

Universiteit
Leiden
The Netherlands

Pharmacology based toxicity assessment : towards quantitative risk prediction in humans

Sahota, T.

Citation
Sahota, T. (2014, October 30). *Pharmacology based toxicity assessment : towards quantitative risk prediction in humans*. Retrieved from https://hdl.handle.net/1887/29414

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/29414

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle http://hdl.handle.net/1887/29414 holds various files of this Leiden University dissertation

Author: Sahota, Tarjinder
Title: Pharmacology based toxicity assessment : towards quantitative risk prediction in humans
Issue Date: 2014-10-30

Pharmacology based toxicity assessment:

Towards quantitative risk prediction in humans

Tarjinder Sahota

PhD. Thesis, Leiden University, October 2014

Pharmacology based toxicity assessment:

Towards quantitative risk prediction in humans

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus Prof. mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 30 oktober 2014

klokke 11.15 uur

door

Tarjinder Sahota

geboren te Letchworth, United Kingdom

in 1982

Promotor	Prof. dr. M. Danhof
Co-promotor	Prof. dr. O.E. Della Pasqua
Referent	Prof. G. Tucker (Sheffield University, Verenigd Koninkrijk)
Overige Leden	Prof. dr. P.H. van der Graaf
	Prof. dr. K. Burggraaf
	Prof. dr. B. vd Water
	Prof. dr. T Hankemeier
	Dr. J.W. van der Laan

The research leading to this thesis has received funding from GlaxoSmithKline (Worldwide Product Development scholarship) and was performed at the Division of Pharmacology of the Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

The printing of this thesis was financially supported by:

Leiden Academic Centre for Drug Research

GlaxoSmithKline

Cover and layout by Roweena Asgalari

Printed by GVO printers & designers B.V., Ede, The Netherlands

ISBN: 978-90-6464-821-2

©2014 Tarjinder Sahota (tarjinder.z.sahota@gsk.com)

No part of this thesis may be reproduced or transmitted in any form or by any means without written permission of the author

CONTENTS

Section 1	**General Introduction**	**7**
Chapter 1	Challenges in the assessment and prediction of safety pharmacology and drug toxicity in humans	9
Chapter 2	Scope and intent of investigation	55
Section 2	**Conceptual framework**	**69**
Chapter 3	The impact of composite AUC estimates on the prediction of systemic exposure in toxicology experiments	71
Chapter 4	Application of optimal design concepts to experimental protocols for the evaluation of toxicokinetics and safety thresholds	120
Chapter 5	Use of biomarkers for the characterisation of long-term safety: advantages of a model-based approach for the analysis of toxicology experiments	148
Chapter 6	Utility of model based approaches to predict the risk of adverse events from preclinical toxicology protocols.	178
Section 3	**Case study and practical application**	**211**
Chapter 7	Model-based analysis of thromboxane B2 and prostaglandin E2 as biomarkers in the safety evaluation of naproxen.	213
Chapter 8	Model-based prediction of the acute and long-term safety profile of naproxen in rats	249
Section 4	**Conclusions and Perspectives**	**280**
Chapter 9	Pharmacology based toxicity assessment: Towards quantitative risk prediction in humans	282
Chapter 10	Nederlandse Samenvatting (Synopsis in Dutch)	320
	Acknowledgements	338
	Curriculum Vitae	339
	List of publications	340

SECTION 1: GENERAL INTRODUCTION

CHAPTER 1

Challenges in the assessment and prediction of safety pharmacology and drug toxicity in humans

Tarjinder Sahota, Meindert Danhof and Oscar Della Pasqua

Abstract

Despite ongoing efforts to better understand the mechanisms underlying safety and toxicity, approximately 30% of the attrition in drug discovery and development is still due to safety concerns. Changes in current practice regarding the assessment of safety and toxicity are required to reduce late stage attrition and enable effective development of novel medicines. This review focuses on the implications of empirical evidence generation for the evaluation of safety and toxicity during drug development. A shift in paradigm is proposed to 1) ensure that pharmacological concepts are incorporated into the evaluation of safety and toxicity; 2) facilitate the integration of historical evidence and thereby the translation of findings across species; and 3) promote the use of experimental protocols tailored to address specific safety and toxicity questions.

Based on historical examples, we highlight the challenges for the early characterisation of the safety profile of a new molecule and discuss how model-based methodology can be applied for the design and analysis of experimental protocols. Issues relative to the scientific rationale are categorised and presented as a hierarchical tree describing the decision making process. Focus is given to four different areas, namely, optimisation, translation, analytical construct, and decision criteria. From a methodological perspective, nonlinear-mixed effects modelling is recommended as a tool to account for such requirements. Its use in the evaluation of pharmacokinetics (PK) and pharmacokinetic-pharmacodynamic relationships (PKPD) has enabled the advance of quantitative approaches in pharmacological research in recent decades. Comparable benefits can be anticipated for the assessment of safety and toxicity.

1. Introduction

The assessment of the safety and toxicity profile of new chemical or biological entities is an integral part of drug development. Despite ongoing efforts to better understand the mechanisms underlying safety and toxicity, approximately 30% of the attrition in drug discovery and development is still due to safety concerns (1,2). Such a high attrition rate is further compounded by the empiricism and entrenched belief which prevails among industry scientists and regulators about the level of evidence and requirements for determining acceptable risk in humans.

In addition to its contribution to the attrition rate, safety and toxicity findings have business, legal and societal consequences, which often lead to speculations and even more empiricism in the evaluation and interpretation of experimental data. Whilst a positive benefit-risk ratio should be anticipated and subsequently demonstrated when administering new drugs to humans, the basis upon which inferences are made still lacks the scientific clarity and rigour one would endeavour. The efficiency and value of current paradigm for the evaluation of safety and toxicity, which relies primarily on standard battery tests at supra-therapeutic exposure levels of the investigational drug, is not questioned by the scientific community. Rather, it is mandated by regulators as a mechanism to minimise liabilities.

A shift in paradigm is required that 1) enables the introduction of pharmacological concepts to the evaluation of safety and toxicity; 2) facilitates the integration of historical evidence and thereby the translation of findings across species; and 3) promotes the value of experimental protocols tailored to address specific safety and toxicity questions.

In this review we will focus on the implications of current practice for drug development and consider the scientific and ethical requirements for the evaluation of safety and toxicity. Of particular interest for us is to demonstrate that despite the assumption that preclinical safety testing, toxicity findings are generally seen as predictive of human toxicity (3), inefficiencies in the experimental design violate the principle of the 3 Rs (reduction,

refinement and replacement) (4). Empirical evidence must be replaced by a model-based approach.

Two recent examples can be used to illustrate the issues with the current paradigm for the evaluation of safety and toxicity, namely the serious adverse events observed with TGN1412 and the increased incidence of myocardial infarction in patients who were prescribed rofecoxib. These two cases encompass most of the critical issues one attempts to address prior to making a commitment to clinical development and subsequently to regulatory submission and marketing of a medicinal product. Albeit neglected in the assessment of the clinical findings and in the subsequent reports in the published literature, the use of a mechanism-based approach in conjunction with some basic pharmacology concepts would be sufficient to predict the consequences of the treatment, whether given as single dose to healthy subjects or chronically to patients; i.e., both examples reflect the immediate consequences of target engagement and the corresponding changes due to the mechanism of action and (patho)physiological pathways. Yet, the experimental evidence generated pre-clinically for these two compounds does not take into account target engagement or exposure-response relationships as the basis for the interpretation of the findings. Instead, it is the characterisation of the maximum tolerated dose (MTD) and /or no-adverse effect level (NOAEL) that ultimately drives the design of safety pharmacology and toxicity experiments. The empirical evidence of MTD and NOAEL does not provide insight into the underlying mechanisms and often obscures the translation of findings across species.

According to published reports, the serious adverse events observed after intravenous administration of TGN1412, a novel monoclonal T-cell agonist, could not have been "predicted" or inferred from non-clinical data. The empiricism in the design of the experimental protocol and in the interpretation of the findings clearly shows the disconnection between pharmacology and toxicology, despite extremely high degree of selectivity and specificity of the biologicals. The failure to predict a systemic inflammatory response by rapid induction of cytokines (a "cytokine storm") with catastrophic multi-organ failure (5) is not surprising when structure homology, target occupancy and pharmacokinetic

principles are disregarded. Despite the availability of in vitro binding assays, there was no attempt to correlate or integrate the results from different experiments with each other. Most importantly, the effects observed with the proposed dosing regimen could have been anticipated even without any experimental data. Knowledge of receptor agonism theory and drug disposition properties would have been sufficient to make inferences about target activation and pharmacological effects.

Tragedies like this provoke reactive measures from industry and regulator (6-11). New guidelines for the assessment of preclinical data were released by regulatory authorities. However, none of them tackle the problem from a scientific, mechanistic perspective. Similarly, changes have been introduced to the design of first-in-man studies (6), which reflect mitigation measures for process-related consequences of safety and toxicity findings. A framework that ensures critical appraisal of the scientific rationale, based on pharmacological concepts and expected biological activity (i.e., target engagement) is still missing.

Rofecoxib, a selective COX-2 inhibitor prescribed to more than 107 million patients in the US (12), is another example of withdrawal from the market because of so-called “unexpected” long-term safety findings. Despite the debate that followed the evidence from clinical trial data on the increased risk of myocardial infarction (13), little effort was made to incorporate very basic pharmacological concepts into the evaluation of the findings and provide a mechanism-based interpretation, which could easily disentangle the core issue: whether this is a class-effect or whether that was a compound specific toxicity. Paracelsus highlighted the importance of the dose more than 500 years ago, and yet none of the published reports considered this critical question: were patients receiving the optimal dose and dosing regimen for the proposed indications? Clinical and scientific experts dwelled on the realm of toxicity as the result of an off-target event, without exploring in a systematic manner the (obvious) connection to dosing regimen, target exposure, the time course of pharmacological effect, the duration of treatment and physiological role of the substrates for COX2 in the heart and other tissues. Evidence of concentration-effect relationship was

not gathered, neither used as basis for interpreting those findings. Instead, allegations of misconduct followed that overruled any comprehensive scientific debate (12).

From a clinical pharmacology perspective, the aforementioned examples reflect the failure in exploring causality and anticipating the biological consequences of target engagement, i.e., in establishing the correlation between target-related events and drug exposure, as defined by the evidence of pharmacokinetic-pharmacodynamic relationships. Post-market withdrawals are not an uncommon occurrence: between 1975 and early 2000 there have been 26 withdrawals from the US market due to safety issues (14). In fact, the withdrawal of a medicinal product seems to have become the expected *course of action* for regulators and industry who are faced with 'unexpected' safety findings. Interestingly, dosage changes, due to safety occurred in approximately one out five drugs in the period from 1980 to 1999 (15,16). On the other hand, from a clinical perspective, the aforementioned landscape appears to result from the lack of a formalised assessment of the benefit-risk ratio in which efficacy and safety are evaluated in an integrated manner. Different stakeholders appraise the problem from a distinct point-of-view without acknowledging the intrinsic, albeit indirect, link between dosing regimen, exposure, target engagement and clinical events.

The incorporation of model-based concepts and pharmacokinetic-pharmacodynamic relationships into the rationale for the design, analysis and interpretation of safety pharmacology and toxicology protocols is vital for the future of screening of novel compounds and for an effective shift in the assessment of safety and acceptable risk in drug discovery and development. More than just enabling a framework for modernisation of outdated methods and techniques, a model-based approach challenges the mainstream scientific views about the role of experimental evidence as the sole basis for the assessment of non-clinical safety; it unravels the strength of inferential methods and evidence synthesis.

In this review, we aim therefore at identifying the pitfalls in current approaches to estimating and predicting safety pharmacology and toxicity in humans. Focus is given to the estimation of safety thresholds and decision making, with special emphasis on the

underlying methodological issues. Our objectives may intersect with the message from other reviews of safety in humans (17,18). However, our concerns go beyond technical aspects of experimental and statistical methods; the objective of the larger research to be presented in this thesis, is to detail improved techniques for data analysis and study design, as well as to illustrate how a mechanism-based approach for risk assessment can formally be applied to support more accurate decision making.

In the subsequent sections, we will cover a wide range of methodological and conceptual issues, starting with low level problems, which usually comprise experimental aspects or relate to the statistical methods. Given their technical nature, implementation of the proposed recommendations requires little effort and can be relatively straightforward, as compared to higher level problems, which involve conceptual features and require a different attitude towards the generation, analysis and interpretation of experimental data regarding safety and toxicity. From a theoretical perspective, different facets of the same problem will be discussed, which relate to four seminal areas of scientific research: 1. **optimisation** (e.g., accuracy, precision), 2. **translation** (e.g., sensitivity, biological substrate, relevance), 3. **analytical construct** (e.g. choice of parameterisation) and 4. **decision criteria** (e.g., acceptable risk level). Each of these points will be addressed separately.

As shown in Figure 1, on the most basic level of the hierarchical tree is the choice of the measure of drug exposure and endpoint selected for the assessment of safety. These issues are compounded by the use of point estimates and by statistical inferences regarding the reporting of safety thresholds. Experimental design considerations in relation to type I and II errors constitute the next level of attention. The drawbacks of the use of empirical approaches as opposed to mechanism-based approaches will be covered. Empiricism here relates to data analysis methods which are primarily descriptive rather than explanatory of the observed phenomena. Of particular interest is the current dichotomisation of the problem using safety thresholds. This will be followed by a critique of allometric scaling to predict exposure in humans and then more generally the manner in which risk is translated into decisions.

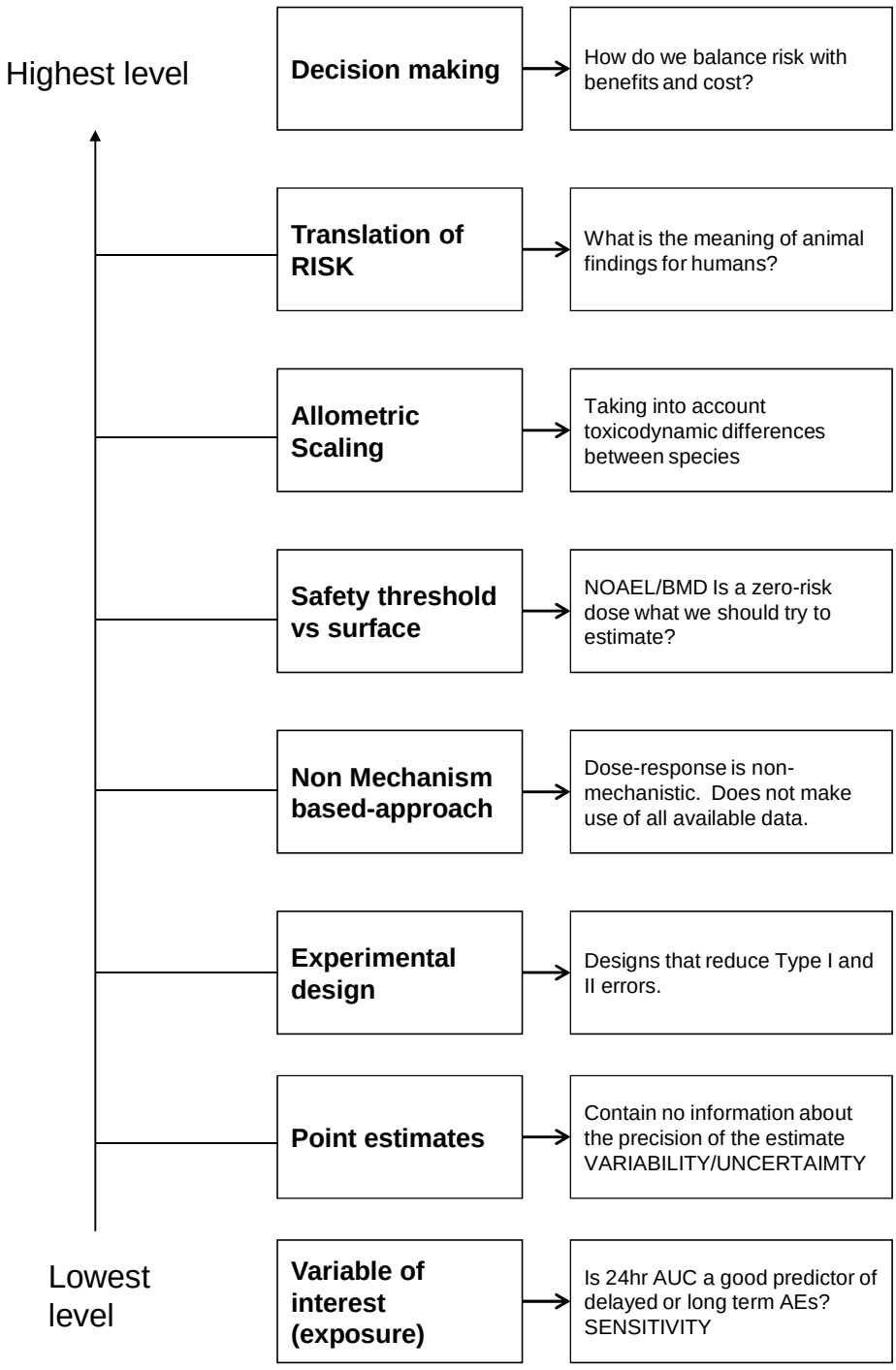


Figure 1: A hierarchical tree describing the different levels and issues underpinning decision making during the assessment of safety and toxicity profile of a new chemical entity.

2. Nonclinical evaluation of safety and toxicity

2.1. Defining variables of interest.

The development of a pharmaceutical is a stepwise process involving an evaluation of both animal and human efficacy and safety information. The goals of the nonclinical safety evaluation generally include a characterisation of toxic effects with respect to target organs, dose dependence, relationship to exposure, and, when appropriate, potential reversibility. This information is used to estimate an initial safe starting dose and dose range for the human trials and to identify parameters for clinical monitoring for potential adverse effects. Toxicity occurs when the drug-induced alteration of biological function overcomes normal repair and homeostatic mechanisms. Toxicity can be measured by its effects on the target (organism, organ, tissue or cell) or indirectly by measuring altered biological function downstream after acute, sub-chronic or chronic exposure to a chemical or biological entity. Drug exposure is then used as a proxy or surrogate for the undesirable effects. It should be noted that an adverse event is any undesirable experience associated with the use of a medical product, irrespective of the evidence of a causal relationship between drug and adverse event. However, from a drug development perspective, different aspects of safety and toxicity need to be evaluated experimentally, which encompass the expected therapeutic and supra-therapeutic dose levels. Although different experimental protocols must be implemented during the development of a new compound, the evaluation of immunotoxicity, genotoxicity, carcinogenicity, phototoxicity, abuse liability and reproductive performance and developmental toxicity are beyond the scope of this review. The nonclinical safety and toxicity studies should be adequate to characterise potential adverse effects that might occur under the conditions of the clinical trial to be supported. Serious nonclinical findings can influence the continuation of the development programme and of clinical trials.

Despite the different protocols for the assessment of safety and toxicity and the myriad of adverse events one may come across, a common practice in this field of research is the assessment of empirical safety thresholds such as the no observed adverse effect level

(NOAEL), which are no more than *qualitative* indicators of acceptable risk. Support for the existence of thresholds has been argued on biological grounds (19-21). The argument is that although any exposure to a chemical will cause some change in the biological system, the change must override homeostatic mechanisms in order for it to be biologically significant. In contrast to the maximum tolerated dose (MTD), which remains the primary endpoint of choice in the evaluation of chronic toxicity, the NOAEL is one of the main indicators of risk in nonclinical safety assessment. Definitions of the NOAEL vary from source to source, however the basis behind all of them is the estimation of "the highest experimental point, without biologically significant adverse effects that are above baseline" (22). In fact, the experimental findings are used to reflect another threshold, i.e., the underlying no adverse event level (NAEL). The calculation involves determination of the lowest observed adverse effect level (LOAEL) which is the lowest observed dosing level for which AEs are recorded. The NOAEL is the dosing level below this. If no LOAEL is found, then the NOAEL cannot be determined. In these cases the LOAEL/10 is sometimes used in place of the NOAEL.

Drug exposure and risk can be represented by a variety of different experimental measures. Usually, in the NOAEL approach, the measures used are dosing level, area-under-concentration-time-curve (AUC) and/or maximum concentration (CMAX). On the other hand, the benchmark dose (BMD) is an alternative to the NOAEL. The method involves the construction of a model of the exposure-AE relationship to predict the dosing level that corresponds to the threshold between non-significant and significant risk of AEs. The quantity is usually expressed as a dose level rather than an AUC or CMAX, but the BMD remains of limited use in Industry (23).

Another common measure is the human equivalent dose (HED), which represents the estimated dose level in humans yielding equivalent drug exposure as observed in animals at the safety threshold (23). In addition, recommendations have been made for the use of the maximum recommended starting dose (MRSD) for the selection of the starting doses in first-in-human studies. The MRSD is believed to minimise the chance of serious adverse events in early clinical studies (7,23). Recently, the minimum anticipated biological effect level

(MABEL) has also been introduced to assist in selection of doses for first in man studies and to supplement existing approaches. MABEL describes the exposure that is anticipated, prior to clinical testing, to produce a minimum biological effect level (24,25).

Given the empirical nature of such safety thresholds, errors in the prediction of safety may arise. Despite the various options, there is still a real safety concern when using these thresholds to extrapolate drug exposure levels from animals to humans and to make inferences from short to long term effects. Unfortunately, instead of pursuing a more mechanistic approach, empirical methods continue to be used. To cope with inaccuracy and poor precision, safety factors, also known as uncertainty factors, have been incorporated on the top of empirical thresholds. Their application in drug development has become widespread (26) and is detailed within the regulatory guidelines. The purpose of such safety factors is to account for variability potentially greater toxicity in humans than predicted by the HED using existing approaches. This is to ensure that the safety threshold is beneath the true threshold. The default safety factor is 10, but it can by modified by considering it as a product of more refined uncertainty factors. These comprise; interspecies uncertainty, UF_A, interindividual uncertainty UF_H, subchronic to chronic uncertainty, UF_S, LOAEL to NOAEL uncertainty, UF_L, and data adequacy UF_D, for when chronic toxicity studies in at least two different species are unavailable (27,28). There is also a modification factor where there is a perceived greater risk of toxicity in humans.

It should be noted that even when safety factors are factored into the estimation of thresholds, the actual risk a treatment represents to humans can be overlooked. Over-conservative attitude may give the wrong perception of caution. Accurate assessment of risk can simply not be performed without some degree of understanding of target engagement and nature of the ligand (i.e., agonistic or antagonistic interaction with the target).

2.2 Measures of drug exposure used as descriptors of acceptable risk

A consequence of the use of safety thresholds is the estimation of drug exposure or dose levels that can be correlated with the adverse events observed beyond that specific threshold, for which the risk for humans is deemed unacceptable. Numerous assumptions are however required to ensure accurate translation of such findings from animals to humans. To be predictive, the exposure levels and the adverse events must reflect pathophysiological processes and pharmacokinetics in humans.

Different measures of exposure are used in reports. The most basic of these is dosing level, which is usually expressed in terms of daily dose (e.g., mg/day). Dose, however may be a poor indicator of response since it does not account for confounders such as bioavailability, differences in metabolic capacity, or other pharmacokinetic processes that alter target exposure despite comparable dose. For this reason, parameters derived from the assessment of systemic drug concentrations are preferred (e.g., AUC and C_{MAX}). The choice for those parameters relies on the assumption that rapid equilibration occurs between systemic circulating drug and the target tissue. Given the fragmented process used for the evaluation of pharmacology and toxicology data, the validity of this assumption is questioned even when evidence from pharmacological and pharmacokinetic data indicates otherwise. Nonlinearity in drug disposition is another important pharmacokinetic aspect which is not accurately captured by the use of dose as a measure of drug exposure. Differences in systemic and target exposure can be large in the case of metabolic saturation, when small increments in dose can produce disproportionately large increases in AUC. This can lead to deceptively safe estimates even if the dose is divided by a safety factor. Conversely, the occurrence of metabolic induction may lead to overly conservative dose selection.

In addition to the aforementioned points, it is also critical to understand the implications of the use of systemic levels as compared to target tissue or target organ exposure. Time-dependent processes take place which cannot be neglected or inferred from conventional measures of exposure. First, one should realise that given that pharmacokinetic equilibration

between plasma and tissue may not always be assumed. Unbound drug concentrations are primarily distributed into tissues. The extent and rate of distribution depend on physicochemical as well as receptor binding properties. The implications of such processes are that irreversible binding, slowly reversible binding and tissue accumulation may not be easily correlated with circulating total concentrations. From a pharmacodynamic perspective, the same considerations must be made when signal transduction and downstream mechanisms are rate limiting for the onset and maintenance of effects (i.e., adverse events). Consequently, the use of AUC and C_{MAX}, expressed over a single day may not accurately reflect the underlying relationship between exposure and adverse event. The implicit assumption that there is a correlation between "daily" drug exposure and risk is suitable mainly for direct and reversible processes; however it is insufficient to account for the complex nature of indirect effects, slowly reversible and irreversible binding.

These complexities can be illustrated by permetrexed-induced neutropenia. Absolute neutrophil count (ANC) is reduced by inhibition of thymidylate synthase, dihydrofolate reductase and glycinamide ribonucleotide formyltransferase (29). The trough of the ANC curve occurs between 8 and 9.6 days after dosing (30), and is followed by an overshoot effect once levels return to baseline (Figure 2). Empirical approaches are in principle able to quantify the PK exposure associated with a particular ANC minimum, however this ignores the complexity of the ANC-curve. The time below a threshold ANC may be a more relevant descriptor of risk and will require a different measure (i.e., parameterisation) of drug exposure. Most importantly, in these circumstances the time course of drug effects (onset, duration and washout) often does not correlate with daily systemic exposures.

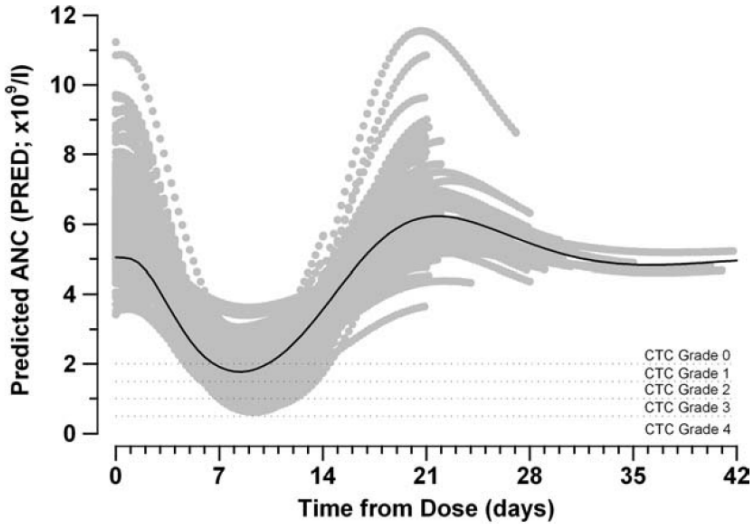


Figure 2. Time course of predicted absolute neutrophil counts (PRED) following 500 mg/m^2 pemetrexed. Lines: Solid black curve the overall ‘‘typical’’ patient in the analysis dataset (i.e., median values for each of the covariates contained in the final PKPD model); gray shading predictions based on the population PKPD model for each of the patients in the analysis dataset, assuming a 500 mg/m^2 dose; dashed horizontal lines hematologic toxicity grades (grade 1 <2, grade 2 <1.5, grade 3 <1, grade 4 <0.5) (30).

Likewise, irreversible binding mechanisms cause drug accumulation at the effect site yielding adverse events that depend primarily on the treatment duration, rather than on daily exposure. Measures that do not capture the cumulative nature of these processes may lead to poor correlation between species. Measures such as cumulative AUC may provide better prediction than 24-hour AUC since the entire dosing history is used. Figure 3 shows an example of such an effect is tardive dyskinesia produced by neuroleptic drugs (31).

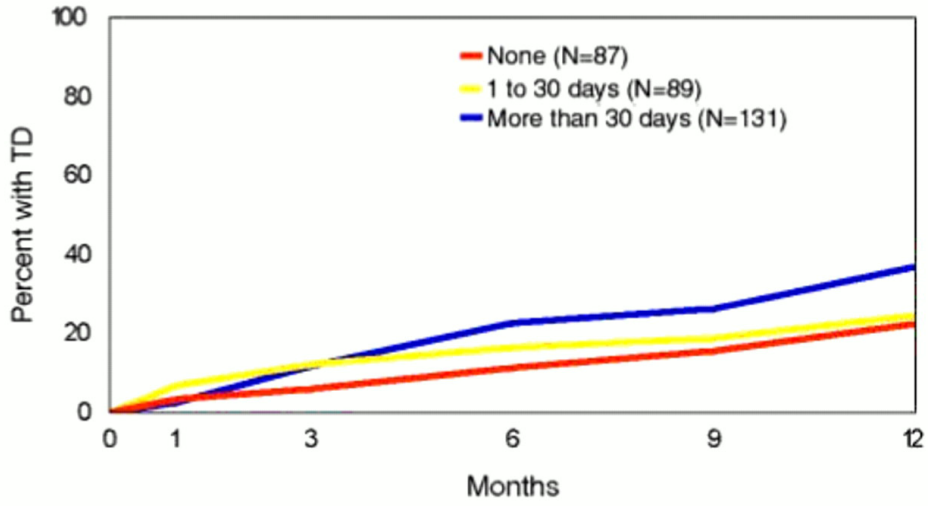


Figure 3. Curve showing incidence of tardive dyskinesia given cumulative neuroleptic exposure. Patients with more than 30 days of neuroleptic use at baseline had a trend for a greater cumulative incidence of tardive dyskinesia than those with 0-30 days of neuroleptic use (31).

Although safety factors have been used to account for possible inaccuracies in the estimates of safety thresholds, there are translational aspects that cannot be factored in by such an empirical approach. A systematic, rational translation of findings across species requires the use of mechanism-based approaches to assess the implications of differences in pharmacokinetics, pharmacodynamics as well as in pathophysiology. Of particular importance is the fact that between species variability in metabolic rate and capacity can lead to completely different safety profiles across specie if metabolites are the moiety underlying adverse events. Likewise, molecules that are substrate to active transporters, carrier-mediated processes and other distribution mechanisms with known species-specific differences will show discrepancies in safety profile.

2.3 Statistical and biological limitations of point estimates

From a statistical perspective, safety thresholds are often presented as point estimates to describe the population. This ignores variability which can be decomposed into two parts; variability associated with estimation methods and real variability in response between and within subjects. There is also lack of best practice in statistical inference. Risk is inferred from toxicology results using statistics that may be imprecise or inaccurate. A statistic is a random variable which is typically a function of the experimental data (such as, e.g., a mean or an observed rate). Statistics are intended to provide an estimate of underlying parameters reflecting physiological processes and/or pharmacokinetics. The implications of such practice can be illustrated by the comparison between *sample* standard deviation and *population* standard deviation. The former is a statistic and the latter is the inferred parameter. The equivalent for NOAEL is the no adverse effect level (NAEL). The term "NOAEL estimate" is a misnomer in that it is the NAEL, which is being estimated by the NOAEL (see figure 4). Based on statistical concepts, it can be shown that meaningful and useful reporting of toxicology findings should be of the estimate of NAEL with its precision (standard errors). However, an empirical approach prevents the estimation of uncertainty in the NAEL.

The use of statistics, in place of model parameters for decision making can treat the estimate as if it were of sufficient precision to give sufficiently narrow confidence intervals. This limitation is believed to be mitigated by the incorporation of safety factors, an assumption which we dispute. As can be seen in Table 1, the parameter precision for the probability of an AE varies from 1587 to 67%, depending on group size and risk. The number of animals in a group exhibiting an adverse event is often reported however, the performance of this estimator is highly dependent on the underlying risk of the AE in question and the sample size.

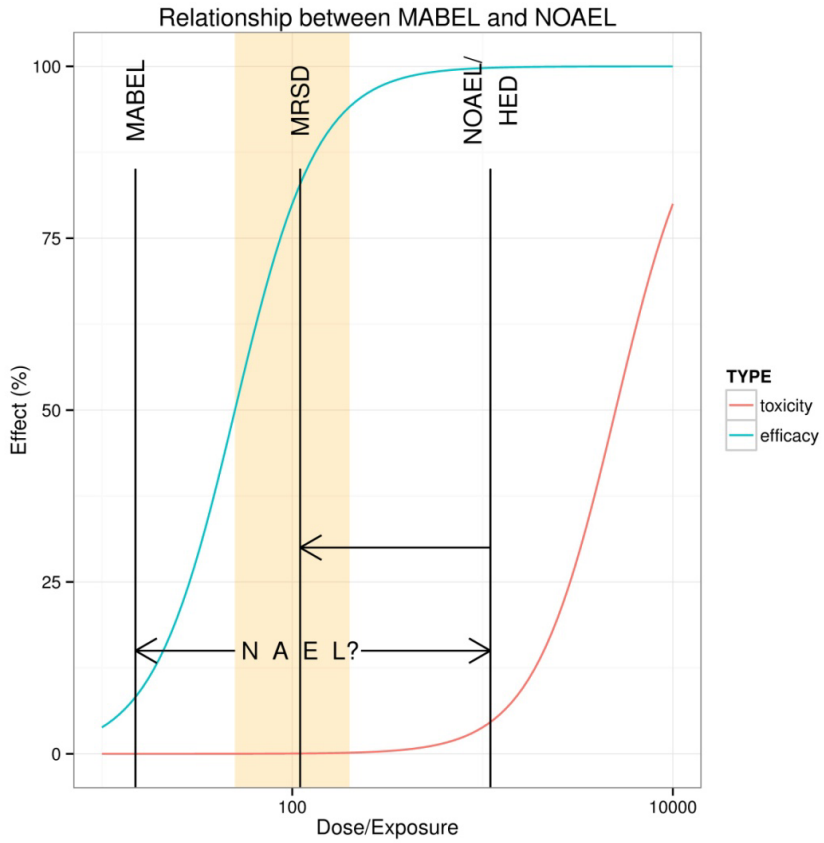


Figure 4. Relationship between MABEL and NOAEL/HED. Shaded region indicates the expected therapeutic range.

Table 1: Parameter precision for probability of adverse events

Risk of AE	n=4	n=8	n=10	n=16	n=20
0.10%	1578.77%	1116.75%	998.08%	790.59%	707.92%
1.00%	497.41%	351.67%	314.71%	248.64%	222.50%
5.00%	217.98%	154.14%	137.83%	108.98%	97.46%
10.00%	150.03%	106.07%	94.87%	74.99%	67.08%

AEs were assumed to be independent binary events. The estimator is the number of animals as a percentage of n. Values depicted coefficients of variation.

The other aspect to missing variability is the real variability in the data. Since sampling in toxicology is often very sparse, exposure levels are calculated from satellite groups which mirror dosing of the animals assigned to the primary treatment group. This ignores real differences that may be present between the two groups. It is equivalent to assuming that all animals have the same exposure and variability in exposure or the underlying physiological processes is not responsible for variability in response.

Given that human variability is typically larger, it is important to understand the role of the different sources of variability. Without quantification of variability and identification of covariates it becomes difficult to predict which groups are more prone to overexposure or more sensitive to adverse events. Furthermore, depending on the actual distribution of drug exposure, the distribution of AEs in this group may not be representative of the risk posed to the overall target population. This is not limited to pharmacokinetics, pharmacodynamic differences have the greater potential for harm and can be more variable than pharmacokinetic differences. In this case, hypersensitive subpopulations can be completely missed. This is the case of abacavir-induced rash and other dose-independent reactions associated with receptor or target polymorphism.

Finally, it should be noted that empirical approaches remain prone to bias. For example, the mean NOAEL is only unbiased if its underlying distribution is symmetrical. This practice ignores that such a summary violates current understanding of pharmacokinetic processes, which are best described by lognormal distributions. Without clear assumptions of the underlying distribution, the choice of measure for central tendency remains unjustifiable and may lead to bias.

2.4 Mechanism-based assessment of safety, toxicity and risk

Whilst the introduction of regulatory policies for the non-clinical evaluation of medicinal products in humans, at a time when understanding about receptor pharmacology and pathophysiology was very limited, partly explains the historical evolution of current

standards and practice in safety and toxicity research, its perpetuation is no longer justifiable. It is evident that the concept of safety thresholds as well as the measures of exposure used as *proxy* for acceptable risk cannot be deemed absolute: they rely upon numerous assumptions, which may not hold true in a considerable number of cases. In principle, information regarding the causal chain between target engagement and adverse events should be used as basis for relevant measures of exposure and risk. This concept can be implemented even in the absence of evidence for the actual target or mechanism underlying a given adverse event or undesirable effect. Sufficient evidence exists to support the use of concentration-effect relationships to identify the rate limiting step in the chain of events from dose to response. In conjunction with tailored experimental protocols and pharmacokinetic-pharmacodynamic modelling, a mechanism-based evaluation of safety findings provides the basis for characterising safety and toxicity. Moreover, it should be noted that safety and toxicity findings may not solely depend on pharmacokinetic drug exposure, but also on the extent of target activation or inhibition, post-receptor amplification and signal transduction processes as well as homeostatic mechanisms. For instance, drug concentrations may be a poor predictor of risk relative to the relevant biomarker concentrations when signal transduction is the rate limiting step for a given response. It is unfortunate that despite the wide discussion regarding the use of biomarkers in the literature (32-35), the focus has primarily been on the assessment of efficacy, not safety.

In addition, experimental protocols and data analysis have not advanced in the same way risk management concepts have evolved over the last decade. Causality has become pivotal for the characterisation of adverse drug reactions, which in contrast to adverse events, are defined as any noxious unintended and undesired effects of a drug that occur at doses used for prevention, diagnosis or treatment. This subtle difference in definition has major consequences for the evaluation of safety, toxicity and risk, including experimental protocol requirements. Rawlins and Thompson devised a classification scheme in 1991, which continues to be the most frequently used in clinical research, which could be used as the basis for the assessment of nonclinical safety. Their scheme, shown in Table 2, defines

adverse drug reactions according to seven different categories, which account for the underlying chain of events. The different categories nicely match the mechanistic classification of biomarkers proposed by Danhof et al., and could form the basis for a new paradigm for the evaluation of nonclinical safety and toxicity (34).

Table 2. Classification of adverse drug reactions, as proposed by Rawlins and Thompson.

Type "A": Predictable, common and related to Pharmacological action of the drug		
Toxicity of overdose:	e.g. hepatic failure	paracetamol
Side effects:	e.g sedation	Antihistaminergic drugs
Secondary effects:	e.g. development of	antibiotic therapy
Drug interaction:	e.g. Theophylline toxicity	erythromycin therapy
Type "B": Unpredictable, uncommon, usually not directly related to the mechanism or pharmacological actions of the drug.		
Intolerance:	e.g. tinnitus	Aspirin
Hypersensitivity:	e.g. anaphylaxis	penicillin
Pseudoallergic:	(Non-Immunological)	radio contrast dye reaction
Idiosyncratic reaction:	e.g. anaemia due to glucose-	anti-oxidant drugs
Type "C": These reactions are associated with long-term drug therapy e.g. Benzodiazepine dependence and Analgesic nephropathy. They are well known and can be anticipated.		
Type "D": These reactions refer to carcinogenic and teratogenic effects. These reactions are delayed in onset and are very rare since extensive mutagenicity and carcinogenicity studies are done before drug is licensed.		
Type "E" : The end of treatment or rebound effects		
Type "F" : Failure of treatment		
Type "G" : due to genetic polymorphism, not immunologically mediated		

As it can be seen from Figure 5, biomarkers can be associated or linked to one of the reaction types in Rawlins and Thompson's classification. Undoubtedly, these concepts allow the causal chain of events to be correlated to the time course of overt symptoms and signs in a quantitative manner. Such an integrated approach is essential for accurate (mechanistic) interpretation of risk in humans. In the next paragraphs, we will highlight how distant these concepts are from the approaches currently used in the assessment of nonclinical safety and toxicity.



Figure 5. Mechanistic classification of biomarkers.

2.5 Minimum Anticipated Biological Effective Level

A first attempt to implement mechanism-based measures of exposure has evolved over the last decade, which relies on the assessment of the minimum anticipated biological effect level (MABEL). In the calculation of the MABEL any biomarker can be used, for example receptor occupancy or even downstream markers such as physiological mediators (25). This has the advantage of allowing measures that correlate to any target-related toxicity (i.e., including off-target or secondary target) when pharmacokinetic processes are not the rate limiting step. The concept relies on the assumption of some knowledge of the putative targets underlying the adverse event to accurately interpret (patho)physiological response and assess causality. However, since the MABEL is defined in terms of biological effect, not toxicity, it is not a measure of risk and not a replacement of the NOAEL. Current practice is therefore to use the NOAEL as a measure of risk to guide maximum doses in dose escalation

studies, but the maximum recommended starting dose in FTIH should now be no higher than both the MABEL and the NOAEL-derived MRSD. If the NOAEL with the addition a safety factors were indeed protective, such a measure would be an unnecessary. Yet, one needs to acknowledge that the MABEL is simply a retrospective risk-mitigation measure that can account for some of the deficiencies of the NOAEL approach.

2.6. Limitations in experimental design

There are methodological aspects that need to be addressed to allow wider use of MABEL or any other mechanism-based measures of 'acceptable risk'. The predictive or prognostic value of statistical correlations depends on satisfying five important criteria, namely: selectivity, specificity, sensitivity, reproducibility and clinical relevance. Currently, despite the characterisation of a correlation between biomarker and response, very little effort has been made to quantify estimators such as false positive and false negative rates. For instance, liver enzyme levels provide an example of a biomarker which has high sensitivity but poor specificity. Interestingly, despite the aforementioned limitations clinical scientists and pathologists will defend the value of ALT, AST and bilirubin as better predictors of risk, as compared to drug exposure. Another aspect of interest is the fact that according to current practice, if e.g., elevated liver enzymes are observed in one individual and acute liver failure in another, an empirical framework ignores the correlation between these adverse events. It may be treated as the same adverse event (i.e., 100% correlation), or a two different adverse events (i.e., uncorrelated). The statistical methods and summary measures of toxicity are unable to account for partial correlation or interaction between events within or between individuals.

So the question is why does one not go further along the causal chain of toxicity for all adverse events, instead of relying on measures of systemic exposure? The answer probably lies in that pharmacokinetics is seen as the primary step along the way for most adverse drug reactions. It is a simple, general purpose measure which fits the criterion of providing

predictive value for many adverse events, despite the exceptions, for which it will perform poorly, with low predictivity.

From a theoretical point of view, it should be highlighted that empirical approaches perform poorly when incidence of a type of adverse event is low (Table 3). This means that pooling data across different types of adverse events is necessary, and this is the root cause behind the choice for a single measure of exposure, rather than more predictive ones.

Table 3: Probability of detection of adverse events with low incidence. Summary data is reflect the occurrence of adverse events according to a Bernoulli random variable. For different incidence rates, value depicts its probability of occurrence given an experimental group size of n.

Risk of adverse events	n=4	n=8	n=10	n=16	n=20
0.10%	0.40%	0.80%	1.00%	1.59%	1.98%
1.00%	3.94%	7.73%	9.56%	14.85%	18.21%
5.00%	18.55%	33.66%	40.13%	55.99%	64.15%
10.00%	34.39%	56.95%	65.13%	81.47%	87.84%

As indicated previously, empirical data analysis does not provide uncertainty estimates to properly account for Type I (false positive) and Type II (false negative). In addition, experimental findings are evaluated in an experiment by experiment basis. This leads to misrepresentation of the estimated population characteristics, which imposes the need for conservative safety factors to account for bias and uncertainty. An immediate consequence of this is illustrated by safety levels identified for tolcapone (36), cerevastin (37), and ximelagatran (38), which were deemed “well-tolerated” at the predefined dose levels, but were later shown to be unsafe (39). It should therefore become clear that the use of the term tolerability ignores the high incidence of false negative results in standard designs.

Based on empirical methods, the absence of adverse events within the experimental group implies that risk is not present all.

Another problem is the fixed design used for in the estimation of the safety thresholds, which relies on a set of arbitrary selection of the dose levels. Consequently, the NOAEL is limited to one of the experimental dose levels. This results in the dose selection having a heavy influence on the precision and accuracy of the NOAEL estimate. Unfortunately, attempts to overcome the uncertainty and bias in the results may prove ineffective even if the number of animal is increased per group. In addition to the dose selection, the duration of the experiments also requires careful consideration and must be factored accordingly into the estimation of safety thresholds. Current approaches consider treatment duration as a constant factor, irrespective of the nature of the underlying adverse event. In general, high doses administered over shorter periods of time are deemed comparable to therapeutic doses administered chronically. This has little pharmacological foundation where time-to-onset may bear little relationship to dose (e.g. neutropenia). At high doses, effects may merely be due to secondary pharmacology. On the other hand, certain effects that can occur at therapeutic levels may be overlooked at higher exposures. Furthermore, if toxicity is delayed, then the likelihood of false negatives will increase if recording of adverse events stops at the end of dosing. A historical example is the case of methylmercury-induced dendritic degeneration in cats (40). Daily dosing for two months results in no differences from control groups, up to month five, when a significant difference becomes evident. If observations had ceased at month two, this effect would have been missed. In brief, the experimental limitations of current approaches can be summarised not only in terms of imprecision and inaccuracy, but also in terms of the lack of integration of the information contained within and between experiments.

From a statistical perspective, the occurrence of an adverse event can be viewed as a multidimensional random process over time, with one dimension for each type of adverse event. In practice, these dimensionalities are reduced to binary processes, leading to loss of information for data arising from continuous processes. Data loss also occurs when all these

binary processes are combined and reduced to a single binary number for each individual: *the animal either had an adverse event or it didn't*. Information about which adverse event occurred, the time-to-onset, duration, frequency and severity is all lost. On top this a further reduction happens at group level whereby the binary numbers for each individual are combined and reduced to a single binary number: *an adverse event occurs at a given dosing level, or it does not*. This approach prevents the use of quantitative methods, as it removes the evidence arising from the number of animals which exhibited adverse events. Data are further reduced by the very definition of NOAEL, which requires only the lowest dose to be considered in the estimation of the NOAEL: *the NOAEL is highest treatment level exhibiting no adverse events*. As a consequence of all the aforementioned steps, important information about the relationship between dose and exposure and adverse events may be lost.

By contrast, an approach which involves longitudinal statistical modelling of continuous and categorical data has the potential use all information in the production of estimates without any loss in information. However, an alternative to the NOAEL, the benchmark dose (BMD) approach has been proposed (41), which permits better use of experimental data. The BMD yields evidence about the entire dose-response curve, rather than a single point. Typically there are also large reductions at an individual and group level, but on a smaller scale. Relevant data across experimental groups are not collated and analysed together (42).

2.7. Additional flaws in the empirical evaluation of safety and toxicity

From a scientific and clinical point of view, one of the main disadvantages of empirical approaches is that extrapolation beyond experimental setting is often unreliable. Paradoxically, the ability to extrapolate or make inferences is central for the evaluation of safety and toxicity. Nonclinical data are generated with the primary objective of data extrapolation in mind.

Another limitation which cannot be easily circumvented is the inability to parameterise risk in a systematic manner, accounting for what is observed and what can be inferred from an intervention, irrespectively of the experimental evidence. Consequently, for instance, one fails to assess the implications of an adverse event arising from two different mechanisms of actions. To make accurate extrapolations, any relevant differences in the mechanism of action must be incorporated into the analysis and interpretation of the data. A similar problem arises in the case of nonlinear kinetics, when extrapolation to dose ranges outside of experimental ranges can lead to very different exposure levels, as compared to those expected from linear kinetics. Hence, it is evident that extrapolations derived from safety factors are doomed to remain inaccurate without further understanding of the mechanisms underlying the overt symptoms and signs.

Lastly, it is important to bear in mind that empirical methods often do not lend themselves well to integrating data and combining results from multiple experiments. This situation forces one to rely on clinical judgment to decide which findings can be deemed relevant. This inflexibility represents another inherent weakness of current approaches for the evaluation of safety, which clashes with one of the primary objectives of the drug development process, i.e., to reduce uncertainty about the safety and efficacy of a compound (43). In theory, more information should to lead to improve precision rather bias.

2.8. Safety threshold vs. risk or hazard surface

Currently, the use of fixed thresholds as a metric of safety ignores the variable nature of continuous processes and potentially prevents accurate interpretation of the underlying phenomena. For example, gastric ulceration is dependent on membrane permeability. Interindividual differences in tissue permeability are perceived as interindividual differences in sensitivity to drug effects, i.e., in the exposure which is required to reach a threshold. Based on current practice, the factor driving such differences often remains obscure. More

sophisticated approaches have been proposed to incorporate toxicodynamic differences through use of a sensitivity parameter (44). However, this suffers from the same weakness as the use of a threshold. Furthermore, thresholds offer no mechanistic basis for extrapolation across species. For example, there is no way to account for interspecies differences in membrane permeability. As such, interspecies differences can only be handled by safety factors.

Another immediate difficulty is the lack of consensus on what is defined as adverse events and how definitions vary across species. These definitions lead to different safety levels, meaning that safety thresholds are sensitive to definitions of events as adverse or non-adverse rather than the risk associated with them. Therefore, it should be noted that even with agreed definitions, the relevance of a threshold for the assessment of risk is questionable since it mostly relates only to the presence of an adverse event, rather than its severity. In this context, the shape and slope of the exposure-risk relationship is an important consideration. Yet, the use of thresholds incurs the danger that risk is treated and thought of as a binary endpoint. Since the only way to truly eliminate risk is to cease the hazard-causing activity, this is at odds with the binary treatment of it. Safety thresholds can also obfuscate more complicate U-shaped or bell-shaped relationships which may be relevant characteristics for consideration in a risk-benefit analysis.

In summary, it should be clear that despite the dichotomous nature of thresholds, all (patho)physiological processes underlying an adverse event are continuous processes. In fact, increasing understanding of the mechanisms underlying drug-target interactions (e.g. receptor pharmacology theory) as well as the identification of downstream pathways (i.e., factors determining post-receptor events) imposes revisiting the utility and relevance of thresholds as basis for the evaluation of drug response, irrespective of whether it involves efficacy or safety. The continuous nature of ligand-target relationships, based upon which target exposure must approach a certain order or magnitude in order to block or transducer a signal, offers the possibility of exploring signal using multidimensional response surfaces, rather than thresholds.

2.9. Translational toxicology: allometric scaling

All the undertaking required to implementing experimental protocols in safety pharmacology and toxicity implies the validity of a set of assumptions regarding the correlation between findings in animals and humans. Unfortunately, these assumptions do not take into account the prerequisite of construct validity to ensure direct comparability of the findings across species.

As indicated previously, uncertainty about differences between species and lack of understanding about the relevance of certain effects in humans, have lead to the introduction of safety factors the estimation of safety thresholds. Whilst many supporters of the approach envisage this as a plausible, cautionary measure, it cannot be ignored that in many cases over-conservatism will prevent the development of compounds that otherwise could be innocuous in humans. The challenge is therefore to identify a mechanistic basis for translating nonclinical safety findings or at least making inferences about drug action based on the results in a different species or experimental system (e.g., in vitro or cell culture). Five different dimensions need to be considered for that purpose: 1) differences in pharmacokinetics (i.e., accounting for physiological processes determining drug absorption, distribution, metabolism and elimination); 2) differences in pharmacodynamics (i.e., accounting for variation or differences in receptor engagement, activation and downstream amplification of the biosignal); 3) differences in homeostasis (i.e., accounting for functional capacity and feedback mechanisms which may compensate for drug-induced changes in physiological processes); 4) differences in response during health vs. disease conditions and 5) differences due to drug delivery properties.

It can be anticipated that accurate assessment of causality is essential for making inferences from one species to another. Furthermore, it is rather evident that in most cases all five dimensions need to be factored in the interpretation of nonclinical findings. However, currently, more focus is given to differences in pharmacokinetics more than any other aspect. As a matter of fact, extrapolation of findings between species often relies on the use of allometric scaling principles (45,46). Allometry requires assumptions about the

relationships between physiological function (e.g., metabolic capacity) and body size. In principle, this concept can also be applied to differences in pharmacodynamics (46,47), but the use of this technique in drug development is usually restricted to pharmacokinetic parameters, and more specifically to volume of distribution and clearance.

Despite its wide use in drug development, one needs to be aware of the limitations allometric methods represent to the prediction of pharmacokinetics and pharmacodynamics in humans. The first point relates to the unawareness of the underlying differences between-species. For example, total clearance can result from multiple routes; metabolism by oxidation and glucuronidation, biliary excretion, and/or renal excretion. The use of allometry assumes that when multiple physiological processes are involved, processes are scaled solely based on size differences and processes that do not scale well are considered clinically irrelevant (48). Biliary excretion is known not to scale well due to the role of ABC transporters expression levels. As such the decision to use scaling is dependent on an overall judgement of its ability to be scaled. For volume of distribution, the assumption is that distribution of drug outside system circulation occurs primarily due to passive diffusion; active transport is not accounted for either. Scaling via the more realistic physiologically based pharmacokinetic (PBPK) models (49), has been shown to account for both size-dependent and size-independent differences.

The second source of error in allometric scaling relate to the use of allometry as a monolithic extrapolation strategy: allometric relationships, even if correct, only relate to size differences between species. It is functionally equivalent to assuming that a human is a large rodent or another non-clinical species. Furthermore, the scaling of parameters assumes that size-related factors influencing systemic exposure are the only important covariate relationships governing drug effects.

Despite the clear flaw in this approach, the evaluation of alternative methods for scaling or translating pharmacokinetics and PKPD relationships remains limited. In fact, size-independent differences compose a much larger part of the differences in

pharmacodynamics and this is not accounted for with allometry. Paradoxically, there is also support for the view that size-independent differences are usually small given that adverse events in humans are predictable in the majority of cases (75%) from information obtained from preclinical experiments (50). This leads to the apparent conclusion that mechanisms of action in animals are similar to humans, however potentially serious differences may exist (51). A related problem is that clinical outcome is dependent on the underlying disease process, which may be different between species. Differences in baseline (physiological) response and in variability due to disease conditions in humans can confound the measurement of drug-induced effects, as compared to animals. Likewise, differences in target distribution can also complicate the interpretation and translation of non-clinical findings. For example, anaphylaxis is observed in the intestine and liver of rats, but in humans these symptoms are primarily observed in the lungs and blood vessels (52). The translational gap becomes even larger if one considers psychiatric or other neurological adverse events, which may not be detected in animals.

2.10. Translation of Risk

Translation of the risk associated with the experimental evidence observed in animals is the ultimate step triggering decisions related to nonclinical safety and toxicity of a novel molecule. Thus far, expert judgment is used by decision makers, which ultimately consists in the use of qualitative criteria for the assessment of risk. These criteria informally include some measure of overall uncertainty, but such an approach makes it difficult to understand the propagation of uncertainty. For instance, to infer that small physiological changes to the binding levels across species can lead to large changes the estimates of safe exposure. Clearly, accurate judgment is even more difficult when dependent on parameters for which uncertainty is unknown or not quantifiable.

Whilst the aforementioned issues have been recognised as important, regulators remain reluctant about the use of quantitative methods for risk assessment (53). There are various

reasons why qualitative risk assessment has been advocated over quantitative methods. However, many of the argued limitations do not necessarily apply when more modern statistical techniques are considered. We will address some of these points later in the next section, where model-based approaches are discussed.

The danger with a qualitative analysis is that the extent of any overall benefit will be left to human intuition. Informed decisions involve taking both benefits and risks of the drug into account. Yet, the consideration of risks and benefits based on safety thresholds is dependent on the nature of the risks in question and as such do not account for the underlying mechanisms, which in turn could be used for subsequent clinical interpretation. An encompassing inferential method is needed which accounts for underlying mechanisms and balance them against benefits. Most importantly, decision making regarding risk should include the contribution of historical data in a statistically and clinically formal manner.

3. Non-linear mixed effects modelling

The use of model-based methods has the ability to address many of the aforementioned criticisms pertinent to the design and analysis of safety pharmacology and toxicology protocols. Nonlinear-mixed effects models are a particular class of models that allow one to handle a variety of parameterisations by integrating stochastic and deterministic components of a problem. Although such models are often referred to as population models, they provide insight at the individual level, separating real variability from estimation uncertainty. They contain the necessary complexities required to assess risk in a manner that translates into scientifically rigorous decisions. In pharmacokinetic-pharmacodynamic data analysis, the use of a parametric approach based on nonlinear mixed-effects models provides a tool for handling repeated-measurement data in which the relationship between the explanatory variable and the response variable can be described by a single function, allowing model parameters to differ between subjects (54). An

immediate advantage of the approach is that the within-subject variability for a given individual can be distinguished from the differences between subjects even in the absence of balanced or frequent sampling of the data.

In hierarchical modelling, the term "mixed" refers to the use of both fixed effects (characterising the typical individual in the population) and random effects (describing the parameter distribution). The latter are divided into two levels: the difference between the individual prediction and the observation (residual error) and the variability between subjects (BSV). There may also be circumstances in which individual parameters vary longitudinally between occasions, randomly or due to some unknown physiological process. In such cases, a third level of variability can be introduced, i.e. the inter-occasion variability (IOV).

The general structure of a hierarchical model is as follows:

$$y_{ijk} = f\left(X_{ijk}, P_{ik}\right) + \varepsilon_{ijk},\ \ \varepsilon_{ijk} \sim N(0, \sigma^2) \qquad \text{Eq. 1}$$

where y_{ijk} is the j^{th} observation at occasion k in individual i. $f(\ )$ is typically a nonlinear function of individual parameter P_{ik} and independent variables X_{ijk}. In PKPD modelling, $f(\ )$ is usually then individual prediction of the observation. Independent variables are usually time, dose or drug exposure and demographic covariates. The ε_{ijk} forms the residual variability with variance σ^2. When the variance is independent of $f\left(X_{ijk}, P_{ik}\right)$, the model is said to have additive variability. On the other hand, when σ is proportional to $f(\ )$, we have a proportional error model (55).

For the i_{th} individual, the individual parameters P_{ik} can by the expression:

$$P_{ik} = \theta \cdot e^{\eta_i},\ \ \eta_i \sim N(0, \omega^2) \quad \text{Eq. 2}$$

This describes a log-normal variation of the individual parameter P, which has a typical value, Θ. The η_i and k_i are the random effects describing the differences between the typical

(population) value and the individual parameter value. η_i is assumed to be normally distributed with mean zero and variance ω^2.

Among other applications, the use of hierarchical models is justified and appropriate when the data available per individual are sparse. In addition, it is recognised as the most effective method to perform meta-analysis of data arising from different studies and to incorporate prior knowledge to the estimation of model parameters. It allows one to adjust for different variances (e.g. presence of influential factor in a given subgroup in the population) and to explore confounding correlations, when the design of the study correlates with the outcome (e.g., effect of weight vs. sex).

3.1. Estimation methods

The field statistical modelling field has developed well-established parameter estimation methods which provide the means not only to estimate the most likely value of the parameters given the data, but also to quantify uncertainty and correlation in estimated parameters and model (mis)specification. This ultimately provides us the opportunity to account for limited information and gaps in our knowledge. For example, if there is little information on the relationship between level of target occupancy and target activation, the corresponding parameters will have an appropriately high uncertainty. This feature is particularly relevant for the estimation and translation of risk as uncertainty can be propagated as high imprecision in exposure-risk relationships. Moreover, the calculation of the propagation of model uncertainty to uncertainty in the risk-benefit profile offers the prospect of efficient data collection.

The standard method for parameter estimation for nonlinear mixed effects models has been the maximum likelihood approach (56-59). This is where parameters are treated as random variables with distribution governed by the likelihood function $p(\mathbf{y}|\boldsymbol{\Theta})$, which represents the probability of the total data arising given the value of the parameters. The reported value for each parameter is the parameter at the maximum of the distribution, and associated

uncertainty given by the variance of the distribution. No data reduction is required; each raw data point directly informs parameter estimation thereby making maximal use of the available data. When multiple studies have been performed in populations which share common physiological processes or treatments, datasets may be aggregated to support integrated analyses across these studies. Furthermore, model-based analysis can handle multiple types of observations (e.g., pharmacokinetics and pharmacodynamics) as well as multiple data types (e.g., continuous and categorical).

Of particular interest for the assessment of safety and toxicity is the possibility of applying extensions of the maximum likelihood, which enable mathematically rigorous incorporation of prior parameter information (e.g., receptor occupancy or blood to plasma binding ratio in vitro to describe in vivo data). The two main methods for achieving this are the penalised likelihood method (49,60) and Bayesian estimation (61). It should also be noted that the advent of exact likelihood methods such as expectation maximisation (EM) methods (62) has provided increased reliability of PKPD analyses, especially in the presence of sparse data, often available from general toxicity protocols.

We should also emphasise that in the context of safety pharmacology and toxicity studies, trial optimisation represents proper adherence to the three R's (reduction, refinement and replacement). When prior information is available for class-specific parameters, a model-based analysis may benefit from this allowing for a reduction experimental cohort sizes or burden to animals. This is possible because model-based analyses are inferential in nature.

3.2. Model parameterisation: empirical vs. mechanistic models

Despite the increasing number of modelling examples in biomedical and pharmaceutical research, the use of pharmacokinetic and pharmacokinetic-pharmacodynamic models has remained primarily descriptive. However, the application of such models for the evaluation of safety requires further consideration of its biological plausibility and predictive or prognostic value. For example, instead of using a simple compartmental model to describe

the observed phases of drug elimination, one may need to consider a physiologically-based pharmacokinetic model (PBPK) (63). Such models can be developed by integrating prior *in vitro* data and literature information.

On the other hand, it is not unusual for components of PKPD models to be statistically correlated to some degree. Therefore, it is important that when identifying at-risk subpopulations based on collected data, covariate selection is guided by a mechanistic or physiological evaluation. There are several methods that allow such an approach (64-67). More recently, these methods have also been applied to describe disease processes (33). Statistically, these models include a response variable that characterises the disease status and its progression over time.

3.3. Simulations, experimental design and optimisation

Model predictions, simulated outside the experimental context are extrapolations subject to model specification bias. Since our primary goal is to show the relevance of such models to analyse data arising from pre-clinical species and eventually from healthy subjects to assess safety and toxicity in patients who will be receiving these drugs, this point is of special importance. In PKPD modelling, computer simulation involves using statistical models to predict the behaviour of the biological system described by the model (68). Clinical trial simulations (CTS) i.e. computer simulation of trials, allows for the investigation of the impact of different design characteristics on the outcome of a trial. It can also be used to investigate the implications of uncertainty and variability in pharmacokinetic and pharmacological processes for recruited individuals, thereby allowing the prior assessment of the robustness of the protocol to known uncertainty and variability (69). More generally, in a CTS it is possible to test the influence of any modelling assumption and design factor beforehand (Figure 6).

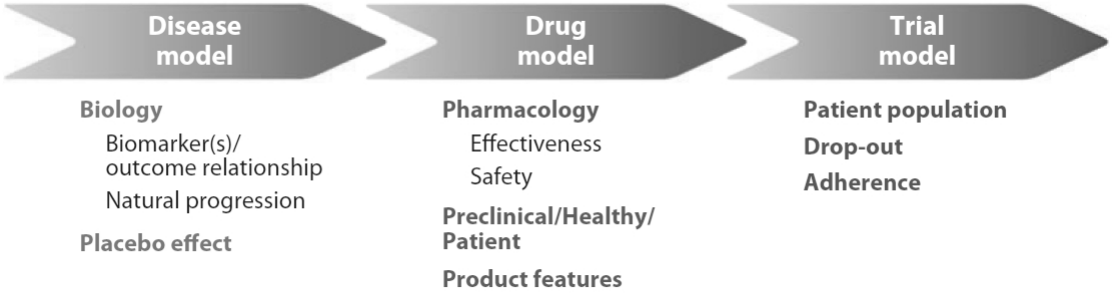


Figure 6: The diagram depicts the major components of a clinical trial simulation (CTS). In model-based drug development, CTS can be used to characterise the interactions between drug and disease, enabling among other things the assessment of disease-modifying effects, dose selection and covariate effects. In conjunction with a trial model, CTS allows the evaluation of such interactions, taking into account uncertainty and trial design factors, including the implications of different statistical methods for the analysis of the data.

Trial design can also benefit from the use of optimal design methodology. The goal of optimal design, specifically the procedure known as D-optimality, is to determine design variables (such as sampling times and dose selection) that optimise the expected information content (usually by maximising the determinant of the Fisher Information Matrix (FIM)) within the desired resource constraints. A variety of software programs exist purpose built for the estimation of PK/PD models (70). Optimal sampling schedules for toxicity experiments can help increase the precision by which drug specific parameters can be estimated and/or reduce the burden to animals by minimising the number of samples needed. This is desirable from an ethical and scientific perspective, as poor experimental design is known to result in biased estimates. Among other advantages, optimal sampling may facilitate the collection of biomarkers in conjunction with pharmacokinetic data when blood volume is limited.

Conclusions

High attrition rates due to a poor safety profile combined with inability to correctly identify risk demand revisiting of concepts and modernisation of the approaches currently used for the assessment of toxicity. Current practices fail to support decision making on multiple levels. Firstly, the parameterisation of drug exposure and available metrics of risk are often justified by historical precedent rather than by an informed scientific rationale. These measures are assumed to be predictive of drug effects in humans, despite the fact that in many cases known pharmacokinetic and pharmacological drug properties contradict such assumptions. Evidence clearly shows that empirical protocols remain primarily descriptive rather than explanatory of the observed phenomena and are therefore unsuitable for extrapolation, an important point to consider when analysing and interpreting safety pharmacology and toxicology data. Moreover, statistically, the use of point estimates and thresholds prevents understanding of the consequences of between subject variability and identification of at-risk subpopulations. Additionally, type I and II errors are also not accounted for in the design or analysis of toxicity data, both of which are critical informed decision making.

In summary, our review has highlighted the implications of empirical data generation for the evaluation of safety and toxicity during drug development. A shift in paradigm was proposed to ensure that pharmacological concepts are incorporated into the evaluation of safety and toxicity. Moreover, we indicate the urgent need to integrate historical evidence, so that findings across species can be effectively translated. Based on historical examples, we have shown some important challenges for the early characterisation of the safety profile of a new molecule and discuss how model-based methodologies can be applied for better design and analysis of experimental protocols. From a methodological perspective, nonlinear-mixed effects modelling is recommended as a tool to account for such requirements. Its use in the evaluation of pharmacokinetics (PK) and pharmacokinetic-pharmacodynamic relationships (PKPD) has enabled the advance of quantitative approaches

in pharmacological research in recent decades. Comparable benefits can be anticipated for the assessment of safety and toxicity.

Reference List

1. Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nat.Rev.Drug Discov. **3**, 711-715.

2. U.S.Food and Drug Administration (CDER). Innovation or stagnation? Challenge and opportunity on the critical path to new medical products. 2004.
URL:
http://www.fda.gov/scienceresearch/specialtopics/criticalpathinitiative/criticalpathopportunitiesreports/ucm077262.htm

3. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul.Toxicol.Pharmacol. **32**, 56-67.

4. Russell, W. M. S. and BURCH, R. L. The principles of humane experimental technique. 1-4-1959. London: Methuen & Co. Ltd.

5. Suntharalingam, G., Perry, M. R., Ward, S., Brett, S. J., Castello-Cortes, A., Brunner, M. D., and Panoskaltsis, N. (2006). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. N.Engl.J.Med. **355**, 1018-1028.

6. Expert Scientific Group. Expert Scientific Group on Phase One Clinical Trials. 2006.

7. European Medicines Agency (CHMP). Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products. 2007.

URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002988.pdf

8. Kenter, M. J., and Cohen, A. F. (2006). Establishing risk of human experimentation with drugs: lessons from TGN1412. Lancet **368**, 1387-1391.

9. Dowsing, T., and Kendall, M. J. (2007). The Northwick Park tragedy--protecting healthy volunteers in future first-in-man trials. J.Clin.Pharm.Ther. **32**, 203-207.

10. Stebbings, R., Poole, S., and Thorpe, R. (2009). Safety of biologics, lessons learnt from TGN1412. Curr.Opin.Biotechnol. **20**, 673-677.

11. Breckenridge, A. (2010). Regulatory challenges, reimbursement, and risk-benefit assessment. Clin.Pharmacol.Ther. **88**, 153-154.

12. Krumholz, H. M., Ross, J. S., Presler, A. H., and Egilman, D. S. (2007). What have we learnt from Vioxx? BMJ **334**, 120-123.

13. Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000). Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N.Engl.J.Med. **343**, 1520-8, 2.

14. Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002). Timing of new black box warnings and withdrawals for prescription medications. JAMA **287**, 2215-2220.

15. Cross, J., Lee, H., Westelinck, A., Nelson, J., Grudzinskas, C., and Peck, C. (2002). Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980-1999. Pharmacoepidemiol.Drug Saf **11**, 439-446.

16. Heerdink, E. R., Urquhart, J., and Leufkens, H. G. (2002). Changes in prescribed drug doses after market introduction. Pharmacoepidemiol.Drug Saf **11**, 447-453.

17. Lowe, P. J., Hijazi, Y., Luttringer, O., Yin, H., Sarangapani, R., and Howard, D. (2007). On the anticipation of the human dose in first-in-man trials from preclinical and prior clinical information in early drug development. Xenobiotica **37**, 1331-1354.

18. Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., and Wurtzen, G. (2002). Mathematical modelling and quantitative methods. Food Chem.Toxicol. **40**, 283-326.

19. Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A. G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., and Younes, M. (2002). Hazard characterisation of chemicals in food and diet. dose response, mechanisms and extrapolation issues. Food Chem.Toxicol. **40**, 237-282.

20. Slob, W. (1999). Thresholds in toxicology and risk assessment. International Journal of Toxicology **18**, 345-367.

21. Kroes, R., Kleiner, J., and Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. Toxicol.Sci. **86**, 226-230.

22. Dorato, M. A., and Engelhardt, J. A. (2005). The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul.Toxicol.Pharmacol. **42**, 265-274.

23. U.S.Food and Drug Administration (CDER). Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. 2005.
URL: http://www.fda.gov/downloads/Drugs/Guidance/UCM078932.pdf

24. Agoram, B. M. (2009). Use of pharmacokinetic/ pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics. Br.J.Clin.Pharmacol. **67**, 153-160.

25. Muller, P. Y., Milton, M., Lloyd, P., Sims, J., and Brennan, F. R. (2009). The minimum anticipated biological effect level (MABEL) for selection of first human dose in clinical trials with monoclonal antibodies. Curr.Opin.Biotechnol. **20**, 722-729.

26. Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Regul.Toxicol.Pharmacol. **3**, 224-238.

27. Swartout, J. C., Price, P. S., Dourson, M. L., Carlson-Lynch, H. L., and Keenan, R. E. (1998). A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. **18**, 271-282.

28. Gaylor, D. W., and Kodell, R. L. (2000). Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal. **20**, 245-250.

29. Shih, C., Gosset, L., and Gates, S. e. a. (1996). LY231514 and its polyglutamates exhibit potent inhibition against both human dihydrofolate reductase (DHFR) and thymidylate synthase (TS): multiple folate enzyme inhibition. Annals of Oncology.

30. Latz, J. E., Rusthoven, J. J., Karlsson, M. O., Ghosh, A., and Johnson, R. D. (2006). Clinical application of a semimechanistic-physiologic population PK/PD model for neutropenia following pemetrexed therapy. Cancer Chemother.Pharmacol. **57**, 427-435.

