CrossMark

**ORIGINAL PAPER**

# The impact of composite AUC estimates on the prediction of systemic exposure in toxicology experiments

Tarjinder Sahota[1] · Meindert Danhof[1] · Oscar Della Pasqua[1,2,3]

**Abstract** Current toxicity protocols relate measures of systemic exposure (i.e. AUC, $C_{max}$) as obtained by non-compartmental analysis to observed toxicity. A complicating factor in this practice is the potential bias in the estimates defining safe drug exposure. Moreover, it prevents the assessment of variability. The objective of the current investigation was therefore (a) to demonstrate the feasibility of applying nonlinear mixed effects modelling for the evaluation of toxicokinetics and (b) to assess the bias and accuracy in summary measures of systemic exposure for each method. Here, simulation scenarios were evaluated, which mimic toxicology protocols in rodents. To ensure differences in pharmacokinetic properties are accounted for, hypothetical drugs with varying disposition properties were considered. Data analysis was performed using non-compartmental methods and nonlinear mixed effects modelling. Exposure levels were expressed as area under the concentration versus time curve (AUC), peak concentrations ($C_{max}$) and time above a predefined threshold (TAT). Results were then compared with the reference values to assess the bias and precision of parameter estimates. Higher accuracy and precision were observed for model-based estimates (i.e. AUC, $C_{max}$ and TAT), irrespective of group or treatment duration, as compared with non-compartmental analysis. Despite the focus of guidelines on establishing safety thresholds for the evaluation of new molecules in humans, current methods neglect uncertainty, lack of precision and bias in parameter estimates. The use of nonlinear mixed effects modelling for the analysis of toxicokinetics provides insight into variability and should be considered for predicting safe exposure in humans.

**Keywords** Toxicokinetics · NOAEL · Safety margin · Model-based drug development

**Abbreviations**
AUC    Area under the concentration versus time curve
$C_{max}$    Peak concentrations
ICH    International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use
PD    Pharmacodynamics
PK    Pharmacokinetics
TAT    Time above a concentration threshold

✉ Oscar Della Pasqua
o.dellapasqua@ucl.ac.uk

1    Division of Pharmacology, Leiden Academic Centre for Drug Research, University of Leiden, Leiden, The Netherlands

2    Clinical Pharmacology Modelling and Simulation, GlaxoSmithKline, Stockley Park West, Uxbridge, UK

3    Clinical Pharmacology & Therapeutics, School of Life and Medical Sciences, University College London, London, UK

## Introduction

The purpose of toxicokinetic studies in the evaluation of safety pharmacology and toxicity is the prediction of the risk that exposure to a new chemical or biological entity represents to humans [1, 2]. Understanding of the relationships between drug exposure, target engagement (i.e., activation or inhibition) and downstream biological effects of a given physiological pathway can provide insight into the mechanisms underlying both expected and 'unexpected' toxicity [3] (Fig. 1). In addition, the use of a

**Fig. 1** Diagram displaying the contribution of toxicokinetics and pharmacology for the characterisation of target-related adverse events and safety risk assessment. The circle depicting target efficacy highlights the role of information regarding the primary target engagement for safety risk assessment. Data on the target efficacy is usually obtained during in vitro and in vivo screening. The arrow indicates that inferences can be made about safety and risk based on the evidence from drug exposure and organ-specific toxicity data. Reprinted with permission from Horii 1998 [3]
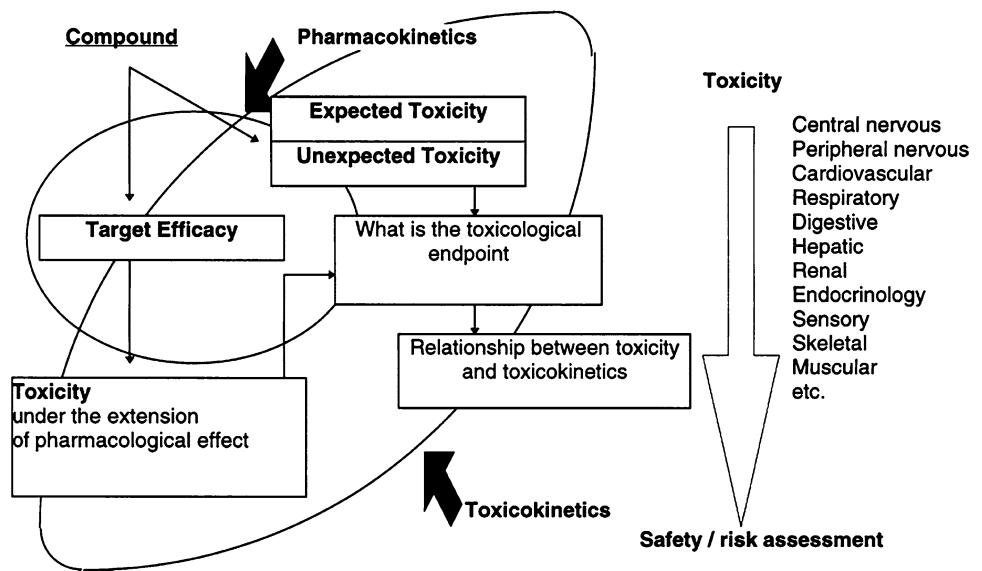


mechanism-based approach has allowed better interpretation of time-dependencies in drug effect, which are often observed following chronic exposure to a drug (e.g., delayed toxicity) [4, 5].

Despite the increased attention to the importance of toxicokinetics in drug discovery and during the early stages of clinical development, the extrapolation and prediction of a safe exposure range in humans from preclinical experiments continues to be one of the major challenges in R&D (Fig. 2) [6]. Irrespective of the choice of experimental protocol, a common practice in toxicology remains the assessment of empirical safety thresholds, in particular the no observed adverse effect level (NOAEL), which is a *qualitative* indicator of acceptable risk. Even though support for the existence of thresholds has been argued on biological grounds [7–9], the NOAEL has been used to establish the safe exposure levels in humans. In fact, this threshold represents a proxy for another threshold, i.e., the underlying no adverse event level (NAEL).

The definition of the NOAEL varies from source to source [6]. Its calculation involves the determination of the lowest observed adverse effect level (LOAEL), which is the lowest observed dose level for which AEs are recorded. The NOAEL is the dose level below this. If no LOAEL is found, then the NOAEL cannot be determined. Usually, in the assessment of the LOAEL measures of systemic exposure are derived, such as area under the concentration versus time curve (AUC) and peak concentrations ($C_{max}$), which serve as basis for the maximum allowed exposure in dose escalation studies in humans [10]. The aforementioned practices in safety and toxicity evaluation are driven by regulatory guidance [11, 12]. The scope of these guidances is to ensure that data on the systemic exposure

achieved in animals is assessed in conjunction with dose level and its relationship to the time course of the toxicity or adverse events (Fig. 2). Another important objective is to establish the relevance of these findings for clinical safety as well as to provide information aimed at the optimisation of subsequent non-clinical toxicology studies.

