluated a latent variable approach for dealing with correlated factors in GSA. An approach was developed that describes the correlation between two model inputs through the causal relationship of three independent factors: the latent variable and the unique variances of the two correlated parameters. The latent variable approach was applied to a set of algebraic models and a case from PBPK. Then, this method was compared to Sobol's GSA assuming no correlations, Sobol's GSA with groups and the Kucherenko approach. For the latent variable approach, GSA was performed with Sobol's method. By using the latent variable approach, it is possible to devise a unique and easy interpretation of the sensitivity indices while maintaining the correlation between the factors. Compared methods either consider the parameters independent, group the dependent variables into one unique factor or present difficulties in the interpretation of the sensitivity indices. In situations where GSA is called upon to support model-informed decision-making, the latent variable approach offers a practical method, in terms of ease of implementation and interpretability, for applying GSA to models with correlated inputs that does not violate the independence assumption. Prerequisites and limitations of the approach are discussed.

Keywords Latent variable · Correlated factors · Global sensitivity analysis · Physiologically based pharmacokinetic models · Model-informed drug discovery and development

Introduction

In pharmaceutical research and development (R&D) decision-making is often supported by modelling and simulation (M&S), referred to as model-informed drug discovery and development (MIDD) [56]. Physiologically-based pharmacokinetic (PBPK) M&S provides a framework for mechanistic predictions of in vivo drug exposure. PBPK M&S has replaced/supplemented clinical trials and

informed labelling for numerous drugs, most notably for dosage recommendations following metabolic drug–drug interactions [22, 59, 59].

Uncertainty and variability are prominent in biological data. In this context, uncertainty mainly relates to inter- and intra-experimental variability and errors, as well as translation of parameters. Variability mainly relates to interindividual variability in physiology, interoccasion variability and more. Correlations between input parameters are often implemented in PBPK models to account for physiological constraints, otherwise causing implausible combinations of parameters [44, 53]. For example, organ weights are constrained by body weight. With the emergence of novel 'omics techniques, the correlation of proteins is also of increasing interest [12, 37].

Sensitivity analysis (SA) and global SA (GSA) are essential instruments for the quality assessment of model-based inference [43] and their use has gained interest from

✉ Adam S. Darwich
darwich@kth.se

¹ Centre for Applied Pharmacokinetic Research, Division of Pharmacy & Optometry, School of Health Sciences, The University of Manchester, Manchester, UK

² Division of Health Informatics and Logistics, Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Stockholm, Sweden

pharmaceutical industry and academia in recent years [10, 26, 34, 36–38, 58, 60]. Moreover, both the United States Food and Drug Administration (FDA) and European Medicines Agency (EMA) have highlighted the importance of SA and GSA as best practice in PBPK to inform model development and refinement [6, 7]. GSA is key for elucidating the relationship between the uncertainty and variability in model inputs and variation in a given model output. For example, GSA can be a valuable method for testing if the model behaves as expected and, if not, it can provide useful information that helps in identifying the reasons and possible errors in the model assumptions or implementation. Moreover, GSA can help identify what parameters may need to be more precisely characterised to allow reliable model predictions [36, 38, 50]. Therefore, by extension, the method is relevant for decision-making informed by modelling in drug development and clinical practice [17, 36, 37, 47, 58].

In this work, we focused on the variance-based GSA (also referred to as Sobol's method) [36, 38]. This choice was made as variance-based GSA is able to handle non-linear and nonmonotonic relationships between the input factors and the model outputs [49–51]. Moreover, with this method it is possible to quantify the effect of each factor taken singularly and the extent of its interaction effects. As we have reported in our previous work, understanding the extent of the interaction effects can be particularly important for an informed use of PBPK models during drug development [38].

The classical variance-based GSA works under the assumption that model inputs (commonly referred to as model parameters in pharmacometrics) are independent [49–51]. Under this assumption, the variance decomposition is unique [51] and reflects the structure of the model itself [40]. In this context, the variance-based sensitivity indices have a clear interpretation [21, 49]. However, it is not uncommon that PBPK models violate the independence assumption [26, 37, 53]. In practice this may lead to correlations being ignored in the analysis, or the use of one of several proposed methods for GSA that deal with dependent inputs. Perhaps, the most simple and elegant way of treating dependent inputs in GSA is by grouping the correlated factors and then performing a GSA with the independent groups. The intrinsic limitation of this approach is that it is not possible to distinguish the contribution of the single variables within each group.

In the literature, several methods have been developed to deal with dependent inputs while retaining the information, or sensitivity indices, of each individual factor. These methods can be classified into two categories: parametric and non-parametric methods [11, 31]. The parametric methods, also called model-based methods, (e.g., [9, 25, 57]) assume an a priori model for the input-

output relation. Instead, the non-parametric approaches do not assume any specific shape for this relation and thus, they are referred to as model-free or non model-based methods [11, 31]. These approaches were considered to be more suitable for computer-based modelling [11]. Generally, the non-parametric methods employ a transformation technique for dealing with correlated factor distributions [11]. For example, Kucherenko et al. [24] used copula transformations to generalise the first order and total Sobol indices for the case of dependent input factors. Mara et al. [31] proposed the use of the Rosenblatt transformation, and Tarantola and Mara [52] used both the Rosenblatt and Nataf transformation within the context of variance-based GSA. Moreover, other methods such as the variogram analysis of response surfaces (VARS) and the Shapely effects have been extended for the case of correlated input factors [11, 21].

The copula-based method, developed by Kucherenko et al. [24], has recently been proposed for PBPK models [26]. However, how to interpret variance-based GSA results in presence of dependent variables is not straightforward and still debated among GSA practitioners. In presence of correlation between the input factors, the correspondence between the variance-based indices and model structure is lost and the variance decomposition can no longer provide a description of the model structure [3, 40, 42]. This was illustrated by Oakley and O'Hagan in 2004 with the use of a simple example [40]. In this context, Pianosi et al. reported that *counterintuitive results may be obtained* [42]. Iooss and Lemaître reported that *SA for dependent inputs has also been discussed by several authors [...], but this issue remains misunderstood* [20]. Moreover, Iooss and Prieur reported that *The so-called Sobol' indices [...], present a difficult interpretation in the presence of statistical dependence between inputs* [21]. Finally, a recent position paper Razavi et al. reported that *The field of SA in terms of methods to handle input constraints and correlation structures is still embryonic* [43].

Several dedicated software platforms exist for PBPK M&S [23], providing accessible tools for non-expert users. As GSA gains use in the community (such as through software implementation) the issue of interpretability becomes increasingly relevant.

Here we propose a latent variable approach for treating correlated input parameters in variance-based GSA. The method expresses the correlation between two parameters as causal relationships between uncorrelated variables. This is done in order to allow the use of classical variance-based GSA and avoids the usage of methods whose interpretation is still a matter of debate. Latent variable models and sub-varieties of them, such as factor analysis, path analysis and structural equation modelling, are widely used in social sciences [28]. In latent variable models, the

correlation between more than one observed measure (or model parameter) is described by one, or more, unobserved (latent) variable(s). Parameters are correlated as they share a common cause [4]. Here we focus on the case of two linearly correlated random variables whose correlation is explained by one latent variable. With this approach, instead of two correlated factors, three independent factors (the latent variable and the two independent variances of the correlated parameters) are considered in the GSA.

The approach is then applied to a set of algebraic models and a whole-body PBPK model for the drug midazolam (MDZ). MDZ is a sedative primarily metabolised by Cytochrome P450 (CYP) 3A4 and CYP3A5 [16]. The expression of CYP3A5 is polymorphic and present in around 10–20% [48] of Caucasians where it is correlated with CYP3A4 through a shared mechanism for expression [29]. The latent variable approach was then compared with the classic Sobol’s variance-based GSA, Sobol’s GSA performed by grouping together the correlated factors, and the Kucherenko approach.

Materials and methods

Variance-based sensitivity analysis and the Kucherenko approach

Let us consider the generic model in Eq. 1:

$$Y = f(\mathbf{X}), \tag{1}$$

where Y is the scalar model output, \mathbf{X} is the vector including the k independent input factors ($X_i, i = 1, \dots, k$) and f is the input–output relationship. In variance-based GSA, two sensitivity indices are derived from the functional decomposition of the variance (V) of Y , in Eq. 2.

$$V(Y) = \sum_{i=1}^k V_i + \sum_i \sum_{j>i} V_{i,j} + \dots + V_{1,\dots,k} \tag{2}$$

The functional decomposition of the variance presented in Eq. 2 is also known as *functional ANOVA* [50, 51]. V_i is called the first order term and it is the portion of $V(Y)$ explained by the variation of each X_i taken alone [49], where E is the expectation operator. $V_{i,j}$ is the second order term and it is the portion of $V(Y)$ explained by the interaction between X_i and X_j . Similarly, it is possible to define all the higher order interaction terms. Variance-based, or Sobol, sensitivity indices can be defined from 2 as in Eq. 3 [18, 50].

$$S_i = \frac{V_i}{V(Y)} = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y | X_i))}{V(Y)} \tag{3}$$

$$S_{T_i} = \frac{E_{\mathbf{X}_{\sim i}}(V_{X_i}(Y | \mathbf{X}_{\sim i}))}{V(Y)}$$

S_i is the so called first order index (or main effect) and S_{T_i} is the total effect. $\mathbf{X}_{\sim i}$ represents a vector including all the factors except X_i . S_i is related with the part of $V(Y)$ explained by the variation of X_i taken singularly and S_{T_i} is the sum of S_i with all the interaction effects of X_i with the other inputs [49, 50]. When the parameters are independent, the relationships $S_i \leq S_{T_i}$ and $\sum S_i \leq 1$ are always valid and $S_{T_i} - S_i$ gives information about the extent of interaction effects involving X_i [49, 50].

