

ORIGINAL RESEARCH

Open Access

# Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification

Markus Krauss<sup>1,2</sup>, Rolf Burghaus<sup>3</sup>, Jörg Lippert<sup>3</sup>, Mikko Niemi<sup>4,5</sup>, Pertti Neuvonen<sup>5</sup>, Andreas Schuppert<sup>1,2</sup>, Stefan Willmann<sup>1</sup>, Lars Kuepfer<sup>1</sup> and Linus Görlitz<sup>1\*</sup>

## Abstract

**Purpose:** Inter-individual variability in clinical endpoints and occurrence of potentially severe adverse effects represent an enormous challenge in drug development at all phases of (pre-)clinical research. To ensure patient safety it is important to identify adverse events or critical subgroups within the population as early as possible. Hence, a comprehensive understanding of the processes governing pharmacokinetics and pharmacodynamics is of utmost importance. In this paper we combine Bayesian statistics with detailed mechanistic physiologically-based pharmacokinetic (PBPK) models. On the example of pravastatin we demonstrate that this combination provides a powerful tool to investigate inter-individual variability in groups of patients and to identify clinically relevant homogenous subgroups in an unsupervised approach. Since PBPK models allow the identification of physiological, drug-specific and genotype-specific knowledge separately, our approach supports knowledge-based extrapolation to other drugs or populations.

**Methods:** PBPK models are based on generic distribution models and extensive collections of physiological parameters and allow a mechanistic investigation of drug distribution and drug action. To systematically account for parameter variability within patient populations, a Bayesian-PBPK approach is developed rigorously quantifying the probability of a parameter given the amount of information contained in the measured data. Since these parameter distributions are high-dimensional, a Markov chain Monte Carlo algorithm is used, where the physiological and drug-specific parameters are considered in separate blocks.

**Results:** Considering pravastatin pharmacokinetics as an application example, Bayesian-PBPK is used to investigate inter-individual variability in a cohort of 10 patients. Correlation analyses infer structural information about the PBPK model. Moreover, homogeneous subpopulations are identified *a posteriori* by examining the parameter distributions, which can even be assigned to a polymorphism in the hepatic organ anion transporter OATP1B1.

**Conclusions:** The presented Bayesian-PBPK approach systematically characterizes inter-individual variability within a population by updating prior knowledge about physiological parameters with new experimental data. Moreover, clinically relevant homogeneous subpopulations can be mechanistically identified. The large scale PBPK model separates physiological and drug-specific knowledge which allows, in combination with Bayesian approaches, the iterative assessment of specific populations by integrating information from several drugs.

**Keywords:** Physiologically-based pharmacokinetic modeling, Bayesian approaches, Markov chain Monte Carlo, Inter-individual variability, OATP1B1, Drug development, Pravastatin

\* Correspondence: Linus.Goerlitz@Bayer.com

<sup>1</sup>Bayer Technology Services GmbH, Computational Systems Biology, Leverkusen 51368, Germany

Full list of author information is available at the end of the article

## Background

Tailor-made therapeutic designs require a functional understanding of the processes governing the distribution of substances within an organism. Anthropometric parameters like age or weight have great influence on the level of drug exposure in the human body (Willmann et al., 2007). Furthermore, the genetic predisposition of a patient is very important, since different genotypes can have significant effects on drug metabolization processes (Eissing et al., 2012; Lippert et al., 2012). In the worst case, side effects due to increased (off-)target tissue drug concentrations become critical for patient safety (Lippert et al., 2012). The early identification of subgroups showing significantly increased adverse event rates is a difficult task since only limited information about a new drug is available but is of utmost importance to prevent costly drug withdrawals in later phases of the drug development process (Kuepfer et al., 2012). Therefore, a mechanistic understanding of pharmacokinetics (PK) is essential in drug development to optimize the risk-benefit profile of a drug. This involves in particular the identification of high-risk subgroups in which an unfortunate combination of predisposition and non-optimal dosing schemes lead to potentially life-threatening side effects. In clinical practice, such subgroups have to be treated with individualized dosing schemes, which need to be designed and surveyed with adequate diagnostics.

The amount and complexity of preclinical and clinical data generated along the drug development process usually represents an immense challenge for the generation of an in-depth mechanistic understanding. Here, *in silico* approaches provide a rational and efficient way to aggregate all data for the determination of drug PK and pharmacodynamics (PD) in support of the drug development process. Once established and validated, computational models allow a detailed analysis of the effect of different dosing schemes or varying anthropometry or physiology by simulating the behavior of a drug in the body. In contrast to the rather descriptive consideration of PK and PD in classical compartmental approaches (Meibohm & Derendorf, 1997), physiologically-based pharmacokinetic (PBPK) models are based on a large amount of prior physiological and anthropometric information which is integrated in the model structure (Nestorov, 2007; Rowland et al., 2011; Schmitt & Willmann, 2004). Since PBPK models explicitly distinguish between properties of the compound and properties of the patients, respectively, they allow separation of physiological and drug-induced effects. Generally, such models consist of several compartments, describing the organs, which are further subdivided in more detailed submodules such as interstitial, intracellular or vascular space. Starting from models with only few equations (Pang & Durk, 2010), they exist on all levels of complexity, up to more than one hundred

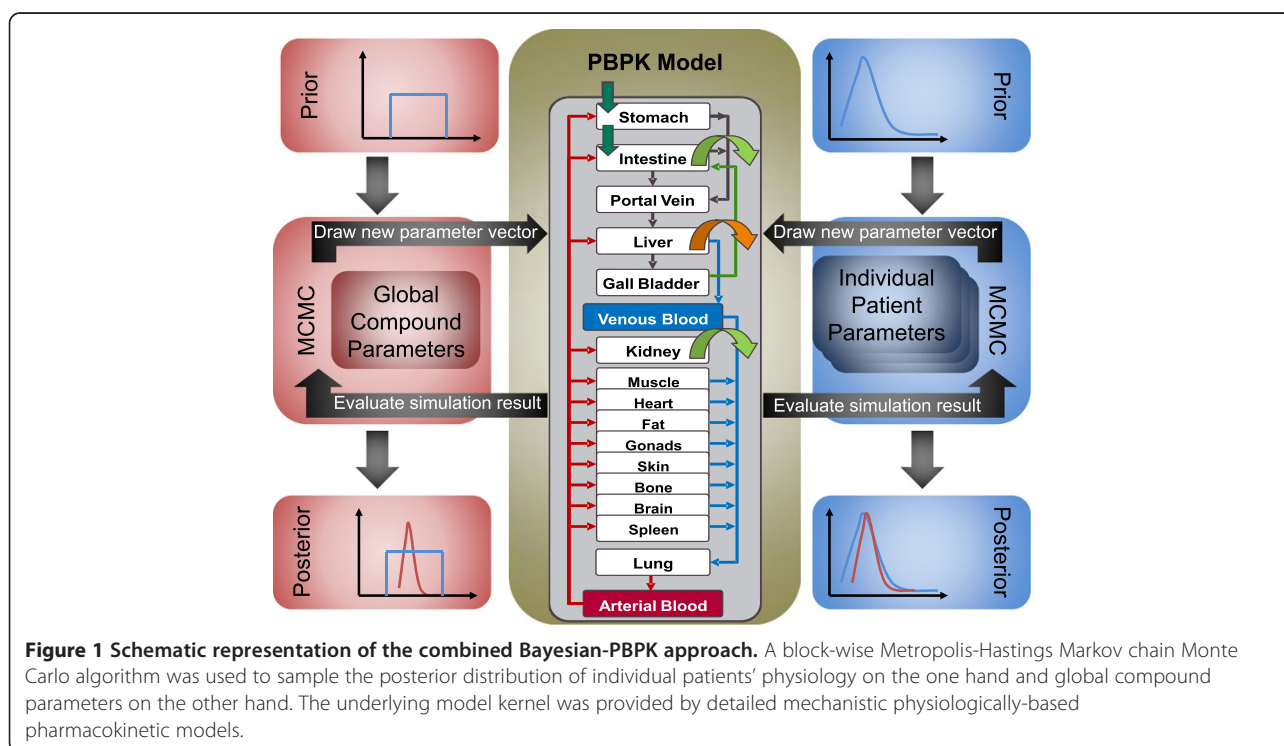
ordinary differential equations (ODEs) and hundreds of parameters (Eissing et al., 2011; Willmann et al., 2003a). PBPK models have previously been used for mechanistic analyses of drug PK (Meyer et al., 2012), pharmacogenomics (Eissing et al., 2012), multiscale modeling (Krauss et al., 2012) or analysis of rare adverse events (Lippert et al., 2012; Willmann et al., 2009). However, current use of such models often provides only a single value time-concentration curve, describing the behavior of a mean patient, neglecting potentially relevant individual properties. This is even more severe as PBPK models allow the creation of personalized models for individual patients by explicitly representing the individual physiological parameters. Thereby it is possible to mechanistically describe special populations (Edginton & Willmann, 2008) or genetic predisposition of patients in pharmacogenomics applications (Eissing et al., 2012; Swen et al., 2007). Nevertheless, PBPK models frequently lack the rigorous quantification of inter-individual variability in parameters which cannot be derived from the patients' anthropometry.