Whilst the scope and intent of such guidance are well described since 1994, when it was introduced by the ICH, there has been much less attention to requirements for the analysis and interpretation of the data. In fact, precise details on the design of toxicokinetic studies or the statistical methods for calculating or estimating the endpoints or variables of interest, are not specified [13–15]. Instead, the assessment of exposure often takes places in satellite groups, which may not necessarily present the (same) adverse events or toxicity observed in the main experimental group. This is because of interferences associated with blood sampling procedures, which may affect toxicological findings. For this same reason, blood sampling for pharmacokinetics is often sparse [16]. Such practice also diverges from efforts in models in environmental toxicology, a field in which deterministic, physiologically-based pharmacokinetic models have been used for a long time [17, 18].

As a consequence, safety thresholds are primarily derived from inferences about the putative pharmacokinetic profiles in the actual treatment group. Furthermore, these thresholds rely on the accuracy of composite profiles obtained from limited sampling in individual animals. Composite profiles consist of pooled concentration data, which is averaged per time point under the assumption that inter-individual differences are simply residual variability, rather than intrinsic differences in pharmacokinetic
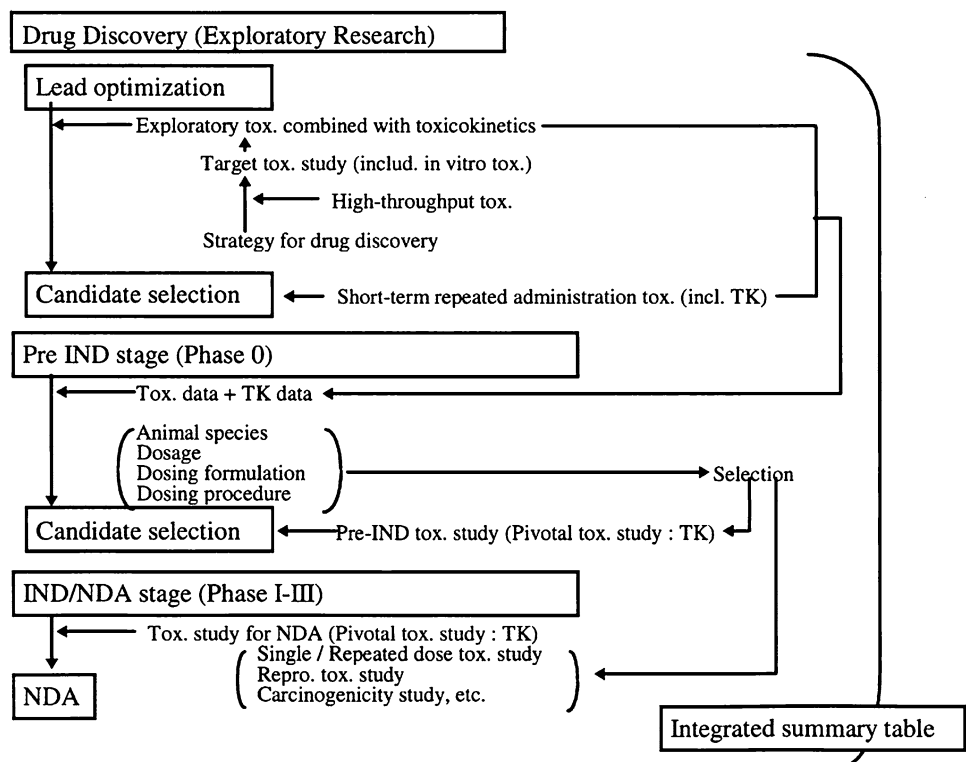
processes [19]. Pharmacokinetic parameters such as area-under-concentration-time (AUC) and observed peak concentrations ($C_{max}$) can then be either derived from the composite profile or by averaging individual estimates from serial profiles in satellite animals when frequent sampling schemes are feasible. Given that the parameters of interest are expressed as point estimates, within- and between-subject variability as well as uncertainty in estimation are not accounted for. In addition, pharmacokinetic data generated from different experiments are not evaluated in an integrated manner, whereby drug disposition (e.g., clearance) can be described mechanistically or at least compartmentally in terms of both first and zero order processes. This is further complicated by another major limitation in the way exposure is described by naïve pooling approaches, i.e., the impossibility to accurately derive parameters such as cumulative exposure, which may be physiologically a more relevant parameter for late onset or cumulative effects (e.g. lead toxicity, aminoglycosides) [20, 21]. Time spent above a threshold concentration may also bear greater physiological relevance for drugs which cause disruption of homeostatic feedback mechanisms. Such parameters cannot be described by empirical approaches due to limitations in sampling frequency.

By contrast, population pharmacokinetic-pharmacodynamic methodologies have the potential to overcome most of the aforementioned problems. Whilst the application of modelling in the evaluation of efficacy is widespread and well-established across different therapeutic areas [22–24], current practices have undoubtedly hampered the development of similar approaches for the evaluation of adverse events, safety pharmacology and toxicity. It should be noted that in addition to the integration of knowledge from a biological and pharmacological perspective, population models provide the basis for the characterisation of different sources of variability, allowing the identification of between-subject and between-occasion variability in parameters [25]. These random effects do not only reflect the evidence of statistical distributions. They can be used for inference about the mechanisms underlying adverse events and toxicity. In fact, recent advancements in environmental toxicology have shown the advantages of PBPK/PD modelling as a tool for quantifying target organ concentrations and dynamic response to arsenic in preclinical species [26].