31. Jeste, D. V., Lacro, J. P., Palmer, B., Rockwell, E., Harris, M. J., and Caligiuri, M. P. (1999). Incidence of tardive dyskinesia in early stages of low-dose treatment with typical neuroleptics in older patients. Am.J.Psychiatry **156**, 309-311.

32. Danhof, M., de Jongh, J., de Lange, E. C., Della, P. O., Ploeger, B. A., and Voskuyl, R. A. (2007). Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. Annu.Rev.Pharmacol.Toxicol. **47**, 357-400.

33. Danhof, M., de Lange, E. C., Della Pasqua, O. E., Ploeger, B. A., and Voskuyl, R. A. (2008). Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. Trends Pharmacol.Sci. **29**, 186-191.

34. Danhof, M., Alvan, G., Dahl, S. G., Kuhlmann, J., and Paintaud, G. (2005). Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. Pharm.Res. **22**, 1432-1437.

35. Rolan, P., Danhof, M., Stanski, D., and Peck, C. (2007). Current issues relating to drug safety especially with regard to the use of biomarkers: a meeting report and progress update. Eur.J.Pharm.Sci. **30**, 107-112.

36. Jorga, K. M., Fotteler, B., Heizmann, P., and Zurcher, G. (1998). Pharmacokinetics and pharmacodynamics after oral and intravenous administration of tolcapone, a novel adjunct to Parkinson's disease therapy. Eur.J.Clin.Pharmacol. **54**, 443-447.

37. Bakri, R., Wang, J., Wierzbicki, A. S., and Goldsmith, D. (2003). Cerivastatin monotherapy-induced muscle weakness, rhabdomyolysis and acute renal failure. Int.J.Cardiol. **91**, 107-109.

38. Wernevik, L. C., Nystrom, P., Johnsson, G., Nakanishi, T., and Eriksson, U. G. (2006). Pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor ximelagatran in young healthy Japanese men. Clin.Pharmacokinet. **45**, 77-84.

39. Cohen, A. (2007). Should we tolerate tolerability as an objective in early drug development? Br.J.Clin.Pharmacol. **64**, 249-252.

40. Oliveira, R. B., Gomes-Leal, W., do-Nascimento, J. L., and Picanco-Diniz, C. W. (1998). Methylmercury intoxication and histochemical demonstration of NADPH-diaphorase activity in the striate cortex of adult cats. Braz.J.Med.Biol.Res. **31**, 1157-1161.

41. Barnes, D. G., Daston, G. P., Evans, J. S., Jarabek, A. M., Kavlock, R. J., Kimmel, C. A., Park, C., and Spitzer, H. L. (1995). Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul.Toxicol.Pharmacol. **21**, 296-306.

42. Budtz-Jorgensen, E., Grandjean, P., Keiding, N., White, R. F., and Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol.Lett. **112-113**, 193-199.

43. Lalonde, R. L., Kowalski, K. G., Hutmacher, M. M., Ewy, W., Nichols, D. J., Milligan, P. A., Corrigan, B. W., Lockwood, P. A., Marshall, S. A., Benincosa, L. J., Tensfeldt, T. G., Parivar, K., Amantea, M., Glue, P., Koide, H., and Miller, R. (2007). Model-based drug development. Clin.Pharmacol.Ther. **82**, 21-32.

44. Kodell, R. L., and Chen, J. J. (2007). On the use of hierarchical probabilistic models for characterizing and managing uncertainty in risk/safety assessment. Risk Anal. **27**, 433-437.

45. West, G. B., Brown, J. H., and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. Science **276**, 122-126.

46. Zuideveld, K. P., Van der Graaf, P. H., Peletier, L. A., and Danhof, M. (2007). Allometric scaling of pharmacodynamic responses: application to 5-Ht1A receptor mediated responses from rat to man. Pharm.Res. **24**, 2031-2039.

47. Mager, D. E., Woo, S., and Jusko, W. J. (2009). Scaling pharmacodynamics from in vitro and preclinical animal studies to humans. Drug Metab Pharmacokinet. **24**, 16-24.

48. Bachmann, K., Pardoe, D., and White, D. (1996). Scaling basic toxicokinetic parameters from rat to man. Environ.Health Perspect. **104**, 400-407.

49. Langdon, G., Gueorguieva, I., Aarons, L., and Karlsson, M. (2007). Linking preclinical and clinical whole-body physiologically based pharmacokinetic models with prior distributions in NONMEM. Eur.J.Clin.Pharmacol. **63**, 485-498.

50. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA **279**, 1200-1205.

51. Ju, C., and Uetrecht, J. P. (2002). Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. Curr.Drug Metab **3**, 367-377.

52. Haley, P. J. (2003). Species differences in the structure and function of the immune system. Toxicology **188**, 49-71.

53. European Medicines Agency (CHMP). Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007.
URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf

54. Davidian, M., and Giltinan, D. M. (1993). Analysis of repeated measurement data using nonlinear mixed effects models. Chemom.Intell.LAb.Syst. **20**, 1-24.

55. Karlsson, M. O., Beal, S. L., and Sheiner, L. B. (1995). Three new residual error models for population PK/PD analyses. J.Pharmacokinet.Biopharm. **23**, 651-672.

56. Beal, S. L. (1984). Population pharmacokinetic data and parameter estimation based on their first two statistical moments. Drug Metab Rev. **15**, 173-193.

57. Lindstrom, M. L., and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. Biometrics **46**, 673-687.

58. Pinheiro, J. C., and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. J.Comput.Graph.Stat. **4**, 12-35.

59. Baverel, P. G., Savic, R. M., Wilkins, J. J., and Karlsson, M. O. (2009). Evaluation of the nonparametric estimation method in NONMEM VI: application to real data. J Pharmacokinet.Pharmacodyn. **36**, 297-315.

60. Gisleskog, P. O., Karlsson, M. O., and Beal, S. L. (2002). Use of prior information to stabilize a population data analysis. J.Pharmacokinet.Pharmacodyn. **29**, 473-505.

61. Lunn, D. J., Best, N., Thomas, A., Wakefield, J., and Spiegelhalter, D. (2002). Bayesian analysis of population PK/PD models: general concepts and software. J.Pharmacokinet.Pharmacodyn. **29**, 271-307.

62. Kuhn E., and Lavielle M. (2004). Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis **49**, 1020-1038.

63. Bois, F. Y. (2010). Physiologically based modelling and prediction of drug interactions. Basic Clin.Pharmacol.Toxicol. **106**, 154-161.

64. Lunn, D. J. (2008). Automated covariate selection and Bayesian model averaging in population PK/PD models. J.Pharmacokinet.Pharmacodyn. **35**, 85-100.

65. Wade, J. R., Beal, S. L., and Sambol, N. C. (1994). Interaction between structural, statistical, and covariate models in population pharmacokinetic analysis. J Pharmacokinet.Biopharm. **22**, 165-177.

66. Mentre, F., and Mallet, A. (1994). Handling covariates in population pharmacokinetics. Int.J Biomed.Comput **36**, 25-33.

67. Mandema, J. W., Verotta, D., and Sheiner, L. B. (1992). Building population pharmacokinetic--pharmacodynamic models. I. Models for covariate effects. J Pharmacokinet.Biopharm. **20**, 511-528.

68. Holford, N. H., Kimko, H. C., Monteleone, J. P., and Peck, C. C. (2000). Simulation of clinical trials. Annu.Rev.Pharmacol.Toxicol. **40**, 209-234.

69. Holford, N., Ma, S. C., and Ploeger, B. A. (2010). Clinical trial simulation: a review. Clin.Pharmacol.Ther. **88**, 166-182.

70. Mentre, F., Duffull, S., Gueorguieva, I., Hooker, A., Leonov, S., Ogungbenro, K., and

Retout, S. Software for optimal design in population pharmacokinetics and

pharmacodynamics: a comparison. PAGE. 2007.

CHAPTER 2

Scope and intent of investigation

Historically, the evaluation of safety pharmacology and toxicity of drugs has largely relied on research in animal models, of which results have been used to extrapolate to potentially harmful events in humans. The research in these models has been developed to evaluate specific toxicological endpoints, (such as oral, dermal and ocular toxicity, immunotoxicity, genotoxicity, reproductive and developmental toxicity and carcinogenicity) rather than specifically designed to understand the exposure response relationships associated with the anticipated adverse event or toxicity (1). Furthermore, even if one considers the information obtained from these experiments useful, they are low throughput and inconsistently predictive of human pharmacology and pathophysiology. Some of these limitations persist in spite of the recognition of toxicokinetics as an important part of the safety assessment (2-3).

More recently, several major new initiatives have begun to utilise *in vitro* methods and a variety of new technologies to develop *in vitro* signatures and computational models predictive of *in vivo* response. These initiatives provide insight and tools to identify a battery of *in vitro* assays to detect perturbations in cellular pathways that are expected to contribute to or result in adverse health effects (4,5). Furthermore, these initiatives represent a welcome movement away from traditional *in vivo* high-dose hazard studies (6,7).

Despite such a continuing improvement in methods to characterise the safety and toxicity of novel medicines, uptake of these new approaches by regulatory agencies remains limited, with quantitative pharmacology concepts still being rarely applied to address clinical and regulatory questions on the safety pharmacology and toxicity of a novel compound. Evidence of the relevance of such concepts has been highlighted with the introduction of structure-activity relationships (SAR) in the absence of adequate toxicity data on the

chemical under certain circumstances, such as when the extent of exposure of humans is extremely low and toxicokinetic data cannot be easily generated (8). Clearly, the lack of a stronger pharmacological basis for the assessment of safety has prevented the implementation of a model-based approach aimed at the characterisation in a strict quantitative manner of the relationship between drug exposure and effects. To date, efforts have been limited to physiologically-based pharmacokinetic modelling, but it is mostly applied to environmental toxicology, rather than to pharmaceutical R&D (9).

Irrespective of the urgently needed changes in regulatory guidance, methodologies that support the translation and prediction of safety pharmacology and toxicity in humans are still required. To this purpose, more than novel experimental protocols and technologies are required. We strongly believe in integrative approaches that enable efficient use of available evidence and facilitate the assessment of pharmacokinetic-pharmacodynamic relationships. Of special importance is the possibility to evaluate and predict long term or rare adverse events, which continue to contribute to high attrition in drug discovery and development (10). In light of the known limitations of current experimental protocols and the implications they represent for hazard characterisation in humans, a range of different approaches is necessary to ensure that the appropriate endpoint is detected and risk evaluated in a precise and accurate manner.

The scientific and regulatory communities should acknowledge that most toxicity tests, as currently designed, are aimed solely at hazard identification at supratherapeutic levels. Data produced using current testing guidelines are not always suitable for robust mathematical exposure–response modelling. We recognise therefore that adequate characterisation of the exposure–response relationship requires a number of doses giving a range of different response levels. On the other hand, mathematical modelling of the exposure–response relationship would represent an important improvement to the risk assessment process.

Here we tackle a number of issues that need to be considered during the course of drug discovery and development to ensure more efficient use of the evidence on safety

pharmacology and toxicity which is generated. Four central questions will form the basis for the work to be presented in the subsequent chapters in this thesis:

1. Can current experimental protocols for safety pharmacology and toxicology evaluation be optimised to support the characterisation of pharmacokinetic-pharmacodynamic relationships?

2. Does a meta-analytical approach based on nonlinear mixed effects modelling provide more precise and accurate estimates of safety thresholds than current methodologies?

3. Can mechanism-based models be used for accurate inferences about safe drug exposure for low frequency, delayed (long term) or rare adverse events?

4. Should biomarkers be used in conjunction with pharmacokinetic data to enable accurate estimation of the safe drug exposure (and consequently of safety thresholds) during chronic therapy?

Our work is presented in a way that both conceptual and practical issues are addressed concurrently. After revisiting the requirements for the implementation of quantitative pharmacology concepts in the evaluation of safety pharmacology and toxicology, we highlight how existing protocols should be redesigned to obtain accurate results from the modelling and emphasise that an appropriate design might even result in a reduction in the total number of animals studied. Moreover, we show that biomarker data may allow translation of the external dose to an internal dose (or target-organ dose), as it reflects a compounds pharmacology. In fact, using naproxen as a paradigm compound for the acute and chronic effects of cyclo-oxygenase inhibition, we explore how biomarkers could be used to provide a full pharmacologically-based exposure-response model, i.e., a PBPKPD model. Our endeavour is complemented by further insight into the implications of modelling for risk prediction purposes, as described by logistic, hazard models. Clear recommendations are provided about the requirements for future refinements regarding the characterisation of

exposure–response relationships, which need to account for the extent of uncertainty and variability in modelling and simulation output.

Section I: General introduction

In **Chapter 1**, we have described the problems with existing practices from a methodological point of view and highlighted the value of mechanism-based PKPD modelling as a tool for the evaluation of safety pharmacology and toxicology. From a methodological perspective, we show that the parameterisation of drug exposure and available metrics of risk are often justified by historical precedent rather than by an informed scientific rationale. These measures are assumed to be predictive of drug effects in humans, despite the fact that in many cases known pharmacokinetic and pharmacological drug properties contradict such assumptions. Evidence clearly shows that empirical protocols remain primarily descriptive rather than explanatory of the observed phenomena and are therefore unsuitable for extrapolation, which is an important point to consider when analysing and interpreting safety pharmacology and toxicology data. Moreover, statistically, the use of point estimates and thresholds prevents understanding of the consequences of between subject variability and identification of at-risk subpopulations. Additionally, type I and II errors are also not accounted for in the design or analysis of toxicity data, both of which are critical informed decision making.

A shift in paradigm is proposed to 1) ensure that pharmacological concepts are incorporated into the evaluation of safety and toxicity; 2) facilitate the integration of historical evidence and thereby the translation of findings across species; and 3) promote the use of experimental protocols tailored to address specific safety and toxicity questions. Three important components have been identified, which will form the framework proposed throughout this thesis., namely, model based optimisation of experimental design, data integration, and incorporation of biomarkers as a way towards the implementation of a

pharmacology-based approach for the characterisation of safety and toxicity in drug discovery and development.

Of particular interest for us is to demonstrate that inefficiencies in the experimental design violate the principle of the 3 Rs (reduction, refinement and replacement) (11,12). Optimality concepts are available that could be implemented even when terminal sampling procedures are used, as is the case of histopathological measures. Using examples, we show that the poor predictive value of experimental data reflects the failure in anticipating the biological consequences of target engagement, i.e., in establishing the correlation between target-related events and drug exposure, as defined by the evidence of pharmacokinetic-pharmacodynamic (PKPD) relationships.

Based on the requirements for the implementation of a pharmacology-based approach, specific issues have been identified which underpin the scope and intent of the investigations described here **in Chapter 2**. Nonlinear-mixed effects modelling will be recommended as a tool for protocol optimisation and knowledge integration (i.e., evidence synthesis). Its use in the evaluation of pharmacokinetics and pharmacokinetic-pharmacodynamic relationships has enabled the advance of quantitative approaches in pharmacological research in recent decades. As shown in the subsequent chapters of this thesis, comparable benefits can be anticipated for the assessment of safety and toxicity.

The overall focus of the work presented in the following sections of this thesis is therefore to illustrate how the proposed methodology can be applied prospectively during the evaluation of a novel molecule in the early stages of development. We also attempt to demonstrate the need and added value of an integrative approach to predict potential long term AEs with respect to performance metrics commonly used in safety pharmacology and toxicology experiments. Where possible, proposals to amend study protocols are kept to a minimum to facilitate acceptance of the proposal by industry and regulatory bodies.

Section II: Conceptual framework

In the section, different aspects, including advantages and limitations of a model-based approach are evaluated. Of interest is the fact that despite the increased attention to the importance of toxicokinetics, the extrapolation and prediction of a safe exposure range in humans from preclinical experiments continues to be based on the assessment of empirical safety thresholds, in particular the no observed adverse effect level (NOAEL), which is a *qualitative* indicator of acceptable risk. 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. As a consequence, safety thresholds are primarily derived from inferences about the putative pharmacokinetic profiles in the actual treatment group. Such an experimental setting has far reaching consequences for the assessment of risk, given the assumption that inter-individual differences are implied to result from residual variability. Pharmacokinetic and pharmacokinetic-pharmacodynamic parameters are treated as point estimates. Factors such as within- and between-subject variability or uncertainty in estimation are not accounted for. 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, indirect or delayed effects, which may be physiologically more relevant depending on type of drug and the mechanism of action.

Therefore, focus is initially given to the opportunities for optimisation of experimental protocols supporting the characterisation of pharmacokinetic properties at therapeutic and supratherapeutic levels. In **Chapter 3**, we show that the estimation of safety thresholds such as the NOAEL can be optimised (13). Using simulation scenarios in which hypothetical compounds with different disposition properties are evaluated, our analysis shows the feasibility and relative performance of a model-based analysis for the characterisation of systemic exposure as compared to empirical, non-compartmental analysis (NCA) methods

currently used in general toxicity protocols. Simulation scenarios are used to illustrate which changes are required in experimental protocols with respect to standard non-compartmental analysis. Expected bias and precision of parameters of interest, such as systemic exposure (AUC) and peak concentrations (C_{MAX}), are then computed with both methodologies. In addition, we also assess the predictive accuracy of cumulative exposure estimates up to three months beyond the study duration. It should be noted that such an extrapolation represents an important advantage of model-based methods, which cannot be derived by descriptive methods such as non-compartmental analysis. Overall, the scope of this evaluation was to show that, despite the need for an iterative process, modelling provides the basis for experimental protocol optimisation.

Given the assumption of unbiased parameter estimation when using a model-based approach for the characterisation of pharmacokinetic properties, a natural question arises with respect to the principle of the 3 Rs in pre-clinical research. Irrespective of the availability of alternative methods that allow evaluation of drug disposition properties *in vitro*, can experimental protocols be optimised to ensure a significant reduction in the number of animals required, whilst still providing sufficient estimation precision for measures of exposure such as AUC and C_{MAX}?

This question is addressed in **Chapter 4**, where an important methodological challenge is overcome, namely the possibility to optimise secondary pharmacokinetic parameters such as AUC and C_{MAX}. In contrast to existing optimality software and algorithms, which support optimisation of experimental design with respect to primary parameter precision, we show that secondary parameters can be optimised without the resource-intensive procedures imposed by D-optimality. Using a range of hypothetical drugs with different pharmacokinetic profiles, we illustrate the implementation of optimisation procedures to select sampling times and define the minimal number of animals per treatment group. By combining the expected Fisher information matrix (FIM) with simulations from uncertainty, this exercise ultimately shows that the precision of secondary parameters can be assessed and minimally sufficient designs obtained, in line with the principle of the 3 Rs. The method

is computationally inexpensive and can provide potential savings to numbers of animals without compromising study objectives.

Still within the scope of protocol optimisation, we also explore the implications of introducing biomarkers into the evaluation of a drug's safety toxicity profile, as biomarkers of pharmacological activity can be crucial for the prediction of long term adverse events and toxicity. In contrast to traditional protocols, which imply a direct relationship between observed systemic exposure and adverse events, in **Chapter 5** we apply a model-based approach to characterise the PKPD correlations and the time course of biomarker responses associated with long-term safety. Our evaluation also compares the analysis of biomarker data based on standard non-compartmental methods. In brief, we propose the collection of biomarkers at the scheduled pharmacokinetic sampling points to facilitate the characterisation of pharmacokinetic-pharmacodynamic relationships.

Study data are simulated for four hypothetical drugs, each with a different mechanism of delayed toxicity. For the purposes of our evaluation, delayed toxicity was parameterised in terms of i) an indirect response mechanism, ii) an indirect response mechanism preceded by biophase equilibration, iii) cumulative effects as a consequence of chronic dosing and iv) formation of a toxic metabolite after repeated dosing. Given the often unknown mechanism of toxicity, model misspecification is also considered to ensure that accurate conclusions are drawn from experimental protocols. Finally, bias and precision of parameter estimates were used as metrics of interest to compare model-based and non-compartmental methods.

The utility of model-based approaches to predict the risk of adverse events from preclinical toxicology protocols is subsequently explored in **Chapter 6**, where pharmacokinetic, biomarker and adverse event data are integrated into a PKPD model. In this investigation, simulation scenarios are used to generate drug-induced adverse events for reversible and irreversible drug effects according to three different pharmacological mechanisms (direct, indirect, and irreversible binding). To ensure real-life conditions, assumptions are made with regard to 1) the presence of background adverse events, including the situation in which

drug-induced and background adverse events are indistinguishable from each other, 2) events occur with low frequency, including rare events, 3) the symptoms evolve over time but can only be detected once per animal during histological examination and 4) adverse events are described by binary data.

Whereas typical toxicology experiments are designed to show evidence of safety thresholds, it can be anticipated that they may not fully support the identification of the underlying mechanisms for adverse drug reactions. Therefore, we show the importance of prior information and more specifically of background rates from placebo and control-treated animals. We also make the effort to quantify model and parameter uncertainty as the basis for subsequent risk assessment. At the same time, we show the technical challenges for characterising exposure-response relationships, which make the validity and reproducibility of models derived by empirical experimentation questionable for predictive purposes.

Section III: Case study and practical application

The third part of the thesis aims to illustrate the implementation of experimental protocols suitable for model-based analysis. Given the ongoing debate of the benefit-risk balance of chronic treatment with non-steroidal anti-inflammatory drugs, naproxen is used as a paradigm compound with known acute and chronic toxicities. Naproxen is a non-selective cyclo-oxygenase inhibitor, whose activity results in the suppression of pro-inflammatory mediators such as prostaglandins and thromboxanes (14). By considering the requirements for a suitable experimental protocol, we also attempt to identify practical challenges and difficulties that one may face for the prospective use of the methodology.

From a clinical pharmacology perspective, the rationale for selecting naproxen is based on the differences in housekeeping function of both isozymes and their contribution to the inflammatory response in acute and chronic inflammatory conditions (15-17). Unfortunately, at present the dose selection of COX inhibitors disregards whether maximum, long-lasting blockade of either enzyme systems is strictly required for anti-inflammatory, analgesic

response and how its pharmacology relates the observed adverse events (18). These considerations become essential when evaluating the side effects associated with long term use of COX inhibitors, which include gastric and cardiac adverse events. Whilst the lack of selectivity of naproxen and the evidence for distinct mechanisms underpinning acute effects (such as bleeding and ulceration) and long term effects (such as renal and cardiovascular damage) have evolved over the years and might not have been understood at time of the development of the compound, such understanding makes it quite didactic in that it demonstrates how human safety and toxicity may require characterisation of drug effects at exposure levels corresponding to the therapeutic doses. Toxicity, and in particular, long term safety is not a matter of supra-therapeutic exposure: it may be determined by time-dependent pharmacological activity.

Using a typical toxicology protocol in rats, in **Chapter 7** naproxen, a non-selective cyclo-oxygenase inhibitor is used as paradigm compound to demonstrate the concept of biomarker-guided safety assessment (19-21). Using pharmacokinetic-pharmacodynamic techniques, we subsequently illustrate how modelling and simulation techniques can be used to ensure accurate estimation of the safe dose levels of naproxen after chronic exposure. Furthermore, the pharmacokinetics of naproxen is evaluated in conjunction thromboxane (TXB_2) and prostaglandin (PGE_2) over short, moderate and long-term treatment. It is assumed that gastrointestinal bleeding is due continuous COX-1 inhibition, whereas ulceration results primarily from the suppression of COX-2, which is known to have an important role in the repair of gastric mucosa. PK and biomarker findings are then integrated with experimental data from historical protocols and published literature to ensure characterisation of drug properties at putative therapeutic levels. From a methodological perspective, it is our endeavour to quantify the impact of nonlinearity in pharmacokinetics and in biomarker response. Given the wealth of clinical data from the published literature, we also take the opportunity to evaluate the predictive value of nonclinical findings and explore whether species differences exist for naproxen effects on TXB_2 and PGE_2.

This evaluation is complemented in **Chapter 8** by further integrating the histological data obtained at completion of treatment to the observed biomarker effects. Here we emphasise the need for evidence synthesis to quantify and explain the risks associated with long term drug exposure. Clearly, efforts are required to ensure the availability of tissue- and mechanism-specific data for accurate interpretation of acute and long term safety findings. Such an objective may be hampered by the use of empirical experimental protocols, as they render the extrapolation of findings across species rather difficult, preventing accurate translation of the pharmacological properties to man. In the current investigation we show therefore how pharmacokinetic-pharmacodynamic (PKPD) modelling can be used to unravel the relationship between chronic drug exposure, pharmacodynamic effects and overt symptoms and signs. The concept is illustrated by the correlation between naproxen concentrations, PGE_2 and TXB_2 inhibition, and gastric ulceration in rats. Through the use of bootstrapping procedures in combination with covariate analysis, we show how model diagnostics can be used for model selection as well as for potential identification of the explanatory factors for the observed gastric ulceration.

Section IV: Conclusions and Perspectives

An overview of the results and conclusions drawn from the various chapters is provided in **Chapter 9.** Most importantly, recommendations are provided for physiologically based quantitative toxicity assessment. Here we also summarise the answers to the initial questions set up at the beginning of this chapter, which underpin the research developed throughout the thesis. We anticipate that the examples used in Section II will make clear that there are numerous opportunities for optimisation of experimental protocols for safety pharmacology and toxicology. The approach should also shed light on the advantages of including biomarkers and characterising PKPD relationships, instead of relying solely on safety thresholds.

Lastly, the issues identified in the various simulation scenarios and the challenges experienced during the implementation of an integrated experimental protocol are discussed. Our findings should make clear that inferences about safe exposure as well as the risk associated with long term use of a compound cannot be achieved by scattered empirical experimentation. Given the increased relevance of evidence synthesis as the basis for decision-making within regulatory and clinical settings, we expect that some of the meta-analytical elements presented across the various simulation scenarios will become embedded into daily practice in safety pharmacology and toxicology. Irrespective of the advancements in the understanding of the mechanisms of toxicity, we envisage that a pharmacology or biomarker-based approach will always be required to allow accurate inferences about safe drug exposure for low frequency, delayed (long term) and rare adverse events. Future perspectives are then presented taking into account ongoing developments in the field of systems pharmacology and its relevance for the prediction of drug toxicity and risk assessment in humans. The work is concluded with a new question being asked with regard to the scientific and ethical basis for current experimental designs in toxicology.

References

1 ICH, 2009. CPMP/ICH/286/95. ICH Topic M 3 (R2). Non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. Step 4. Note for guidance on non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals (cpmp/ich/286/95). June 2009. EMEA, London.

2. ICH, 1995. CPMP/ICH/384/95. ICH Topic S 3 A. Toxicokinetics: a guidance for assessing systemic exposure in toxicology studies. Note for guidance on toxicokinetics: a guidance for assessing systemic exposure in toxicology studies. ICH Harmonised Tripartite Guideline. June 1995. EMEA, London

3. Ploemen JP, Kramer H, Krajnc EI, Martin I. (2007) Use of toxicokinetic data in preclinical safety assessment: a toxicologic pathologist perspective. Toxicol. Pathol. 35, 834–837.

4. Andersen ME, Krewski D. (2009) Toxicity testing in the 21st century: bringing the vision to life. Toxicol. Sci. 107, 324–330.

5 Krewski D, Andersen ME, Mantus E, Zeise L. (2009) Toxicity testing in the 21st century: implications for human. health risk assessment. Risk Anal. 29, 474–479.

6. Magda G, Vaudano E, Goldman M. (2012) The rational use of animals in drug development: contribution of the innovative medicines initiative. Altern Lab Anim. 40(6):307-12.

7. Goldman M, Compton C, Mittleman BB. (2013) Public-private partnerships as driving forces in the quest for innovative medicines.Clin Transl Med. 2(1):2.

8. Edler L, Poirier K, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K, Würtzen G. (2002) Mathematical modelling and quantitative methods. Food Chem Toxicol. 40(2-3):283-326.

9. Yang,R, Thomas RS, Gustafson DL, Campain J, Benjamin SA, Verhaar HJ, Mumtaz MM. (1998) Approaches to developing alternative and predictive toxicology based on PBPK/PD and QSAR modeling. Environmental Health Perspectives 106, 1385–1393.

10. Kola I, Landis J. (2004)Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 3(8):711-5.

11. Olson H, Betton G, Robinson G, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, van Deun K, Smith P, Berger B, Heller A. (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul.Toxicol.Pharmacol. 32:56-67.

12. Russell WMS, Burch RL. The principles of humane experimental technique. 1959. London: Methuen & Co. Ltd.

13. Dorato MA, Engelhardt JA. (2005) The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 42(3):265-74.

14. Vane J, Botting R. (1998) Mechanism of Action of Nonsteroidal Anti-inflammatory Drugs Am J Med. 104(3A):2S–8S.

15. Mukherjee D, Nissen SE, Topol EJ. (2001) RIsk of cardiovascular events associated with selective cox-2 inhibitors. JAMA 286(8):954-9.

16. Khan M, Fraser A. (2012) Cox-2 inhibitors and the risk of cardiovascular thrombotic events. Ir Med J. 105(4):119-21.

17. Atukorala I, Hunter DJ. (2013) Valdecoxib : the rise and fall of a COX-2 inhibitor. Expert Opin Pharmacother. 14(8):1077-86.

18. Huntjens DR, Danhof M, Della Pasqua OE. (2005) PK/PD correlations and biomarkers in the development of COX-2 inhibitors. Rheumatology 44:846-59.

19. Berger SI, Iyengar R. (2011) Role of systems pharmacology in understanding drug adverse events. Wiley Interdiscip. Rev Syst Biol Med. 3, 129-35

20. Bai JP, Abernethy DR. (2013) Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu Rev Pharmacol Toxicol. 53, 451-73.

21. Bai JP, Fontana RJ, Price ND, Sangar V. (2013) Systems pharmacology modeling: an approach to improving drug safety. Biopharm Drug Dispos. doi: 10.1002/bdd.1871.

SECTION 2: CONCEPTUAL FRAMEWORK

CHAPTER 3

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

Tarjinder Sahota, Meindert Danhof and Oscar Della Pasqua

Submitted for publication (Journal of Pharmacokinetics and Pharmacodynamics)

Abstract

Purpose: Current toxicity protocols relate measures of systemic exposure (i.e. AUC, Cmax) as obtained by non-compartmental analysis to observed toxicity. A complicating factor in this practice is potential bias in the exposure estimates. 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 model*l*ing for the evaluation of toxicokinetics and b) to assess the bias and accuracy of systemic exposure for each method.

Methods: Simulation scenarios were evaluated, which mimic standard toxicology protocols in rodents. To ensure differences in pharmacokinetic properties were accounted for, hypothetical drugs with varying disposition properties were considered, including a one-compartment pharmacokinetics with linear and nonlinear elimination as well as a two-compartment pharmacokinetics. Data analysis was performed using non-compartmental methods and nonlinear mixed effects modelling. Exposure levels were summarised as area under the concentration vs. time curve (AUC), peak concentrations (Cmax) and time above a predefined threshold (TAT). Results were then compared with the reference values to assess the bias and precision of parameter estimates.

Results*:* Population pharmacokinetic modelling yields higher accuracy and precision of estimates for AUC, CMAX and TAT irrespective of group or treatment duration, as compared with non-compartmental analysis. Moreover, population pharmacokinetics modelling constitutes a basis for PKPD based analysis of safety outcomes.

Conclusions*:* Despite the focus of toxicology 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 in toxicology provides insight into variability and should be considered for predicting safe exposure in humans.

Abbreviations:

AUC - area under the concentration vs. time curve

Cmax – peak concentrations

PD - pharmacodynamics

PK – pharmacokinetics

TAT – time above a concentration threshold

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) (Figure 1). In addition, the use of a 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).

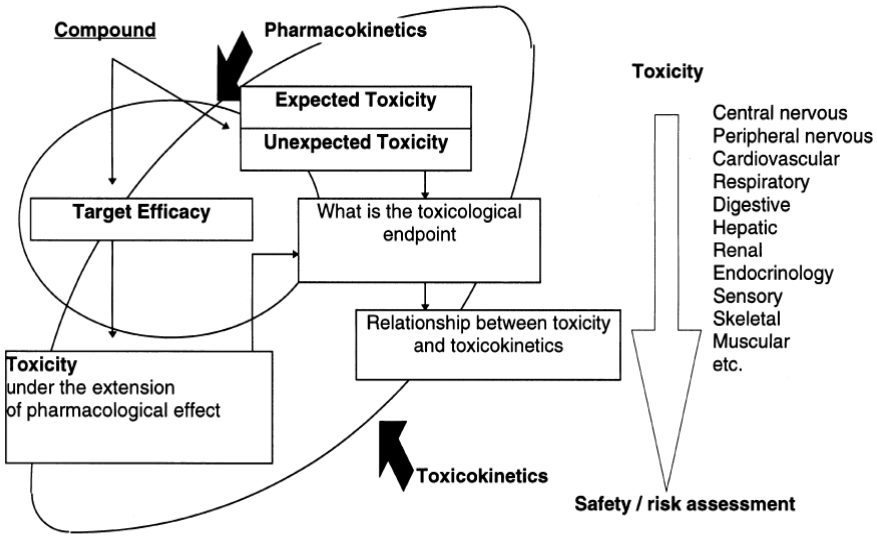


Figure 1 – Safety risk assessment based on toxicokinetics and pharmacological basis for target-related adverse events. Target efficacy: target engagement endpoint on in vitro or in vivo screening. Reprinted with permission from Horii, 1998 (3).

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 (Figure 2) (71). 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 (19-21), the NOAEL has been used to establish 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 (22). 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 vs. time curve (AUC) and peak concentrations (Cmax), 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 (72). The scope of these guidances is to ensure that the systemic exposure achieved in animals is assessed in conjunction with its relationship to dose level and the time course of the toxicity or adverse events (Figure 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 toxicity studies.

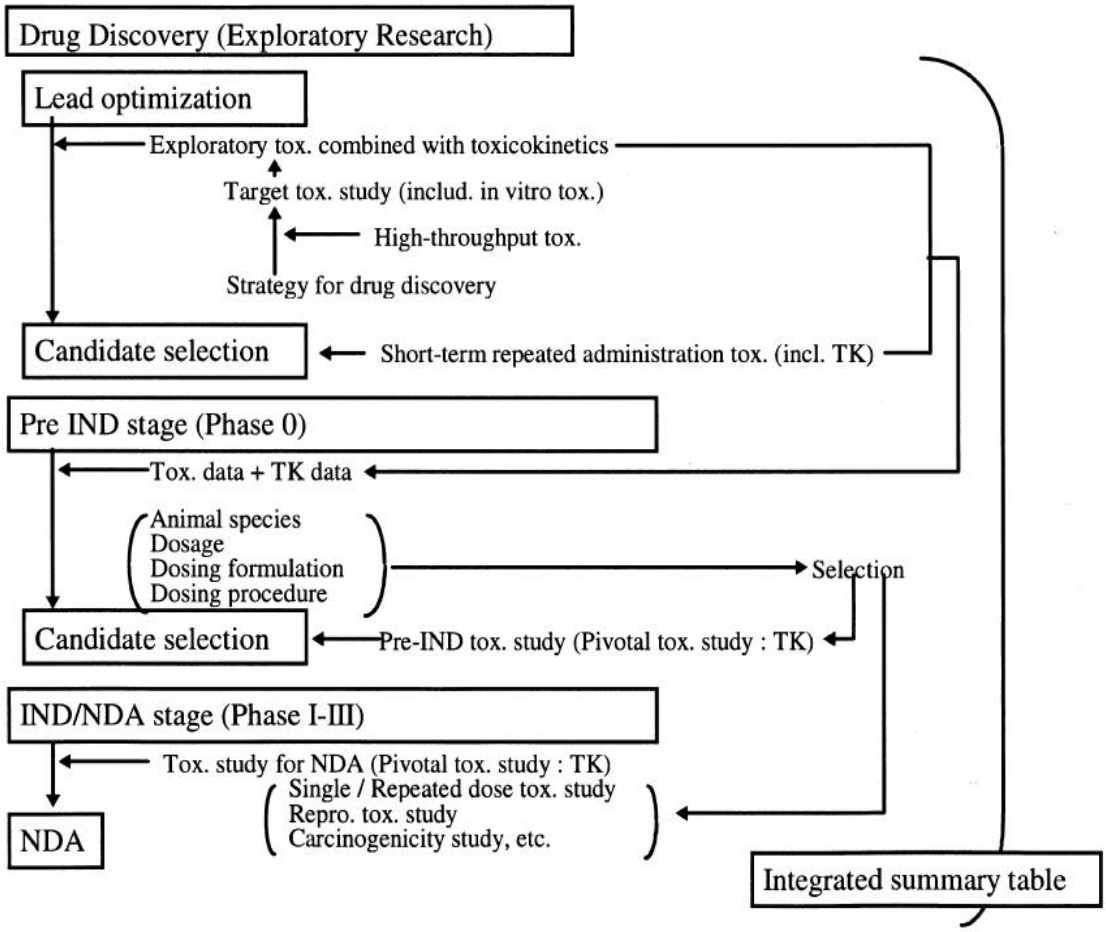


Figure 2 – TK; toxicokinetic study in drug-development process. IND; investigational new drug application, NDA; new drug application. Reprinted with permission from Horii, 1998 (3). General toxicity data used for supporting early clinical trials is gathered in the pre-IND stage. After IND submission, the FDA will confirm whether adequate evidence of safety has been generated for human trials.

Whilst the scope and intent of such guidance are well described since 1994, when it was introduced by 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 (73).

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 (74). 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 cumulative effects (e.g. lead toxicity, aminoglycosides) (18-19). 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 (20-22), 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 (23). 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 (24).

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 hypothetical drugs with different disposition properties are evaluated.

Methods

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. Experimental procedures were defined according to current guidelines for the assessment of toxicity. Given the pre-defined pharmacokinetic parameters used in the simulations, true exposure and biomarker levels for each individual animal were computed in accordance with Table 3. These values were subsequently used as reference for comparison of the methodologies and assessment of bias and precision of the parameters of interest. The sampled data obtained according to a sampling matrix was analysed using non-compartmental methods and by nonlinear mixed effects modelling. All simulations and fitting procedures described below were performed in NONMEM 7.1 (25). Data manipulation and statistical and graphical summaries were performed in R 3.0.0 (26).

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 nonlinear (Michaelis-Menten) elimination as well as a two-compartment pharmacokinetics. Parameter values for each scenario are shown in Table 1. In all scenarios, residual variability was assumed to be 15%. For the purposes of this exercise, we have assumed a homogeneous population, avoiding the need to explore covariate relationships in any of the models.

Table 1 - Pharmacokinetic models used to assess the implications of molecules with varying disposition properties.

Model A: One-compartment model (1 CMT)

Parameter	**Pop Estimate**	**BSV**
KA	13.46 h^{-1}	50%
V	49.4 ml/kg	16%
CL	2.72 ml/hr	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**
Vmax	2.72 mg/hr	20%
Km	1 mg/ml	-
Ka	13.46 h^{-1}	50%
V	49.4 ml/kg	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 two weeks.

Parameter	Pop Estimate	BSV
Ka	0.55 h^{-1}	50%
V	49.4 ml/kg	16%
CL	2.72 ml/hr	20%
K12	0.3 h^{-1}	-
K21	0.05 h^{-1}	32%

Experimental design: A summary of the sampling schemes and experimental conditions is shown in 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 (one week, one month or three months). Satellite groups each were used to characterize 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 hours 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 was sampled an equal number of times. Figure 3 shows PK observations from a typical dataset.

Table 2 - Experimental design of satellite groups in a general toxicity study with serial and composite sampling

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

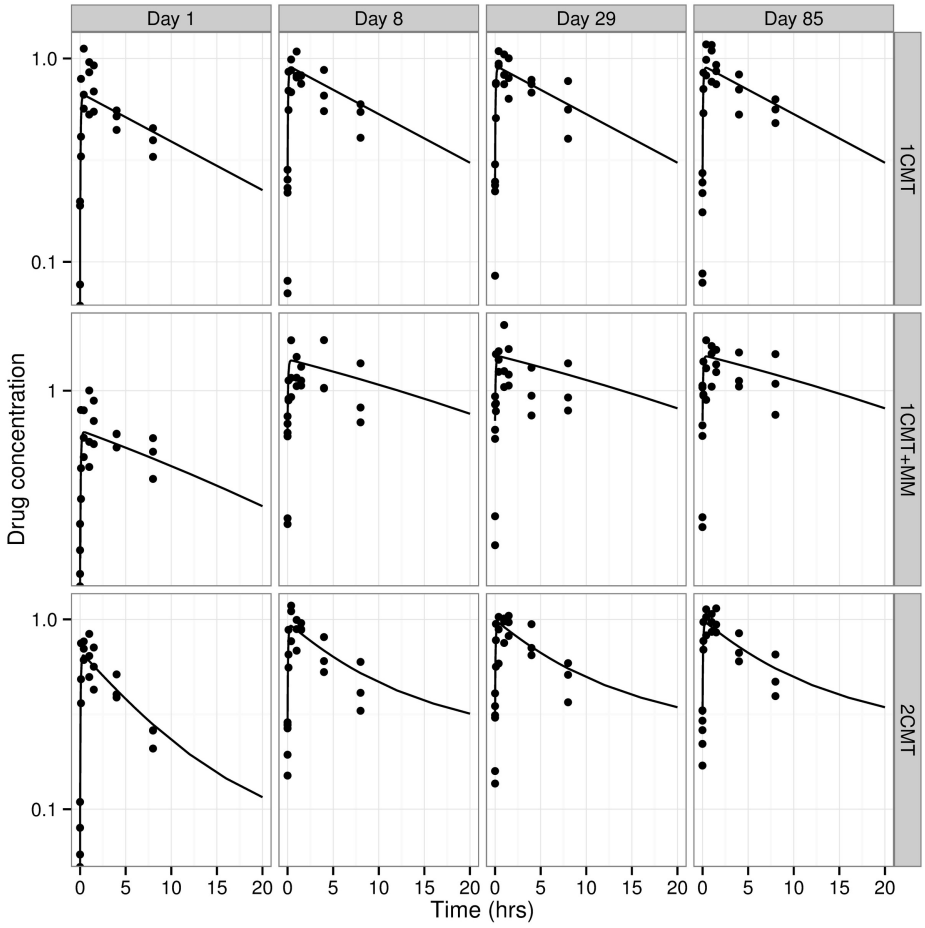


Figure 3 – Overview of a simulated dataset for each of the experimental scenarios, in which 3 animals/sampling time point design are assessed. Dots represent simulated concentrations at the pre-defined sampling times, whereas the solid black line depicts the population predicted profile after a dose of 30 mg/kg for hypothetical drugs with different pharmacokinetic characteristics.

Exposure calculations: Five different measures of exposure were used for calculation of exposure, using the predicted concentration profiles obtained from the models used for simulation. These exposure measures can be seen alongside the formula used for their calculation in Table 3. The simulations (n = 200 replicates) were performed assuming repeat dosing for up to six months (three months beyond the treatment duration presented the investigated studies) in order to evaluate the implications of longer periods of drug exposure.

Table 3 - Individual predicted drug concentrations are denoted by $C_p(t)$.

Covariate name	Model based exposure calculation
24-hour AUC	$\int_{t-24}^{t} C_p dt$
24-hour C_{MAX}	$\max\left(\{C_p(s): t-24 < s < t\}\right)$
24-hour time above threshold drug concentration. (TAT)	$\int_{t-24}^{t} 1_{C_p > thresh} dt$
Predicted 6-month cumulative AUC	$\int_{0}^{6\ months} C_p dt$
Predicted 6-month C_{MAX}	$\max\left(\{C_p(s): 0 < s < 6\ months\}\right)$

For composite sampling, non-compartmental analysis was used to determine overall drug exposure, which consisted in averaging the simulated concentrations at each sampling time point. For serial sampling, drug exposure was calculated for each individual animal and then averaged over the cohort. In both cases, the arithmetic mean and geometric mean were calculated. Three different non-compartmental exposure measures were derived, the AUC, estimated using the linear-logarithmic trapezoidal rule, the C_{MAX}, and the time above threshold drug concentration, 0.01mg/ml. This value was used based on the assumption that adverse events were likely to occur above those levels.

Population pharmacokinetic modelling: Drug concentration profiles were fitted to pharmacokinetic models using first-order conditional estimation method with interaction, 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 three 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. model-based methods) 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 (27):

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

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

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 vs. time profiles, model convergence rates were usually high, with successful completion of the covariate step. An overview of the convergence rates is presented in Table 4.

Table 4 – Rates of successful model convergence and successful covariance (parameter precision) estimation.

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

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, Cmax 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 5).

In Figure 4, the relative errors are presented for the estimates for AUC, C_{MAX} and TAT. The relative errors were 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 exposure showed significantly higher bias and less precision in all scenarios. The performance for model-based exposure estimates obtained in the 3-month protocol is summarised in figure 5.

Our results also reveal the impact of composite versus serial sampling on bias and precision. For both model-based and NCA methods, the coefficient of variation increased with composite designs (with 8 animals) compared to serial sampling designs (with 3 animals), however the increase in precision for NCA 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 1CMT+MM scenario showed increased bias compared to the 1CMT and 2CMT scenarios.

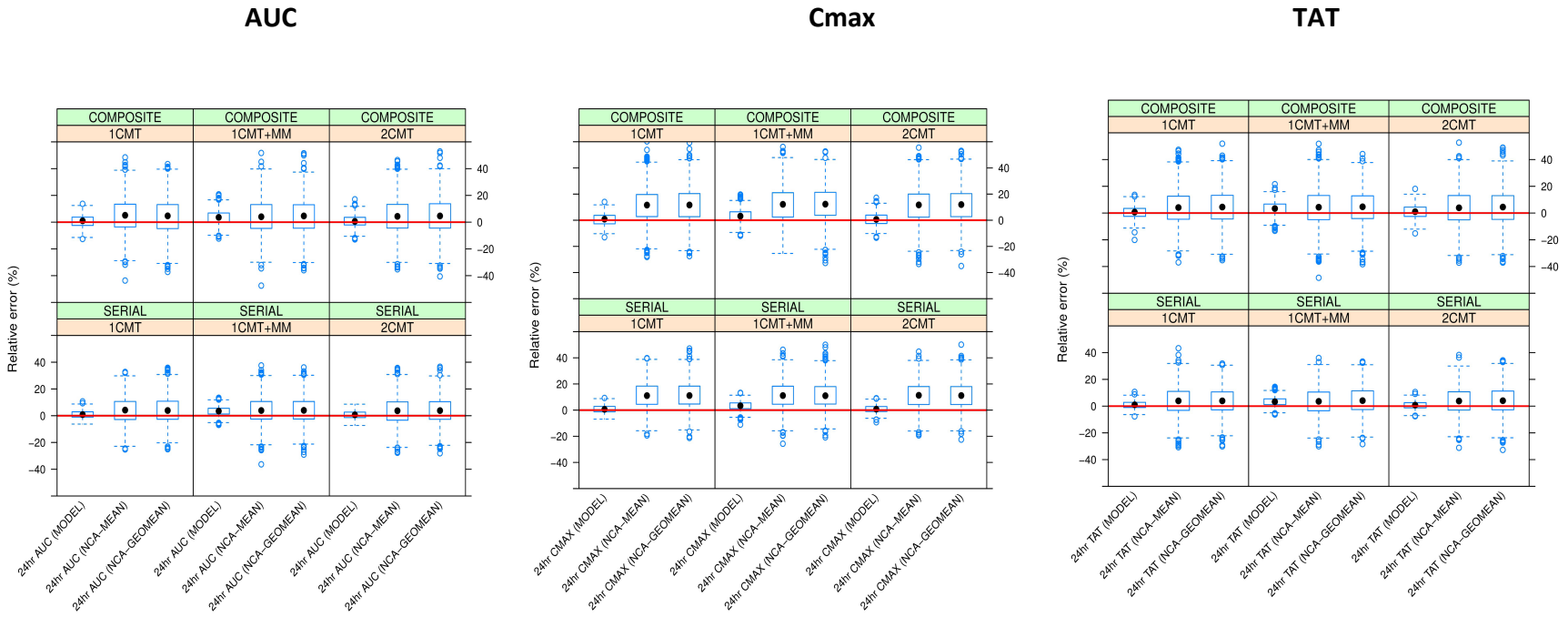


Figure 4 – Relative errors of parameter estimates for AUC (A, left panel), C_{MAX} (B, mid panel) and TAT (C, right panel). Data refers only to the 3-month toxicology protocol design following administration of 30 mg/kg/day of three hypothetical drugs with different pharmacokinetic profiles. Similar 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, and error bars denote the 5th and 95th percentiles. The horizontal line shows the reference level for relative error equal to zero. Composite – composite sampling; GEOMEAN – geometric mean; MEAN – arithmetic mean; MODEL- nonlinear mixed effects modelling; NCA – non-compartmental analysis and Serial – serial sampling.

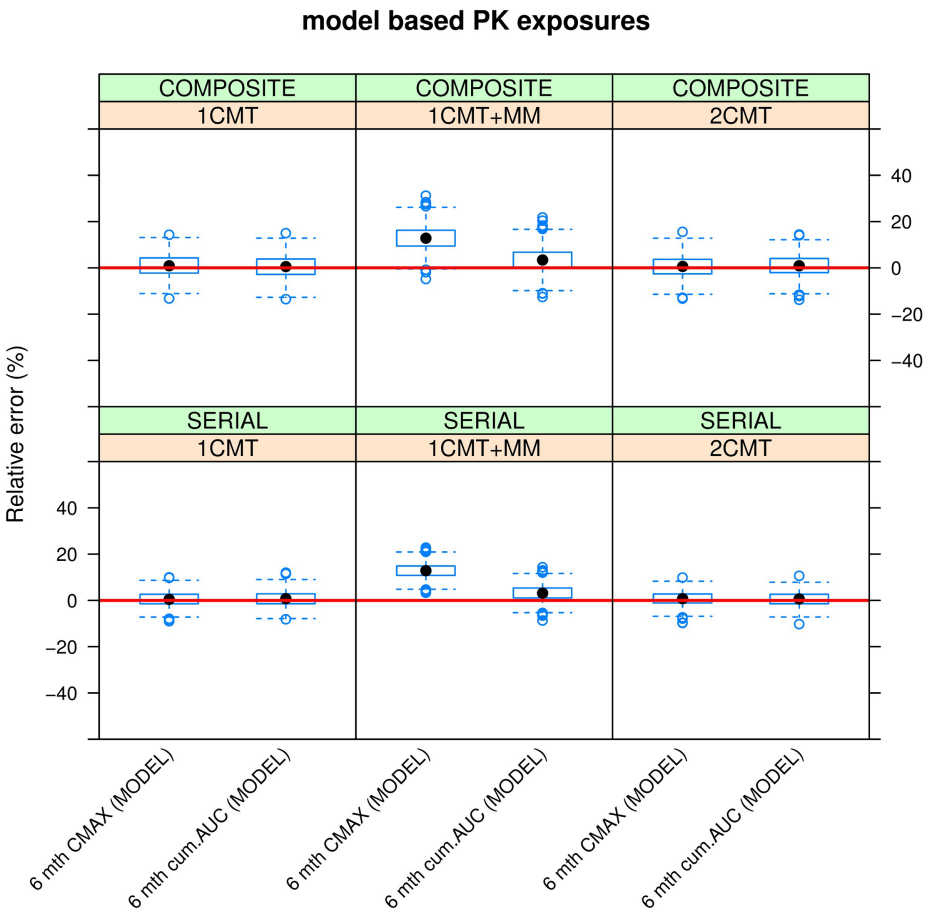


Figure 5 – Overview of the relative errors of model-based estimators of long-term exposure, as determined by a 3-month toxicology protocol following administration of 30 mg/kg/day of three hypothetical drugs with different pharmacokinetic profiles. Similar results were found for other cohorts in which 10 and 100 mg/kg/day were evaluated. Dots represent the median, boxes show the 25th and 75th percentiles, and error bars denote 1.5 times the interquartile range from the median. The horizontal line shows the reference level for relative error equal to zero.

Discussion

In this investigation we have attempted to identify important limitations in existing methodologies for the analysis of toxicokinetic data. Most importantly, we have illustrated the feasibility of a model-based approach for the estimation of toxicokinetic profile using a well-established parameterisation for drug disposition processes. Furthermore, given that model performance of toxicokinetic data has been previously evaluated (75), we have been able to focus the performance of measures of exposure that cannot be derived from empirical approaches, i.e., non-compartmental methods (29).

The chosen models for the hypothetical drugs reflect the likely toxicokinetic profile of many compounds in general toxicology studies. Taking into account the sampling schemes, the choice of one- and two-compartment models may apply to the majority compounds exhibiting linear pharmacokinetics. 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. The results presented here should therefore be indicative of the most common toxicokinetic profiles and as such 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 of computation of the covariance matrix for the scenarios tested here confirm the feasibility and reliability of results obtained using nonlinear mixed-effects modelling. Despite variations in bias and precision parameter precision was consistently high. The model-based approach performed particular well (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 suggest that a model-based approach may enable the reductions to the numbers of animals and/or samples whilst still 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 over-estimated. The cause is due to the definition of the NCA-based Cmax, $\max_t Cp(t)$ being necessarily greater than or equal to $Cp(t = \mathrm{T_{MAX}})$, 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 leads, actually lead to more bias. In other words, with NCA 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 of that increased sampling uniformly increases precision. Model specification issues which is discussed further in the limitations sections. 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.

.

Data integration

In contrast to non-compartmental methods, the data was analysed in an integrated manner, by combining the results from all experimental cohorts. This is undoubtedly the primary driver of the increased accuracy and precision in model-based estimates (30-32). In fact, we envisage further improvement by incorporating pharmacokinetic data from other experiments in the same species, which are normally collected during preclinical evaluation of the 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 (4). 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 advancements in toxicology, as they provide the basis for mechanism-based inferences about unwanted effects, irrespective of their incidence or occurrence in the actual experimental protocol (4, 35).

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 (76-78), 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 (79).

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 (39).

Given the success of PKPD modelling to aid in drug development (40-42), some attention must be paid to why the field of toxicology has yet to embrace it. There is sometimes scepticism of model based approach from a view they require knowing the model in advance (80). This argumentation is however flawed. The inference principles used for hypothesis generation and characterisation of PKPD relationships relies on the use of statistical criteria that are sophisticated enough to allow model identification and its suitability for subsequent parameter estimation purposes, irrespective of amounts of data available. Moreover, it should be noted that non-compartmental methods also make implicit assumptions about the underlying concentration vs. 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 the validity of these assumptions for the dataset at hand cannot be checked during the analysis.

However, NLME is specifically intended to efficiently process sparse data. The performance of the NLME-based PK exposure estimates in the composite designs is illustrative of this.

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 exposure 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) (44,45) . This is a particular 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 the above model selection criteria, clear reporting of such model uncertainty is necessary. Model averaging should be discouraged when predictions arising from different model differ significantly (64). 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 of presence of risk. This assumption disregards growing evidence of bias and poor precision of the derived measures of exposure, which should be avoided if data are subsequently used to define safety margins or thresholds. 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 will be essential to ensure experimental data is unbiased. Most importantly, it represent further adherence to the 3Rs principle, enabling significant reduction in number of animals required for the evaluation of toxicokinetics.

Reference List

1. Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nat.Rev.Drug Discov. **3**, 711-715.

2. U.S.Food and Drug Administration (CDER). Innovation or stagnation? Challenge and opportunity on the critical path to new medical products. 2004.

URL:
http://www.fda.gov/scienceresearch/specialtopics/criticalpathinitiative/criticalpathopportunitiesreports/ucm077262.htm

3. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul.Toxicol.Pharmacol. **32**, 56-67.

4. Russell, W. M. S. and BURCH, R. L. The principles of humane experimental technique. 1-4-1959. London: Methuen & Co. Ltd.

5. Suntharalingam, G., Perry, M. R., Ward, S., Brett, S. J., Castello-Cortes, A., Brunner, M. D., and Panoskaltsis, N. (2006). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. N.Engl.J.Med. **355**, 1018-1028.

6. Expert Scientific Group. Expert Scientific Group on Phase One Clinical Trials. 2006.

7. European Medicines Agency (CHMP). Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products. 2007.
URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002988.pdf

8. Kenter, M. J., and Cohen, A. F. (2006). Establishing risk of human experimentation with drugs: lessons from TGN1412. Lancet **368**, 1387-1391.

9. Dowsing, T., and Kendall, M. J. (2007). The Northwick Park tragedy--protecting healthy volunteers in future first-in-man trials. J.Clin.Pharm.Ther. **32**, 203-207.

10. Stebbings, R., Poole, S., and Thorpe, R. (2009). Safety of biologics, lessons learnt from TGN1412. Curr.Opin.Biotechnol. **20**, 673-677.

11. Breckenridge, A. (2010). Regulatory challenges, reimbursement, and risk-benefit assessment. Clin.Pharmacol.Ther. **88**, 153-154.

12. Krumholz, H. M., Ross, J. S., Presler, A. H., and Egilman, D. S. (2007). What have we learnt from Vioxx? BMJ **334**, 120-123.

13. Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000). Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N.Engl.J.Med. **343**, 1520-8, 2.

14. Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002). Timing of new black box warnings and withdrawals for prescription medications. JAMA **287**, 2215-2220.

15. Cross, J., Lee, H., Westelinck, A., Nelson, J., Grudzinskas, C., and Peck, C. (2002). Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980-1999. Pharmacoepidemiol.Drug Saf **11**, 439-446.

16. Heerdink, E. R., Urquhart, J., and Leufkens, H. G. (2002). Changes in prescribed drug doses after market introduction. Pharmacoepidemiol.Drug Saf **11**, 447-453.

17. Lowe, P. J., Hijazi, Y., Luttringer, O., Yin, H., Sarangapani, R., and Howard, D. (2007). On the anticipation of the human dose in first-in-man trials from preclinical and prior clinical information in early drug development. Xenobiotica **37**, 1331-1354.

18. Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., and Wurtzen, G. (2002). Mathematical modelling and quantitative methods. Food Chem.Toxicol. **40**, 283-326.

19. Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A. G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., and Younes, M. (2002). Hazard characterisation of

chemicals in food and diet. dose response, mechanisms and extrapolation issues. Food Chem.Toxicol. **40**, 237-282.

20. Slob, W. (1999). Thresholds in toxicology and risk assessment. International Journal of Toxicology **18**, 345-367.

21. Kroes, R., Kleiner, J., and Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. Toxicol.Sci. **86**, 226-230.

22. Dorato, M. A., and Engelhardt, J. A. (2005). The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul.Toxicol.Pharmacol. **42**, 265-274.

23. U.S.Food and Drug Administration (CDER). Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. 2005.
URL: http://www.fda.gov/downloads/Drugs/Guidance/UCM078932.pdf

24. Agoram, B. M. (2009). Use of pharmacokinetic/ pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics. Br.J.Clin.Pharmacol. **67**, 153-160.

25. Muller, P. Y., Milton, M., Lloyd, P., Sims, J., and Brennan, F. R. (2009). The minimum anticipated biological effect level (MABEL) for selection of first human dose in clinical trials with monoclonal antibodies. Curr.Opin.Biotechnol. **20**, 722-729.

26. Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Regul.Toxicol.Pharmacol. **3**, 224-238.

27. Swartout, J. C., Price, P. S., Dourson, M. L., Carlson-Lynch, H. L., and Keenan, R. E. (1998). A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. **18**, 271-282.

28. Gaylor, D. W., and Kodell, R. L. (2000). Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal. **20**, 245-250.

29. Shih, C., Gosset, L., and Gates, S. e. a. (1996). LY231514 and its polyglutamates exhibit potent inhibition against both human dihydrofolate reductase (DHFR) and thymidylate synthase (TS): multiple folate enzyme inhibition. Annals of Oncology.

30. Latz, J. E., Rusthoven, J. J., Karlsson, M. O., Ghosh, A., and Johnson, R. D. (2006). Clinical application of a semimechanistic-physiologic population PK/PD model for neutropenia following pemetrexed therapy. Cancer Chemother.Pharmacol. **57**, 427-435.

31. Jeste, D. V., Lacro, J. P., Palmer, B., Rockwell, E., Harris, M. J., and Caligiuri, M. P. (1999). Incidence of tardive dyskinesia in early stages of low-dose treatment with typical neuroleptics in older patients. Am.J.Psychiatry **156**, 309-311.

32. Danhof, M., de Jongh, J., de Lange, E. C., Della, P. O., Ploeger, B. A., and Voskuyl, R. A. (2007). Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. Annu.Rev.Pharmacol.Toxicol. **47**, 357-400.

33. Danhof, M., de Lange, E. C., Della Pasqua, O. E., Ploeger, B. A., and Voskuyl, R. A. (2008). Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. Trends Pharmacol.Sci. **29**, 186-191.

34. Danhof, M., Alvan, G., Dahl, S. G., Kuhlmann, J., and Paintaud, G. (2005). Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. Pharm.Res. **22**, 1432-1437.

35. Rolan, P., Danhof, M., Stanski, D., and Peck, C. (2007). Current issues relating to drug safety especially with regard to the use of biomarkers: a meeting report and progress update. Eur.J.Pharm.Sci. **30**, 107-112.

36. Jorga, K. M., Fotteler, B., Heizmann, P., and Zurcher, G. (1998). Pharmacokinetics and pharmacodynamics after oral and intravenous administration of tolcapone, a novel adjunct to Parkinson's disease therapy. Eur.J.Clin.Pharmacol. **54**, 443-447.

37. Bakri, R., Wang, J., Wierzbicki, A. S., and Goldsmith, D. (2003). Cerivastatin monotherapy-induced muscle weakness, rhabdomyolysis and acute renal failure. Int.J.Cardiol. **91**, 107-109.

38. Wernevik, L. C., Nystrom, P., Johnsson, G., Nakanishi, T., and Eriksson, U. G. (2006). Pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor ximelagatran in young healthy Japanese men. Clin.Pharmacokinet. **45**, 77-84.

39. Cohen, A. (2007). Should we tolerate tolerability as an objective in early drug development? Br.J.Clin.Pharmacol. **64**, 249-252.

40. Oliveira, R. B., Gomes-Leal, W., do-Nascimento, J. L., and Picanco-Diniz, C. W. (1998). Methylmercury intoxication and histochemical demonstration of NADPH-diaphorase activity in the striate cortex of adult cats. Braz.J.Med.Biol.Res. **31**, 1157-1161.

41. Barnes, D. G., Daston, G. P., Evans, J. S., Jarabek, A. M., Kavlock, R. J., Kimmel, C. A., Park, C., and Spitzer, H. L. (1995). Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul.Toxicol.Pharmacol. **21**, 296-306.

42. Budtz-Jorgensen, E., Grandjean, P., Keiding, N., White, R. F., and Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol.Lett. **112-113**, 193-199.

43. Lalonde, R. L., Kowalski, K. G., Hutmacher, M. M., Ewy, W., Nichols, D. J., Milligan, P. A., Corrigan, B. W., Lockwood, P. A., Marshall, S. A., Benincosa, L. J., Tensfeldt, T. G., Parivar, K., Amantea, M., Glue, P., Koide, H., and Miller, R. (2007). Model-based drug development. Clin.Pharmacol.Ther. **82**, 21-32.

44. Kodell, R. L., and Chen, J. J. (2007). On the use of hierarchical probabilistic models for characterizing and managing uncertainty in risk/safety assessment. Risk Anal. **27**, 433-437.

45. West, G. B., Brown, J. H., and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. Science **276**, 122-126.

46. Zuideveld, K. P., Van der Graaf, P. H., Peletier, L. A., and Danhof, M. (2007). Allometric scaling of pharmacodynamic responses: application to 5-Ht1A receptor mediated responses from rat to man. Pharm.Res. **24**, 2031-2039.

47. Mager, D. E., Woo, S., and Jusko, W. J. (2009). Scaling pharmacodynamics from in vitro and preclinical animal studies to humans. Drug Metab Pharmacokinet. **24**, 16-24.

48. Bachmann, K., Pardoe, D., and White, D. (1996). Scaling basic toxicokinetic parameters from rat to man. Environ.Health Perspect. **104**, 400-407.

49. Langdon, G., Gueorguieva, I., Aarons, L., and Karlsson, M. (2007). Linking preclinical and clinical whole-body physiologically based pharmacokinetic models with prior distributions in NONMEM. Eur.J.Clin.Pharmacol. **63**, 485-498.

50. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA **279**, 1200-1205.

51. Ju, C., and Uetrecht, J. P. (2002). Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. Curr.Drug Metab **3**, 367-377.

52. Haley, P. J. (2003). Species differences in the structure and function of the immune system. Toxicology **188**, 49-71.

53. European Medicines Agency (CHMP). Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007.
URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf

54. Davidian, M., and Giltinan, D. M. (1993). Analysis of repeated measurement data using nonlinear mixed effects models. Chemom.Intell.LAb.Syst. **20**, 1-24.

55. Karlsson, M. O., Beal, S. L., and Sheiner, L. B. (1995). Three new residual error models for population PK/PD analyses. J.Pharmacokinet.Biopharm. **23**, 651-672.

56. Beal, S. L. (1984). Population pharmacokinetic data and parameter estimation based on their first two statistical moments. Drug Metab Rev. **15**, 173-193.

57. Lindstrom, M. L., and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. Biometrics **46**, 673-687.

58. Pinheiro, J. C., and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. J.Comput.Graph.Stat. **4**, 12-35.

59. Baverel, P. G., Savic, R. M., Wilkins, J. J., and Karlsson, M. O. (2009). Evaluation of the nonparametric estimation method in NONMEM VI: application to real data. J Pharmacokinet.Pharmacodyn. **36**, 297-315.

60. Gisleskog, P. O., Karlsson, M. O., and Beal, S. L. (2002). Use of prior information to stabilize a population data analysis. J.Pharmacokinet.Pharmacodyn. **29**, 473-505.

61. Lunn, D. J., Best, N., Thomas, A., Wakefield, J., and Spiegelhalter, D. (2002). Bayesian analysis of population PK/PD models: general concepts and software. J.Pharmacokinet.Pharmacodyn. **29**, 271-307.

62. Kuhn E., and Lavielle M. (2004). Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis **49**, 1020-1038.

63. Bois, F. Y. (2010). Physiologically based modelling and prediction of drug interactions. Basic Clin.Pharmacol.Toxicol. **106**, 154-161.

64. Lunn, D. J. (2008). Automated covariate selection and Bayesian model averaging in population PK/PD models. J.Pharmacokinet.Pharmacodyn. **35**, 85-100.

65. Wade, J. R., Beal, S. L., and Sambol, N. C. (1994). Interaction between structural, statistical, and covariate models in population pharmacokinetic analysis. J Pharmacokinet.Biopharm. **22**, 165-177.

66. Mentre, F., and Mallet, A. (1994). Handling covariates in population pharmacokinetics. Int.J Biomed.Comput **36**, 25-33.

67. Mandema, J. W., Verotta, D., and Sheiner, L. B. (1992). Building population pharmacokinetic--pharmacodynamic models. I. Models for covariate effects. J Pharmacokinet.Biopharm. **20**, 511-528.

68. Holford, N. H., Kimko, H. C., Monteleone, J. P., and Peck, C. C. (2000). Simulation of clinical trials. Annu.Rev.Pharmacol.Toxicol. **40**, 209-234.

69. Holford, N., Ma, S. C., and Ploeger, B. A. (2010). Clinical trial simulation: a review. Clin.Pharmacol.Ther. **88**, 166-182.

70. Mentre, F., Duffull, S., Gueorguieva, I., Hooker, A., Leonov, S., Ogungbenro, K., and Retout, S. Software for optimal design in population pharmacokinetics and pharmacodynamics: a comparison. PAGE. 2007.

71. Dorato, M. A., and Engelhardt, J. A. (2005). The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul.Toxicol.Pharmacol. **42**, 265-274.

72. U.S.Food and Drug Administration. Guidance for Industry M3(R2) Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. 1-1-2010.
URL:
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073246.pdf

73. Tse, F. L., and Nedelman, J. R. (1996). Serial versus sparse sampling in toxicokinetic studies. Pharm.Res. **13**, 1105-1108.

74. Millard, S., and Krause, A. (2001). Applied statistics in the pharmaceutical industry: with case studies using S-Plus. Springer-Verlag.

75. Hing, J. P., Woolfrey, S. G., Greenslade, D., and Wright, P. M. (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.

76. Gagnon, R. C., and Peterson, J. J. (1998). Estimation of confidence intervals for area under the curve from destructively obtained pharmacokinetic data. J.Pharmacokinet.Biopharm. **26**, 87-102.

77. Nedelman, J. R., Gibiansky, E., and Lau, D. T. (1995). Applying Bailer's method for AUC confidence intervals to sparse sampling. Pharm.Res. **12**, 124-128.

78. Bonate, P. L. (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.

79. ICH. Toxicokinetics: a guidance for assessing systemic exposure in toxicology studies. 1-3-1995.
URL: http://www.emea.eu.int

80. Pai, S. M., Fettner, S. H., Hajian, G., Cayen, M. N., and Batra, V. K. (1996). Characterization of AUCs from sparsely sampled populations in toxicology studies. Pharm.Res. **13**, 1283-1290.

81. Aronson, J. K., and Ferner, R. E. (2003). Joining the DoTS: new approach to classifying adverse drug reactions. BMJ **327**, 1222-1225.

82. Contrera, J. F., Matthews, E. J., Kruhlak, N. L., and Benz, R. D. (2004). Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul.Toxicol.Pharmacol. **40**, 185-206.

83. Retout, S., Duffull, S., and Mentre, F. (2001). Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. Comput.Methods Programs Biomed. **65**, 141-151.

84. Waterhouse, T. H., Redmann, S., Duffull, S. B., and Eccleston, J. A. (2005). Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. J.Pharmacokinet.Pharmacodyn. **32**, 521-545.

85. Duffull, S. B., Mentre, F., and Aarons, L. (2001). Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm.Res. **18**, 83-89.

86. Ette, E. I., Howie, C. A., Kelman, A. W., and Whiting, B. (1995). Experimental design and efficient parameter estimation in preclinical pharmacokinetic studies. Pharm.Res. **12**, 729-737.

87. Hooker, A. C., Foracchia, M., Dodds, M. G., and Vicini, P. (2003). An evaluation of population D-optimal designs via pharmacokinetic simulations. Ann.Biomed.Eng **31**, 98-111.

88. Foracchia, M., Hooker, A., Vicini, P., and Ruggeri, A. (2004). POPED, a software for optimal experiment design in population kinetics. Comput.Methods Programs Biomed. **74**, 29-46.

89. Retout, S., Mentre, F., and Bruno, R. (2002). Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Stat.Med. **21**, 2623-2639.

90. Ogungbenro, K., Gueorguieva, I., Majid, O., Graham, G., and Aarons, L. (2007). Optimal design for multiresponse pharmacokinetic-pharmacodynamic models - dealing with unbalanced designs. J.Pharmacokinet.Pharmacodyn. **34**, 313-331.

91. Huang, C., Zheng, M., Yang, Z., Rodrigues, A. D., and Marathe, P. (2008). Projection of exposure and efficacious dose prior to first-in-human studies: how successful have we been? Pharm.Res. **25**, 713-726.

92. Gibbs, J. P. (2010). Prediction of exposure-response relationships to support first-in-human study design. AAPS.J. **12**, 750-758.

93. Budtz-Jorgensen, E., Grandjean, P., Keiding, N., White, R. F., and Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol.Lett. **112-113**, 193-199.

94. Nyberg, J., Karlsson, M. O., and Hooker, A. C. (2009). Simultaneous optimal experimental design on dose and sample times. J.Pharmacokinet.Pharmacodyn. **36**, 125-145.

95. D'Argenio, D. Z. (1981). Optimal sampling times for pharmacokinetic experiments. J.Pharmacokinet.Biopharm. **9**, 739-756.