Up to now, population simulations try to assess inter-individual variability in PK in groups of patients (Schüttler & Ihmsen, 2000; Willmann et al., 2007). This is examined by *a priori* variation of physiological parameters and cross correlations to other model parameters. Such correlations are estimated by means of scaling laws depending on the anthropometry, since no literature information on inter-individual variability of organ weights or blood flows is available (Willmann et al., 2007). Therefore, since the physiology of every individual is calculated *before* the simulation, such population simulations cannot be processed if for example special groups of patients are investigated where little prior information about their anthropometry, (patho-)physiology or genotype-phenotype correlation is available.

An alternative approach to analyze the inter-individual variability and to perform population simulations is Bayesian modeling (Bolstadt, 2010). Bayesian statistics is based on Bayes' theorem, which provides a rational way to combine prior information on parameters with the information contained in data to infer the variability of parameters and therefore also the PK variability *a posteriori*, even with little prior information. The key idea of Bayesian statistics is to define unknown parameters as random variables, which is in contrast to the general approach in statistics, where parameters are defined as fixed, but unknown constants. In Bayes' theorem, prior knowledge about the parameters is updated with new experimental data in the so-called posterior distribution (Bolstadt, 2010). Determining the posterior distribution explicitly is very difficult or even impossible with nonlinear model kernels or when many parameters are considered simultaneously. In such cases, Markov-chain Monte-Carlo (MCMC) methods can be used to estimate the posterior distribution.

MCMC covers a large group of algorithms containing for example usual (Geman & Geman, 1984; Hastings, 1970; Metropolis, 1953), adaptive (Atchadé & Rosenthal, 2005; Gilks et al., 1998; Haario et al., 2001; Roberts & Rosenthal, 2009) and particle (Andrieu et al., 2010) MCMC approaches to take samples from the posterior distribution of a parameter vector. The core idea of MCMC is to sample the unknown variables along a Markov chain, which has the posterior distribution as its stationary distribution. If several parameters are considered, such probability distributions are high dimensional. Thorough analysis of this posterior distribution quantifies inter-individual variability of a group of patients as well as the co-variability of the parameters, allowing the identification of homogenous subgroups. Bayesian approaches have already been used in conjunction with PBPK modeling, especially in toxicological questions (Bernillon & Bois, 2000; Bois et al., 2010), but also for population PK (Gelman et al., 1996; Gueorguieva et al., 2006; Yang et al., 2009). However, often the PBPK models used have been comparatively small and have contained lumped parameters carrying mixed information of different physiological or drug specific parameters. By using a large scale PBPK model, which separates drug specific from population specific information, in combination with Bayesian approaches, an iterative characterization of special populations by optimally leveraging information from different drugs can be achieved.

In this work, we present a new approach applying MCMC to Bayesian-PBPK modeling for the assessment of inter-individual variability in groups of patients (Figure 1). Notably, we use a highly detailed and mechanistic PBPK model, where every organ is divided into four sub-compartments describing the intracellular space, the interstitial space, the blood plasma and the blood cells (Willmann et al., 2003a). Due to a segregated representation of the physiology of the patient and the underlying distribution model which is related to the physico-chemistry of the drug, physiological parameters, genotype-specific (in the following also referred to as physiological parameters) and drug-specific parameters, respectively, can be considered separately. This model representation together with Bayesian approaches allows direct inference of physiological and drug-specific information, such as the variability in organ volumes or the uncertainty in the lipophilicity of a drug. Therefore, the main sources of variability within the PK of a drug may be quantified by analyzing the posterior distribution. Moreover, we present a way to analyze the posterior to identify clinically relevant homogenous subgroups, such as patients with a specific genotype linking to a PK (or PD) phenotype. Additionally, the use of such a mechanistic model bears great extrapolation capacity, by enabling an iterative use of the posterior as the prior distribution of a new run. Since physiology and drug are treated independently, posterior physiological information of a run with a known drug can be



used for the investigation of a new drug candidate in the same group of patients such that physiological knowledge is conserved. The same applies for using the same drug in different populations. This allows for example the construction of a large database, wherein prior information about physiological as well as drug-specific parameters may be updated with experimental data of lots of experiments. Since only little literature information about specific parameters is available, informative prior distributions can be 'learned' after several MCMC runs with different experimental data.

Taken together, besides the assessment of inter-individual variability and co-variability of physiological parameters, our presented approach additionally provides a very valuable tool for long-term characterization of special populations as well as drug physicochemistry.

## Methods

### Physiologically-based pharmacokinetic modeling

PBPK models quantitatively consider the absorption, distribution, metabolism and excretion (ADME) of exogenous and endogenous substances at a very high level of detail (Nestorov, 2007; Rowland et al., 2011; Schmitt & Willmann, 2004; Willmann et al., 2003a). They mechanistically describe all relevant processes based on a large amount of prior physiological information. The models consist of compartmental representations of all relevant organs, tissues and the vascular system. The underlying model structure which is based on generic distribution models quantifies the mass transfer between the vascular system and the organs (Poulin et al., 2001; Rodgers et al., 2005; Rodgers & Rowland, 2006; Willmann et al., 2003b; Willmann et al., 2004). Parameters in the PBPK model can be divided into two types of parameters: (1) physiological parameters such as organ volumes or blood flow rates which are obtained from large collections of physiological data integrated into the PBPK software database and (2) substance-specific parameters describing the physicochemistry of a compound such as the molecular weight or the lipophilicity. Moreover, the large amount of prior physiological information constraints the number of independent parameters in the PBPK model which need to be identified (usually less than ten). The PBPK model of the present work consists of more than one hundred ordinary differential equations containing hundreds of parameters. The clear separation of physiology and drug-specific parameters due to the mechanism-based approach and the size of the model also allow the separated inference in parameter-identification processes.