The GSA method proposed by Kucherenko et al. [24] extends the variance based methods for models with dependent input factors. Here, the main and total effects of the variance-based GSA are calculated with a copula-based method. With this approach, S_i includes the effects of the dependence of X_i with other factors [31] and can be higher than S_{T_i} . As reported by [31], S_{T_i} includes only the effects of X_i that are not due to its dependence with $\mathbf{X}_{\sim i}$. A given factor whose importance is only due to the correlation with another factor would have $S_{T_i} = 0$, but S_i can differ from 0 [31]. Moreover, S_{T_i} approaches 0 as the correlation $|\rho| \rightarrow 1$ [24]. A possible explanation for this behaviour is that as the correlation approaches 1, the value of X_i is completely informed by $\mathbf{X}_{\sim i}$ and thus $V_{X_i}(Y | \mathbf{X}_{\sim i})$ will tend to 0.

Latent variable approach for GSA

The latent variable approach expresses the inter-correlation between two parameters as causal relationships between uncorrelated variables and therefore, it allows the use of classical variance-based GSA.

Latent variable methods partition the *observed variance* of each correlated parameter (observed variable) into two parts: a *common variance*, caused by the latent variable and a *unique variance*, specific to the parameter itself [4]. In this work, we focus on the case of two linearly correlated random variables whose correlation is explained by one latent variable. The relationship between the observed, common and unique variances for two correlated parameters and one latent variable is reported through a path diagram as shown in Fig. 1 [28]. Following the notation of latent-variable methodology, η is the latent variable, and is conventionally represented by a circle in the path diagram. Unidirectional arrows represent the causal relationships between latent and dependent factors $X_i, i = 1, 2$ (depicted by a box) and ε_i represents the unique variance associated with X_i [4]. X_1 and X_2 are considered linearly correlated, with a linear (Pearson) correlation coefficient of ρ_{12} . Here

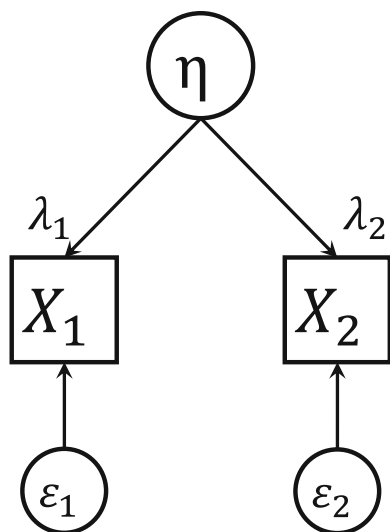


Fig. 1 Relationship between the observed, common and unique variances for two correlated parameters and one latent variable. X_1 and X_2 are the observed variables, η is the latent variable, ε_1 and ε_2 are the unique variances and λ are the factor loadings

we assume that η , X_i and ε_i are distributed as in Equation system 4 and that η and ε_i are independent.

$$\begin{aligned} \eta &\sim \mathcal{N}(0, 1) \\ X_i &\sim \mathcal{N}(0, 1) \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma_i^2) \end{aligned} \tag{4}$$

A common assumption is that the causal relationships between η and X_i are linear. In this case, it is possible to write the following Equation system 5 [4, 28].

$$\begin{aligned} X_1 &= \lambda_1 \eta + \varepsilon_1 \\ X_2 &= \lambda_2 \eta + \varepsilon_2 \end{aligned} \tag{5}$$

λ_1 and λ_2 are called the *factor loadings* and represent the correlations of X_1 and X_2 with η [14]. Given that our hypothesis is that η and X_i are standard normal random variables, and that ε_i is distributed normally with a mean equal to 0 and variance σ_i^2 , by calculating the variance of both sides of the equations in Equation system 5, it is possible to derive that $\sigma_i^2 = (1 - \lambda_i^2)$, $i = 1, 2$.

Now, to correctly express X_1 and X_2 as functions of η , we need to define λ_1 , λ_2 and σ_1^2 , σ_2^2 . According to *path analysis* theory, the correlation between X_1 and X_2 can be expressed as $\rho_{12} = \lambda_1 \cdot \lambda_2$ [28]. With the hypotheses that $\rho_{12} > 0$ and that X_1 and X_2 have the same relationship with η , thus $\lambda_1 = \lambda_2 = \lambda$, it is possible to define λ as in Eq. 6 [28].

$$\lambda = \sqrt{\rho_{12}} \tag{6}$$

Another possible solution is $\lambda = -\sqrt{\rho_{12}}$, where the latent variable has a negative correlation with both X_1 and X_2 . In case of $\rho_{12} < 0$, the absolute values of both factors loadings are equal to $\sqrt{\rho_{12}}$, while their signs are opposite.

According to Eq. 5, λ^2 is the portion of the variance of X_i that is attributed to the latent factor. With our approach, λ^2 is the average variance extracted (AVE). AVE can be defined as *the average amount of variation that a latent construct is able to explain in the observed variables* [14]. Intuitively, this is the overall amount of variance that ‘is taken’ from our dependent factors X_i and attributed to the latent variable η , in order to define the causal relationships in Eq. 5. As shown in the Appendix 5, with our hypothesis that X_1 and X_2 have the same relationship with η , the AVE is minimised. This means that we are explaining the correlation between two observed variables by attributing (on average) the minimum variance possible to the latent construct.

With the latent variable approach, instead of two correlated random variables (X_1 and X_2), three independent random variables (η , ε_1 and ε_2) will be considered in the variance-based GSA. In this context, the impact of ε_1 and ε_2 on the model output can be uniquely attributed to X_1 and X_2 , respectively. Instead, it would be impossible to distinguish if the impact of η on the model output is primarily mediated by X_1 or X_2 .

For simplicity, we have considered standardised variables. However, the latent variable approach can easily be extended to data in original units with the use of simple transformations. Nevertheless, in order to use this method several assumptions must be satisfied (summarised in Table 1) and some limitations still exist. The sums of the random variables representing the latent and independent variances must follow the distributions of X_i . This condition is satisfied if both the parameters are normally distributed and it can easily be extended to the case of the two parameters being log-normally distributed. However, the condition in Equation system 5 is not easily satisfied for

Table 1 Assumptions for the use of the latent variable approach

Assumptions ^a
Only two correlated input factors X_1 and X_2
A linear correlation between X_1 and X_2
η , ε_1 , ε_2 , X_1 , X_2 normally distributed as in Eq. 4
Linear relation between η and X_1 , X_2 , as in Eq. 5
Same relation between X_1 , X_2 and η , thus $ \lambda_1 = \lambda_2 = \lambda $ in Eq. 5
^a X_1 , X_2 are the dependent input factors
η is the latent variable
ε_1 , ε_2 are the unique variances

other types of distributions. The method presented here is valid when considering two correlated factors and it can be extended to three mutually correlated factors, by using the so called *method of triads* to derive a unique solution for the factor loadings [28]. However, it is possible that there is no unique solution when more than three mutually correlated factors are considered [28]. In this situation, the application of the latent variable approach for GSA would become more challenging.

The practical implementation of the latent variable approach is relatively straightforward. First, λ is defined as per Eq. 6, where ρ_{12} is the linear correlation between the two variables of interest, X_1 and X_2 . Then, the values for η are extracted from a standard normal distribution, while the ones for ε_1 and ε_2 are extracted from a normal distribution, with mean 0 and variance $\sigma^2 = 1 - \lambda^2$. X_1 and X_2 are then defined as per Eq. 5. By doing this, X_1 and X_2 would be standard normal random variables. Then, they can be easily transformed to normal variables with the desired mean and standard deviation. As previously stated, the approach can be extended to X_1 and X_2 being log-normally distributed, although in this case $\log(X_1)$ and $\log(X_2)$ should be linearly correlated.

Algebraic models

The latent variable approach was initially tested on three algebraic models, namely model 1, 2 and 3, in Eqs. 7, 8 and 9 respectively.

$$Y = X_1 + X_2 + X_2 \cdot X_3 \quad (7)$$

$$Y = X_1 + X_2 + X_1 \cdot X_3 \quad (8)$$

$$Y = X_1 + X_2 + X_3 + X_4 \quad (9)$$

For all the models, all factors were considered to be normally distributed with means equal to 0 and variances equal to 1, $X_i \sim \mathcal{N}(0, 1)$, $i = 1, 2, 3, 4$. X_1 and X_4 were considered linearly correlated, with a Pearson correlation coefficient of ρ_{14} . Model 1 and model 2 differ in the fact that in model 1, X_1 is not involved in any interaction, while in model 2, X_1 interacts with X_3 .

X_4 does not appear in the model 1 or model 2 equations, consequently, its ‘causal impact’¹ on the model output Y must be null. Intuitively, for both model 1 and 2, the results of a variance-based GSA in absence of correlation, considering only X_1 , X_2 and X_3 , will correctly reflect the structure of the model.