96. al Banna, M. K., Kelman, A. W., and Whiting, B. (1990). Experimental design and efficient parameter estimation in population pharmacokinetics. J.Pharmacokinet.Biopharm. **18**, 347-360.

97. Wang, J., and Endrenyi, L. (1992). A computationally efficient approach for the design of population pharmacokinetic studies. J.Pharmacokinet.Biopharm. **20**, 279-294.

98. Thai, H., Mentre, F., Holford, N. H., Veyrat-Follet, C., and Comets, E. Residual-based bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models with heteroscedastic error. PAGE meeting . 2012.
URL: http://www.page-meeting.org/pdf_assets/2550-Thai_poster_Page12_vfinal.pdf

99. Hooker, A., and Vicini, P. (2005). Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. AAPS.J. **7**, E759-E785.

100. Edler, D., Glimelius, B., Hallstrom, M., Jakobsen, A., Johnston, P. G., Magnusson, I., Ragnhammar, P., and Blomgren, H. (2002). Thymidylate synthase expression in colorectal cancer: a prognostic and predictive marker of benefit from adjuvant fluorouracil-based chemotherapy. J.Clin.Oncol. **20**, 1721-1728.

101. European Medicines Agency. Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007.

102. Vane, J. R. (1971). Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. Nat.New Biol. **231**, 232-235.

103. Langman, M. J., Jensen, D. M., Watson, D. J., Harper, S. E., Zhao, P. L., Quan, H., Bolognese, J. A., and Simon, T. J. (1999). Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA **282**, 1929-1933.

104. Hooker, A. C., Staatz, C. E., and Karlsson, M. O. (2007). Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. Pharm.Res. **24**, 2187-2197.

105. Zhang, L., Beal, S. L., and Sheiner, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J.Pharmacokinet.Pharmacodyn. **30**, 387-404.

106. Zhang, L., Beal, S. L., and Sheinerz, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data II: robustness of methods. J.Pharmacokinet.Pharmacodyn. **30**, 405-416.

107. Lindbom, L., Pihlgren, P., and Jonsson, E. N. (2005). PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput.Methods Programs Biomed. **79**, 241-257.

108. Capone, M. L., Tacconelli, S., Sciulli, M. G., Anzellotti, P., Di, F. L., Merciaro, G., Di, G. P., and Patrignani, P. (2007). Human pharmacology of naproxen sodium. J.Pharmacol.Exp.Ther. **322**, 453-460.

109. Konstam, M. A., Weir, M. R., Reicin, A., Shapiro, D., Sperling, R. S., Barr, E., and Gertz, B. J. (2001). Cardiovascular thrombotic events in controlled, clinical trials of rofecoxib. Circulation **104**, 2280-2288.

110. Fitzgerald, G. A. (2007). COX-2 in play at the AHA and the FDA. Trends Pharmacol.Sci. **28**, 303-307.

Appendix

The following tables contain all SRME and CV values for all scenarios, cohorts and analysis methods.

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
1CMT	SERIAL	24hr AUC	10	1	0.8066	2.787	3.993	10.74	4.15	9.824
1CMT	SERIAL	24hr AUC	10	4	1.403	2.982	5.107	9.256	4.053	10.26
1CMT	SERIAL	24hr AUC	10	12	0.784	2.822	3.055	9.367	3.948	9.826
1CMT	SERIAL	24hr AUC	30	1	0.8066	2.787	4.565	9.472	4.298	9.344
1CMT	SERIAL	24hr AUC	30	4	1.403	2.982	3.721	9.74	4.934	10.58
1CMT	SERIAL	24hr AUC	30	12	0.784	2.822	3.065	9.873	4.258	10.48
1CMT	SERIAL	24hr AUC	100	1	0.8066	2.787	4.615	9.845	3.538	9.214
1CMT	SERIAL	24hr AUC	100	4	1.403	2.982	3.02	9.474	3.939	8.846
1CMT	SERIAL	24hr AUC	100	12	0.784	2.822	4.038	9.928	3.59	10.71
1CMT	COMPOSITE	24hr AUC	10	1	0.4114	4.651	5.245	13.41	4.815	12.93
1CMT	COMPOSITE	24hr AUC	10	4	0.6365	4.324	5.634	12.73	3.838	13.51
1CMT	COMPOSITE	24hr AUC	10	12	1.115	4.373	3.815	13.17	4.564	12.98
1CMT	COMPOSITE	24hr AUC	30	1	0.4114	4.651	6.231	12.56	4.027	13.28
1CMT	COMPOSITE	24hr AUC	30	4	0.6365	4.324	5.212	12.68	4.35	12.98
1CMT	COMPOSITE	24hr AUC	30	12	1.115	4.373	3.564	13.24	4.632	12.19
1CMT	COMPOSITE	24hr AUC	100	1	0.4114	4.651	5.29	12.85	4.466	13.68
1CMT	COMPOSITE	24hr AUC	100	4	0.6365	4.324	4.209	12.18	3.833	13.03
1CMT	COMPOSITE	24hr AUC	100	12	1.115	4.373	4.564	12.71	3.903	13.63

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA -GEOMEAN
1CMT+MM	SERIAL	24hr AUC	10	1	3.454	3.152	4.042	9.245	4.066	9.429
1CMT+MM	SERIAL	24hr AUC	10	4	3.177	3.255	3.784	9.352	5.347	8.897
1CMT+MM	SERIAL	24hr AUC	10	12	3.457	3.167	4.354	9.667	3.889	9.966
1CMT+MM	SERIAL	24hr AUC	30	1	3.165	3.309	3.883	10.35	3.277	9.953
1CMT+MM	SERIAL	24hr AUC	30	4	3.416	3.281	4.477	10.15	3.199	10.62
1CMT+MM	SERIAL	24hr AUC	30	12	3.572	3.226	3.282	11.27	4.358	9.756
1CMT+MM	SERIAL	24hr AUC	100	1	2.993	3.069	4.03	9.883	3.608	10.66
1CMT+MM	SERIAL	24hr AUC	100	4	3.369	3.37	4.945	9.871	3.363	9.435
1CMT+MM	SERIAL	24hr AUC	100	12	3.459	3.024	3.657	9.409	5.191	9.725
1CMT+MM	COMPOSITE	24hr AUC	10	1	3.176	4.932	3.502	12.07	5.057	13.09
1CMT+MM	COMPOSITE	24hr AUC	10	4	3.192	4.768	3.777	12.38	4.02	12.54
1CMT+MM	COMPOSITE	24hr AUC	10	12	3.399	4.88	3.864	13.15	4.442	12.16
1CMT+MM	COMPOSITE	24hr AUC	30	1	4.214	5.316	3.68	13.83	3.438	12.73
1CMT+MM	COMPOSITE	24hr AUC	30	4	3.262	5.196	5.853	12.8	4.427	12.42
1CMT+MM	COMPOSITE	24hr AUC	30	12	3.496	5.407	3.888	13.18	7.559	13.46
1CMT+MM	COMPOSITE	24hr AUC	100	1	3.868	5.214	4.33	12.11	3.738	13.41
1CMT+MM	COMPOSITE	24hr AUC	100	4	3.502	4.82	5.149	13.12	2.725	13.61
1CMT+MM	COMPOSITE	24hr AUC	100	12	3.312	4.733	3.812	12.58	5.262	12.38

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
2CMT	SERIAL	24hr AUC	10	1	0.5484	2.819	3.009	10.21	4.508	9.854
2CMT	SERIAL	24hr AUC	10	4	0.8587	3.256	4.84	10.01	5.346	9.577
2CMT	SERIAL	24hr AUC	10	12	0.5528	2.943	3.429	9.583	3.767	10.48
2CMT	SERIAL	24hr AUC	30	1	0.5484	2.819	4.379	10.31	3.721	9.561
2CMT	SERIAL	24hr AUC	30	4	0.8587	3.256	3.079	10.8	3.865	9.479
2CMT	SERIAL	24hr AUC	30	12	0.5528	2.943	3.966	10.32	4.311	9.826
2CMT	SERIAL	24hr AUC	100	1	0.5484	2.819	4.732	10.38	3.649	10.28
2CMT	SERIAL	24hr AUC	100	4	0.8587	3.256	2.261	9.635	3.11	9.859
2CMT	SERIAL	24hr AUC	100	12	0.5528	2.943	4.047	10.43	3.432	9.983
2CMT	COMPOSITE	24hr AUC	10	1	0.6988	4.286	2.96	13.31	4.885	13.99
2CMT	COMPOSITE	24hr AUC	10	4	0.8458	4.847	5.062	12.59	6.413	12.44
2CMT	COMPOSITE	24hr AUC	10	12	0.5273	4.514	5.177	12.45	4.647	13.75
2CMT	COMPOSITE	24hr AUC	30	1	0.6988	4.286	4.258	13.9	4.463	12.16
2CMT	COMPOSITE	24hr AUC	30	4	0.8458	4.847	3.757	12.9	2.889	14.08
2CMT	COMPOSITE	24hr AUC	30	12	0.5273	4.514	4.381	12.81	4.018	12.59
2CMT	COMPOSITE	24hr AUC	100	1	0.6988	4.286	5.057	13.23	3.703	13.07
2CMT	COMPOSITE	24hr AUC	100	4	0.8458	4.847	3.908	14.25	6.301	13.07
2CMT	COMPOSITE	24hr AUC	100	12	0.5273	4.514	4.586	12.94	4.2	12.94

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
1CMT	SERIAL	24hr CMAX	10	1	0.7306	2.899	11.54	9.945	10.8	9.747
1CMT	SERIAL	24hr CMAX	10	4	0.5798	2.893	10.74	9.382	11.41	10.37
1CMT	SERIAL	24hr CMAX	10	12	0.8339	3.045	11.58	10.51	10.32	9.832
1CMT	SERIAL	24hr CMAX	30	1	0.7306	2.899	11.72	8.873	11.42	9.831
1CMT	SERIAL	24hr CMAX	30	4	0.5798	2.893	10.6	9.846	12.38	10.74
1CMT	SERIAL	24hr CMAX	30	12	0.8339	3.045	9.94	10.25	11.82	10.21
1CMT	SERIAL	24hr CMAX	100	1	0.7306	2.899	11.92	8.828	11.23	10.8
1CMT	SERIAL	24hr CMAX	100	4	0.5798	2.893	11.55	10.67	12.12	9.98
1CMT	SERIAL	24hr CMAX	100	12	0.8339	3.045	12.1	10.27	11.61	9.536
1CMT	COMPOSITE	24hr CMAX	10	1	0.6322	4.557	10.61	12.68	11.85	12.11
1CMT	COMPOSITE	24hr CMAX	10	4	0.647	4.659	9.818	12.3	11.59	14.16
1CMT	COMPOSITE	24hr CMAX	10	12	0.3693	4.73	9.869	13.41	10.37	13.76
1CMT	COMPOSITE	24hr CMAX	30	1	0.6322	4.557	12.7	12.29	10.56	13.33
1CMT	COMPOSITE	24hr CMAX	30	4	0.647	4.659	12.34	12.77	11.76	13.51
1CMT	COMPOSITE	24hr CMAX	30	12	0.3693	4.73	12.21	13.67	13.15	13.48
1CMT	COMPOSITE	24hr CMAX	100	1	0.6322	4.557	12.12	12	10.8	12.03
1CMT	COMPOSITE	24hr CMAX	100	4	0.647	4.659	11.19	13.89	12.08	12.21
1CMT	COMPOSITE	24hr CMAX	100	12	0.3693	4.73	9.968	13.04	11.67	12.97

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
1CMT+MM	SERIAL	24hr	10	4	3.916	3.461	7.387	13.15	14.32	15.12
1CMT+MM	SERIAL	24hr	10	12	1.814	2.586	15.26	6.701	11.92	8.455
1CMT+MM	SERIAL	24hr	30	1	2.997	2.533	12.89	9.863	6.667	7.948
1CMT+MM	SERIAL	24hr	30	4	2.195	3.013	11.01	12.71	13.32	8.925
1CMT+MM	SERIAL	24hr	30	12	2.997	4.675	12.79	7.512	11.2	9.27
1CMT+MM	SERIAL	24hr	100	1	2.539	2.99	9.542	6.405	4.643	10.42
1CMT+MM	SERIAL	24hr	100	4	2.236	4.535	13.72	9.105	13.44	11.59
1CMT+MM	SERIAL	24hr	100	12	3.108	3.104	8.219	10.94	16.4	9.72
1CMT+MM	COMPOSITE	24hr	10	1	4.193	3.414	22.85	17.18	12.08	13.98
1CMT+MM	COMPOSITE	24hr	10	4	3.428	6.017	13.7	14.06	13.61	14.29
1CMT+MM	COMPOSITE	24hr	10	12	3.076	4.608	14.43	17.99	13.83	15.88
1CMT+MM	COMPOSITE	24hr	30	1	3.89	4.507	16.05	12.46	17.38	18.27
1CMT+MM	COMPOSITE	24hr	30	4	2.88	3.301	17.05	12	17.62	10.72
1CMT+MM	COMPOSITE	24hr	30	12	4.101	7.039	12.73	17.45	14.67	10.93
1CMT+MM	COMPOSITE	24hr	100	1	5.108	3.725	15.38	10.59	14.11	16.02
1CMT+MM	COMPOSITE	24hr	100	4	0.8694	3.595	10.94	9.899	10.14	8.477
1CMT+MM	COMPOSITE	24hr	100	12	2.504	6.026	8.776	13.82	13.18	17.37

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
2CMT	SERIAL	24hr CMAX	10	1	0.7472	3.015	10.9	10.33	12.4	9.994
2CMT	SERIAL	24hr CMAX	10	4	0.8701	2.938	11.06	10.1	11.89	9.653
2CMT	SERIAL	24hr CMAX	10	12	0.6139	2.959	11.32	10.4	10.68	10.7
2CMT	SERIAL	24hr CMAX	30	1	0.7472	3.015	11.35	9.376	10.76	10.76
2CMT	SERIAL	24hr CMAX	30	4	0.8701	2.938	11.19	10.82	10.64	9.939
2CMT	SERIAL	24hr CMAX	30	12	0.6139	2.959	11.19	9.61	11.63	9.976
2CMT	SERIAL	24hr CMAX	100	1	0.7472	3.015	10.86	8.862	11.59	9.721
2CMT	SERIAL	24hr CMAX	100	4	0.8701	2.938	11.34	9.274	10.92	9.915
2CMT	SERIAL	24hr CMAX	100	12	0.6139	2.959	10.53	10.36	10.95	10.01
2CMT	COMPOSITE	24hr CMAX	10	1	0.9441	4.693	10.16	14.9	9.881	12.51
2CMT	COMPOSITE	24hr CMAX	10	4	0.7777	4.718	10.49	12.52	11.22	12.73
2CMT	COMPOSITE	24hr CMAX	10	12	0.6393	5.002	11.36	14.06	11.31	13.68
2CMT	COMPOSITE	24hr CMAX	30	1	0.9441	4.693	11.82	14.28	13.13	12.84
2CMT	COMPOSITE	24hr CMAX	30	4	0.7777	4.718	12.67	12.09	12.69	12.89
2CMT	COMPOSITE	24hr CMAX	30	12	0.6393	5.002	11.43	13.28	11.23	12.88
2CMT	COMPOSITE	24hr CMAX	100	1	0.9441	4.693	11.76	11.98	11.6	14.2
2CMT	COMPOSITE	24hr CMAX	100	4	0.7777	4.718	12.52	13.27	12.34	11.92
2CMT	COMPOSITE	24hr CMAX	100	12	0.6393	5.002	10.43	13.29	12.31	13.51

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
1CMT	SERIAL	24hr TAT	10	1	1.104	2.855	3.516	10.72	3.06	9.718
1CMT	SERIAL	24hr TAT	10	4	0.9731	2.95	3.725	9.746	4.782	9.684
1CMT	SERIAL	24hr TAT	10	12	0.7183	2.738	3.945	9.307	3.261	10.24
1CMT	SERIAL	24hr TAT	30	1	1.104	2.855	3.846	10.61	3.263	9.204
1CMT	SERIAL	24hr TAT	30	4	0.9731	2.95	3.112	10.22	4.793	10.36
1CMT	SERIAL	24hr TAT	30	12	0.7183	2.738	3.944	10.07	4.197	9.843
1CMT	SERIAL	24hr TAT	100	1	1.104	2.855	5.182	10.92	4.328	9.798
1CMT	SERIAL	24hr TAT	100	4	0.9731	2.95	4.906	10.05	3.802	10.18
1CMT	SERIAL	24hr TAT	100	12	0.7183	2.738	3.106	10.58	4.077	9.474
1CMT	COMPOSITE	24hr TAT	10	1	0.1905	4.511	3.011	12.53	3.635	12.95
1CMT	COMPOSITE	24hr TAT	10	4	1.022	4.251	4.875	12.75	5.447	12.87
1CMT	COMPOSITE	24hr TAT	10	12	0.1963	4.784	6.422	12.15	4.639	12.27
1CMT	COMPOSITE	24hr TAT	30	1	0.1905	4.511	2.807	13.83	4.701	13.07
1CMT	COMPOSITE	24hr TAT	30	4	1.022	4.251	6.049	12.4	2.792	13.49
1CMT	COMPOSITE	24hr TAT	30	12	0.1963	4.784	3.696	11.97	4.605	14.31
1CMT	COMPOSITE	24hr TAT	100	1	0.1905	4.511	3.725	11.97	3.86	12.92
1CMT	COMPOSITE	24hr TAT	100	4	1.022	4.251	4.092	13.38	5.199	13.95
1CMT	COMPOSITE	24hr TAT	100	12	0.1963	4.784	2.898	13.19	3.727	12.68

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
1CMT+MM	SERIAL	24hr TAT	10	1	3.51	2.891	3.178	10.52	5.906	9.741
1CMT+MM	SERIAL	24hr TAT	10	4	3.492	3.447	4.244	10.57	4.249	10.57
1CMT+MM	SERIAL	24hr TAT	10	12	2.956	3.273	3.632	10.48	4.573	10.21
1CMT+MM	SERIAL	24hr TAT	30	1	3.473	3.036	2.326	10.07	4.185	9.986
1CMT+MM	SERIAL	24hr TAT	30	4	3.747	3.168	4.568	10.9	3.938	10.14
1CMT+MM	SERIAL	24hr TAT	30	12	2.577	3.079	3.717	9.856	2.89	10.71
1CMT+MM	SERIAL	24hr TAT	100	1	2.873	3.179	3.713	10.02	4.367	9.592
1CMT+MM	SERIAL	24hr TAT	100	4	3.447	2.947	3.826	10.31	4.132	9.726
1CMT+MM	SERIAL	24hr TAT	100	12	3.263	3.185	3.388	10.08	3.888	10.4
1CMT+MM	COMPOSITE	24hr TAT	10	1	3.786	4.678	4.308	13.54	2.909	11.91
1CMT+MM	COMPOSITE	24hr TAT	10	4	3.544	4.288	5.699	13.18	4.995	12.28
1CMT+MM	COMPOSITE	24hr TAT	10	12	3.001	4.644	4.643	13.52	3.367	11.96
1CMT+MM	COMPOSITE	24hr TAT	30	1	3.472	4.949	2.865	14.24	5.8	12.81
1CMT+MM	COMPOSITE	24hr TAT	30	4	3.562	4.826	4.148	13.6	2.608	12.83
1CMT+MM	COMPOSITE	24hr TAT	30	12	3.763	5.2	5.31	13.39	4.779	13.14
1CMT+MM	COMPOSITE	24hr TAT	100	1	3.168	4.877	3.489	14.78	5.073	11.72
1CMT+MM	COMPOSITE	24hr TAT	100	4	3.176	5.163	3.82	13.59	5.091	13.09
1CMT+MM	COMPOSITE	24hr TAT	100	12	3.286	5.171	3.485	13.09	4.549	12.53

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL	SRME% : NCA-MEAN	CV% : NCA-MEAN	SRME% : NCA-GEOMEAN	CV% : NCA-GEOMEAN
2CMT	SERIAL	24hr TAT	10	1	0.7214	3.072	5.377	10.74	4.592	10.37
2CMT	SERIAL	24hr TAT	10	4	0.5198	2.962	4.408	10.03	4.979	9.128
2CMT	SERIAL	24hr TAT	10	12	0.7113	2.8	2.965	10.73	5.301	10.43
2CMT	SERIAL	24hr TAT	30	1	0.7214	3.072	3.161	9.763	3.817	9.643
2CMT	SERIAL	24hr TAT	30	4	0.5198	2.962	3.984	9.413	4.117	9.773
2CMT	SERIAL	24hr TAT	30	12	0.7113	2.8	3.268	9.713	4.137	9.669
2CMT	SERIAL	24hr TAT	100	1	0.7214	3.072	3.36	10.37	4.292	10.07
2CMT	SERIAL	24hr TAT	100	4	0.5198	2.962	5.182	9.81	2.393	10.03
2CMT	SERIAL	24hr TAT	100	12	0.7113	2.8	3.436	10.14	3.129	9.555
2CMT	COMPOSITE	24hr TAT	10	1	0.63	4.583	4.218	11.94	3.532	13.95
2CMT	COMPOSITE	24hr TAT	10	4	0.9004	4.939	4.626	14.21	3.189	12.9
2CMT	COMPOSITE	24hr TAT	10	12	1.214	4.747	3.387	14.45	4.046	12.71
2CMT	COMPOSITE	24hr TAT	30	1	0.63	4.583	3.724	12.65	4.101	12.58
2CMT	COMPOSITE	24hr TAT	30	4	0.9004	4.939	5.745	12.5	5.615	13.23
2CMT	COMPOSITE	24hr TAT	30	12	1.214	4.747	4.28	12.28	4.449	13.19
2CMT	COMPOSITE	24hr TAT	100	1	0.63	4.583	3.513	12.72	5.004	11.98
2CMT	COMPOSITE	24hr TAT	100	4	0.9004	4.939	4.212	11.98	4.553	12.29
2CMT	COMPOSITE	24hr TAT	100	12	1.214	4.747	2.435	13.04	5.184	14.3

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL
1CMT	SERIAL	6 mth CMAX	10	1	0.3527	2.86
1CMT	SERIAL	6 mth CMAX	10	4	0.4266	3.159
1CMT	SERIAL	6 mth CMAX	10	12	0.5896	3.19
1CMT	SERIAL	6 mth CMAX	30	1	0.3527	2.86
1CMT	SERIAL	6 mth CMAX	30	4	0.4266	3.159
1CMT	SERIAL	6 mth CMAX	30	12	0.5896	3.19
1CMT	SERIAL	6 mth CMAX	100	1	0.3527	2.86
1CMT	SERIAL	6 mth CMAX	100	4	0.4266	3.159
1CMT	SERIAL	6 mth CMAX	100	12	0.5896	3.19
1CMT	SERIAL	6 mth cum.AUC	10	1	0.6984	2.862
1CMT	SERIAL	6 mth cum.AUC	10	4	0.7434	3.223
1CMT	SERIAL	6 mth cum.AUC	10	12	0.8577	3.317
1CMT	SERIAL	6 mth cum.AUC	30	1	0.6984	2.862
1CMT	SERIAL	6 mth cum.AUC	30	4	0.7434	3.223
1CMT	SERIAL	6 mth cum.AUC	30	12	0.8577	3.317
1CMT	SERIAL	6 mth cum.AUC	100	1	0.6984	2.862
1CMT	SERIAL	6 mth cum.AUC	100	4	0.7434	3.223
1CMT	SERIAL	6 mth cum.AUC	100	12	0.8577	3.317
1CMT	COMPOSITE	6 mth CMAX	10	1	1.031	4.775
1CMT	COMPOSITE	6 mth CMAX	10	4	0.8082	4.768
1CMT	COMPOSITE	6 mth CMAX	10	12	0.7882	4.773
1CMT	COMPOSITE	6 mth CMAX	30	1	1.031	4.775
1CMT	COMPOSITE	6 mth CMAX	30	4	0.8082	4.768
1CMT	COMPOSITE	6 mth CMAX	30	12	0.7882	4.773
1CMT	COMPOSITE	6 mth CMAX	100	1	1.031	4.775
1CMT	COMPOSITE	6 mth CMAX	100	4	0.8082	4.768
1CMT	COMPOSITE	6 mth CMAX	100	12	0.7882	4.773
1CMT	COMPOSITE	6 mth cum.AUC	10	1	0.1272	4.794
1CMT	COMPOSITE	6 mth cum.AUC	10	4	0.808	4.862
1CMT	COMPOSITE	6 mth cum.AUC	10	12	0.7279	4.937
1CMT	COMPOSITE	6 mth cum.AUC	30	1	0.1272	4.794
1CMT	COMPOSITE	6 mth cum.AUC	30	4	0.808	4.862
1CMT	COMPOSITE	6 mth cum.AUC	30	12	0.7279	4.937
1CMT	COMPOSITE	6 mth cum.AUC	100	1	0.1272	4.794
1CMT	COMPOSITE	6 mth cum.AUC	100	4	0.808	4.862
1CMT	COMPOSITE	6 mth cum.AUC	100	12	0.7279	4.937

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL
1CMT+MM	SERIAL	6 mth CMAX	10	1	12.84	2.748
1CMT+MM	SERIAL	6 mth CMAX	10	4	12.57	2.88
1CMT+MM	SERIAL	6 mth CMAX	10	12	13.02	3.121
1CMT+MM	SERIAL	6 mth CMAX	30	1	12.78	3.111
1CMT+MM	SERIAL	6 mth CMAX	30	4	13.11	3.019
1CMT+MM	SERIAL	6 mth CMAX	30	12	12.6	3.09
1CMT+MM	SERIAL	6 mth CMAX	100	1	12.62	3.252
1CMT+MM	SERIAL	6 mth CMAX	100	4	13.14	2.978
1CMT+MM	SERIAL	6 mth CMAX	100	12	13.02	3.437
1CMT+MM	SERIAL	6 mth cum.AUC	10	1	3.455	3.015
1CMT+MM	SERIAL	6 mth cum.AUC	10	4	3.266	3.138
1CMT+MM	SERIAL	6 mth cum.AUC	10	12	3.178	3.18
1CMT+MM	SERIAL	6 mth cum.AUC	30	1	3.387	3.581
1CMT+MM	SERIAL	6 mth cum.AUC	30	4	3.396	3.254
1CMT+MM	SERIAL	6 mth cum.AUC	30	12	2.9	3.18
1CMT+MM	SERIAL	6 mth cum.AUC	100	1	3.306	3.238
1CMT+MM	SERIAL	6 mth cum.AUC	100	4	3.021	3.041
1CMT+MM	SERIAL	6 mth cum.AUC	100	12	2.919	3.32
1CMT+MM	COMPOSITE	6 mth CMAX	10	1	12.83	5.214
1CMT+MM	COMPOSITE	6 mth CMAX	10	4	13.02	5.138
1CMT+MM	COMPOSITE	6 mth CMAX	10	12	12.93	5.137
1CMT+MM	COMPOSITE	6 mth CMAX	30	1	12.9	5.171
1CMT+MM	COMPOSITE	6 mth CMAX	30	4	13.25	4.735
1CMT+MM	COMPOSITE	6 mth CMAX	30	12	12.7	5.002
1CMT+MM	COMPOSITE	6 mth CMAX	100	1	12.74	5.135
1CMT+MM	COMPOSITE	6 mth CMAX	100	4	12.55	4.757
1CMT+MM	COMPOSITE	6 mth CMAX	100	12	12.34	4.913
1CMT+MM	COMPOSITE	6 mth cum.AUC	10	1	3.005	4.78
1CMT+MM	COMPOSITE	6 mth cum.AUC	10	4	3.895	4.873
1CMT+MM	COMPOSITE	6 mth cum.AUC	10	12	3.81	5.018
1CMT+MM	COMPOSITE	6 mth cum.AUC	30	1	3.572	4.78
1CMT+MM	COMPOSITE	6 mth cum.AUC	30	4	3.798	5.084
1CMT+MM	COMPOSITE	6 mth cum.AUC	30	12	3.044	4.721
1CMT+MM	COMPOSITE	6 mth cum.AUC	100	1	3.439	4.822
1CMT+MM	COMPOSITE	6 mth cum.AUC	100	4	3.656	5.019
1CMT+MM	COMPOSITE	6 mth cum.AUC	100	12	3.076	5.491

MODEL	SAMP	LABEL	DOSE	WEEK	SRME% : MODEL	CV% : MODEL
2CMT	SERIAL	6 mth CMAX	10	1	0.6298	2.803
2CMT	SERIAL	6 mth CMAX	10	4	1.158	3.014
2CMT	SERIAL	6 mth CMAX	10	12	0.6211	3.021
2CMT	SERIAL	6 mth CMAX	30	1	0.6298	2.803
2CMT	SERIAL	6 mth CMAX	30	4	1.158	3.014
2CMT	SERIAL	6 mth CMAX	30	12	0.6211	3.021
2CMT	SERIAL	6 mth CMAX	100	1	0.6298	2.803
2CMT	SERIAL	6 mth CMAX	100	4	1.158	3.014
2CMT	SERIAL	6 mth CMAX	100	12	0.6211	3.021
2CMT	SERIAL	6 mth cum.AUC	10	1	0.3169	2.891
2CMT	SERIAL	6 mth cum.AUC	10	4	0.811	3.139
2CMT	SERIAL	6 mth cum.AUC	10	12	0.6379	2.703
2CMT	SERIAL	6 mth cum.AUC	30	1	0.3169	2.891
2CMT	SERIAL	6 mth cum.AUC	30	4	0.811	3.139
2CMT	SERIAL	6 mth cum.AUC	30	12	0.6379	2.703
2CMT	SERIAL	6 mth cum.AUC	100	1	0.3169	2.891
2CMT	SERIAL	6 mth cum.AUC	100	4	0.811	3.139
2CMT	SERIAL	6 mth cum.AUC	100	12	0.6379	2.703
2CMT	COMPOSITE	6 mth CMAX	10	1	0.611	4.481
2CMT	COMPOSITE	6 mth CMAX	10	4	0.4159	4.665
2CMT	COMPOSITE	6 mth CMAX	10	12	0.8378	4.972
2CMT	COMPOSITE	6 mth CMAX	30	1	0.611	4.481
2CMT	COMPOSITE	6 mth CMAX	30	4	0.4159	4.665
2CMT	COMPOSITE	6 mth CMAX	30	12	0.8378	4.972
2CMT	COMPOSITE	6 mth CMAX	100	1	0.611	4.481
2CMT	COMPOSITE	6 mth CMAX	100	4	0.4159	4.665
2CMT	COMPOSITE	6 mth CMAX	100	12	0.8378	4.972
2CMT	COMPOSITE	6 mth cum.AUC	10	1	0.9684	4.76
2CMT	COMPOSITE	6 mth cum.AUC	10	4	1.059	4.849
2CMT	COMPOSITE	6 mth cum.AUC	10	12	0.6126	4.281
2CMT	COMPOSITE	6 mth cum.AUC	30	1	0.9684	4.76
2CMT	COMPOSITE	6 mth cum.AUC	30	4	1.059	4.849
2CMT	COMPOSITE	6 mth cum.AUC	30	12	0.6126	4.281
2CMT	COMPOSITE	6 mth cum.AUC	100	1	0.9684	4.76
2CMT	COMPOSITE	6 mth cum.AUC	100	4	1.059	4.849
2CMT	COMPOSITE	6 mth cum.AUC	100	12	0.6126	4.281

CHAPTER 4

Application of optimal design concepts to experimental protocols for the evaluation of toxicokinetics and safety thresholds

Tarjinder Sahota, Núria Buil-Bruna, Meindert Danhof and Oscar Della Pasqua

Submitted for publication (Pharmaceutical Research)

Abstract

Purpose: In toxicology experiments measures of drug exposure are calculated using non-compartmental methods, despite evidence that population pharmacokinetic (PK) modelling can provide accurate estimates of the parameters of interest. Here we explore the utility of optimised protocol design and PK modelling on the precision of exposure measures for a variety of hypothetical compounds.

Methods: Optimal design concepts were applied to a range of hypothetical drugs with different pharmacokinetic profiles. Protocol designs were optimised both in terms of sampling schedule and number of animals per group. The precision of secondary parameters, namely AUC and C_{MAX} was used as target for optimization purposes. Adequate precision levels were defined as expected CV% < 40%. Absolute changes in expected precision of less than 10% were deemed acceptable.

Results: Independent of differences in drug disposition, our results show that the number of animals used in experimental protocols can be reduced by 2/3 with acceptable loss of precision in AUC and C_{MAX} estimates. Even though some PK parameters were found to be imprecisely estimated when drug disposition involves more than one compartment, this does not significantly affect the secondary parameters describing systemic exposure, which showed adequate precision (all CVs <36%).

Conclusions: The accuracy and precision of measures of systemic exposure such as AUC and C_{MAX} are essential to ensure appropriate interpretation of experimental findings and make inferences about safety risk in humans. However, our analysis reveals that for composite methods, which are commonly used in toxicology protocols, sample size does not determine the precision of the pharmacokinetic parameters of interest. Rather, it is the sampling scheme and dose levels which matter. In contrast to current practice, precise calculation of safety thresholds can be obtained with a considerable reduction in the number of animals used in a typical protocol.

Introduction

Despite the evidence for important limitations in the assessment of non-clinical safety and toxicology, experimental protocols and data analysis have not advanced in the same way risk management concepts have evolved over the last decade (81). Drug exposure remains a proxy for risk even when other markers of safety and toxicity might be better predictors of adverse drug reactions (5). In fact, the establishment of safe exposure levels prior to first time in human studies is still one of the most important milestones in drug development (6,7). Yet, the reliability of these estimates depends on the quality, accuracy and precision of the data obtained from preclinical toxicology experiments. Even though statistical considerations are described in current guidelines, these methodological aspects appear to remain beyond the scope of the scientific debate on the relevance of safety thresholds.

Undoubtedly, prediction of safety thresholds is fraught with various challenges from a scientific, statistical and practical perspective. As shown in Table 1, strengths and weaknesses exist for the different methods currently used for the assessment of safe exposure, whether based on thresholds or not (8). These challenges are often compounded by the restrictive nature of regulatory guidelines for the evaluation of safety pharmacology and toxicity. Typically, experimental protocols for general toxicity used for defining safe exposure ranges in dose escalation (i.e., first-time-in-humans) studies rely on sparse sampling of pharmacokinetic data and other relevant safety measures. Samples are collected according to a pre-defined sampling matrix with a fixed number of animals per time point. Measures of drug exposure are then derived by naive pooling of the data to generate using composite parameters such as AUC and C_{MAX}. Subsequently, these parameters are used to establish the no-adverse-event-level (NOAEL), which determines the maximum allowed exposure during dose escalation in clinical trials (82).

Table 1 Safety thresholds and prediction of risk in humans. Reprinted with permission from Edler *et al.* (7).

	Strengths	Limitations and Weakness
SAR and TTC	• Avoids unnecessary animal testing	• Assumes that structure predicts toxicity • Depends on current exposure estimates for the population
Threshold	• Is simple to apply and readily understood	• Assumes the existence of a threshold • The NOAEL does not exclude biologically significant effects below the sensitivity of the test • The value of the NOAEL depends on experimental conditions such as group size, sensitivity of measurement of the adverse effect, and dose spacing. • Does not make full use of the dose-response information • Uses default UFs
CSAF modelling	• Chemical specific data can be incorporated to reduce uncertainty	• Depends on the validity of the subdivision of the 10-fold factors • Is a data intensive method
Non-threshold	• Linear extrapolation is simple to apply	• Linear extrapolation is thought to be highly conservative. • LMS cannot be validated as a model for low doses and extrapolation is model dependent • Differing balances between reactivity and repair between low and high doses are not accommodated.
BMD	• Makes full use of the dose-response data • Allows confidence limits for point estimates • An optimal experimental design may allow reduction of the number of animals tested (does not require a large number of	• Obtaining consensus defining a benchmark response level for the adverse effect (e.g. 5 or 10%) is difficult • Is not applicable to studies with few dose groups

	animals per group)	
Probabilistic RA	• Uncertainties associated with all aspects of the quantitative methods of the RA process can be taken into account • Appropriate chemical specific information can be incorporated to reduce uncertainty • Provides effect estimates at actual exposure levels	• Requires use of default distributions in most cases
Categorical regression	• Takes all studies into account and not only the most sensitive one • Allows the prediction of a severity effect category at a particular dose (e.g. above ADI)	• Requires toxicological judgement for the categorisation. • The interpretation of fitted model (different endpoints, observer variation etc.) is difficult
PBTK	• Is able to model the time course of the amount of the active compound at the target site • Is possible for any species and for different exposure (e.g. route to route extrapolation) and lifetime conditions • Allows extrapolation from animal to human without having to have human exposure data • Allows target organ dose-response relationships to be used for low-dose extrapolation	• Is a data intensive method • Does not address the dynamics

Given the importance to explore pharmacologically relevant exposure levels in humans, it should be clear that the accuracy of such estimates can become a critical factor during the dose escalation. To date, current guidelines do not describe the implications of variability or bias in these estimates. Yet, the NOAEL is often presented as point estimates to describe the population (22). This ignores variability which can be decomposed into two parts; variability associated with estimation methods and biological variation in pharmacokinetics which arises from inter- and intra-individual differences. Most importantly the exposure estimates from composite measures such as AUC do not allow accurate inferences about the underlying pharmacokinetic processes and individual concentration-effect relationships.

In a previous investigation we have shown that lack of precision exists in exposure estimates derived from the empirical methods currently used for the estimation of toxicokinetic (Sahota et al, unpublished results). One of the main problems is that drug exposure levels observed in satellite animals do not necessarily mirror those assigned to the primary treatment group, in which safety pharmacology and toxicity are evaluated. Evidence form long-standing pharmacokinetic research in pre-clinical species clearly shows that such an approach ignores important differences that may exist between the two experimental groups (11, 12). It is equivalent to assuming that all animals have the same exposure and variability in exposure, i.e., that the underlying physiological processes do not vary between animals. By contrast, the use of a model-based approach enables one to incorporate prior knowledge and additional data from other experiments into the analysis, providing accurate estimates of between- and within-subject variability. This information is essential to ensure a more quantitative, unbiased evaluation of safety pharmacology and toxicology findings.

Arguably, one should not consider only the implications of the statistical method for the analysis and interpretation of safety thresholds, but also question whether experimental protocols are informative enough to allow accurate estimation of the parameters of interest.

In this context, there has been an increase in the awareness about the relevance of optimality concepts for the optimisation and selection of suitable protocol designs for the

evaluation of pharmacokinetic data in conjunction with non-linear mixed-effects modelling. The statistical method was first proposed by Fedorov and later adopted into the PKPD field (83). The approach enables the prospective prediction of parameter precision in the protocol development phase using the expected fisher information matrix (FIM). Variations or adaptations to the original methods have been introduced, which have enable further use of optimality concepts in experimental protocols involving different types of continuous, repeated measurements (84,85). In addition to enhancing the informative value of experimental protocols, the use of optimal design has proven to be an opportunity for reduction in total sample size and consequently in the number of animals required for an experiment (86). Of particular relevance for the evaluation of safety protocols is the possibility of building robust designs to prior uncertainty in pharmacokinetic parameters. Model uncertainty can be explored via sensitivity analysis or by of applying ED-optimality which assumes a prior distribution around the parameters of interest (87).

In the current investigation, simulations are used to illustrate how a model-based approach can be implemented in conjunction with D-optimality software to improve the design of protocols for safety pharmacology and toxicology experiments. It can be anticipated that improved parameter precision and accuracy will allow appropriate dose escalation with less uncertainty about the safety thresholds (20). In fact, our analysis includes an evaluation of the sensitivity to model and parameter uncertainty (21). Furthermore, we also show how to account for the principle of the 3 Rs to ensure that the optimisation procedures do not represent an additional burden to animals required for the experiments (4).

Methods

Currently available software programs have two major limitations for optimising general toxicity protocols. The first is that optimisation is performed with respect to primary model parameters (e.g. CL, Vd). This is restrictive because measures of interest in toxicology are

secondary parameters such AUC and C_{MAX}. For instance, for AUC estimation, the precision of KA is of little importance. Similarly, for most drugs, precise estimation of C_{MAX} will not depend on the precision of CL and peripheral compartment parameters. An optimisation routine that optimises over all parameters may not be suitable either. Ideally, it would be useful to reparameterise the model so that derived measures of exposure are treated as optimisation variables, but this is not always possible as there may be no closed form solution relating primary and secondary model parameters.

The second problem arises from the tendency of software to only provide optimal solutions. In practice there are many other factors to consider (e.g. logistical, ethical, financial, and/or minimal false positive rate) which can be difficult to account for within the optimisation options in a software program. For example, there may be suboptimal designs (in terms of expected parameter precision) that are much more cost effective or ethical. It is therefore important to be able to explore the space of candidate study designs achieving a desired level of precision.

To address the aforementioned problems we proposed to use a simulation-re-estimation approach to study design. However, this is computationally intensive and can quickly become unfeasible when applied to variety of candidate designs and proposal models. For this reason, here we employ a hybrid approach where candidate designs are evaluated in PopED v. 2.10 (University of Uppsala, Sweden) and then expected primary parameter (co)variances are converted to secondary parameter variances using traditional PKPD simulation procedures, as implemented in NONMEM v.6.2 (ICON Development Solutions. Hanover, Maryland).

The studies under consideration were a one week, one month, and three-month general toxicology protocol, in which toxicokinetic data for three different hypothetical drugs were evaluated. Given the pre-defined pharmacokinetic parameters used in the simulations, the true exposure for each individual animal was computed using a variety of measures which were subsequently set as reference for further assessment of the no adverse effect level (NOAEL).

Finally, it should be noted that one of the main issues with the estimation of the NOAEL is that it is limited to the computed exposure at one of the pre-specified experimental doses (22). Consequently, the estimated exposure at any one of the dose levels is a candidate threshold depending on the observed adverse events. To overcome this limitation, the assessment of experimental designs was primarily based on the estimates from secondary parameters (AUC and C_{MAX}) across all treatment groups. In addition, our design space was limited to sampling schedule and number of animals per group to ensure that the NOAEL estimates could be obtained both by NCA and non-linear mixed effects methods. In fact, only experimental designs which allowed for the analysis of the data according to both methods were evaluated.

Given that in typical experimental protocols, three animals are sampled per time point for toxicokinetic analysis, alternative candidate designs were aimed at reducing total sample size, including two or even one animal per sampling time point. These alternative designs represent therefore a reduction in the total number of samples and in the number of animals required per study. Details of the experimental protocols, pharmacokinetic models and optimisation procedures are described in details in the next paragraphs.

Experimental protocols: Three hypothetical drugs were considered to account for differences in disposition properties. We assumed the availability of prior information in the form of single dose pharmacokinetic experiments performed across a range of doses with putative pharmacological activity (1, 3, and 10 mg/kg), in which 8 animals were tested per cohort. The toxicology protocol design was based on an initial set-up commonly used for chronic toxicity evaluation. Four treatment groups (N= 8 per group) receiving oral daily doses of vehicle, 10, 30, and 100 mg/kg/day were tested throughout this set of virtual experiments, which lasted either one week, one month or three months. Satellite groups with 3 animals/time point were used to mimic the dosing conditions in the animals used for the assessment of toxicity (see Figure 1 for a simulation of typical satellite group data). This procedure ensures the availability of more frequent blood samples for toxicokinetics. Blood sampling scheme included four occasions based on feasibility, namely days 1, 8, 25, and 89.

Sampling times on those days were determined by ED-optimality. For the purposes of optimisation, we assumed that all three hypothetical drugs could be fitted by a one-compartment model (model A1) and assumed a 50% CV on all parameters. This was intended to represent standard use of ED-optimality for the optimisation of sampling times. Sampling times were rounded to the nearest 15 minutes.

Pharmacokinetic models: To ensure accurate evaluation of the impact that differences in drug disposition may have on the requirements for experimental design optimisation, three different scenarios were considered in which hypothetical drugs showing on a one-compartment pharmacokinetics with linear and nonlinear (Michaelis-Menten) elimination as well as a two-compartment pharmacokinetics were tested. Parameter values for each scenario are shown in Table 2. In all scenarios, residual variability was assumed to be 15%. Moreover, for the purposes of this exercise, we have assumed a homogeneous population, avoiding the need to explore covariate relationships in any of the models.

Optimisation criteria: See the appendix for background information on the optimality concepts used in this investigation. ED-optimality can be used to incorporate parameter uncertainty into the optimisation process. However, ED optimality only provides an assessment of expected parameter precision and provides no basis for exploration of suboptimal, yet sufficient designs, i.e. reduced designs. Therefore, our decision to use the expected FIM explicitly for the prediction of parameter precision is motivated by a need to have a fast, reliable and flexible method to assess and optimise experimental designs for a model-based analysis whilst adhering to the principle of the 3 Rs. The expected FIM provides a close approximation of expected parameter uncertainty (23,24). In addition, we have favoured the practice of explicitly running the optimisation at different perturbations in model parameters (Table 3). Model parameters were changed in the three PK models tested (one compartment with linear and nonlinear elimination and two compartments), yielding to a total of 27 different models. These models are labelled A1...9, B1....9 and C1....9.

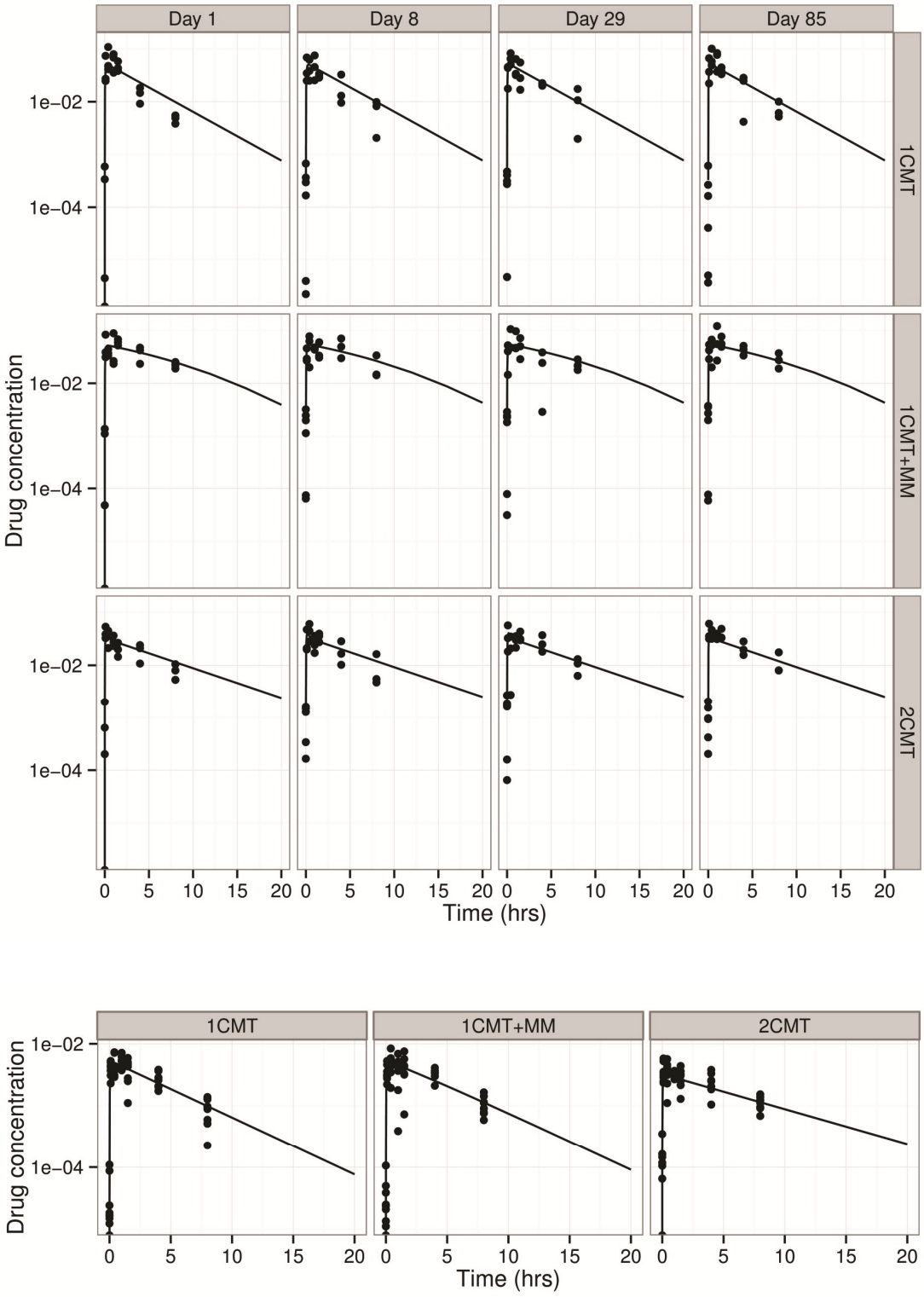


Figure 1: Plots of simulated data for scenarios A1, B1, and C1 overlaid with population prediction (black line). Top panel shows 10mg/kg dosing group using the 3 samples per time point. Bottom panel shows pharmacokinetic profiles at the lower dose level (1 mg/kg) with 8 animals per cohort.

Table 2: Parameters and corresponding between-subject variability used to characterise the pharmacokinetic profiles of hypothetical compounds showing one-compartment, two-compartment and Michaelis-Menten disposition in rats. Doses were defined according to a general toxicology protocol design. Ke: first order rate constant of elimination, Ka: first order rate constant of absorption V: volume of distribution, K_{12}: hybrid constant, K_{21}: hybrid constant; Vmax: maximum metabolic rate ; Km: Michaelis-Menten constant (substrate concentration corresponding to 0.5 V_{max})

MODEL A:

Parameter	Value	BSV (%)
CL (ml/h)	10	20
Ka (h-1)	14.82	50
V (mL)	49	16

MODEL B:

Parameter	Value	BSV (%)
CL (ml/h)	10	20
Ka (h-1)	14.82	50
V (mL)	49	16
K12(h-1)	2.17	16
K21(h-1)	3.554	69

MODEL C:

Parameter	Value	BSV (%)
Vmax (mg/h)	0.3	20
Ka (h-1)	14.82	50
V (mL)	49	16
Km(mg/L)	30	0 FIX

Table 3: Perturbations in the parameters for the three different pharmacokinetic models. CL: clearance, Ka: first order rate constant of absorption V: volume of distribution, Vmax: maximum metabolic rate.

Model	KA	V	CL	Model	KA	V	CL	Model	KA	V	VMAX
A1	-	-	-	B1	-	-	-	C1	-	-	-
A2	-	+50%	+50%	B2	-	+50%	+50%	C2	-	+50%	+50%
A3	-	+50%	-50%	B3	-	+50%	-50%	C3	-	+50%	-50%
A4	-	-50%	+50%	B4	-	-50%	+50%	C4	-	-50%	+50%
A5	-	-50%	-50%	B5	-	-50%	-50%	C5	-	-50%	-50%
A6	-80%	+50%	+50%	B6	-80%	+50%	+50%	C6	-80%	+50%	+50%
A7	-80%	+50%	-50%	B7	-80%	+50%	-50%	C7	-80%	+50%	-50%
A8	-80%	-50%	+50%	B8	-80%	-50%	+50%	C8	-80%	-50%	+50%
A9	-80%	-50%	-50%	B9	-80%	-50%	-50%	C9	-80%	-50%	-50%

All evaluations were performed in PopED v.2.10 (University of Uppsala, Sweden) (88), a software developed in O-Matrix® (Harmonic Software Inc., Seattle, WA, USA). Data manipulation and statistical and graphical summaries were performed in R 2.10.0 (26). In our analysis, the expected FIM was used to compute the expected covariance matrix from which, the expected precision of primary pharmacokinetic parameters was quantified (89,90).

The expected precision of the derived parameters of interest, namely AUC and C_{MAX}, were calculated from the expected covariance matrix of primary parameters in NONMEM 6.2 (ICON Development Solutions. Hanover, Maryland) (27). First, 1000 pharmacokinetic profiles were simulated from the primary parameters uncertainty distributions by including the covariance information in the $PRIOR subroutine. For each pharmacokinetic profile, the AUC and C_{MAX} were calculated as follows:

$$AUC = \int_{t-24}^{t} C_p dt$$

$$C_{MAX} = \max\left(\{C_p(s): t-24 < s < t\}\right)$$

where individual predicted drug concentrations are denoted by $C_p(t)$.

The expected precision (standard error) of the parameters was then summarised. Adequate precision was defined as expected CV% < 40%. Absolute changes in expected precision of less than 10% were deemed biologically irrelevant.

Results

Our analysis shows that optimal design concepts can be used in toxicology research to improve the precision of the parameters of interest whilst allowing for a reduction in the total number of animals required per experiment. As shown in figure 1, plots of the

simulated profiles for a typical individual together with simulated samples, representing "observed" data are depicted to illustrate the impact of different disposition characteristics on the concentration vs. time profiles.

The optimised sampling times for all scenarios were 0.25, 0.5,0.75, 1, 1.5, 2, 8 and 24 hours after dosing. Results show that for all designs the precision of AUC and C_{MAX} associated with a reduced sample size of 2/3 from the initial sample size resulted in an acceptable loss of precision (the absolute difference in expected precision was <10% for all scenarios for sample size reduction of 2/3). Therefore, optimised protocols result in a reduction of up to 2/3 in the number of animals utilised in toxicokinetic experiments.

An overview of the point estimates and coefficient of variation (CV%) obtained for AUC and C_{MAX} is presented in Table 5. The differences in parameter precision associated with varying sample size, including the NOAEL, is summarised for each model in Figures 2, 3 and 4). We show how precision changes when one or two animals are sampled at each time point instead of using 3 animals per sampling time point. Interestingly, the expected precision was very high for the one-compartmental model but there was less precision for the two-compartmental model, where a distribution phase is evident. In addition, our analysis reveals that metabolic saturation, as described by Michaelis-Menten kinetics does not further affect the precision of parameter estimates. Further assessment of the precision of the primary parameters indicates that the parameters governing peripheral compartment distribution will be the least precisely estimated, with a loss of precision as high as 75% for some parameter perturbations. Between-subject variability was also found to be imprecisely estimated and would have to be fixed to 0 for some parameters during data analysis. Yet, despite these differences, AUC and C_{MAX} imprecision was <36% for the two-compartmental models.

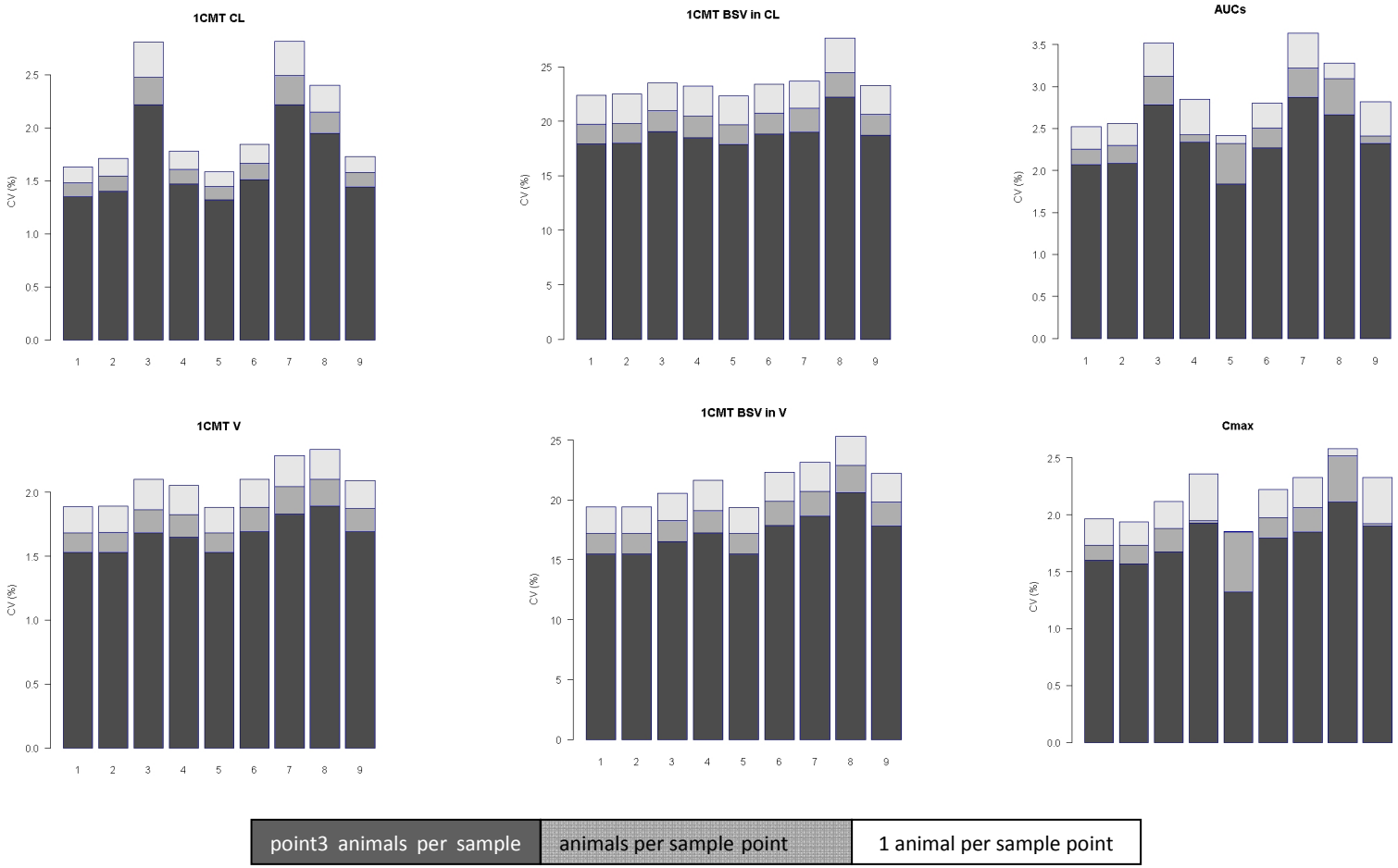


Figure 2. Bar charts of CVs of selected parameters for models A.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

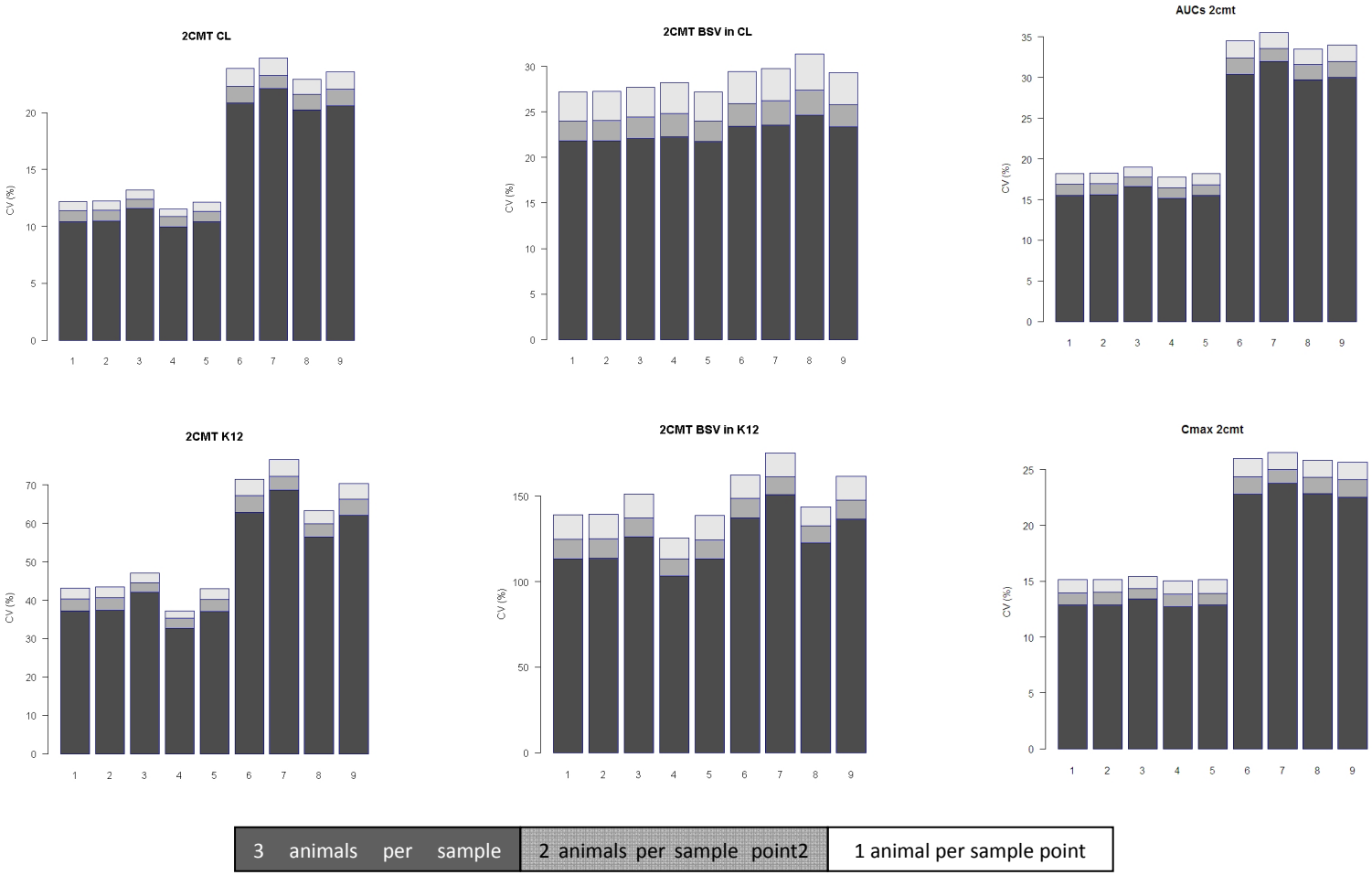


Figure 3. Bar charts of CVs of selected parameters for models B.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

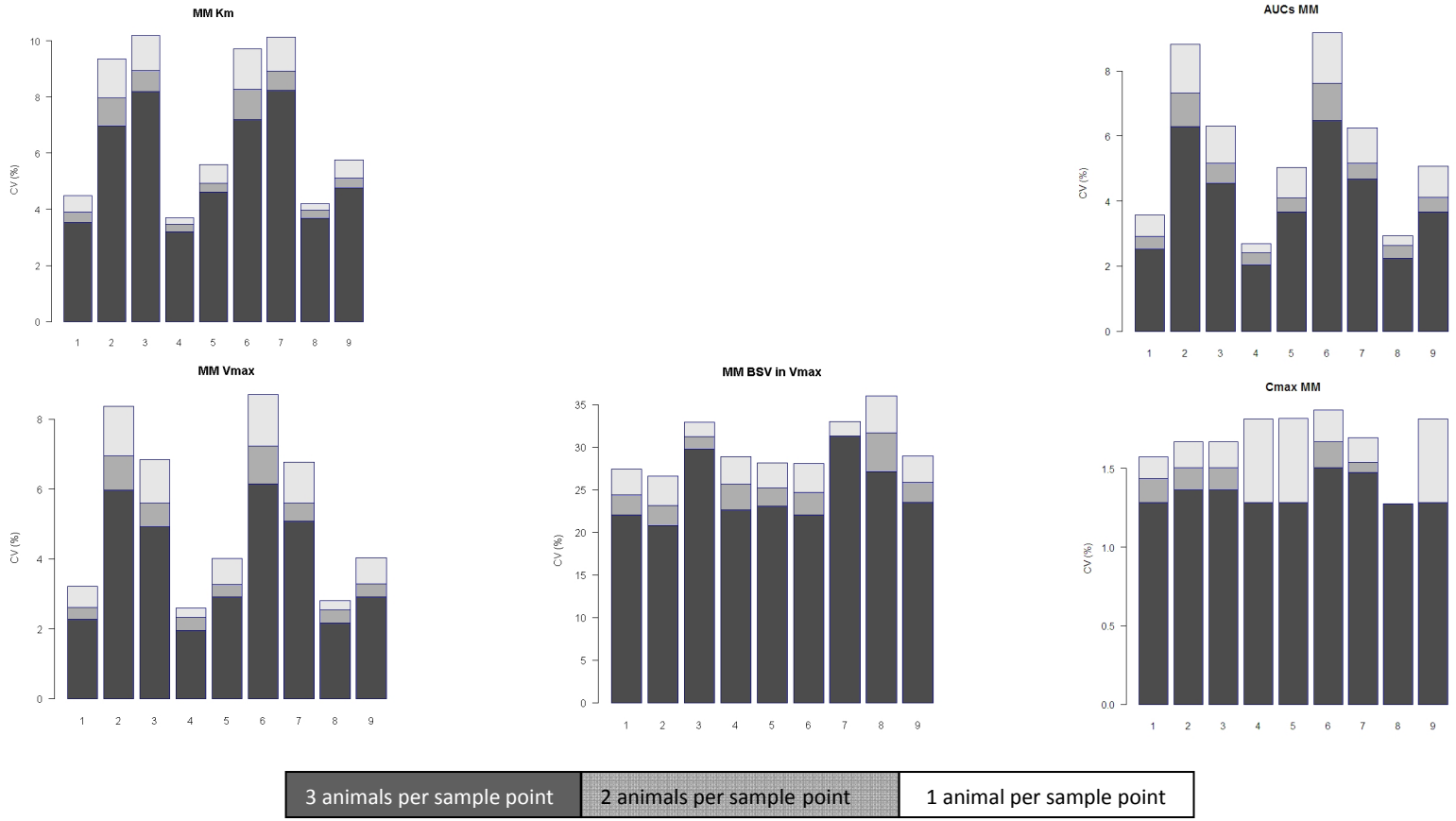


Figure 4. Bar charts of CVs of selected parameters for models C.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

Discussion

Experimental protocols based on repeated-dose treatment arms are essential for accurate inferences about the risk associated with the exposure to new chemical entities in the early phase of clinical development. These studies provide the basis for the calculation of safety thresholds such as the no-observed-adverse-effect level (NOAEL) or lowest-observed-adverse-effect level (LOAEL), which are used to extrapolate the concentration or exposure above which adverse effects can be expected in humans (82,91).

Despite the efforts and attention given to different methodologies for the estimation of such safety thresholds, it is now acknowledged that the use of NOAEL or LOAEL as traditional thresholds or point of departure for risk assessment has significant limitations. The NOAEL and LOAEL are determined by the selected dose levels and intervals used in an experimental protocol.

To date, these measures remain a requirement for regulatory purposes (2). However, there is a wide consensus that they do not mathematically relate to the underlying exposure-response curve (92). In addition, it has been shown that differences in protocol design can influence the precision and accuracy of the parameters of interest, yielding biased NOAEL and LOAEL estimates. In fact, the bench mark dose (BMD) as the threshold or point of departure has been proposed as an alternative method to avoid many of these pitfalls (41). Unfortunately, similar challenges exist with regard to the accuracy and precision of estimates obtained by the BMD (18,93). The experimental data are not integrated nor parameterised in a mechanistic manner so as to benefit from the advantages of a model-based approach.

Whilst risk assessment methods need undoubtedly to incorporate mechanistic aspects of drug action to ensure better characterisation of potential hazards to humans, it should be noted that improvements are also required from a statistical perspective. Thus far empiricism and regulatory-related issues have dominated traditional toxicological testing

paradigms (32-35). Minimal efforts have been made to introduce optimality concepts in experimental design as a means to increase accuracy and precision of the parameters of interest.

In this investigation we have attempted to show the feasibility of implementing a model-based approach in conjunction with optimal design based on techniques, which have been developed for the field of pharmacokinetics for more than two decades ago (13,36,37). By considering a number of hypothetical scenarios in which drugs with different disposition properties were simulated, we have demonstrated that accurate estimates of AUC and C_{MAX} can be obtained for drugs showing different pharmacokinetic profiles. Our results also highlight the impact of optimisation procedures on the estimation of secondary parameters. We have shown that even when precision of the primary pharmacokinetic parameter is poor, as in the case of parameters governing distribution into peripheral compartments, the precision of the secondary parameters remains unaffected. This can be attributed to the fact that the selected candidate designs systematically yield estimates of clearance and volume of distribution with acceptable precision. These two parameters ultimately determine systemic exposure and peak concentrations, respectively.

Although it may seem a disadvantage to use model-dependent estimates for the assessment of safety thresholds, this approach presents various important advantages (38-39). First, it is unbiased and predictive, allowing for the incorporation of the physiological factors underlying the pharmacokinetic properties of the drug under investigation. Moreover, it enables ne to integrate prior information, including data from other experiments. We anticipate that many areas in toxicology research which can benefit from such an approach. New methodology does not necessarily mean that human safety will be placed at risk. On the contrary, newer methods provide an opportunity to remove much of the guess work involved with older methodologies, which rely on assumptions which clearly prevent the uptake of evolving knowledge about pharmacokinetic and pharmacodynamic properties of a drug.

Methodological aspects

In assessing and optimising the protocol we found that existing routines in optimality software were insufficient to meet our assessment criteria. In particular, existing software did not enable the assessment and optimisation over arbitrary secondary parameters, and did not allow for the impact of parameter perturbations on expected precision to be assessed. The alternative brute force approach to account for these limitations would have been to perform multiple simulation-re-estimation procedures across our design and model space. However, this would have involved extensive computation times. Our approach instead consisted of FIM evaluations followed by calculation of the expected secondary parameter precision. This exercise ultimately showed that optimisation can be performed on secondary parameters of interest, and minimally sufficient designs can be obtained. Both of these procedures are computationally inexpensive. Our approach therefore enables exploration of large design and model spaces without the aforementioned limitations in current optimality software.

Limitations

Our work does involve a number of assumptions, which may represent potential theoretical and practical limitations. First, it should be noted that we have constrained ourselves to candidate designs that enable estimation of exposure using non-compartmental methods for each treatment group. Further gains in terms of reduced burden and/or parameter precision are likely to be achieved if a model-based analysis was the only intended analysis of the data.

Another requirement is the availability of a well-defined population pharmacokinetic model, which is feasible, but in practice not used in routine pre-clinical research. It should be clear that the computation of expected (co)variance by means of the FIM, cannot directly account for the possibility of unidentifiably of parameters. Hence, the validity of any optimisation procedures implies accurate knowledge of the pharmacokinetic properties and corresponding parameterisation. Parameter unidentifiability will likely manifest in terms of

large standard errors, high correlations in the correlations and/or large differences in eigenvalues. On the other hand, optimal design does tackle another common issue observed during data fitting and parameter estimation, i.e., numerical unidentifiability, which may be caused by poor experimental design.

An additional assumption is that parameter estimates will be unbiased. This assumption may not hold true for more complex models, but the reader should be aware that this issue may be equally important when non-compartmental methods are used to describe complex pharmacokinetic profiles, as for instance in the case of metabolic inhibition or drugs with long elimination half-life (40). To ensure further characterisation of bias, a full bootstrap (simulation-re-estimation) procedure is recommended. Lastly, one should realise the implications of our own objectives, i.e., to compare designs which are suitable for both non-compartmental and model-based methods. Further gains in terms of reduced burden and/or parameter precision are likely to be achieved if a model-based analysis was the only intended analysis of the data.

In summary, it can be concluded that despite the biological debate about the relevance of safety thresholds, the accuracy and precision of estimates are essential to ensure appropriate interpretation of experimental findings and make inferences about risk in humans. We have shown that the use of a model-based approach is critical for appropriate data integration and informative value of experimental protocols. Our work also demonstrates that population size is not the critical variable when evaluating precision and accuracy of the parameters of interest. This feature allows for comparable results to be obtained with considerable lower number of animals and consequently reduction in the cost of experiments. Overall, these results make the need to explore the requirements for further implementation of optimal design in toxicology research an ethical and scientific imperative.