The pravastatin model considered in this work was built with the software tools PK-Sim and MoBi. Academic licenses for both tools are available free of charge and both PK-Sim and MoBi have been explained in detail before (Eissing et al., 2011; Willmann et al., 2003a). The

anthropometric information of the patients regarding age, weight and height further specifies the selection of physiological parameters as provided in the software, which allows a specific parameterization of the PBPK model.

### Bayesian approach in combination with PBPK modeling

PBPK models can be parameterized for individuals with defined anthropometries such as age, sex, weight or height. Nevertheless, every model represents a mean value model, assuming that a group of individuals with the same anthropometry also has the same parameterization. However, even in-between defined groups of patients, parameters such as organ volumes or blood flow rates, can show substantial variation from individual to individual. Additionally, the determination of substance-specific parameters often contains uncertainties since such parameters are determined *ex vivo*. Thus, substance-specific parameters also vary, but in contrast to the individual parameters, their value is the same for all patients. We call this type of parameter global parameters, in contrast to the parameters we call individual parameters, which need to be randomized separately in every patient.

Both, variability and uncertainty in individual parameters  $\theta^I$  and uncertainty in global parameters  $\theta^G$  influence the PK of endogenous and exogenous compounds in an organism. Uncertainty can be reduced for example by increasing the number of experiments or by an optimized experimental design. In contrast, inter-individual variability is a characteristic property and cannot be reduced (Bernillon & Bois, 2000).

But how to determine such uncertainty and variability in a group of patients? Classical parameter identification relies on optimization-based approaches. They determine only *the* parameter vector with the highest probability to fit to the data, called maximum likelihood estimator. In contrast, Bayesian approaches aim for the identification of a probability distribution of the parameter vector. Furthermore, they also consider prior knowledge about the parameters, and 'update' this prior knowledge by integration of new experimental data based on Bayes' theorem given by

$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)} \quad (1)$$

The posterior distribution  $p(\theta|D)$  combines prior knowledge  $p(\theta)$  about the parameter vector  $\theta$  ( $\theta \in \mathbb{R}^{P \times 1}$ ) with the likelihood function  $p(D|\theta)$ . The likelihood function represents the closeness to the data  $D$ , where  $D = \{x_{i,k}, t_{i,k}\}$  with  $i=1, \dots, N_k$ .  $x_{i,k}$  represents the measurement of individual  $k$  ( $k=1, \dots, K$ ) at time point  $t_{i,k}$ . The idea of Bayesian approaches is to model any unknown parameter as random variable, since the true value of the parameter is unknown. However, especially in high

dimensional problems, the numerical determination of the posterior is almost impossible. Therefore, several methods have been developed which draw a sample from the posterior distribution to estimate its probability density. Markov chain Monte Carlo (MCMC) approaches are a large group of sampling algorithms which can be divided to a great extent into two groups: Metropolis-Hastings (MH) algorithms (Hastings, 1970; Metropolis, 1953) and the Gibbs sampler (Gelfand & Smith, 1990; Geman & Geman, 1984). In contrast to classical Monte Carlo sampling, MCMC samples from a special Markov chain which, in our case, is constructed to have the posterior distribution as its long-run stationary distribution (Andrieu et al., 2003).

For our combined Bayesian-PBPK approach, we considered a block-wise MH algorithm to sample from the posterior distribution. In contrast to a single MH block containing all parameters, dividing the parameter space in blocks improves the convergence speed of the Markov chain. Thus, one MH step was applied to every block of parameters, conditional on knowing the other parameter values which were not in this block. We considered  $K+1$  main blocks, one for every individual and one containing the global parameters. One MH step was performed as follows:

- i. Let  $\theta_k^l(n)$  be the parameter vector of the individual parameters of individual  $k$  after  $n$  steps. Propose a new parameter vector  $\theta_k^l$  by random sampling from proposal density  $Q(\theta_k^l(n), \cdot)$ .
- ii. Generate  $u \in [0, 1]$  uniformly distributed. Examine if

$$u \leq \frac{p(\theta_k^l | D, \theta^G(n)) \cdot Q(\theta_k^l(n), \theta_k^l)}{p(\theta_k^l(n) | D, \theta^G(n)) \cdot Q(\theta_k^l, \theta_k^l(n))} \quad (2)$$

- iii. If true:  $\theta_k^l(n+1) = \theta_k^l$ ; else:  $\theta_k^l(n+1) = \theta_k^l(n)$ ;  $\Rightarrow \theta_k(n+1) = [\theta_k^l(n+1), \theta^G(n)]$ .

In another MH block, only the global parameters  $\theta^G$  were sampled and the individual parameters  $\theta_k^l(n)$  were fixed to the value of the last individual MH step:

- i. Let  $\theta^G(n)$  be the parameter vector of the global parameters after  $n$  steps. Propose a new parameter vector  $\theta^G$  by random sampling from proposal density  $Q(\theta^G, \cdot)$ .
- ii. Generate  $u \in [0, 1]$  uniformly distributed. Examine if

$$u \leq \frac{\prod_k (p(\theta^G | D, \theta^l(n))) \cdot Q(\theta^G(n), \theta^G)}{\prod_k (p(\theta^G(n) | D, \theta^l(n))) \cdot Q(\theta^G, \theta^G(n))} \quad (3)$$

- iii. If true:  $\theta^G(n+1) = \theta^G$ ; else:  $\theta^G(n+1) = \theta^G(n)$ ;  $\Rightarrow \theta_k(n+1) = [\theta_k^l(n), \theta^G(n+1)]$ .

Notably, a truncated normal distribution centered around the current value  $\theta$  was chosen for the proposal density  $Q(\theta, \cdot)$ , since sampling was constrained by physiological constraints ( $\theta_{min}$   $\theta_{max}$ ) of each parameter.

For every individual  $k$ , informative prior distributions were defined as lognormal distributions for every of  $M$  parameters for which enough literature information about the population wide distribution was available (Gelman et al. 1996). Otherwise, flat non-informative prior distributions (for parameters  $P-M$ , remember that  $\theta \in \mathbb{R}^{P \times 1}$ ) were chosen:

$$p(\theta_k) = \prod_{m=1}^M p(\theta_{km}) \cdot \prod_{n=1}^{P-M} p(\theta_{kn}) \cdot p(\sigma_k) \quad (4)$$

$$p(\theta_{km}) \propto e^{-\frac{(\log \theta_{km} - \mu_m)^2}{\sigma_m^2}} \quad (5)$$

$$p(\theta_{kn}) = \begin{cases} \frac{1}{b_n - a_n} & \text{if } a_n \leq \theta_{kn} \leq b_n \\ 0 & \text{else} \end{cases} \quad (6)$$

$$p(\sigma_k) \propto \frac{1}{\sigma_k^2} \quad (7)$$

The uncertainties  $\sigma_k$  represented the measurement error of the experimental data  $D$ . Since their original values were unknown they were also considered as individual-specific random variables and have been assigned a prior distribution. Notably, the prior distributions were assumed to be independent from each other, since no information about co-variances was available at the beginning. However, after several runs with the same population as described in the introduction, prior distributions may be updated with the information about the co-variances between the parameters.