The aim of this investigation was therefore to assess the relative performance of model-based approaches as compared to empirical methods currently used to analyse toxicokinetic data. We show that, modelling is an iterative process which allows further insight into relevant biological processes as well as into data gaps, providing the basis for experimental protocol optimisation. We illustrate the concepts by exploring a variety of scenarios in which

**Fig. 3** Overview of a simulated dataset along with the predicted pharmacokinetic profiles for each of the experimental scenarios, in which blood samples are collected from 3 animals per sampling time point. *Dots* represent simulated concentrations at the pre-defined sampling times, whereas the *solid black line* depicts the population predicted profile after a dose of 30 mg/kg for hypothetical drugs with different pharmacokinetic characteristics
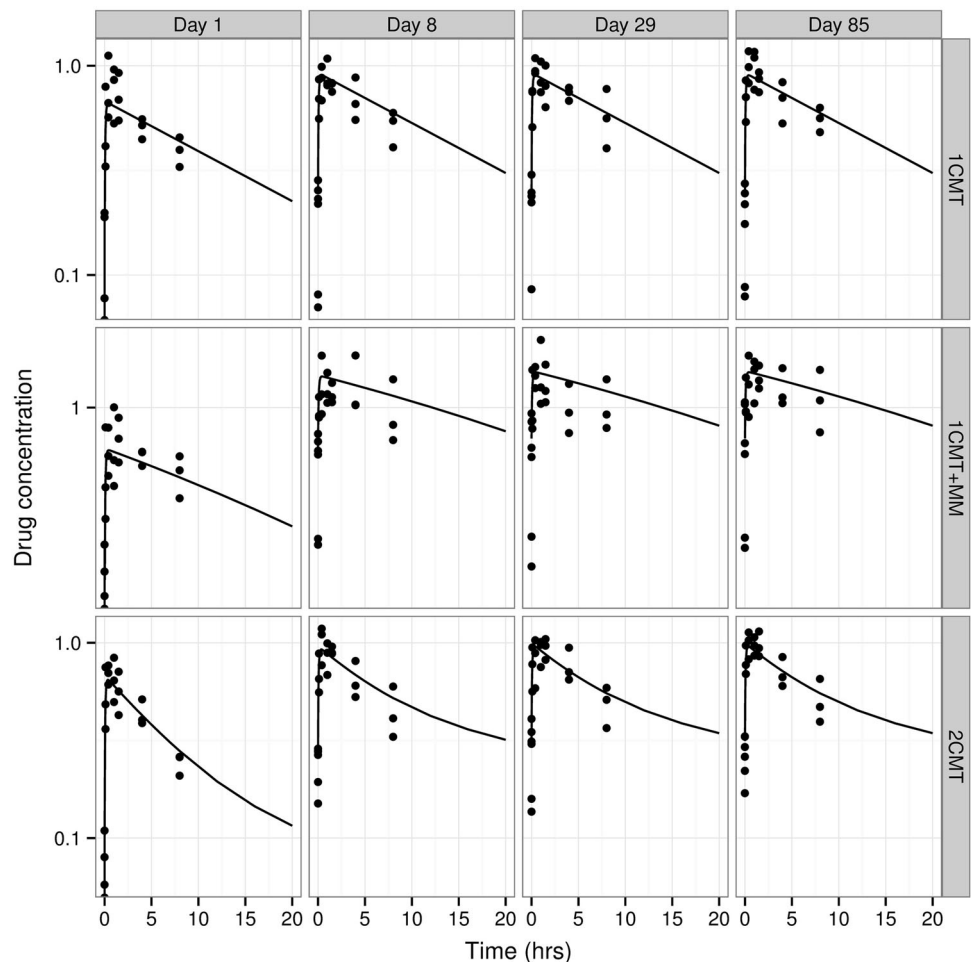


hypothetical drugs with different disposition properties are evaluated.

## Methods

Using historical reference data from a range of non-steroidal anti-inflammatory compounds for which pharmacokinetic parameter estimates were known in rodents, a model-based approach was used to simulate the outcomes of a 3-month study protocol, in which toxicokinetic data for three hypothetical drugs were evaluated. The selection of non-steroidal anti-inflammatory compounds as paradigm for this analysis is due to the mechanisms underlying both short and long term adverse events as well as the evidence for a correlation between drug levels and incidence of such events in humans. In fact, a relationship has been identified between the degree of inhibition of cyclooxygenase at the maximum plasma concentration ($C_{max}$) of individual non-steroidal anti-inflammatory drugs and relative risk (RR) of upper gastrointestinal bleeding/perforation [27].

## Simulation of drug profiles using predefined pharmacokinetic models

The impact of differences in drug disposition on bias and precision of the typical measures of systemic exposure was explored by including three different scenarios based on a one-compartment pharmacokinetics with linear and non-linear (Michaelis–Menten) elimination as well as a two-compartment pharmacokinetics. Parameter values for each scenario are shown in Table 1. In all simulation scenarios, residual variability was set to 15 %. For the purposes of this exercise, we have assumed that the models used as reference show no misspecification. In addition, we have considered the use of a homogeneous population of rodents, avoiding the need to explore covariate relationships in any of the models.

### Experimental design

Experimental procedures were defined according to current guidelines for the assessment of toxicity. A summary of the sampling schemes and experimental conditions is shown in

**Table 1** Pharmacokinetic models used to assess the impact of varying disposition properties on the estimation of safety thresholds

*Model A*: One-compartment model (1 CMT)

| Parameter | Pop estimate | BSV (%) |
| --- | --- | --- |
| KA ($h^{-1}$) | 13.46 | 50 |
| V (ml/kg) | 49.4 | 16 |
| CL (ml/h) | 2.72 | 20 |

*Model B*: One compartment model with Michaelis–Menten elimination (1 CMT + MM). Parameter values were chosen to ensure departure from dose proportionality at the highest dose

| Parameter | Pop estimate | BSV (%) |
| --- | --- | --- |
| $V_{max}$ (mg/h) | 2.72 | 20 |
| Km (mg/ml) | 1 | – |
| Ka ($h^{-1}$) | 13.46 | 50 |
| V (ml/kg) | 49.4 | 16 |

*Model C*: Two-compartment model (2 CMT). The values for the absorption and elimination rate constants were selected in such a way that slow accumulation of drug is observed at stead-state conditions after daily dosing for approximately 2 weeks

| Parameter | Value | Variability (%) |
| --- | --- | --- |
| Ka ($h^{-1}$) | 0.55 | 50 |
| V (ml/kg) | 49.4 | 16 |
| CL (ml/h) | 2.72 | 20 |
| K12 ($h^{-1}$) | 0.3 | – |
| K21 ($h^{-1}$) | 0.05 | 32 |

Table 2. The protocol design for each experiment with the three hypothetical drugs was based on protocols typically used for chronic toxicity evaluation. Four treatment groups receiving oral daily doses of vehicle, 10, 30, and 100 mg/kg/day were tested throughout this set of virtual experiments. The same treatment groups were present in all duration cohorts (1 week, 1 month or 3 months). Satellite groups each were used to characterise the pharmacokinetics under the dosing conditions in the animals used for the assessment of toxicity. This procedure ensures the availability of more frequent blood samples for toxicokinetics, while not influencing the assessment of the toxicity. Two different sampling schedules were investigated, namely, composite sampling and serial sampling. For the sake of

comparison, the same number of samples was collected in both cases. For composite sampling, blood was collected from three animals in the satellite group at predetermined sampling time points, namely, 0.1, 0.4, 1, 1.5, 4, 8, 24 h after drug administration on sampling days (see Table 2). The allocation of animals to each sampling time point was random within the constraint that all animals were sampled an equal number of times. An overview of a simulated dataset along with the predicted pharmacokinetic profiles for each of the experimental scenarios is shown in Fig. 3.