Reference List

1. ICH Expert Working Group Guideline S3A. Note for guidance on toxicokinetics: The assessment of systemic exposure in toxicity studies, 1994.

2. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry M3(R2). Nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. 2010.

3 Hartford CG, Petchel KS, Mickail H, Perez-Gutthann S, McHale M, Grana JM, Marquez P. Pharmacovigilance during the pre-approval phases: an evolving pharmaceutical industry model in response to ICH E2E, CIOMS VI, FDA and EMEA/CHMP risk-management guidelines. Drug Saf. 2006; 29(8): 657-73.

4. Cooper AJ, Lettis S, Chapman CL, Evans SJ, Waller PC, Shakir S, Payvandi N, Murray AB. Developing tools for the safety specification in risk management plans: lessons learned from a pilot project. Pharmacoepidemiol Drug Saf. 2008; 17(5):445-54. doi: 10.1002/pds.1576.

5. Aronson JK, Ferner RE. Joining the DoTS: new approach to classifying adverse drug reactions. BMJ 2003; 327:1222-1225.

6. DeGeorge JJ. Food and Drug Administration viewpoints on toxicokinetics: the view from review. Toxicol Pathol. 1995; 23(2): 220-5.

7. Bessems JG, Geraets L. Proper knowledge on toxicokinetics improves human hazard testing and subsequent health risk characterisation. A case study approach. Regul Toxicol Pharmacol. 2013; 67(3):325-34.

8. Edler L, Poirier K, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K, Wurtzen G. Mathematical modelling and quantitative methods. Food Chem Toxicol. 2002; 40:283-326.

9. Contrera J F, Matthews EJ, Kruhlak NL, Benz RD. Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul Toxicol Pharmacol. 2004; 40:185-206.

10. Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol. 2005; 42:265-274.

11. Singh SS. Preclinical pharmacokinetics: an approach towards safer and efficacious drugs. Curr Drug Metab. 2006; 7(2):165-82.

12. Ding X, He M, Kulkarni R, Patel N, Zhang X. Investigation of clinical pharmacokinetic variability of an opioid antagonist through physiologically based absorption modeling. J Pharm Sci. 2013; 102(8):2859-74.

13. Fedorov V. Optimal experimental design. WIREs Comp Stat 2010; 2: 581-589.

14. Retout S, Mentre F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Stat Med. 2002; 21:2623-2639.

15. Duffull SB, Mentre F, Aarons L. Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm Res. 2001; 18:83-89.

16. Waterhouse TH, Redmann S, Duffull SB, Eccleston JA. Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. J Pharmacokinet Pharmacodyn. 2005; 32:521-545.

17. Ette EI, Howie CA, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in preclinical pharmacokinetic studies. Pharm Res. 1995; 12:729-737.

18. Tod M, Rocchisani JM. Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics. J Pharmacokinet Biopharm. 1997; 25:515-537

19. Hooker AC, Foracchia M, Dodds MG, Vicini P. An evaluation of population D-optimal designs via pharmacokinetic simulations. Ann Biomed Eng 2003; 31:98-111.

20. 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. The Three Rs: the way forward. Alternatives to Laboratory Animals 1995; 23: 838-866.

21. Nestorov I. Modelling and simulation of variability and uncertainty in toxicokinetics and pharmacokinetics. Toxicol Lett. 2001; 120:411-420.

22. Ogungbenro K, Gueorguieva I, Majid O, Graham G, Aarons L. Optimal design for multiresponse pharmacokinetic-pharmacodynamic models - dealing with unbalanced designs. J Pharmacokinet Pharmacodyn. 2007; 34:313-331.

23. Thai H, Mentre F, Holford NH, Veyrat-Follet C, Comets E. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics . J Pharmacokinet Pharmacodyn. (2014) 41:15–33

24. Hooker A, Vicini P. Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. AAPS J. 2005; 7:E759-E785.

25. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. Comput Methods Programs Biomed. 2004; 74:29-46.

26. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

27. Beal, S., Sheiner, L.B., Boeckmann, A., & Bauer, R.J., NONMEM User's Guides. (1989-2009), Icon Development Solutions, Ellicott City, MD, USA, 2009.

28. Huang C, Zheng M, Yang Z, Rodrigues AD, Marathe P. Projection of exposure and efficacious dose prior to first-in-human studies: how successful have we been? Pharm Res. 2008; 25:713-726.

29. Gibbs JP. Prediction of exposure-response relationships to support first-in-human study design. AAPSJ. 2010; 12:750-758.

30. Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C, Spitzer HL. Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul Toxicol Pharmacol. 1995; 21:296-306.

31. Budtz-Jorgensen E, Grandjean P, Keiding N, White RF, Weihe P. Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol Lett. 2000; 112-113: 193-199.

32. indgren S, Bass AS, Briscoe R, Bruse K, Friedrichs GS, Kallman MJ, Markgraf C, Patmore L, Pugsley MK. Benchmarking safety pharmacology regulatory packages and best practice. J Pharmacol Toxicol Methods. 2008; 58(2):99-109.

33. Rowbotham AL, Gibson RM. Exposure-driven risk assessment: applying exposure-based waiving of toxicity tests under REACH. Food Chem Toxicol. 2011; 49(8):1661-73.

34. Kooijman M, van Meer PJ, Moors EH, Schellekens H. Thirty years of preclinical safety evaluation of biopharmaceuticals: Did scientific progress lead to appropriate regulatory guidance? Expert Opin Drug Saf. 2012; 11(5):797-801

35. Schellekens H. Overregulation and unnecessary animal testing: requirements for market approval of biopharmaceuticals too rigid. Ned Tijdschr Geneeskd. 2012; 156(50):A5723.

36. D'Argenio, DZ. Optimal sampling times for pharmacokinetic experiments. J Pharmacokinet Biopharm. 1981; 9:739-756

37. al Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in population pharmacokinetics. J Pharmacokinet Biopharm. 1990; .18:347-360.

38. Beatty DA, Piegorsch WW. Optimal statistical design for toxicokinetic studies. Stat Methods Med Res. 1997; 6(4):359-76.

39. Dodds MG, Hooker AC, Vicini P. Robust population pharmacokinetic experiment design. J Pharmacokinet Pharmacodyn. 2005; 32(1):33-64.

40. Lehr T, Staab A, Trommeshauser D, Schaefer HG, Kloft C. Semi-mechanistic population pharmacokinetic drug-drug interaction modelling of a long half-life substrate and itraconazole. Clin Pharmacokinet. 2010; 49(1):53-66.

41. Nyberg J, Karlsson MO, Hooker AC. Simultaneous optimal experimental design on dose and sample times. J Pharmacokinet Pharmacodyn. 2009; 36:125-145.

42. Wang J, Endrenyi L. A computationally efficient approach for the design of population pharmacokinetic studies. J Pharmacokinet Biopharm. 1992; 20:279-294.

Appendix

In an optimal design exercise, design variables are variables that describe properties of the biological system, drug or experimental protocol which can be changed to explore their impact on the information contents of the experiment. Typically these include dose, sampling scheme, number of samples, number of individuals or other covariates (94). Even though the number of animals is constrained (88), the main use of this technique is to optimise sampling times. It has been shown that sample times can have significant influence in the accuracy and precision of parameters (95,96). By optimising sampling times it is possible therefore to improve the overall efficiency of PK experiments (96,97).

Here we summarise the statistical framework for the evaluation and optimisation of experimental designs using D-optimality. There are various software programs for optimal design, making them equally suitable for the purposes of this type of analysis. They differ primarily in the features available for optimisation and in the optimisation method.

Statistical summary

There are various numerical methods to fit a model to data. The mostly commonly used is the maximum likelihood (ML) estimator. The maximum likelihood is calculated by maximising the following likelihood function (L):

$$L(\theta) = p(D|\theta)$$

where θ is the vector of parameters, *D* is the data. The results of a maximum likelihood estimation are $\hat{\theta}$, the maximum likelihood estimate and cov($\hat{\theta}$), the covariance matrix determining the parameter precision. The information contents within the study data, *D* is what determines cov($\hat{\theta}$). Prior to running the experiment, assuming the availability of a model, it is possible to compute an expected covariance matrix by the use of the Cramer-Rao inequality:

$$cov(\hat{\theta}) \geq \frac{1}{FIM(\hat{\theta})}$$

where the Fisher Information Matrix (FIM) is given by

$$FIM(\hat{\theta}) = E\left[\left(\frac{\partial}{\partial\theta}L(\theta)\right)^{T}\left(\frac{\partial}{\partial\theta}L(\theta)\right)\right]$$

Although this function constrains the lower bound of cov($\hat{\theta}$), in practice such a lower bound is reached as indicated by comparisons with bootstrapped expected covariance estimates (98,99). Thus, by computing the FIM of a given design, under the assumption of no or minor model and parameter misspecification, one can estimate the covariance matrix and consequently assess parameter precision values. By maximising the determinant of the FIM over design variables, such, as for instance the sampling schedule, it is possible to identify experimental conditions or design(s) that maximise the expected parameter precision.

CHAPTER 5

Use of biomarkers for the characterisation of long-term safety: advantages of a model-based approach for the analysis of toxicology experiments

Tarjinder Sahota, Meindert Danhof and Oscar Della Pasqua

Submitted for publication (Pharmaceutical Research)

Abstract

Purpose: Toxicology assessment relies on the evidence of a direct relationship between observed systemic exposure and adverse events. This empirical approach prevents the identification and the use of suitable biomarkers associated with the underlying pharmacodynamic processes, which ultimately determine delayed toxicity. The objective of this investigation was therefore to explore the feasibility of applying a model-based approach to characterise the PKPD correlations and the time course of biomarker responses associated with long-term safety as compared to standard non-compartmental methods.

Methods: A hypothetical toxicology protocol was designed by simulating the pharmacokinetics and pharmacodynamics (biomarkers responses) of four different drugs, each with a different mechanism of delayed toxicity. The mechanisms of delayed toxicity were: i) indirect response mechanism, ii) indirect response mechanism preceded by biophase equilibration, iii) cumulative effects of chronic dosing and iv) formation of a toxic metabolite. In the simulations data were sampled according to standard experimental designs. Data for each drug were then analysed using non-compartmental methods and by nonlinear mixed effects modelling, as implemented in NONMEM v7.1. Given the often unknown mechanism of toxicity, a variety of models was evaluated to explore model misspecification. Finally, bias and precision of parameter estimates were compared for each method.

Results: The true underlying model was often unidentifiable. However, model approximations were identified for each scenario with satisfactory performance. NCA-derived estimates showed more bias and less precision for all methods in all scenarios. The relative errors were smaller for parameter estimates obtained by data fitting.

Conclusions: Integration of toxicokinetic and biomarker data is essential for the evaluation of long-term safety and toxicity. Despite issues due experimental protocol design, the use of a model-based approach enables the assessment of putative mechanisms of toxicity. Traditional techniques, such as non-compartmental methods are unsuitable for the characterisation of long term, delayed effects.

Introduction

Understanding of toxicokinetics during the evaluation of safety pharmacology and nonclinical toxicity has been considered essential for accurate prediction of safety thresholds for a new chemical or biological entity (1,2) (Figure 1). Increasingly, however, it has become evident that characterisation of the relationship between drug exposure, target engagement (i.e., activation or inhibition) and downstream biological effects associated with a given physiological pathway can provide further insight into the mechanisms underlying both expected and 'unexpected' toxicity. In fact, several novel toxicity biomarkers have emerged as sensitive tools for detection, monitoring, quantification and prediction of organ toxicity (3-5) (Figure 2). In addition, the use of a more mechanism-based approach for the evaluation of drug effects has allowed better interpretation of time-dependencies, which are often observed following chronic exposure to a drug (e.g., delayed toxicity) (6).

Whilst the availability of tissue-specific data can provide valuable information for decision making during toxicological assessment (7), empirical safety thresholds based on systemic drug exposure continue to prevail as the mainstream approach for assessing the safety profile of new chemical entities, preventing wider use of biomarkers and potential translation of pharmacological properties of a molecule from animals to man (8,9). These hurdles are perpetuated by the existing view or notion that experimental data represent the basis for characterising phenomena arising from causes that are unknown or uncertain, as is often the case in early drug development.

Thus far, little attention has been given to the possibility of evaluating toxicity using a mechanism-based approach whereby adverse events are assessed from a pharmacological perspective. Such an approach would allow information from putative biomarkers to be integrated with pharmacokinetic data to support inferences about observed and unobserved adverse events (10-12). In addition, the use of modelling and simulation would provide the opportunity to predict sub chronic and chronic safe exposure range in humans from preclinical experiments as well as investigate short and long term treatment effects.

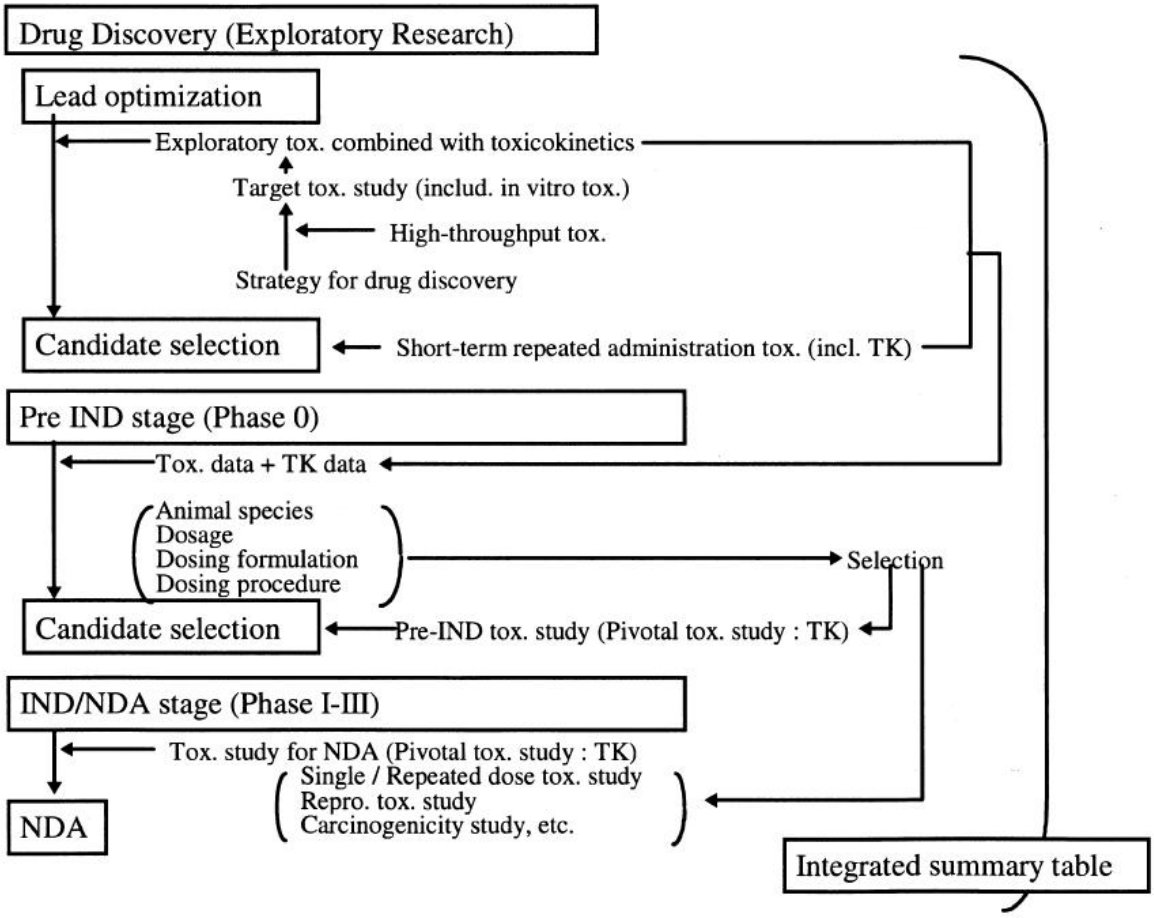


Figure 1 TK toxicokinetic studies in drug-development process. IND; investigational new drug application, NDA; new drug application. Reprinted with permission from Toxicology Letters 102-103, pages: 657-664 (1998)

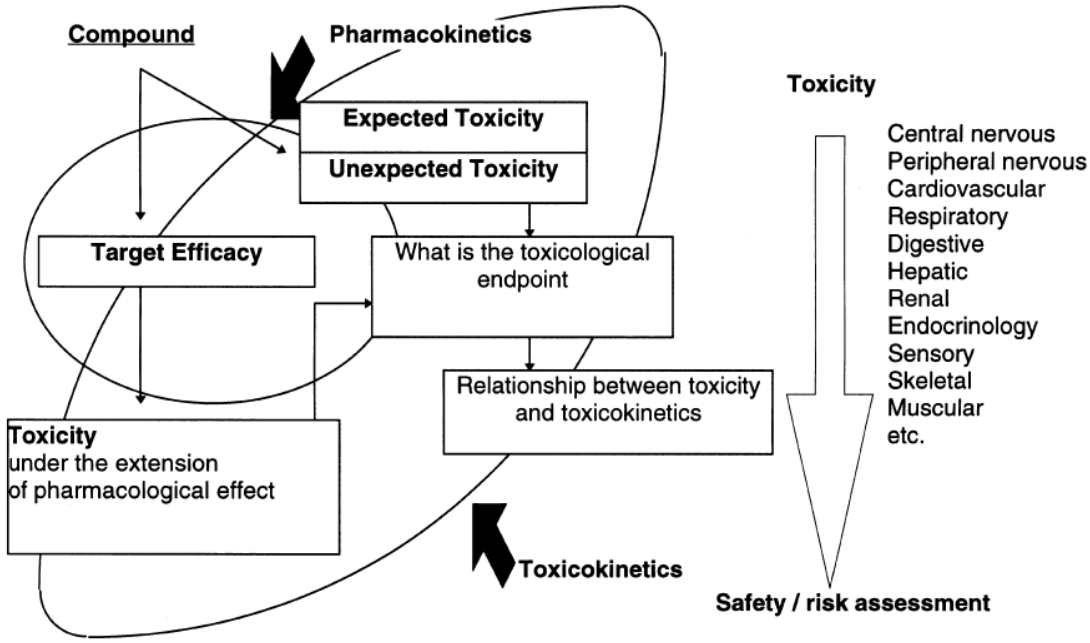


Figure 2 Safety risk assessment based on toxicokinetics. Target efficacy: target engagement endpoint on in vitro or in vivo screening. Reprinted with permission from Toxicology Letters 102-103, pages: 657-664 (1998)

One of the problems non-clinical scientists face when considering the implementation of alternative methodologies is, however, the fact that pre-clinical toxicity studies are not designed for the assessment of concentration-effect relationships, i.e., they are aimed primarily at establishing a safety threshold (e.g., NOAEL) (13, 14). A common justification for current experimental protocol designs is often the complexity and limited understanding of the biological processes involved on one hand and the challenges to obtain regulatory acceptance of an alternative method on the other (15). This is further compounded by the shortcomings of non-compartmental data analysis methods, which are currently recommended for estimating and summarising measures of exposure such area under the concentration vs. time curve (AUC) or peak concentrations (C_{MAX}). These methods cannot be easily adapted to account for nonlinearities in the time course of drug effects, nor allow for extrapolation or interpolation procedures. Such limitations pose important questions about the rationale and relevance of such experiments for the translation of findings across species and accurate inferences about the risk associated with the proposed treatment or intervention in humans.

In the current investigation we explore the feasibility of using a model-based approach to describe time-dependent pharmacokinetic-pharmacodynamic relationships and incorporate biomarkers as a proxy of drug exposure in general toxicity studies. In addition, we show how the accuracy and precision of experimental parameters compare when analysing data based on nonlinear mixed effects modelling instead of the traditional non-compartmental methods. We illustrate the concepts using simulations in which hypothetical drugs with different pharmacological properties are tested in a variety of scenarios. For the sake of simplicity, in all scenarios the biomarker is assumed to be inhibited by the active treatment. Although a myriad of pathological mechanisms may exist, our scenarios are limited to a few examples, including biophase equilibration, (re)active metabolite formation, irreversible binding and indirect response mechanisms, which can be easily expanded or generalised, enabling accurate inferences about known causes of nonlinearity and time-dependencies regarding the onset, maintenance and waning of unwanted drug effects.

Methods

A model-based approach was used to generate drug exposure and safety biomarker data for five hypothetical drugs. Experimental protocols were defined according to current guidelines for the evaluation of toxicity and safety pharmacology with the exception of additional safety biomarker data collected in parallel to the scheduled pharmacokinetic sampling points. All simulations and fitting procedures described below were performed in NONMEM 7.1 (ICON Development Solutions. Hanover, Maryland) (16). Data manipulation and statistical and graphical summaries were performed in R 3.0.0 (17).

For the purposes of our investigation, pharmacokinetic and pharmacodynamic parameters were considered to accurately reflect the risk of adverse events and toxicity. Whilst all drugs were assumed to have the same pharmacokinetic characteristics, different scenarios were used to explore five toxicodynamic mechanisms leading to biomarker inhibition. No covariate relationships were included in any of the models to facilitate the interpretation and comparison of the results. Data was subsequently analysed using standard non-compartmental methods and by nonlinear mixed effects modelling. The estimates obtained from these virtual experiments were then compared to the true values used initially to allow the assessment of bias and precision. Methods regarding the simulation and reanalysis of the PK data can be found in Chapter 3 of this thesis, in which the feasibility of PK modelling in general toxicity study is evaluated (18).

Experimental design: The protocol design used for each of the hypothetical drugs was based on an initial set-up commonly used for chronic toxicity evaluation. Four treatment groups (N= 8 per group) receiving oral daily doses of vehicle, 10, 30, and 100 mg/kg/day were tested throughout this set of virtual experiments, which lasted either one week, one month or three months. Satellite groups with 24 animals each were used to mimic the dosing conditions in the animals used for the assessment of toxicity. This procedure ensures the

availability of more frequent blood samples for toxicokinetics. The sampling schedule investigated was composite sampling where 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 hours after dosing. It was assumed that sufficient blood could be collected for plasma drug concentration and biomarker measurements. The allocation of animals to each sampling time point was random within the constraint that all animals must be sampled an equal number of times. An overview of the experimental conditions is summarised in Table 1.

Table 1: Experimental design of treatment and satellite groups in a general toxicity study with serial and composite sampling of blood for the evaluation of drug concentrations and biomarker levels in plasma.

Duration	Sampling approach	No. of animals per dose group	Sampling scheme	Sampling time
1 week	*Composite:*	8	3 per animal. 3 per time point	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose
	Serial:	3	Serial profiles from Day 1 only	
1 month	*Composite:*	8	3 per animal. 3 per time point	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose
	Serial:	3	Serial profiles from Day 1 and 12	
3 month	*Composite:*	8	3 per animal. 3 per time point	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose
	Serial:	3	Serial profiles from Day 1, Wk 4, 12.	

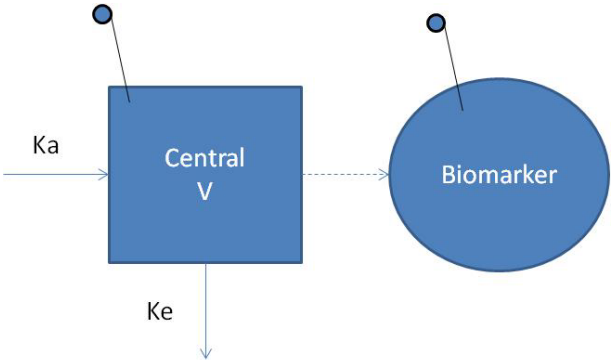
Pharmacokinetics: The pharmacokinetic model for all scenarios was a one-compartment pharmacokinetic model with first order absorption and first order elimination. This corresponded to Model A in Chapter 3 of this thesis (18). Parameter values for each scenario shown in table 2. Residual variability for pharmacokinetic data was assumed to be 15%.

Table 2: Pharmacokinetic model used to simulate concentrations and derive measure of drug exposure in the experimental groups. For the sake of simplicity, a one-compartment model (1 CMT) was selected for the purpose of this analysis. Parameters reflect data previously reported in Chapter 3 of this thesis (18).

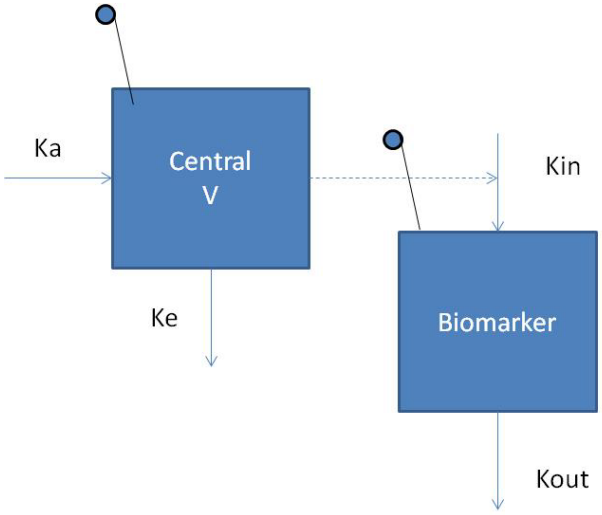
Parameter	Pop Estimate	BSV
KA	13.46 h^{-1}	50%
V	49.4 ml/kg	16%
CL	2.72 ml/hr	20%

Pharmacodynamic effects: Five hypothetical mechanisms of drug-induced toxicity were simulated. Their parameterisation is summarised in Table 3. In brief, a number of scenarios were included, which are representative of onset and dynamics of the effect, i.e., that take into account the time dependencies and delays between the start of treatment, the onset, maintenance and waning of the pharmacodynamic effects: 1) a direct I_{MAX} model, describing immediate onset of effect and direct relationship between drug exposure and biomarker inhibition at the target site; 2) an indirect response model, describing the presence of turnover mechanisms with a delayed onset of effect and disconnect between drug exposure and biomarker inhibition; 3) indirect response model preceded by biophase equilibration processes, which emphasise the role of tissue kinetics for the characterisation of pharmacodynamic effects; 4) a model describing the cumulative effects of chronic dosing regimen associated with slow-offset and irreversible binding and 5) a model describing delays due to metabolite formation with immediate inhibitory effects on biomarker levels. All scenarios were evaluated under the assumption that assay error was small in relation to the magnitude of the drug effects on biomarker levels.

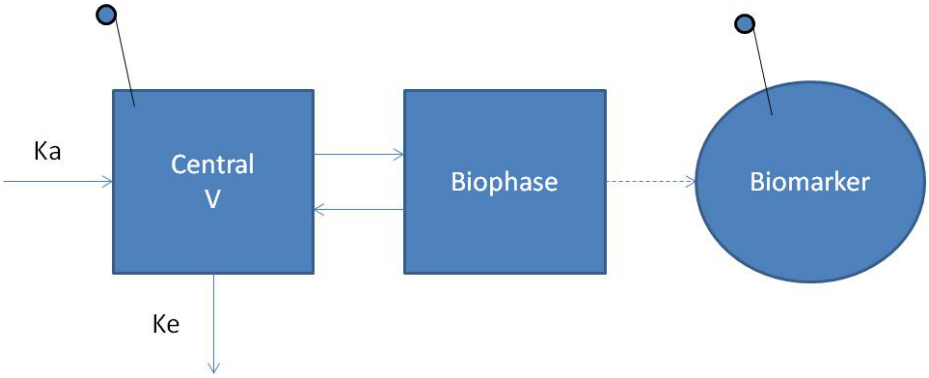
Imax model



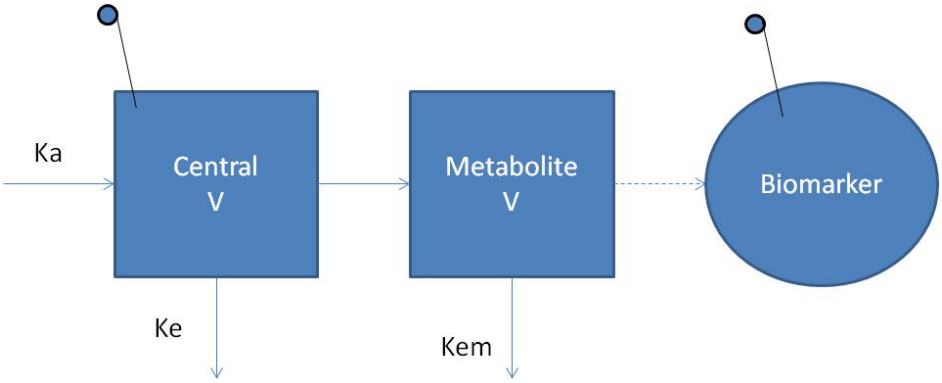
Indirect model



Biophase equilibration + Imax



Metabolite formation + Imax



Irreversible binding

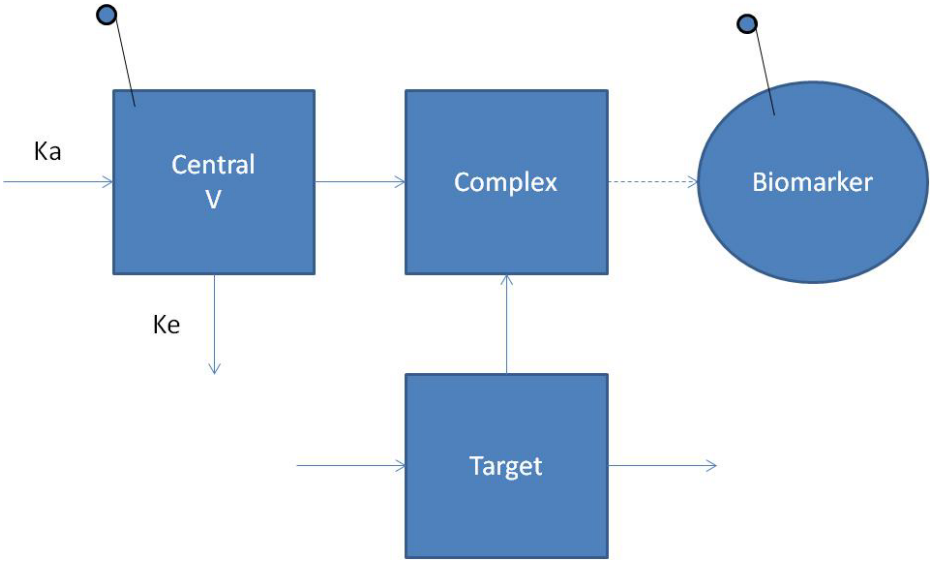


Figure 3 Diagrams depict the different pharmacokinetic-pharmacodynamic models associated with the hypothesised pharmacological mechanisms leading to toxicity.

Table 3: Simulation scenarios modelled a range of pharmacological mechanisms. Base: baseline effect, IC_{50}: concentration required for 50% inhibition of biomarker response, Kout: first order elimination of biomarker response, Kelm: first order elimination rate of metabolite, Kon: receptor-ligand association rate: Imax: maximum inhibitory effect; CAOC: cumulative area above the effect (biomarker) vs. time curve.

Simulation Scenario	Parameter names	Parameter values	Rationale	Modelling strategies
1) Direct effect	Base Imax IC_{50} (mg/ml)	1 1 0.1	- IC_{80} similar to C_{trough} at lowest dosing level. -Normalised biomarker concentrations.	1) Direct Imax
2) Indirect response	Base Kout (h-1) IC_{50} (mg/ml) Imax	1 1 0.1 1	-Fast elimination of biomarker	1) Indirect 2) Direct Imax
3) Biophase equilibration + Imax	Imax IC_{50} (mg/ml) Ke0 (h-1)	1 0.1 0.25	- K_{e0} selected to give similar biomarker levels to indirect response model	1) Biophase + Indirect response 2) Indirect response 3) Direct Imax
4) Metabolite formation + Imax	Kelm (h-1) Imax IC_{50} (mg/ml)	0.0866 1 3	- 100% conversion to metabolite assumed. metabolite t_{hh}= 8h	1) Metabolite + Imax 2) Direct Imax
5) Irreversible binding	Kout (h-1) IC_{50} (mg/ml) Kon	0.029 0.2 0.0005	- 24 hr turnover for (off)target assumed	1) Irreversible binding 2) Cumulative AOC (CAOC) + Imax

Biomarker exposure measures: Five different measures of biomarker exposure were used for calculation of the true pharmacodynamic effects, as determined by the simulated profiles. These measures can be seen in Table 4 alongside their calculation method. AOC_{24} and C_{MIN24} are intended to mimic the AUCτ and $C_{MAX\tau}$ exposure variables typically calculated for the analysis of pharmacokinetic data. They provide a measure biomarker inhibition on the final day of measurement (i.e., at steady state conditions). On the other hand, time under threshold biomarker concentration (TUT) is a measure of time spent, on final day of measurement, under a clinically significant threshold. The threshold in this case was 0.2mg/ml, which represents 20% change from baseline and which was assumed to be physiologically meaningful. Given that at three months the scenario describing slow-offset and irreversible binding is not at steady-steady at the end of treatment, simulations were performed assuming repeat dosing up to six months (three months beyond the time frames of the investigated studies). This procedure was required to ensure comparability of the results obtained for all five mechanisms of action.

Non-compartmental analysis: Biomarker exposures were calculated on the composite profile. Two different, commonly used, averages were investigated, the arithmetic mean and geometric mean. Since the standard sampling scheme is limited to a particular day during the course of treatment, composite profiles over six months cannot be estimated. Therefore, only AOC, C_{MIN}, TUT (in Table 4) were calculated by non-compartmental analysis.

Model-based estimates: Each simulated dataset was an integration of all pharmacokinetic and pharmacodynamic data for all experimental groups. Then drug concentration and biomarker profiles were fitted to multiple PKPD models (as shown by the multiple modelling strategies in Table 3) using the FOCEI estimation method. In Table 3, modelling strategies for each scenario are ordered by decreasing numbers of parameters, starting with the true model. Model convergence for each modelling strategy was determined by standard minimisation success criteria. Below quantification limit (BQL) data were omitted to mimic experimental conditions in which imputation methods are not applied. The model-based

calculation of biomarker and exposures measures are summarised in Table 4. Estimates of exposure were all calculated by using same methods as for the true exposure calculation but with the estimated model. Overall performance of competing modelling strategies was assessed by convergence rate and bias/precision of exposure/biomarker level.

Bias/precision of exposure/biomarker levels: The process of simulation and estimation of exposure (using non-compartmental and model-based methods) was repeated 200 times. Bias and precision were assessed via the scaled relative mean error (SRME) and the coefficient of variation (%CV), respectively. Relative error was also calculated for graphical comparison.

Table 4: Measures of biomarker exposure obtained with the simulated and estimated models for calculation of the true pharmacodynamic effects

Covariate name	**Symbol**	**Model based biomarker level calculation**
Area above biomarker levels vs. time profile	AOC24	$B_{C^p}(0) - \int_{t-24}^{t} B_{C^p} dt$
Minimum biomarker level over 24 hour period	C_{MIN24}	$\min\left(\{B_{C^p}(s): t-24 < s < t\}\right)$
Time under threshold (80% inhibition)	TUT	$\int_{0}^{t} 1_{B_{C^p} < .B_{C^p}(0)} dt$
Predicted 6-month cumulative area above biomarker concentration vs. time profile	CAOC	$B_{C^p}(0) * 6months - \int_{0}^{6\ months} B_{C^p} dt$
Predicted 6-month trough biomarker levels	C_{MIN}	$\min\left(\{B_{C^p}(s): 0 < s < 6\ months\}\right)$

Individual predicted biomarker levels are denoted by $B_{C^p}(t)$.

Table 5: Convergence rate of different modelling strategies (as determined by NONMEM minimization success criteria).

Simulation Scenario	Modelling strategies	Convergence rate (%)
1) Direct effect	1) Direct Imax	94.2
2) Indirect response	1) Indirect	100
	2) Direct Imax	13.5
3) Biophase equilibration + Imax	1) Biophase + Indirect response	0
	2) Indirect response	88.5
	3) Direct Imax	100
4) Metabolite formation + Imax	1) Metabolite + Imax	0
	2) Direct Imax	100
5) Irreversible binding	1) Irreversible binding	0
	2) Cumulative AOC (CAOC) + Imax	100

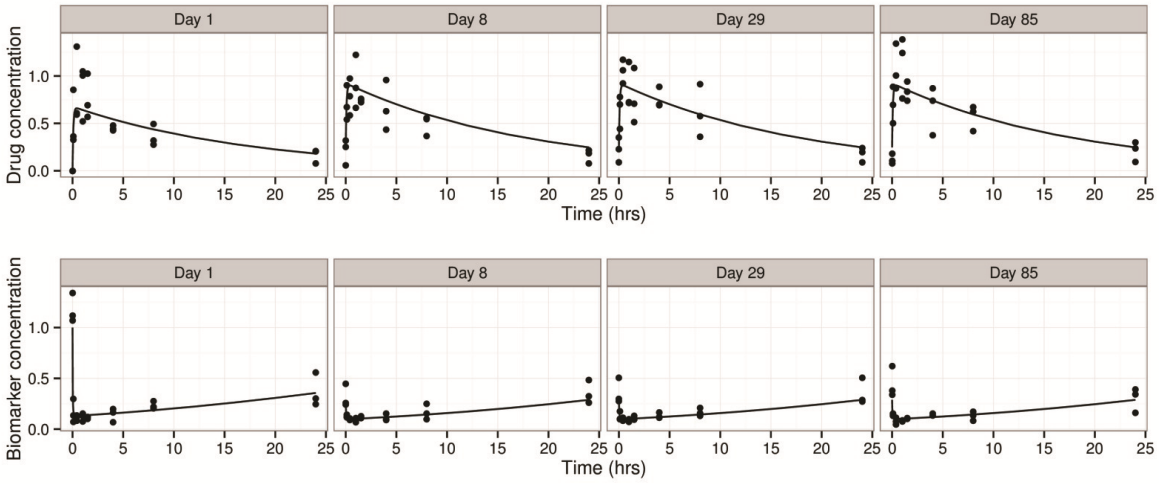
Results

As described below, the use of simulated data for the evaluation of hypothetical scenarios provided clear insight of the impact of current practices on the identification of putative mechanisms underlying the observed pharmacological effects as well as on the accuracy and precision of safety thresholds, and in particular of the NOAEL. Results from modelling are shown together with the parameter values obtained from naïve pooling and non-compartmental analysis where applicable.

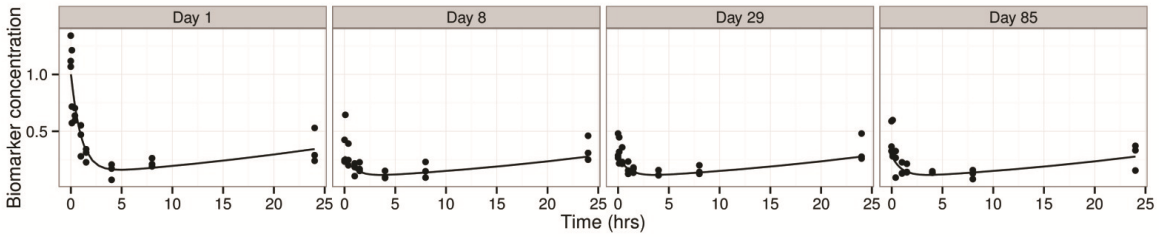
It is clear from the different profiles (Figure 4) that not only the dose level under investigation, but also the mechanism of action underlying drug toxicity, contribute to differences in the onset, magnitude and duration of the effects. Moreover, these

differences may or may not be evident depending on the dose rationale and sampling scheme used in the experimental protocol (Figure 5).

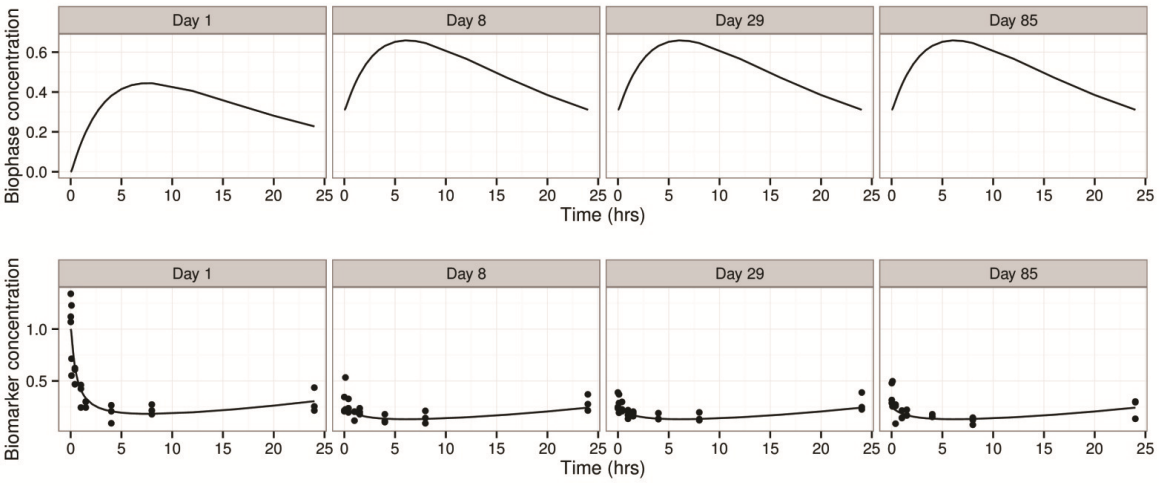
Model 1 - IMAX



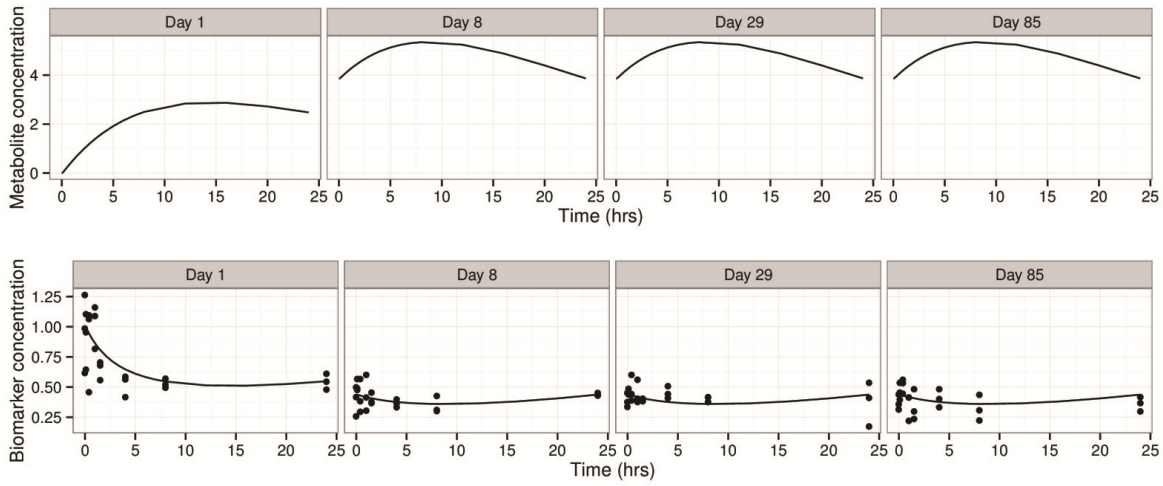
Model 2 – Indirect response



Model 3- Biophase + IMAX



Model 4: Metabolite + IMAX



Model 5: Irreversible binding

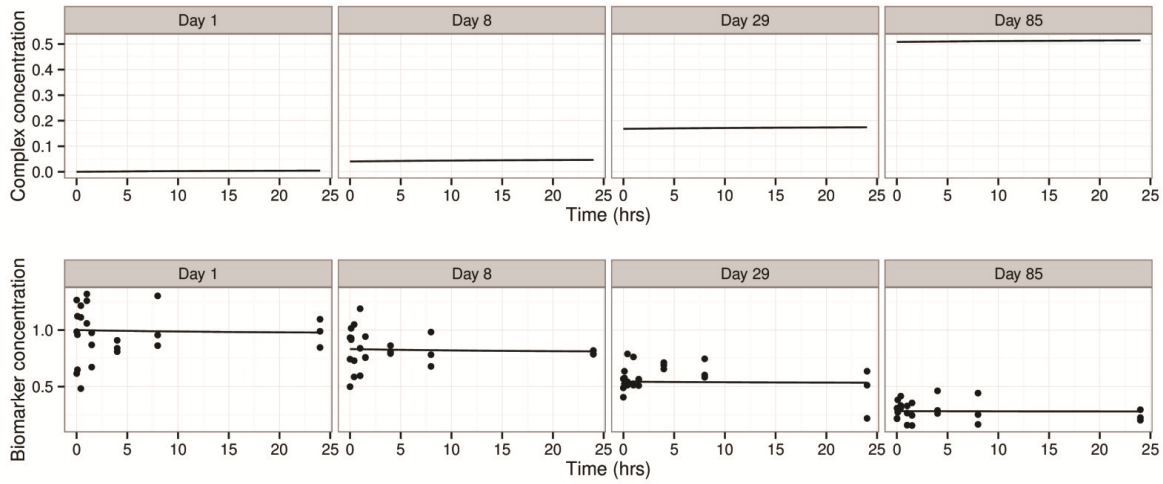


Figure 4: Full pharmacokinetic and pharmacodynamic profiles observed for each hypothetical mechanisms on selected sampling days. Lines represent the typical population estimates. Dots represent simulated concentrations at the pre-defined sampling times. Since all simulation scenarios share the same pharmacokinetics, PK is only shown for scenario 1.

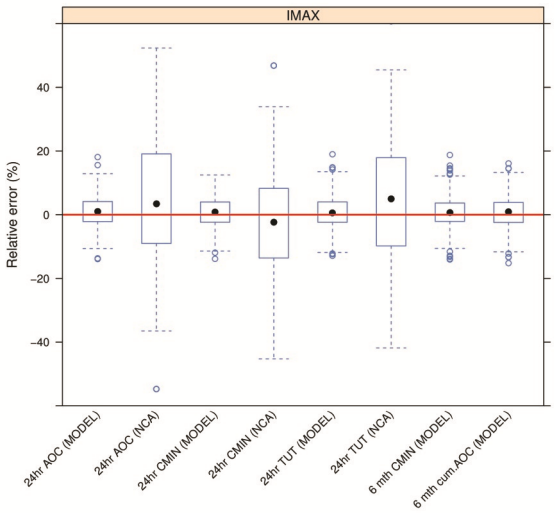
As can be observed from the summary of convergence rates in Table 5, the inability to discriminate the underlying mechanism of action based on the available experimental data can lead to obvious issues with model identifiability. On the other hand, despite this

limitation, the large differences in convergence rate suggest that model-based estimates might be suitable to explore or exclude possible or plausible causes of toxicity.

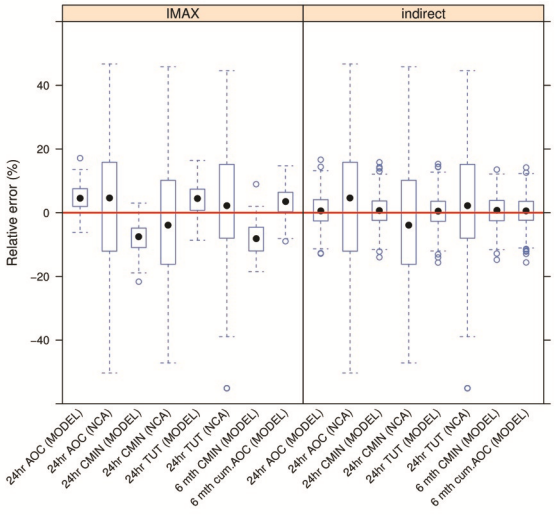
Despite variations in bias and precision between analysis methods and sampling schemes, parameter precision was relatively high (<30%). This suggests that when bias is acceptable, reductions in numbers of animals may be possible whilst still achieving study objectives. For all scenarios tested, we have assumed that the safety biomarkers levels are closely related to target engagement of targets relevant to downstream toxicity findings. In other words, the safety biomarkers are in the causal pathway between drug exposure and manifest toxicity. Therefore, for all scenarios, AOC, C_{MIN}, TOT, CAOC are expected to be more highly correlated with toxicity than their pharmacokinetic equivalents (AUC, C_{MAX}, TUT and CAUC, respectively). The relative relevance of AOC, C_{MIN}, TUT and CAOC will depend on the downstream pathway between target engagement and toxicity finding. For chronic enzyme inhibitors and receptor antagonists and/or long term toxicity, cumulative biomarker inhibition is likely to be more pharmacologically relevant. For toxicity that involves overriding homeostatic control, TUT using a physiologically relevant threshold may be most relevant.

With regard to the method of analysis, our results show that the accuracy and precision of model-based estimates for AOC, C_{MIN} and TUT were similar across different dosing groups and treatment durations. Non-compartmental estimates showed more bias and less precision in all scenarios. In addition, relative errors were also smaller for model-based estimates (Figure 5). For both model-based and compartmental methods, the coefficient of variation increased with composite designs (with 8 animals), as compared to serial sampling designs (with 3 animals). Interestingly, the use of arithmetic and geometric means for non-compartmental methods had minor impact on the parameter estimates. C_{MIN} was consistently over- estimated by non-compartmental methods.

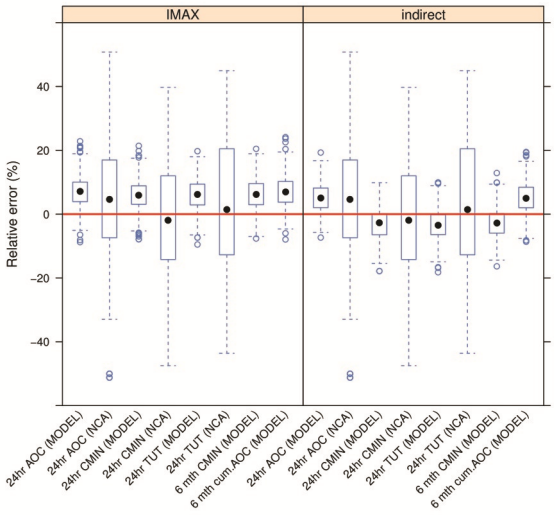
Scenario : IMAX



Scenario : indirect



Scenario : biophase+indirect



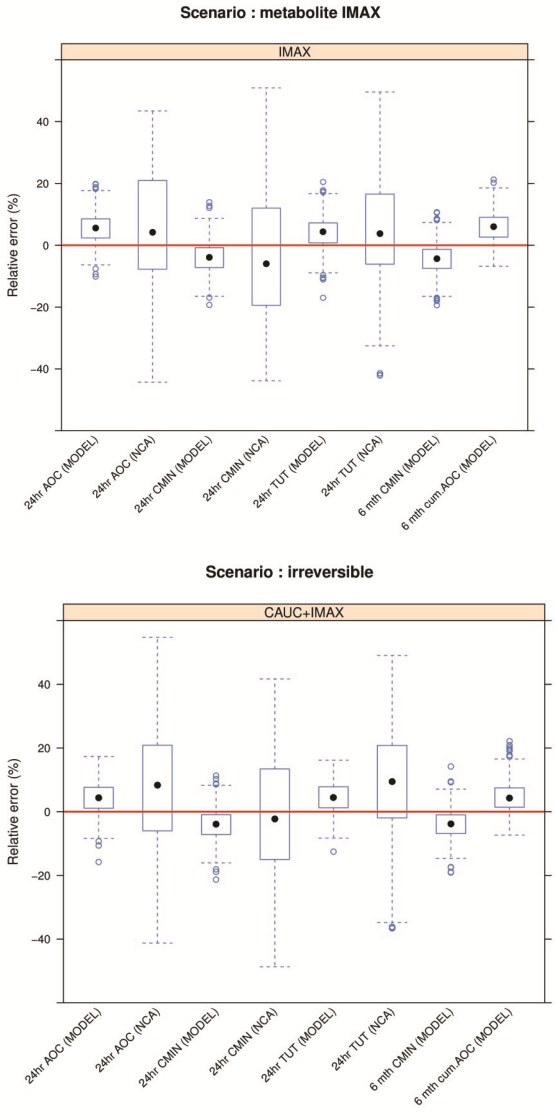


Figure 5. Relative errors of model-based and NCA estimators of exposure obtained for the different models: I_{MAX} model (a) Imax (b), indirect model (c), biophase equilibration + I_{MAX} model (d) prodrug+I_{MAX} , (e), the irreversible binding. X-axis shows the different measures of exposure, as described in Table 1. NCA estimates are repeated in each panel for comparison purposes. For the sake of clarity, only data from the 30mg/kg/day following 3-month treatment are summarised. Similar results were observed across other cohorts.

Discussion

Any drug can produce an adverse response at therapeutic or supratherapeutic exposures. It is imperative therefore to identify not only the response but also the exposure at which the effect is observed (19). Yet, over the past decade, it has also become clear that detection of organ-specific toxicity is critical, both for improved preclinical/clinical translatability and accurate prediction of toxicity at early stages of development. Despite the scientific rationale, few successful examples exist that demonstrate the development and consequent use of specific markers of organ toxicity during preclinical safety evaluation (20, 21). The limited impact of biomarkers has been associated with the fact that early prediction of specific organ function, such as hepatic, dermal or immunologic, is not well established. With the possible exception of cardiac function, very few novel biomarkers have been identified and accepted over the past decade. On the other hand, markers of tissue injury have been identified, but they are not predictive of overall organ function and often do not correlate with overt pathology (22). These conclusions have been drawn without careful consideration of the impact current experimental protocol designs and data analysis methods have on the characterisation of the underlying pharmacokinetic-pharmacodynamic relationships. Instead, here we have shown how a model-based approach can be used to integrate toxicokinetic and biomarker data for the evaluation of long-term safety and toxicity. Based on a series of hypothetical drugs, simulation scenarios have been used to show the feasibility of introducing biomarkers as a proxy of drug exposure in general toxicity studies. Furthermore, our work highlights the impact that modelling can have on the evaluation of exposure measures that cannot be derived from empirical protocols.

From a conceptual perspective, the evaluation of hypothetical compounds whose mechanisms of action reflect nonlinearity and time-dependencies in the onset, maintenance and waning of drug effects also sheds light on the shortcomings of current protocols and data analysis methods, for which data and knowledge integration have remained marginal. Current mainstream research in toxicology and safety pharmacology is performed under the

assumption that evidence from data or lack thereof is sufficient to make inferences about the risk or hazard in humans. Our approach comes from a quite different perspective, in that it incorporates oncoming data into a modelling framework, i.e., a mathematical representation of existing knowledge. Whilst scepticism exists about the predictive reliability of models due to uncertainties (23), they facilitate the assessment of causation and provide the basis for the exploring the plausibility of alternative mechanisms or causes (24). Most importantly, models when used as an ancillary tool during planning and design of experimental protocols can significantly increase the informative value and reduce bias. In fact, mechanism-based pharmacokinetic-pharmacodynamic modelling has evolved successfully as an important tool for the evaluation of exposure-response relationships and as such has represented a major contribution to the dose rationale in clinical research (25,26). In conjunction with nonlinear mixed effects techniques, it has become possible to integrate efficacy and safety measures under the assumption that wanted and unwanted pharmacological activity is directly or indirectly associated with drug action on primary or secondary targets, rather than treating such effects by default as the result of an unknown off-target binding site, which is often assumed to be the cause of toxicity (27,28). It should be noted, however, that thanks to the use of model parameterisation describing (patho)physiological phenomena in terms of zero, first and second order processes, it is possible to establish correlations between drug exposure, biomarkers and effects even if the underlying mechanisms are not fully understood.

As indicated by the differences in convergence rate (table 5), our findings reveal that even with the incorporation of biomarkers, it may be sometimes impossible to identify the true model and consequently, characterise the true mechanism of toxicity. Yet, despite model identifiability issues, these results also show the potential benefits of model parameters to rank compounds (e.g., by differences in potency) and quantify the effects associated with a given exposure or effect. Moreover, the various scenarios can be used to elucidate how differences in mechanism of action may lead to biased estimation of the relationship between drug exposure and toxicity, as well as to inaccurate safety thresholds. We believe, therefore, that greater awareness is required about the limitations of current experimental

protocols, particularly in a period in which long-term safety have become a major clinical and regulatory concern (29-31). On the other hand, model misspecification, even when convergence is successful, may lead to significant bias when predictions are made beyond the experimental context (i.e., longer timescale or different dosing regimens). To mitigate such effects we recommend careful consideration of model selection during model development and model uncertainty (32-34). Model selection criteria should be guided not only by ability to describe data but also by assessing the physiological relevance of model assumptions. When model development ends in multiple competing models performing similarly with respect to the above model selection criteria, clear reporting of such model uncertainty is necessary. In any case, model averaging should be discouraged when predictions arising from different model differ significantly (35). Finally, parameter uncertainty should be incorporated when performing simulations or using the model to make predictions.

Since the model-based methods outperformed non-compartmental analysis, further refinement of experimental protocols can be achieved if the data are analysed using nonlinear mixed effects modelling. Despite the conceptual challenges, maximum likelihood based model estimates, statistically speaking, are asymptotically efficient. This means that model parameters extract maximum information from the dataset when compared to any other statistical technique. In this respect, the use of a model-based approach is not only an improvement on non-compartmental methods in terms bias and precision, but is optimal for the datasets under consideration.

Limitations

All scenarios depicted here corresponded to the case where substantial inhibition of a biomarker, relative to assay and normal physiological variability, was correlated with toxicity. In the present study we focused on biological systems with built in control for minor fluctuations in biomarker levels so that substantial inhibition could be seen in the data. An example of such a biomarker and system would be prostaglandins, which exert their protective role in conjunction with other mediators (36-39). The findings presented

here are expected to also reflect other mechanisms such as induction or tolerance, if the profiles and magnitude of effect are similarly large relative to the assay error.

We also acknowledge that detection of signals above variability may not always be possible due to a variety of factors, including large between or within subject variability or poor assay precision. In this case non-compartmental analysis and model-based approaches will fail to detect systematic variations without additional information.

Lastly, we believe that complex systems pose little problem to estimation procedures as long as viable simplifications are available and able to describe important trends in the data. Clearly, there may be instances where the biological response to drug exposure manifest in trends in data which cannot be described accurately by more simplistic models. In these circumstances, knowledge that an experiment cannot be used to describe the underlying exposure-effect relationships could be invaluable as the basis for further improvement of experimental design in subsequent phases of drug development.

In summary, toxicology need to evolve from a discipline largely devoted to routine performance and interpretation of safety tests, to a quantitative discipline in which advances in pharmacology and molecular biology can be applied in an integrated manner, enabling better understanding the nature and mechanism of adverse effects caused by chemicals. Model-based analysis of biomarkers and toxicokinetic data provides the basis for differentiating settled toxicological knowledge of risk from mere possibility, and facilitating the translation of safety thresholds and safe exposure from animals to humans.

Reference List

1. DeGeorge JJ. Food and Drug Administration viewpoints on toxicokinetics: the view from review. Toxicol Pathol. 1995; 23(2): 220-5.

2. Bessems JG, Geraets L. Proper knowledge on toxicokinetics improves human hazard testing and subsequent health risk characterisation. A case study approach. Regul Toxicol Pharmacol. 2013; 67(3):325-34

3. Campion S, Aubrecht J, Boekelheide K, Brewster DW, Vaidya VS, Anderson L, Burt D, Dere E, Hwang K, Pacheco S, Saikumar J, Schomaker S, Sigman M, Goodsaid F. The current status of biomarkers for predicting toxicity. Expert Opin Drug Metab Toxicol. 2013; 9(11):1391-408.

4. Li AP. Biomarkers and human hepatocytes. Biomark Med. 2014; 8(2):173-83.

5. Sasseville VG, Mansfield KG, Brees DJ. Safety biomarkers in preclinical development: translational potential. Vet Pathol. 2014; 51(1):281-91.

6. Csajka C, Verotta D. Pharmacokinetic–Pharmacodynamic Modelling: History and Perspectives. J Pharmacokin Pharmacodyn 2006; 33(3): 227-279.

7. Leighton JK. Application of emerging technologies in toxicology and safety assessment: regulatory perspectives. Int J Toxicol. 2005; 24(3):153-5.

8. Green AR, King MV, Shortall SE, Fone KCF. Lost in translation: preclinical studies on MDMA provide information on mechanisms of action, but do not allow accurate prediction of adverse events in humans. British Journal of Pharmacology

9. Coruzzi G, Menozzi A, Dobrilla G. Novel Non-Steroidal Anti-Inflammatory Drugs: what we have learned from animal studies. Current Drug Targets - Inflammation & Allergy, 2004; 3: 43-61.

10. Abernethy DR, Altman RB, Brouwer KLR, Califano A, D'Argenio DZ, Iyengar R, Jusko WJ, Lalonde R, Lauffenburger DA, Shoichet B, Stevens JL, Subramaniam S, van der Graaf P, Vicini P. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. An NIH white paper by the QSP workshop group – October, 2011.

11. Bai JP, Fontana RJ, Price ND, Sangar V. Systems pharmacology modeling: an approach to improving drug safety. Biopharm Drug Dispos. 2014; 35(1):1-14.

12. Bai JP, Abernethy DR. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu Rev Pharmacol Toxicol. 2013; 53:451-73.

13. Slob W. Thresholds in toxicology and risk assessment. International Journal of Toxicology 1999; 18:345-67.

14. Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 2005; 42(3):265-74.

15. Lilienblum W, Dekant W, Foth H, Gebel T, Hengstler JG, Kahl R, Kramer PJ, Schweinfurth H, Wollin KM. Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH). Arch Toxicol. 2008; 82(4):211-36.

16. Beal S, Sheiner LB, Boeckmann A, Bauer RJ. NONMEM User's Guides. (1989-2009), Icon Development Solutions, Ellicott City, MD, USA, 2009.

17. R Core Development Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

18. Sahota T, Danhof M, Della Pasqua O. The impact of composite AUC estimates on the prediction of systemic exposure in toxicology experiments. In: Pharmacology-based

assessment of adverse drug effects. A quantitative approach to the prediction of risk in humans. PhD Thesis University of Leiden, 2014.

19. Horii I. Advantages of toxicokinetics in new drug development. Toxicol Lett 1998; 102-103:657-64.

20. Sistare FD, DeGeorge JJ. Preclinical predictors of clinical safety: opportunities for improvement. Clin Pharmacol Ther. 2007; 82(2):210–214

21. Amur S, Frueh FW, Lesko LJ, Huang SM. Integration and use of biomarkers in drug development, regulation and clinical practice: a US regulatory perspective. Biomarkers Med. 2008; 2(3):305–311.

22. Campion S, Aubrecht J, Boekelheide K, Brewster DW, Vaidya VS, Anderson L, Burt D, Dere E, Hwang K, Pacheco S, Saikumar J, Schomaker S, Sigman M, Goodsaid F. The current status of biomarke#rs for predicting toxicity. Expert Opin Drug Metab Toxicol. 2013; 9(11):1391-408

23. Oreskes N. Evaluation (not validation) of quantitative models. Environ Health Perspect. 1998; 106 Suppl 6:1453-60.

24. Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005; 24(4):161-201.

25. Bender BC, Schindler E, Friberg LE. Population pharmacokinetic pharmacodynamic modelling in oncology: a tool for predicting clinical response. Br J Clin Pharmacol. 2013; doi: 10.1111/bcp.12258.

26. Chain AS, Dieleman JP, van Noord C, Hofman A, Stricker BH, Danhof M, Sturkenboom MC, Della Pasqua O. Not-in-trial simulation I: Bridging cardiovascular risk from clinical trials to real-life conditions. Br J Clin Pharmacol. 2013; 76(6):964-72.

27. Bumbaca D, Wong A, Drake E, Reyes AE 2nd, Lin BC, Stephan JP, Desnoyers L, Shen BQ, Dennis MS. Highly specific off-target binding identified and eliminated during the humanization of an antibody against FGF receptor 4. MAbs. 2011; 3(4):376-86.

28. Green MR, Newton MD, Fancher KM. Off-Target Effects of BCR-ABL and JAK2 Inhibitors. Am J Clin Oncol. 2013 [Epub ahead of print]

29. Stevens JL. Future of toxicology--mechanisms of toxicity and drug safety: where do we go from here? Chem Res Toxicol. 2006; 19(11):1393-401.

30. Houck KA, Kavlock RJ. Understanding mechanisms of toxicity: insights from drug discovery research. Toxicol Appl Pharmacol. 2008; 227(2):163-78.

31. Mielke H, Gundert-Remy U. Physiologically based toxicokinetic modelling as a tool to support risk assessment: three case studies. J Toxicol. 2012; 2012:359471.

32. Nedelman JR, Gibiansky E, Lau DT. Applying Bailer's method for AUC confidence intervals to sparse sampling. Pharm Res 1995; 12(1):124-8.

33. Gagnon RC, Peterson JJ. Estimation of confidence intervals for area under the curve from destructively obtained pharmacokinetic data. J Pharmacokinet Biopharm 1998; 26(1):87-102.

34. Bonate PL. Coverage and precision of confidence intervals for area under the curve using parametric and non-parametric methods in a toxicokinetic experimental design. Pharm Res 1998; 15(3):405-10.

35. Lunn DJ. Automated covariate selection and Bayesian model averaging in population PK/PD models. J Pharmacokinet Pharmacodyn 2008; 35(1):85-100.

36. Kargman S, Charleson S, Cartwright M, Frank J, Riendeau D, Mancini J et al. Characterization of Prostaglandin G/H Synthase 1 and 2 in rat, dog, monkey, and human gastrointestinal tracts. Gastroenterology 1996; 111(2):445-54.

37. Fiorucci S, Distrutti E, Cirino G, Wallace JL. The emerging roles of hydrogen sulfide in the gastrointestinal tract and liver. Gastroenterology 2006; 131(1):259-71.

38. Wallace JL, Miller MJ. Nitric oxide in mucosal defense: a little goes a long way. Gastroenterology 2000; 119(2):512-20.

39. Wallace JL. Prostaglandins, NSAIDs, and gastric mucosal protection: why doesn't the stomach digest itself? Physiol Rev 2008; 88(4):1547-65.

CHAPTER 6

Utility of model based approaches to predict the risk of adverse events from preclinical toxicology protocols.

Tarjinder Sahota, Meindert Danhof and Oscar Della Pasqua

Abstract

Purpose: Current toxicity protocols use the NOAEL approach to relate observed systemic exposure to the observed AEs. However, biomarker data can provide information on mechanisms of toxicity and historical placebo data can help distinguish non-drug induced AEs to ADRs. The objective here is to determine the feasibility of model-based risk assessment with the aforementioned data and to compare this with the NOAEL approach.

Methods: An in-silico approach based on simulation scenarios and nonlinear mixed effects models was used to generate drug-induced and background adverse events. The test species was rats and data was generated according to standard preclinical toxicological designs. A total of six scenarios were simulated, in which reversible and irreversible drug effects were evaluated under the assumption of three different pharmacological mechanisms (direct, indirect, and irreversible binding). Data was then analysed using standard NOAEL approach and by nonlinear mixed effects modelling in NONMEM 7.1 and WinBUGS 1.4.3.

Results: Three out of six scenarios had a viable therapeutic window. The NOAEL approach showed significant bias by overestimation of toxicity. The potential impact of bias to drug development programs is summarised for each scenario. Model-based approaches showed high convergence rates, however model identifiability prevented model discrimination indicating that although risk can be predicted the underlying causes of risk cannot be determined.

Conclusions: Our results indicate that standard toxicology experiments are likely to provide enough information to detect drug related ADRs with a model based approach, but are unlikely to have the power to precisely indentify the mechanisms of AE formation for rare events. Quantifying model uncertainty enables this uncertainty to be reported to aid project teams in future study planning. A model-based approach outperforms the NOAEL methodology in terms bias and precision and should therefore be recommended as method of choice for the purposes of safety assessment.

Introduction

One of the main purposes of safety pharmacology and toxicology screening is the prediction of risk that exposure to a new chemical or biological entity represents to humans. A major challenge in this endeavour is the prediction of the safe exposure in humans based on preclinical experiments. Historically, numerous approaches have been considered for the assessment of safety and risk, which differ in their data requirements, degree of complexity, their applicability in different situations and the type and quality of resulting risk estimates (100). Among the accepted methods, safety thresholds have been derived under the assumption that there is a level of exposure below which a biologically significant effect is unlikely to occur, i.e., no-observed adverse- effect level (NOAEL) (2,3). Even though estimation of such a threshold has little or no mechanistic basis and is greatly influenced by experimental design factors, it has become one of preferred methods for regulatory evaluation of risk. This choice has been made irrespective of the frequency of the events of interest or whether the occurrence of events is delayed relative to the duration of treatment. In these circumstances the evidence generated from small experiments may be affected by censoring or other shortcomings in the experimental design (4,5). As a consequence, derived measures of safe exposure may become biased and imprecise.

By contrast, a model-based analysis rooted in statistical inference and mechanistic description of physiological processes can have several advantages over safety thresholds, but its uptake has been very limited (6). Another important point to consider is that the application of pharmacokinetic-pharmacodynamic modelling and simulation concepts enables one to explore the relationship between drug exposure and pharmacological or toxicological effects in a mechanistic manner, relating experimental findings to target engagement (See Figure 1) (7). By using pharmacokinetic models, factors that are known or expected to influence the relationship between the administered dose and the target exposure may be accounted for (8). Pharmacokinetic models may also be used to optimize protocol design and strengthen the extraction of information from experimental results by

linking data obtained under different experimental conditions in a uniform model (9,10). Thus, modelling is often hypothesis generating and may have utility for discriminating between markers of exposure and markers of risk (Figure 2). Thereby, some of the uncertainty factors associated with the true hazard or risk may be reduced. Furthermore, specific questions on mode of action may be addressed, and these models can provide a stronger basis for extrapolation across species, routes of exposure, dosing patterns, and ultimately human risk assessment.