The likelihood function was defined with the help of a least squares error model. Additionally it was assumed that errors were distributed normal on a log-scale and were independent.

$$\log(p(D|\theta_k)) \propto -N \cdot \log \sigma_k^2 - \frac{1}{\sigma_k^2} \cdot \sum_i^N (\log(x_{i,k}) - \log(f^k(\theta_k, t_{i,k})))^2 \quad (8)$$

$f^k(\theta_k, t_{i,k})$  represents the time-resolved evaluation of the PBPK model with the respective parameter vector  $\theta_k$ .

### Results: Inter-individual variability in pravastatin pharmacokinetics

As an application example for the presented Bayesian-PBPK approach we here considered the PK of the 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase inhibitor pravastatin. This drug has been known for long and its genotype mediated inter-individual variability is well-characterized (Everett et al., 1991; Kivisto & Niemi,

2007; Serajuddin et al., 1991; Singhvi et al., 1990). Therefore, this case study is well suited to demonstrate the advantages of our approach, in particular the assessment of inter-individual variability as well as the identification of homogenous subgroups even with small amounts of experimental data.

Pravastatin is a HMG-CoA reductase inhibitor which lowers the cholesterol level within the body and thereby prevents cardiovascular diseases. Compared to other statins, it has a low lipophilicity (Serajuddin et al., 1991) such that pravastatin uptake is mainly distributed by active transporters (Kivisto & Niemi, 2007): On the one hand, the organic anion transporting polypeptide (OATP1B1) transports pravastatin into the intracellular space of the liver and on the other hand the organic anion transporter 3 (OAT3) inserts pravastatin in the intracellular space of the kidneys (Kivisto & Niemi, 2007). In the liver, pravastatin is excreted by biliary excretion, leading to enterohepatic circulation, while tubular secretion is the main pathway to excrete pravastatin from the kidneys (Hatanaka, 2000). Thereby, both routes of excretion are also performed by an active transporter, the multidrug resistance-associated protein 2 (MRP2) (Additional file 1: Figure S1). MRP2 is also significantly expressed in the apical membrane of enterocytes in the duodenum and jejunum. The bioavailability of pravastatin is low due to an incomplete absorption in the small-intestine (Kivisto & Niemi, 2007).

Notably, significant alterations in pravastatin PK are associated to three different genotypes (SNP; c.521T→C, p.Val174Ala) of SLCO1B1 encoding for OATP1B1 (Kivisto & Niemi, 2007; Niemi et al., 2006). This genotype determines the transporter activity (Lippert et al., 2012); the CC genotype has decreased activity compared to the normal TT genotype, which leads to higher pravastatin concentrations in the body. In contrast, no such effect is known for MRP2.

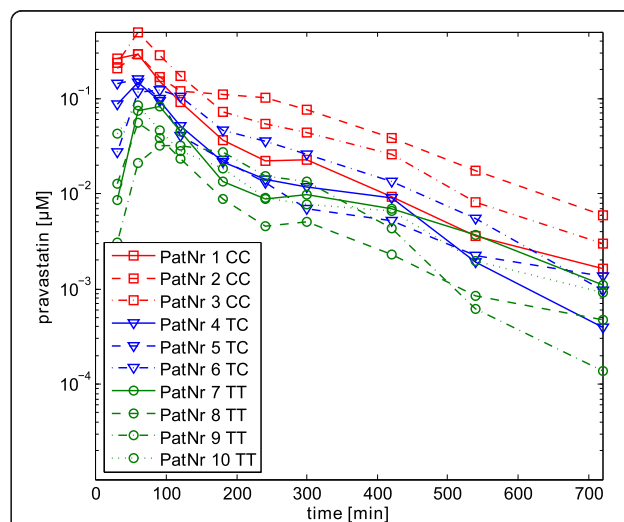
For our analyses we considered a previously established and validated PBPK model of pravastatin (Lippert et al., 2012) for an oral dose of 40 mg. In this model, active transport processes have been established in the interstitial (OATP1B1) and the intracellular space (MRP2) of the liver as well as in the interstitial space of the kidneys (OAT3). Additionally, MRP2 mediated transport was considered in the gastrointestinal compartment of our model as well as the intracellular space of the kidneys. Tissue specific enzyme activity was estimated by using gene expression data as a proxy for protein abundance. Notably, this allows the discrimination between organ-specific protein levels and the global catalytic rate constant  $k_{cat}$  (Meyer et al., 2012). A luminal clearance reaction in the small intestine accounted for the low bioavailability of pravastatin.

The experimental data was provided from previously published studies (Niemi et al., 2006). Out of the dataset of 32 patients, 10 patients have been chosen randomly

to lower computational costs (Figure 2). Nevertheless, the three genotypes of OATP1B1 are distributed equally in the chosen population. To describe the variability in all relevant ADME processes, 8 individual parameters together with 4 global parameters were chosen for the Bayesian analysis (Table 1), which means the variation of 84 parameters in total. During the separation of the parameters into different blocks, it is very important to know if parameters are correlated, since correlated parameters have to be sampled in one block (Smith et al., 1992). Our block structure is driven by the clear separation between substance and individual physiology in the PBPK model, therefore, we can assume that all parameters of different blocks are independent and uncorrelated (see also the discussion) and we can assure that no lumped parameters exist which depend on physiological and substance-specific information.

With the established PBPK model, the combined Bayesian-PBPK approach was processed and 300000 iteration steps were calculated. The computation time was 3.6 s/iteration and was performed on a quad core i5 processor running under Windows 7. Although the process is independent from the initial guesses, parameter start values have been estimated with the help of an identification process with a single patient to reduce convergence time (Additional file 2: Table S1).

During the first 150000 steps the parameter vectors have not been sampled from the correct distribution. For this so-called burn-in period the samples were discarded. By subsampling 200 parameter vectors of each patient from the remaining 150000 steps, an



**Figure 2** Experimental data of ten patients which were considered for the assessment of their inter-individual variability. The patients have been chosen out of a dataset of 32 patients provided by Niemi et al. (Niemi et al., 2006) such that all three possible genotypes of the hepatic uptake transporter OATP1B1 occurred equally.