¹ Here we refer to ‘causal impact’ as the impact of an input factor X_i on the model output Y that is not due to the dependence of X_i with other factors.

Whole-body PBPK model for midazolam

A whole-body PBPK model was developed, describing the pharmacokinetics of the drug MDZ following an intravenous (IV) bolus injection in a population of human healthy subjects. The model is represented in Fig. 2. This section provides a brief description of the model. For a detailed account of the model equations, the parameters used for the PBPK construction and the algorithm used for generating the population, see the Supplementary Material.

The typical equation used to describe the mass balance in a given organ or tissue t within a PBPK model is reported in Eq. 10. For a detailed description and the underlying theories of this model, called *well-stirred perfusion-limited PBPK*, please refer to [2].

$$\frac{dx_t}{dt} = Q_t \left(\frac{x_{art}}{V_{art}} - \frac{x_t/V_t}{P_{t,p}/B:P} \right) \quad (10)$$

Equation 10 is valid for all organs and tissues except the liver, the lungs, the arterial and venous blood. x_t is the drug amount in compartment t , while V_t is the volume. Subscript *art* stands for arterial blood. Q_t is the blood flow to compartment t . $B:P$ is the blood-to-plasma ratio and $P_{t,p}$ is the tissue-to-plasma partition coefficient.

MDZ is primarily metabolised in the liver by the two enzymes, CYP3A4 and CYP3A5. For MDZ both enzymes catalyse two reactions, leading to the formation of two metabolites, *1-hydroxy midazolam* (1-OH-MDZ) and *4-hydroxy midazolam* (4-OH-MDZ) [16, 54]. For this reason, two mass flows corresponding to MDZ metabolism leave the PBPK system from the liver compartment, as represented in Eq. system 11.

$$\begin{aligned} \frac{dx_{liv}}{dt} = & Q_{liv} \left(\frac{x_{art}}{V_{art}} - \frac{x_{liv}/V_{liv}}{P_{liv,p}/B:P} \right) + \sum_{t \in \mathcal{S}} \left[Q_t \left(\frac{x_t/V_t}{P_{t,p}/B:P} \right) \right] \\ & - MET_{3A4} - MET_{3A5} \end{aligned} \quad (11)$$

Subscript *liv* stands for liver, \mathcal{S} represents the splanchnic organs (spleen, pancreas, stomach, small and large intestine). $c_{u,liv}$ is the unbound liver concentration. MET_{3A4} and MET_{3A5} are the fluxes representing the reactions catalysed by CYP3A4 and CYP3A5. All the chemical reactions are described using *Michaelis–Menten* equations [39]. The *Michaelis–Menten* parameters for MDZ are taken from in vitro studies [16] and they are scaled to the in vivo context as per [46]. One of the main parameters used for the in vitro to in vivo scaling is the microsomal protein per gram of liver (*MPPGL*) (see supplementary materials for a detailed description of this process).

The population variability of physiological parameters such as the compartment volumes and blood flow was

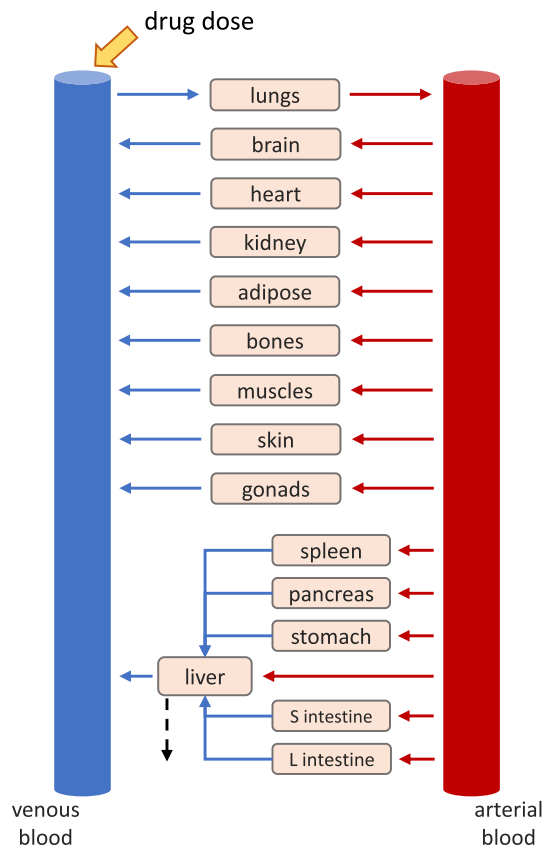


Fig. 2 Structure of a general whole-body PBPK model. Each box corresponds to a specific compartment. The red and blue arrows represent the arterial and venous blood flow, respectively. The black-dashed arrow represents elimination through metabolism in the liver. The yellow arrow represent the drug intravenous administration. *S intestine* and *L intestine* are the small and large intestine, respectively

generated with a simple algorithm having as inputs the sex of the subject, the height and the body mass index (BMI).

To simulate an IV bolus injection of 5 mg of MDZ, the initial condition of the venous blood compartment was set equal to 5, while the remaining compartments were set to equal 0. The *area under the curve* (AUC) of the venous plasma compartment from time 0 to $24 \cdot 7$ h was considered the output of interest for the GSA. The distributions of the model parameters considered in this analysis are reported in Table 2.

Analysis overview

For the GSA, the following methods were applied to both the algebraic and the PBPK models:

- classical variance-based GSA considering all the parameters uncorrelated;
- variance-based GSA grouping together the two correlated parameters;

- the method developed by Kucherenko for computing the variance-based GSA indices in presence of correlation [24];
- the latent variable approach.

Concerning the algebraic models, the analysis was carried out varying ρ_{14} , from -0.9 to 0.9 . When $\rho_{14} > 0$, the latent variable was considered to be positively correlated with both X_1 and X_4 ($\lambda > 0$). Instead, when $\rho_{14} < 0$, the latent variable was considered to be positively correlated with X_1 and negatively correlated with X_4 .

For the PBPK model, the (Pearson) correlation between the logarithms of CYP3A4 and CYP3A5 abundances $\rho_{3A4,3A5}$ was considered to equal 0.52, based on proteomic data from human liver samples [1], for the variance-based GSA with grouped factors, for the Kucherenko and the latent variable approaches. In this analysis, all simulated individuals were assumed to express CYP3A5.

All analysis was performed in MATLAB R2020a² [33]. The systems of differential equations were solved with the ode15s MATLAB solver, for a timespan ranging from 0 to $24 \cdot 7$ h. GSA was performed using the software UQLab [32] except for the variance-based GSA with groups, where an ‘ad hoc’ MATLAB code was developed. For the numerical estimation of the sensitivity indices, within UQLab, the *homma* estimator was used for the Sobol approach, while the default estimator embedded in the software was used for the Kucherenko approach. Concerning the ‘ad hoc’ MATLAB code, we used the estimator reported in [50] (the *errata corrige* version). For all the methods, the sample size was fixed to 10,000. The uncertainty of the sensitivity indices estimates was assessed by using 1000 bootstrap samples, with the exception of the Kucherenko method, where the convergence plots were used.

Results

Algebraic models

The GSA results for the algebraic models 1, 2 and 3, with $\rho_{14} = 0.7$ and $\rho_{14} = 0.9$, are reported in Tables 3, 4 and 5, respectively. In Fig. 3 the GSA results obtained with the latent variable and the Kucherenko approaches for the algebraic model 1 are given as a function of ρ_{14} , ranging from -0.9 to 0.9 . For the models 2 and 3, the equivalent information is shown in Figs. 4 and 5, respectively. Here we begin by reporting the results of model 1 and 2 and then, model 3.

² The codes are made available at the following link https://github.com/NicolaMelillo/latent_variable_GSA.

Table 2 Variable parameters used for the MDZ PBPK model

Parameters	Distribution parameters	Distribution type	Units	References
Sex ^d	0, 1			
height (male) ^e	176.7 (6.15)	Normal ^a	cm	[5]
Height (female) ^e	163.3 (5.85)	Normal ^a	cm	[5]
BMI ^f	18.5, 24.9	Uniform ^b	kg/m ²	[55]
[CYP3A4] ^g	137 (41%)	log-normal ^c	(pmol CYP)/(mg MP)	[8]
[CYP3A5] ^g	103 (65%)	log-normal ^c	(pmol CYP)/(mg MP)	[8]
MPPGL ^h	39.79 (27%)	log-normal ^c	(mg prot)/(g liver)	[30]

^aFor distribution parameters, *mean (standard deviation)* of the normal variable

^bFor distribution parameters, *minimum, maximum* of the parameter

^cFor distribution parameters, *mean (CV)* of the log-normal variable

^dSamples from an uniform distribution are extracted: if the extracted value is <0.5 the subject is female (0), otherwise male (1)

^eHeight for a 20 years old Italian population

^fBody mass index corresponding to the nutritional status of ‘Normal weight’ according to the World Health Organization

^gCYP abundance per mg of microsomal protein

^hmg of microsomal proteins for gram of liver

The parameter X_4 does not appear in Eqs. 7 and 8. Regardless of presence or absence of correlation between X_1 and X_4 its ‘causal’ impact on the output should therefore be null. Hence, intuitively, the results of a variance-based GSA with the classic Sobol’s method considering only X_1 , X_2 and X_3 should be the ones that truly represent the model structure. Any differences in main and total effects for the Kucherenko approach, the latent variable approach and the variance based GSA with grouped factors are therefore due to how these methods handle the correlation.