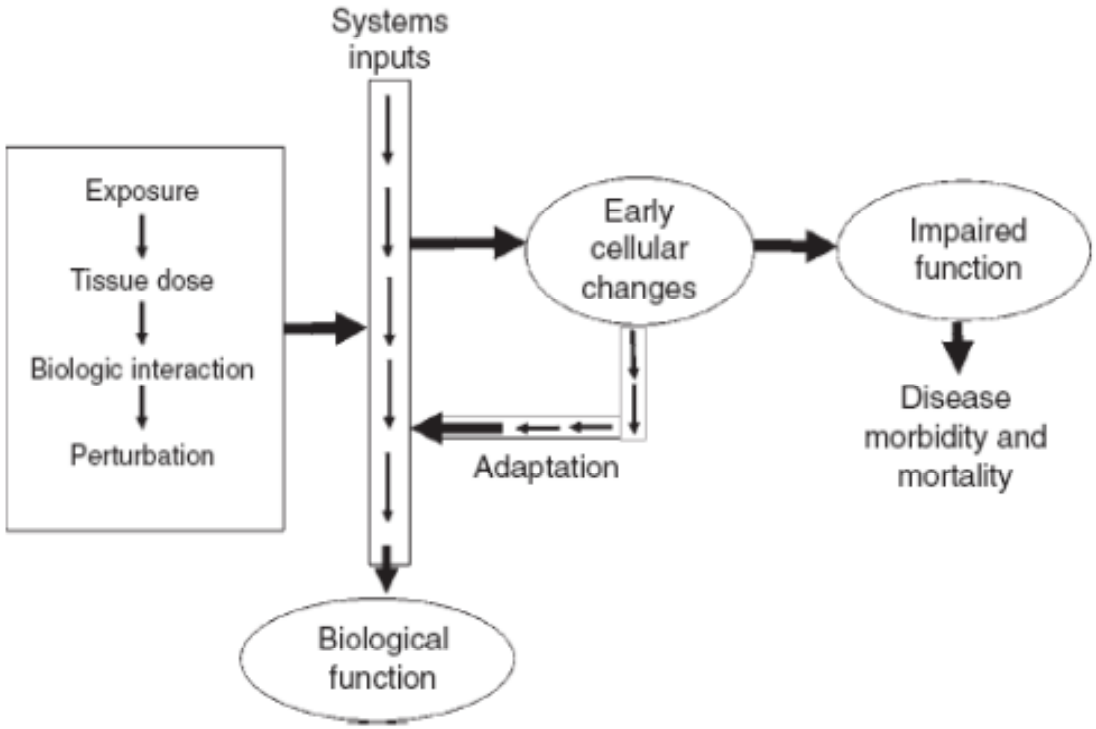


Figure 1 The diagram illustrates different steps that lead to disruptions of biologic pathways: ''biologic responses are results of an intersection of exposure and biologic function. The intersection results in perturbation of biologic pathways. When perturbations are sufficiently large or when the host is unable to adapt because of underlying nutritional, genetic, disease, or life-state status, biological function is compromised; this leads to toxicity and disease''. A model-based approach can be used to parameterised both pharmacokinetic and pharmacodynamic processes. Of particular interest is the evaluation of the outcome from function impairment when incidence of events is low or processes rate are such that the events are delayed relative to the period of intervention. (From Anderson et al.,2005 (7))

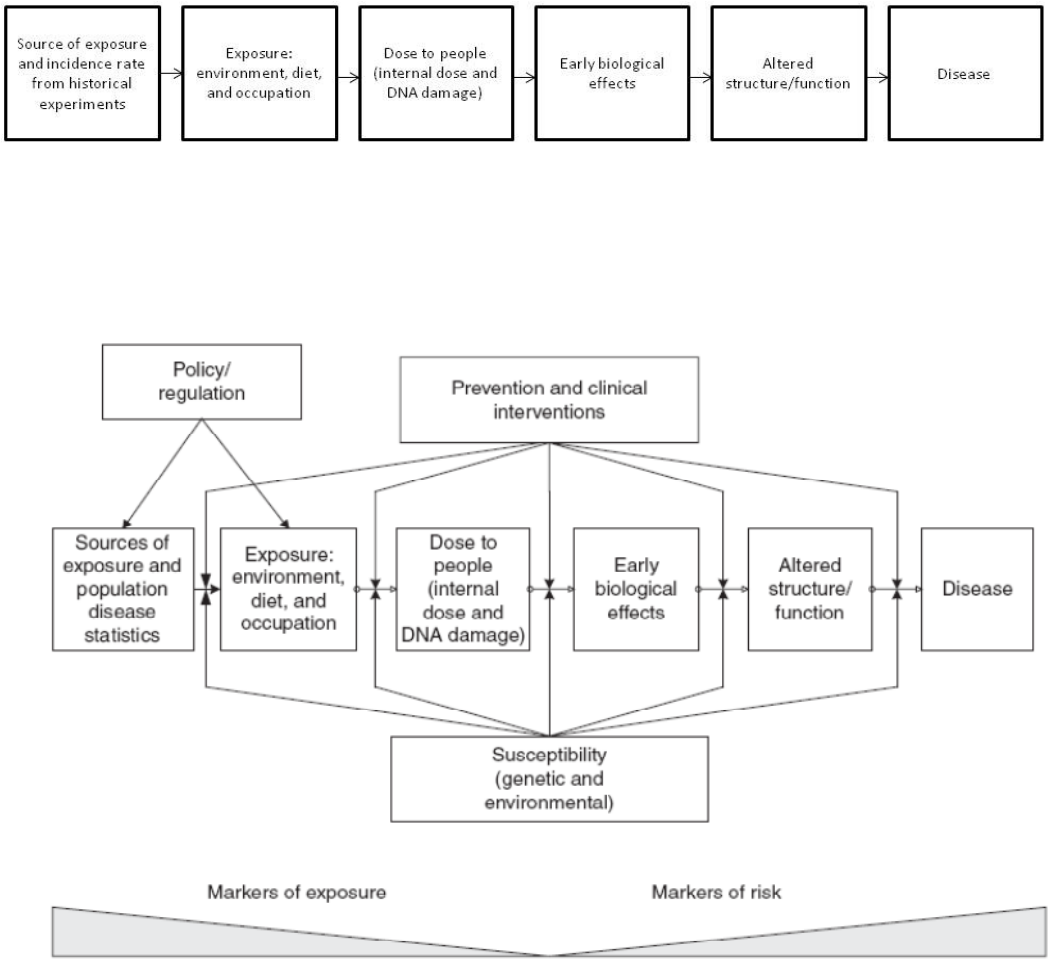


Figure 2 Inferences from risk of toxicity or disease from drug exposure data. Different approaches can be considered in which markers of exposure are used in conjunction or independently of markers of risk to predict safe exposure in humans. This diagram clearly indicates the need to discriminate drug reactions from adverse events during drug screening and early characterisation of the safety profile of a new chemical or biological entity.

Despite the aforementioned advantages, regulatory agencies still tend to favour the view that risk assessment should remain qualitative until important issues, primarily those related to quantitative decision-making concepts, have been addressed (101). From a scientific and clinical perspective, the main concern, however, is the potential for overconfidence in the numerical answers obtained from small experiments. At the same time, it must be acknowledged that characterising exposure-response relationships does face technical challenges when data is too uninformative (12). One of the issues is model and parameter identifiability, which make the validity and reproducibility of models derived by empirical experimentation questionable for predictive purposes. Another important point to consider is that toxicology studies are designed to show evidence for safety not for risk.

To address the issue of uncertainty and data sparseness arising from safety pharmacology and toxicology screening, here we illustrate how population pharmacokinetic-pharmacodynamic modelling can be implemented to characterise the relationship between drug exposure and the risk of adverse events. The ultimate goal of our investigation is to demonstrate how limitations in the informative value of experimental data can be overcome by integrating non-linear mixed effects modelling with MCMC sampling algorithms. An important advantage is that by simulating from uncertainty, one can eliminate the need for empirical safety factors when scaling up findings from animals to humans. In addition, relevant biomarker data can be integrated into the analysis either as priors or as historical baseline data, allowing incorporation of pharmacodynamic processes and other covariates in the overall estimates of drug-induced risk (13-15). Focus is given to the evaluation of events with low incidence under the assumption of different mechanisms of action for the observed events. For the sake of completeness, model-based estimates are subsequently compared with the results based on the traditional NOAEL approach.

Methods

Overview

A model-based approach was used to generate three month toxicokinetic data for a variety of hypothetical drugs. The experimental protocols for simulation were defined according to current guidelines for the evaluation of toxicity and safety pharmacology with the one exception that data characterising a safety biomarker be collected at the scheduled pharmacokinetic sampling points. Information on the occurrence of adverse events was assumed to be limited to terminal observations upon post-mortem examination. Each simulation scenario, detailed below, was intended to detect the NOAEL in an unbiased manner, i.e. when the computed NOAEL was most likely to be associated with the treatment group receiving the lowest dose level being test0ed. In this respect, the experimental design and the selection of the dose levels were such that further analysis could be performed using standard methods, i.e., the NOAEL approach.

The proposed scenarios have taken into account conditions in which adverse events are rare or have very low incidence. In this respect, this also represents a challenge for model-based analysis techniques due to low information content of the datasets. Two different PKPD models were used for the simulation of pharmacokinetic and biomarker data. On the other hand, a variety of models were considered for the simulation of adverse events, based on different incidence rates and mechanisms for the onset and cessation of adverse events. Data was then analysed using standard NOAEL approach and by a variety of model-based analysis techniques. Results were assessed for accuracy, precision and suitability for informed decision making. A schematic showing the general workflow is shown in Figure 3.

Data

Toxicology Protocols: The experimental design of the general toxicity studies was chosen to mimic existing practices including treatment and satellite groups, as shown in Table 1. Different treatment durations were evaluated, in which four dose levels are considered

(vehicle, 3, 10, and 30 mg/kg/day). The experimental protocol was based on the assumption that all animals are dosed daily via oral administration. Animals in the treatment groups are sacrificed at the end of the experiment, after the last sampling time. Satellite animals receive the same dose levels used in the main experimental groups. AE information from these groups is not used due to the potential confounding effect of frequent blood sampling. In contrast to the traditional sampling schemes for pharmacokinetics, biomarkers of pharmacology are also sampled at pre-defined 0.1, 0.4, 1,1.5, 4, 8, 24 hours after dosing) and random time points so that an equal number samples are taken at each time point.

Table 1: Experimental design of general toxicity study

Duration	Sampling approach	No. of animals per dose group	Sampling scheme	Sampling time
1 week	*Treatment:* *Satellite:*	4 3	Composite with 2 animals per time point Serial profiles from Day 1 only	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose
1 month	*Treatment:* *Satellite:*	10 3	Composite with 2 animals per time point Serial profiles from Day 1 and 28	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose
3 month	*Treatment:* *Satellite:*	12 3	Composite with 2 animals per time point on Week 4 and Week 13. Serial profiles from Day 1, Week 4 and Week 13.	0.1, 0.4, 1, 1.5, 4, 8, 24 hours after dose

Ancillary pharmacology protocols (PK): It was assumed that additional data was available from drug metabolism and pharmacokinetics studies. Typical experimental protocols were assumed which provided serial blood sampling based on eight animals per dose level, which received a for single dose oral administration (0.3, 1, and 3 mg/kg). Only drug concentrations were obtained from these animals, no biomarker concentrations.

Ancillary pharmacology protocols (placebo AE): Monitoring of placebo animals enables the assessment of non-drug induced risk. However, the quantification of rare adverse events requires a far larger database than general toxicity studies provide. Therefore, it was assumed that historical placebo datasets were also available from acute and chronic general toxicity experiments consisting of 400 animals for each type of study and treatment duration.

Simulation scenarios: All simulations and fitting procedures described below were performed in NONMEM 7.1 (ICON Development Solutions. Hanover, Maryland) (25). Data manipulation and statistical and graphical summaries were performed in R 3.0.0 (26). We assumed a population with high homogeneity and therefore no covariate relationships were included in any of the models. 200 simulated datasets were produced per scenario.

Simulation of PK data: The pharmacokinetic model for all scenarios was a one-compartment pharmacokinetic model with first order absorption and first order elimination. Parameter values for each scenario shown in Table 2. Residual variability for PK observations was assumed to be 15%.

Table 2: Pharmacokinetic model used to simulate concentrations and derive measure of drug exposure in the experimental groups. For the sake of simplicity, a one-compartment model (1 CMT) was selected for the purpose of this analysis.

Parameter	Pop Estimate	BSV
KA	13.46 h^{-1}	50%
V	49.4 ml/kg	16%
CL	2.72 ml/hr	20%

Simulation of biomarker data: Two different pharmacodynamic models were investigated which related drug concentration to safety biomarker to risk of adverse events. A schematic diagram of the two models is shown in Figure 4. The rationale was to see how both empirical and model-based methodologies performed when time dependencies exist, i.e., the onset of the adverse events is delayed with regard to the start of treatment. The first

model was an indirect response model where the biomarker was assumed to be directly related to risk of adverse events. The second was an irreversible binding model where the formation of an unmeasured biomarker was assumed to be related to risk of adverse events.

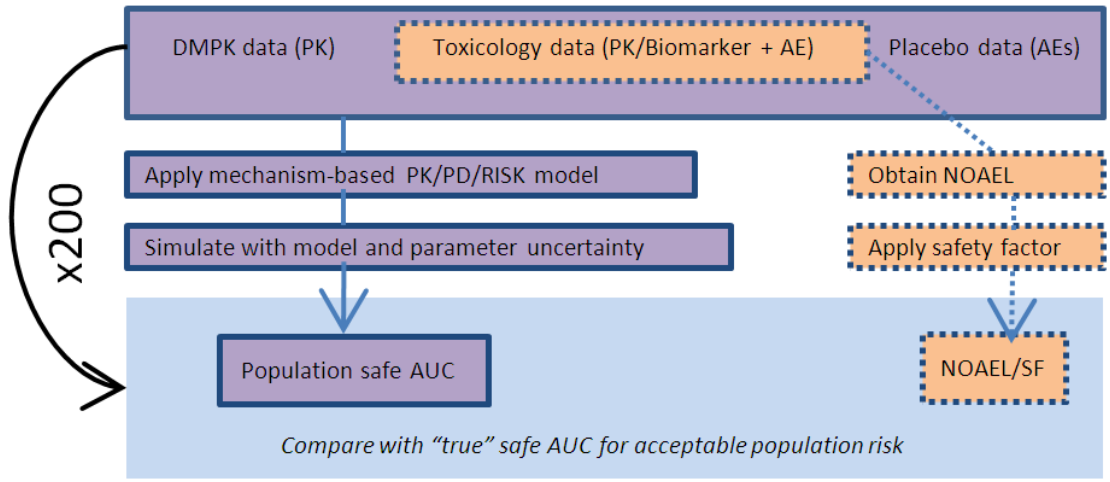
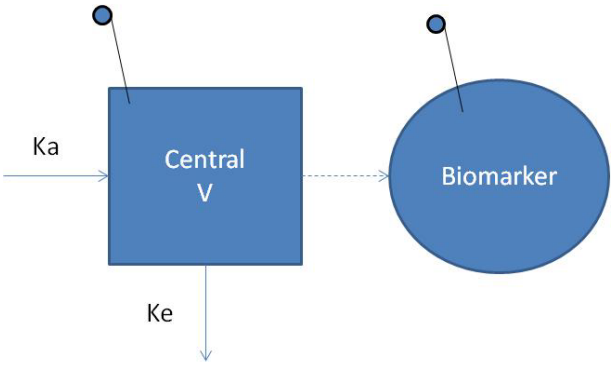


Figure 3 Schematic representation of the simulations performed for the evaluation and comparison of model-based vs. standard approach.

Simulation of adverse event data: In total, six scenarios were investigated (Table 3). All adverse events were modelled as a two state continuous time Markov process, in which a state 0 corresponds to health and a state 1 corresponds to the presence of toxicity. This assumption provided for simulated data yielding low information for subsequent model-based analyses. For both hypothetical mechanisms, multiple Markov models were used to define different types of adverse events. These adverse events were classified into reversible and irreversible. Irreversible adverse events were events for which the remission rate was set to zero. On the other hand, reversible adverse events always maintained a non-zero probably of spontaneous remission. The irreversible adverse event scenarios also happen to reflect a reversible event with left censored observation time, e.g. where

histological examination provides evidence of incidence (e.g. scarred tissue). For all adverse events, it was assumed that symptoms and signs could only be detected at the end of the study duration. This further challenges a model-based approach to analyse the data without compromising the standard methods.

Indirect



Cumulative effect (irreversible binding)

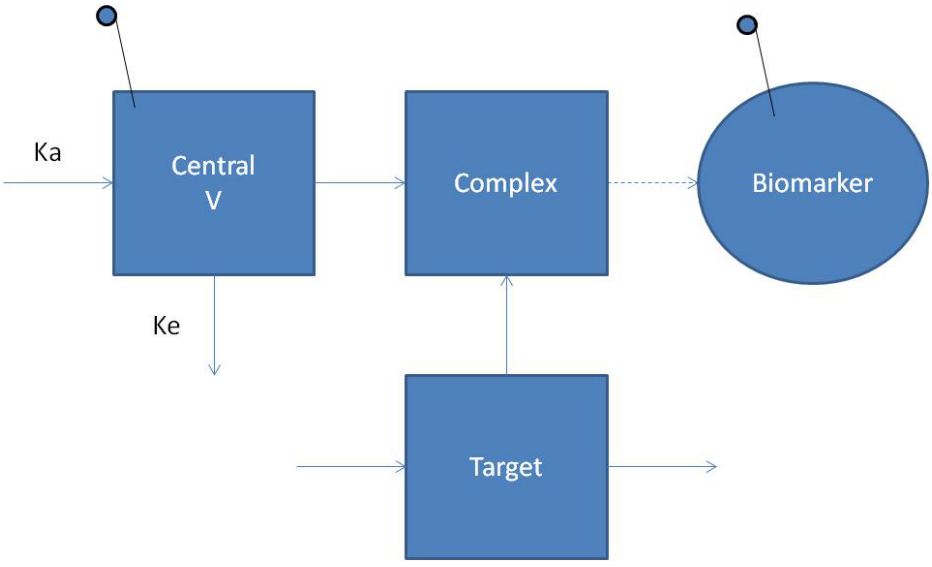


Figure 4 Diagrams depict the different pharmacokinetic-pharmacodynamic models associated with the hypothesised pharmacological mechanisms leading to toxicity.

Table 3: Summary of simulation scenarios for a range of putative mechanisms associated with reversible and irreversible adverse events. *None* corresponds to conditions in which the compound is not toxic; *Low* corresponds to scenarios in which the occurrence of false positives and false negative is most relevant. Here the probability of NOAEL being the lowest dose level in the 1 month study is maximised.

Scenario	PKPD model	AE type	Drug toxicity
n1	NA	Reversible	None
n2	NA	Irreversible	None
a1	Indirect	Reversible	Low
a2	Indirect	Irreversible	Low
b1	Irreversible	Reversible	Low
b2	Irreversible	Irreversible	Low

With regard to the assessment of the relationship between exposure and drug-induced risk, two possibilities were considered. A category "none" corresponded to drug effect having no influence on toxicity, so only non-adverse events were observed. For the purposes of subsequent data analysis, it was assumed that drug-induced and non-drug induced adverse events were indistinguishable from each other. The NOAEL analysis therefore treated all events as adverse drug reactions. The second category was "low" toxicity. Here, the drug effect parameters governing the simulation of adverse events in these scenarios were optimised so that the likelihood of the low dose of the 1 month study being the NOAEL dose was maximised. The maximisation of NOAEL likelihood was performed by the R function optimise().

For each scenario the occurrence of adverse events was described by a time-inhomogeneous Markov model for transition rates (R_{xy}):

$$\mathrm{R}_{01} = \mathrm{B}_{01}\left(1-(BIO(0)-BIO)\frac{1}{\mathrm{EB}_{50}+BIO}\right)$$

$$R_{10} = \begin{cases} B_{10} & \text{reversible} \\ 0 & \text{irreversible} \end{cases}$$

where BIO is an independent variable representing putative (bio)markers. B_{01} and B_{10} correspond to baseline transition rates. For irreversible adverse events, B_{10} was fixed to 0. For reversible adverse events it was set to 1/168, so that mean duration of an adverse event was 1 week. B_{10} was fixed to a value which corresponded to a prevalence of adverse events of 1% at three months. Transition times were simulated by sampling from their cumulative distributions (i.e. cumulative hazard function) to obtain a continue state vs. time relationship for all subjects.

Estimation steps

Estimation of NOAEL: A NOAEL was obtained for each study duration. It corresponded to the maximum daily dosing level for which no adverse event were observed. The NOAEL was expressed as the area under the concentration vs. time curve (AUC) at that dosing level, as determined by the composite method. Calculation of composite AUC values entails naïve pooling of drug concentration data on the day of sacrifice. AUCs were calculated by the trapezoidal rule using mean concentrations for each sampling time point.

Model-based risk assessment: Pharmacokinetics and biomarker concentration-effect relationships were refitted using NONMEM 7.1 with FOCEI estimation. Predicted exposure and biomarker estimates were obtained using empirical Bayes estimates (EBEs). Derived exposure and biomarker variables were then used independent variables for the characterisation of the relationship between exposure and risk of adverse events.

A logit transformation was used to describe the incidence of AEs. The general equation describing the incidence of ulcers is given by:

$$P(U_{i,j} = 1) = \frac{EXP(\theta_1 + \theta_2 * COV_{i,j})}{1 + EXP(\theta_1 + \theta_2 COV_{i,j})} \qquad \text{(equation 5)}$$

where $P(U_{i,j})$ represents the probability of the presence of ulceration in individual *i* at time t_j. $COV_{i,j}$ is the aforementioned independent variable individual *i* and time t_j. Possible values for COV were 24 hour AUC (AUC24), cumulative AUC (CAUC), 24hr area under biomarker-time curve (AUEC24) and cumulative AUEC (CAUEC). θ_1, is a parameter governing the background logit probability and θ_2 is the slope parameter. A basic model was also tested where θ_2 was fixed to 0 and no covariate was used.

All adverse events were modelled in WinBUGS 1.4.3 with time, exposure and biomarker levels as independent variables using a proportional hazards model for left censored adverse events and an exponential model for current state observations. To assess the implications of different strategies for the analysis of adverse event data, multiple strategies were used to refit the data. These are listed in Table 5. In addition to the five covariates models, two different averaged models were attempted on each simulated dataset. The first averaging approach was where model predictions from all five covariate models were weighted equally. The second approach was to use the Bayesian information criterion (BIC) to weight models. This is consistent with weighting according to posterior model probability assuming a uniform prior weighting of models. The unweighted average equally weighted all models ignoring model performance for the weighting scheme.

For each model refit, prediction intervals for the risk of each observation were obtained, for plotting and calculation of predicted coverage. Predictive coverage was defined as the number of observations where the true simulated risk falls within the 95% prediction interval. Model convergence was determined by stationarity of the MCMC chain for all parameters by calculation of the Geweke statistic. Unsuccessful runs were given a 0 weight in all average and discarded from summaries.

Table 5: Overview of the methods and measurements applied to the different scenarios. The table also summarises how data are integrated to distinguish between adverse drug reactions (ADRs) and adverse events (AEs).

	Assessment	**Data to be generated**	**Data contributing to evidence synthesis**	***Estimation of uncertainty***	***Separation of ADR and AE risk***	***Quantification of delayed effects***
Current approach (NOAEL)	Observed PK exposure (over 1-day snapshot) ↓ Observed AEs	- PK (satellite groups) - Presence/absence (binary) of AEs in experimental cohorts	- None	*No*	*No*	*No*
Model-based approach (PK)	PK exposure (over course of therapy) + PK variability ↓ ADR risk	- PK (satellite groups + toxicology groups) - Predicted AE incidence rate in experimental cohorts	- Historical PK data - AE incidence rate in historical placebo	*Parameter and model uncertainty*	*Yes*	*Yes*
Model-based approach (Biomarker)	PK exposure (over course of therapy) + PK variability ↓ Biomarkers (over course of therapy) + variability ↓ ADR risk	- PK + biomarkers (satellite groups + toxicology groups) - Predicted AE incidence rate in experimental cohorts	- Historical PK data - Historical biomarker data - AE incidence rate in historical placebo	*Parameter and model uncertainty*	*Yes*	*Yes*

Results

As described below, the use of simulated data for the evaluation of hypothetical scenarios provided further insight into the limitations of current practices for the assessment of safety thresholds, and in particular of the NOAEL when taking into account differences in the underlying mechanisms of toxicity. Results from modelling are shown together with the parameter values obtained from naïve pooling and non-compartmental analysis where applicable.

It is clear from the profiles observed for scenario a1 (Figure 5) that the risk of adverse events changes in a time-dependent manner, irrespectively of the point estimates for drug exposure or effect, as determined by pharmacodynamics (biomarker levels). Such time-dependencies impose further attention to the experimental design as not only the dose level under investigation, but also the mechanism of action underlying drug toxicity will contribute to the experimental results. Moreover, these differences may or may not be captured by typical variables of interest (Table 6). The inability to discriminate the underlying mechanism of action based on the available experimental data can lead to obvious issues with model identifiability. On the other hand, despite this limitation, the large differences in convergence rate suggest that model-based estimates might be suitable to explore or exclude possible or plausible causes of toxicity.

Table 6: Overview of the predicted coverage (%) for model predictions, and corresponding model selection rates (%) for each of the scenarios (lower panel). The basic model was not included in tests for model selection.

Scenario	BMA -all models	BMA - excluding basic model	Basic	AUC24	CAUC	CAUEC	AUEC24
a1	65	78	48	78	65	65	78
a2	65	78	48	48	65	65	96
b1	48	65	30	65	65	78	78
b2	48	78	30	48	30	48	96
n1	78	78	91	78	78	78	78
n2	96	96	96	96	96	96	96

Scenario	AUC24	CAUC	CAUEC	AUEC24
n1	23	29	40	22
n2	61	25	22	6
a1	11	0	13	76
a2	0	29	28	43
b1	3	0	10	87
b2	0	24	36	40

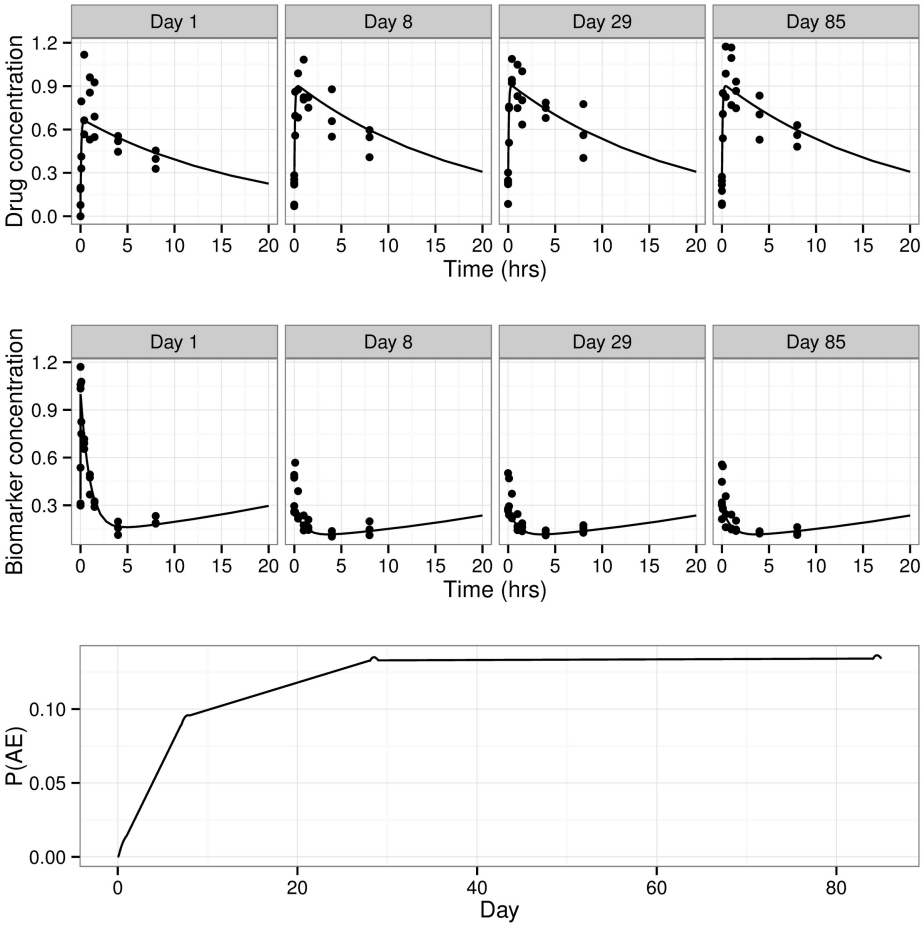


Figure 5. Plot of study data for scenario a1. This increase in risk represents approximately a 10-fold increase from baseline risk, however, overall risk is low (risk = 13% per individual; expected number of events per cohort = 1.04).

With regard to the method of analysis, our results show that the accuracy and precision of model-based estimates for AUC24, CAUC, AUEC24 and CAUEC were similar across different dosing groups and treatment durations. For all scenarios, BICs for competing models with different covariate relationships (AUC24, AUC, AUEC and AUEC24) were broadly similar. This suggests that whilst the parameters of a given model may be estimated accurately, the binary adverse events datasets simulated here provide insufficient information for model discrimination between competing models.

The success rates for convergence of models was high, the success rate being less than 100% was due to false negatives in the Geweke diagnostics. Given the alpha level of 0.05, this would result in a 9.8% failure by random chance for a model including two parameters. On the other hand, the observation that the BIC, AIC and DIC tend to overweight models without drug affect for rare adverse events implies that a conservative approach should be taken where the weighting of models should be decided by a priori confidence in a model specification rather than data fitting criteria (Table 6). A general strategy, supported by our results, to account for model uncertainty is model averaging with model weighting being independent of data fitting criteria (such as the BIC). Modellers should define models to fit to the data before conducting the analysis focusing on physiologically plausible specifications. Prior model weights should be assigned and reflect the prior belief in a model specification. Deviation from these model weights should only take place when it is strongly justified by the data (i.e. a large drop in BIC).

Safety threshold vs. exposure-risk relationships

NOAEL probability distributions are depicted in figure 6. The mostly likely outcome for the NOAEL is the peak of the distributions. An overview of the most likely outcome of the analysis of each scenario is shown in Table 7.

Lastly, Table 8 shows the likelihood for each scenario, of concluding that the compound is safe, given that based on the lack evidence of adverse events by traditional methods would imply progression of the compound. The results show that despite the characterisation of a viable therapeutic window for scenarios n1, n2, and a1, a different conclusion would be drawn by empirical methods. In fact, scenario n2 is only 43% likely to be deemed safe, whereas this figure would be even lower for scenario a1 (27%). This is likely due to the inability of these methods to quantitative account for background adverse events.

Table 7 - Likely study outcome using a standard approach (NOAEL) for a range of putative mechanisms associated with reversible and irreversible adverse events.

Scenario	Likelihood outcome given strict adherence to NOAEL approach
n1	No adverse events will be observed. Drug will appear safe at all duration levels.
n2	Drug will appear safe at 1 month and below, but at three months, the NOAEL will most likely fall below the lowest tested dose
a1	Most likely NOAEL dose at 1 week treatment is overestimated, i.e., it appears to be the mid dose level. Most likely NOAEL dose at 1 month is the low dose, but it has only 47% chance of being selected. At three months, the NOAEL will most likely fall below the lowest tested dose .
a2	At 1 week no NOAEL can be established due to lack of adverse events. Most likely NOAEL dose at 1 month is the low dose, but it has only 47% chance of being selected. At three months, the NOAEL will most likely fall below the lowest tested dose
b1	At 1 week no NOAEL can be established due to lack of adverse events. Although low dose at 1 month has been maximised, the NOAEL will most likely fall below the lowest tested dose for treatment duration > 1month.
b2	At 1 week no NOAEL can be established due to lack of adverse events. Although low dose at 1 month has been maximised, the NOAEL will most likely fall below the lowest tested dose for treatment duration > 1month

Table 8. Expected probability (%) of progression beyond the different stages of development assuming an empirical analysis where AEs observed at the low dose would lead to the compound being discontinued. To better interpret these figures, the reader is advised to compare the results presented in Figure 6, where the true NOAEL levels are presented. As it can be noticed scenarios n1,n2 and a1 are shown to have a viable therapeutic window and therefore high probability of progression. Scenarios a2, b1, and b2 are not deemed safe and therefore have low probability of progression into further development.

SCENARIO	1 week	1 month	3 months	total
n1	97.37	90.14	87.98	77.22
n2	98.14	83.76	52.73	43.34
a1	87.87	59.74	51.02	26.79
a2	95.26	61.29	15.68	9.156
b1	94.44	50.06	8.076	3.818
b2	96.85	52.49	3.757	1.91

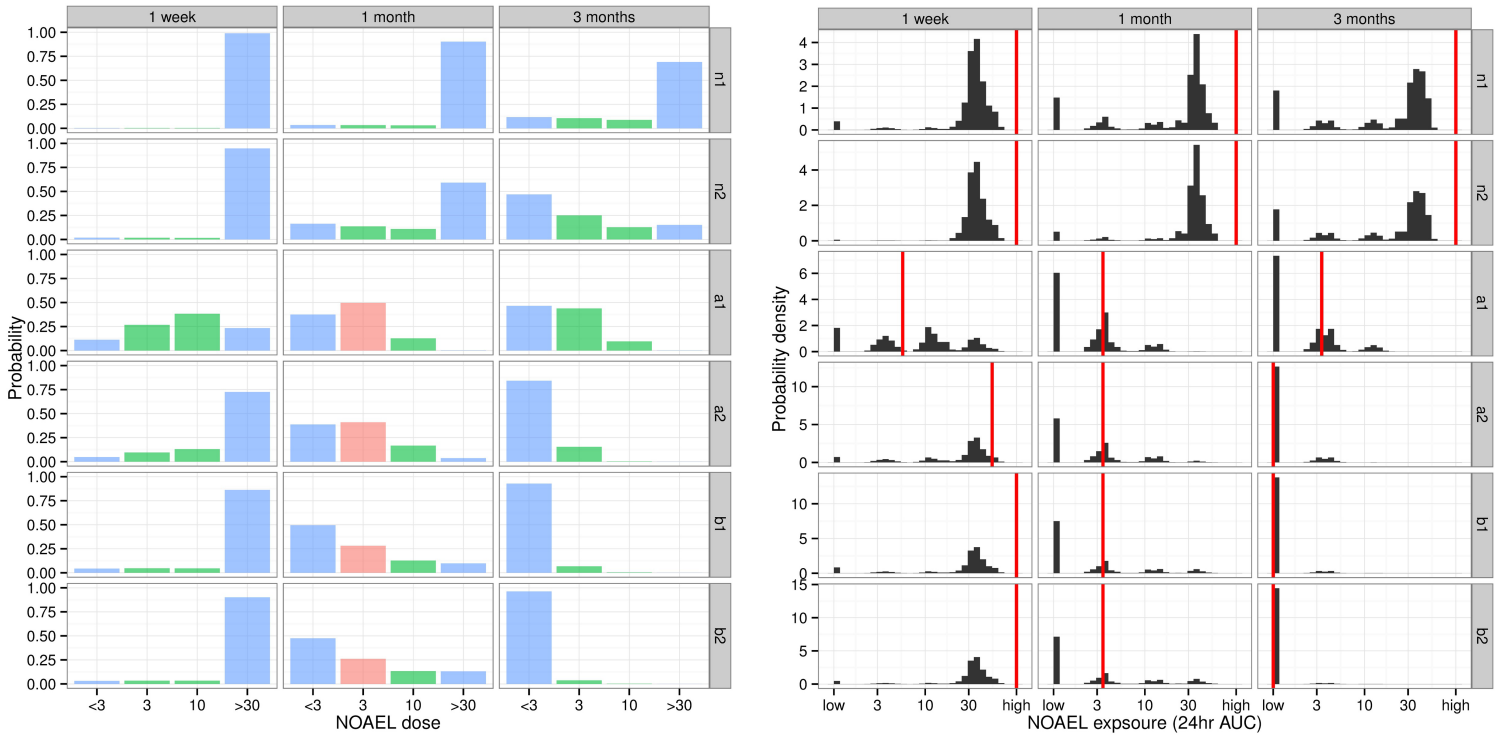


Figure 6: Probability distributions of the NOAEL dose (left) and exposure (right) for each simulated scenario. (left) Pink indicates the cohort where p(dose=NOAEL dose) is maximised. Blue indicates that a NOAEL could not be determined where either no cohorts exhibit AEs or the lowest dose cohort exhibits AEs. (Right) Red line indicates true exposure corresponding to the threshold ADR risk (threshold set to ADR risk at low dose of 1 month cohort). "Low" and "high" indicates exposures below and above the exposure range of the study, respectively.

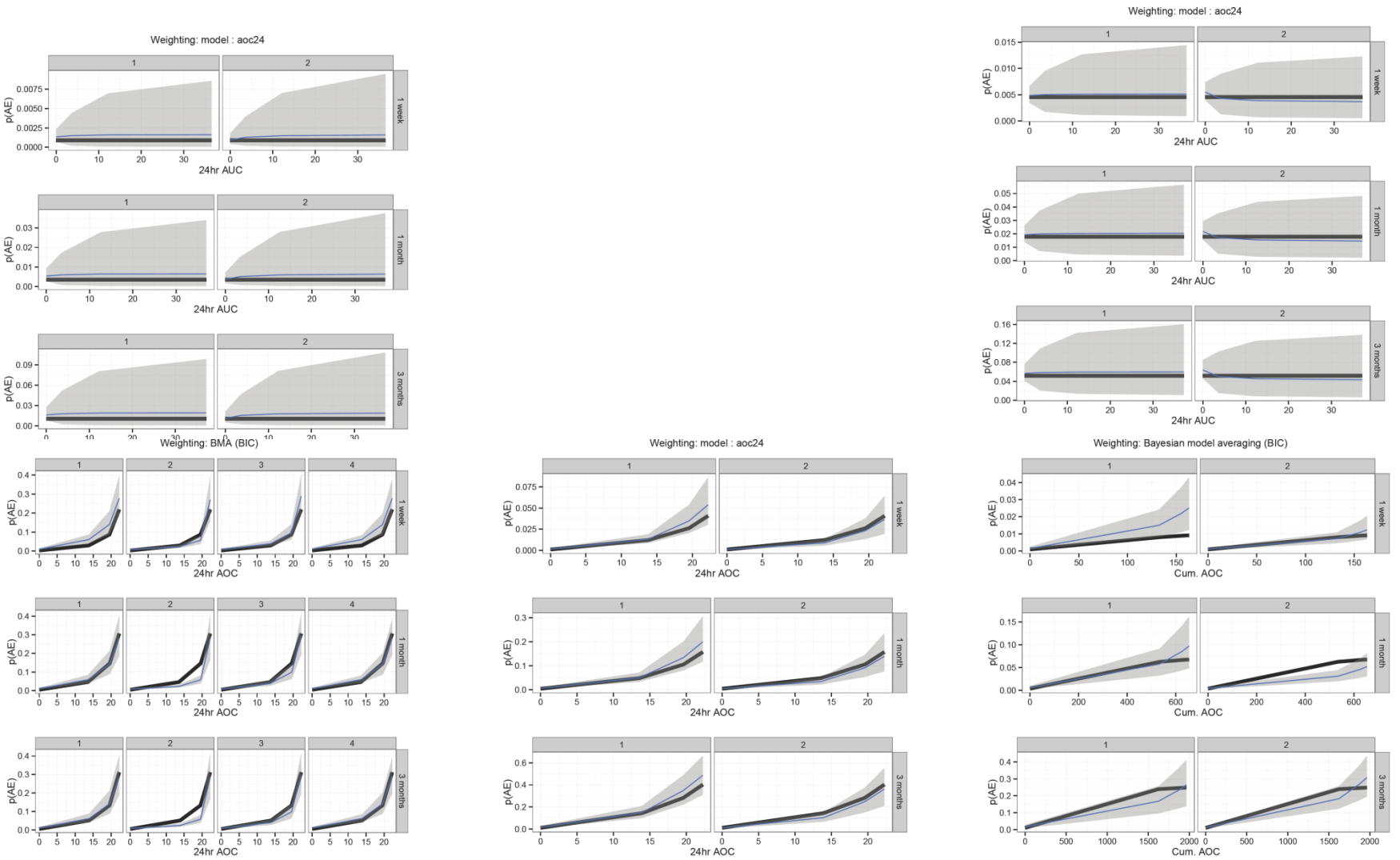


Figure 7. Relationship between predicted risk of adverse event and different measures of drug exposure, as described by the different scenarios investigated in this paper. Upper panel, scenario n1 (left), n2 (right); lower panel, scenario a1 (left), a2 (middle), b2 (right).

Discussion

Empirical experimental protocols are used to evaluate safety pharmacology and toxicity in preclinical species as the basis for defining safe exposure in humans. These protocols are not designed to understand toxicological mechanisms or provide insight into mechanism-based extrapolation across species (2,16,17). Such a limitation in experimental protocols can lead to biased conclusions about drug safety especially when the events of interest occur at low frequency or are delayed. In these circumstances, shortcomings in current approaches cannot be ignored.

In the current investigation we have illustrated the implication of low incidence adverse events on the estimates of a safety threshold (i.e., NOAEL) as compared to those obtained by the characterisation of the relationship between drug exposure and risk of adverse events. Although there may be fundamental differences in toxicity pathways at different parts of the exposure–response curve, we have assumed that examples based on a single mechanism would be sufficiently realistic to introduce the concept. The phenomenon of exposure-dependent transitions in mechanisms of toxicity can be explored in a similar manner by introducing interaction factors (18,19). Thanks to the statistical features of nonlinear mixed effects modelling, we have also shown how individual susceptibility can be incorporated into the evaluation of the exposure-response relationships, whilst taking into account differences in the underlying mechanisms involved in the continuum between exposure and adverse events.

In this regard, it is helpful to think of a multistage process, which starts with systemic exposure and progresses through target exposure, yielding early biological effects (e.g., at the sub-cellular level), altered structure or function and subsequently clinical disease (20-22). The introduction of biomarkers of pharmacology can therefore not only contribute to further understanding of target exposure, but it also enables discrimination between dose-dependency, class effects and regimen-related mechanisms without the risk of inaccuracies and poor precision which seem to prevail when relying on NOAEL estimates (23).

From a pharmacological perspective, the selection of exposure measures (i.e. the parameter of interest) forms the basis for defining "*safe exposure*". Arbitrary selection of the measure of exposure to be used as a marker of safety can add unnecessary, correlated noise into the data, which may subsequently lead to bias and loss of precision (24,25). In this context, our analysis has shown that toxicity findings associated with direct effects were most accurate represented by AUC or C_{MAX}. Moreover, in these cases, it appears that performance relative to the NOAEL approach is improved even when biomarker data was not available. By contrast, the availability of biomarker data was shown to help in the estimation of adverse events when delays occur between the beginning of treatment and onset of the effects. Other measures of exposure such as cumulative AUC proved more effect with indirect mechanisms.

The results presented here also provide guidance for prospective use of model-based approaches in the evaluation of safety pharmacology and toxicology. In contrast to current practice, in which experimental data is generated to define a safety threshold, we have shown how current understanding pharmacokinetic processes can be integrated with knowledge about the putative mechanisms of action to characterise exposure-risk relationships during safety screening in early drug development. In fact, we demonstrate how important additional, ancillary data can be when dealing with rare or low frequency events. Statistical methods are available that enable formal inclusion of such knowledge as informative priors (26, 27). However, prior distributions need be defined in advance of the analysis to minimise subjective bias. The alternative strategy of aggregating datasets is also useful if additional study data is available from a population which is exchangeable with experimental data under investigation. This underpinned the use of the aggregated placebo data set in our analysis.

From a methodological perspective, the various scenarios have shown that despite the known advantages of parametric, model-based approaches, model identifiability can be an important issue. The inability to separate models based on diagnostic criteria may lead to bias in predicted risk. Care therefore needs to be taken not to over-interpret good

performance on model diagnostics as an indication of a well specified structural model. The specification of the structural models for rare adverse events needs to be primarily justified on biological grounds. It is also important to emphasise that reporting of model outputs (e.g. parameters, model predictions) should clearly incorporate parameter and model uncertainty. Clear and accurate reporting of uncertainty enables us to understand and formally account for uncertainties arising from limitations in data collection and model estimation.

One last methodological aspect that deserves attention concerns the accuracy and precision of the estimates of safe exposure. The incorporation of placebo aggregated data allowed for quantification of baseline and drug-induced risk in a similar manner to what is currently performed for the clinical evaluation of drug safety (28,29). Parameters for baseline risk were likely to be large without sufficiently large amounts of placebo data. This would in turn inflate uncertainty around remaining parameters determining drug-induced risk. Therefore, we have modelled both components of risk together. Nevertheless, it is also feasible to model baseline risk a priori to ensure the uncertainty distributions will be narrow enough to precisely estimate drug-induced risk.

The high success rates observed with a model-based approach also shows that incorporating model uncertainty is feasible. In turn, realistic estimation of uncertainty enables more informed decision-making with regard to risk. There is a caveat that outside the experimental range, the positioning of the 95% risk bound will likely depend on a single model if equal weighting is used. This will be the model that produces highest predictions of risk at low exposures. If this proposed model is physiologically plausible and a priori equally likely to other proposed models, then this is appropriate (30). However, a linear model in this context was the least plausible and provided overly conservative estimates. A way of handling plausible, but a priori unlikely mechanisms would be to assign appropriately small model priors.

Finally, it should be noted that missed adverse events were also easily quantified using the proposed strategies. Differently, from the empirical approach to treating missing events as

absent, the use of MCMC methods provides evidence of the parameter distribution, enabling imputation of the events, even if they have not been observed.

Limitations

The scenarios used in this investigation were not intended to provide a comprehensive review of all possible toxicological mechanisms. Therefore, it should be noted that no covariate effects other than baseline incidence and treatment itself were considered to drive adverse events. As the number of potential covariates increases, the chance of selecting a false positive covariate relationship increases. This is an important consideration whenever several competing models perform similarly. Our recommendation is a pragmatic approach of restricting the model search to physiologically plausible models and the use statistical tools to guard against over-fitting (31). In the present study, this would have involved an increase in computational time by a factor of more 1000 times.

The simulated adverse events were related to descriptors of occurrence such as incidence and prevalence. These were idealised situations that represent two extremes, either where none or complete information was available. In reality adverse event data will contain a spectrum of varying degrees of information on incidence for example with interval censoring or imperfect sensitivity in detection. For instance, gastric ulceration may form and heal before the end of treatment, making histological data inaccurate for the estimation of risk. Poor specificity in detection for another adverse event may similarly overestimate risk. Resolving this uncertainty, however, is only possible with additional information regarding the pathophysiology, sensitivity and specificity of detection of the methods used to investigate these events.

In conclusion, evaluation of safety is paramount for the progression of new molecules into humans. However, current methods in preclinical toxicology do not support the integration

of pharmacokinetic and pharmacodynamic data as basis for predicting safe exposure in humans. By contrast, a model-based approach represents a viable tool for characterising risk-exposure relationships, including estimates of parameter and model uncertainty. A benefit this strategy lends to decision-making is that clinical judgment can be applied to consider the entire risk-response relationship of each adverse event, rather than a point estimate or threshold.

References

(1) Organisation for Economic Co-operation and Development (OECD) Secretariat. Chapter 4: Initial Assessment of Data. In: Manual for the Assessment of Chemicals. Organisation for Economic Co-operation and Development (OECD). April 2011.

(2) Slob W. Thresholds in toxicology and risk assessment. International Journal of Toxicology 1999; 18:345-67.

(3) Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 2005; 42(3):265-74.

(4) Hatoum NS, Leach CL, Talsma DM, Gibbons RD, Garvin PJ. A statistical Basis for using fewer rabbits in dermal irritation testing. J Amer Coll Toxicol. 1990; 9: 49-60.

(5) Gibbons, R.D., Coleman, D.E. Statistical Methods for Detection and Quantification of Environmental Contamination. John Wiley and Sons, New York, USA, 2001

(6) Bai JP, Abernethy DR. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu Rev Pharmacol Toxicol. 2013; 53:451-73.

(7) Andersen ME, Dennison JE, Thomas RS, Conolly RB. New directions in incidence-dose modeling. Trends Biotechnol. 2005; 23:122-127.

(8) Horii I. Advantages of toxicokinetics in new drug development. Toxicol Lett 1998; 102-103:657-64.

(9) al Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in population pharmacokinetics. J Pharmacokinet Biopharm. 1990; .18:347-360

(10) Nyberg J, Karlsson MO, Hooker AC. Simultaneous optimal experimental design on dose and sample times. J Pharmacokinet Pharmacodyn. 2009; 36:125-145.

(11) European Medicines Agency. Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007. (available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf)

(12) Chapter 10.2. Statistical framework for estimating population pharmacokinetics. *In:* Pharmacometrics: The Science of Quantitative Pharmacology Ette EI, Williams PJ (eds), John Willey & Sons, Inc, Hoboken, USA, 2007.

(13) Woosley RL, Romero K. Assessing cardiovascular drug safety for clinical decision-making. Nat Rev Cardiol. 2013; 10(6):330-7.

(14) Forsby A, Blaauboer B. Integration of in vitro neurotoxicity data with biokinetic modelling for the estimation of in vivo neurotoxicity. Hum Exp Toxicol. 2007; 26(4):333-8.

(15) Chen B, Dong JQ, Pan WJ, Ruiz A. Pharmacokinetics/pharmacodynamics model-supported early drug development. Curr Pharm Biotechnol. 2012; 13(7):1360-75.

(16) Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 2005; 42(3):265-74.

(17) Dorato MA, Buckley LA. Toxicology testing in drug discovery and development. Curr Protoc Toxicol. 2007; Chapter 19:Unit19.1

(18) Slikker W Jr, Andersen ME, Bogdanffy MS, Bus JS, Cohen SD, Conolly RB, David RM, Doerrer NG, Dorman DC, Gaylor DW, Hattis D, Rogers JM, Woodrow Setzer R, Swenberg JA, Wallace K. Dose-dependent transitions in mechanisms of toxicity. Toxicol Appl Pharmacol. 2004; 201(3): 203-25.

(19) Slikker W Jr1, Andersen ME, Bogdanffy MS, Bus JS, Cohen SD, Conolly RB, David RM, Doerrer NG, Dorman DC, Gaylor DW, Hattis D, Rogers JM, Setzer RW, Swenberg JA,

Wallace K. Dose-dependent transitions in mechanisms of toxicity: case studies.Toxicol Appl Pharmacol. 2004; 201(3):226-94.

(20) Crump KS. The linearized multistage model and the future of quantitative risk assessment. Hum Exp Toxicol. 1996; 15(10):787-98.

(21) Greim H.Mechanistic and toxicokinetic data reducing uncertainty in risk assessment. Toxicol Lett. 2003; 138(1-2):1-8

(22) Clewell HJ, Andersen ME. Applying mode-of-action and pharmacokinetic considerations in contemporary cancer risk assessments: an example with trichloroethylene. Crit Rev Toxicol. 2004; 34(5):385-445.

(23) Gundert-Remy U, Sonich-Mullin C; IPCS Uncertainty and Variability Planning Workgroup and Drafting Group (International Program on Chemical Safety's). The use of toxicokinetic and toxicodynamic data in risk assessment: an international perspective. Sci Total Environ. 2002; 288(1-2):3-11

(24) Wolfsegger MJ, Jaki T. Assessing systemic drug exposure in repeated dose toxicity studies in the case of complete and incomplete sampling. Biom J. 2009; 51(6):1017-29.

(25) Smith DA, Morgan P, Vogel WM, Walker DK. The use of C(av) rather than AUC in safety assessment. Regul Toxicol Pharmacol. 2010 ;57(1):70-3.

(26) Dragalin V, Fedorov V, Cheuvart B. Statistical approaches to establishing vaccine safety. Stat Med. 2002; 21(6):877-93.

(27) Bhaumik DK, Amatya A, Normand SL, Greenhouse J, Kaizar E, Neelon B, Gibbons RD. Meta-Analysis of Rare Binary Adverse Event Data. J Am Stat Assoc. 2012; 107(498):555-567.

(28) O'Quigley J, Conaway M. Extended model-based designs for more complex (dose-finding studies. Stat Med. 2011; 30(17):2062-9.

(29) Thall PF, Lee SJ. Practical model-based dose-finding in phase I clinical trials: methods based on toxicity. Int J Gynecol Cancer. 2003; 13(3):251-61.

(30) Arie Bruinink. In vitro toxicokinetics and dynamics: Modeling and interpretation of toxicity data. Pharmaceutical Sciences Encyclopedia Drug Discovery, Development, and Manufacturing. John Wiley & Sons. 2010; pp. 1-42.

(31) Lindbom L, Pihlgren P, Jonsson EN. PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput Methods Programs Biomed 2005; 79(3):241-57.

SECTION 3: CASE STUDY AND PRACTICAL APPLICATION

CHAPTER 7

Model-based analysis of thromboxane B2 and prostaglandin E2 as biomarkers in the safety evaluation of naproxen.

Tarjinder Sahota, Ian Sanderson, Meindert Danhof and Oscar Della Pasqua

Published in *Toxicology and Applied Pharmacology* **278** (3):209-219

Abstract

The assessment of safety in traditional toxicology protocols rely on evidence arising from observed adverse events (AEs) in animals and on establishing their correlation with different measures of drug exposure (e.g., Cmax and AUC). Such correlations, however, ignore the role of biomarkers, which can provide further insight into the underlying pharmacological mechanisms. Here we use naproxen as a paradigm drug to explore the feasibility of a biomarker-guided approach for the prediction of AEs in humans. A standard toxicology protocol was set up for the evaluation of effects of naproxen in rat, in which four doses were tested (7.5, 15, 40 and 80mg/kg). In addition to sparse blood sampling for the assessment of exposure, thromboxane B2 and prostaglandin E2 were also collected in satellite groups. Nonlinear mixed effects were performed to evaluate the predictive performance of the approach. A one-compartmental model with first order absorption was found to best describe the pharmacokinetics of naproxen. A nonlinear relationship between dose and bioavailability was observed which leads to a less than proportional increase in naproxen concentrations with increasing doses. The PD of TXB and PGE was described by direct inhibition models with maximum pharmacological effects achieved at doses > 7.5 mg/kg. The predicted PKPD relationship in humans was within 10-fold of the previously published values. Moreover, our results indicate that biomarkers can be used to assess interspecies differences in PKPD and extrapolated data from animals to humans. Biomarker sampling should be used systematically in general toxicity studies.

Introduction

Long term safety issues have increasingly become a cause of late stage attrition (1), prompting regulatory authorities to increase requirements for sponsors to demonstrate a favourable benefit-risk balance for new medicines. Such a prerequisite has implications for current practices in early drug discovery and development. Thus far, pharmaceutical companies seem to have adopted the concept of measuring any markers based on known pharmacology of the drug under development. One challenge is the expectation/early identification of unknown mechanisms and the timely implementation of the assessment, the other challenge is the effective translation or interpretation of the data accumulated.

Both from a clinical perspective and a pharmacological perspective, the demonstration of safety can only be tackled by a strategy that ensures the characterisation, in a mechanistic manner, of the relation between drug dosing and response (Danhof et al., 2008). The vast majority of experimental protocols currently used for the evaluation of toxicity and safety pharmacology use dose and systemic drug exposure as a proxy for risk. However, other markers of safety and toxicity may be better predictors of adverse drug reactions. This is particularly important given the high degree of nonlinearity in the relation between pharmacokinetics and pharmacodynamics as well as the potential interspecies differences in these relations. Ultimately, these nonlinearities may cause downstream biomarkers, other than target or systemic exposure to a drug, to better describe and predict the outcome or response to treatment (Bai *et al*., 2013).

There is therefore an urgent need to evaluate and refine the methodology for the assessment of safety. To this purpose, the classification scheme devised by Rawlins and Thompson, 1991, constitutes a scientific basis for the establishing correlations between adverse drug reactions and pharmacological effects. Briefly, this scheme defines adverse drug reactions according to seven different categories, which correspond to the underlying pharmacological effects. As shown in Figure 1, the different categories nicely match the mechanistic classification of biomarkers proposed by Danhof *et al.*, 2005, which defines the

requirements for establishing further correlations with drug exposure. As such, this mechanistic classification could be used for the evaluation of safety and toxicity and consequently for the accurate assessment of (long term) risk in humans.



Figure 1. Mechanistic classification of biomarkers (Reprinted with permission from Danhof *et al.,* 2005). This concept can be linked to the classification for adverse events proposed by Rawlins and Thompson, which clusters unwanted pharmacological effect into seven types, based on their mechanism of action or characteristics of their manifestation. A type A event is one that is due to an extension of the active pharmacologic properties of the drug (A indicates augmented). They are also called predictable or anticipated events. They are generally less severe and more frequent than type B events. This augmented pharmacologic action may occur at the targeted receptors or at other nontargeted receptors producing lateral effects, parallel effects, or side effects. Types C, D, E, F and G are not mechanisms but characteristics of their manifestations. Type C refers to reactions associated with long-term drug therapy. Type D is linked to carcinogenic and teratogenic effects. These reactions are delayed in onset and are very rare since extensive mutagenicity and carcinogenicity studies are done before drug is licensed. Type E refers to end of use or rebound effects. Type "F" reactions indicate failure of treatment. Type "G" reactions are due to genetic polymorphism.

In the current investigation we therefore explore the feasibility of a model-based approach for the evaluation of long term adverse events in which biomarkers of pharmacology are used as proxy of drug exposure. Naproxen, a non-selective cyclo-oxygenase inhibitor is used as paradigm compound to demonstrate the concept of biomarker-guided safety assessment (Berger *et al*., 2011; Bai *et al.*, 2013). Using pharmacokinetic-pharmacodynamic techniques, we subsequently illustrate how modelling and simulation techniques can be used to ensure accurate estimation of the safe dose levels of naproxen after chronic exposure.

Non steroidal anti-inflammatory drugs exert their actions though an interaction with cyclo-oxygenase (COX). Selective blockade of COX-1 and COX-2 activity results in direct suppression of the formation of pro-inflammatory mediators such as thromboxanes (TXB)

and prostaglandins (PG) (102). The perceived role of COX-2 in inflammation has substantiated the extensive use of selective COX-2 inhibitors as analgesic drugs in acute and chronic inflammation. Yet, despite a putative reduction in gastrointestinal bleeding and ulceration by selective inhibition (13,103), cardiovascular events have arisen after prolonged use of rofecoxib, a selective COX-2 inhibitor which led to its withdrawal from the market, followed by considerable changes in the regulatory requirements for approval of novel non-steroidal anti-inflammatory drugs (Fitzgerald, 2007).

Continuous COX-1 inhibition following prolonged administration of non-selective COX inhibitors is known to induce gastrointestinal adverse effects, in particular ulcerations and haemorrhagic bleeding. Unfortunately, at present the dose selection of COX inhibitors disregards whether maximum, long-lasting blockade of either enzyme is strictly required for response (Huntjens *et al.,* 2005). An important question that needs to be answered is therefore how much and how long COX-2 and COX-1 should be inhibited to ensure an optimal risk-benefit balance allowing for a sustained analgesic response and an appropriate safety margin in the treatment of chronic inflammatory conditions. Given the mechanism of COX inhibition and the nature of the inflammatory response, PG and TXB can be used as biomarkers of pharmacological effects (Huntjens *et al*., 2006).

Using the biomarker classification proposed by Danhof at al., 2005 and a model-based approach for the analysis and interpretation of the results, we show how a relationship can be established between drug exposure, biomarkers and safety findings (i.e., gastric ulceration).

Methods

In this investigation the safety of the non-selective COX inhibitor naproxen was evaluated at three different treatment durations using a slightly modified version of a typical general toxicology protocol. Endpoints included adverse events, including GI histology, pharmacokinetics and pharmacodynamics. The intent of the study was explore the feasibility and impact of proposed methodology. Briefly, rats received varying daily oral doses of naproxen. A predefined sampling scheme was used to monitor for adverse events, which included sacrifice of individual animals for histopathology of the GI tract. Satellite animals receiving identical doses had blood samples collected at various time points during the course of treatment for the assessment of both pharmacokinetics in plasma and biomarkers (TXB_2 and PGE_2). An overview of the study protocol is depicted in Figure 2. The diagram shows a typical experimental protocol including different treatment duration and satellite animals sampled according to composite sampling scheme.

Animals

The experimental protocol was approved by the Ethical Committee on Animal Experimentation of the University of Leiden. Experiments were performed on male Sprague-Dawley (SD) rats (Charles River B.V., Maastricht, The Netherlands) with an initial weight of 256 ± 19 g. The animals, 4 per cage, were housed in standard plastic cages with a normal 12-hour day/night schedule (lights on 07.00 a.m.) and a temperature of 21^0C. The animals had access to standard laboratory chow (RMH-TM; Hope Farms, Woerden, The Netherlands) and acidified water *ad libitum*.

Drug administration

Naproxen Sodium (Sigma Aldrich BV, Zwijndrecht, The Netherlands) was dissolved in sterile Millipore distilled H_2O. The animals received daily doses via oral gavage for periods of between 1 and 4 weeks at dose levels of 0, 7.5, 15, 40 and 80 mg/kg.

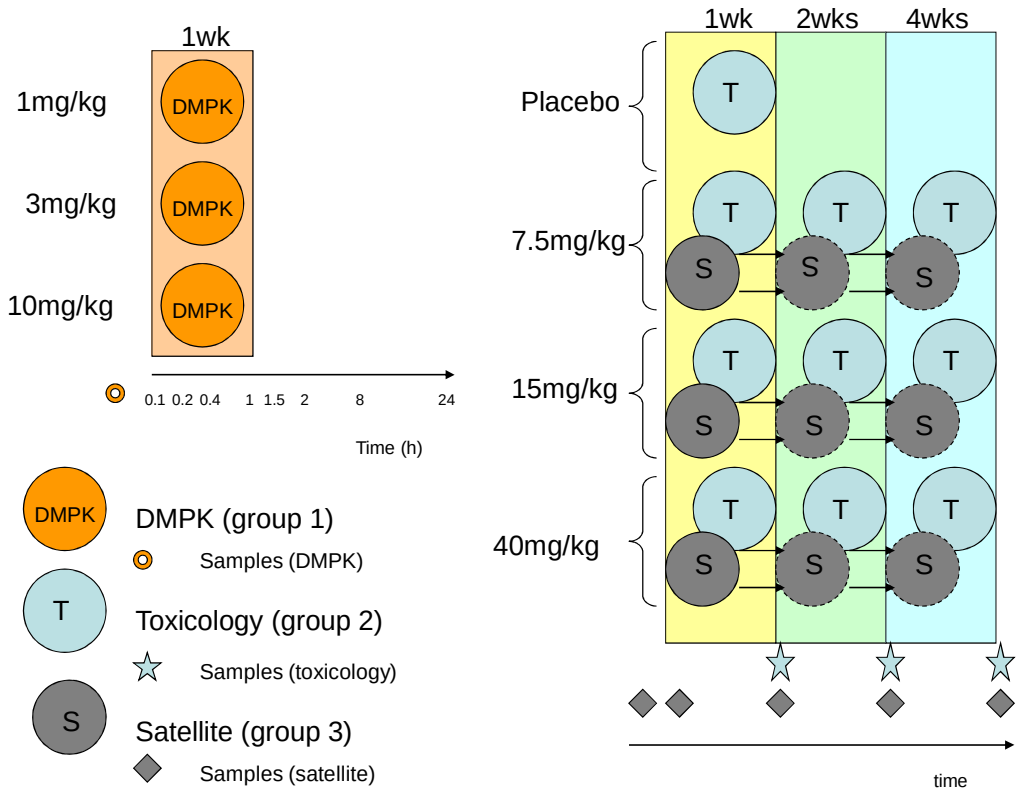


Figure 2: Schematic representation of the experimental protocol design. As the primary purpose of the study was to investigate the utility of biomarker data collection, modifications were made to a standard toxicology protocol. Only male animals were investigated and organ histology and pathology were limited to the known drug-induced toxicology findings, i.e., stomach ulceration. The use of an integrated approach implied the combination of additional data from (standard) pharmacokinetic experimental data (DMPK experiments). Animals were stratified into groups by treatment duration and dose level. Treatment duration varied between one, two and four weeks. For each treatment duration, four groups were tested each of which received four different dose levels (n = 8 animals/dose level in the toxicology group and n= 24 animals /dose level in the satellite arm, i.e., 3 animals/sampling time point).

Study Design

As the primary purpose of the study was to investigate the utility of biomarker data collection, modifications were made to a standard toxicology protocol. Only male animals were investigated and organ histology and pathology were limited to the known drug-induced toxicology findings, i.e., stomach ulceration. Animals were stratified into groups by treatment duration and dose level. Treatment duration varied between one, two and four

weeks. For each treatment duration, four groups were tested each of which received four different dose levels (n = 8 animals/dose level in the toxicology group and n= 24 animals /dose level in the satellite arm, i.e., 3 animals/sampling time point). Details of the final experimental protocol are depicted in Figure 2. It should be noted that the initial regimens included oral daily doses of 0, 15, 40 and 80 mg/kg naproxen. However due to adverse events observed in the 1-week 80 mg/kg group, the protocol was amended to 0, 7.5, 15, and 40 mg/kg cohorts. The animals receiving 80 mg/kg suffered from unacceptable weight loss and were sacrificed immediately after the first week on treatment. Unacceptable weight loss was defined as either a weight loss on three or more consecutive day or a total weight loss of more than 10% relative to the baseline value. Histological evaluation of the stomach was performed to establish a correlation between acute and long term adverse events. After euthanasia, stomachs were removed immediately and were cut open along the greater curvature and washed with warm saline. The inner surface was photographed to allow the measurement of the area covered by hemorrhagic ulceration. The area of ulceration was determined under a dissecting microscope. Gastric ulceration was measured as percentage stomach surface area affected by ulceration. A software (Image J version 1.43) was used for calculating ulcer area and total stomach surface area. The person who performed the ulceration measurement was blinded as to treatment group.

Given the need to establish a correlation between drug exposure, biomarkers, and adverse events, optimality concepts were used to ensure accurate characterisation of pharmacokinetics and biomarkers. In addition, an integrated approach was used which takes into account the pharmacokinetics of the naproxen at putative therapeutic levels (i.e., the so-called DMPK group). In contrast to standard protocols, sparse and serial blood sampling schemes were considered. For the DMPK group, serial samples were collected at 0, 0.25, 0.75, 1.5, 8, 24 hours after dosing. The sampling times in this set of animals were not optimised for subsequent modelling. On the other hand, optimal design methodology was used to select a sampling scheme and individual sampling times for satellite and toxicity animals (Josa *et al.*, 2001). Optimisation of the sampling scheme has been performed according to D-optimality principles, as implemented in PopED (University of Uppsala,

Sweden). Due to practical constraints, optimized sampling time points were rounded to the nearest 15 min after dosing. The final schedule included therefore the following sampling times: 0, 0.25, 0.5, 0.75, 1, 1.5, 2, 8, 24 hours post dose. Blood samples of 250µl were taken from the tail vein and split into aliquots, namely 100µl for naproxen concentrations and PGE2 and 50µl for TXB_2. Animals allocated to the toxicology groups were sampled only at the beginning and end of the treatment, prior to sacrifice. Satellite animals were each sampled four times throughout the study after dosing on day 1, 7, 14 and 28. Blood samples for pharmacokinetics were placed into heparinised tubes and centrifuged at 5000 rpm for 10 min. Plasma was stored at -20^0C until analysis. Blood samples for TXB_2 analysis were placed into tubes and allowed to clot for 1 hour at 37^0C in a stirring water bath. Serum was collected after centrifugation and stored at -20^0C until analysis. Tubes for the analysis of PGE_2 were prepared by evaporating aspirin (10 µg/ml in methanol and heparin (10 IU)). Blood samples were placed in tubes together with 10 µg/ml lipopolysaccharide (LPS). Samples were incubated and stirred for 24 hours at 37^0C in a water bath. Plasma was separated by centrifugation and stored at -20^0C until analysis.

Bioanalysis of naproxen

Naproxen concentrations were analysed via HPLC in accordance with the method described by Satterwhite and Boudinot (1988). 50µl Plasma samples were spiked with 50µl internal standard (1000mg/ml ketoprofen in methanol). The pH was then adjusted via addition of 0.2ml 1M phosphate solution at pH 2. The extraction process was performed with 5ml diethyl ether, after which the residue was then dissolved in 100µl mobile phase and then 50µl of this solution was injected into the HPLC system. The HPLC system consisted of a Water 501 solvent pump, a Waters 717plus autosampler (Millipore-Waters, Milford, MA, USA), Superflow 757 Kratus UV absorbance detector (Shimadzu, Kyoto, Japan). A C18 3µm cartridge column (100 x 4.6mm i.d., Chrompack, Bergen op Zoom, The Netherlands) was equipped with a guard column for the chromatography process. Mobile phase was made up of an 82:18v/v of 0.02M phosphate buffer (pH 7.0) and acetonitrile and was set to a flow

rate of 1.0ml/min. Measurement of ultraviolet absorbance was performed at a wavelength of 258nm. The data was acquired and processed using a Chromatopac CR3A integrator (Shimadzu, Kyoto, Japan). The calibration curves showed linearity over range of known concentrations (250-100,000ng/ml). Validation was carried out and the analytical process was shown to have a mean accuracy and precision of 96.3% and 2.94%, respectively. The intra-assay variability was shown to be 2.97%.

Analysis of TXB_2 and PGE_2

PGE_2 and TXB_2 were quantified by an *in vitro* whole blood assay (WBA) using a validated enzyme immunoassay (Amersham Biosciences Europe GmbH, Freiburg, Germany). Samples were diluted in assay buffer (2-50 times for PGE_2, 200-2000 times for TXB_2) and a 50 µl sample was transferred into a coated well plate. After addition of 50 µl antibody and 50 µl peroxidase conjugate, samples were incubated for 1 hour, washed four times and incubated for 15 min (TXB_2) or 30 min (PGE_2) after which 150 µl substrate was added. The enzyme reaction was halted by addition of 100 µl 1M sulphuric acid and optical density was measured in a plate reader at 450 nm.

Data analysis

Pharmacokinetic and biomarker data from all experimental groups were combined for an integrated analysis of pharmacokinetics and pharmacodynamics of naproxen using nonlinear mixed effects modelling, as implemented in NONMEM version 7.2.0. Convergence was determined by successful minimisation and covariance step. Final model parameters were estimated by the first order conditional estimation method with interaction (FOCEI). This approach allows the estimation of inter- and intraindividual variability in model parameters. All fitting procedures were performed on a computer (AMD-Athlon XP-M 3000+) running under Windows XP with a FORTRAN compiler (Compaq Visual Fortran, version 6.1). Data

processing, management and graphical display were performed in R (R Development Core Team, 2012). Model diagnostics and validation were performed according to graphical and statistical criteria. Goodness-of-fit plots, including observed (OBS) *versus* individual prediction (IPRED), OBS *versus* population prediction (PRED), conditional weighted residuals (CWRES) *versus* time and CWRES *versus* OBS were used for diagnostic purposes (104). Model validation included numerical predictive checks (NPC), visual predictive checks (VPC) and normalised prediction distribution errors (NPDE). If shrinkage was found to be high (>20%), diagnostics using empirical Bayes estimates (EBEs), e.g. plots involving IPRED or individual parameter estimates, were not performed due to their reduced diagnostic value.

Pharmacokinetic (PK) model: The pharmacokinetics of naproxen was described initially by a one-compartmental model with first order absorption and first order elimination assuming a (relative) bioavailability of 1. Additional compartments and dose-dependent kinetics were also evaluated during model building. Model selection and identification was based on the likelihood ratio test, parameter point estimates and their respective 95% confidence intervals, parameter correlations and goodness-of-fit plots. For the likelihood ratio test, the significance level was set at $p<0.01$, which corresponds with a decrease of 6.6 points, after the inclusion of one parameter, in the minimum value of the objective function (MVOF) under the assumption that the difference in MVOF between two nested models is χ^2 distributed.

Based on model selection criteria, naproxen pharmacokinetics was best described by a one compartment model including absorption rate constant (Ka), clearance (CL) and volume of distribution (V) as primary parameters. The analysis was performed by use of the ADVAN13 routine in NONMEM. Variability in pharmacokinetic parameters was assumed to be log-normally distributed in the population. An exponential distribution model was used to account for inter-individual variability:

$$P_i = \theta_i \cdot \exp(\eta_i) \qquad \text{equation (1)}$$

where θ is the population estimate for parameter *P*, P_i is the individual estimate and η_i is the normally distributed interindividual random variable with mean zero and variance ω^2. The coefficient of variation (CV %) of the structural model parameters is expressed as percentage of the root mean square of the interindividual variance term.