**Table 1 Parameters to be varied in the coupled Bayesian-PBPK approach**

Parameter	Unit	Abbreviation	Type
Intestinal permeability	cm/min	$P_{int}$	Individual
Intestinal transit time	min	ITT	Individual
Gastric emptying time	min	GET	Individual
Luminal clearance factor	$\mu\text{M}/\text{min}$	$CL_{lum}$	Individual
$k_{cat}$ OATP1B1 (factor)	-	$k_{cat,O}$	Individual
$k_{cat}$ MRP2 (factor)	-	$k_{cat,M}$	Individual
Lag Time of enterohepatic circulation	min	EHC <sub>lagtime</sub>	Individual
Measurement error	-	sd	Individual
Lipophilicity (logP)	-	lip	Global
Unbound protein fraction	%	fu	Global
Km OATP1B1	$\mu\text{M}$	$Km_O$	Global
Km MRP2	$\mu\text{M}$	$Km_M$	Global

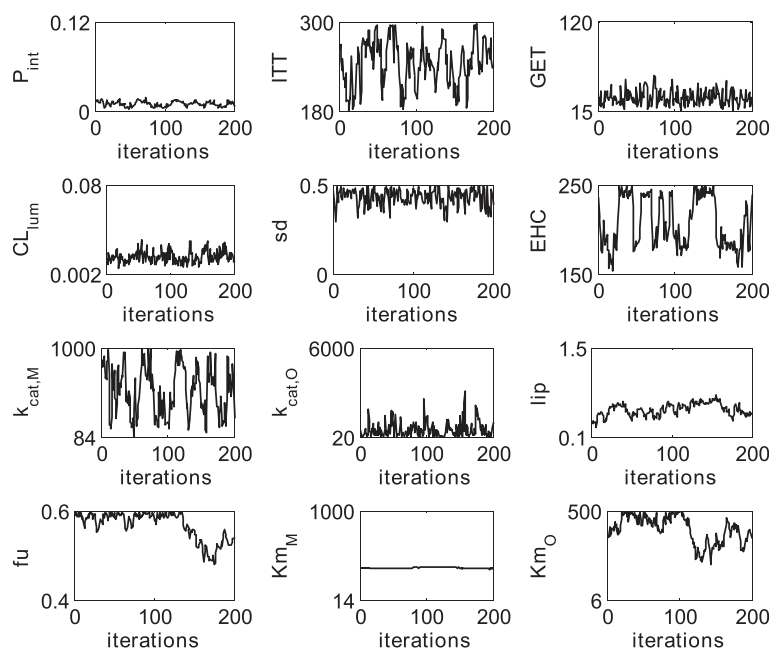
independent sample of the posterior distribution was drawn. The resulting traces are exemplarily shown for one patient (Figure 3). Notably, the four global parameters remain the same for every patient as described above, since they only depend on the physicochemistry of the drug.

Next, the PK which described the inter-individual variability of the whole population and the mean PK of the corresponding patients were simulated (Figure 4A). The inter-individual variability was estimated by calculating the 5–95% range of all patients. To demonstrate that the

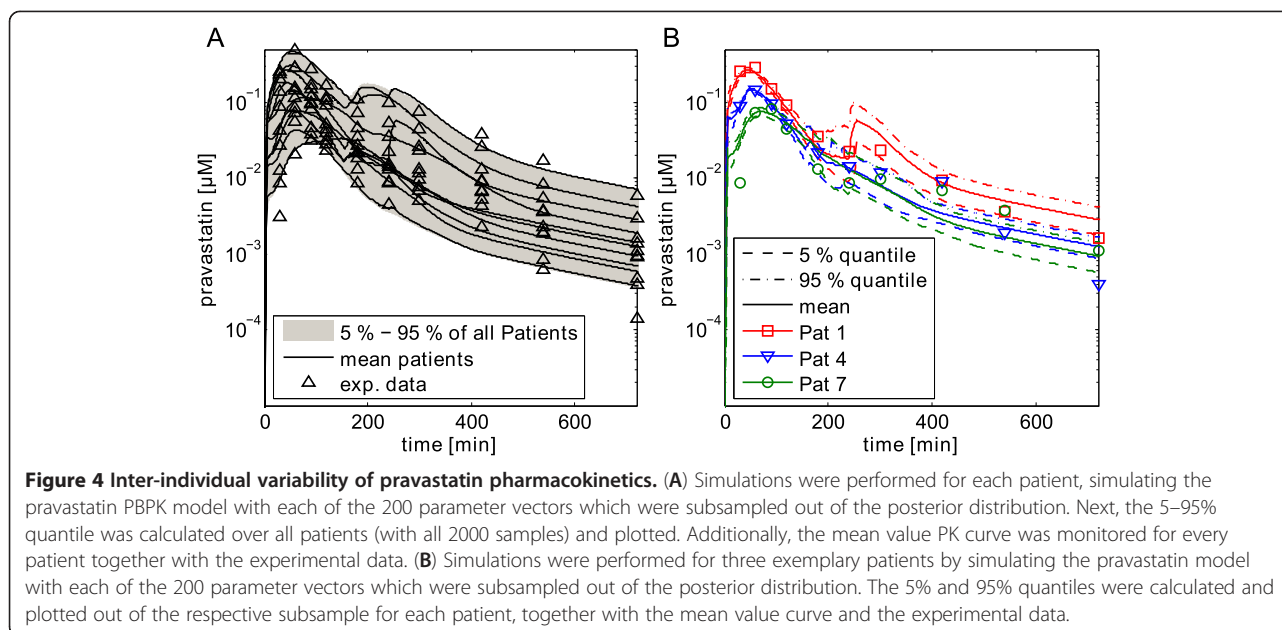
depicted inter-individual variability did not already result from large variability and uncertainty of the single patients, the 5% and 95% quantiles and mean values for three exemplary patients were illustrated (Figure 4B). Additionally, the patient-specific mean value curves show good agreement to the experimental data (Figure 5). Notably, beside the PK range which is kind of a ‘macroscopic’ result of the posterior parameter distribution a lot of other information can be obtained by directly analyzing the posterior. The calculation of correlations between the 8 individual parameters provided information about dependencies between the various parameters in the model. For example, a strong correlation between  $P_{int}$  and  $k_{cat,M}$  was observed (Figure 6).

We next asked whether our approach can also be used for the identification of specific subgroups within a population. This is a challenging task in particular in early phases of drug development, since only little prior knowledge may be available. Therefore, we asked if our Bayesian-PBPK approach enabled the identification of such homogenous groups of patients even if no additional information was taken into account and considered the transporter activities of MRP2 and OATP1B1 as a putative source for subgroup stratification.

First, we performed a Shapiro-Wilk test for normal distribution (Shapiro & Wilk, 1965) of the logarithmic mean values of the 200 samples of every patient, since protein expression has to be log-normally distributed



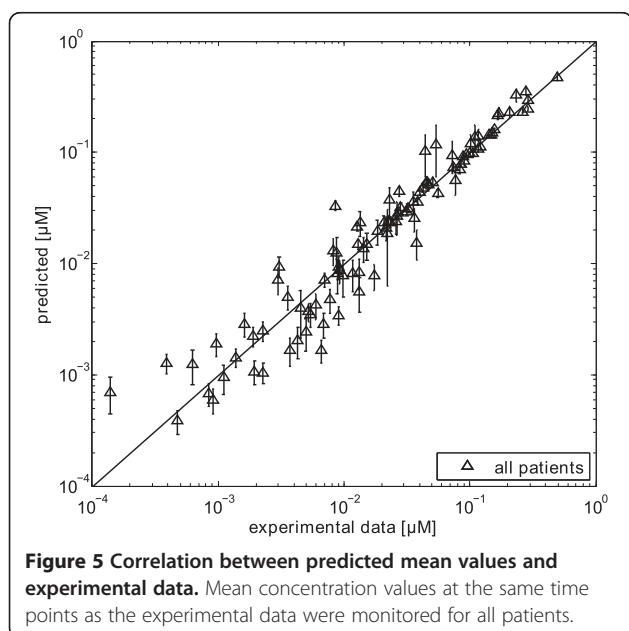
**Figure 3 Exemplary representation of a subsample of the posterior distribution.** After a burn-in period of 150000 steps, a subsample of 200 parameter vectors was drawn for each patient. The figure shows the traces for all eight individual parameters exemplarily for one patient as well as the four global parameters which were the same for all patients. The limits on the y-axis represent the physiological constraints ( $\theta_{min}$ ,  $\theta_{max}$ ).