Concerning the Kucherenko approach, in Fig. 3 the higher the absolute value of ρ_{14} is, the higher the main effect of X_4 is, while its total effect always remains equal to 0. This substantially confirms the findings of [31], where it was highlighted that an input whose importance is due to the dependencies with other inputs has a total effect equal to 0, but a main effect that can be higher than 0. Moreover, as the absolute value of the correlation increases, the total effect of X_1 decreases, while the main effect remains stable. From [31] we know that S_1 includes the impact of the correlation of X_1 with X_4 , while $S_{T,1}$ just includes the ‘uncorrelated’ effects. From our example is possible to appreciate that the higher $|\rho_{14}|$ is, the lower the ‘uncorrelated’ effect of X_1 is. In this context it is actually challenging to distinguish between the ‘causal’ effect of X_1 and X_4 on Y and the effect due to their dependence. Similar conclusions can be made for the model 2. By limiting the analysis to the Kucherenko indices, it is challenging to understand how much X_1 is involved in interaction effects and, ultimately, to determine any ranking of importance of the parameters as can be used in practical applications.

Concerning the latent variable approach, presented in Figs. 3 and 4, the higher the absolute value of ρ_{14} is, the higher the importance of the latent variable over the unique variances. Ultimately, with ρ_{14} approaching 1 the whole variance of both X_1 and X_4 becomes fully explained by the latent factor and thus, the residual variances’ effect on the output variance tends to 0. Given that the latent variable affects both the correlated factors equally, it is not possible to elucidate if the impact of η on the output variance is primarily mediated by X_1 or X_4 . However, the impact of the unique variances can be uniquely attributed to the correlated factors. In fact, for both models 1 and 2, both the main and total effect of ε_4 are always equal to zero, as seen in Figs. 3 and 4. This is unlikely the case for traditional variance-based GSA with groups (see Tables 3 and 4), where, independently of the values of ρ_{14} , it is not possible to determine the impact of the variable within the groups. Notably, if $|\rho|$ is close to 1, the latent variable will fully explain both X_1 and X_4 , resembling the case of the grouping approach. Given that in both the grouping and the latent variable approach we are performing a standard Sobol’s GSA with uncorrelated factors, the interpretation of the sensitivity indices and the factor ranking is straightforward.

In model 1, X_1 is not involved in any interactions. This is discernible when $S_i = S_{T,i}$. In this case, $S_1 = S_{T,1}$, as seen in Table 3 and Fig. 3. Neither η or ε_1 are involved in any interactions. This is quite intuitive as the model is linear and X_1 is defined as the sum of the latent variable and the unique variance in the latent variable approach. However, interaction effects between the latent variable

Table 3 Sensitivity indices for the algebraic model 1

Factor	Sobol ^a		Kucherenko ^b		Latent variable ^a		Grouped ^a	
	Main	Total	Main	Total	Main	Total	Main	Total
			$\rho_{14} = 0.7$					
X_1^c	0.34 (0.32,0.36)	0.33 (0.31,0.34)	0.33	0.17	0.11 (0.09,0.13)	0.1 (0.9,0.11)	0.31 ^d (0.29,0.33)	0.34 ^d (0.31,0.37)
X_2	0.33 (0.31,0.35)	0.67 (0.64,0.7)	0.32	0.64	0.32 (0.3,0.35)	0.65 (0.63,0.67)	0.31 (0.28,0.33)	0.7 (0.67,0.72)
X_3	0 (- 0.03,0.02)	0.33 (0.31,0.35)	0	0.34	0.02 (- 0.01,0.04)	0.33 (0.31,0.35)	- 0.03 (- 0.06,0)	0.32 (0.29,0.34)
X_4^c	0 (- 0.02,0.02)	0 (0,0)	0.16	0	0.02 (0,0.03)	0 (0,0)		
η					0.26 (0.24,0.28)	0.23 (0.22,0.25)		
			$\rho_{14} = 0.9$					
X_1^c	0.33 (0.31,0.35)	0.35 (0.33,0.37)	0.33	0.06	0.05 (0.03,0.07)	0.04 (0.03,0.04)	0.33 ^d (0.31,0.35)	0.34 ^d (0.31,0.37)
X_2	0.32 (0.29,0.34)	0.66 (0.64,0.69)	0.33	0.69	0.33 (0.31,0.35)	0.65 (0.63,0.68)	0.35 (0.33,0.38)	0.67 (0.64,0.7)
X_3	- 0.01 (- 0.04,0.02)	0.33 (0.31,0.36)	- 0.01	0.35	0.02 (- 0.01,0.04)	0.35 (0.33,0.37)	0 (- 0.03,0.02)	0.33 (0.31,0.36) (- 0.03,0.02)
X_4^c	- 0.01 (- 0.03,0.01)	0 (0,0)	0.27	0	0.01 (- 0.01,0.03)	0 (0,0)		
η					0.3 (0.28,0.32)	0.29 (0.27,0.3)		

^aValues reported in the table are mean (2.5,97.5) percentiles calculated with 1000 bootstrap samples

^bConvergence plots are shown in the supplementary materials

^cFor the latent variable model, this refers to the unique variance

^dThis refers to the X_1 and X_4 group

and the unique variance will arise, for example, in case of X_1 having a nonlinear effect (e.g., quadratic) on Y^3 . In model 2, X_1 and X_3 show interaction effects, as noted in the Sobol's GSA results. This happens when $S_{T,i} > S_i$. In Table 4 and Fig. 4 we can see that both the latent variable and the unique variance of X_1 show interaction effects.

Concerning model 3, Table 5 and Fig. 5, we observe that the sensitivity indices of X_2 and X_3 change in function of ρ_{14} . The traditional variance-based GSA that considers all the factors uncorrelated does not capture this effect. With this simple example, we can see that ignoring the correlation within GSA could potentially bias the overall results of the analysis. Traditional GSA with groups can capture this effect and thus, it can be an easy and reliable method for treating correlations. However, as explained for

models 1 and 2, it has the limitation of not distinguishing the impact of the variables within the groups of correlated factors.

Concerning the Kucherenko approach, S_1 and S_4 are close to 0 when ρ_{14} is close to -1 and they both grow as $|\rho_{14}|$ grows. Instead, $S_{T,1}$ and $S_{T,4}$ have almost a parabolic shape. Both the main and total effects of X_1 and X_4 are low for strong negative correlation, probably because in this model the effect of X_1 tends to cancel the one of X_4 on Y and vice versa. For a high positive correlation the total effects tend to zero, while the main effects are close to 0.6.

Regarding the latent variable approach, one interesting observation is that the overall tendency of the unique variances and latent variable sensitivity indices are similar to those of the total and main effects of X_1 and X_4 of the Kucherenko approach, respectively. This probably happens because the unique variances represents the impact of the 'uncorrelated' part of the factors, similarly to the total effect of the Kucherenko approach. Instead, both the latent

³ If $X_1 = \lambda\eta + \varepsilon$ and $Y = X_1^2$, it is straightforward to derive that $Y = \lambda^2\eta^2 + \varepsilon^2 + 2\lambda\eta\varepsilon$. In this case, there are interaction effects between η and ε .

Table 4 Sensitivity indices for the algebraic model 2

Factor	Sobol ^a		Kucherenko ^b		Latent variable ^a		Grouped ^a	
	Main	Total	Main	Total	Main	Total	Main	Total
			$\rho_{14} = 0.7$					
X_1^c	0.34 (0.32,0.36)	0.68 (0.66,0.71)	0.32	0.34	0.11 (0.09,0.13)	0.2 (0.18,0.21)	0.33 ^d (0.3,0.35)	0.68 ^d (0.65,0.71)
X_2	0.32 (0.3,0.34)	0.33 (0.32,0.35)	0.33	0.32	0.33 (0.31,0.35)	0.33 (0.31,0.35)	0.32 (0.3,0.34)	0.34 (0.31,0.37)
X_3	0 (- 0.02,0.03)	0.34 (0.32,0.36)	0	0.34	0.01 (- 0.01,0.04)	0.33 (0.3,0.35)	- 0.03 (- 0.05, - 0.01)	0.33 (0.31,0.35)
X_4^c	- 0.01 (- 0.03,0.01)	0 (0,0)	0.16	0	0.01 (- 0.01,0.02)	0 (0,0)		
η					0.24 (0.22,0.26)	0.47 (0.45,0.49)		
			$\rho_{14} = 0.9$					
X_1^c	0.33 (0.31,0.35)	0.66 (0.63,0.69)	0.32	0.13	0.03 (0.01,0.05)	0.06 (0.05,0.07)	0.36 ^d (0.33,0.38)	0.65 ^d (0.62,0.68)
X_2	0.32 (0.3,0.34)	0.34 (0.32,0.35)	0.32	0.33	0.32 (0.3,0.34)	0.33 (0.32,0.35)	0.35 (0.33,0.37)	0.32 (0.29,0.35)
X_3	0 (- 0.03,0.03)	0.34 (0.32,0.37)	0	0.35	- 0.01 (- 0.03,0.01)	0.34 (0.32,0.37)	0.01 (- 0.01,0.04)	0.34 (0.32,0.37)
X_4^c	0 (- 0.02,0.02)	0 (0,0)	0.25	0	0 (- 0.02,0.02)	0 (0,0)		
η					0.29 (0.27,0.32)	0.61 (0.59,0.64)		

^aValues reported in the table are mean (2.5,97.5) percentiles calculated with 1000 bootstrap samples

^bConvergence plots are shown in the supplementary materials

^cFor the latent variable model, this refers to the unique variance

^dThis refers to the X_1 and X_4 group

variable and the main effect include the ‘dependent’ part of the factors. However, one important difference is that the latent variable approach is a variance-based GSA performed with independent variables and thus, the indices are easily understandable, this is unlikely the case for the Kucherenko approach. Finally, it is interesting to observe that for negative correlations the impact of the latent variable is zero. This happens because the factor loadings (λ) are equal in module, but opposite in sign and thus, the latent variable term is cancelled from Eq. 9.