Selection of an appropriate residual error model was based on inspection of the goodness-of-fit plots. A combination of a proportional and an additive error model was then proposed to describe residual error in the plasma drug concentration:

$$C_{obs,ij} = C_{pred,ij} \cdot (1 + \varepsilon_{ij,1}) + \varepsilon_{ij,2} \qquad \text{equation (2)}$$

where *Cobs,ij* is the j[th] observed concentration in the i[th] individual, *Cpred,ij* is the predicted concentration, and *εij* is the normally distributed residual random variable with mean zero and variance σ^2. The residual error term contains all the error terms that cannot be explained by other fixed effects including experimental error (e.g., error in recording sampling times) and structural model misspecification.

Pharmacokinetic-pharmacodynamic (PKPD) model: The PKPD data were analysed sequentially using the so-called PPP&D approach (105,106). In the PPP&D sequential analysis, population pharmacokinetic parameters are fixed, but individual pharmacokinetic parameters are estimated simultaneously with pharmacodynamic parameters based on both PK and PD data. Even though more computationally intensive, we have preferred this strategy to the more common usage of simulated plasma concentration from empirical Bayes estimates as an independent variable since the high expected shrinkage would result in overestimation of the variance of PKPD random effects parameters (Karlsson et al., 2007).

PGE_2 and TXB_2 concentrations were used in this study as markers of the underlying pharmacological effects with the aim of identifying their relevance as a proxy for safety after naproxen exposure. The sigmoid I_{max} model was used to relate naproxen plasma concentration (C) to the drug effect by the equation:

$$Effect = I_0 - (I_0 - I_{\max}) * (C^n / (C^n + IC_{50}^n)) \qquad \text{equation (3)}$$

where I_{-max} represents the maximal inhibitory effect to naproxen, I_0 the baseline production of PGE2 or TXB_2 and *n* the Hill coefficient. This equation is an adaptation from the E_{max} model in order to obtain the absolute values for I_0 and I_{max} for the direct calculation of maximal inhibition in percentages. To allow for further comparison with historical data, the model was re-parameterised during the final analysis to obtain estimates of IC_{80}, i.e., the concentration corresponding to 80% biomarker inhibition. The relationship between IC_{50} and IC_{80} can be implemented by the following equation:

$$IC_{80} = IC_{50} \cdot \sqrt[n]{4} \qquad \text{equation (4)}$$

Covariates: The role of potential covariate factors on PK and PKPD model parameters was evaluated using the stepwise covariate method (SCM) in PsN (107). Potential influential factors included clock time, body weight, age and biomarker levels at baseline. Covariates were incorporated into the model by stepwise forward inclusion. A significance level of $p<0.01$ was used for inclusion, which represented a drop of least 6.63 units in the objective function for each additional parameter. A final evaluation of the statistical significance of all factors identified during the previous step was performed by subtracting each covariate individually (backward elimination). The final structural model (i.e., fixed effects model) included only those covariates whose subtraction resulted in a decrease of at least 3.84 units in the objective function ($p<0.05$).

Posterior predictive performance evaluation: The performance of the population PK and PKPD models were assessed by numerical and visual predictive checks. To that purpose, 1000 data sets were simulated with the final model parameter estimates. The mean and the 95 % confidence intervals were calculated for naproxen, PGE_2 and TXB_2 concentrations at the pre-defined sampling time points used in the experimental protocols. Validation procedures also included normalised prediction distribution errors (NPDE), which are based on the assumption that the normalised (decorrelated) prediction distribution errors (discrepancies) are normally distributed (Comets *et al*., 2008). One hundred datasets were simulated using

the final model, which was then tested for the assumption of normality of the prediction distribution errors.

Exposure calculation: In addition to the use of a model-based approach to estimate relevant pharmacokinetic parameters, naproxen plasma data from the satellite groups were also analysed using traditional non-compartmental (NCA) methods for comparison with the predicted values of systemic exposure. Data for each cohort was aggregated by time point to produce composite, geometric mean naproxen concentrations at each time point. Summary statistics were performed on this composite profile for the peak concentrations (C_{MAX}) and area under the concentration *vs*. time curve (AUC). Composite C_{MAX} was taken to be the highest point on the composite profile (at time T_{MAX}). Composite AUC was calculated via the log-linear trapezoidal rule where a linear increase in the concentration vs. time profile was assumed till T_{MAX}, and a log-linear decline thereafter.

Model performance was first assessed by means of a predictive check using 1000 simulations. Composite AUCs and composite C_{MAX} values were calculated on each simulated dataset and then compared to the observed values. The model performing well on this predictive check was used to compute model-based AUC and C_{MAX}. To ensure higher precision of the predicted measures of systemic exposure, 100 animals were simulated per cohort. Empirical exposure calculations were then compared to model-based results.

Simulations**:** As described previously, inferences about the safety profile of a drug in humans may be more accurate if biomarkers are considered in conjunction with or eventually as proxy for naproxen exposure. Extrapolation of preclinical findings into drug effects in humans was therefore based on the predicted exposure-biomarker relationships in humans (108) Using the PKPD models developed for PGE_2 and TXB_2 inhibition, simulation scenarios were evaluated for wide range of naproxen concentrations. Despite some evidence of differences in the homeostasis of prostacyclins in rats, for the purpose of this investigation downstream effects were assumed to reflect the mechanisms by which adverse drug reactions emerge both in rats and humans. The inhibition levels were then compared to previously reported data on the PKPD relationship obtained in vitro by Huntjens *et al.*, 2006

using healthy subject blood and to the AUC and C_{MAX} observed after the recommended 500mg b.i.d. dose, as described in the naproxen prescriber information (Roche Pharmaceuticals Australia, 2012).

Results

In total, the dataset consists of 550 samples were collected for the evaluation of the pharmacokinetics (n=113) and toxicokinetics of naproxen (n=437). 16.7% of the samples had concentrations below the lower limit of quantification. Due to calibration curve issues, not all samples could be evaluated in a pairwise manner with the pharmacokinetic data. In total, 65 samples were analysed for PGE_2 and 73 TXB_2 levels, none falling below the limit of quantification.

As shown in Figure 3, histological examination revealed gastric ulceration in all dose levels; therefore no NOAEL could be obtained for any of the treatment durations. Ulceration occurred at an incidence of 11% after administration of the 7.5 mg/kg dose, which was defined as the lowest observed adverse effect level (LOAEL).

All animals receiving 80 mg/kg in the 1 week cohorts suffered from unacceptable weight loss. The animals in this cohort were immediately sacrificed, and due to ethical reasons this dose level was discontinued. Different dose levels of naproxen were considered for the two-week and one-month cohorts, which received lower doses (7.5, 15, and 40mg/kg). Histological examination and terminal blood samples were performed on these animals. No other adverse events were reported.

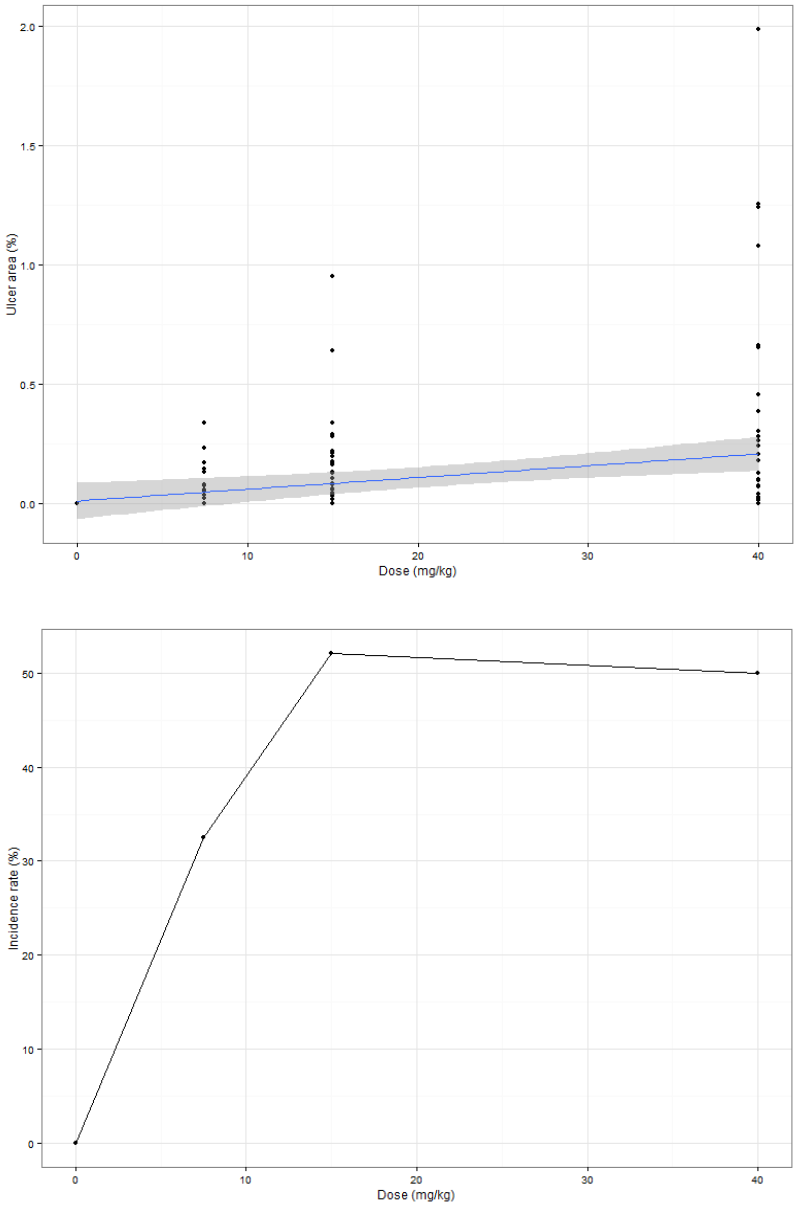


Figure 3. Plots of gastric ulcer incidence and severity after oral administration of 7.5, 15 and 40 mg/kg naproxen to rats. Ulcer severity is measured as % of stomach area covered in ulcers. The dots in the upper panel represent observed events in individual animals, whereas the solid line and shaded grey area represent the regression line and 95% confidence interval.

Pharmacokinetic model

Naproxen pharmacokinetics in plasma was best described by a one-compartment model with first order absorption, first order elimination and dose-dependent bioavailability (Figure 4). Interindividual variability was identified on all model parameters, whereas residual variability was described by a combined proportional error model. As expected, η and ε shrinkage was high (>20%) due to the sparseness of the data. Weight was found to show a statistically significant effect on both clearance and volume of distribution. An overview of the final parameter estimates is presented in Table 1.

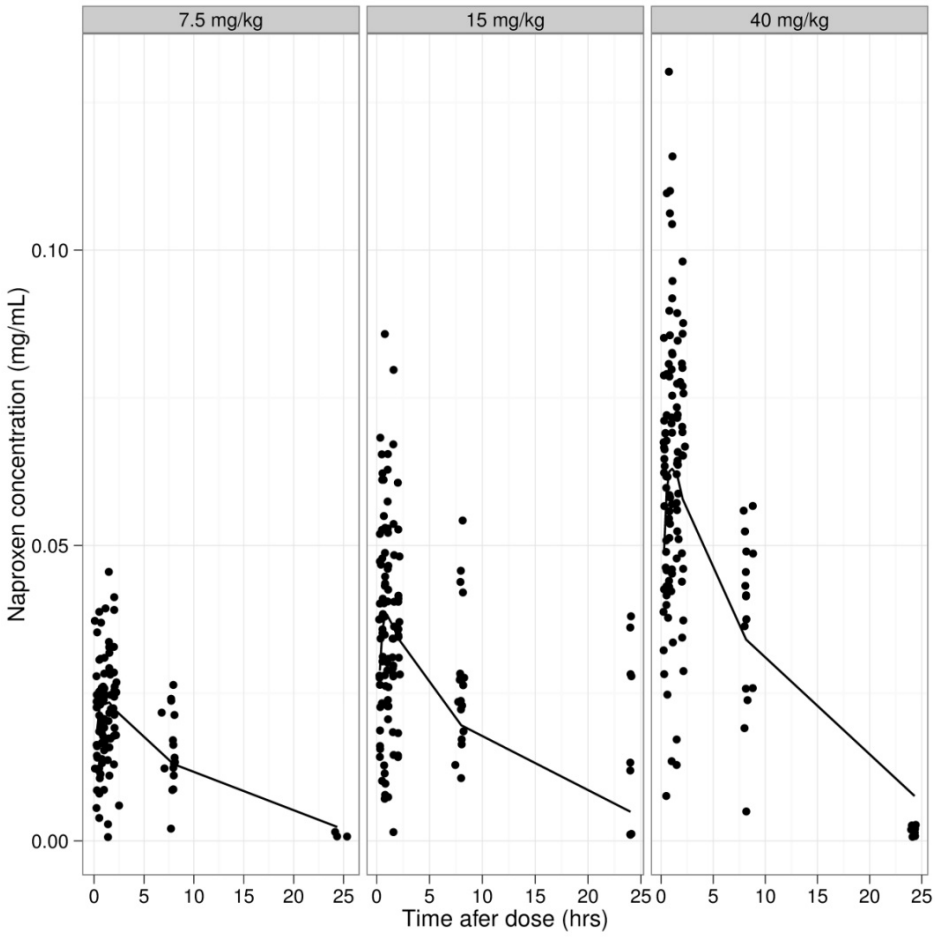


Figure 4: Naproxen concentrations vs. time profiles after oral administration of 7.5, 15 and 40 mg/kg to rats. The dots represent observed concentrations, whereas the solid line represents the population predicted profiles.

Table 1. Final parameter estimates

PK Parameter	Final estimate	Precision (CV%)
Ka (h^{-1})	3.7	14.95
CL/F (L/h)	6.42	7.85
V/F (L)	67	7.43
$F1_{diff}$	2.47	5.75
$SLOPE_{WT,CL/F}$[a]	0.00531	-
$SLOPE_{WT,V/F}$ [b]	0.00229	-
BSV in Ka (%CV)	15.8	196
BSV in CL/F (%CV)	43.0	13.46
BSV in V/F (%CV)	12.7	66.25
Residual variability	41.59	8.97
PGE Parameter	**Final estimate**	**Precision (CV%)**
I0 (ng/ml)	57.97	5.09
IC50 (mg/L)	0.0132	8.27
HILL	1.51	-
BSV in I0 (%CV)	12.69	49.3
BSV in IC50 (%CV)	33.91	63.0
Residual variability	14.76	41.4
Residual variability (additive)	1.43	35.4
TXB Parameter	**Final estimate**	**Precision (CV%)**
I0 (ng/ml)	192.48	-
IC50 (mg/L)	0.000599	-
HILL	1 FIX	-
BSV in I0 (%CV)	42.78	-
Residual variability	14.76	-
Residual variability (additive)	33.62	-

[a]Covariate relationship was modelled as $CL/F*(1 + SLOPE_{WT,CL/F} *(WT - 296.60))$

[b]Covariate relationship was modelled as $V/F*(1 + SLOPE_{WT,V/F} *(WT - 296.60))$

Pharmacokinetic-pharmacodynamic models

Given the fast onset of effect, the naproxen-induced inhibition of PGE_2 and TXB_2 could be characterised by direct sigmoid I_{MAX} models. The final model for PGE_2 included estimates of Hill coefficient different from 1, whereas for the TXB this parameter was fixed to 1. Clearly, the doses used in this experimental protocol has led to considerable level of inhibition of TXB_2, which nears complete suppression at the highest concentrations. A similar pattern was observed for PGE_2, but the profiles are much more variable (Figure 5). No relevant deviation or model misspecification was observed in any of the diagnostics measures. In addition, NPDE plots suggested no significant discrepancies across the range of predicted concentrations (Figure 1S, supplemental material). Yet, it should be noted that 24-hour TXB_2 samples were not well predicted due to a rebound effect at the end of treatment, which could not be captured by the model.

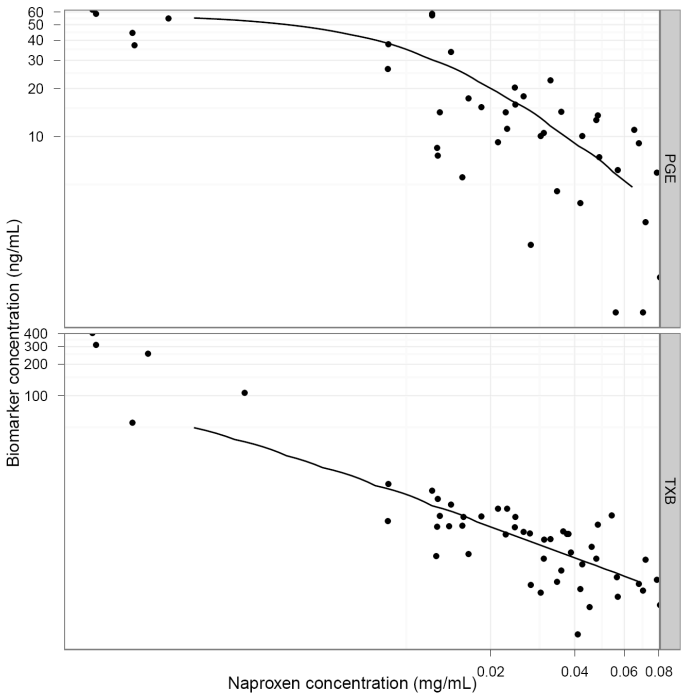


Figure 5: Pharmacokinetic-pharmacodynamic relationships for PGE_2 and TXB_2, after administration of increasing doses of naproxen (7.5, 15 and 40 mg/kg). Data pooled from animals treated during 1, 2 and 4 weeks. Pharmacological effects are assumed to be time independent, i.e., no tolerance or hypersensitisation is observed at the different durations of treatment. Dots represent observed levels of PGE_2 and TXB_2, whereas the solid line depicts the population predicted inhibition.

Exposure calculation

The predictive check shown in Figure 6 was performed to assess the model's ability to accurately predict drug exposure, as defined non-compartmentally in terms of AUC and $C_{MAX.}$ It shows that model predictions are slightly different from the observed exposure estimates obtained parametrically. As depicted in the predictive check for derived measure of exposure (see Figure 2S, supplemental material), this bias may remain undetectable when data analysis is performed by non-compartmental methods, which handle variability as random noise. By contrast, hierarchical modelling of pooled data assumes part of variation to be caused by inter-individual differences in the underlying parameters that determine the time course of drug concentrations. With the exception of the last time point of the 40 mg/kg dose group, C_{MAX} and AUC values derived by NCA are systematically overestimated. In addition, it should be noted that one cannot discriminate the impact of drug accumulation based on NCA results. Naproxen accumulation over time upon repeated dosing is evident from the model-predicted. Model predicted-curves shows also reveal a risk of significantly higher than average AUCs for some individuals in the 40 mg/kg group.

Simulations

Figure 7 shows the PKPD relationship for PGE_2 and TXB_2 obtained from the pooling of data from the present study as compared to the ex vivo results published by Huntjes and collaborators (Huntjes et al., 2006). In contrast to the observed pharmacological profile using human blood, which suggests similar IC_{80}s for the inhibition of both PGE_2 and TXB_2, naproxen was found to show higher potency in terms of TXB inhibition in rats. These differences strengthen our assumption that the differences in homeostasis in pre-clinical species must be considered when interpreting toxicology and safety pharmacology findings. On the other hands, the PKPD curves reveal two important features of the safety pharmacology of naproxen. First, it can be observed that the exposure range associated with the LOAEL dose, where GI toxicity was evident in rats lies above the predicted IC_{80} values in rats. Second, one can see that the exposure range observed after the currently recommended doses of naproxen also lies above the IC_{80} estimates in humans. This finding

is in agreement with the assumption that GI toxicity is induced by high degree of COX-1 inhibition.

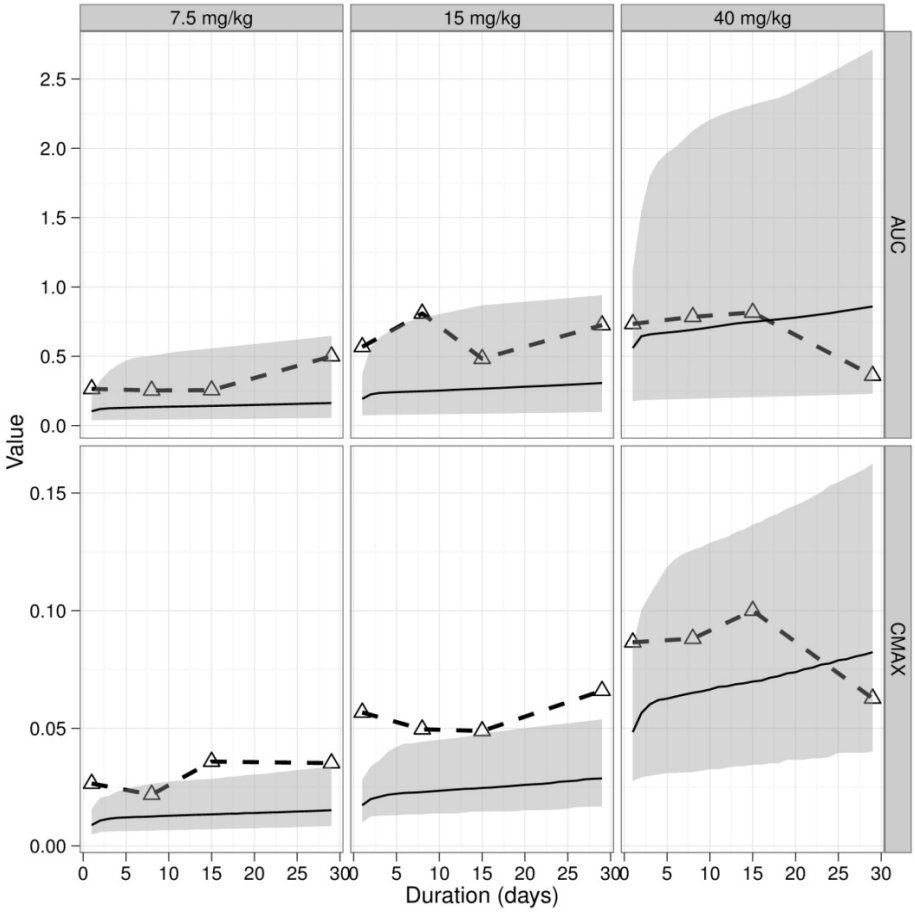


Figure 6: Model-based predictions and estimated non-compartmental values for systemic exposure (AUC) and peak concentrations (C_{MAX}) after single and repeated oral administration of 7.5, 15 and 40 mg/kg naproxen to rats. Observed exposure (triangles and dotted line) is shown together with model predicted parameter estimates. Solid line represents the predicted median values, whereas the shaded area indicates the 95% prediction intervals. The discrepancy between model predictions and observed AUC and C_{MAX} values is not a result of model misspecification. It is caused by the bias from non-compartmental analysis.

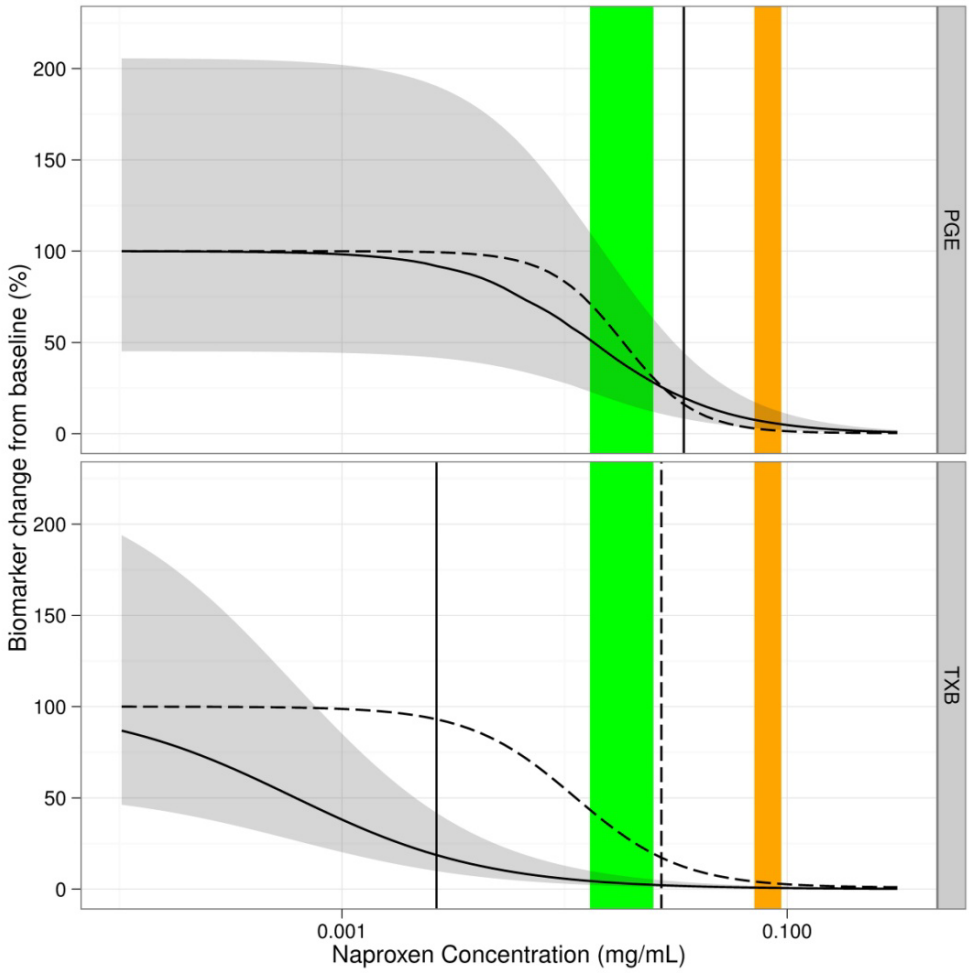


Figure 7: Estimated PKPD relationships for TXB_2 and PGE_2 in rats. The solid curve with shaded area representing 95% prediction interval, vertical line shows corresponding IC_{80}) whilst the green-shaded bar depicts the drug levels associated with the LOAEL dose (7.5 mg/kg) observed in animals. The preclinical findings are compared to the corresponding concentration-effect relationships in humans, as determined by *ex vivo* assays in whole blood using a wide range of naproxen doses. The dashed line depicts PKPD curves in healthy subjects along with the IC_{80} values (dashed vertical line). The distance between the solid and dashed lines shows the magnitude of inter-species differences in terms of the sensitivity to the thromboxane (anti-platelet aggregation) effects. The orange-shaded bar corresponds to therapeutic levels (Css to Cmax) observed at the approved doses of naproxen.

Discussion

Historically, general toxicity experiments have been designed with the primary objective of deriving estimates of systemic drug exposure and a safety threshold, i.e., the NOAEL (Parasuraman, 2011). Another major goal of repeat-dose general toxicology experiments is to identify target organs. However, important limitations in protocol design such as the sparseness of the data collected, the inferences made from separate satellite groups and the descriptive nature of the data analysis preclude their use for further characterisation of concentration-effect relationships. There is barely any consideration about the degree of receptor occupancy or target engagement at tissue and organ levels. In addition, experimental and statistical methods rely on sparse sampling schemes which prevent the identification of the different sources of variability at the proposed dosing regimens (Chain and Dubois *et al.*, 2013). Here we have attempted to circumvent these conceptual and experimental limitations using a biomarker guided approach in which the primary objective is not to obtain a NOAEL, but rather to characterise the exposure-effect relationships associated with the observed adverse events.

Firstly, it should be noted that by incorporating data from typical pharmacokinetic studies at pharmacological levels with serial sampling we were able to accurately describe the changes in drug absorption and disposition which occur with increasing dose levels. Secondly, we have analysed all the data generated on toxicokinetics and toxicodynamics in a single, integrated model, rather than separately according to the traditional group by group comparison. Finally, we performed sequential pharmacokinetic-pharmacodynamic modelling of the data using the so called PPPD approach, which ensures that pharmacodynamic parameter estimation properly accounts for the individual uncertainty in pharmacokinetics (due to sparse sampling) in those animals where biomarkers information was available (105). In summary, this approach combines the necessary statistical rigour for accurate characterisation of the exposure-effect relationship.

From a conceptual point of view, it is worth mentioning that despite current understanding about the contribution of drug target to the safety profile of a compound, chronic toxicology protocols still rely on the assumption that unwanted events will occur at some frequency, making the conclusions about risk highly dependent on the experimental conditions (Lazarou *et al.*, 1998; Guzelian *et al.*, 2005; van Vliet, 2011). We understand that during drug development there will be instances in which the mechanisms of toxicity may not or cannot be established. In fact, from a regulatory perspective, this information is actually only very rarely obtained. In addition, in many cases toxicity may result from off-target effects and biomarkers may not be available. This is a common feature in some therapeutic areas where intended pharmacology does not involve host targets (e.g., antiviral drugs, antibiotics). In such circumstances, a model-based approach would still be preferred to standard methods, but drug exposure rather than biomarkers should be considered. Yet, these limitations should not preclude us from advancing developing more integrated protocols, incorporating measures of primary and secondary pharmacological activity into the assessment of safety and toxicity.

Our experiments yielded suitable data for modelling of the inhibitory effects of naproxen on TXB_2 and PGE_2, as shown by the goodness of fit diagnostics. Moreover, our analysis enabled the incorporation of non-linearity in the pharmacokinetics of naproxen, which occurs at high dose levels (Runkel *et al.*, 1974; Josa *et al.*, 2001). Unfortunately, due to the lack of intravenous data, it was not possible to establish whether dose-dependent pharmacokinetics results from incomplete absorption, possibly limited by surface area, saturations of transporters or by first-pass metabolism. Our findings also corroborate the data published previously by Huntjens *et al.* (2006). Moreover, the characterisation of the PKPD relationships for TXB_2 and PGE_2 provides a more useful summary of findings than one normally can deduce from the reporting of observed drug exposure, NOAEL and LOAEL, which, in general, are gender and strain-dependent (Urushidani *et al.*, 1978; Nicolson *et al.*, 2010). The predicted levels of biomarker inhibition across a wide concentration range and the evidence from human *in vitro* experiment indicates how interspecies differences in the

underlying pharmacological effects may be used to translate safety findings. The incidence of adverse events at exposure levels that correspond to IC_{80} values suggests a possible causal association between adverse events and prostanoids (Laine *et al.*, 2008). Hence, it can be concluded that both analgesia and gastrointestinal adverse events seem to occur at therapeutic drug levels.

Biomarkers of drug effects as proxy of drug exposure

Here we make a plea for the use of biomarkers of pharmacology as the basis for defining interspecies validity and interpreting risk in humans. Our investigation shows how a model-based approach can be used to integrate pharmacokinetic and pharmacodynamic data for the evaluation of safety pharmacology and long term toxicity, enabling the incorporation of biomarkers of pharmacological activity into the assessment of safety margin and other measures of risk. In addition, these results emphasise the role of construct validity to account for the potential impact of interspecies differences in the underlying exposure-response relationships (Knight, 2007). As indicated by the level of biomarker inhibition observed at the selected doses inferences from the preclinical data may be used to infer drug effects at comparable levels of inhibition in humans. Naproxen's prescriber information provides data on the incidence of gastro-intestinal side effects varying between 1 and 4% with increasing doses. Despite comparable drug concentrations in rats receiving doses up to 40 mg/kg naproxen and in patients taking therapeutic doses, interspecies differences in the sensitivity to the effects of naproxen on TXB_2 may explain the lower incidence of adverse events in humans as compared to the findings in rats.

Clearly strategies are needed in the evaluation of long term safety and toxicity that increase full and impartial examination of existing data before generating new evidence using experimental protocols. Understanding of the underlying pharmacokinetic-pharmacodynamic (PKPD) relationships becomes therefore a pre-requisite to improve the methodological quality and minimise the consumption of animal and other resources within

experiments of questionable utility. As stated by Perel et al. (2007), the failure of animal models to adequately represent chronic disease processes in humans may be one of the fundamental causes of the poor predictive value of preclinical data. Yet, the authors seem to overlook the relevance of the underlying PKPD relationships to explain concordance or discrepancy between animal and clinical data. In addition establishing the correlation between drug exposure and pharmacodynamics, another advantage of a biomarker-guided evaluation of safety is the possibility to make inferences about long term effects. The implications of chronic treatment, expressed in terms of total daily dose or systemic concentrations may not be as sensitive to allow characterisation of risk. In our investigation with naproxen, one of the major concerns has been the potential for cardiovascular risk associated with the chronic use of COX inhibitors (Figure 3S). More specifically, an issue that remains unanswered is how to best predict the implications of long term suppression of COX-2 activity. Evidence exists for the role of PGE_2 and other prostanoids, which suggests their contribution to tissue healing and repair (109). Information on the levels of PGE2 inhibition (instead of systemic concentrations or dose level) may facilitate the interpretation and translation of chronic safety data.

Integrated design and analysis of safety pharmacology and toxicology protocols

Several challenges exist to successfully translating the outcomes from animal research to humans in a clinical setting. Despite the efforts to account for biological and genetic differences between species and strains in the interpretation of findings, these differences are often disregarded in the design of animal studies (Hooijmans et al., 2013). In addition, the statistical methods used to analyse results are often questionable (Kilkenny *et al*., 2009). These failures have prompted to the use of systematic reviews to assess the predictive value of non-clinical experiments. Yet these reviews have not provided a solution to the source problem, i.e., the rationale for evidence generation in safety pharmacology and toxicology.

There are a few additional limitations in the current investigation which have not been previously mentioned. Some mechanisms of action may be too complex or poorly understood to be characterised by PKPD modelling of data arising from a general toxicity study, even if biomarkers have been collected. Multiple downstream markers may present a significant confounder problem which cannot be avoided without additional data. This cannot be easily addressed by the proposed analysis method and becomes a drug development issue. Yet, the use of a parametric approach, and more specifically of hierarchical mixed-effects modelling, to inform experimental design and dose selection represents an important step in the advancement of translational toxicology, both from a biological and statistical perspective. In this context, the design of the present study was not intended to replicate a full toxicology programme for a new chemical entity. Our intent was to show how data obtained from different experimental protocols can be integrated to optimise the design of new experimental protocols as well as to characterise drug exposure and the underlying pharmacological effects in a strict quantitative manner. In fact, we acknowledge that evaluation of the gastrointestinal effects without prior consideration of expected pharmacology and available assays would have been far less informative.

Important lessons and recommendations can however be derived from our study which are applicable other compounds across a wide range of mechanisms of action. First is the need to revisit the dose rationale for the evaluation of safety pharmacology and toxicity. Whilst the concept of safety margin is appealing, it does not address the main issue one faces with regard to the therapeutic use of drugs, which is the understanding of the impact of sustained pharmacological effects associated with the primary target or receptor system on which the drug acts. Currently, doses are selected in experimental protocols, which exceed by far the levels required to achieve maximum pharmacological effects and often even the levels required for maximum receptor binding. Secondly, safety pharmacology and toxicity findings are analysed independently from existing data on the pharmacokinetics and pharmacodynamics of the compound of interest, making the interpretation of findings an empirical process. Pharmacokinetic and pharmacokinetic-pharmacodynamic modelling provides a framework for data integration, enabling a distinction between drug- and system

specific properties (Danhof *et al*., 2008). Of relevance is the possibility to accurately characterise background adverse event rates as well as to establish correlations between primary and secondary adverse events. Lastly, the use of different measures of (systemic) exposure as a proxy for the underlying risk or hazard needs to be revisited. Advancements in imaging, pathology, genetic and genomic research clearly show that overt symptoms and signs arise from drug action as well as from the pharmacological effects induced at cellular and tissue levels. The availability of physiologically-based or semi-mechanistic pharmacokinetic and pharmacokinetic-pharmacodynamic models may provide a stronger basis for the assessment of risk in humans. In addition to gaining further understanding of possible nonlinearity in drug disposition, the possibility to estimating drug-specific parameters, such as the estimates of potency or IC_{80} values, offers measures of the pharmacological activity during the course of treatment, which cannot be intuitively derived from systemic exposure data, such as Cmax or AUC values.

In summary, we have shown the benefits of implementing a model-based approach for the evaluation of the safety profile of naproxen after chronic administration. Furthermore, our investigation illustrates how PKPD relationships can be used to translate pre-clinical findings taking into account interspecies differences in the underlying pharmacological effects.

Conflict of interest

The authors declare that there are no conflicts of interest.

Reference List

Bai JP, Abernethy DR, 2013. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu. Rev. Pharmacol. Toxicol. 53, 451-73.

Bai JP, Fontana RJ, Price ND, Sangar V, 2013. Systems pharmacology modeling: an approach to improving drug safety. Biopharm. Drug Dispos. doi: 10.1002/bdd.1871.

Berger SI, Iyengar R, 2011. Role of systems pharmacology in understanding drug adverse events. Wiley Interdiscip. Rev. Syst. Biol. Med. 3, 129-35

Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ; VIGOR Study Group, 2000. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N. Engl. J. Med. 343, 1520-8.

Bowes J, Brown AJ, Hamon J, Jarolimek W, Sridhar A, Waldron G, Whitebread S, 2012. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. Nat. Rev. Drug Discov. 11, 909-22

Capone ML, Tacconelli S, Sciulli MG, Anzellotti P, Di Francesco L, Merciaro G, Di Gregorio P, Patrignani P, 2007. Human pharmacology of naproxen sodium. J. Pharmacol. Exp. Ther. 322, 453-60.

Chain A, Dubois V, Danhof M, Sturkenboom M, Della Pasqua O, 2013. Identifying the translational gap in the evaluation of drug-induced QTc-interval prolongation. Br. J. Clin. Pharmacol. 76, 708-24.

Comets E, Brendel K, Mentre F, 2008. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. Comput. Methods Programs Biomed. 90, 154-166.

Danhof M, Alvan G, Dahl SG, Kuhlmann J, Paintaud G, 2005. Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. Pharm. Res. 22, 1432-7.

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, 186-91.

Fitzgerald GA , 2007. COX-2 in play at the AHA and the FDA. Trends Pharmacol. Sci. 28,303-7.

Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP, 2005. Evidence-based toxicology: a comprehensive framework for causation. Hum. Exp. Toxicol. 24, 161-201.

Hooijmans CR, Ritskes-Hoitinga M, 2013. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 10, e1001482.

Hooker AC, Staatz CE, Karlsson MO, 2007. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. Pharm. Res. 24, 2187-97.

Huntjens DR, Danhof M, Della Pasqua OE, 2005. Pharmacokinetic-pharmacodynamic correlations and biomarkers in the development of COX-2 inhibitors. Rheumatology (Oxford) 44, 846-59.

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, 396-404.

Josa M, Urizar JP, Rapado J, Dios-Viéitez C, Castañeda-Hernández G, Flores-Murrieta F, Renedo MJ, Trocóniz IF, 2001. Pharmacokinetic/pharmacodynamic modeling of antipyretic and anti-inflammatory effects of naproxen in the rat. J. Pharmacol. Exp. Ther. 297, 198-205.

Karlsson MO, Savic RM, 2007. Diagnosing model diagnostics. Clin. Pharmacol. Ther. 82, 17-20.

Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG, 2009. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One 4: e7824.

Knight A, 2007. Animal experiments scrutinised: systematic reviews demonstrate poor human clinical and toxicological utility. ALTEX 24, 320-5.

Kola I, Landis J, 2004. Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov. 3, 711-5.

Konstam MA, Weir MR, Reicin A, Shapiro D, Sperling RS, Barr E, Gertz BJ, 2001. Cardiovascular thrombotic events in controlled, clinical trials of rofecoxib. Circulation 104, 2280-8.

Lacroix BD, Friberg LE, Karlsson MO, 2012. Evaluation of IPPSE, an alternative method for sequential population PKPD analysis. J. Pharmacokinet. Pharmacodyn. 39, 177-93.

Laine L, Takeuchi K, Tarnawski A, 2008. Gastric mucosal defense and cytoprotection: bench to bedside. Gastroenterology 135, 41-60.

Langman MJ, Jensen DM, Watson DJ, Harper SE, Zhao PL, Quan H, Bolognese JA, Simon TJ, 1999. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA 282, 1929-33.

Lazarou J, Pomeranz BH, Corey PN, 1998. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. JAMA 279, 1200-5.

Lindbom L, Pihlgren P, Jonsson EN, 2005. PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput. Methods Programs Biomed. 79, 241-57.

Naproxen sodium prescriber information (PI 120814). Roche Pharmaceuticals. http://www.roche-australia.com/fmfiles/re7229005/downloads/anti-inflammatory/anaprox-pi.pdf, last accessed on 01/11/13)

Nicolson TJ, Mellor HR, Roberts RR, 2010. Gender differences in drug toxicity. Trends Pharmacol. Sci. 31, 108-14.

Parasuraman S, 2011. Toxicological screening. J. Pharmacol. Pharmacother. 2, 74–79.

Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS, 2007. Comparison of treatment effects between animal experiments and clinical trials: systematic review. Brit. Med. J. 334, 197-200.

Pirmohamed M, Breckenridge AM, Kitteringham NR, Park BK, 1998. Adverse drug reactions. BMJ 361, 1295-8.

Rawlins MD, Thompson JW, 1991. Mechanisms of adverse drug reactions. *In*: Davies DM, ed. Textbook of adverse drug reactions. Oxford: Oxford University Press,18-45.

R Development Core Team, 2012. A language and Environment for Statistical Computing, R foundation for Statistical computing, Vienna Austria.

Rolan P, 1997. The contribution of clinical pharmacology surrogates and models to drug development--a critical appraisal. Br. J. Clin. Pharmacol. 44, 219-25.

Runkel R, Forchielli E, Sevelius H, Chaplin M, Segre E, 1974. Nonlinear plasma level response to high doses of naproxen. Clin. Pharmacol. Ther. 15, 261-6.

Satterwhite JH, Boudinot FD, 1988. High-performance liquid chromatographic determination of ketoprofen and naproxen in rat plasma. J. Chromatogr. 431, 444-9.

Scheuren N, Jacobs M, Ertl G, Schorb W, 2002. Cyclooxygenase-2 in myocardium stimulation by angiotensin-II in cultured cardiac fibroblasts and role at acute myocardial infarction. J. Mol. Cell. Cardiol. 34,29-37.

Urushidani T, Okabe S, Takeuchi K, Takagi K, 1978. Strain differences in aspirin-induced gastric ulceration in rats. Jpn. J. Pharmacol. 28, 569-78.

van Vliet E, 2011. Current standing and future prospects for the technologies proposed to transform toxicity testing in the 21st century. ALTEX 28, 17-44.

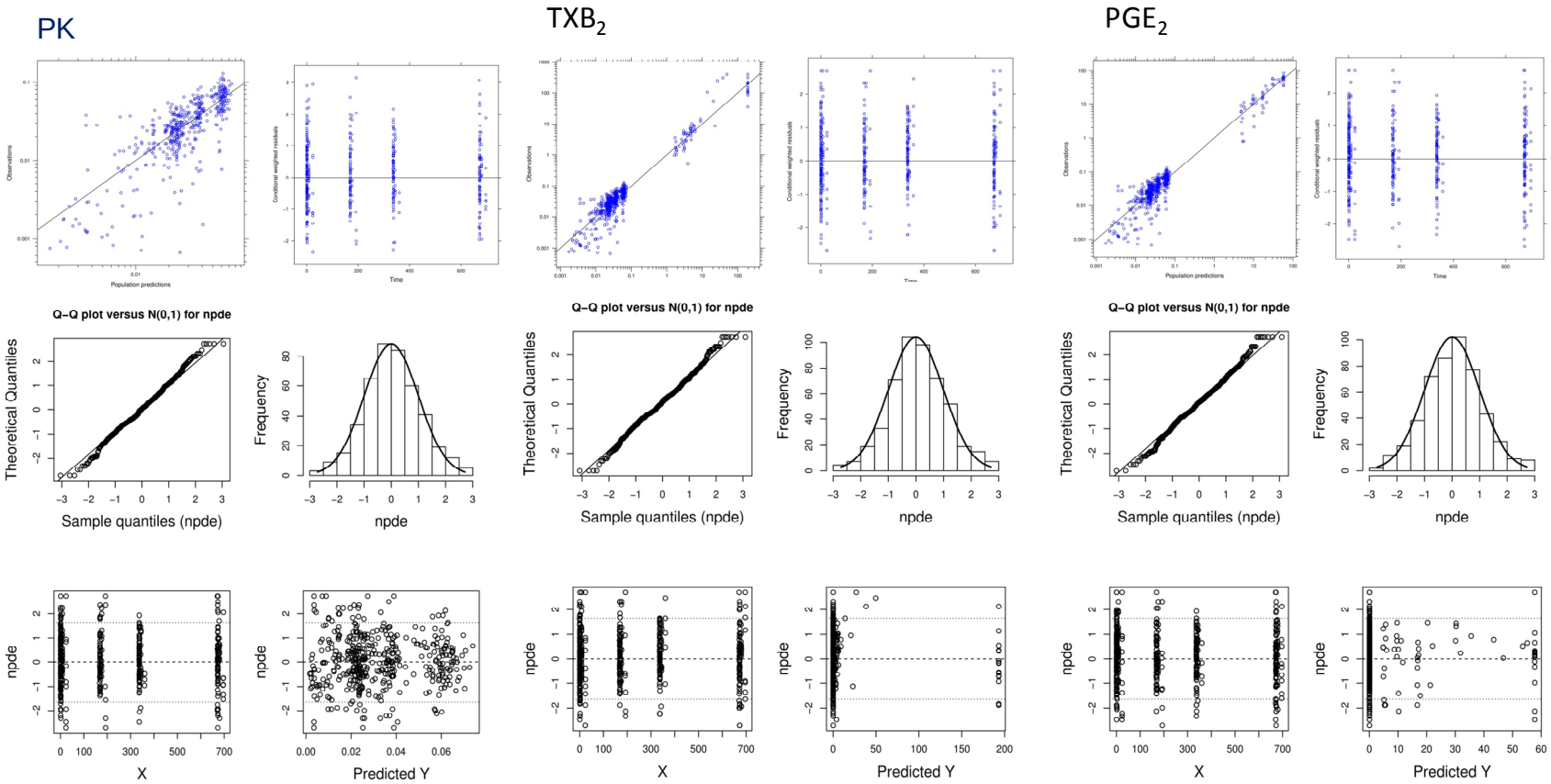
Vane JR, 1971. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. Nat. New Biol.231, 232-5.

Zhang L, Beal SL, Sheiner LB, 2003. Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J. Pharmacokinet. Pharmacodyn. 30, 387-404.

Zhang L, Beal SL, Sheinerz LB, 2003. Simultaneous vs. sequential analysis for population PK/PD data II: robustness of methods. J. Pharmacokinet. Pharmacodyn. 30, 405-16.

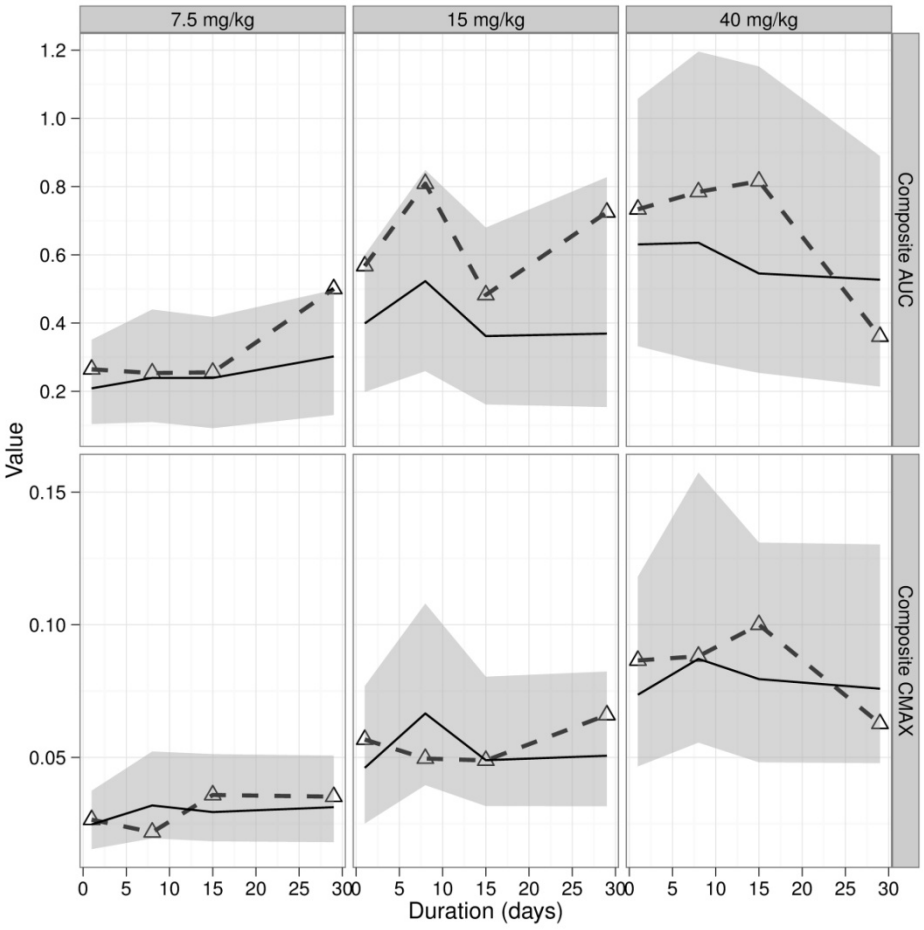
Supplemental material

Figure 1S: Population pharmacokinetics and pharmacokinetic-pharmacodynamic modelling of TXB_2 and PGE_2 inhibition by naproxen. For each biomarker, goodness-of-fit plots show observed vs. population predicted concentrations (upper left) and conditional weighted residuals *vs.* time (upper right). Mid and lower panels depict the NPDE summary, including QQ plot and histogram of the normalised discrepancy between observed and predicted values. X denotes the independent variable, i.e., time. Samples are clustered around 0, 1, 2, and 4 weeks.



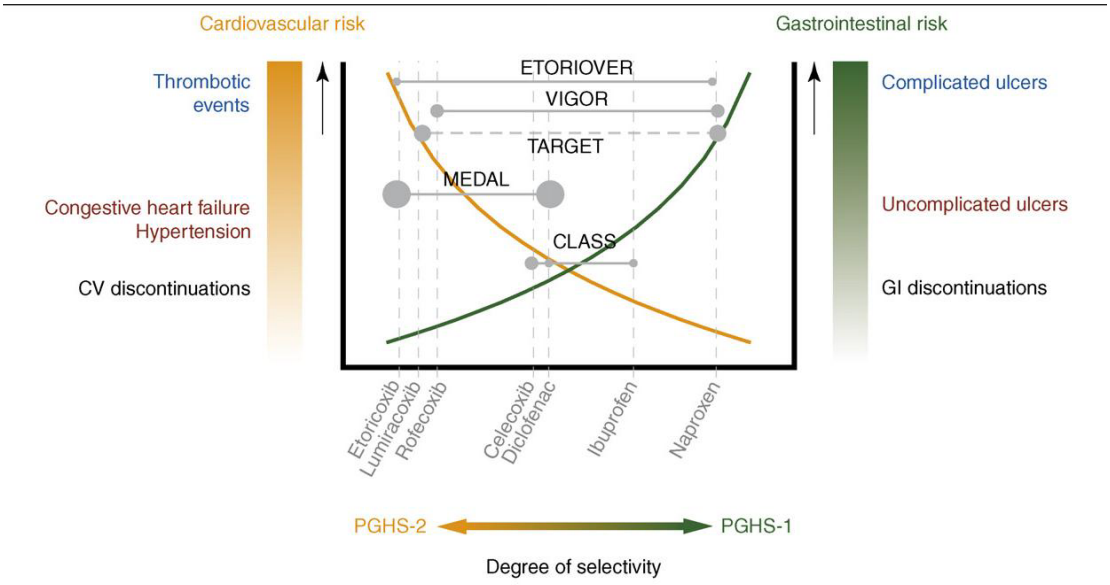
Supplemental material

Figure 2S. Predictive check for derived measures of exposure, as determined by non-compartmental analysis. Observed values (triangles and dotted line) are shown to occur within the 95% prediction intervals. Solid line depicts the predicted median, whereas shaded region indicates the 2.5th and 97.5th percentiles. Note that predicted and observed estimates obtained by non-compartmental analysis differ from model-based predictions shown in Figure 6.



Supplemental material

Figure 3S: Schematic representation of COX-2 selectivity with incidence of cardiovascular (CV) and gastrointestinal (GI) risk. Increasing degrees of selectivity for COX-2 are associated with augmented CV risk, whereas increasing degrees of selectivity for COX-1 are associated with augmented GI risk. The relative size of the circles indicates the variation in sample sizes among the trials. The average selectivity for each drug is presented ranging from drugs that are highly selective for inhibition of COX-2 (e.g., etoricoxib) to those that are more selective for COX-1 (e.g., naproxen). Given the interindividual variability in response to these drugs, selectivity is a continuous variable at the individual level. ETORIOVER, VIGOR, MEDAL, TARGET and CLASS refer to the overview of Phase II and III trials with COX inhibitors. (Reprinted with permission from Fitzgerald, 2007)



CHAPTER 8

Model-based prediction of the acute and long-term safety profile of naproxen in rats

Tarjinder Sahota, Ian Sanderson, Meindert Danhof and Oscar Della Pasqua

Submitted for publication (British Journal of Clinical Pharmacology)

Abstract

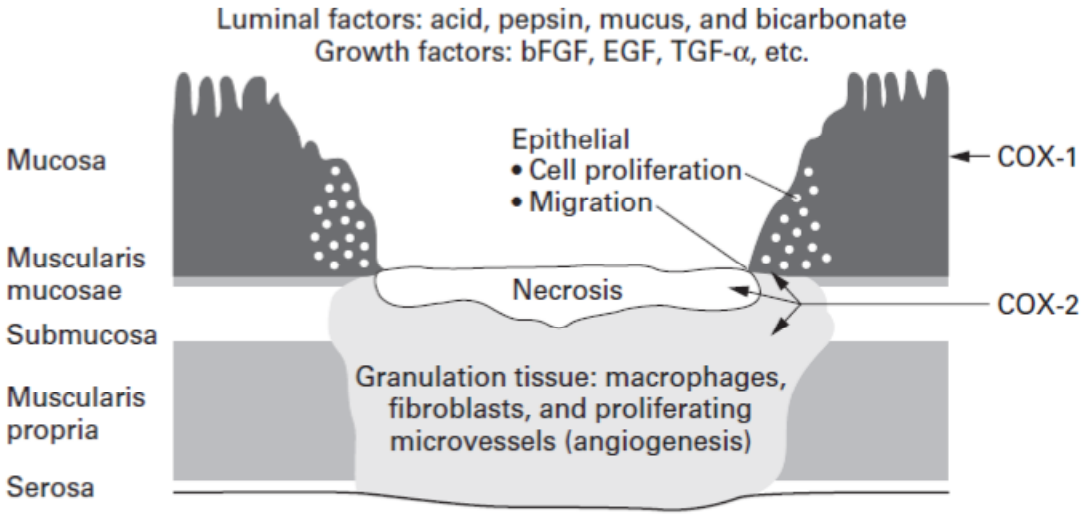
Despite increasing relevance of the use of biomarkers as predictors of drug effects, traditional toxicology protocols continue to rely on the experimental evidence of the link between adverse events (AEs) in animals and estimates of systemic drug exposure (e.g., Cmax and AUC). Furthermore, biomarkers may facilitate the translation of findings from animals to humans. Thus, combined with a model-based approach, biomarker data has the potential to predict long term pharmacodynamic effects arising from prolonged drug exposure. Here, we use naproxen as a paradigm drug to explore the feasibility of a biomarker-guided approach for the prediction of long term AEs in humans. An experimental toxicology protocol was set up for the evaluation of effects of naproxen in rats, in which four doses were tested (7.5, 15, 40 and 80mg/kg). In addition to AE monitoring and histology, sparse blood sampling for the assessment of exposure, thromboxane B2 and prostaglandin E2 were also collected. Nonlinear mixed effects modelling was used to analyse the data and identify covariate factors on the incidence and severity of AEs. Modelling results show that besides drug exposure, maximum PGE_2 inhibition and treatment duration are also predictors of GI ulceration. Although PGE_2 levels were clearly linked to the incidence rates, it appears that ulceration severity is better predicted by measures of drug exposure. These results show that the use of a model-based approach provides the opportunity to integrate pharmacokinetics, pharmacodynamics and toxicity data, enabling optimisation of the design, analysis and interpretation of toxicology experiments.

Introduction

A key purpose of preclinical general toxicity and safety pharmacology studies is to support the safe dose selection in humans. In particular, the need to understand the risks associated with long term drug exposure falls within the remit of these two disciplines. Preclinical toxicity data consists of a mixture of acute, mid-term and chronic toxicity data, however, identification of long term risks often happens in Phase IV post marketing surveillance. Earlier identification of potential risks would enable the use of evidence-based risk mitigation strategies. However, understanding of time-dependent physiological changes arising from repeated exposure to a drug is required to identify and assess risks associated with long term use of medicinal products. Such an objective may be hampered by the use of empirical experimental protocols, as they render the extrapolation of findings across species and across molecules rather difficult, preventing accurate translation of the pharmacological properties to man (Bai et al. 2013, Della Pasqua, 2013). Among other things, differences in sensitivity and target organ specificity continue to represent drawbacks for most clinical pathology parameters traditionally used for monitoring organ integrity both during preclinical toxicological assessment and clinical safety testing (Connelly et al., 1991). Clearly, efforts are required to ensure the availability of tissue- and mechanism-specific data for accurate interpretation of acute and long term safety findings. Over the last few years, several novel toxicity biomarkers have emerged as sensitive tools for detection, monitoring, quantification and prediction of safety and toxicity (O'Brien, 2008, Xie et al., 2013). Nevertheless, little attention has been given to the possibility of evaluating safety and toxicity using a mechanism-based approach whereby adverse events are assessed taking into account the underlying pharmacokinetic-pharmacodynamic (PKPD) properties of the molecule (McGonigle et al., 2013). In this context biomarkers can be of great relevance for drug discovery and development as they offer the possibility to discriminate between acute and chronic term treatment effects.

In the current investigation we show therefore how pharmacokinetic-pharmacodynamic (PKPD) modelling can be used to unravel the relationship between chronic drug exposure, pharmacodynamic effects and overt symptoms and signs. The concept is illustrated by the correlation between naproxen concentrations, inhibition of prostaglandin E_2 (PGE_2) and thromboxane B_2 (TXB_2) and gastric ulceration in rats. Non-selective NSAIDs, such as naproxen, act by blocking cyclo-oxygenase (COX), which catalyses the rate-limiting step in the formation of prostanoids from arachidonic acid (Chakraborti et al., 2010). Continuous COX-1 inhibition following prolonged administration of non-selective COX inhibitors is known to induce gastrointestinal adverse effects, especially ulceration and haemorrhagic bleeding. Unfortunately, at present the dose selection of COX inhibitors disregards whether maximum, long-lasting blockade of either enzyme systems is strictly required for anti-inflammatory, analgesic response and how its pharmacology relates the observed adverse events (Huntjens et al., 2006). These considerations become essential when evaluating the side effects associated with long term use of COX inhibitors, which include gastric and cardiac adverse events. From a pharmacological perspective, various investigations have shown that both COX-1 and COX-2 mRNA and protein are either constitutive or inducible in specific areas of the stomach of animals and humans (Morita, 2002, Coruzzi et al., 2007) (Figure 1). Hence, it can be anticipated that some balance between the activity of either isoform may be required to ensure normal physiological function. On the other hand, COX-1-deficient mice show no evidence of spontaneous gastric injury despite the absence of COX-1-derived prostaglandins (Langenbach et al., 1999). Yet, the administration of NSAIDs-induced gastric damage can be invariably related to COX-2 inhibition (Loftin et al., 2002, Wallace, 2008, Takeuchi, 2012). In a previous investigation, we have shown how these safety biomarkers can be used in conjunction with general toxicity protocols to predict the safety window in humans using an empirically derived safety threshold; the no-observed adverse- effect level (NOAEL) (Sahota et al., 2014). The NOAEL approach has many statistical and experimental limitations which have been documented elsewhere (Dorato et al.,2005, Sahota et al., 2014). Most importantly, by dichotomising the exposure-risk relationship using a threshold, the NOAEL approach precludes quantitative risk assessment. Here we demonstrate that the availability of a mechanism-based PKPD model together with the application of probabilistic modelling

to the adverse event data not only provides a quantitative rationale for determining effective and safe dosages following chronic treatment in humans, it also enables effective data integration, offering a stronger basis for extrapolating pre-clinical findings into humans (Rohatagi et al., 2007). Moreover, predictive modelling enables testing of different parameterisations of biomarker response and drug exposure to enable exploration of causal factors driving risk. This is ultimately provides a flexible evidence-based framework for risk management and risk mitigation strategies in humans.



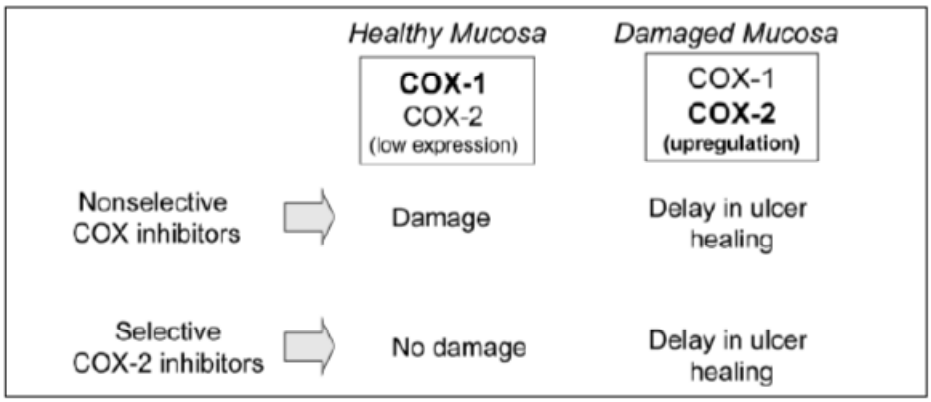


Figure 1: (Upper panel) Diagrammatic presentation of ulcer healing and factors affecting ulcer healing. In the intact mucosa, cyclooxygenase 1 (COX-1) is the predominant COX isoform in the gastrointestinal tract. In contrast, during wound healing, expression of cyclooxygenase 2 (COX-2),

rather than COX-1, is strongly increased in the repair zone. (Lower panel) Gastric effects of non-selective and COX-2 selective NSAIDs in normal or damaged gastric mucosa. The different effects of non-selective or selective COX-2 inhibition are explained by differences in COX-2 tissue expression (printed with permission from Halter et al., 2001, Coruzzi et al., 2007)

In contrast to traditional safety extrapolation methods such as allometric scaling, which relies primarily on the estimation of safe exposures based on the human equivalent dose (HED), a mechanism-based approach can account for the variability in drug elimination or differences with respect to physiological, biochemical (e.g., expression of drug metabolizing enzymes), and other time-variant factors (e.g., disease). These time-variant factors may become more important as clinical trials move from acute to chronic interventions in patients (in Phase II and III).

In spite of known interspecies differences exist in GI-related morbidity, we hypothesise that the characterisation of the relationship between markers of COX inhibition and adverse events enables the prediction of safety windows for chronic treatment with selective and non-selective COX inhibitors. In fact, various studies provide further evidence of a multistage pathogenic mechanism for NSAID enteropathy by which the topical action of NSAIDs may initiate mucosal damage, which is then converted to macroscopic damage by the concomitant inhibition of COX, with decreased mucosal prostaglandins, presumably because of their effect on the microvasculature (Fornai et al., 2014).

Methods

The present investigation is based on a previously published general toxicity study in rats by Sahota et al. (2014), with the non-selective COX inhibitor naproxen. Detailed description of the study design, strain of rats, sample collection and analysis and PKPD modelling details can be found in Sahota et al. (2014).

Summary of study design: Three different treatment durations were investigated (1 week, 2 weeks and 4 weeks). Rats were given daily doses of naproxen by oral gavage. There were four cohorts per treatment duration receiving 0, 15, 40 and 80 mg/kg/day doses. Satellite

animals received identical doses to toxicology groups and were used for plasma drug concentration (PK) and biomarker (PD) data measures, TXB_2 and PGE_2. A optimised composite sampling scheme was used and sampling too place on days 1, 7, 14 and 28. Details regarding sample analysis can be found in Sahota et al. (2014). Endpoints in toxicology groups included adverse events, including GI histology and terminal PK and PD measurements. An overview of the study protocol is depicted in Figure 2.

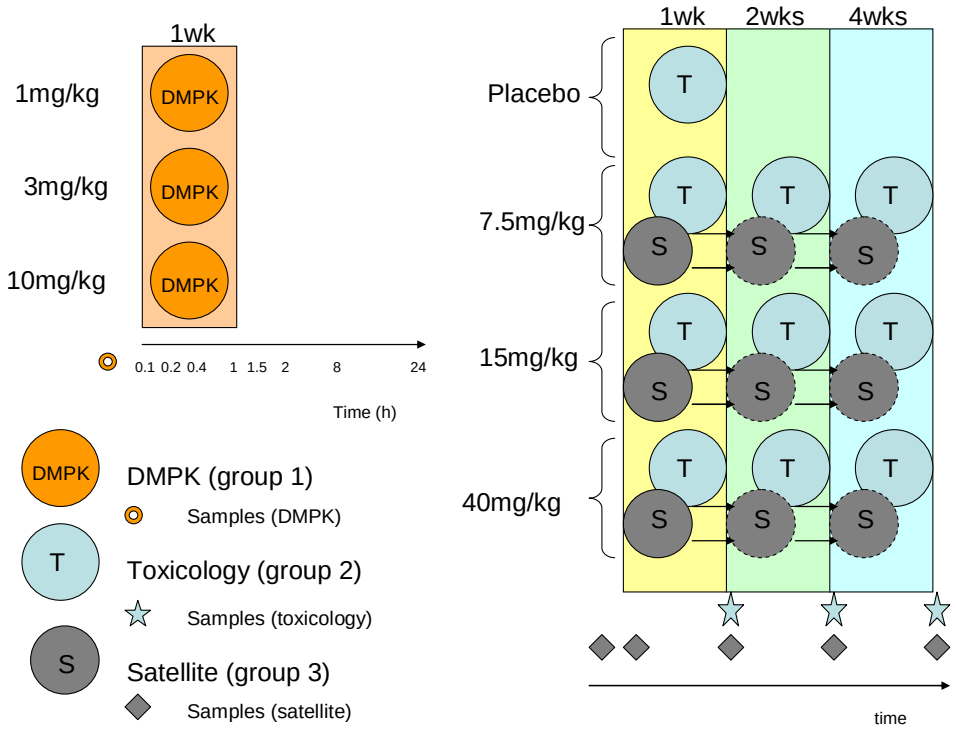


Figure 2 : Schematic representation of general toxicity study.

Histology: Histological evaluation of the stomach was performed to establish a correlation between acute and long term adverse events. After euthanasia, stomachs were removed immediately and were cut open along the greater curvature and washed with warm saline. The inner surface was photographed to allow the measurement of the area covered by hemorrhagic ulceration. The area of ulceration was determined under a dissecting microscope. Gastric ulceration was measured as percentage stomach surface area affected

by ulceration the software Image J version 1.43 (Abramoff et al., 2004) was used for calculating ulcer area and total stomach surface area. The person who performed the ulceration measurement was blinded as to animal ID and treatment group.

PKPD model: The PK and PD data of naproxen were assessed by nonlinear mixed effects modelling, as implemented in NONMEM version 7.2.0. The pharmacokinetics of naproxen in plasma were best described by a one-compartment model with first order absorption, first order elimination and nonlinear dose-dependent bioavailability. Weight was included as a significant covariate on clearance and volume of distribution. The PK/PD models for both biomarkers, PGE_2 and TXB_2 were characterised by direct sigmoid I_{MAX} models. Parameter values, precision estimates and goodness of fit diagnostics are described in Sahota et al. (in press).

Data analysis

Final model parameters describing the gastric ulceration incidence and percentage gastric area affected were performed via the numerical integration routine ADVAN13 in NONMEM 7.2.0 using FOCE with Laplacian estimation. Convergence was determined by successful minimisation and covariance step. All fitting procedures were performed on a computer (AMD-Athlon XP-M 3000+) running under Windows XP with a FORTRAN compiler (Compaq Visual Fortran, version 6.1). Data processing, management and graphical display were performed in R (R Development Core Team, 2012). Model diagnostics and validation were performed according to graphical and statistical criteria. Goodness-of-fit plots, including observed (OBS) *versus* individual prediction (IPRED), OBS *versus* population prediction (PRED), conditional weighted residuals (CWRES) *versus* time and CWRES *versus* OBS were used for diagnostic purposes (104).

Given the purpose of the study in discriminating between acute and long term effects of naproxen, different parameterisations were considered for describing drug effects during the course of treatment. Model-based exposure and biomarker levels from the final PKPD

model were calculated for each individual animal using post-hoc empirical Bayes estimates (using MAXEVAL=0). Details of the calculation methods are described in Table 1.

Table 1: Calculation of biomarker response and exposure variables. Individual predicted naproxen concentrations and biomarker levels are denoted by $C_p(t)$ and $BC_p(t)$, respectively.

Parameter name	**Symbol**	**Calculation**
Area under drug concentration vs. time profile	AUC	$\int_{t-24}^{t} C_p dt$
Area above biomarker concentration vs. time profile	AOC	$BC_p(0) - \int_{t-24}^{t} BC_p dt$
Time under threshold (80% inhibition)	TUT	$\int_{0}^{t} 1_{BC_p<.2BC_p(0)} dt$
Cumulative area under drug concentration vs. time profile	CAUC	$\int_{0}^{t} C_p dt$
Cumulative area over biomarker concentration vs. time profile	CAOC	$BC_p(0) - \int_{0}^{t} BC_p dt$
Maximum drug concentration over 24 hour period	CMAX	$\max\left(\{C_p(s): t-24 < s < t\}\right)$
Maximum biomarker inhibition over 24 hour period	CMIN	$\min\left(\{BC_p(s): t-24 < s < t\}\right)$

Ulceration model: Since each histological examination was performed once per animal, no between-subject variability could be estimated. All random effects are therefore accounted for with the residual variability structure. Nevertheless, both the incidence and severity of ulceration were considered during modelling. Incidence was modelled as the probability of occurrence of ulceration, $U_{i,j}$ at the time of sacrifice, t_j, and severity was modelled as $PER_{pred,i,j}$, the % gastric surface area affected, when ulceration is observable at the time of assessment .

A logit transformation was used to describe the incidence of stomach ulcers. The general equation describing the incidence of ulcers is given by:

$$P(U_{i,j} = 1) = \frac{EXP(\theta_1 + \sum_k \theta_k * COV_{i,j,k})}{1 + EXP(\theta_1 + \sum_k \theta_k * COV_{i,j,k})} \qquad \text{(equation 5)}$$

where $P(U_{i,j})$ represents the probability of the presence of ulceration in individual *i* at time t_j. $COV_{i,j,k}$ is the K^{th} covariate value for individual *i* and time t_j. θ_1 is a parameter governing the baseline logit probability and θ_k is the coefficient of the K^{th} covariate relationship.