in homogenous groups of patients (Sigal et al., 2006; Spencer et al., 2009). The results supported the hypothesis of lognormal distribution for MRP2 ( $p > 0.75$ ) and gave a strong indication of rejection of the hypothesis for OATP1B1 ( $p < 0.1$ ). Visual inspection of the estimated kernel densities (Bowman & Azzalini, 1997) of the logarithmic mean values (Figure 7) supported this, since two groups of patients were monitored for OATP1B1 but the density of MRP2 is clearly normally distributed. Thus, with regard to OATP1B1 the patient mean values were

analyzed individually to examine which patient can be assessed to which group (Figure 7). A clear separation into two groups of four and six patients, respectively, was found. It should be noted that this separation of the OATP1B1 transporter activity was not an implicit property of the model structure but emerged as a result during the Bayesian-PBPK approach.

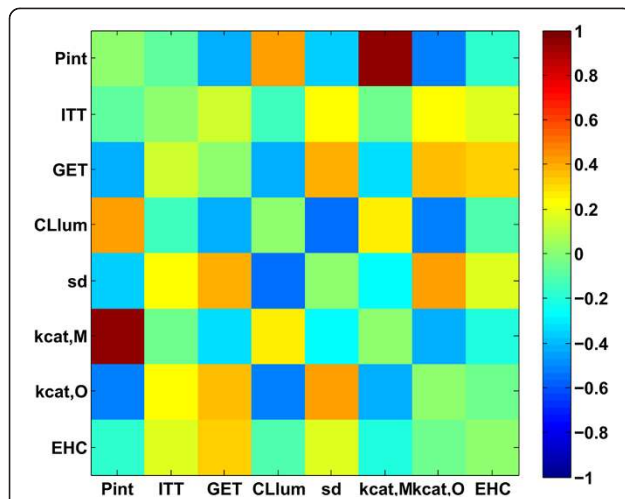
This grouping of patients was compared to the different genotypes in OATP1B1, which is known to significantly influence PK. This consideration led to a clear separation of the two homozygous genotypes, which demonstrated the capability of the approach to give strong hints about the reasons for subgroup stratification, even when only little experimental data of a small population was available.



## Discussion

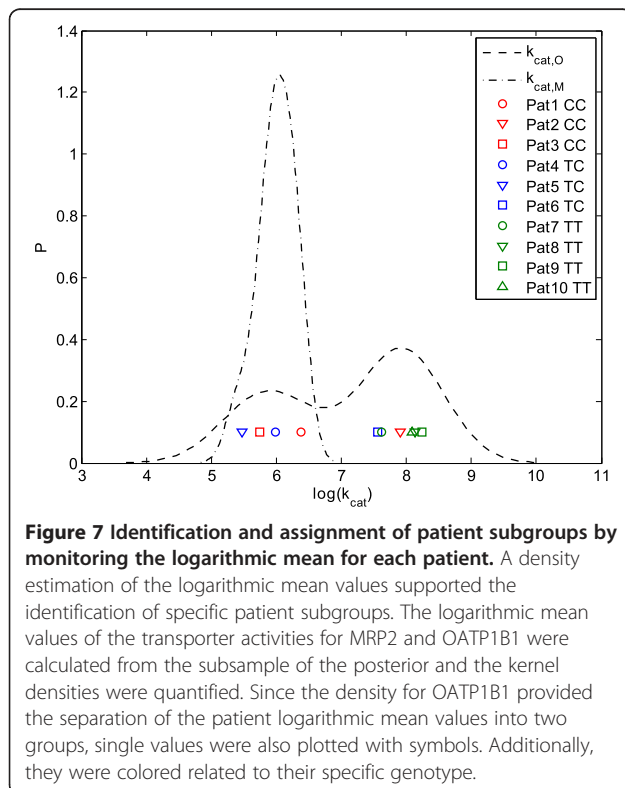
In the present work, we introduced a combined Bayesian-PBPK approach to quantify inter-individual variability in groups of patients. In former work such approaches have been mainly used in the context of toxicokinetics (Bois et al., 2010; Jonsson & Johanson, 2002). For PK simulations in virtual groups of patients, usually PBPK population models have been considered (Willmann et al., 2007). One drawback of combined Bayesian approaches so far has been that many PBPK models were relatively small and not fully physiology-based (Willmann et al., 2007). Due to model reduction processes, parameters contain mixed information about the patients' physiology and the substance. This ambiguity in parameter information prevents extrapolation to other drugs or groups of individuals. In contrast, the here used detailed mechanistic PBPK model is





**Figure 6 Correlation matrix of all individual parameters.** Spearman correlation coefficients were calculated from the overall subsample of 2000 parameter vectors for all parameter combinations to identify structural connections. To improve the visualization of the correlations the main diagonal was set to zero.

fully physiology-based and substance-specific parameters, physiological parameters and genotype-specific parameters are considered separately. The consideration of a Bayesian approach in combination with such models enables the inference of both physiological variations in a



**Figure 7 Identification and assignment of patient subgroups by monitoring the logarithmic mean for each patient.** A density estimation of the logarithmic mean values supported the identification of specific patient subgroups. The logarithmic mean values of the transporter activities for MRP2 and OATP1B1 were calculated from the subsample of the posterior and the kernel densities were quantified. Since the density for OATP1B1 provided the separation of the patient logarithmic mean values into two groups, single values were also plotted with symbols. Additionally, they were colored related to their specific genotype.

population and intra-individual parameter uncertainty of single patients and in contrast to the PBPK population approaches even when little experimental data is available.

Generally, one challenge in the performance of the approach is the identification of the convergence of the Markov chain. To prove that a finite sample of the posterior is representative to the posterior distribution, several tools have been developed. Most of them, however, are very difficult to use and have a relatively high probability to fail (Cowles & Carlin, 1996). Moreover, they were developed for less complex models and of lower dimensionality. To decide after how many steps the burn-in period ended, we visually inspected the traces of the parameters in all patients. Nevertheless, since this is a crucial point in MCMC, high quality convergence analyses should be considered in future work.

The gold standard for the assessment of inter-individual variability would be the identification of whole patients' physiology and the integration of as much experimental data as possible. In our model this would lead to the identification of hundreds of parameters per patient and thousands of parameters for a large population. Due to computational restrictions we here chose only a population of 10 patients and varied only several parameters per patient, however, the parameters have been chosen in a way such that all the important ADME processes were represented. A possible concept for computational reduction would be the parallelization of the individual Metropolis-Hastings blocks, which would reduce the computation time by the number of patients if enough computational power is available. Notably, the presented concept is not constrained in its dimensionality, therefore also the investigation of large populations and hundreds of parameter is possible, which provides great opportunities for the assessment of inter-individual variability in clinical trials.