Whole-body PBPK model for midazolam

The simulated MDZ plasma concentration-time profiles and AUCs for a population of 10,000 subjects are shown in Supplementary Figs. 8 and 9, respectively. The GSA results of Sobol’s method without accounting for the correlation, of the Kucherenko method, of the traditional

variance-based GSA with groups and of the latent variable approach are presented in Table 6.

According to the results from Sobol’s GSA, the most important parameters in explaining the variability in AUC are (in order of importance) the MPPGL, CYP3A4 and CYP3A5 abundances. These factors are important because they control the rate of metabolism in the liver. The fact that the metabolism-related parameters are the most important for explaining variability in AUC suggests that the rate-limiting step of drug elimination is the metabolism and not, for example, liver blood flow. Given that exposure drives drug effect, the interindividual variability in efficacy, due to PK, is mainly explained by genetics in this case example. However, we need to consider that our population is composed by healthy adults with a BMI corresponding to the nutritional status of ‘normal weight’ [55]. The inclusion of overweight or obese subjects may impact the results of the GSA.

Table 5 Sensitivity indices for the algebraic model 3

Factor	Sobol ^a		Kucherenko ^B		Latent variable ^a		Grouped ^a	
	Main	Total	Main	Total	Main	Total	Main	Total
$\rho_{14} = 0.7$								
X_1^c	0.25 (0.23,0.26)	0.25 (0.23,0.26)	0.55	0.1	0.07 (0.05,0.09)	0.05 (0.04,0.06)	0.62 ^d (0.6,0.64)	0.63 ^d (0.61,0.65)
X_2	0.25 (0.23,0.26)	0.24 (0.23,0.26)	0.19	0.19	0.18 (0.16,0.2)	0.19 (0.18,0.2)	0.19 (0.17,0.21)	0.18 (0.16,0.2)
X_3	0.26 (0.24,0.27)	0.25 (0.24,0.27)	0.18	0.19	0.2 (0.18,0.22)	0.19 (0.18,0.2)	0.18 (0.16,0.2)	0.19 (0.17,0.2)
X_4^c	0.26 (0.24,0.28)	0.25 (0.24,0.27)	0.54	0.1	0.06 (0.04,0.08)	0.05 (0.05,0.06)		
η					0.51 (0.49,0.53)	0.51 (0.49,0.53)		
$\rho_{14} = 0.9$								
X_1^c	0.24 (0.22,0.26)	0.25 (0.24,0.26)	0.63	0.03	0.02 (0,0.04)	0.02 (0.02,0.02)	0.65 ^d (0.63,0.67)	0.65 ^d (0.63,0.67)
X_2	0.24 (0.22,0.26)	0.25 (0.24,0.26)	0.17	0.17	0.18 (0.16,0.2)	0.17 (0.16,0.18)	0.17 (0.15,0.19)	0.17 (0.15,0.19)
X_3	0.26 (0.24,0.28)	0.24 (0.23,0.25)	0.18	0.17	0.17 (0.15,0.19)	0.17 (0.16,0.18)	0.18 (0.16,0.2)	0.19 (0.17,0.21)
X_4^c	0.25 (0.23,0.27)	0.26 (0.25,0.28)	0.62	0.03	0.02 (0,0.04)	0.02 (0.01,0.02)		
η					0.63 (0.61,0.64)	0.62 (0.6,0.64)		

^aValues reported in the table are mean (2.5,97.5) percentiles calculated with 1000 bootstrap samples

^bConvergence plots are shown in the supplementary materials

^cFor the latent variable model, this refers to the unique variance

^dThis refers to the X_1 and X_4 group

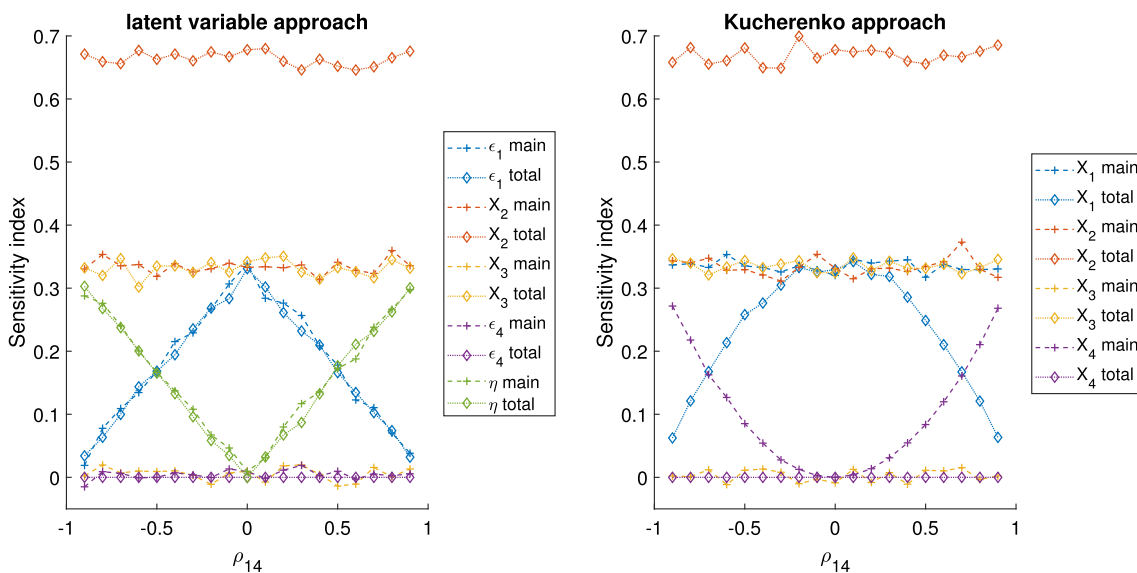


Fig. 3 Algebraic model 1 GSA results of the latent variable and the method presented by Kucherenko 2012 [24]

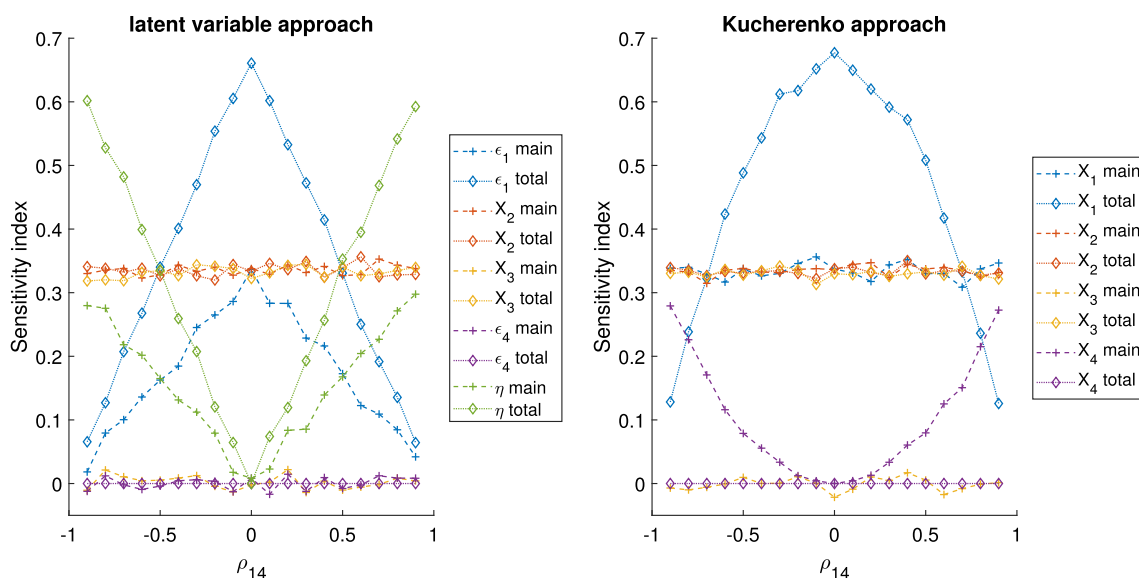


Fig. 4 Algebraic model 2 GSA results of the latent variable and the method presented by Kucherenko 2012 [24]