### Derivation of true exposure levels

Five different measures of exposure were derived from the simulated concentration profiles obtained from the models used for simulation. They included the 24-h area under the concentration versus time curve (AUC), the maximum concentration ($C_{max}$), the time above a threshold drug concentration (TAT), the predicted 6-month cumulative AUC and the predicted 6-month $C_{max}$. These exposure measures can be seen alongside the formula used for their calculation in Table 3. The threshold for adverse events was assumed to be 10 µg/ml. This arbitrary value was selected for illustrative purposes only. The simulations ($n = 200$ replicates) were performed assuming repeat dosing for up to 6 months (3 months beyond the treatment duration presented the investigated studies) in order to evaluate the implications of longer periods of drug exposure.

### Calculation of measures of exposure by non-compartmental analysis

Data from composite sampling across all satellite animals were used to determine the overall drug exposure, which consisted in averaging the simulated concentrations at each sampling time point. A similar approach was used for serial sampling, but in this case, drug exposure was calculated for each individual animal and then averaged over the cohort. In both cases, the arithmetic mean and geometric mean were used as summary statistics. As non-compartmental methods do not allow extrapolation beyond the actual experimental conditions, only three of the five measures of exposure were derived, namely, the AUC, estimated using

**Table 2** Experimental design characteristics of treatment and satellite groups in a general toxicity study with serial and composite pharmacokinetic sampling

| Duration | Numbers of animals | Sampling scheme |
| --- | --- | --- |
| 1 week | *Toxicity*: 4 per dose group | *Toxicity*: Composite 2 per animal |
| | *Satellite*: 3 per dose group | *Satellite*: Serial profiles from day 1 only |
| 1 month | *Toxicity*: 10 per dose group | *Toxicity*: Composite 2 per animal |
| | *Satellite*: 3 per dose group | *Satellite*: Serial profiles from day 1 and 28 |
| 3 months | *Toxicity*: 12 per dose group | *Toxicity*: Composite: week 4, week 13 |
| | *Satellite*: 3 per dose group | *Satellite*: Serial profiles from day 1, week 4, week 13 |

**Table 3** Exposure measures derived from the simulated concentration vs. time profiles. The assessment of bias and precision in the estimates of safety thresholds was based on these secondary pharmacokinetic parameters, which are shown alongside the formula used for their calculation. Individual predicted drug concentrations are denoted by $C_p(t)$

| Variable name | Symbol | Model based exposure calculation |
|---|---|---|
| 24-h AUC | AUC24 | $\int_{t-24}^{t} C_p \, dt$ |
| 24-h $C_{max}$ | $C_{max24}$ | $\max\left(\left\{C_p(s) : t - 24 < s < t\right\}\right)$ |
| 24-h time above threshold drug concentration | TAT | $\int_{t-24}^{t} 1_{C_p > thresh} \, dt$ |
| Predicted 6-month cumulative AUC | CAUC | $\int_{0}^{6months} C_p \, dt$ |
| Predicted 6-month $C_{max}$ | $C_{max24}$ | $\max\left(\left\{C_p(s) : 0 < s < 6months\right\}\right)$ |

the linear-logarithmic trapezoidal rule, the $C_{max}$, and the TAT.

## Calculation of measures of exposure by nonlinear mixed effects modelling

For each simulation replicate, drug concentration profiles were fitted to pharmacokinetic models using the first-order conditional estimation method with interaction (FOCEI), as implemented in NONMEM. Model building steps were limited to the same structural models used for the initial simulations under the assumption that pharmacokinetic properties of the drugs are known at the time toxicology experiments are performed. Model convergence was determined by successful minimisation and estimation of the covariance step. Data below the lower quantification limit (BQL) were omitted to mimic experimental conditions in which imputation methods are not applied. Estimates for all five measures of exposure were calculated by using same procedures applied for the reference values obtained during the initial simulation step (see Table 3).

### Comparison

To ensure accurate estimates of bias and precision of the two methodologies, the process of simulation and estimation of exposure (using non-compartmental vs. nonlinear mixed effects) was repeated 200 times. Bias and precision were assessed by the relative error, scaled relative mean error (SRME) and the coefficient of variation (CV) respectively [28]:

$$SMRE = \frac{1}{N} \sum_{i=1}^{N} \frac{(estimated_i - true)}{true} \times 100$$

$$CV = \frac{1}{N} \sqrt{\sum_{i=1}^{N} \left(\frac{estimated_i - mean}{mean}\right)^2} \times 100$$

All simulations and fitting procedures described above were performed in NONMEM 7.1 [29]. Data manipulation

**Table 4** Rates of convergence and covariance (parameter precision) estimation based on nonlinear mixed effects modelling. Simulated drug concentrations collected at the predefined sampling times were used as input for the pharmacokinetic analysis