By using a block-wise MH algorithm, a standard MCMC algorithm was chosen for this combined Bayesian-PBPK approach. This enabled first analyses of the behavior of the results as well as the simulation process itself under consideration of large mechanistic PBPK models. The several MH blocks allowed the separation between individual parameters and global parameters and reduced the convergence time of the run since every block could converge faster as if all parameters would have been varied in one large block. In following investigations, different algorithms such as adaptive approaches (Gilks et al., 1998; Haario et al., 2005; Roberts & Rosenthal, 2009) could be tested to identify the ones which for example further reduce convergence time or improve the mixing of the Markov chains. Furthermore, the use of Bayesian population approaches could be an option to make better inferences about the whole population, especially when only few patients are considered (Bernillon & Bois, 2000; Bois et al., 2010).

Concerning the global parameters it has to be noted that such parameters have to be chosen very carefully, since they have by definition a large effect on all obtained individuals. In our application example, the unbound protein fraction was defined as a global parameter. Since it is also determined by the composition of the blood serum it can as such also be defined as individual parameter. However, the unbound fraction also depends on the lipophilicity of the drug, which is varied in our approach. Therefore, both parameters had to be sampled in the same MH block to consider the covariance between these parameters (Bernillon & Bois, 2000; Smith et al., 1992).

Advantages of using the highly-detailed mechanistic PBPK model were demonstrated by analyzing the example of pravastatin. Relationships between the physiological parameters were provided directly from the posterior and could be easily identified, for example a strong correlation was found between the enzyme activity of the MRP2 transporter and the interstitial permeability in all patients. This results from a contrary transport of pravastatin in the gastrointestinal tract, because MRP2 transports pravastatin back into the intestinal lumen. Therefore, by the analysis of the posterior, structural information about the model can be inferred.

Furthermore, beside the derivation of structural information about the PBPK model the identification of clinically relevant subgroups within the population is possible. By investigating the logarithmic mean values of the single patients with a Shapiro-Wilk test the assumption of more than one homogenous group was confirmed for OATP1B1. Additionally, the two groups of patients were assigned to different homozygous genotypes. This demonstrates the ability of our approach to make physiological inferences with very little prior information and only few individuals. The heterozygous genotype could not be assigned to an own group. However, Niemi et al. also showed that a significant separation of the heterozygous genotype is not possible (Niemi et al., 2006). Notably, the separation of different subgroups itself may also be possible with smaller models. However, the use of a mechanistic PBPK model can point out the relation between subgroup and genotype which makes our Bayesian-PBPK approach a suitable alternative to rather phenomenological methods (Link et al., 2008).

## Conclusions

Altogether, our presented Bayesian-PBPK approach provides many opportunities for the assessment of inter-individual variability in groups of patients. The advantages of MCMC compared to classical population PK lie in its usability even when only little prior information or experimental data is available. Especially for early phases of clinical development the identification of subgroups as well as

special physiological properties can increase both patient safety and the level of information about the benefit-risk profile of a new drug candidate. The full physiological PBPK models allow the inference of physiological, drug-specific and genotype-specific knowledge separately, which therefore bears large extrapolation capacity to both, other drug candidates and populations. This will be greatly supported by the creation of a large database, where posterior knowledge about physiological parameter distributions can be collected iteratively. This would allow the consideration of posterior distributions of former Bayesian-PBPK runs as prior information in new runs, providing a framework for improving a mechanistic understanding of drug action, inter-individual variability and genotype-phenotype correlations with the help of growing amounts of different drugs or different populations.

## Additional files

**Additional file 1: Figure S1.** Schematic representation of the enterohepatic circulation and the key transporting enzymes in pravastatin pharmacokinetics. It has to be noted, that this is only a simplified consideration for a better representation of the processes. However, the enterohepatic cycle and the transporting enzymes are integrated into the mechanistic whole-body physiologically-based pharmacokinetic model.

**Additional file 2: Table S1.** Parameter start values used for the initialization of the combined Bayesian-PBPK approach.

## Competing interests

RB, JL are employees of Bayer Pharma AG. AS, SW, LK, LG are employees of Bayer Technology Services GmbH the company developing PK-Sim and MoBi.

## Authors' contributions

Recorded the data, conceived and designed the approach and analyses: MK RB JL MN PN AS LK LG. Performed the analyses: MK LK LG. Analyzed the results: MK RB JL AS SW LK LG. Wrote the paper: MK RB JL MN PN AS SW LK LG. All authors read and approved the final manuscript.

## Acknowledgments

The authors acknowledge financial support by the German Federal Ministry of Education and Research for the grant #0315747 (Virtual Liver). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author details

<sup>1</sup>Bayer Technology Services GmbH, Computational Systems Biology, Leverkusen 51368, Germany. <sup>2</sup>Aachen Institute for Advanced Study in Computational Engineering Sciences, RWTH Aachen, Schinkelstr, Aachen 2, 52062, Germany. <sup>3</sup>Bayer Pharma AG, Clinical Pharmacometrics, Wuppertal 42117, Germany. <sup>4</sup>Department of Clinical Pharmacology, University of Helsinki, Helsinki, Finland. <sup>5</sup>HUSLAB, Helsinki University Central Hospital, Helsinki, Finland.

Received: 3 January 2013 Accepted: 24 March 2013

Published: 11 April 2013

## References

- Andrieu C, de Freitas N, Doucet A, Jordan MI (2003) An Introduction to MCMC for Machine Learning. *Machine Learning* 50:5–43
- Andrieu C, Doucet A, Holenstein R (2010) Particle Markov chain Monte Carlo methods. *J R Stat Soc: Series B (Statist Methodol)* 72:269–342
- Atchadé YF, Rosenthal JS (2005) On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* 11:815–828