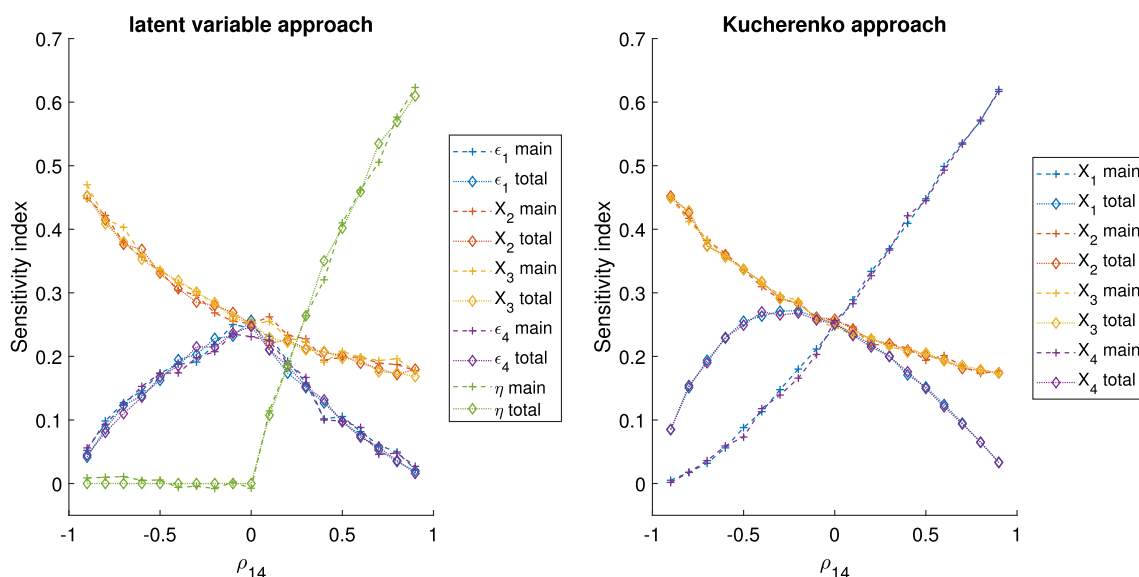


Fig. 5 Algebraic model 3 GSA results of the latent variable and the method presented by Kucherenko 2012 [24]

Concerning the GSA results obtained with the Kucherenko, the variance-based GSA with groups and the latent variable approach, the sensitivity indices of MPPGL are slightly reduced as compared to Sobol’s GSA. This is most likely related with the fact that the correlation between CYP3A4 and CYP3A5 tends to generate more ‘extreme’ individuals, i.e., poor metabolisers (with low CYP3A4 and low CYP3A5 abundances) and rapid metabolisers (with high CYP3A4 and high CYP3A5 abundances). Thus, as it is possible to observe in Supplementary Fig. 9, the AUC distribution in case of correlation is slightly wider with respect to the case of no correlation. These results are in agreement with our previous studies, where we showed

how a positive correlation between two enzymes metabolising a given compound can cause a widening of the systemic AUC distribution [37].

Concerning the Kucherenko analysis, it is difficult to confidently use either the main or the total effects for the purpose of factor ranking. For example, by observing the main effect the two most important parameters are CYP3A4 and CYP3A5 abundances. However, it is difficult to understand what the contributions of the variables themselves are and what is due to the correlation. For this reason, in our example, there is a risk of overestimating the importance of the enzymatic abundances and, by extension, underestimating the importance of the other factors. By

Table 6 Sensitivity indices for the MDZ PBPK model

Factor	Sobol ^a		Kucherenko ^b		Latent variable ^a		Grouped ^a	
	Main	Total	Main	Total	Main	Total	Main	Total
			$\rho_{3A4,3A5} = 0.52$					
sex	0 (− 0.02,0.02)	0.02 (0.01,0.03)	0.01	0.02	0.03 (0.01,0.05)	0.02 (0.01,0.02)	0 (− 0.02,0.02)	0.01 (− 0.03,0.04)
height	0.01 (− 0.01,0.03)	0.05 (0.04,0.05)	0.02	0.03	0.04 (0.02,0.06)	0.03 (0.02,0.04)	0.01 (− 0.01,0.03)	0.01 (− 0.02,0.05)
BMI	0.03 (0.01,0.05)	0.05 (0.04,0.06)	0.03	0.03	0.04 (0.02,0.07)	0.03 (0.02,0.05)	0.01 (− 0.01,0.03)	0.03 (− 0.01,0.06)
MPPGL	0.29 (0.27,0.31)	0.39 (0.37,0.41)	0.25	0.3	0.26 (0.24,0.29)	0.3 (0.27,0.32)	0.24 (0.22,0.27)	0.29 (0.26,0.32)
CYP3A4 ^c	0.27 (0.25,0.3)	0.33 (0.31,0.35)	0.49	0.22	0.12 (0.1,0.15)	0.15 (0.13,0.17)	0.61 ^d (0.58,0.64)	0.69 ^d (0.67,0.72)
CYP3A5 ^c	0.23 (0.2,0.25)	0.29 (0.27,0.31)	0.42	0.15	0.09 (0.07,0.09)	0.1 (0.09,0.12)		
η					0.43 (0.41,0.46)	0.48 (0.46–0.5)		

^aValues reported in the table are mean (2.5,97.5) percentiles calculated with 1000 bootstrap samples

^bConvergence plots are shown in the supplementary materials

^cFor the latent variable model, this refers to the unique variance

^dRefers to the group of CYP3A4 and CYP3A5

using the total effect for the factor ranking, there is instead the risk of underestimating the importance of the correlated factors and overestimating the importance of the remaining inputs, as the total effects for the factors involved in the correlation tend to 0 as $|\rho| \rightarrow 1$ [24]. Moreover, by using these two indices, given that for both CYP3A4 and CYP3A5 abundances the total effect is lower than the main effect, it is difficult to understand the effect of interactions.

In the latent variable approach, the factor ranking can be done by examining either the main or at the total effects. This is possible because the correlation between CYP3A4 and CYP3A5 was expressed in terms of a functional relationship between three independent factors, the latent variable and two independent variances. Thus, the classical variance-based GSA was used. With this approach, the most important factor in explaining the AUC is η , followed by MPPGL and the independent components of CYP3A4 and CYP3A5. By using either the main or the total effect for the factor ranking, we can confidently assess that the main drivers for the plasma AUC are the metabolism-related parameters. Moreover, with this method it is possible to appreciate the interaction effects, that in this case are mild and do not have a great impact on the factor ranking. A downside of this approach is that η drives both CYP3A4 and CYP3A5 variability. For this reason, given that the latent variable is one of the two most important parameters,

it is not possible to appreciate if its importance is primarily caused by the CYP3A4 or CYP3A5 mediated pathway. By investigating the independent components of CYP3A4 and CYP3A5 abundances, it is noted that they do have a similar impact. Intuitively, if one of the two factors was not important for the AUC, the independent component would be equal to zero (however, it is not necessarily true for the opposite case).

The results of the PBPK simulations presented here aim to illustrate a GSA methodology, only. Therefore, we do not recommend their use for other purposes.

Discussion

GSA is gaining use in modelling for pharmaceuticals, especially in the field of PBPK M&S. Recent applications in the literature [10, 26, 34, 36–38, 58, 60] and regulatory discussions [6, 7] have indicated the usefulness of these methods and it is likely that GSA will become an important feature of modelling in pharmaceutical R&D and for regulatory decision-making. This development is welcomed, indeed in the field of toxicology GSA is an important part of best practices for risk assessment of dose metric predictions [19, 34, 35, 41].

In order for GSA to gain wider use, the issues of usability and interpretation of the results need to be considered. PBPK M&S is an interdisciplinary effort highly reliant on experts in several domains, including medicinal chemistry, *in vitro* drug metabolism, pharmacokinetics, pharmacology, toxicology, statistical and mathematical modelling, and more. Further, modelling activities are an important tool for supporting a wide variety of decisions in R&D and regulatory submissions. For this reason, dedicated user-friendly software platforms are widely used [13], facilitating standardisation and easy access for non-expert users. We suspect that this is likely to hold true across many different domains, and therefore relevant across areas of application. In this context, particular attention in communicating GSA results should be paid.

Most whole-body PBPK models include several sets of correlated parameters, many of which constrain the models to realistic parameter combinations. It is therefore important that these correlations are accounted for when performing GSA. Several GSA methodologies have been proposed to account for dependent inputs [11, 24, 31, 52, 57] and the method developed by Kucherenko was proposed for PBPK models [27]. However, considerable debate is still ongoing amongst GSA practitioners on how to appropriately interpret the outcomes of these methods. We believe that the use of methodologies whose interpretation is still a matter of debate, require appropriate care in cases where GSA is called upon to support critical decisions, such as those relating to patient safety. The use of such methods may in fact lead to results that are uninterpretable, or, even worse, open to misinterpretation by non-expert GSA users. Certain applications of PBPK M&S require reliable, robust, well-characterised and tested models [45]. We believe that these requirements should apply for GSA methods and algorithms as well.

Here we propose a relatively simple method using a latent variable approach that deals with correlated input variables in variance-based GSA. The method expresses the correlation between two factors as causal relationships between a latent factor, η , and two unique variances. As a result this allows the use of classical Sobol's GSA with uncorrelated factors. In our opinion, the approach provides an intuitive process for implementation and interpretation as illustrated in the analysis for MDZ. By ranking the factors according to the total effects of Sobol's GSA, it was possible to clearly interpret the sensitivity indices. This allows insights into the model behaviour and to understand what the main drivers of variability are in a given output. By having a unique, easy and universally recognised interpretation of the sensitivity indices, it is possible to use GSA for supporting decision-making with increased confidence.