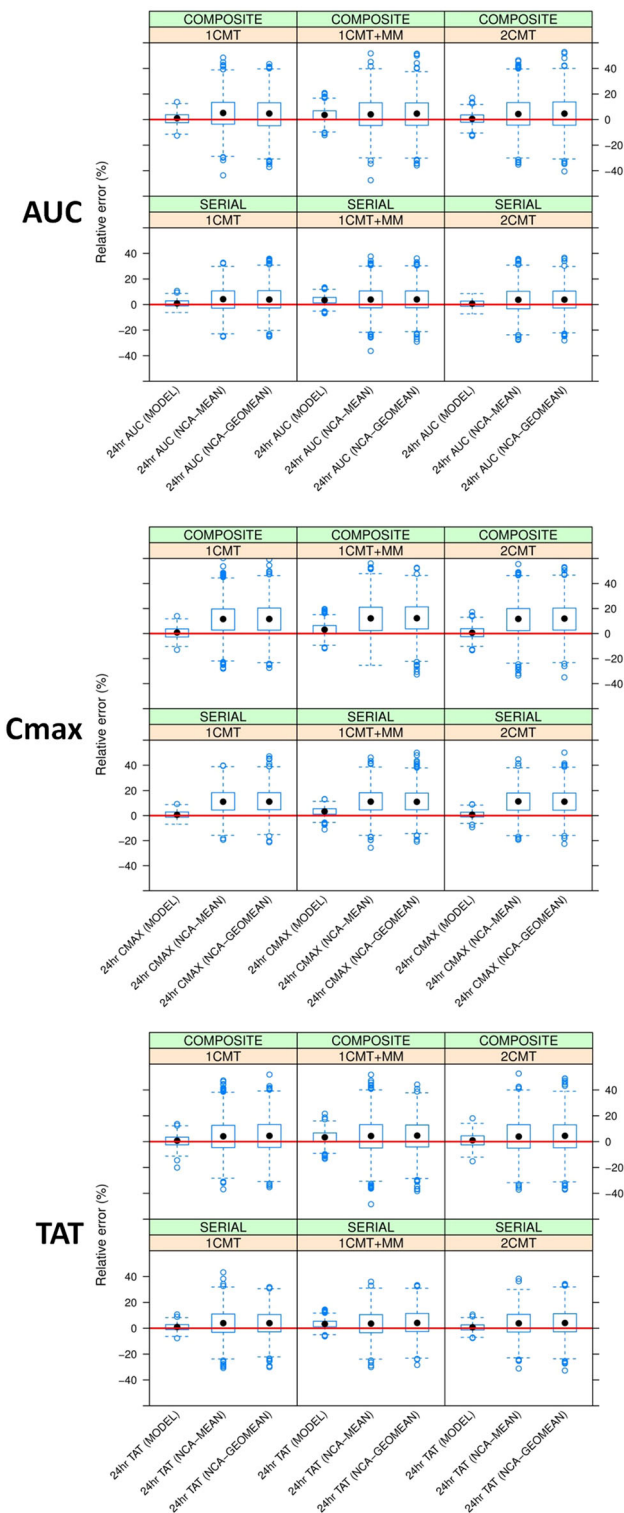| Model | Successful convergence | Successful covariance step |
|---|---|---|
| 1 CMT | 99.75 | 99.75 |
| 1 CMT + MM | 99.75 | 99.75 |
| 2 CMT | 100 | 100 |

and statistical and graphical summaries were performed in R 3.0.0 [30].

## Results

The use of simulated data for the evaluation of hypothetical scenarios provided clear insight of the impact of current practices on the accuracy and precision of safety thresholds, and in particular of the NOAEL. Irrespective of the use of serial or sparse sampling schemes for the characterisation of the concentration versus time profiles, model convergence rates were usually high, with successful completion of the covariance step. An overview of the convergence rates is presented in Table 4.

To facilitate the comparison of the magnitude of bias and precision, results from modelling are shown together with the parameter values obtained from non-compartmental analysis where applicable. Due to the large number of experimental conditions to be summarised, here we present a brief description of the relative errors obtained in the 3-month protocol, for AUC, $C_{max}$ and TAT. All other experimental conditions, including an overview of the scaled relative mean error (SRME) and the coefficient of variation (CV) are presented in tabular format as supplemental material (Table S1).

In Fig. 4, the relative errors are presented for the estimates for AUC, $C_{max}$ and TAT. The relative errors were

exposure showed significantly higher bias and less precision in all scenarios. The performance for model-based exposure estimates derived from the 3-month protocol is summarised in Fig. 5.

Our results also reveal the impact of composite versus serial sampling on bias and precision. For both model-based and non-compartmental methods, the coefficient of variation increased with composite designs (with 8 animals), as compared to serial sampling designs (with 3 animals). However, the increase in precision for non-compartmental method was larger than for model-based estimates. It should also be noted that C$_{max}$ was consistently over-estimated by the non-compartmental method. We also demonstrate that the use of arithmetic and geometric means for NCA had minor impact in these relatively small groups.

Lastly, it was found that that nonlinearity in pharmacokinetics also has an important effect on bias and precision when sparse samples and limited number of dose levels are evaluated experimentally. Model-based estimates in the 1 CMT + MM scenario showed increased bias compared to the 1 CMT and 2 CMT scenarios.

## Discussion

In this investigation we have attempted to identify important limitations in the use of non-compartmental methods for the analysis of toxicokinetic data. Irrespective of the limited number of scenarios, our findings illustrate the feasibility of using hierarchical models for the evaluation of toxicokinetic data using a well-established parameterisation for drug disposition processes. Furthermore, given that model performance in the analysis of toxicokinetic data has been previously evaluated [31], we have been able to focus on the performance of measures of exposure that cannot be derived from empirical approaches, i.e., non-compartmental methods [32].

It is important to highlight that the use of compartmental models, instead of physiologically-based pharmacokinetic models in this exercise was required to avoid issues such as parameter identifiability [33], which would arise from the

clearly smaller when measures of exposure were derived by modelling, as compared to the results obtained by non-compartmental analysis. In fact, the accuracy and precision of model-based estimates for all three measures of exposure were similar across the different dosing groups and treatment durations. Non-compartmental estimates of
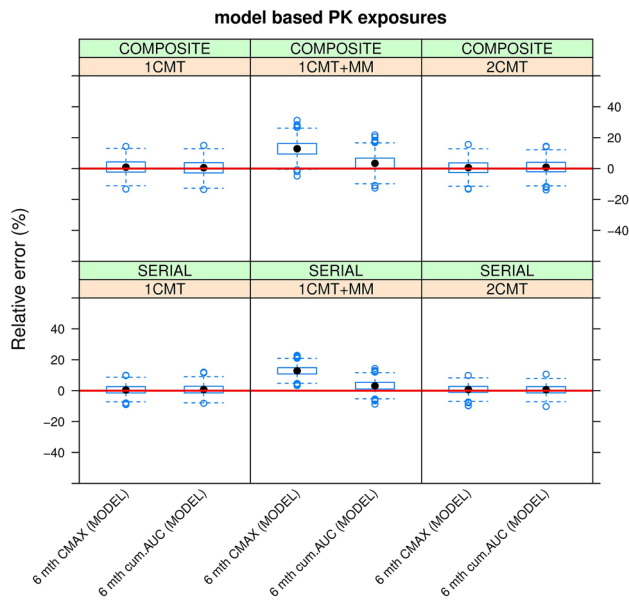
**Fig. 5** Overview of the relative errors of model-based estimators of long-term exposure, i.e. predicted peak concentrations after 6 months (6 mth Cmax) and cumulative area under the concentration vs. time curve (6 mth cum. AUC). The analysis is based on the data from a 3-month toxicology protocol following administration of 30 mg/kg/day of three hypothetical drugs with different pharmacokinetic profiles. Similar results were found for other cohorts in which 10 and 100 mg/kg/day were evaluated. *Dots* represent the median, *boxes* show the 25th and 75th percentiles, *error bars* denote 1.5 times the interquartile range from the median. The *red line* shows the reference level for relative error equal to zero (Color figure online)