For technical reasons, the severity of ulceration, i.e., percentage gastric surface area affected was log-transformed. The basic model did not include any covariates on response. Two fixed effect model parameters were used, θ_1 and θ_2.

$$P(U_{i,j} = 1) = \frac{EXP(\theta_1)}{1 + EXP(\theta_1)} \qquad \text{(equation 6)}$$

$$\mathrm{PER}_{\mathrm{pred,i,j}} = \begin{cases} 0, \ if \ U_{i,j} = 0 \\ \theta_2, \ if \ U_{i,j} = 1 \end{cases} \qquad \text{(equation 7)}$$

$$\log\left(\mathrm{PER}_{\mathrm{obs,i,j}}\right) = \log\left(\mathrm{PER}_{\mathrm{pred,i,j}}\right) + \varepsilon_{\mathrm{i,j}} \qquad \text{(equation 8)}$$

where $PER_{obs,i,j}$ and $PER_{pred,i,j}$ represent observed and predicted percentage ulceration, respectively, in individual *i* at time t_j. $\varepsilon_{i,j}$ is the random effect describing residual variability with mean 0 and estimated standard deviation.

Covariates: To explore the relationship between drug exposure, biomarkers and adverse events over the course of treatment, different secondary pharmacokinetic and pharmacodynamic parameters expressing systemic exposure and pharmacological activity were explored as covariates on the logistic model parameters using the stepwise covariate method (SCM) in PsN (107). Potential influential factors on the incidence of ulcers included body weight, age. Time measured in days (DAY) was also tested used as a covariate as surrogate for time-dependent effects such as healing, tolerance or other mechanisms influencing ulceration incidence and/or severity. For the percentage gastric area affected, PER, the specification of the covariate relationship was based on the diagnostic plots of the basic model. Linear, exponential and hyperbolic (sigmoid Emax) functions were considered during covariate model building. A hockey-stick function was also tested to describe toxicity only manifesting above a threshold exposure/biomarker level. The linear relationship was characterised by:

$$\mathrm{PER_{pred,i,j}} = \begin{cases} 0, \ if \ U_{i,j} = 0 \\ \theta_1 + \left(\theta_{SLOPE} * (COV - median(COV))\right), \ if \ U_{i,j} = 1 \end{cases}$$

where θ_2 is population prediction and θ_{slope} is the slope of relationship between parameter and (centred) covariate

The exponential relationship was similarly characterised by:

$$\mathrm{PER_{pred,i,j}} = \begin{cases} 0, \ if \ U_{i,j} = 0 \\ \theta_1 * \exp\left(\theta_{SLOPE} * (COV - median(COV))\right), \ if \ U_{i,j} = 1 \end{cases}$$

Since data were sparse and maximum effect may not have been reached, maximum effect was fixed to 100% during the evaluation of the sigmoid Emax function.

$$\mathrm{PER}_{\mathrm{pred,i,j}} = \begin{cases} 0, & if\ U_{i,j} = 0 \\ \theta_2 * \frac{100 * COV^{\gamma}}{COV_{50}{}^{\gamma} + COV^{\gamma}}, & if\ U_{i,j} = 1 \end{cases}$$

where γ is an estimated Hill coefficient.

Covariate relationships were centred by the median value of the covariate so that in this case $\mathrm{PER}_{\mathrm{pred,i,j}} = \theta_2$, as in the basic model.

$$COV_{50}{}^{\gamma} = \frac{(100-\theta) median(COV)^{\gamma}}{\theta}$$

The hockey stick function was implemented according to the following function:

$$\mathrm{PER}_{\mathrm{pred,i,j}} = \begin{cases} \theta_1, & if\ U_{i,j} = 0 \text{ and COV} < \theta_{THRESH} \\ \theta_1 + \left(\theta_{SLOPE} * (COV - \theta_{THRESH})\right), & if\ U_{i,j} = 1 \text{ and COV} \geq \theta_{THRESH} \end{cases}$$

where θ_{THRESH} is the threshold value of the covariate where toxicity begins.

Covariates were incorporated into the model by stepwise forward inclusion. A significance level of p<0.01 was used for inclusion, which represented a drop of least 6.63 units in the objective function for each additional parameter. A final evaluation of the statistical significance of all factors identified during the previous step was performed by subtracting each covariate individually (backward elimination). The final structural model (i.e., fixed effects model) included only those covariates whose subtraction resulted in a decrease of at least 6.63 units in the objective function (p<0.01). Finally, to investigate model uncertainty a bootstrap SCM was performed to estimate covariate inclusion probabilities.

Model validation: The performance of the ulceration models were assessed by numerical and visual predictive checks. To that purpose, 1000 data sets were simulated with the final model parameter estimates. The mean and the 95 % confidence intervals were calculated for the incidence and percentage gastric area affected. Validation procedures also included normalised prediction distribution errors (NPDE), which are based on the assumption that

the normalised (de-correlated) prediction distribution errors (discrepancies) are normally distributed (Comets *et al*., 2008). One hundred datasets were simulated using the final model, which was then tested for the assumption of normality of the prediction distribution errors.

Results

In total, 80 histological examinations were performed on ontoxicology group animals. These revealed gastric ulceration in all dose levels; therefore no NOAEL could be obtained for any of the treatment durations (Figure 3). All animals receiving 80 mg/kg in the 1 week cohorts suffered from moderate weight loss. The animals in this cohort were immediately sacrificed, and due to ethical reasons this dose level was discontinued. Different dose levels of naproxen were considered for the two-week and one-month cohorts, which received lower doses (7.5, 15, and 40mg/kg). Histological examination and terminal blood samples were performed on these animals. No other adverse events were reported.

Logistic models for gastric ulcerations

Empirical analysis of this data revealed some peculiarities in data where there was no significant dose-response until week 4. In fact, the data revealed a possible negative dose-response relationship before week 4. Moreover, the incidence of ulcers was much lower in the week 4 cohort than in shorter treatment durations. Furthermore, exploratory evaluation of the relationship between naproxen exposure and biomarker levels instead of dose did not provide further evidence of an apparent relationship. Physiologically, interpreting such data is difficult. However, it is plausible that the ulcerative effect is acute and diminishes with sustained long term exposure. After an initial attempt to describe the data without the use of covariates, a clear model misspecification was observed. As shown in Figure 4, the apparent negative dose-toxicity relationship for week 1 and week 2 was not replicated by final model predictions.

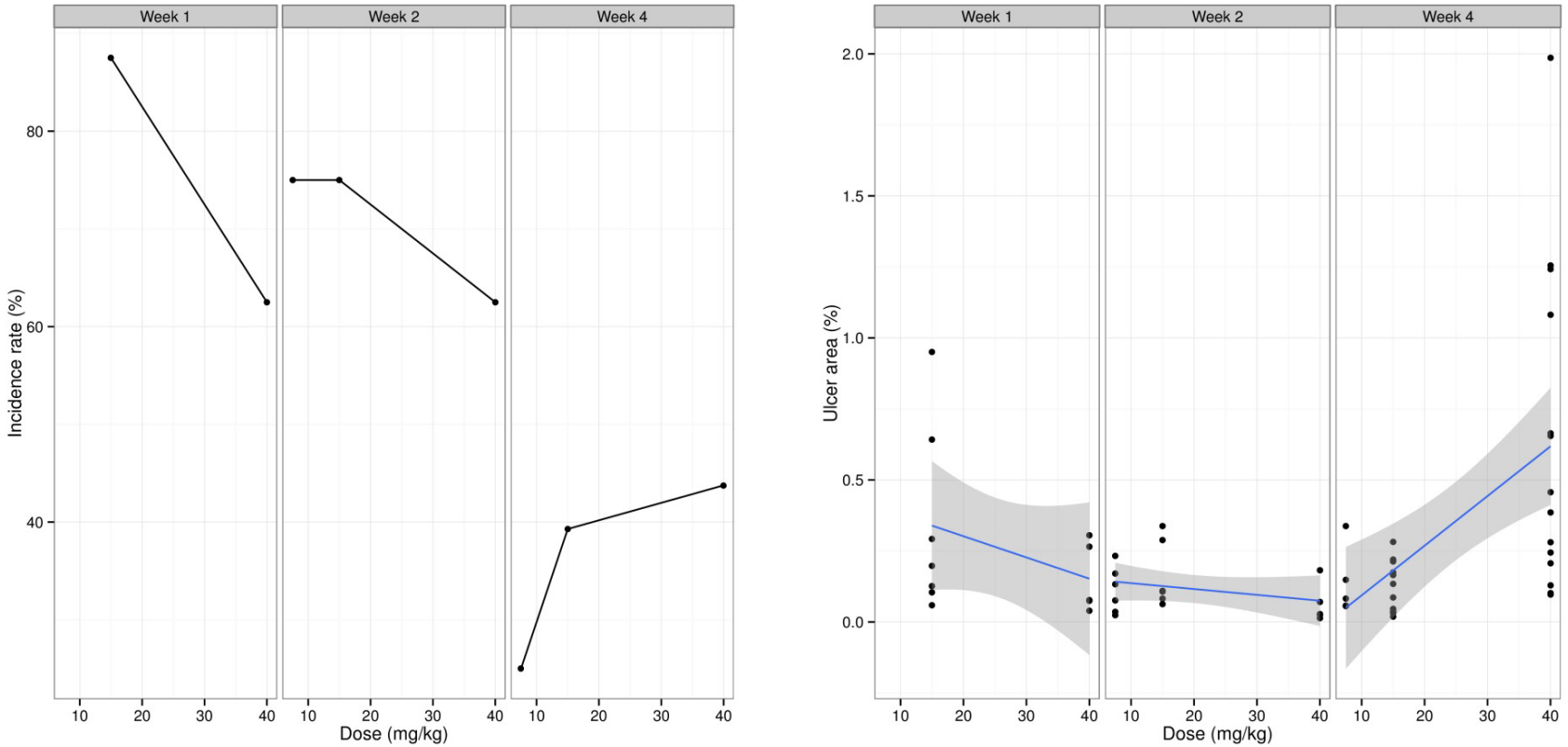


Figure 3. Plots of the observed ulcer incidence (left) and severity (right). Dots in the left plot show observed percentage of total animals in each cohort manifesting GI toxicity. Ulcer severity is measured as % of stomach area affected by ulceration. Given the time-dependent effect on the accuracy of this measure, the uncertainty (shaded area) of the regression line is also shown together with the data.

Stepwise model building combined with bootstrap methods showed a possible negative correlation between the incidence of ulceration and treatment duration, indicating an acute effect dissipating over time (see Figure 5). By contrast, upon incorporation of this time-dependent effect into the final model, the overall fit improved (Figure 6). CMAX and AUC were shown not be the primary drivers of toxicity, although these parameters may be indirectly correlated with risk.

Our attempt to establish a relationship between drug exposure/ biomarker levels and adverse events revealed clear differences in the sensitivity of explanatory variables used to describe the incidence of ulcers and ulceration severity. Out of tested relationships, the maximum inhibition of PGE_2 was the best predictor of adverse event incidence, with the bootstrap SCM showing low model uncertainty. On the other hand, cumulative TXB_2 inhibition was found to be the best explanatory variable for the severity of ulceration (Figure 7). Other physiologically plausible explanatory factors, such as maximum PGE_2 inhibition or DAY (treatment duration) were found to be fraught with significant model uncertainty. The model parameters for the final model are summarised in Table 2 and figure 5.

Table 2: Logistic model parameters. SE = standard error.

Relationship	Description	Parameter value	SE
LOGIT	Typical value: Logit ($P(U_{i,j} = 1)$)	-0.226	0.305
LOGIT – $IMAX_{PGE}$	Covariate: Logit (additive)	0.042	64.8%
LOGIT – DAY	Covariate: Logit (additive)	-0.066	0.022
PER	Typical value: % gastric area affected	0.21%	17.2%
PER – $CAUC_{TXB}$	Covariate: Hockey stick	Threshold= 94.3% Slope = 149.9	13% 74%
Residual variability (%)	Proportional error	78%	16%

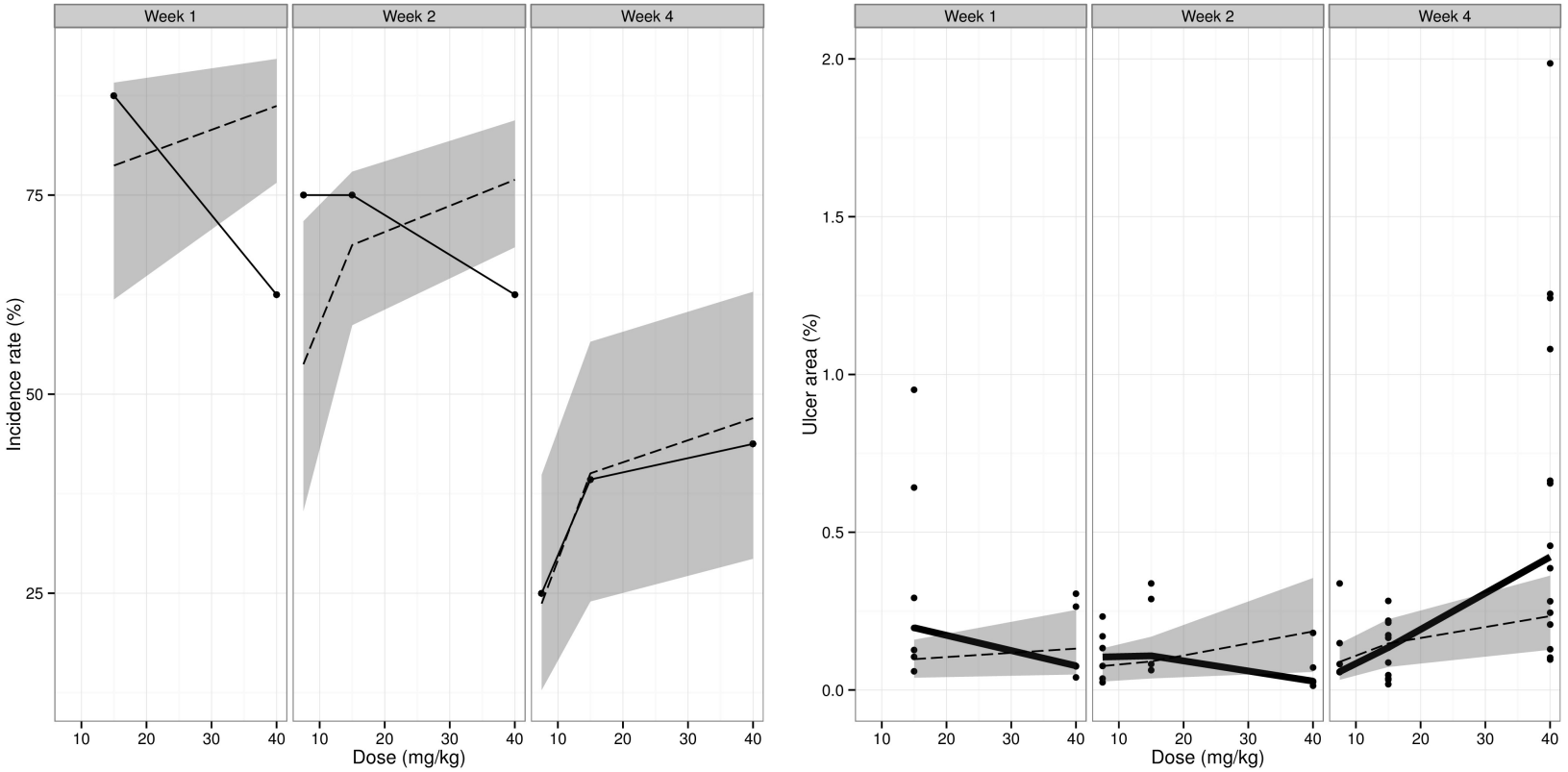


Figure 4: Plots of the observed and predicted ulcer incidence (left) and severity (right). Dots in the left plot show observed percentage of total animals in each cohort manifesting GI toxicity. Ulcer severity is measured as % of stomach area affected by ulceration. The shaded area is depicts the 95% uncertainty in population prediction of the model (dotted lines depict the 50th percentile). The model is unable to describe the apparent negative dose-response trend observed after short treatment durations (i.e., 1 and 2 weeks), indicating that time-independent, long-lasting or irreversible processes may appear only after long term treatment.

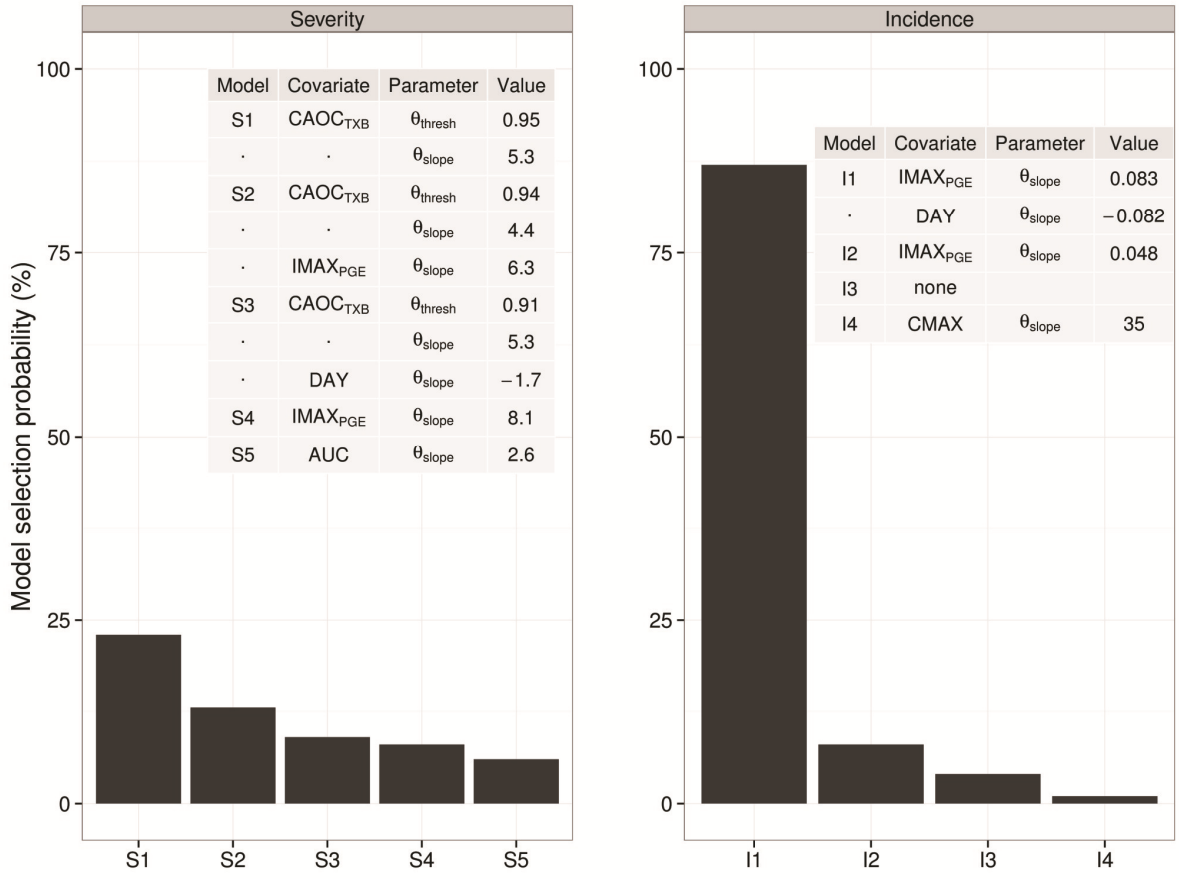


Figure 5: Model specification uncertainty. Results of boostrap SCM. Bars indicate model selection probability as determined by the bootstrap SCM ordered from most to least probable. Only top 5 most probable displayed. A wider, flatter distribution reflects high model specification uncertainty, i.e., two or more different models may be indistinguishable. The overlaid table in the insert shows numerical details of model selection probability.

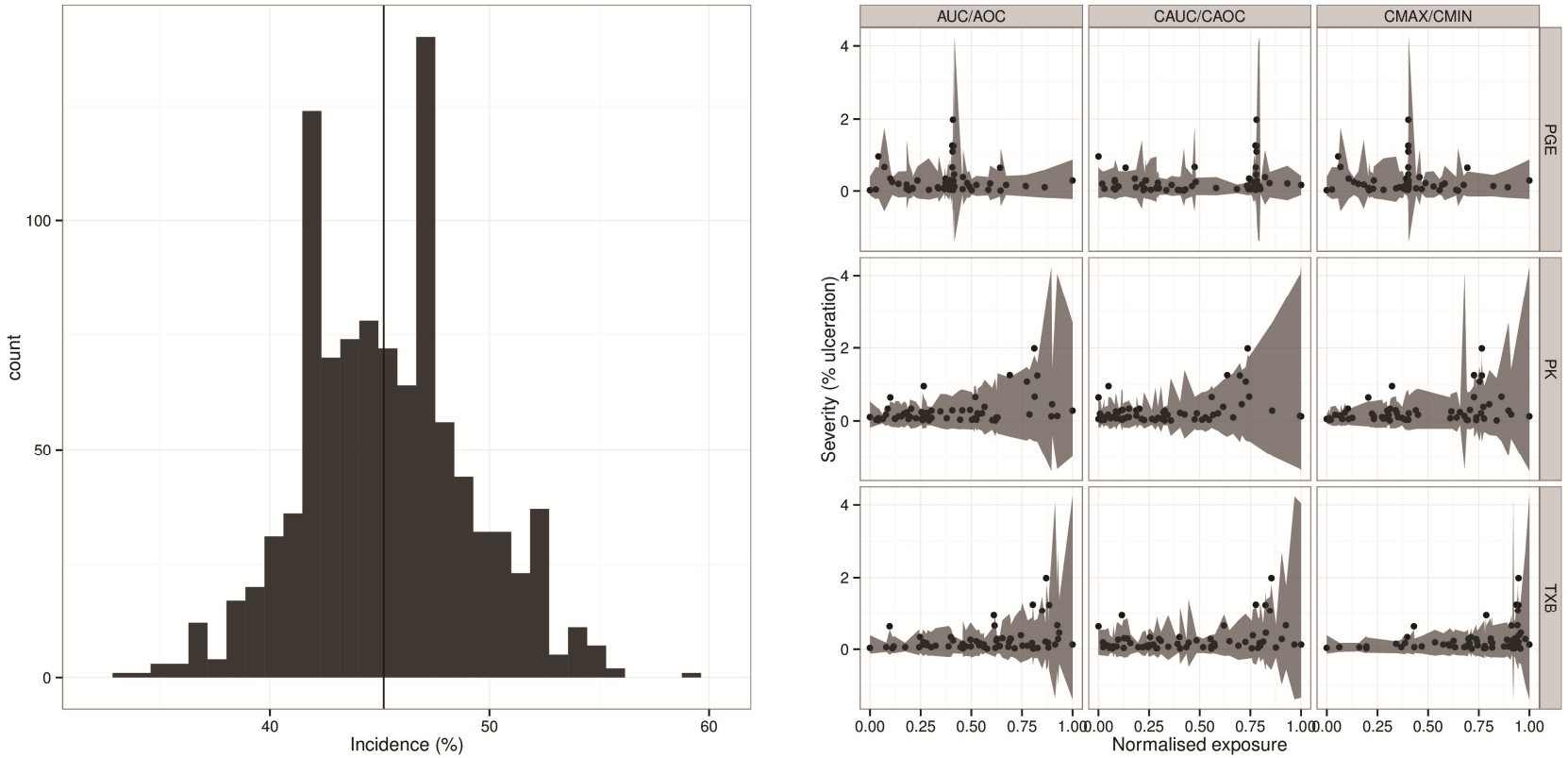


Figure 6: Visual predictive checks. Left panel depicts the prediction distribution for the incidence of ulcers, whereas the solid line indicates the observed percentage of total animals manifesting GI toxicity. Right panel shows ulceration severity against predicted normalised exposure for different variables of interest (AUC= area under the concentration vs. time curve; AOC= area above biomarker concentration vs. time profile; CAUC= cumulative area under the concentration vs. time curve; CAOC= cumulative area over biomarker concentration vs. time profile; CMAX = maximum drug concentration over the period of 24h; CMIN= maximum biomarker inhibition over the period of 24h). The shaded area represents the 95% prediction interval; points represent the actual data. PGE = prostaglandin E_2; PK= naproxen concentrations; TXB= thromboxane E_2.

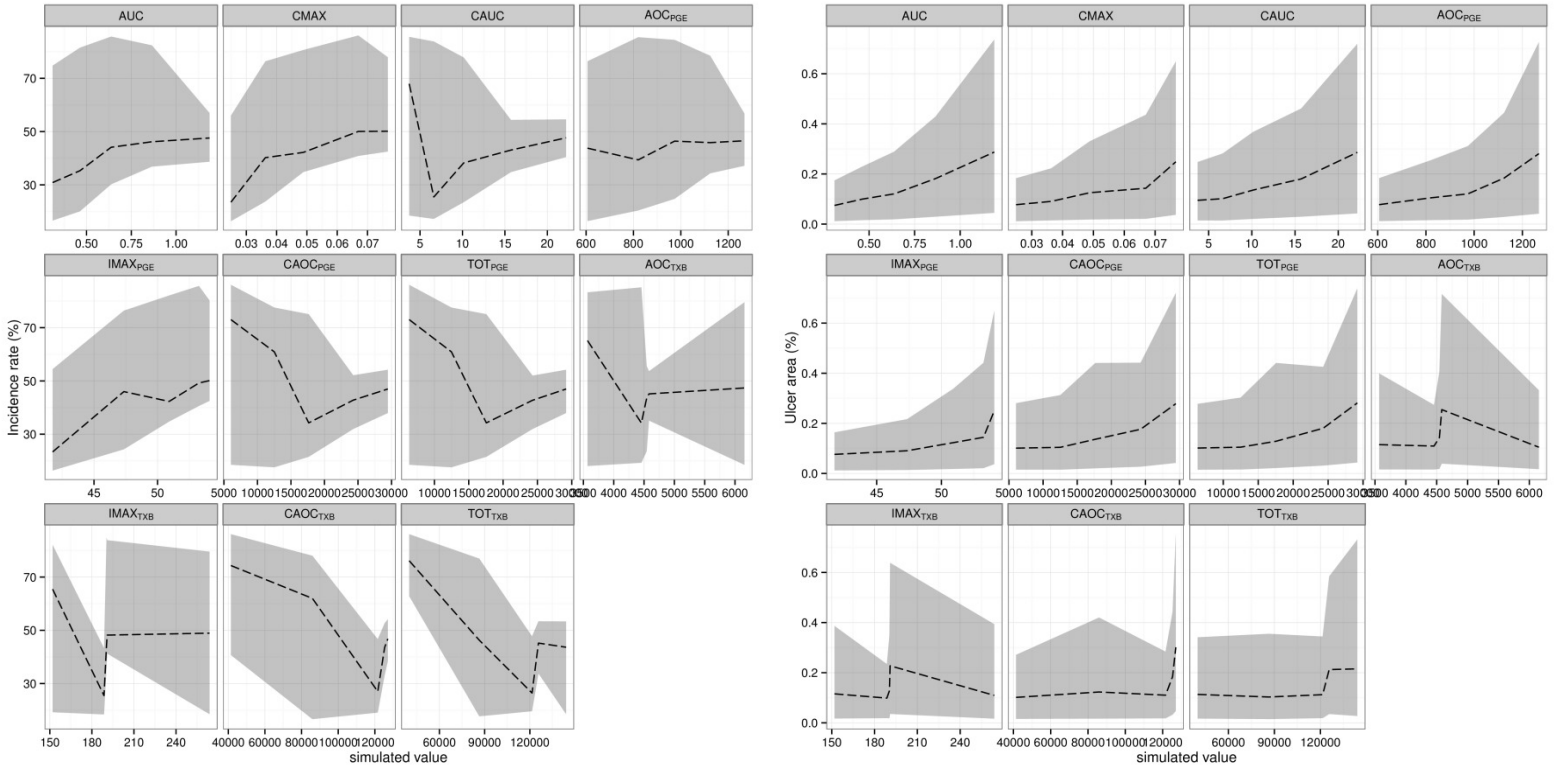


Figure 7: Differences in the sensitivity of explanatory variables describing the relationship between drug exposure/ biomarker levels and adverse events, as determined by the incidence of ulcers and ulceration severity. Dashed lines represented median profile of simulated values using the final model, whereas shaded represents the 95% prediction intervals. AUC= area under the concentration vs. time curve; AOC= area above biomarker concentration vs. time profile; CAUC= cumulative area under the concentration vs. time curve; CAOC= cumulative area over biomarker concentration vs. time profile; CMAX = maximum drug concentration over the period of 24h; CMIN= maximum biomarker inhibition over the period of 24h; IMAX= maximum biomarker inhibition; TOT= time over threshold (i.e., 80% biomarker inhibition); PGE = prostaglandin E_2; TXB= thromboxane E_2). See text for details on the units of the independent variables (x-axis). Based on statistical criteria, it appears that maximum inhibition of PGE_2 was the best predictor of adverse event incidence. On the other hand, cumulative TXB_2 inhibition was found to be the best explanatory variable for the severity of ulceration.

Discussion

Current practices in toxicology and safety pharmacology rely on the concept of thresholds of drug exposure (e.g., NOAEL) as a proxy for the risk of adverse events, which are treated in a mechanism- and time-independent manner. The disadvantage of such an approach is that long term toxicity can become conflated with acute toxicity, which in turn could be mitigated or related to entirely different physiological mechanisms (Blantz, 1996, Dom et al., 2012). Another hurdle to overcome in the assessment of risk is that general toxicity studies are not designed to characterise the relationship between drug exposure and toxicity, but rather to explore the boundary between therapeutic and toxic exposures. As such, data can be uninformative with respect to understanding the causal factors and underlying mechanisms associated with unwanted pharmacological effects. Clearly, these inefficiencies in experimental protocol design also violate the principle of the 3 Rs (reduction, refinement and replacement) and ultimately contribute to biased conclusions about the long term benefit-risk ratio of an intervention (Balls, 1994). By contrast, the use of a model-based approach provides the opportunity to integrate safety and toxicity data and assess in a strictly quantitative manner the contribution of influential factors, namely drug exposure and biomarkers of pharmacological activity to potential adverse events (Danhof et al., 2005, Danhof et al., 2008, Bai et al., 2013).

Mechanism-based analysis of long-term safety and toxicology data

From a methodological perspective, general toxicology studies represent a challenge for model-based analysis techniques since sparse pharmacokinetic data, which are often derived from satellite animals, need to be linked to adverse event data, which are also typically sparse. In addition, lack of individual exposure profiles often prevents further evaluation of the role of relevant physiological or pathophysiological measures, such as biochemistry, haematology or biomarker data as influential covariates on treatment outcome. Typical experimental protocols in toxicology research yield therefore less

informative datasets, as compared to studies aimed at the characterisation of PKPD relationships, which are now commonly used in early drug development (Knight, 2007). In fact, the impact of such limitations has been highlighted in a separate investigation, where focus is given to the statistical aspects of protocol optimisation and to the use of nonlinear mixed effects modelling of safety data (Sahota et al., unpublished results). Among other things, we have identified important design requirements for ensuring accuracy and precision of parameter estimates for safety thresholds.

An important aspect our analysis was to show that without major modification to existing general toxicity protocols, it is possible to explore and eventually elucidate the causal relationship between drug administration, exposure and the incidence and severity of adverse events associated with chronic therapy. In addition to the integrated analysis of pharmacokinetic, pharmacodynamic and toxicity data, here we have shown that the lack of NOAEL in the present study (due the presence of adverse events at all tested dosing levels) has not prevented us from further characterising the exposure-adverse event relationships. Yet, the proposed modifications to the study protocol were designed not to prevent existing empirical analysis methods, including the estimation of non-compartmental parameters such as composite AUCs. The main modifications consisted in the additional collection of biomarker data from animals and the choice for treating histological observations as a continuous data type. The incorporation of biomarkers into the assessment of long term toxicity enables us to further understand time dependencies and nonlinearities in down-stream effects related to the primary pharmacological target (Huntjens et al., 2010). Here we have characterised COX-1 (TXB assay) and COX-2 (PGE assay) activity given their role in maintaining the homeostasis and integrity of the gastric mucosa (Jackson et al., 2000). As shown in Sahota et al. (2014), considerable inhibition of both isoforms occurs at all experimental dose levels.

PKPD relationships as translational factor for the evaluation of risk in humans.

Characterisation of the relationship between chronic exposure and the incidence and severity of adverse events is a critical but not sufficient requirement to predict safety and toxicity in humans. PKPD models need to be parameterised in such a way that it is possible to discriminate between drug-specific and system-specific parameters. Understanding of pharmacokinetic differences in conjunction with detailed information on potential system specific differences, such as varying metabolic capacity, are sine qua non conditions to translate and accurately interpret safety findings (Zuideveld et al., 2007, Chain and Dubois et al., 2013). Even when estimation of such parameters may be impractical, inferences can be made about their magnitude. Undoubtedly, a mechanism-based approach is likely to yield more reliable predictions than the currently accepted use of empirical cover or safety margin which disregard any possible pharmacological basis for both observed and unobserved adverse events.

Specifically with regard to naproxen-induced ulceration, our results need to be interpreted with caution. First, it should be noted that formal extrapolation of our findings requires further information on system-specific properties, including potential differences in gastric mucosa susceptibility to ulceration and expression and activity of isozymes during maintenance and repair processes. Rats appear to be more susceptible to GI toxicity than humans and show gender specific differences in ulceration, so any prediction without correcting for such differences is therefore likely to overestimate risk (Urushidani et al., 1978, Lanza et al., 1979). In addition, from a methodological perspective, the use of non-linear mixed effects modelling as a tool to characterise the determinants of drug effects and concurrently explain variability, imposes a different approach to statistical inference and interpretation of experimental results. Here we have shown that multiple models, with different explanatory variables meet the statistical criteria used for fitting procedures. These apparently conflicting findings can be interpreted as model uncertainty due to design or even imprecision in parameter estimation. On the other hand, these same results can also be considered hypotheses generating, i.e., they shed light into the possible or even plausible combination of mechanisms underpinning the causal path(s) between drug exposure and toxicity. This latter aspect is essential for extrapolating data from animals to

humans. In fact, a study performed by Huntjens *et al.* (Huntjens et al., 2006). The authors conclude that the main determinant of the primary anti-inflammatory, analgesic effect is the degree of target engagement at the tested dose ranges, as defined by the inhibition e of PGE_2 and TXB_2.

It is known that inhibition of both isoforms is required for GI toxicity. Hence, despite our attempt to identify a single biomarker as explanatory variable or covariate on the incidence of ulceration is likely a result of the interaction between them (White, 2004). However, our model was unable to estimate interaction terms. Given that selective COX-2 inhibitors cause less GI toxicity than non-selective inhibitors (Brzozowski et al., 2001, Rostom et al., 2007), this limitation could be overcome by incorporation of toxicity data from compounds with high selectivity for COX-2. Accounting for this interaction will allow prospective prediction of new compounds with varying selectivity for COX-1 and COX-2. Such integration could be achieved either within a Bayesian framework through the use of informative prior distributions, or through simultaneous analysis of the aggregated dataset. Ultimately, such an analysis may shed light on the optimum degree of selectivity to be obtained for the selection of future compounds with a superior risk-benefit profile. Moreover, we anticipate the possibility to extend the approach for the evaluation of NSAID-induced cardiovascular effects (McGettigan et al., 2006, Schneeweiss et al., 2006, Fitzgerald, 2007). A comprehensive risk management strategy prior to market authorisation is now in place for the development of new selective and non-selective COX-inhibitors, which is aimed at the detection of late onset cardiovascular events associated with long term use of a compound (Solomon et al., 2004; Motsko et al., 2006). As such, these adverse drug reactions are not likely to be observed pre-clinically in a traditional chronic toxicity protocol. Efforts are required to predict the implications of continuous target engagement, instead of simply exploring safety signals in wide cohorts of patients (Mukherjee et al., 2001, Solomon et al., 2004).

Potential limitations

The findings of this study demonstrate the feasibility and potential benefits of proposed model-based approach for the evaluation of chronic safety pharmacology and toxicity. However, it should be noted that the accuracy, precision and validity of the method still relies on the experimental data, which is maximised in terms of its informative value. The adverse events we have assessed in this study were relatively frequent. Characterisation of rare or low frequency events may still be difficult, particularly if one cannot make use of historical data (e.g., unprecedented mechanism) or make inferences about class effects. The quantification of model uncertainty is not currently routine practice in traditional PKPD analyses. The present work has shown that even for relatively frequent adverse events, model uncertainty can be significant and therefore one should quantify it. This likely arises from the fact that toxicity studies are generally designed to find safety windows and not to explore the entire exposure-risk profile. We also acknowledge that the absence of ulcerations in vehicle treated animals and the lack of additional cohort with lower exposure levels may represent a weakness in our investigation. True baseline rates for ulceration could not be factored into the analysis, nor was it possible to accurately establish the adverse event rates at lower doses.

In summary, identification of long term adverse events often arises in Phase IV post marketing surveillance. Our investigation has shown how a model-based approach can be used to support early identification of long term adverse events, enabling further integration and translation of pre-clinical data. Our results also illustrate the importance of quantitative methods for further understanding of the mechanisms of toxicity. Moreover, the availability of PKPD relationships may allows us to make inferences about untested doses and dosing regimens, providing an opportunity for risk mitigation, independently from available experimental data.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

Abramoff MD, Magalhaes PJ, Ram SJ, 2004. Image Processing with ImageJ". Biophotonics International. 11, 36-42.

Bai JP, Fontana RJ, Price ND, Sangar V, 2013. Systems pharmacology modeling: an approach to improving drug safety. Biopharm Drug Dispos. doi: 10.1002/bdd.1871.

Balls M, 1994. Replacement of animal procedures: alternatives in research, education and testing. Lab Anim. 28, 193-211.

Blantz RC, 1996. Acetaminophen: acute and chronic effects on renal function. Am J Kidney Dis., 28(1 Suppl 1):S3-6

Brzozowski T, Konturek PC, Konturek SJ, Sliwowski Z, Pajdo R, Drozdowicz D, Ptak A, Hahn EG, 2001. Classic NSAID and selective cyclooxygenase (COX)-1 and COX-2 inhibitors in healing of chronic gastric ulcers. Microsc Res Tech. 53,343-353.

Chain A, Dubois V, Danhof M, Sturkenboom M, Della Pasqua O, 2013. Identifying the translational gap in the evaluation of drug-induced QTc-interval prolongation. Br J Clin Pharmacol. 76, 708-24.

Chakraborti AK, Garg SK, Kumar R, Motiwala HF, Jadhavar PS, 2010. Progress in COX-2 inhibitors: a journey so far. Curr Med Chem. 17(15), 1563-93.

Capone ML, Tacconelli S, Sciulli MG, Anzellotti P, Di Francesco L, Merciaro G, Di Gregorio P, Patrignani P, 2007. Human pharmacology of naproxen sodium. J Pharmacol Exp Ther. 322, 453-60.

Comets E, Brendel K, Mentre F, 2008. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. Comput. Methods Programs Biomed. 90, 154-166.

Connelly JC, Bridges JW, 1991. Species variation in target organ toxicity. In: Target Organ Toxicity. Cohen GM (ed). CRC Press, Boca Raton, Florida, 89-115.

Coruzzi G, Venturi N, Spaggiari S, 2007. Gastrointestinal safety of novel nonsteroidal antiinflammatory drugs: selective COX-2 inhibitors and beyond. Acta Biomed. 78, 96-110

Danhof M, Alvan G, Dahl SG, Kuhlmann J, Paintaud G, 2005. Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. Pharm Res. 22, 1432-7.

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, 186-91.

Della Pasqua O, 2013. Translational pharmacology: from animal to man and back. Drug Discov Today Technol. 10(3), e315-7.

Dom N, Knapen D, Blust R, 2012. Assessment of aquatic experimental versus predicted and extrapolated chronic toxicity data of four structural analogues. Chemosphere. 86, 56-64.

Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol 2005; 42(3):265-74.

Famaey JP, 1997. In vitro and in vivo pharmacological evidence of selective cyclooxygenase-2 inhibition by nimesulide: an overview. Inflamm Res. 46, 437-46.

Fitzgerald GA, 2007. COX-2 in play at the AHA and the FDA. Trends Pharmacol Sci. 28, 303-307.

Fornai M, Antonioli L, Colucci R, Pellegrini C, Giustarini G, Testai L, Martelli A, Matarangasi A, Natale G, Calderone V, Tuccori M, Scarpignato C, Blandizzi C, 2014. NSAID-Induced Enteropathy: Are the Currently Available Selective COX-2 Inhibitors All the Same? J Pharmacol Exp Ther. 348, 86-95.

Halter F, Tarnawski AS, Schmassmann A, Peskar BM, 2001. Cyclooxygenase 2—implications on maintenance of gastric mucosal integrity and ulcer healing: controversial issues and perspectives. Gut 49, 443-53.

Hooker AC, Staatz CE, Karlsson MO, 2007. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. Pharm Res. 24, 2187-97.

Huntjens, D. R., D. J. Spalding, M. Danhof, and O. E. Della Pasqua, 2006. Correlation between in vitro and in vivo concentration-effect relationships of naproxen in rats and healthy volunteers. Br J Pharmacol. 148,396-404.

Jackson LM, Wu KC, Mahida YR, Jenkins D, Hawkey CJ, 2000. Cyclooxygenase (COX) 1 and 2 in normal, inflamed, and ulcerated human gastric mucosa. Gut 47, 762-770.

Josa M, Urizar JP, Rapado J, Dios-Viéitez C, Castañeda-Hernández G, Flores-Murrieta F, Renedo MJ, Trocóniz IF, 2001. Pharmacokinetic/pharmacodynamic modeling of antipyretic and anti-inflammatory effects of naproxen in the rat. J Pharmacol Exp Ther. 297, 198-205.

Kargman, S, Charleson S, Cartwright M, Frank M, Riendeau M, Mancini J, Evans J, O'Neill G, 1996. Characterization of Prostaglandin G/H Synthase 1 and 2 in rat, dog, monkey, and human gastrointestinal tracts. Gastroenterology 111,445-454.

Karlsson MO, Savic RM, 2007. Diagnosing model diagnostics. Clin Pharmacol Ther. 82, 17-20.

Knight A, 2007. Animal experiments scrutinised: systematic reviews demonstrate poor human clinical and toxicological utility. ALTEX 24, 320-5.

Lacroix BD, Friberg LE, Karlsson MO, 2012. Evaluation of IPPSE, an alternative method for sequential population PKPD analysis. J. Pharmacokinet Pharmacodyn. 39, 177-93.

Langenbach R, Loftin C, Lee C, Tiano H, 1999. Cyclooxygenase knockout mice: models for elucidating isoform-specific functions. Biochem Pharmacol. 58, 1237-46.

Lanza FL., Royer Jr GL., Nelson RS, Chen TT, Seckman CE, Rack MF, 1979. The effects of ibuprofen, indomethacin, aspirin, naproxen, and placebo on the gastric mucosa of normal volunteers. Digest Dis Sci. 24:823-828.

Lindbom L, Pihlgren P, Jonsson EN, 2005. PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput Methods Programs Biomed. 79, 241-57.

Loftin CD, Tiano HF, Langenbach R, 2002. Phenotypes of the COX-deficient mice indicate physiological and pathophysiological roles for COX-1 and COX-2. Prostaglandins Other Lipid Mediat. 68-69, 177-85.

McGettigan P, Henry D, 2006. Cardiovascular risk and inhibition of cyclooxygenase: A systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. JAMA 296, 1633-1644.

McGonigle P, Ruggeri B, 2014. Animal models of human disease: Challenges in enabling translation. Biochem Pharmacol. 87, 162–171.

Morita I, 2002. Distinct functions of COX-1 and COX-2. Prostaglandins Other Lipid Mediat. 68-69, 165-75.

Motsko S, Rascati K, Busti A, Wilson J, Barner J, Lawson K, Worchel J, 2006. Temporal Relationship between Use of NSAIDs, Including Selective COX-2 Inhibitors, and Cardiovascular Risk. Drug Safety 29, 621-632.

Mukherjee D, Nissen SE, Topol EJ, 2001. RIsk of cardiovascular events associated with selective cox-2 inhibitors. JAMA 286, 954-959.

Nyman AM, Schirmer K, Ashauer R, 2012. Toxicokinetic-toxicodynamic modelling of survival of Gammarus pulex in multiple pulse exposures to propiconazole: model assumptions, calibration data requirements and predictive power. Ecotoxicology 21, 1828–1840.

O'Brien PJ, 2008. Cardiac troponin is the most effective translational safety biomarker for myocardial injury in cardiotoxicity. Toxicology 245, 206-18.

R Development Core Team. R: A language and Environment for Statistical computing. In R Foundation for Statistical Computing. Vienna Austria, 2012.

Reagan-Shaw S, Nihal M, Ahmad N. 2008. Dose translation from animal to human studies revisited. FASEB J. 22, 659-661.

Rohatagi S, Kastrissios H, Gao Y, Zhang N, Xu J, Moberly J, Wada R, Yoshihara K, Takahashi M, Truitt K, Salazar D, 2007. Predictive population pharmacokinetic/pharmacodynamic model for a novel COX-2 inhibitor. J Clin Pharmacol. 47, 358-370.

Rostom A, Muir K, Dubé C, Jolicoeur E, Boucher M, Joyce J, Tugwell P, Wells GW, 2007. Gastrointestinal Safety of Cyclooxygenase-2 Inhibitors: A Cochrane Collaboration Systematic Review. Clin Gastroenterol Hepatol. 5, 818-828.

Satterwhite JH, Boudinot FD, 1988. High-performance liquid chromatographic determination of ketoprofen and naproxen in rat plasma. J Chromatogr. 431, 444-9.

Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA, 2006. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: An instrumental variable analysis. Arthrit Rheumat 54, 3390-3398.

Solomon DH, Schneeweiss S, Glynn RJ, KiyotaY, Levin R, Mogun H, Avorn J, 2004. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation 109, 2068-2073.

Sahota T, Sanderson I, Danhof M, Della Pasqua O. Model-based analysis of thromboxane B_2 and prostaglandin E_2 as biomarkers in the safety evaluation of naproxen. Toxicol Appl Pharmacol. in press, 2014

Sahota T, Danhof M, Della Pasqua O. Utility of model based approaches to predict the risk of adverse events from preclinical toxicology protocols. In: Pharmacology-based assessment of adverse drug effects. A quantitative approach to the prediction of risk in humans. PhD Thesis University of Leiden, 2014.

Takeuchi K, 2012. Pathogenesis of NSAID-induced gastric damage: Importance of cyclooxygenase inhibition and gastric hypermotility. World J Gastroenterol. 18, 2147-2160.

Urushidani T, Okabe S, Takeuchi K, Takagi K, 1978. Strain differences in aspirin-induced gastric ulceration in rats. Jpn J Pharmacol. 28, 569-578.

Xie HG, Wang SK, Cao CC, Harpur E, 2013. Qualified kidney biomarkers and their potential significance in drug safety evaluation and prediction. Pharmacol Ther. 137, 100-7.

Wallace JL, 2008. Prostaglandins, NSAIDs, and gastric mucosal protection: Why doesn't the stomach digest itself? Physiol Rev. 88, 1547–1565.

Whittle B J, 2004. Mechanisms underlying intestinal injury induced by anti-inflammatory COX inhibitors. Eur J Pharmacol. 500, 427-439.

Zhang L, Beal SL, Sheiner LB, 2003. Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J Pharmacokinet Pharmacodyn. 30, 387-404.

Zhang L, Beal SL, Sheinerz LB, 2003. Simultaneous vs. sequential analysis for population PK/PD data II: robustness of methods. J Pharmacokinet Pharmacodyn. 30, 405-16.

Zuideveld KP, Van der Graaf PH, Peletier LA, Danhof M, 2007. Allometric scaling of pharmacodynamic responses: application to 5-HT_{1A} receptor mediated responses from rat to man. Pharm Res. 24, 2031-9

SECTION 4: CONCLUSIONS AND PERSPECTIVES

CHAPTER 9

Pharmacology-based assessment of toxicity: towards quantitative risk prediction in humans

Undoubtedly, the main objective of toxicology studies during the course of drug discovery and development is to support scientists, clinicians and regulators in establishing the likely risks posed to humans and more specifically to patients. Challenges exist not only when interpreting the results and making extrapolations to predict risk, but also at the planning and design stage, including the choice of most relevant species, choice of the doses to be investigated and duration of treatment.

Despite the requirement for extrapolations and more quantitative measures of what represents safe exposure, limited attention has been given to the role of alternative methodologies that have emerged in pharmacological sciences. Over the last decades, most of the empirical evidence generated as part of general toxicity package in drug development has been treated in a descriptive manner. Yet, numerous statistical modelling tools have been developed over that same period that have substantially improved our understanding of human exposure, pharmacokinetics, pharmacodynamics and disease processes (32-35).

Opportunities exist for toxicology to transition from a qualitative science to a discipline capable of quantitatively describing relevant biological and pharmacological processes that determine the exposure-effect relationships in animals and in humans. However, there are multiple methodological obstacles to overcome before efficient and early prediction of chronic toxicity of new chemical and biological entities becomes routine practice in pharmaceutical R&D. First of all, the application of quantitative modelling concepts to toxicology imposes the need for an integrative approach in that the evaluation of toxicity

and adverse events become part of continuum that encompasses primary and secondary pharmacology as start point (5,6).

Integration of quantitative tools within experimental design, data collection, analysis and interpretation has become more important than ever in pharmacology research. Evidence so far supports the use of such tools to 1) optimise experimental protocols, 2) refining and reducing the burden and number animals required and most importantly 3) translating drug effects from animals to humans.

The scientific and regulatory communities should acknowledge that most toxicity tests, as currently designed, provide only a qualitative estimate of the hazard associated with supratherapeutic exposure (7). This is clearly not the most important question that needs to be addressed from a clinical perspective. The safety and toxicity profile of a medicinal product needs to include an assessment of the risk at therapeutic levels, especially in chronic disease conditions (8). Yet, data produced using current testing guidelines are not always suitable for robust mathematical exposure–response modelling. As stated at the beginning of this thesis, we recognise therefore that adequate data integration and optimised protocols are required before quantitative modelling can be applied as mainstream tool for the analysis and interpretation of toxicity and adverse events.

The research performed in this thesis is therefore focused on a number of issues that need to be considered during the course of drug discovery and development to ensure more efficient use of the data generated in safety pharmacology and toxicology protocols. We have attempted to address four questions that can be considered enablers for the implementation of a systems approach for the characterisation of physiological and pharmacological responses induced by chronic exposure to a drug.

In **Chapter 1** we reviewed mainstream safety assessment practices in drug development and the consequences of empirical evidence generation. Based on historical examples we identified methodological flaws in the current paradigm and categorised issues relative to

the scientific rationale as a hierarchical tree describing the decision making process. From a theoretical perspective, different facets of the same problem were discussed, which relate to four seminal areas of scientific research: 1. optimisation, 2. translation, 3. analytical construct and 4. decision criteria. The implications of each of these points for the implementation of model-based methods were addressed separately. We showed that errors in the prediction of safety may arise due to the use of empirical safety thresholds, which are used as a proxy or surrogate for toxicity or undesirable effects. Published data make it clear that instead of pursuing a more mechanistic approach, empirical methods continue to be used. To cope with inaccuracy and poor precision, safety factors, also known as uncertainty factors, have been incorporated on the top of empirical thresholds. Their application in drug development has become widespread and is detailed within the regulatory guidelines. Based on historical examples, we have shown some important challenges for the early characterisation of the safety profile of a new molecule and discuss how model-based methodologies can be applied for better design and analysis of experimental protocols. An initial conclusion can be drawn in support of the efforts presented throughout the thesis, in that current practices fail to support decision making on multiple levels.

A shift in paradigm was then proposed to ensure that pharmacological concepts are incorporated into the evaluation of safety and toxicity. In **chapter 2**, we presented the conceptual and methodological aspects that underpin the work presented in the subsequent chapters of this thesis. Our goal was to explore the feasibility of pharmacologically based quantitative toxicology assessment and risk prediction in humans and, where possible, to compare the performance of this approach to traditional safety assessment approaches. We have also highlighted an important difference in the objective of current experimental protocols, which are aimed at confirming safety rather than characterising the range of toxicity. Four important questions were highlighted which define the scientific framework presented in the subsequent chapters, which can be defined as opportunities for optimisation and knowledge integration. We set a constraint that existing experimental protocols would be viewed as a starting point, and any proposals to deviate from these

protocols would be minimal. Although similarities exist between efficacy and toxicology assessment from a pharmacological point of view, here we proposed an investigational plan to determine the methodological requirements of toxicological data analysis. Furthermore, we set an often forgotten objective in non-clinical research, i.e., the ethical duty to refine, reduce and replace the use of animals in experimental protocols (9). The investigational plan of the thesis was detailed and divided into two distinct sections (sections 2 and 3), in which the development of methodology is followed by a case study with real data.

Conceptual framework

In **Section II**, the advantages and limitations of a model-based approach were evaluated. Conceptually, we have demonstrated how factors such as within- and between-subject variability or uncertainty in estimation can be accounted for when descriptive statistics are replaced by pharmacokinetic and pharmacokinetic-pharmacodynamic parameters. Using simulations to replicate experimental protocols we have illustrated how different measures of exposure can be obtained which may be physiologically more relevant for the characterisation of delayed or late onset adverse events. Particular focus was given to the feasibility of assessing long term risk from shorter duration studies. In addition, we have identified alternative options for the design and analysis of preclinical general toxicology protocols.

Initially, focus was given to the use of non-linear mixed-effect (NLME) modelling as a data analysis tool for the evaluation of toxicokinetic experiments and parametric estimation of safety thresholds. In **Chapter 3** we simulated toxicokinetic data from satellite treatment groups in general toxicity protocols using three hypothetical drugs, with distinctly different pharmacokinetic properties. Analysis of the simulated datasets with traditional non-compartmental analysis and NLME models allowed us to measure the performance of both methodologies and compare them in terms of bias and precision. The main source of the bias in the parameters of interest was found to be intrinsic to the non-compartmental method, especially when looking at the estimation of Cmax. Our results also revealed 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. As demonstrated by the simulations, this issue could be circumvented by model-based methods, which enable the assessment of uncertainty associated with a variety of causes such as uninformative study design, large variability and/or unknown covariates. The use of hypothetical drugs with different pharmacokinetic properties also allowed us to illustrate how obtaining a pharmacokinetic model provides opportunities for different parameterisations or metrics of drug exposure, as for example, the estimation of cumulative AUC to describe irreversible or chronic toxicity, including predictions beyond the study duration. This entails an increase in the quality of the decision-making process and ultimately in the interpretation of the estimated safety thresholds.

Since experimental protocols for the evaluation of general toxicity are not optimised for model-based analysis, and more specifically for population pharmacokinetic modelling, an important question to be addressed is whether they can be optimised to ensure a reduction in the number of animals required, whilst still providing sufficient estimation precision for measures of exposure, which are often secondary pharmacokinetic parameters, such as AUC and Cmax. In contrast to existing optimality software and algorithms, which support optimisation of experimental design with respect to primary parameter precision, in **Chapter 4** we show that secondary parameters can be optimised without the resource-intensive procedures imposed by D-optimality. Our approach instead consisted of FIM evaluations followed by calculation of the expected secondary parameter precision. Both of these procedures were found to be computationally inexpensive. Most importantly, our results highlight the impact of optimal protocol design on parameter estimation. The proposed method for optimisation of sampling time and group size indicates that a reduction of approximately 30% in the number of animals can be obtained for composite sampling designs without significant loss of precision in the estimates of interest. This improvement was found to be independent of differences in drug disposition, as assessed by the different profiles derived for the hypothetical compounds. Our analysis also suggests that for

composite methods sample size does not determine the precision of the pharmacokinetic parameters of interest. Rather, it is the sampling scheme and dose levels which matter. Interestingly, we have observed that the precision of the secondary parameters remains unaffected even when some of the primary pharmacokinetic parameters are poorly estimated.

Whereas the use of model-based estimates for the assessment of safety thresholds may be perceived as complicated, this approach was shown to be unbiased and predictive, allowing for the incorporation of the physiological factors underlying the pharmacokinetic properties of the drug under investigation, such as metabolic saturation. Moreover, our simulation scenarios provided evidence of the feasibility to integrate prior information, including data from other experiments.

Still within the scope of protocol optimisation, in **Chapter 5** we explored the implications of introducing biomarkers into the evaluation of a drug's safety toxicity profile. Here we emphasised the fact that accurate prediction of long term adverse events and toxicity may require one to identify not only the exposure at which the effects are observed, but also biomarkers of pharmacological activity. In contrast to traditional protocols, which imply a direct relationship between observed systemic exposure and adverse events, we have proposed the collection of biomarkers at the scheduled pharmacokinetic sampling points to facilitate the characterisation of pharmacokinetic-pharmacodynamic relationships. Our evaluation also compared the analysis of biomarker data based on standard non-compartmental methods. We simulated toxicokinetic and biomarker data from satellite groups using a variety of hypothetical drugs. The analysis of the simulated data showed that the true underlying model was often unidentifiable particularly in scenarios with delayed PD effects (hysteresis). However, in all scenarios, model approximations could be made which led to satisfactory performance in predicting biomarker levels. We believe, therefore, that greater awareness is required about the limitations of current experimental protocols, particularly in a period in which long-term safety have become a major clinical and regulatory concern. To mitigate such effects we recommend careful consideration of model

uncertainty. Our analysis showed that model selection criteria should be guided not only by ability to describe data but also by assessing the physiological relevance of model assumptions. When model development ends in multiple competing models performing similarly with respect to model selection criteria, clear reporting of such model uncertainty is necessary. In any case, model averaging should be discouraged when predictions arising from different model differ significantly. Finally, parameter uncertainty should be incorporated when performing simulations or using the model to make predictions.

Our goal with **Chapter 6** was to investigate the feasibility of integrating the aforementioned mechanistic PKPD models with adverse event data for model-based toxicology assessment. Similar *in silico* methods used in chapters 3 and 5 were used to simulate drug induced and background adverse events according to three different pharmacological mechanisms (direct, indirect, and irreversible binding). We focused on rare and chronic adverse drug reactions to provide the largest methodological challenge, including reversible and irreversible drug effects. To ensure real-life conditions, assumptions were made with regard to situations 1) in which drug-induced and background adverse events are indistinguishable from each other, 2) the time interval elapsed between onset and diagnosis was large and symptoms can be detected only once per animal during histological examination and 3) the adverse event can be treated as binary data. Our results showed that estimation of safety thresholds, as determined by the NOAEL, was highly biased and imprecise. Moreover, in two out of three scenarios where the effects of safe and effective hypothetical compounds were simulated, we found that strict use of the NOAEL as go/no-go criteria would lead to a more than 50% probability of concluding that the compound is unsafe and consequently leading to wrongful termination of the development program. Upon investigating the feasibility of model-based analysis, we found that we required two important components for successful quantification of rare drug-induced effects: a) the availability of prior information on background adverse events and b) MCMC-based estimation algorithms. Regarding the first requirement, we showed that without prior information, adverse drug reactions are confounded with background incidence rates, preventing parameter identifiability. We found that an aggregated historical placebo data was sufficient to resolve this confounding.

On the other hand, when evaluating the performance of available parameter estimation methods, we found that maximum likelihood algorithms are unstable and unreliable. By contrast, MCMC-based estimation provided stable and accurate measures of parameter uncertainty. The use of the BIC as a model comparison and averaging criteria showed consistently high model specification uncertainty. Our results highlighted that traditional model selection and averaging techniques based on the penalizing models for complexity were not appropriate as they heavily weighted models featuring no drug effect. Finally, it should be noted that missed adverse events were also easily quantified using the proposed strategies. Differently, from the empirical approach to treating missing events as absent, the use of MCMC methods provided evidence of the parameter distribution, enabling imputation of the events, even if they have not been observed.

In summary, the conceptual framework presented throughout this section provides evidence regarding the feasibility and relevance of a model-based approach for the, evaluation of safety pharmacology and toxicology profile of new molecules prior to their progression into humans. It has become clear that current methods in preclinical toxicology do not support the integration of pharmacokinetic and pharmacodynamic data as basis for predicting safe exposure in humans. By contrast, a model-based approach represents a viable tool for characterising PKPD relationships, including estimates of parameter and model uncertainty. A benefit this strategy lends to decision-making is that clinical judgment can be applied to consider the entire relationship between drug exposure and adverse event, rather than a point estimate or threshold.

In the third part of the thesis (**Section III**) we attempted to illustrate the implementation of experimental protocols that meet the requirements for model-based analysis. Given the continuous debate regarding the benefit-risk balance of chronic treatment with non-steroidal anti-inflammatory drugs, naproxen was used as a paradigm compound to evaluate the known acute and chronic toxicities. Whilst the lack of selectivity of naproxen and the evidence for distinct mechanisms underpinning acute effects (such as bleeding and ulceration) and long term effects (such as renal and cardiovascular damage) have evolved

over the years and might not have been understood at time of the development of the compound, our investigation made it quite didactic in that it demonstrates how human safety and toxicity may require characterisation of drug effects at exposure levels corresponding to the therapeutic doses. Furthermore, by considering the requirements for a suitable experimental protocol, we also took the opportunity to identify practical challenges and difficulties that one may face for the prospective use of the methodology.

Practical application

Using a typical toxicology protocol in rats, in **Chapter 7** we have explored how two biomarkers, namely thromboxane (TXB_2) and prostaglandin (PGE_2), can be used in conjunction with drug exposure data to evaluate short, moderate and long-term treatment effect. It was assumed that gastrointestinal bleeding is due continuous COX-1 inhibition, whereas ulceration results primarily from the suppression of COX-2. Pharmacokinetic and biomarker data were integrated with data from historical protocols and published literature to ensure characterisation of drug properties at putative therapeutic levels. We found that the pharmacokinetics were best described a one-compartmental model with first-order absorption. A nonlinear relationship between dose and bioavailability was included into the model which led to a less than proportional increase in exposure with respect to dose. Toxicity findings showed gastric ulceration at all tested dosing levels (7.5, 15, 40 and 80 mg/kg) meaning that no NOAEL could be established. Despite the lack of a safety threshold, we have demonstrated that experimental data can be used to characterise the underlying PKPD relationships for both TXB_2 and PGE_2, which were best described by direct inhibition models. Estimation of all parameters was precise and models performed well in diagnostics and predictive checks, confirming the feasibility claims of chapters 3 and 5. In addition, our results emphasised the role of construct validity to account for the potential impact of interspecies differences in the underlying exposure-response relationships. As indicated by the level of biomarker inhibition observed at the selected doses, inferences from the preclinical data can be made to predict drug effects at comparable levels of inhibition in

humans. In fact, we found that the PKPD relationship was within 10-fold range of published human values, raising questions about differences in the sensitivity of rats to the cyclo-oxygenase inhibition.

Whereas the use of a parametric approach and more specifically of hierarchical mixed-effects modelling to inform experimental design and dose selection represents an important step in the advancement of translational toxicology, both from a biological and statistical perspective, we have also identified a few limitations that are worth mentioning. Some mechanisms of action may be too complex or poorly understood to be characterised by PKPD modelling of data arising from a general toxicity study, even if biomarkers are collected. Multiple downstream markers may present a significant confounder problem which cannot be avoided without additional data. This cannot be easily addressed by the proposed analysis method and becomes a drug development issue.

Our feasibility evaluation was complemented **in Chapter 8** by further integrating the histological data obtained at completion of treatment to the observed biomarker effects. In this investigation we showed how the pharmacokinetic-pharmacodynamic (PKPD) model obtained in the previous chapter can be incorporated into a formal analysis to describe adverse event incidence and severity. The adverse events (gastric ulceration) were quantified as continuous measures of ulcerative area. Ulceration incidence (binary) and severity (severity) were modelled as two separate variables or endpoints of interest. The final model parameters describing the incidence of adverse events showed that ulceration was an acute effect driven primarily by maximum inhibition of PGE_2 levels corresponding to maximum blockade of COX-2. The implications of model uncertainty highlighted in chapter 6 prompted us to combine model selection criteria with bootstrap methodology to obtain model uncertainty estimates. We found that there was minimal model uncertainty with regard to the characterisation of ulceration incidence but high specification uncertainty when describing ulceration severity. Despite such high uncertainty, cumulative suppression of TXB_2 levels, assumed to result from to long term blockade of constitutive COX-1 could be identified as an influential covariate of ulceration severity. In summary, our investigation has

shown how a model-based approach can be used to support early identification of long term adverse events, enabling further integration and translation of pre-clinical data. Our results also illustrated how the availability of PKPD relationships may allows us to make inferences about untested doses and dosing regimens, providing an opportunity for risk mitigation, independently from available experimental data.

From a methodological perspective, the findings arising from this experimental protocol demonstrated the feasibility and potential benefits of proposed model-based approach for the evaluation of chronic safety pharmacology and toxicity. However, it should be noted that the accuracy, precision and validity of the method still relies on the experimental data. The adverse events we have assessed in this study were relatively frequent. Characterisation of rare or low frequency events may still be difficult, particularly if one cannot make use of historical data (e.g., unprecedented mechanism) or make inferences about class effects. We also acknowledge that the absence of ulcerations in vehicle treated animals and the lack of additional cohort with lower exposure levels may represent a weakness in our investigation. True baseline rates for ulceration could not be factored into the analysis, nor was it possible to accurately establish the adverse event rates at lower doses.

Practical recommendations for safety assessment

Given the challenges and limitations for the characterisation of exposure-effect relationships using data arising from typical experimental protocols, we have compiled a list of points to consider regarding methodological and practical issues, including recommendations for further protocol optimisation which may facilitate the implementation of model-based techniques in safety pharmacology and toxicology research.

- Sampling scheme and dose selection to be used in the safety pharmacology and toxicology protocols need to take into account the underlying mechanism or mode of action associated with the primary pharmacological target or receptor system. It is imperative to ensure that different levels of target engagement (i.e., receptor occupancy) within and beyond the expected therapeutic exposure are included.
- Study protocols should be analysed in an integrated manner to ensure accurate conclusions are drawn about the safety and toxicity profile of the compound. This implies the combination of data arising from all experimental protocols where pharmacokinetic and biomarker data are collected.
- Integration of historical data as priors (describing parameter distributions) may be required to reduce the degree of uncertainty associated with models predictions across the exposure-response curve describing the adverse event or toxicity.
- Better, continuous inference metrics (e.g., EC_{10}, cumulative biomarker levels and other derived parameters from the underlying PKPD relationship), are required to extrapolate findings from toxicological dose levels to clinically relevant therapeutic exposure ranges. Safety thresholds are conservative and biased.
- Optimisation of study design should be performed on parameters of interest (i.e. AUC< Cmax) rather than primary model parameters. Standard optimality algorithms (e.g. D-optimality) are not suitable for that purpose, as current software programs maximise the overall expected parameter precision within design constraints (10). Acceptability criteria for precision of parameters of interest should be defined in advance and evaluated within the design space taking into account feasibility aspects. Selected designs should be parsimonious in that further reduction does not produce a sufficient design.
- The impact of prior model and parameter uncertainty should be investigated during the study design phase (e.g., by simulation) to ensure uncertainty is factored accordingly into the expected study outcomes.
- Lack of model identifiability represents a risk for PKPD analyses based on standard experimental toxicology and safety pharmacology protocols. Therefore, to ensure model and parameter identifiability, simulation re-estimation (SSE), bootstrapping and sensitivity

analysis with respect to initial estimates are essential steps to be considered throughout model development and validation (11, 12).

- Maximum likelihood methodology is insufficient to model rare adverse events. The use of a Monte Carlo-Markov Chain algorithm is required for stability and accurate parameter uncertainty estimation (13).
- Model uncertainty is a likely outcome of adverse event modelling. Model uncertainty should therefore be accounted for and quantified e.g. using the bootstrap covariate method. Final model predictions should also be displayed along with prediction intervals to account for parameter uncertainty (14-16).
- Traditional model selection and averaging criteria which penalise for model complexity (e.g. BIC and AIC) are inappropriate when modelling rare adverse event data as models without drug effects are overweighed. A conservative approach to model selection should instead be guided by pharmacological plausibility and data fitting metrics without penalisation (e.g. -2 log likelihood).

Future perspectives

The methodological issues identified through simulation scenarios and the lessons learned from the integrated experimental protocol developed for naproxen have highlighted the limitations of current practice in the evaluation of the safety profile of new chemical entities. More specifically, our findings reveal that inferences about safe exposure as well as the risk associated with long term use of a compound cannot be achieved by scattered empirical experimentation. A framework is required that enables integration, in a parametric manner, of experimental data and theoretical knowledge. As shown in figure 1, such a framework would encompass multidimensional data, allowing for the incorporation of not only in vivo, but also in vitro data as input for computational models.

-

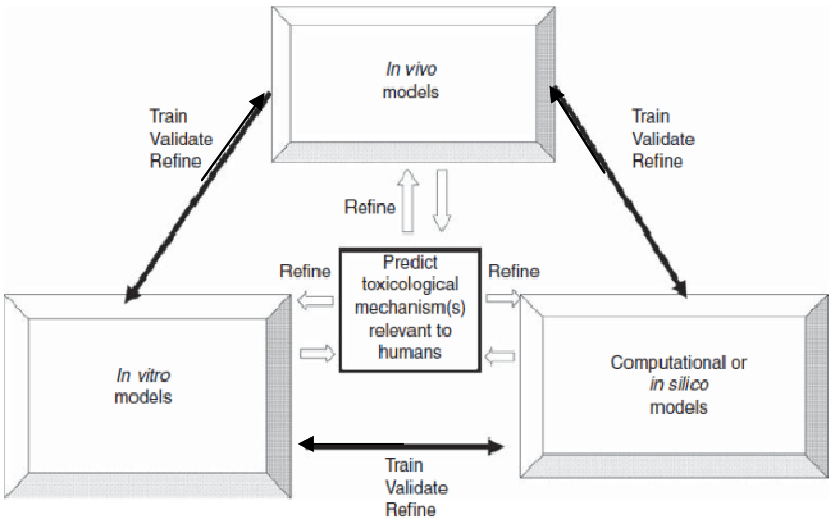


Figure 1: Integration of *in vitro* and *in vivo* data as input for *in silico* models. A model's ability to predict toxicity in humans is used as reference for further refinement of the model as well as of the experimental design (modified with permission from 17).

In order for computational models to be used to predict long term safety and toxicity in humans, methods are required that incorporate the mechanisms associated with primary and known secondary targets. In addition, general parameterisations also need be considered to describe drug action beyond the receptor or target level, including broader concepts, such as inappropriate cell signalling, mutagenesis and carcinogenesis (17). The emerging field of systems pharmacology could hold promise in this respect by providing a systematic framework which accounts for all relevant processes from target-drug interaction at the biophase to downstream cellular and organ level processes (110). In fact, one of the first examples of the approach for the characterisation of general toxicity is the case of vitamin D, which has been used to establish target tissues for 1,25-$(OH)_2$ vitamin D3 (19). Systems pharmacology makes evident that the actions of most of the target tissues are unrelated to systemic calcium regulation and are instead related to the regulation of endocrine and exocrine secretion, cell proliferation and cell differentiation. It can be easily seen that many, if not all, target tissues of the vitamin D system will be activated in patients

treated with a vitamin D-related compound – whether taken against osteoporosis, tumour growth or any other single condition. However, physiologic dosing of vitamin D does not cause hypercalcemia – hypercalcemia is related to overdosing only (19, 20).