- Bernillon P, Bois FY (2000) Statistical Issues in Toxicokinetic Modeling: A Bayesian Perspective. *Environ Heal Perspect* 108:883–893
- Bois FY, Jamei M, Clewell HJ (2010) PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. *Toxicology* 278:256–267
- Bolstad WM (2010) *Understanding Computational Bayesian Statistics*. John Wiley & Sons, New Jersey
- Bowman AW, Azzalini A (1997) *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Clarendon Press; Oxford University Press, Oxford; New York
- Cowles MK, Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J Am Stat Assoc* 91:883–904
- Edginton AN, Willmann S (2008) Physiology-Based Simulations of a Pathological Condition: Prediction of Pharmacokinetics in Patients with Liver Cirrhosis. *Clin Pharmacokinet* 47:743–752
- Eissing T, Kuepfer L, Becker C, Block M, Coboeken K, Gaub T, Goerlitz L, Jaeger J, Loosen R, Ludewig B, Meyer M, Niederalt C, Sevestre M, Siegmund H-U, Solodenko J, Thelen K, Telle U, Weiss W, Wendl T, Willmann S et al (2011) A computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and molecular reaction networks. *Front Physio* 2:4
- Eissing T, Lippert J, Willmann S (2012) Pharmacogenomics of Codeine, Morphine, and Morphine-6-Glucuronide: Model-Based Analysis of the Influence of CYP2D6 Activity, UGT2B7 Activity, Renal Impairment, and CYP3A4 Inhibition. *Mol Diagn Ther* 16:43–53
- Everett DW, Chando TJ, Didonato GC, Singhvi SM, Pan HY, Weinstein SH (1991) Biotransformation of pravastatin sodium in humans. *Drug Metab Dispos* 19:740–748
- Gelfand AE, Smith AF (1990) Sampling-Based Approaches to Calculating Marginal Densities. *J Am Stat Assoc* 85:398–409
- Gelman A, Bois F, Jiang J (1996) Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions. *J Am Stat Assoc* 91:1400–1412
- Geman S, Geman D (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans Pattern Anal Machine Intell* 6:721–741
- Gilks WR, Roberts GO, Sahu SK (1998) Adaptive Markov Chain Monte Carlo through Regeneration. *J Am Stat Assoc* 93:1045–1054
- Gueorguieva I, Aarons L, Rowland M (2006) Diazepam Pharmacokinetics from Preclinical to Phase I Using a Bayesian Population Physiologically Based Pharmacokinetic Model with Informative Prior Distributions in Winbugs. *J Pharmacokinet Pharmacodyn* 33:571–594
- Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. *Bernoulli* 7:223–242
- Haario H, Saksman E, Tamminen J (2005) Componentwise adaptation for high dimensional MCMC. *Comput Statist* 20:265–273
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Hatanaka T (2000) Clinical Pharmacokinetics of Pravastatin: Mechanisms of Pharmacokinetic Events. *Clin Pharmacokinet* 39:397–412
- Jonsson F, Johanson G (2002) Physiologically Based Modeling of the Inhalation Kinetics of Styrene in Humans Using a Bayesian Population Approach. *Toxicol Appl Pharmacol* 179:35–49
- Kivisto KT, Niemi M (2007) Influence of drug transporter polymorphisms on pravastatin pharmacokinetics in humans. *Pharm Res* 24:239–247
- Krauss M, Schaller S, Borchers S, Findeisen R, Lippert J, Kuepfer L (2012) Integrating Cellular Metabolism into a Multiscale Whole-Body Model. *PLoS Comput Biol* 8:e1002750
- Kuepfer L, Lippert J, Eissing T (2012) Multiscale Mechanistic Modeling in Pharmaceutical Research and Development. *Adv Exp Med Biol* 736:543–561
- Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R (2008) SLCO1B1 variants and statin-induced myopathy—a genomewide study. *N Engl J Med* 359:789–799
- Lippert J, Brosch M, von Kampen O, Meyer M, Siegmund HU, Schafmayer C, Becker T, Laffert B, Gorlitz L, Schreiber S, Neuvonen PJ, Niemi M, Hampe J, Kuepfer L (2012) A Mechanistic, Model-Based Approach to Safety Assessment in Clinical Development. *CPT: Pharmacomet Syst Pharmacol* 1:e13
- Meibohm B, Derendorf H (1997) Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling. *Int J Clin Pharm Th* 35:401–413
- Metropolis N (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087
- Meyer M, Schneckener S, Ludewig B, Kuepfer L, Lippert J (2012) Using expression data for quantification of active processes in physiologically-based pharmacokinetic modeling. *Drug Metab Dispos* 40:892–901
- Nestorov I (2007) Whole-body physiologically based pharmacokinetic models. *Expert Opin Drug Metab Toxicol* 3:235–249
- Niemi M, Pasanen MK, Neuvonen PJ (2006) SLCO1B1 polymorphism and sex affect the pharmacokinetics of pravastatin but not fluvastatin. *Clin Pharmacol Ther* 80:356–366
- Pang KS, Durk MR (2010) Physiologically-based pharmacokinetic modeling for absorption, transport, metabolism and excretion. *J Pharmacokinet Pharmacodyn* 37:591–615
- Poulin P, Schoenlein K, Theil FP (2001) Prediction of adipose tissue: plasma partition coefficients for structurally unrelated drugs. *J Pharm Sci* 90:436–447
- Roberts GO, Rosenthal JS (2009) Examples of Adaptive MCMC. *J Comput Graph Stat* 18:349–367
- Rodgers T, Leahy D, Rowland M (2005) Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci* 94:1259–1276
- Rodgers T, Rowland M (2006) Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci* 95:1238–1257
- Rowland M, Peck C, Tucker G (2011) Physiologically-Based Pharmacokinetics in Drug Development and Regulatory Science. *Annu Rev Pharmacol Toxicol* 51:45–73
- Schmitt W, Willmann S (2004) Physiology-based pharmacokinetic modeling: ready to be used. *Drug Discov Today: Technol* 1:449–456
- Schüttler J, Ihmsen H (2000) Population Pharmacokinetics of Propofol A Multicenter Study. *Anesthesiology* 92:727–738
- Serajuddin AT, Ranadive SA, Mahoney EM (1991) Relative lipophilicities, solubilities, and structure-pharmacological considerations of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors pravastatin, lovastatin, mevastatin, and simvastatin. *J Pharm Sci* 80:830–834
- Shapiro SS, Wilk MB (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52:591–611
- Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N, Alon U (2006) Variability and memory of protein levels in human cells. *Nature* 444:643–646
- Singhvi SM, Pan HY, Morrison RA, Willard DA (1990) Disposition of pravastatin sodium, a tissue-selective HMG-CoA reductase inhibitor, in healthy subjects. *Br J Clin Pharmacol* 29:239–243
- Smith AE, Ryan PB, Evans JS (1992) The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Anal* 12:467–474
- Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459:428–432
- Swen JJ, Huizinga TW, Gelderblom H, de Vries EGE, Assendelft WJJ, Kirchheiner J, Guchelaar H-J (2007) Translating Pharmacogenomics: Challenges on the Road to the Clinic. *PLoS Med* 4:e209
- Willmann S, Edginton AN, Coboeken K, Ahr G, Lippert J (2009) Risk to the breast-fed neonate from codeine treatment to the mother: a quantitative mechanistic modeling study. *Clin Pharmacol Ther* 86:634–643
- Willmann S, Hohn K, Edginton A, Sevestre M, Solodenko J, Weiss W, Lippert J, Schmitt W (2007) Development of a physiology-based whole-body population model for assessing the influence of individual variability on the pharmacokinetics of drugs. *J Pharmacokinet Pharmacodyn* 34:401–431
- Willmann S, Lippert J, Sevestre M, Solodenko J, Fois F, Schmitt W (2003a) PK-Sim<sup>®</sup>: a physiologically based pharmacokinetic 'whole-body' model. *Biosilico* 1:121–124
- Willmann S, Schmitt W, Keldenich J, Dressman JB (2003b) A physiologic model for simulating gastrointestinal flow and drug absorption in rats. *Pharm Res* 20:1766–1771
- Willmann S, Schmitt W, Keldenich J, Lippert J, Dressman JB (2004) A physiological model for the estimation of the fraction dose absorbed in humans. *J Med Chem* 47:4022–4031
- Yang Y, Xu X, Georgopoulos PG (2009) A Bayesian population PBPK model for multiroute chloroform exposure. *J Expo Sci Environ Epidemiol* 20:326–341

doi:10.1186/2193-9616-1-6

Cite this article as: Krauss et al.: Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification. *In Silico Pharmacology* 2013 **1**:6.