data generated in standard toxicology protocols. The plasma pharmacokinetic profiles derived for the hypothetical compounds were considered realistic enough to reflect the time course of drug levels observed in many toxicology studies. In fact, these profiles are greatly affected by the standard sampling schemes in toxicology experiments, which may not allow one to identify more than one- and two-compartment models. Moreover, consideration was given to the implications that high doses may have on drug metabolism and elimination. A pharmacokinetic model with Michaelis–Menten elimination was also included to ensure accurate characterisation of dose- and concentration-dependent pharmacokinetics, which is likely to occur for many compounds at least in one experimental dose level. Saturation of metabolism has implications for the interpretation of safety thresholds, especially if nonlinearity is not observed at pharmacologically relevant levels. The results presented here should therefore be indicative of the most common toxicokinetic profiles. Given the evidence of the superiority of nonlinear mixed effects modelling to describe sparse pharmacokinetic data [34–37], we anticipate the possibility to generalise the lessons learned to a much wider range of drugs,

for which pharmacokinetic parameter values may differ considerably from those presented here.

**Parameter precision and bias**

As shown in Table 4, the high convergence rates of models and high success rate for the computation of the covariance matrix for the scenarios tested here confirm the robustness of results obtained using nonlinear mixed-effects modelling. Despite variations in bias and precision, parameter precision was consistently high. Whilst these results must be interpreted under the assumption of minor or no model misspecification, the use of modelling showed particularly good performance (CV < 10 % and SRME < 10 % for within study exposure predictions and SRME < 15 % for long term exposure predictions). Such high levels of precision may not be required for safe exposure evaluation where between-subject variability in humans is expected to be larger and comparatively large uncertainty factors are routinely used. This suggests that a model-based approach will enable considerable reductions in the numbers of animals and/or samples to be used in experimental protocols whilst providing acceptable parameter precision. Moreover since optimal design methodologies for model-based analysis are well established, further refinement of the experimental protocol design is feasible if experimentalists and statisticians choose nonlinear mixed effects modelling as the primary method of analysis.

On the other hand, the presence of bias in some of the experimental conditions presented here has clear implications for the so-called safety margin and toxicological cover to be used as proxy for risk during clinical development, especially for $C_{max}$, which is consistently overestimated. This is due to the definition of peak concentrations in non-compartmental analysis where $C_M$ is necessarily greater than or equal to $Cp(t = T_{max})$, where $T_{max}$ represents the time point which maximises the true concentration–time profile. When the sampling scheme contains other observations in the region of $T_{max}$ there is potential for neighbouring sampling times to produce higher than predicted concentrations due to natural variability. This is a fundamental limitation in the methodology in that more samples around $T_{max}$ which intuitively should increase confidence, actually leads to more bias. In other words, with non-compartmental analysis precisely estimating $T_{max}$ comes at the unavoidable cost of biased estimation of $C_{max}$. Model-based analysis has an additional advantage in this respect. Without model misspecification issues, maximum likelihood estimates are (asymptotically) unbiased and have the property that increased sampling uniformly increases precision. The implications of model specification issues are discussed further in the limitations section. Given that the residual variability in the scenarios

was not large (i.e., fixed at 15 %), the bias seen here may increase with larger residual noise, which may occur in real life. The issue of bias can be further mitigated by the appropriate use of predictive checks. An example of the procedures can be found in the supplemental material for naproxen [38], where predictive checks including data analysed by non-compartmental methods illustrate how to assess bias in AUC and $C_{max}$.

## Data integration

In contrast to non-compartmental methods, our investigation was based on an integrated analysis of the data, i.e. by combining the results from all experimental cohorts. This is undoubtedly the primary driver of the increased accuracy and precision in model-based estimates [38–40]. In fact, we envisage further improvement by incorporating pharmacokinetic data from other experiments in the same species, which are normally collected during the preclinical evaluation of a molecule, as for instance during the characterisation of drug metabolism. Such an increase in precision would represent further adherence to the reduction, refinement and replacement principle (3 Rs) in ethical animal studies [41, 42]. It should also be noted that the possibility of data integration provides the basis for combining safety pharmacology and adverse event data, enabling the development of toxicokinetic-toxicodynamic models and consequently allowing for the evaluation of exposure–response relationships in a continuous manner. Such models would represent an advancement in toxicology and risk management and mitigation, as they provide the basis for mechanism-based inferences about unwanted effects, irrespective of their incidence or occurrence in the actual experimental protocol [4, 43].

It is important to realise that the typical point estimates of parameters derived from empirical methods to describe drug exposure give an undue measure of certainty, allowing for the propagation of uncertainty from estimation to uncertainty in safety thresholds, such as NOAEL. Whilst there exist methods for estimating uncertainty in a composite or destructive sampling approach [44–46], their adoption in experimental research has not been widespread due in part to the requirement of normality assumptions on toxicokinetic parameters, and an acceptance in guidelines towards possibly large amounts of imprecision [12].

As demonstrated here, model-based methods allow simulations to be performed in conjunction with estimation procedures, enabling the assessment of uncertainty associated with a variety of causes such as uninformative study design, large variability and/or unknown covariates. This entails an increase in the quality of the decision-making process and ultimately in the interpretation of the estimated safety thresholds [47].

Given the success of modelling and simulation in drug development [48–50], one should ask why the field of toxicology has yet to embrace it. The scepticism regarding the value of model-based approaches often arises from a view that knowledge about the model is required in advance [51, 52]. This argumentation is however flawed. Non-linear mixed effects modelling is specifically intended to efficiently process sparse data. The performance of the model-based exposure estimates in the composite designs is illustrative of this. Moreover, the inference principles used for hypothesis generation and characterisation of drug disposition parameters relies on the use of statistical criteria that are sophisticated enough to allow model identification and its suitability for subsequent parameter estimation purposes. Moreover, it should be noted that non-compartmental methods also make implicit assumptions about the underlying concentration versus time profile. For instance, with a linear-logarithmic analysis of AUC, first-order elimination kinetics is assumed. The suitability of measures of central tendency will also depend on the assumed distribution characteristics and on residual variability. These assumptions are often implicit and their validity regarding the dataset at hand cannot be checked during the analysis. There are no strong statistical justification to support the choice for non-compartmental methods, other than the lack of technical knowledge and familiarity with hierarchical modelling by toxicologists in industry and regulatory agencies. The persistence in the use of non-compartmental methods bears an unnoticed cost, i.e., the ethical cost of utilising more animals than what is really necessary.