More recent developments have allowed for a more quantitative characterisation of system and drug behaviour *in vivo*. Quantitative systems pharmacology represents a convergence of systems biology and pharmacology, combining computational and experimental methods to elucidate and predict disease progression and drug effects. The approach does not only take into account the underlying pharmacokinetic-pharmacodynamic relationships, but also potentially multiple components of the biological systems leading to changes in biological or disease state. This feature is particularly relevant, both from a clinical and methodological perspective, for the parameterisation of long term adverse events, which may originate from a perturbation of homeostatic mechanisms, from cellular changes or cell injury (21). In contrast to empirical and probabilistic models, in systems pharmacology one can introduce both mechanistic and physiological elements as parameters for the characterisation of acute, delayed or late safety signals, which in turn can be correlated with global clinical measures, such as morbidity (figure 2). In conjunction with physiologically-based pharmacokinetic (PBPK) models, systems pharmacology can provide the basis for determining the impact of observed variations in physiological and biochemical factors, as well as discriminate pharmacokinetic from pharmacodynamic or biological variability. Instead of compartments defined solely by experimental kinetic data, compartments in a PBPK model are based on realistic organ and tissue groups, with weights and blood flows obtained from the literature. Moreover, instead of compartmental rate constants determined solely by fitting data, actual physicochemical and biochemical properties of the compound can often be used to define parameters in the model. In particular, a properly validated PBPK model can be used to perform the high-to-low dose, dose-route, and interspecies extrapolations necessary for estimating human risk on the basis of experimental protocols in animals.

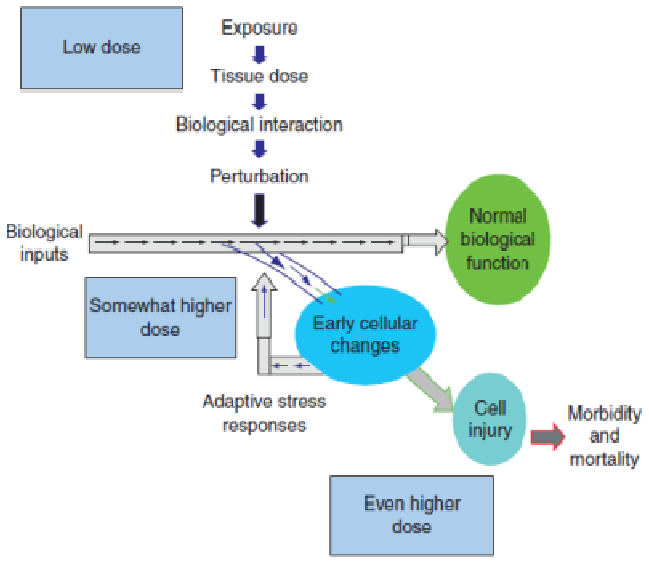


Figure 2. Dose-exposure-response paradigm for toxic effects, relating observed response as consequence of perturbations of the normal control processes in the cell. Low doses are largely without functional consequences; intermediate doses activate adaptive stress responses with attendant homeostatic controls; and high-enough exposures lead to overt toxicity (reprinted with permission from 22).

One example of modelling incorporating systems pharmacology, which can be deemed relevant for the evaluation of target-mediated efficacy and safety, regards the effects of steroids. The work developed by Ramakrishnan and collaborators shows how experimental data, including transcription and gene mediated effects can be parameterised to describe the binding of steroidal drugs to the cytosolic glucocorticoid receptor and subsequent translocation of the complex into the nucleus where it binds as a dimer to the glucocorticoid responsive element (GRE) in the DNA (23) (Figure 3). This leads to the enhanced or repressed expression of numerous genes. At the same time, binding of the activated steroid-receptor complex to the GRE results in reduced levels of receptor mRNA. This further leads to decrease in the free receptor density in the cytosol. The concept nicely illustrates how long term use of corticosteroids may lead to suppression of normal physiological function at cellular and whole organ levels.

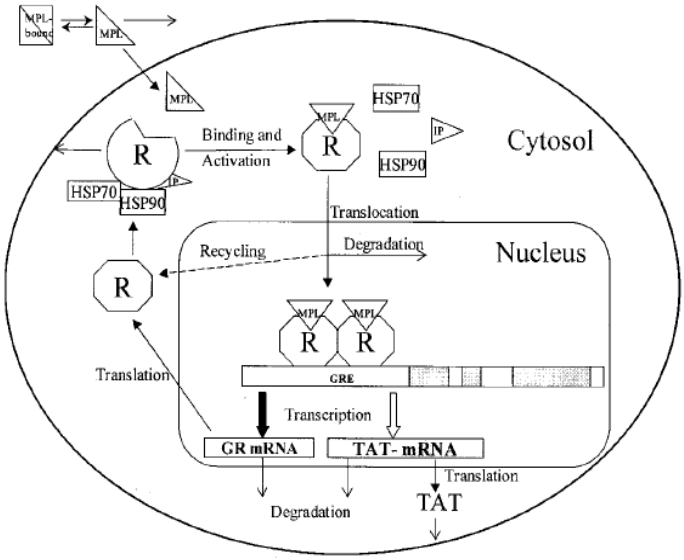


Figure 3 – Schematic representation of the cellular/molecular mechanism of steroid action in the hepatocyte. The thick open and solid arrows indicate induction and repression of gene transcription (reprinted with permission from 23)

Whilst there are relatively few examples specific to safety pharmacology and toxicology, Berger et al. have recently shown how a systems pharmacology approach can be used to characterise and predict QT prolongation (111). Their work shows that the QT prolonging effects of multiple drugs targeting very different indications could be estimated with a network analysis approach. They were also able to account for multiple off-target binding sites for each drug showing that target related and off-target effects could be assessed within the same framework. It should be noted that the model was Boolean in nature, which implies the need for further refinement for quantitative toxicological predictions. Another interesting application has been shown by Timchalk et al., who illustrate the development of a model-based approach to describe the pharmacodynamics of cholinesterase inhibiting compounds (25). Their model accounts for the synthesis and loss rates of the enzyme in vivo, enabling prediction of the brain synthesis.

Still in the realm of systems pharmacology, attention should be given to the contribution of mechanism-based models describing biomarkers as predictors of drug response, even when the underlying toxicological mechanisms are poorly understood. Increasing emphasis has been given by academic researchers and regulators on the relevance of biomarker selection and early risk prediction (112-114). In fact, the use of biomarkers to predict liver and kidney toxicity has been the subject of numerous public-private initiatives. Unfortunately, little has been done to integrate biomarkers as covariates or into PKPD models. In principle, one could consider the prediction of acute and chronic toxicity by parameterising biomarker response in a similar manner to what creatinine clearance currently represents in renal impairment.

In addition to the development of more physiological, mechanism-base models, another avenue for future extension of the proposed methodology in this thesis lies in risk-benefit analysis. There are numerous examples of risk-benefit assessment in the published literature in which pharmacological and physiological models have been applied. Yassen et al. (2008) performed an analysis on buprenorphine and fentanyl to assess risk benefit for antinociceptive and respiratory depressant effects (115). The development of a population PKPD model enabled both effects to be probabilistically modelled as a function of the predicted biophase concentrations. By constructing a clearly defined utility function, they were able to obtain therapeutic indices consistent with known literature at the time. The most difficult hurdle to overcome with the acceptance of utility functions is in demonstrating construct validity. Ultimately, it should function as a mathematical description of the subjective risk-benefit criteria held by patients and physicians. Methods to assess the degree of construct validity however are not currently well established and widespread acceptance of utility functions to define therapeutic windows is still lacking. When utility functions are too subjective, overlaying exposure-benefit and exposure-risk relationships will possibly aid in the selection of safe and efficacious doses.

The advantages of quantitative models in toxicology are unquestionable, as they facilitate the characterisation of exposure, biomarkers, and pharmacodynamics both at organ, tissue and cellular levels. However, a model can only be validated for its predictive performance for

some aspects/ modules, but not for other due to the difficulty in obtaining experimental measures (e.g., free concentration in a given organ). Yet, the primary advantage of a biologically based model is the possibility to make predictions of variables that cannot be easily accessed with the available methodologies or that are impossible to measure in an intact biological system using current technologies.

Efforts must therefore be made to define the endpoints as well as the purpose of the biological model even before its development. In this context, one last aspect that deserves further attention is the need to replace uncertainty factors by a more formal, systematic measure of the lack of construct validity or discrepancy between experimental conditions and the expected therapeutic use of a drug. The predictive performance of a model must include the uncertainty about the model itself (e.g., identifiability) and about the translational gap (e.g., differences between species or experimental conditions). A central premise of toxicology has been that adverse effect are examined on the basis of higher doses and then extrapolated to lower doses. There is enough evidence showing that responses occurring a lower exposure may not be predicted from higher doses when homeostatic regulation (e.g., oscillatory, antagonistic balance) is involved (7,31). Dose and time considerations in the development and use of a drug are important for assessing actions and side effects, as well as predictions of safety and toxicity. We believe that lack of observance of this axiom will probably be the main source of uncertainty in any integrative approach, such as proposed throughout this thesis. This point has been raised by the Swiss-German physician, Theophrastus of Hohenheim in 1538, who stated that all things are poison and nothing is without poison: only the dose makes a thing not to be poison (20). (Figure 4)

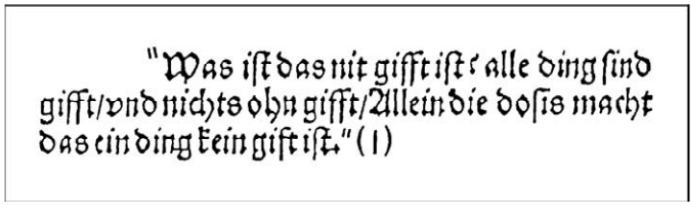


Figure 4 – Statement by the Swiss-German physician, Theophrastus of Hohenheim (Paracelsus). What is not a poison? All things are poison and nothing is without poison. Only the dose makes a thing not to be poison (reprinted with permission from 20).

Given the increased relevance of evidence synthesis as the basis for decision-making within regulatory and clinical practice, we anticipate that some of the meta-analytical elements presented across the various simulation scenarios will become embedded into daily practice in safety pharmacology and toxicology. Irrespective of the degree of understanding of the mechanisms of toxicity, a model-based approach appears to outperform standard methods for the prediction of the safe drug exposure of novel molecules in early drug development, especially those events that show low frequency or have delayed onset. Despite the narrow scope of the scenarios and limitations intrinsic to the selected experimental protocols presented in this thesis, our findings raise a new, potentially even more important question regarding the ethical basis for using empirical protocols in safety pharmacology and toxicology.

Reference List

1. Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nat.Rev.Drug Discov. **3**, 711-715.

2. U.S.Food and Drug Administration (CDER). Innovation or stagnation? Challenge and opportunity on the critical path to new medical products. 2004.
URL:
http://www.fda.gov/scienceresearch/specialtopics/criticalpathinitiative/criticalpathopportunitiesreports/ucm077262.htm

3. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul.Toxicol.Pharmacol. **32**, 56-67.

4. Russell, W. M. S. and BURCH, R. L. The principles of humane experimental technique. 1-4-1959. London: Methuen & Co. Ltd.

5. Suntharalingam, G., Perry, M. R., Ward, S., Brett, S. J., Castello-Cortes, A., Brunner, M. D., and Panoskaltsis, N. (2006). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. N.Engl.J.Med. **355**, 1018-1028.

6. Expert Scientific Group. Expert Scientific Group on Phase One Clinical Trials. 2006.

7. European Medicines Agency (CHMP). Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products. 2007.
URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002988.pdf

8. Kenter, M. J., and Cohen, A. F. (2006). Establishing risk of human experimentation with drugs: lessons from TGN1412. Lancet **368**, 1387-1391.

9. Dowsing, T., and Kendall, M. J. (2007). The Northwick Park tragedy--protecting healthy volunteers in future first-in-man trials. J.Clin.Pharm.Ther. **32**, 203-207.

10. Stebbings, R., Poole, S., and Thorpe, R. (2009). Safety of biologics, lessons learnt from TGN1412. Curr.Opin.Biotechnol. **20**, 673-677.

11. Breckenridge, A. (2010). Regulatory challenges, reimbursement, and risk-benefit assessment. Clin.Pharmacol.Ther. **88**, 153-154.

12. Krumholz, H. M., Ross, J. S., Presler, A. H., and Egilman, D. S. (2007). What have we learnt from Vioxx? BMJ **334**, 120-123.

13. Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000). Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N.Engl.J.Med. **343**, 1520-8, 2.

14. Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002). Timing of new black box warnings and withdrawals for prescription medications. JAMA **287**, 2215-2220.

15. Cross, J., Lee, H., Westelinck, A., Nelson, J., Grudzinskas, C., and Peck, C. (2002). Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980-1999. Pharmacoepidemiol.Drug Saf **11**, 439-446.

16. Heerdink, E. R., Urquhart, J., and Leufkens, H. G. (2002). Changes in prescribed drug doses after market introduction. Pharmacoepidemiol.Drug Saf **11**, 447-453.

17. Lowe, P. J., Hijazi, Y., Luttringer, O., Yin, H., Sarangapani, R., and Howard, D. (2007). On the anticipation of the human dose in first-in-man trials from preclinical and prior clinical information in early drug development. Xenobiotica **37**, 1331-1354.

18. Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., and Wurtzen, G. (2002). Mathematical modelling and quantitative methods. Food Chem.Toxicol. **40**, 283-326.

19. Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A. G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., and Younes, M. (2002). Hazard characterisation of chemicals in food and diet. dose response, mechanisms and extrapolation issues. Food Chem.Toxicol. **40**, 237-282.

20. Slob, W. (1999). Thresholds in toxicology and risk assessment. International Journal of Toxicology **18**, 345-367.

21. Kroes, R., Kleiner, J., and Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. Toxicol.Sci. **86**, 226-230.

22. Dorato, M. A., and Engelhardt, J. A. (2005). The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul.Toxicol.Pharmacol. **42**, 265-274.

23. U.S.Food and Drug Administration (CDER). Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. 2005.
URL: http://www.fda.gov/downloads/Drugs/Guidance/UCM078932.pdf

24. Agoram, B. M. (2009). Use of pharmacokinetic/ pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics. Br.J.Clin.Pharmacol. **67**, 153-160.

25. Muller, P. Y., Milton, M., Lloyd, P., Sims, J., and Brennan, F. R. (2009). The minimum anticipated biological effect level (MABEL) for selection of first human dose in clinical trials with monoclonal antibodies. Curr.Opin.Biotechnol. **20**, 722-729.

26. Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Regul.Toxicol.Pharmacol. **3**, 224-238.

27. Swartout, J. C., Price, P. S., Dourson, M. L., Carlson-Lynch, H. L., and Keenan, R. E. (1998). A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. **18**, 271-282.

28. Gaylor, D. W., and Kodell, R. L. (2000). Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal. **20**, 245-250.

29. Shih, C., Gosset, L., and Gates, S. e. a. (1996). LY231514 and its polyglutamates exhibit potent inhibition against both human dihydrofolate reductase (DHFR) and thymidylate synthase (TS): multiple folate enzyme inhibition. Annals of Oncology.

30. Latz, J. E., Rusthoven, J. J., Karlsson, M. O., Ghosh, A., and Johnson, R. D. (2006). Clinical application of a semimechanistic-physiologic population PK/PD model for neutropenia following pemetrexed therapy. Cancer Chemother.Pharmacol. **57**, 427-435.

31. Jeste, D. V., Lacro, J. P., Palmer, B., Rockwell, E., Harris, M. J., and Caligiuri, M. P. (1999). Incidence of tardive dyskinesia in early stages of low-dose treatment with typical neuroleptics in older patients. Am.J.Psychiatry **156**, 309-311.

32. Danhof, M., de Jongh, J., de Lange, E. C., Della, P. O., Ploeger, B. A., and Voskuyl, R. A. (2007). Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. Annu.Rev.Pharmacol.Toxicol. **47**, 357-400.

33. Danhof, M., de Lange, E. C., Della Pasqua, O. E., Ploeger, B. A., and Voskuyl, R. A. (2008). Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. Trends Pharmacol.Sci. **29**, 186-191.

34. Danhof, M., Alvan, G., Dahl, S. G., Kuhlmann, J., and Paintaud, G. (2005). Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. Pharm.Res. **22**, 1432-1437.

35. Rolan, P., Danhof, M., Stanski, D., and Peck, C. (2007). Current issues relating to drug safety especially with regard to the use of biomarkers: a meeting report and progress update. Eur.J.Pharm.Sci. **30**, 107-112.

36. Jorga, K. M., Fotteler, B., Heizmann, P., and Zurcher, G. (1998). Pharmacokinetics and pharmacodynamics after oral and intravenous administration of tolcapone, a novel adjunct to Parkinson's disease therapy. Eur.J.Clin.Pharmacol. **54**, 443-447.

37. Bakri, R., Wang, J., Wierzbicki, A. S., and Goldsmith, D. (2003). Cerivastatin monotherapy-induced muscle weakness, rhabdomyolysis and acute renal failure. Int.J.Cardiol. **91**, 107-109.

38. Wernevik, L. C., Nystrom, P., Johnsson, G., Nakanishi, T., and Eriksson, U. G. (2006). Pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor ximelagatran in young healthy Japanese men. Clin.Pharmacokinet. **45**, 77-84.

39. Cohen, A. (2007). Should we tolerate tolerability as an objective in early drug development? Br.J.Clin.Pharmacol. **64**, 249-252.

40. Oliveira, R. B., Gomes-Leal, W., do-Nascimento, J. L., and Picanco-Diniz, C. W. (1998). Methylmercury intoxication and histochemical demonstration of NADPH-diaphorase activity in the striate cortex of adult cats. Braz.J.Med.Biol.Res. **31**, 1157-1161.

41. Barnes, D. G., Daston, G. P., Evans, J. S., Jarabek, A. M., Kavlock, R. J., Kimmel, C. A., Park, C., and Spitzer, H. L. (1995). Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul.Toxicol.Pharmacol. **21**, 296-306.

42. Budtz-Jorgensen, E., Grandjean, P., Keiding, N., White, R. F., and Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol.Lett. **112-113**, 193-199.

43. Lalonde, R. L., Kowalski, K. G., Hutmacher, M. M., Ewy, W., Nichols, D. J., Milligan, P. A., Corrigan, B. W., Lockwood, P. A., Marshall, S. A., Benincosa, L. J., Tensfeldt, T. G., Parivar, K., Amantea, M., Glue, P., Koide, H., and Miller, R. (2007). Model-based drug development. Clin.Pharmacol.Ther. **82**, 21-32.

44. Kodell, R. L., and Chen, J. J. (2007). On the use of hierarchical probabilistic models for characterizing and managing uncertainty in risk/safety assessment. Risk Anal. **27**, 433-437.

45. West, G. B., Brown, J. H., and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. Science **276**, 122-126.

46. Zuideveld, K. P., Van der Graaf, P. H., Peletier, L. A., and Danhof, M. (2007). Allometric scaling of pharmacodynamic responses: application to 5-Ht1A receptor mediated responses from rat to man. Pharm.Res. **24**, 2031-2039.

47. Mager, D. E., Woo, S., and Jusko, W. J. (2009). Scaling pharmacodynamics from in vitro and preclinical animal studies to humans. Drug Metab Pharmacokinet. **24**, 16-24.

48. Bachmann, K., Pardoe, D., and White, D. (1996). Scaling basic toxicokinetic parameters from rat to man. Environ.Health Perspect. **104**, 400-407.

49. Langdon, G., Gueorguieva, I., Aarons, L., and Karlsson, M. (2007). Linking preclinical and clinical whole-body physiologically based pharmacokinetic models with prior distributions in NONMEM. Eur.J.Clin.Pharmacol. **63**, 485-498.

50. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA **279**, 1200-1205.

51. Ju, C., and Uetrecht, J. P. (2002). Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. Curr.Drug Metab **3**, 367-377.

52. Haley, P. J. (2003). Species differences in the structure and function of the immune system. Toxicology **188**, 49-71.

53. European Medicines Agency (CHMP). Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007.
URL:
http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf

54. Davidian, M., and Giltinan, D. M. (1993). Analysis of repeated measurement data using nonlinear mixed effects models. Chemom.Intell.LAb.Syst. **20**, 1-24.

55. Karlsson, M. O., Beal, S. L., and Sheiner, L. B. (1995). Three new residual error models for population PK/PD analyses. J.Pharmacokinet.Biopharm. **23**, 651-672.

56. Beal, S. L. (1984). Population pharmacokinetic data and parameter estimation based on their first two statistical moments. Drug Metab Rev. **15**, 173-193.

57. Lindstrom, M. L., and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. Biometrics **46**, 673-687.

58. Pinheiro, J. C., and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. J.Comput.Graph.Stat. **4**, 12-35.

59. Baverel, P. G., Savic, R. M., Wilkins, J. J., and Karlsson, M. O. (2009). Evaluation of the nonparametric estimation method in NONMEM VI: application to real data. J Pharmacokinet.Pharmacodyn. **36**, 297-315.

60. Gisleskog, P. O., Karlsson, M. O., and Beal, S. L. (2002). Use of prior information to stabilize a population data analysis. J.Pharmacokinet.Pharmacodyn. **29**, 473-505.

61. Lunn, D. J., Best, N., Thomas, A., Wakefield, J., and Spiegelhalter, D. (2002). Bayesian analysis of population PK/PD models: general concepts and software. J.Pharmacokinet.Pharmacodyn. **29**, 271-307.

62. Kuhn E., and Lavielle M. (2004). Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis **49**, 1020-1038.

63. Bois, F. Y. (2010). Physiologically based modelling and prediction of drug interactions. Basic Clin.Pharmacol.Toxicol. **106**, 154-161.

64. Lunn, D. J. (2008). Automated covariate selection and Bayesian model averaging in population PK/PD models. J.Pharmacokinet.Pharmacodyn. **35**, 85-100.

65. Wade, J. R., Beal, S. L., and Sambol, N. C. (1994). Interaction between structural, statistical, and covariate models in population pharmacokinetic analysis. J Pharmacokinet.Biopharm. **22**, 165-177.

66. Mentre, F., and Mallet, A. (1994). Handling covariates in population pharmacokinetics. Int.J Biomed.Comput **36**, 25-33.

67. Mandema, J. W., Verotta, D., and Sheiner, L. B. (1992). Building population pharmacokinetic--pharmacodynamic models. I. Models for covariate effects. J Pharmacokinet.Biopharm. **20**, 511-528.

68. Holford, N. H., Kimko, H. C., Monteleone, J. P., and Peck, C. C. (2000). Simulation of clinical trials. Annu.Rev.Pharmacol.Toxicol. **40**, 209-234.

69. Holford, N., Ma, S. C., and Ploeger, B. A. (2010). Clinical trial simulation: a review. Clin.Pharmacol.Ther. **88**, 166-182.

70. Mentre, F., Duffull, S., Gueorguieva, I., Hooker, A., Leonov, S., Ogungbenro, K., and Retout, S. Software for optimal design in population pharmacokinetics and pharmacodynamics: a comparison. PAGE. 2007.

71. Dorato, M. A., and Engelhardt, J. A. (2005). The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul.Toxicol.Pharmacol. **42**, 265-274.

72. U.S.Food and Drug Administration. Guidance for Industry M3(R2) Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. 1-1-2010.

URL:

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073246.pdf

73. Tse, F. L., and Nedelman, J. R. (1996). Serial versus sparse sampling in toxicokinetic studies. Pharm.Res. **13**, 1105-1108.

74. Millard, S., and Krause, A. (2001). Applied statistics in the pharmaceutical industry: with case studies using S-Plus. Springer-Verlag.

75. Hing, J. P., Woolfrey, S. G., Greenslade, D., and Wright, P. M. (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.

76. Gagnon, R. C., and Peterson, J. J. (1998). Estimation of confidence intervals for area under the curve from destructively obtained pharmacokinetic data. J.Pharmacokinet.Biopharm. **26**, 87-102.

77. Nedelman, J. R., Gibiansky, E., and Lau, D. T. (1995). Applying Bailer's method for AUC confidence intervals to sparse sampling. Pharm.Res. **12**, 124-128.

78. Bonate, P. L. (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.

79. ICH. Toxicokinetics: a guidance for assessing systemic exposure in toxicology studies. 1-3-1995.

URL: http://www.emea.eu.int

80. Pai, S. M., Fettner, S. H., Hajian, G., Cayen, M. N., and Batra, V. K. (1996). Characterization of AUCs from sparsely sampled populations in toxicology studies. Pharm.Res. **13**, 1283-1290.

81. Aronson, J. K., and Ferner, R. E. (2003). Joining the DoTS: new approach to classifying adverse drug reactions. BMJ **327**, 1222-1225.

82. Contrera, J. F., Matthews, E. J., Kruhlak, N. L., and Benz, R. D. (2004). Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul.Toxicol.Pharmacol. **40**, 185-206.

83. Retout, S., Duffull, S., and Mentre, F. (2001). Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. Comput.Methods Programs Biomed. **65**, 141-151.

84. Waterhouse, T. H., Redmann, S., Duffull, S. B., and Eccleston, J. A. (2005). Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. J.Pharmacokinet.Pharmacodyn. **32**, 521-545.

85. Duffull, S. B., Mentre, F., and Aarons, L. (2001). Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm.Res. **18**, 83-89.

86. Ette, E. I., Howie, C. A., Kelman, A. W., and Whiting, B. (1995). Experimental design and efficient parameter estimation in preclinical pharmacokinetic studies. Pharm.Res. **12**, 729-737.

87. Hooker, A. C., Foracchia, M., Dodds, M. G., and Vicini, P. (2003). An evaluation of population D-optimal designs via pharmacokinetic simulations. Ann.Biomed.Eng **31**, 98-111.

88. Foracchia, M., Hooker, A., Vicini, P., and Ruggeri, A. (2004). POPED, a software for optimal experiment design in population kinetics. Comput.Methods Programs Biomed. **74**, 29-46.

89. Retout, S., Mentre, F., and Bruno, R. (2002). Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Stat.Med. **21**, 2623-2639.

90. Ogungbenro, K., Gueorguieva, I., Majid, O., Graham, G., and Aarons, L. (2007). Optimal design for multiresponse pharmacokinetic-pharmacodynamic models - dealing with unbalanced designs. J.Pharmacokinet.Pharmacodyn. **34**, 313-331.

91. Huang, C., Zheng, M., Yang, Z., Rodrigues, A. D., and Marathe, P. (2008). Projection of exposure and efficacious dose prior to first-in-human studies: how successful have we been? Pharm.Res. **25**, 713-726.

92. Gibbs, J. P. (2010). Prediction of exposure-response relationships to support first-in-human study design. AAPS.J. **12**, 750-758.

93. Budtz-Jorgensen, E., Grandjean, P., Keiding, N., White, R. F., and Weihe, P. (2000). Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol.Lett. **112-113**, 193-199.

94. Nyberg, J., Karlsson, M. O., and Hooker, A. C. (2009). Simultaneous optimal experimental design on dose and sample times. J.Pharmacokinet.Pharmacodyn. **36**, 125-145.

95. D'Argenio, D. Z. (1981). Optimal sampling times for pharmacokinetic experiments. J.Pharmacokinet.Biopharm. **9**, 739-756.

96. al Banna, M. K., Kelman, A. W., and Whiting, B. (1990). Experimental design and efficient parameter estimation in population pharmacokinetics. J.Pharmacokinet.Biopharm. **18**, 347-360.

97. Wang, J., and Endrenyi, L. (1992). A computationally efficient approach for the design of population pharmacokinetic studies. J.Pharmacokinet.Biopharm. **20**, 279-294.

98. Thai, H., Mentre, F., Holford, N. H., Veyrat-Follet, C., and Comets, E. Residual-based bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models with heteroscedastic error. PAGE meeting . 2012.
URL: http://www.page-meeting.org/pdf_assets/2550-Thai_poster_Page12_vfinal.pdf

99. Hooker, A., and Vicini, P. (2005). Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. AAPS.J. **7**, E759-E785.

100. Edler, D., Glimelius, B., Hallstrom, M., Jakobsen, A., Johnston, P. G., Magnusson, I., Ragnhammar, P., and Blomgren, H. (2002). Thymidylate synthase expression in colorectal cancer: a prognostic and predictive marker of benefit from adjuvant fluorouracil-based chemotherapy. J.Clin.Oncol. **20**, 1721-1728.

101. European Medicines Agency. Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. 19-1-2007.

102. Vane, J. R. (1971). Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. Nat.New Biol. **231**, 232-235.

103. Langman, M. J., Jensen, D. M., Watson, D. J., Harper, S. E., Zhao, P. L., Quan, H., Bolognese, J. A., and Simon, T. J. (1999). Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA **282**, 1929-1933.

104. Hooker, A. C., Staatz, C. E., and Karlsson, M. O. (2007). Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. Pharm.Res. **24**, 2187-2197.

105. Zhang, L., Beal, S. L., and Sheiner, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J.Pharmacokinet.Pharmacodyn. **30**, 387-404.

106. Zhang, L., Beal, S. L., and Sheinerz, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data II: robustness of methods. J.Pharmacokinet.Pharmacodyn. **30**, 405-416.

107. Lindbom, L., Pihlgren, P., and Jonsson, E. N. (2005). PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput.Methods Programs Biomed. **79**, 241-257.

108. Capone, M. L., Tacconelli, S., Sciulli, M. G., Anzellotti, P., Di, F. L., Merciaro, G., Di, G. P., and Patrignani, P. (2007). Human pharmacology of naproxen sodium. J.Pharmacol.Exp.Ther. **322**, 453-460.

109. Konstam, M. A., Weir, M. R., Reicin, A., Shapiro, D., Sperling, R. S., Barr, E., and Gertz, B. J. (2001). Cardiovascular thrombotic events in controlled, clinical trials of rofecoxib. Circulation **104**, 2280-2288.

110. Zhao, S., and Iyengar, R. (2012). Systems Pharmacology: Network Analysis to Identify Multiscale Mechanisms of Drug Action. Annual Review of Pharmacology and Toxicology **52**, 505-521.

111. Berger, S. I., Ma'ayan, A., and Iyengar, R. (2010). Systems Pharmacology of Arrhythmias. Science Signaling **3**, ra30.

112. Berger, S. I., and Iyengar, R. (2011). Role of systems pharmacology in understanding drug adverse events. Wiley Interdisciplinary Reviews: Systems Biology and Medicine **3**, 129-135.

113. Bai, J. P. F., and Abernethy, D. R. (2013). Systems Pharmacology to Predict Drug Toxicity: Integration Across Levels of Biological Organization*. Annual Review of Pharmacology and Toxicology **53**, 451-473.

114. van der Greef, J., and McBurney, R. N. (2005). Innovation: Rescuing drug discovery: in vivo systems pathology and systems pharmacology. Nat.Rev.Drug Discov. **4**, 961-967.

115. Yassen, A., Olofsen, E., Kan, J., Dahan, A., and Danhof, M. (2008). PharmacokineticΓÇôPharmacodynamic Modeling of the Effectiveness and Safety of Buprenorphine and Fentanyl in Rats. Pharm Res **25**, 183-193.

CHAPTER 10

Nederlandse Samenvatting (Synopsis in Dutch)

De toxiciteit en veiligheid van nieuwe geneesmiddelen wordt voornamelijk bestudeerd in proefdier modellen. De resultaten verkregen met deze modellen worden vervolgens geëxtrapoleerd om bijwerkingen in de mens te voorspellen. Bij dit onderzoek richt men zich vooral op lever en nierschade, schade aan het oog, immuno- en genotocixiteit alsmede aan carcinogeniteit. Daarbij wordt slechts beperkt aandacht besteed aan de analyse van de relatie tussen de blootstelling (in termen van het beloop van de geneesmidelconcentratie in het lichaam) en de toxiciteit. Zelfs in het geval dat resultaten van het veiligheidsonderzoek bruikbaar zijn voor een dergelijke analyse, is het uitvoeren van dit type experimenten tijdrovend en zijn extrapolaties van dier naar de mens vaak niet robuust, accuraat en/of precies. Een aantal van deze beperkingen is blijven bestaan ondanks het feit dat toxicokinetiek wordt beschouwd als een essentieel onderdeel van de beoordeling van de veiligheid van nieuwe farmaca.

In dit proefschrift laten we zien dat het mogelijk is om kennis uit farmacologische experimenten te combineren met specifieke informatie uit *in vitro* test systemen en computer modellen om zowel de werkzaamheid als de veiligheid *in vivo* te voorspellen. Daarbij is het ook mogelijk om verstoringen *in vitro* te bestuderen die de oorzaak zijn van bijwerkingen van een geneesmiddel of die daaraan een bijdrage leveren. Dit heeft als belangrijk extra voordeel dat de focus van traditionele toxicologie studies, waar risico's van een hoge dosis van een stof *in vivo* bestudeerd worden, vervangen kunnen worden door experimentele protocollen waarin de de veiligheid van geneesmiddelen wordt bestudeerd bij een klinisch relevante blootstelling.

Op dit moment zijn er methodologische en conceptuele ontwikkelingen voor het bestuderen van de veiligheid en de toxiciteit van nieuwe farmaca, waardoor de risico's van

het gebruik bij de therapeutische toepassing efficiënter kunnen worden gekarakteriseerd. Helaas worden deze methoden nog niet vaak toegepast om bij de registratie van nieuwe geneesmiddelen, belangrijke vragen op het gebied van de veiligheid en toxiciteit te beantwoorden. De betekenis van kwantitatieve concepten is eerder aangetoond oa. voor de structuur-werkingsrelaties, als basis voor het voorspellen van de toxiciteit van een chemische verbinding onder bepaalde omstandigheden. Op basis daarvan kan men uitspraken doen over de mogelijke effecten van een behandeling zelfs wanneer de blootstelling in de mens erg laag is en toxicokinetische gegevens niet gemakkelijk verkregen kunnen worden.

Door het ontbreken van een sterke farmacologische basis voor de experimentele protocollen heeft de implementatie van een modelmatige benadering voor het karakteriseren van de veiligheids- en toxiciteitsprofiel van een nieuw geneesmiddel nooit op grote schaal plaatsgevonden. Daardoor is de analyse van de relatie tussen blootstelling en effect nog steeds niet het primaire doel van de desbetreffende protocollen. Men gaat ervan uit dat veiligheid gekarakteriseerd kan worden door een veiligheidsdrempel. Ook het concept van op fysiologie-gebaseerde farmacokinetische (PBPK) modellering voor het voorspellen van de blootstelling wordt nog niet veel toegepast en wanneer het wordt toegepast is dit voornamelijk op het terrein van milieu-toxicologische vraagstukken en in mindere mate voor de ontwikkeling van geneesmiddelen.

Methoden die veiligheid en de toxicologische effecten van geneesmiddelen in de mens kunnen voorspellen of op basis waarvan een vertaling vanuit *in vivo* diermodellen of *in-vitro* modellen naar de mens kan worden gemaakt, zijn van erg groot belang. Er is een grote vraag naar de ontwikkeling van deze methoden, ongeacht de huidige richtlijen voor toxicologisch onderzoek of de urgentie voor verandering van de eisen die gesteld worden door de registratie autoriteiten. Om dit te kunnen bewerkstelligen is er meer nodig dan de ontwikkeling nieuwe experimentele protocollen en technologieën. Een integrale benadering waar efficiënt gebruik wordt gemaakt van de beschikbare informatie en de toepassing van farmacokinetische-farmacodynamische modellering kan van grote betekenis zijn. Hierbij is

het essentieel dat ook bijwerkingen die op lange termijn optreden en zeldzame bijwerkingen bestudeerd kunnen worden, omdat die nog steeds een belangrijke oorzaak zijn van het falen van nieuwe medicijnen. Daarnaast ontbreekt er ook een geïntegreerde benadering om de juiste eindpunten te kiezen om zo de risico's die het gebruik van het geneesmiddel met zich meebrengt nauwkeurig en precies te kunnen beoordelen.

De huidige toxiciteit testen zijn voornamelijk bedoeld om de gevaren van het gebruik van een geneesmiddel op supraherapeutisch niveau te kunnen bestuderen. De data die wordt gegenereerd op basis van de bestaande richtlijnen is daardoor niet altijd geschikt om de relatie tussen blootstelling en effect vast te stellen. Noch om deze relatie op basis van wiskundige modellen te kunnen karakteriseren. Om de relatie tussen blootstelling en respons te bestuderen zou er eerst aan enkele voorwaarden moeten worden voldaan. Ten eerste, zouden verschillende doses moeten worden toegediend, waardoor een breed bereik kan worden verkregen in termen van blootstelling niveaus en respons. Daarbij moet men ook de samenhang tussen de primaire en secondaire farmacologische mechanismen en de gewenste en ongewenste effecten proberen te identificeren.

Het onderzoek dat is beschreven in dit proefschrift heeft betrekking op een aantal onderdelen in het proces van de ontwikkeling van nieuwe geneesmiddelen, die overwogen dienen te worden om dit efficiënter te laten verlopen. De nadruk ligt bij het efficiënt gebruik van data die worden verkregen bij het veiligheids- en toxiciteitsonderzoek. Vier onderzoeksvragen vormen de basis van het werk zoals gepresenteerd in de volgende hoofdstukken:

1. Kunnen experimentele protocollen voor veiligheids- en toxiciteitsevaluaties worden geoptimaliseerd om de relaties tussen farmacokinetiek en farmacodynamiek te karakteriseren?

2. Kan het gebruik van meta-analytische methoden gebaseerd op niet-lineair gemengde effecten modellen bijdragen aan een verhooging van de nauwkeurigheid en precisie van veiligheidsdrempels in vergelijking met de methoden die thans worden toegepast?

3. Kan een op mechanisme gebaseerd model worden gebruikt om de veilige blootstelling aan een geneesmiddel nauwkeuriger te definiëren, waardoor vertraagde (lange termijn) of zeldzame bijwerkingen voorspeld en voorkomen kunnen worden?

4. Kunnen biomarkers gecombineerd met farmacokinetische data bijdragen aan het vastellen van de veilige blootstelling bij langdurig gebruik van een geneesmiddel?

Het proefschrift heeft betrekking op de hierboven benoemde onderzoeksvragen zowel van conceptueel als van praktisch oogpunt.

Sectie I: Algemene inleiding

In **Hoofdstuk 1** wordt een overzicht gepresenteerd van de gangbare veiligheid toetsen en methoden die worden toegepast bij de ontwikkeling van nieuwe geneesmiddelen. Hierbij wordt er aandacht besteed aan de gevolgen van het genereren van empirisch bewijs. Vervolgens worden methodologische beperkingen voor het vastellen van de relatie tussen blootstelling en ongewenste effecten (bijwerkingen) geïdentificeerd op basis van voorbeelden. Tevens wordt een hiërarchische beslisboom ontwikkeld die het beslissingsproces weergeeft en de daarbij behorende experimentele data samenvat. Vanuit een theoretisch perspectief worden verschillende facetten van hetzelfde vraagstuk besproken, die gerelateerd kunnen worden aan vier aspecten van wetenschappelijk onderzoek, te weten: 1. optimalisatie 2. vertaling 3. analyse en 4. beslissingscriteria. De relevantie van deze punten voor de implementatie van experimentele protocollen en voor de schatting en interpretatie van parameters die de veiligheid en toxiciteit van een geneesmiddel beschrijven wordt apart besproken. We hebben laten zien dat foutieve voorspellingen van veiligheid kunnen ontstaan door gebruik te maken van empirische veiligheidsdrempelwaarden, indien die beschouwd worden als voorspellend voor toxiciteit of ongewenste effecten. Om rekening te kunnen houden met slechte precisie en vertekende nauwkeurigheid van deze methodes worden in de praktijk veiligheidsfactoren (ook wel bekend als onzekerheidsfactoren) geïmplementeerd bovenop de empirische criteria. We laten ook zien dat ondanks bovengenoemde beperkingen het gebruik van een

veiligheidsdrempel binnen de ontwikkeling van geneesmiddelen breed geaccepteerd is en gedetailleerd wordt beschreven in de richtlijnen van registratie autoriteiten.

Op basis van historische voorbeelden hebben wij een aantal belangrijke uitdagingen geïdentificeerd om het karakteriseren van een veiligheids profiel van een nieuw molecuul in een vroeg stadium mogelijk te maken. Voor de toepassing van farmacologische concepten in het onderzoek naar de veiligheid en de toxiciteit van nieuwe geneesmiddelen werd een nieuw paradigma voorgesteld.

In **Hoofdstuk 2** presenteren wij de conceptuele en methodologische aspecten die op de daaropvolgende hoofstukken worden uitgewerkt. Het doel van het onderzoek was om de uitvoerbaarheid van op farmacologische concepten gebaseerde analyse van toxicologische gegevens het risico van een behandeling in de mens te voorspellen, en waar mogelijk de uitkomsten te vergelijken met die verkregen met traditionele veiligheidstoetsen en methoden. Daarbij komt een fundamenteel verschil in het doel van de veiligheidsevaluatie aan de orde. Het doel van de nieuwe benadering is om de veiligheid van een nieuw geneesmiddel bij normaal gebruik vast te stellen. Dat is een belangrijk verschil met de huidige praktijk die erop gericht is om vast te stellen dat een bepaalde toxische limiet niet wordt overschreden. Vier belangrijke vragen die het wetenschappelijke raamwerk vormen van het onderzoek, worden in de hieropvolgende hoofdstukken besproken, tegen de achtergrond van de mogelijkheden voor optimalisatie en integratie van kennis. De in de praktijk gehanteerde experimentele protocollen vormden daarbij het uitgangspunt. Hoewel er, vanuit een farmacologische perspectief, overeenkomsten bestaan tussen de toetsing van de werkzaamheid enerzijds en toetsing van de toxiciteit anderzijds stellen wij uitsluitend een onderzoeksplan voor de analyse van toxicologische data voor. Verder hebben wij een vaak vergeten aspect van het pre-klinisch onderzoek gedefinïeerd: het ethische belang van het verfijnen, reduceren en vervangen van experimenten met proefdieren.

Sectie II: Conceptueel kader

De voordelen en beperkingen van een modelmatige benadering voor veiligheids- en toxiciteitsonderzoek werden geëvalueerd in **Sectie II**. Binnen een conceptueel kader laten

we zien hoe er rekening kan worden gehouden met de variabiliteit binnen en tussen patiënten en met de onzekerheid in de respons (bijwerkingen) door gebruik te maken van farmacokinetische en farmacokinetisch-farmacodynamische modellering in plaats van beschrijvende statistische methoden. Op basis van simulaties van experimentele protocollen wordt aangetoond hoe waarden van de blootstelling kunnen worden verkregen die fysiologisch relevant zijn voor het karakteriseren van vertraagde of late bijwerkingen. De nadruk lag hier vooral op de uitvoerbaarheid van het voorspellen van lange termijn bijwerkingen gebruik makend van de gegevens uit een studie met een korte duur. Verder wordt er een alternatieve manier beschreven om pre-klinische algemene toxicologische protocollen te ontwerpen en analyseren.

In eerste instantie werd er gefocused op niet-lineair gemengde effecten modellen als een data analyse methode voor de evaluatie van toxicokinetische gegevens en het vaststelllen van (parametrische) veiligheidsdrempels. In **hoofdstuk 3** werd voor drie hypothetische geneesmiddelen, met verschillende farmacokinetische eigenschappen, toxicokinetische data gesimuleerd voor dieren in de satellietgroepen van een algemene toxiciteitsstudie. Deze analyse maakte het mogelijk om de juistheid en nauwkeurigheid van zowel de veiligheidsdrempels als de secondaire farmacokinetische parameters zoals de opervlakte onder de concentratie vs. tijd curve (AUC) te vergelijken en de beperkingen van de traditionele niet-compartmentele analyse methode aan te tonen, ten opzichte van de resultaten die verkregen zijn op basis van populatie farmacokinetische modellen. De grootste foutmarges in de geschatte farmacokinetische parameters bleken intrinsiek verbonden te zijn met de niet-compartmentele analyse methode, vooral als de maximale concentratie (Cmax) geschat moeten worden. Deze resultaten laten ook zien dat de typische puntschatter afgeleid van empirische methoden om medicijn blootstelling te beschrijven een te grote mate van onzekerheid bevatten, die ongeïdentificeerd blijft. Dit onderstreept het belang van de toepassing van benaderingen voor het vaststellen van de onzekerheid in veiligheidsdrempels zoals de drempel voor het niet-ongewenst effect (no adverse effect level), oftewel de NOAEL. Met de simulaties werd aangetoond dat dit fenomeen kan worden omzeild door gebruikt te maken van modelmatige methoden, die de onzekerheid

parametrisch beschrijven ongeacht of deze als gevolg van een ongeschikt protocolontwerp, een grote biologische variabiliteit en/of onbekende covariaten voorkomt. De analyse op basis van de drie hypothetische geneesmiddelen heeft ook aangetoond hoe een farmacokinetisch model mogelijkheden biedt om de blootstelling op verschillende manieren parametrisch af te leiden. Een voorbeeld daarvan is de cumulatieve opervlak onder de concentratie vs. tijd curve (CAUC) bij de beschrijving van toxiciteit na langdurige behandeling, die ook gebruikt kan worden om voorspellingen te maken buiten de studie duur. Dit draagt bij aan een toename in kwaliteit van het beslissingsproces en uiteindelijk in de klinische interpretatie van de veiligheidsdrempels.

Data uit in de praktijk gehanteerde experimentele protocollen voor de evaluatie van algemene toxiciteit zijn niet geoptimaliseerd voor de analyse met behulp van modelmatige methoden, inclusief populatie farmacokinetisch modellen. De vraag is of protocollen zo kunnen worden geoptimaliseerd dat een vermindering van het aantal benodigde proefdieren kan worden bereikt zonder dat dit gepaard gaat met een verlies van de juistheid en de nauwkeurigheid van farmacokinetische parameters (zoals AUC en Cmax). Anders dan door gebuik te maken van bestaande optimalisatie software en algoritmes, waar de optimalisatie van het experimenteel ontwerp wordt bereikt op basis van de precisie van de primaire parameters, wordt in **Hoofdstuk 4** een methode voorgesteld die het mogelijk maakt om secundaire parameters te optimaliseren. Onze benadering bestond uit evaluaties van de Fisher informatiematrix gevolgd door berekeningen van verwachte juistheid of betrouwbaarheid van de secundaire parameter, zonder de nadelen van intensieve procedures zoals D-optimality. De resultaten laten zien dat de opzet van de studie, inclusief de keuze van de doses en het aantal monsters, grote invloed heeft op de juistheid van de parameter schattingen. De voorgestelde methode om de tijdstippen voor het nemen van bloedmonsters en de groepsgrootte te optimaliseren kan leiden tot een afname van het benodigde aantal proefdieren (ongeveer 30%) zonder verlies van de juistheid van de parameters die relevant zijn voor het karakteriseren van de veiligheidsdrempels. Deze verbetering was onafhankelijk van de verschillen in de farmacokinetiek sche profielen, zoals bestuudeerd op basis van de drie hypothetische geneesmiddelen. Tevens suggereert onze

analyse dat de juistheid van de farmacokinetische parameters niet door de groepsomvang wordt beïnvloedt maar dat het juist de tijdstippen van bloedafname en de toegediende doses zijn, die dit bepalen. Daarnaast laten deze resultaten zien dat de betrouwbaarheid van de secundaire parameters niet beïnvloed wordt als een aantal primaire parameters met onvoldoende nauwkeurigheid geschat wordt.

Hoewel het gebruik van een modelmatige benadering voor de evaluatie van de veiligheidsdrempels als gecompliceerd wordt gezien bleek onze methode juist te beschikken over een goed voorspellend vermogen zonder veel fouten of problemen met de identificeerbaarheid van het model en de daarbij behorende parameters. Verder laten we zien hoe bestaande kennis, inclusief data van andere experimenten, op een formele manier opgenomen kan worden tijdens de analyse en intrepretatie van de resultaten. De mogelijkheid om zgn. 'priors' te gebruiken voor het schatten van parameters in een model biedt vele kansen om oa. fysiologische factoren, die ten grondslag liggen aan farmacokinetische eigenschappen, mechanistisch te bestuderen. Men zou bijvoorbeeld de invloed van de verzadiging van metaboliserende enzymen kunnen evalueren, ongeacht de doses die gebruikt zijn tijdens een experiment.

Hoofdstuk 5 heeft betrekking op het gebruik van biomarkers in de evaluatie van de veiligheid en toxiciteit van nieuwe geneesmiddelen. Hier ligt de nadruk op het feit dat voor een nauwkeurige voorspelling van de veiligheid en toxiciteit na langdurig gebruik van een geneesmiddel, biomarkers belangrijke informatie kunnen opleveren voor het voorkomen van een bijwerking. In tegenstelling tot traditionele protocollen, die een directe relatie tussen de blootstelling en de bijwerking veronderstellen, hebben wij het voorgesteld om het nemen van bloedmonsters voor de bepaling van de farmacokinetiek te combineren met de bepaling van biomarkers. Het uiteindelijk doel van zo'n aanpak is het ontrafelen van de vaak vertraagde of indirecte relatie tussen de farmacokinetiek in plasma en de bijwerkingen of ongewenste effecten. Om de voordelen van een modelmatige aanpak te kunnen aantonen werden farmacokinetische en biomarker data gesimuleerd voor een aantal hypothetische geneesmiddelen met verschillende PKPD relaties. Nog eens proberen we de voordelen en

nadelen van de voorgestelde benadering te vergelijken met een analyse van de data op basis van een standaard niet-compartmentele methode. De analyse van de gesimuleerde data laat zien dat, op basis van gegevens verkregen met een traditioneel monsterafname schema, het echte onderliggende PKPD model vaak niet identificeerbaar is. Dat is in het bijzonder het geval wanneer er sprake is van een complexe PKPD relatie (o.m. door het optreden van *hysterese*). Niettemin werden in alle gevallen bevredigende resultaten verkregen met betrekking tot de geschatte biomarker concentraties. Wij zijn daarom van mening dat men zich meer bewust moet zijn van de beperkingen die de huidige experimentele protocollen met zich mee brengen. Dit is vooral van belang omdat de veiligheid van geneesmiddelen na langdurig gebruik steeds meer een prominente rol krijgt binnen de ontwikkeling en toelating van nieuwe medicijnen. Om de gevolgen van zo'n vertekening te beperken benadrukken wij de behoefte aan methoden die de mate van onzekerheid bepalen en daardoor modellen en voorspellingen betrouwbaarder kunnen maken.

Een andere bevinding van onze analyse is dat de criteria voor de selectie van parameters tijdens het ontwikkelen van een model niet alleen aan statistische eisen moeten voldoen maar ook de mogelijkheid moeten bieden om de fysiologische relevantie van bepaalde aannames te kunnen beoordelen. Duidelijke rapportage van alle modellen is van belang wanneer de ontwikkeling van het hierarchische model eindigt in een verzameling van modellen met vergelijkbare selectie criteria.

Het doel van het onderzoek dat in **Hoofdstuk 6** is beschreven was om de uitvoerbaarheid te evalueren van de integratie van mechanistische PKPD modellen met toxicologische gegevens uit standaard experimentele protocollen. Rekening houdende met de achtergrond incidentie van verschijnselen en fysiologische veranderingen die op lange termijn voorkomen en vaak met bijwerkingen kunnen worden verwisseld, werden vergelijkbare *in silico* methoden zoals eerder beschreven in hoofdstukken 3 en 5 toegepast om door geneesmiddelen geïnduceerde bijwerkingen te simuleren voor drie verschillende farmacologische mechanismen (directe werking, indirecte werking en irreversibele binding). Wij hebben ons hierbij geconcentreerd op zeldzame en chronische bijwerkingen die pas na

langdurige inname van een geneesmiddel ontstaan, inclusief het onderscheid tussen reversibele en irreversibele effecten, om de grootste methodologische uitdaging aan te gaan. Om er zeker van te zijn dat realistische condities werden gecreëerd zijn verschillende scenario's getoetst, namelijk: 1) de door het geneesmiddel geïnduceerde bijwerking en het achtergrond fysiologische verschijnsel waren niet van elkaar te onderscheiden. 2) het tijdsinterval tussen het begin van de behandeling en de diagnose was groot en de symptomen konden per proefdier maar een keer worden vastgesteld op basis van histologisch onderzoek (na afloop van het experiment). 3) de bijwerking kon beschreven worden als binaire respons. Onze resultaten lieten zien dat veiligheidsdrempels, zoals vastgesteld door de NOAEL, onbetrouwbaar en onnauwkeurig zijn. In twee van de drie scenario's, waar het effect van veiligheid en effectiviteit van de hypothetische geneesmiddelen was gesimuleerd, vonden wij dat het rigoureus toepassen van de NOAEL als beslissingscriterium zou leiden tot een foutieve classificatie in 50% van de gevallen. Dit zou vervolgens leiden tot een onterechte beeïndiging van de ontwikkeling van het geneesmiddel. Tijdens het vastellen van de haalbaarheid van de toepassing van modelmatige methoden voor de analyse van toxicologische gegevens hebben wij twee belangrijke componenten geïdentificeerd, die essentieel zijn voor de voorspelling van de geïnduceerde effecten: a) de beschikbaarheid van onafhankelijke informatie over achtergrond verschijnselen en verandering die meegeteld kunnen worden als bijwerkingen en b) het gebruik van Markov keten Monte Carlo (MCMC)-gebaseerde algorithmen. Zo hebben wij laten zien dat historische placebo data doeltreffend genoeg is om achtergrond verschijnselen en andere fysiologische veranaderingen te kunnen onderscheiden van de onderliggende farmacologische effecten en bijwerkingen. Aan de andere kant, toen de prestatie van statistische methoden geëvalueerd werd, vonden wij dat modellen die gebaseerd zijn op het 'maximum waarschijnlijkheid'kriterium onstabiel en onbetrouwbaar zijn. Daarentegen, bleken de MCMC-gebasseerde resultaten stabieler en nauwkeuriger, inclusief het schatten van de model- en parameteronzekerheid. Door gebruik te maken van het Bayes informatie criterium, oftewel BIC, om modellen te vergelijken konden wij de hoge mate van model onzekerheid blootleggen. Onze resultaten tonen aan dat traditionele technieken die gebruikt worden voor de selectie van een model en de daarbij behorende parameter verdelingen niet

geschikt waren om modellen met een zekere mate van complexiteit te identificeren. Het dient verder gezegd te worden dat gemiste bijwerkingen ook gemakkelijk gekwantificeerd konden worden middels de voorgestelde modelmatige aanpak. Door MCMC methoden toe te passen was het mogelijk om verschijnselen en bijwerkingen te beschrijven zelfs als deze niet waren waargenomen tijdens het experiment.

Het conceptuele raamwerk gepresenteerd in deze sectie van het proefcshrift draagt het bewijs aan voor de uitvoerbaarheid van een modelmatige benadering voor de evaluatie van de veiligheid en toxiciteit van nieuwe moleculen voordat ze in de mens worden getest. De integratie van farmacokinetische en farmacodynamische data als basis voor het voorspellen van de veilige blootstelling in de mens vereist enkele belangrijke voorwaarden waaraan de huidige methoden voor pre-klinische toxicologisch onderzoek niet voldoen. Een modelmatige benadering is daarvoor de geschikte oplossing. In plaats van het vastellen van de correlatie tussen één enkele waarde en de geschatte veiligheidsdrempel, biedt deze strategie het voordeel dat het klinische oordeel over de kans op toxiciteit wordt gebaseerd op de gehele relatie tussen blootstelling en bijwerkingen.

In het derde deel van dit proefschrift (**Sectie III**) hebben wij geprobeerd de implementatie van experimentele protocollen die voldoen aan de eisen van modelmatige dataanalyse methoden te illustreren op basis experimentele studies. Vanwege de lopende discussies over de risico's en de baten omtrent de chronische behandeling met niet-steroïdale ontstekingsremmers hebben wij naproxen gebruikt als voorbeeldstof om de bekende acute en chronische toxiciteit van deze klasse geneesmiddelen te evalueren. Gedurende de afgelopen jaren en tijdens de ontwikkeling van naproxen was men niet op de hoogte van het gebrek aan de selectiviteit van werking. Daarnaast ontbrak ook het bewijs voor de mechanismen, die de effecten na acuut gebruik (zoals bloedingen en maagzweren) en langdurige behandeling (zoals renale en cardiovasculaire schade) onderschrijven. Desalnietemin, laat onze analyse zien dat kennis over de farmacokinetiek en de blootstelling aan naproxen die bereikt wordt na de toediening van therapeutische doses van belang zijn voor het karakteriseren van het veiligheids- en toxiciteitsprofiel. Verder hebben wij ook

praktische uitdagingen geïdentificeerd, die men tegenkomt bij het gebruik van nieuwe experimentele protocollen.

Sectie III: Toepassing in de praktijk

In **Hoofdstuk 7** wordt een typisch toxicologie protocol gebruikt om de effecten van acute, middellange en langdurige behandeling aan naproxen te bestuderen. Naast het karakteriseren van de blootstelling in plasma worden twee biomarkers gemeten, namelijk thromboxane (TXB2) en prostaglandin (PGE2). Er werd vanuit gegaan dat gastro-intestinale bloedingen veroorzaakt werden door continue remming van cyclo-oxygenase 1 (COX-1), terwijl zweren voornamelijk door de inhibitie van cyclo-oxygenase 2 (COX-2) onstaan. Farmacokinetische en biomarker data werden geïntegreerd met gegevens uit de literatuur om de vermeende therapeutische effecten te correleren met zowel de blootstelling als de veranderingen in de biomarkers thromboxane en prostaglandin. De farmacokinetiek van naproxen werd het best beschreven met een een-compartiment model met eerste-orde absorptie. Er werd een niet-lineaire relatie vastgesteld tussen de dosis en biologische beschikbaarheid, die ertoe leidde dat de toename van de blootstelling in plasma minder dan proportinieel was met toenemende doses van naproxen. In tegenstelling tot eerdere bevindingen werd zweer vorming in de maag gezien bij alle doses (7.5, 15, 40, 80 mg/kg), waardoor geen NOAEL vastgesteld kon worden. Desalniettemin, hebben wij aangetoond dat de beschikbare experimentele data gebruikt kan worden om de onderliggende PKPD relaties for TXB2 en PGE2 te karakteriseren. Door middel van inibitiemodellen hebben we de dalende bloedspiegels van zowel thromboxane als prostaglandin kunnen correleren met naproxen concentraties. Er waren geen problemen met de identificeerbaarheid van de modellen en de schatting van parameters was precies, overeenkomend met de resultaten van hoofdstukken 3 en 5. Daarnaast toonden onze resultaten aan, dat farmacokinetische-pharmacodynamische relaties het mogelijk maken om potentiele verschillen tussen diersoorten te onderscheiden en desnoods daarvoor te corrigeren. De bijwerkingen van naproxen in de mens kunnen worden voorspeld op basis van pre-klinische data mits rekening wordt gehouden met de onderliggende blootstelling-effect relaties. Eigenlijk hebben wij gevonden dat bij de rat de PKPD relaties ongeveer tienvoud afwijken van de

waardes die verkregen zijn voor de mens. Hierdoor ontstaat de vraag over de vertaling naar de mens en de geschiktheid van de zogenoemde “meest gevoelige species” bij toxicologisch onderzoek als de gevoeligheid van ratten voor cyclo-oxygenase inhibitie aanzienlijk verschilt van de mens.

De toepassing van een parametrische aanpak en meer specifiek, van niet-lineair gemengde effecten modellen om experimenten te ontwerpen en de keuze voor een dosis te kunnen onderbouwen kan een belangrijke stap zijn voor translationeel toxicologisch onderzoek vanuit zowel het biologisch als statistisch perspectief. Tegelijkertijd hebben wij ook een aantal beprekingen kunnen vaststellen. De werkingsmechanismen van sommige geneesmiddelen zijn niet voldoende onderzocht of begrepen om op een parametrische manier vertaald te kunnen worden in een PKPD model, zelfs als biomarkers beschikbaar zijn. Daarnaast is het beschrijven van effecten die via meerdere pathways tot stand komen buitengewoon moeilijk. Zo’n situatie kan een significant probleem blootleggen dat niet voorkomen kan worden zonder extra experimentele data.

De uitvoerbaarheid en geschiktheid van een modelmatige benadering voor het bestuderen van het veiligheids- en toxiciteitsprofiel van een geneesmiddel is in **hoofdstuk 8** aangevuld door histologische data te integreren met de PKPD relaties die op het voorafgaande hoofdstuk zijn beschreven. Hier lieten wij zien dat het mogelijk is om farmacokinetische-farmacodynamische modellen, die de relatie tussen blootstelling en het effect of biomarkers beschrijven, te koppelen aan de analyse van de frequentie en intensiteit van bijwerkingen. In dit experiment werd de ernst van een zweer gecorreleerd met het oppervlak daarvan en als een continue variabele uitgedrukt. Daarnaast werd de frequentie van maagzweren als een discrete variabele geanalyseerd. De modelparameter die de frequentie van bijwerkingen beschreef wijst aan dat zweer vorming na toediening van naproxen een acuut effect is dat gepaard gaat met de maximale inhibitie van PGE_2, welke een maat is van de blokkade van COX-2. Gegeven de implicaties van model onzekerheid, zoals beschreven in hoofdstuk 6, zijn we gedwongen geweest om bootstrap methodes te gebruiken om zo de onzekerheid te kunnen schatten. De onzekerheid met betrekking tot de karakterisatie van de frequentie van

zweren was aanzienlijk lager dan wanneer de ernstigtheid van zweren werd beschreven. Ondanks de hoge mate van onzekerheid hebben we kunnen vaststellen dat het oppervlak van maagzweren het gevolg is van cumulatieve remming van TXB_2. Samengevat, ons onderzoek laat zien hoe een modelmatige benadering gebruikt kan worden om bijwerkingen die zowel na acute als langdurige behandeling voorkomen, vroeg te kunnen identificeren en zo pre-klinische data als basis te kunnen gebruiken voor de vertaling van het veiligheids en toxiciteitsprofiel naar de mens. Verder illustreren onze resultaten ook hoe PKPD relaties het mogelijk maken om conclusies te trekken over doses and dosis schema's die niet experimenteel geëvalueerd zijn.

Vanuit een methodologisch oogpunt hebben de bevindingen van dit onderzoeksprotocol de haalbaarheid en voordelen van een modelmatige benadering voor de evaluatie van chronische veiligheid en toxiciteit onderschreven. Hierbij dient echter de kanttekening te worden gemaakt, dat precisie, nauwkeurigheid en validiteit van de voorgestelde methoden nog altijd afhankelijk zijn van de experimentele data. In deze studie waren de bijwerkingen die wij beschreven hebben relatief frequent. Het karakteriseren van effecten die met lage frequentie voorkomen kan aanzienlijk moeilijker zijn, vooral als historische data ontbreken of waaruit ook maar enige gevolgtrekking gemaakt kan worden met betrekking tot klasse effecten. Wij erkennen ook dat de afwezigheid van zweer vorming in de controlegroep en het gebrek aan een cohort met lagere blootstelling een zwak punt in ons onderzoek is. Daardoor kon de echte basislijn voor het onstaan van maagzweren niet worden vastgesteld. Als gevolg daarvan was het ook niet mogelijk om de frequentie van bijwerkingen bij lagere dosissen nauwkeurig te voorspellen.

Sectie IV: Conclusies, aanbevelingen en perspectieven

Een overzicht van de resultaten en conclusies zoals beschreven in de verschillende hoofdstukken is samengevat in **Hoofdstuk 9**. Het meest belangrijk is dat aanbevelingen verstrekt zijn voor de analyse van veiligheids en toxicologie protocollen met behulp van farmacologisch gebaseerde kwantitatieve methoden. Hier hebben wij de antwoorden op de initiele vragen zoals beschreven aan het begin van dit proefschrift samengevat. Er is een lijst

is samengesteld van punten, die overwogen dienen te worden, gezien de uitdagingen en beperkingen die men tegenkomt bij het kakrakteriseren van relatie tussen blootstelling en effect wanneer er gebruik gemaakt wordt van traditionele toxicologische onderzoeksprotocollen. Deze lijst bevat aanbevelingen met betrekking tot methodologische en praktische aspecten die die ertoe leiden dat modelmatige data analyse technieken, toegepast kunnen worden bij veiligheids en toxicologisch onderzoek.

Noemenswaardig is de rol van het primaire (farmacologische) werkingsmechanisme en de daarbij betrokken receptorsystemen, die overwogen dienen te worden bij de keuze van de dosis en bij het vaststellen van het benodigde tijdschema voor monster afname. Daarnaast kan het gebruik van historische data (die de distributie van parameters kan bescrijven) van belang zijn om de onzekerheid omtrent de relatie tussen blootstelling en bijwerkingen of toxiciteit te reduceren. Vanuit een statistisch oogpunt, kan het gebrek aan model en parameter identificeerbaarheid een risico zijn voor de interpretatie van resultatent uit een PKPD analyse. Daarom zijn techknieken zoals simulaties, bootstrap en gevoeligheidsanalyse essentieel om zowel de betrouwbaarheid als de nauwkeurigheid van de voorspelingen te kunnen waarborgen. Tenslotte, benadrukken wij de relevantie van de zogenoemde selectie criteria op de identificeerbaarheid van een model. Onze bevindingen wijzen er op dat traditionele criteria, die gebruikt worden om de complexiteit van een model beoordelen zoals BIC en AIC, ongeschikt zijn voor het modelleren van effecten die met lage of zeer lage frequentie voorkomen. Daarom stellen wij voor om farmacologische plausibiliteit naast minder conservatieve statistische criteria toe te passen bij de selectie van een PKPD model.

Het hoofdstuk wordt afgerond met een beknopte discussie over de ontwikkelingen op het terrein van PKPD modellering en hoe de methodologische problemen die wij geïdentificeerd hebben verholpen kunnen worden. Onze bevindingen onthullen dat gevolgtrekkingen over de zogenoemde “veilige blootstelling” en de daarop geassocieerde risico’s niet geschat of voorspeld kunnen worden op basis van losse empirische experimenten waarin uitsluitend het effect van supratherapeutische concentraties is bestudeerd. Het gebruik van een

computationeel modelmatig raamwerk is onvermijdelijk als men data en kennis over het lotgeval en de farmacologische eigenschappen van een stof tracht te integreren.

Om het lange termijn veiligheids en toxiciteitsprofiel van een geneesmiddel in mensen te kunnen voorspellen zijn methodes nodig die de werkingsmechanismen geassocieerd met zowel primaire als secundaire receptorsystemen verbinden. Verder moet er rekening worden gehouden met de factoren die deel uitmaken van het biologische systeem na de receptor. Daarbij zal kwantitatieve systeem farmacologie een belangrijk rol kunnen spelen. Dit vakgebid vertegenwoordigd de integratie van systeem biologie en farmacologie waar computationele en experimentele methoden gecombineerd kunnen worden om de progressie van ziekten en effecten van geneesmiddelen te bestuderen en/of voorspellen. Deze aanpak is met name relevant, zowel vanuit klinisch als methodologisch perspectief, voor de parameterisatie van lange termijn bijwerkingen, die hun oorsprong vinden in de pertubatie van homeostatische mechanismen, door cellulaire veranderingen of bij weefsel en cel schade.

Naast het ontwikkelen van meer fysiologisch en mechanistisch-gebasseerde modellen dient ook de toekomstige uitbreiding van de risico-baten analyse zoals voorgesteld in dit proefschrift overwogen te worden. Er zijn verschillende voorbeelden gepubliceerd van risico-baten analyses waarin gebruik wordt gemaakt van PKPD modellen. Het is aannemelijk dat de integratie van deze methoden tot een veel betrouwbaarder raamwerk kan leiden, dat vervolgens gebruikt zou kunnen worden ter beoordeling van nieuwe en bestaande geneesmiddelen, alsmede het optimale gebruikt ervan. Bij de evaluatie van het veiligheids en toxiciteitsprofiel van een geneesmiddel dienen de dosis en de tijd als bepalende factoren te worden overwogen. Dit punt werd aangehaald al in 1538 door de Zwits-Duitse arts, Theophrastus of Hohenheim, die stelde dat alles giftig is en dat niets niet giftig is: alleen de dosis maakt iets niet giftig.

Ter conclusie: ondanks het gebrek aan een breed scala van scenario's en beprekingen die intrinsiek verbonden zijn aan het selecteren van experimentele protocollen zoals gepresenteerd in dit proefschrift, werpen onze bevindingen een nieuwe, wellicht

belangrijkere vraag op met betrekking tot de ethische basis voor het gebruik van proefdieren bij empirische experimentele protocollen, met als rechtvaardiging dat daarmee de veiligheid en toxiciteit van een geneesmiddel gekarakteriseerd kunnen worden.

Acknowledgements

First, I would like to thank my supervisor Prof. Oscar Della Pasqua for providing the chance to work on so many challenging projects in Leiden and London. His enthusiasm has led to ongoing projects within GlaxoSmithKline where we have successfully piloted the methodology presented in this thesis and developed it into active projects. Prof. Danhof, I am grateful for your support and pragmatic viewpoint. During my time at the Division of Pharmacology I had the opportunity to meet many people I respect. Chiara and Francesco, my paranimfen, whom I consider to be very good friends, thank you for all the good times inside and outside university. Gijs, I am thankful to have been mentored by someone of your brilliance. I would also like give a special thanks to Vincenzo, for his willingness to share his technical expertise and time with the group. There are many other colleagues whom I am also very glad to have met: Elke, Massimo, Anne, Roosmarijn, Ibrahim, Maurice, Amit and Donato. For the simulation work described in the theoretical portion of this thesis, the computational resource that was required was so large that I needed to borrow the "idle time" of most of the computers in the department. I would like to thank all those that have allowed me to do this. Chapters 7 and 8 required me to conduct a very large animal experiment. I'm grateful to the *in-vivo experts* for their willingness to put up with a mathematician in their labs (Maartin, Dirk-Jan, Robin, Liesbeth, Joost, Jasper). A special thanks to Ian, Núria and Chantal for staying late into the evenings on many days to collect samples and helping to make sure the study went according to schedule.

During my time at GlaxoSmithKline in London, in the fourth year of my research programme, I had the opportunity to meet great people: Alex, Alienor, Chao, Chris, Claire, Daren, Georgios, Hugo, Isabelle, Lorenzo, Misba, Monica, Sophie and Stefano. I am grateful for the opportunity to work with you all.

Frank Klopprogge has recently spent his holiday working on the translation of my Dutch summary. I hope it wasn't too difficult. I am also grateful to Rowena Asgarali for designing my cover and being a good friend over the years. She's a very talented artist. Finally, I would like to thank my parents, Pardeep and Núria for their continued support**.**

Curriculum Vitae

Tarjinder Sahota was born on the 16th of January 1982 in Hitchin, UK. He attended Fearnhill School where he obtained his A levels in August 2000. Subsequently he started his training in Mathematics and Physics at King's College London, where he obtained his MSc degree in July 2004 where he was awarded prizes for "best final year project" and "best results". Following his graduation, he spent two years teaching privately and in secondary schools for his PGCE qualification. In July 2006 he joined GSK as a contract analyst where he worked on optimising study design. He started his PhD research programme at the division of Pharmacology in Leiden in October 2006, under direct supervision of Dr Oscar Della Pasqua, who together with Prof Meindert Danhof acted as co-promotor and promoter of the work presented in this PhD thesis.

In November 2010, he was employed at GlaxoSmithKline R&D in the United Kingdom to conduct a research project on the optimisation of fixed dosing regimens for the treatment of tuberculosis. In November 2011 he transitioned to his current role as Pharmacometrician, supporting early drug development projects in the "Biopharm and Immunoinflammatory" therapeutic area and also leading the NONMEM estimation methods sub-group within GSK's internal Pharmacometrics group.

List of publications

Sahota, T., & Della Pasqua, O. (2012). Feasibility of a fixed-dose regimen of pyrazinamide and its impact on systemic drug exposure and liver safety in patients with tuberculosis. *Antimicrobial Agents and Chemotherapy* **56**: 5442-5449.

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. *Toxicology and Applied Pharmacology* **278** (3):209-219