One of several alternatives to the latent variable approach would be the use of traditional variance-based GSA with groups. The main advantage is that this method allows treating more than two, or three, dependent factors and other dependencies than the linear correlations. However, as highlighted in the results section, with this approach is not possible to separately distinguish the impact of the dependent variables within a given group. Another alternative could be to assign causal dependencies between the correlated factors as we have done in a previous study in the context of PBPK models [37]. However, by doing so to describe the dependency, this will affect the relative significance of one input over the other. The potentially arbitrary choice of assigning dependency will increase the importance of the independent variable in the GSA and may produce misleading results. With the latent variable approach we renounce any attempt to completely distinguish the impact of the two correlated inputs on a given model output. Instead, we highlight the impact of the latent variable η (as the 'common cause') along with the independent part.

Here we also attempt to examine the shortcomings of the latent variable approach. In fact, the method presents some limitations with regards to the number and the distribution of the factors that are mutually correlated, as described in section 2. Moreover, the results of the latent variable approach need to be interpreted in light of the assumptions summarised in Table 1. In case one or more of these assumption are not satisfied (e.g., for bespoke PBPK platforms), the use of traditional GSA with groups is likely a better choice. Despite this, we believe that the latent variable approach can be of use. In conclusion, further research should be performed to find a reliable and interpretable method for handling multiple correlated inputs in GSA. This can be achieved, for example, by overcoming the current limitations of the latent variable approach to expand its use to more than two or three correlated input factors per latent variable. Alternatively, a clear and universally recognised interpretation should be agreed for more general GSA methods for dependent inputs, such as the approaches proposed by Kucherenko et al. [24] and Mara et al. [31].

Appendix: Average variance extracted

The general AVE expression, corresponding to one latent variable and k unique variances, is reported in Eq. 12 [15]. Considering that $\sigma_i^2 = 1 - \lambda_i^2$, AVE can be calculated as the average of the squares of the factor loadings associated with the latent variable [14].

$$AVE = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^k \lambda_i^2 + \sum_{i=1}^k \sigma_i^2} = \frac{1}{k} \sum_{i=1}^k \lambda_i^2 \quad (12)$$

Considering that in our case $k = 2$ (two dependent factors) and $\lambda_1 \cdot \lambda_2 = \rho$, we can derive the expression in Eq. 13.

$$AVE = \frac{1}{2}(\lambda_1^2 + \lambda_2^2) = \frac{1}{2} \left(\lambda_1^2 + \frac{\rho^2}{\lambda_1^2} \right) \quad (13)$$

If we calculate the first derivative of AVE over λ_1 and set it equal to zero, we can obtain the following expression in Eq. 14.

$$\begin{aligned} \frac{dAVE}{d\lambda_1} &= \lambda_1 - \frac{\rho^2}{\lambda_1^3} = 0 \\ \lambda_1 &= \sqrt{|\rho_{12}|} \\ \lambda_2 &= \text{sign}(\rho_{12}) \cdot \sqrt{|\rho_{12}|} \end{aligned} \quad (14)$$

Where, $\text{sign}(\rho_{12})$ is equal to +1 if $\rho_{12} > 0$, while it is equal to -1 if $\rho_{12} < 0$. If we calculate the second derivative we can see that it is always positive, thus $|\lambda_1| = |\lambda_2|$ corresponds to a minimum.

Supplementary Information The online version supplementary material available at <https://doi.org/10.1007/s10928-021-09764-x>.

Acknowledgements We would like to thank Professor Paolo Magni and Dr. Elena Maria Tosca of Università degli Studi di Pavia for valuable discussions and suggestions.

Funding Open access funding provided by Royal Institute of Technology.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Achour B, Russell MR, Barber J, Rostami-Hodjegan A (2014) Simultaneous Quantification of the Abundance of Several Cytochrome P450 and Uridine 5-Diphospho-Glucuronosyltransferase

Enzymes in Human Liver Microsomes Using Multiplexed Targeted Proteomics. *Drug Metab Dispos* 42(4):500–510

2. Berezhkovskiy LM (2010) A valid equation for the well-stirred perfusion limited physiologically based pharmacokinetic model that consistently accounts for the blood-tissue drug distribution in the organ and the corresponding valid equation for the steady state volume of distribution. *J Pharm Sci* 99(1):475–485. <https://doi.org/10.1002/jps.21798>
3. Borgonovo E, Plischke E (2016) Sensitivity analysis: a review of recent advances. *Eur J Oper Res* 248(3):869–887. <https://doi.org/10.1016/j.ejor.2015.06.032>
4. Brown TA (2015) *Confirmatory factor analysis for applied research*, 2nd edn. Guilford Press, New York
5. Cacciari E, Milani S, Balsamo A, Spada E, Bona G, Cavallo L, Cerutti F, Gargantini L, Greggio N, Tonini G, Cicognani A (2006) Italian cross-sectional growth charts for height, weight and BMI (2 to 20 yr). *J Endocrinol Invest* 29(7):581–593. <https://doi.org/10.1007/BF03344156>
6. CDER: Physiologically based pharmacokinetic analyses - format and content: Guidance for industry (2018). <https://www.fda.gov/files/drugs/published/Physiologically-Based-Pharmacokinetic-Analyses-%E2%80%94-Format-and-Content-Guidance-for-Industry.pdf>
7. CHMP (EMA): Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. Tech. Rep. EMA/CHMP/458101/2016, Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency (EMA), London, UK (2018). <https://www.ema.europa.eu/en/reporting-physiologically-based-pharmacokinetic-pbpb-modeling-simulation>
8. Cubitt HE, Yeo KR, Howgate EM, Rostami-Hodjegan A, Barter ZE (2011) Sources of interindividual variability in IVIVE of clearance: an investigation into the prediction of benzodiazepine clearance using a mechanistic population-based pharmacokinetic model. *Xenobiotica* 41(8):623–638
9. Da Veiga S, Wahl F, Gamboa F (2009) Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics* 51(4):452–463. <https://doi.org/10.1198/TECH.2009.08124>
10. Daga PR, Bolger MB, Haworth IS, Clark RD, Martin EJ (2018) Physiologically based pharmacokinetic modeling in lead optimization. 2. Rational bioavailability design by global sensitivity analysis to identify properties affecting bioavailability. *Mol Pharma* 15(3):831–839. <https://doi.org/10.1021/acs.molpharma.7b00973>
11. Do NC, Razavi S (2020) Correlation effects? A major but often neglected component in sensitivity and uncertainty analysis. *Water Resour Res* 56(3):e2019WR025436
12. Doki K, Darwich AS, Achour B, Tornio A, Backman JT, Rostami-Hodjegan A (2018) Implications of intercorrelation between hepatic cyp3a4-cyp2c8 enzymes for the evaluation of drug–drug interactions: a case study with repaglinide. *Br J Clin Pharmacol* 84(5):972–986
13. El-Khateeb E, Burkhill S, Murby S, Amirat H, Rostami-Hodjegan A, Ahmad A (2020) Physiological-based pharmacokinetic modeling trends in pharmaceutical drug development over the last 20-years; in-depth analysis of applications, organizations, and platforms. *Biopharm Drug Dispos* 42(4):107–117. <https://doi.org/10.1002/bdd.2257>
14. Farrell AM (2010) Insufficient discriminant validity: a comment on Bove, Pervan, Beatty, and Shiu (2009). *J Bus Res* 63(3):324–327
15. Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50