## Potential limitations

In the present investigation, the impact of model misspecification in the analysis of general toxicity data was not investigated. For exposure measures which have a corresponding estimate based on non-compartmental methods (e.g. AUC and $C_{max}$), the impact is likely to be small as long as the model fit to the data is good. This is because these measures are highly dependent on the observations. Therefore, accurate prediction of the observed profiles during model evaluation is likely to result in accurate prediction of these exposure variables. Model misspecification however, may lead to significant bias when exposure predictions are made outside the experimental context (i.e. longer timescales or different dosing regimens) [53, 54]. This is a risk when the pharmacokinetics of the drug is nonlinear or shows metabolic saturation. To mitigate such effects we recommend that model selection criteria take into account not only the ability to describe data, but also the physiological relevance of model assumptions. When model development ends in multiple competing models

performing similarly with respect to statistical selection criteria, clear reporting of such model uncertainty is necessary. Model averaging should be discouraged when predictions arising from different model differ significantly [55]. Finally, parameter uncertainty should be incorporated into the predictions of exposure to ensure accurate evaluation of risk and potential therapeutic window of the compound.

In summary, evaluation of safety is paramount for the progression of new molecules into humans. Historically, toxicology experiments have evolved based the assumption that experimental findings suffice to demonstrate the absence or presence of risk. This assumption disregards growing evidence of the advantages of data integration for the characterisation of drug properties. Whilst the challenges R&D faces to translate toxicity findings from animals to humans may remain, the use of an integrated approach to the analysis and interpretation of toxicokinetic data represents further adherence to the 3Rs principle, enabling significant reduction in number of animals required for the evaluation of toxicokinetics.

# References

1. Walker DK (2004) The use of pharmacokinetic and pharmacodynamic data in the assessment of drug safety in early drug development. Br J Clin Pharmacol 58(6):601–608
2. Singh SS (2006) Preclinical pharmacokinetics: an approach towards safer and efficacious drugs. Curr Drug Metab 7(2):165–182
3. Horii I (1998) Advantages of toxicokinetics in new drug development. Toxicol Lett 102–103:657–664
4. Bai JP, Fontana RJ, Price ND, Sangar V (2013) Systems pharmacology modeling: an approach to improving drug safety. Biopharm Drug Dispos 35(1):1–14. doi:10.1002/bdd.1871
5. Danhof M, de Lange EC, Della Pasqua OE, Ploeger BA, Voskuyl RA (2008) Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. Trends Pharmacol Sci 29:186–191
6. Dorato MA, Engelhardt JA (2005) The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 42:265–274
7. Dybing E, Doe J, Groten J, Kleiner J, O'Brien J, Renwick AG, Schlatter J, Steinberg P, Tritscher A, Walker R, Younes M (2002) Hazard characterisation of chemicals in food and diet. Dose response, mechanisms and extrapolation issues. Food Chem Toxicol 40:237–282
8. Kroes R, Kleiner J, Renwick A (2005) The threshold of toxicological concern concept in risk assessment. Toxicol Sci 86:226–230
9. Slob W (1999) Thresholds in toxicology and risk assessment. Int J Toxicol 18:345–367
10. Wexler D, Bertelsen KM (2011) A brief survey of first-in-human studies. J Clin Pharmacol 51(7):988–993

11. U.S. Department of Health and Human Services (2010) Food and Drug Administration. Guidance for Industry M3(R2). Nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals
12. ICH Expert Working Group Guideline S3A (1994) Note for guidance on toxicokinetics: the assessment of systemic exposure in toxicity studies
13. DeGeorge JJ (1995) Food and Drug Administration viewpoints on toxicokinetics: the view from review. Toxicol Pathol 23(2):220–225
14. Dahlem AM, Allerheiligen SR, Vodicnik MJ (1995) Concomitant toxicokinetics: techniques for and interpretation of exposure data obtained during the conduct of toxicology studies. Toxicol Pathol 23(2):170–178
15. Ploemen J-P, Kramer H, Krajnc EI, Martinhe I (2007) Use of toxicokinetic data in preclinical safety assessment: a toxicologic pathologist perspective. Toxicol Pathol 35:834–837
16. Tse FL, Nedelman JR (1996) Serial versus sparse sampling in toxicokinetic studies. Pharm Res 13:1105–1108
17. Leung HW (1991) Development and utilization of physiologically based pharmacokinetic models for toxicological applications. J Toxicol Environ Health 32(3):247–267
18. Sweeney LM (2000) Comparing occupational and environmental risk assessment methodologies using pharmacokinetic modeling. Hum Ecol Risk Assess. 6(6):1101–1124
19. Hoffman WP, Heathman MA, Chou JZ, Allen DL (2001) Analysis of toxicokinetic and pharmacokinetic data from animal studies. In: Millard S, Krause A (eds) Applied statistics in the pharmaceutical industry: with case studies using S-Plus. Springer, New York, pp 75–106
20. Taib NT, Jarrar BM, Mubarak M (2004) Ultrastructural alterations in hepatic tissues of white rats (*Rattus norvegicus*) induced by lead experimental toxicity. Saudi J Biol Sci. 11(1):11–20
21. Sedó-Cabezón L, Boadas-Vaello P, Soler-Martín C, Llorens J (2014) Vestibular damage in chronic ototoxicity: a mini-review. Neurotoxicology 43:21–27
22. Huntjens DR, Spalding DJ, Danhof M, Della Pasqua OE (2006) Correlation between in vitro and in vivo concentration-effect relationships of naproxen in rats and healthy volunteers. Br J Pharmacol 148(4):396–404
23. Rocchetti M, Simeoni M, Pesenti E, De Nicolao G, Poggesi I (2007) Predicting the active doses in humans from animal studies: a novel approach in oncology. Eur J Cancer 43(12):1862–1868
24. Beier H, Garrido MJ, Christoph T, Kasel D, Trocóniz IF (2008) Semi-mechanistic pharmacokinetic/pharmacodynamic modelling of the antinociceptive response in the presence of competitive antagonism: the interaction between tramadol and its active metabolite on micro-opioid agonism and monoamine reuptake inhibition, in the rat. Pharm Res 25(8):1789–1797
25. Miller R (2012) Population pharmacokinetics. In: Atkinson AJ, Huang SM, Lertora JJ, Markey SP (eds) Principles of clinical pharmacology, 3rd edn. Elsevier Inc, London, pp 139–149
26. Liao CM, Liang HM, Chen BC, Singh S, Tsai JW, Chou YH, Lin WT (2005) Dynamical coupling of PBPK/PD and AUC-based toxicity models for arsenic in tilapia *Oreochromis mossambicus* from blackfoot disease area in Taiwan. Environ Pollut 135(2): 221–233
27. Massó González EL, Patrignani P, Tacconelli S, García Rodríguez LA (2010) Variability among nonsteroidal antiinflammatory drugs in risk of upper gastrointestinal bleeding. Arthritis Rheum 62(6):1592–1601
28. Ahn JE, Karlsson MO, Dunne A, Ludden TM (2008) Likelihood based approaches to handling data below the quantification limit using NONMEM VI. J Pharmacokinet Pharmacodyn 35(4):401–421
29. Beal S, Sheiner LB, Boeckmann A, Bauer RJ (2009) NONMEM user's guides (1989–2009). Icon Development Solutions, Ellicott City

30. Core Team R (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
31. Hing JP, Woolfrey SG, Greenslade D, Wright PM (2001) Is mixed effects modeling or naive pooled data analysis preferred for the interpretation of single sample per subject toxicokinetic data? J Pharmacokinet Pharmacodyn 28:193–210
32. Pai SM, Fettner SH, Hajian G, Cayen MN, Batra VK (1996) Characterization of AUCs from sparsely sampled populations in toxicology studies. Pharm Res 13:1283–1290
33. Poggesi I, Snoeys J, Van Peer A (2014) The successes and failures of physiologically based pharmacokinetic modeling: there is room for improvement. Expert Opinion Drug Metabol Toxicol 10(5):631–663
34. Sheiner LB, Beal SL (1980) Evaluation of methods for estimating population pharmacokinetic parameters, I. Michelis-Menten model: routine clinical data. J Pharmacokinet Biopharm 8:553–571
35. Sheiner LB, Beal SL (1981) Evaluation of methods for estimating population pharmacokinetic parameters, II. Biexponential model and experimental pharmacokinetic data. J Pharmacokinet Biopharm 9:635–651
36. Sheiner LB, Beal SL (1983) Evaluation of methods for estimating population pharmacokinetic parameters, III. Monoexponential model and routine clinical data. J Pharmacokinet Biopharm 11:303–319
37. Mahmood I, Miller R (1999) Comparison of the Bayesian approach and a limited sampling model for the estimation of AUC and $C_{max}$: a computer simulation analysis. Int J Clin Pharmacol Ther 37(9):439–445
38. Sahota T, Sanderson I, Danhof M, Della Pasqua O (2014) Model-based analysis of thromboxane $B_2$ and prostaglandin $E_2$ as biomarkers in the safety evaluation of naproxen. Toxicol Appl Pharmacol 278(3):209–219
39. Richard AM (2006) Future of toxicology–predictive toxicology: an expanded view of "chemical toxicity". Chem Res Toxicol 19(10):1257–1262
40. Edelstein M, Buchwald F, Richter L, Kramer S (2010) Integrating background knowledge from internet databases into predictive toxicology models. SAR QSAR Environ Res 21(1):21–35
41. Briggs K, Cases M, Heard DJ, Pastor M, Pognan F, Sanz F, Schwab CH, Steger-Hartmann T, Sutter A, Watson DK, Wichard JD (2012) Inroads to predict in vivo toxicology—an introduction to the eTOX project. Int J Mol Sci 13(3):3820–3846
42. Balls M, Goldberg AM, Fentem JH, Broadhead CL, Burch RL, Festing MFW, Frazier JM, Hendriksen CFM, Jennings M, van der Kamp MDO, Morton DB, Rowan AN, Russell C, Russell WMS, Spielmann H, Stephens ML, Stokes WS, Straughan DW, Yager JD, Zurlo J, van Zutphen BFM (1995) The three Rs: the way forward. Altern Lab Anim 23:838–866
43. Lapenna S, Gabbert S, Worth A (2012) Training needs for toxicity testing in the 21st century: a survey-informed analysis. Altern Lab Anim 40(6):313–320
44. Della Pasqua O (2013) Translational pharmacology: from animal to man and back. Drug Discov Today Technol 10(3):e315–e317
45. Bonate PL (1998) Coverage and precision of confidence intervals for area under the curve using parametric and non-parametric methods in a toxicokinetic experimental design. Pharm Res 15:405–410
46. Gagnon RC, Peterson JJ (1998) Estimation of confidence intervals for area under the curve from destructively obtained pharmacokinetic data. J Pharmacokinet Biopharm 26:87–102
47. Nedelman JR, Gibiansky E, Lau DT (1995) Applying Bailer's method for AUC confidence intervals to sparse sampling. Pharm Res 12:124–128
48. Chain AS, Dubois VF, Danhof M, Sturkenboom MC, Della Pasqua O, Cardiovascular Safety Project Team, TI Pharma PKPD Platform (2013) Identifying the translational gap in the evaluation of drug-induced QTc interval prolongation. Br J Clin Pharmacol 76(5):708–724
49. Zineh I, Woodcock J (2013) Clinical pharmacology and the catalysis of regulatory science: opportunities for the advancement of drug development and evaluation. Clin Pharmacol Ther 93(6):515–525
50. Jones HM, Mayawala K, Poulin P (2013) Dose selection based on physiologically based pharmacokinetic (PBPK) approaches. AAPS J. 15(2):377–387
51. van der Graaf PH, Benson N (2011) Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. Pharm Res 28(7):1460–1464
52. Geenen S, Taylor PN, Snoep JL, Wilson ID, Kenna JG, Westerhoff HV (2012) Systems biology tools for toxicology. Arch Toxicol 86(8):1251–1271
53. Cella M, Zhao W, Jacqz-Aigrain E, Burger D, Danhof M, Della Pasqua O (2011) Paediatric drug development: are population models predictive of pharmacokinetics across paediatric populations? Br J Clin Pharmacol 72(3):454–464
54. Cella M, Knibbe C, de Wildt SN, Van Gerven J, Danhof M, Della Pasqua O (2012) Scaling of pharmacokinetics across paediatric populations: the lack of interpolative power of allometric models. Br J Clin Pharmacol 74(3):525–535
55. Lunn DJ (2008) Automated covariate selection and Bayesian model averaging in population PK/PD models. J Pharmacokinet Pharmacodyn 35:85–100