16. Galetin A, Brown C, Hallifax D, Ito K, Houston JB (2004) Utility of recombinant enzyme kinetics in prediction of human clearance: impact of variability, cyp3a5, and cyp2c19 on cyp3a4 probe substrates. *Drug Metab Dispos* 32(12):1411–1420. <https://doi.org/10.1124/dmd.104.000844>
17. Garcia-Cremades M, Melillo N, Troconiz IF, Magni P (2020) Mechanistic multiscale pharmacokinetic model for the anticancer drug 2,2-difluorodeoxycytidine (gemcitabine) in pancreatic cancer. *Clin Transl Sci* 13(3):608–617. <https://doi.org/10.1111/cts.12747>
18. Homma T, Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliabil Eng Syst Saf* 52(1):1–17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6)
19. International Programme on Chemical Safety (IPCS): Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment. Harmonization Project Document 9, World Health Organization (WHO) (2010). https://www.who.int/ipcs/methods/harmonization/areas/pbpbk_models.pdf
20. Iooss B, Lemaitre P (2015) A Review on Global Sensitivity Analysis Methods. In: *Uncertainty Management in Simulation-Optimization of Complex Systems, Operations Research/Computer Science Interfaces Series*, pp. 101–122. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7547-8_5
21. Iooss B, Prieur C (2019) Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol indices, numerical estimation and applications. *International Journal for Uncertainty Quantification* 9(5). <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019028372>. <http://www.dl.begellhouse.com/journals/52034eb04b657aea,23ab8f375b210514,706e4963504bc249.html>. Publisher: Begel House Inc
22. Jamei M (2016) Recent advances in development and application of physiologically-based pharmacokinetic (pbpk) models: a transition from academic curiosity to regulatory acceptance. *Curr Pharmacol Rep* 2:161–169
23. Kostewicz ES, Aarons L, Bergstrand M, Bolger MB, Galetin A, Hatley O, Jamei M, Lloyd R, Pepin X, Rostami-Hodjegan A, Sjogren E, Tannergren C, Turner DB, Wagner C, Weitschies W, Dressman J (2014) Pbpk models for the prediction of in vivo performance of oral dosage forms. *Eur J Pharm Sci* 57:300–21. <https://doi.org/10.1016/j.ejps.2013.09.008>
24. Kucherenko S, Tarantola S, Annoni P (2012) Estimation of global sensitivity indices for models with dependent variables. *Comput Phys Commun* 183(4):937–946. <https://doi.org/10.1016/j.cpc.2011.12.020>
25. Li G, Rabitz H, Yelvington PE, Oluwole OO, Bacon F, Kolb CE, Schoendorf J (2010) Global sensitivity analysis for systems with independent and/or correlated inputs. *J Phys Chem A* 114(19):6022–6032. <https://doi.org/10.1021/jp9096919>
26. Liu D, Li L, Rostami-Hodjegan A, Bois FY, Jamei M (2020) Considerations and caveats when applying global sensitivity analysis methods to physiologically based pharmacokinetic models. *AAPS J* 22(5):93. <https://doi.org/10.1208/s12248-020-00480-x>
27. Liu D, Li L, Rostami-Hodjegan A, Jamei M (2019) Investigating Impacts of Model Parameters Correlations in Global Sensitivity Analysis: Determining the most influential parameters of a Minimal PBPK Model of Midazolam. In: *PAGE 28, Abstr 8875* [www.page-meeting.org/?abstract=8875]. Stockholm, Sweden
28. Loehlin JC, Beaujean AA (2017) *Latent variable models: an introduction to factor, path, and structural equation analysis*, 5th edn. Routledge, New York
29. Lolodi O, Wang YM, Wright WC, Chen T (2017) Differential regulation of cyp3a4 and cyp3a5 and its implication in drug discovery. *Curr Drug Metab* 18(12):1095–1105. <https://doi.org/10.2174/1389200218666170531112038>
30. L.P. C (2017) Simcyp Simulator - Version 17 (2017). <https://www.certara.com/software/physiologically-based-pharmacokinetic-modeling-and-simulation/simcyp-simulator/?ap%5B0%5D=PBPk>
31. Mara TA, Tarantola S, Annoni P (2015) Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environ Modell Softw* 72:173–183. <https://doi.org/10.1016/j.envsoft.2015.07.010>
32. Marelli S, Sudret B (2014) UQLab: a framework for uncertainty quantification in MATLAB. In: *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)*, Liverpool, United Kingdom
33. MathWorks T (2019) MATLAB R2019b, the Mathworks, inc. Natick, Massachusetts
34. McNally K, Cotton R, Loizou GD (2011) A workflow for global sensitivity analysis of PBPK models. *Front Pharmacol* 2:31. <https://doi.org/10.3389/fphar.2011.00031>
35. Meek MEB, Barton HA, Bessems JG, Lipscomb JC, Krishnan K (2013) Case study illustrating the WHO IPCS guidance on characterization and application of physiologically based pharmacokinetic models in risk assessment. *Regul Toxicol Pharmacol* 66(1):116–129. <https://doi.org/10.1016/j.yrtph.2013.03.005>
36. Melillo N, Aarons L, Magni P, Darwich AS (2019) Variance based global sensitivity analysis of physiologically based pharmacokinetic absorption models for BCS i–iv drugs. *J Pharmacokinetic Pharmacodyn* 46(1):27–42. <https://doi.org/10.1007/s10928-018-9615-8>
37. Melillo N, Darwich AS, Magni P, Rostami-Hodjegan A (2019) Accounting for inter-correlation between enzyme abundance: a simulation study to assess implications on global sensitivity analysis within physiologically-based pharmacokinetics. *J Pharmacokinetic Pharmacodyn* 46(2):137–154. <https://doi.org/10.1007/s10928-019-09627-6>
38. Melillo N, Grandoni S, Cesari N, Brogin G, Puccini P, Magni P (2020) Inter-compound and intra-compound global sensitivity analysis of a physiological model for pulmonary absorption of inhaled compounds. *AAPS J* 22(5):116. <https://doi.org/10.1208/s12248-020-00499-0>
39. Michaelis L, Menten ML (1913) Die Kinetik der Invertinwirkung. *Biochemistry Zeitung* 49:333–369
40. Oakley JE, O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J R Stat Soc B* 66(3):751–769
41. OECD: Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes. Tech. Rep. 331, Environment, Health and Safety, Environment Directorate, OECD (2021)
42. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T (2016) Sensitivity analysis of environmental models: a systematic review with practical workflow. *Environ Model Softw* 79:214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
43. Razavi S, Jakeman A, Saltelli A, Prieur C, Iooss B, Borgonovo E, Plischke E, Lo Piano S, Iwanaga T, Becker W, Tarantola S, Guillaume JHA, Jakeman J, Gupta H, Melillo N, Rabitti G, Chabridon V, Duan Q, Sun X, Smith S, Sheikholeslami R, Hosseini N, Asadzadeh M, Puy A, Kucherenko S, Maier HR (2021) The future of sensitivity analysis: an essential discipline for systems modeling and policy support. *Environ Model Softw* 137:104954. <https://doi.org/10.1016/j.envsoft.2020.104954>
44. Rostami-Hodjegan A (2018) Reverse translation in PBPK and QSP: going backwards in order to go forward with confidence. *Clin Pharmacol Therap* 103(2):224–232
45. Rostami-Hodjegan A, Bois FY (2021) Opening a debate on open-source modelling tools: pouring fuel on fire vs. extinguishing the flare of a healthy debate. *CPT: Pharmacometr Syst Pharmacol*. <https://doi.org/10.1002/psp4.12615>

46. Rostami-Hodjegan A, Tucker GT (2007) Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nat Rev Drug Discov* 6(2):140–148
47. Rowland A, van Dyk M, Hopkins AM, Mounzer R, Polasek TM, Rostami-Hodjegan A, Sorich MJ (2018) Physiologically based pharmacokinetic modeling to identify physiological and molecular characteristics driving variability in drug exposure. *Clin Pharmacol Ther* 104(6):1219–1228
48. Roy JN, Lajoie J, Zijenah LS, Barama A, Poirier C, Ward BJ, Roger M (2005) Cyp3a5 genetic polymorphisms in different ethnic populations. *Drug Metab Dispos* 33(7):884–7. <https://doi.org/10.1124/dmd.105.003822>
49. Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. *Comput Phys Commun* 145(2):280–297. [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1)
50. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) *Global sensitivity analysis. The primer*. Wiley, Hoboken
51. Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. *Math Modell Comput Exp* 1(4):407–414
52. Tarantola S, Mara TA (2017) Variance-based sensitivity indices of computer models with dependent inputs: the Fourier amplitude sensitivity test. *International Journal for Uncertainty Quantification* 7(6). <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2017020291>. <http://www.dl.begellhouse.com/journals/52034eb04b657aea.25688f033da19d10.6769c5736b9bbb65.html>. Publisher: Begel House Inc
53. Tsamandouras N, Wendling T, Rostami-Hodjegan A, Galetin A, Aarons L (2015) Incorporation of stochastic variability in mechanistic population pharmacokinetic models: handling the physiological constraints using normal transformations. *J Pharmacokinetic Pharmacodyn* 42(4):349–373. <https://doi.org/10.1007/s10928-015-9418-0>
54. Vossen M, Sevestre M, Niederalt C, Jang IJ, Willmann S, Edgington AN (2007) Dynamically simulating the interaction of midazolam and the CYP3A4 inhibitor itraconazole using individual coupled whole-body physiologically-based pharmacokinetic (WB-PBPK) models. *Theoret Biol Med Model* 4(1):13. <https://doi.org/10.1186/1742-4682-4-13>
55. (WHO), W.H.O.: Body Mass Index - BMI, (accessed September 28, 2020) (2020). <https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>
56. Workgroup EM, Marshall S, Burghaus R, Cosson V, Cheung S, Chenel M, DellaPasqua O, Frey N, Hamrén B, Harnisch L, Ivanow F, Kerbusch T, Lippert J, Milligan P, Rohou S, Staab A, Steimer J, Tornøe C, Visser S (2016) Good practices in model-informed drug discovery and development: practice, application, and documentation. *CPT: Pharmacometr Syst Pharmacol* 5(3):93–122
57. Xu C, Gertner G (2007) Extending a global sensitivity analysis technique to models with correlated parameters. *Comput Stat Data Anal* 51(12):5579–5590. <https://doi.org/10.1016/j.csda.2007.04.003>
58. Yau E, Olivares-Morales A, Gertz M, Parrott N, Darwich AS, Aarons L, Ogungbenro K (2020) Global sensitivity analysis of the Rodgers and Rowland model for prediction of tissue: plasma partitioning coefficients: Assessment of the key physiological and physicochemical factors that determine small-molecule tissue distribution. *AAPS J* 22(2):41. <https://doi.org/10.1208/s12248-020-0418-7>
59. Zhang X, Yang Y, Grimstein M, Fan J, Grillo JA, Huang SM, Zhu H, Wang Y (2020) Application of PBPK modeling and simulation for regulatory decision making and its impact on US prescribing information: an update on the 2018–2019 Submissions to the US FDAs Office of Clinical Pharmacology. *J Clin Pharmacol* 60(S1):S160–S178
60. Zhang XY, Trame M, Lesko L, Schmidt S (2015) Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. *CPT Pharmacometr Syst Pharmacol* 4(2):69–79. <https://doi.org/10.1002/psp4.6>

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