BJP British Journal of Pharmacology

DOI:10.1111/bph.13167
www.brjpharmacol.org

RESEARCH PAPER

Model-based prediction of the acute and long-term safety profile of naproxen in rats

Correspondence
Oscar Della Pasqua, Clinical Pharmacology & Therapeutics, University College London, London WC1H 9JP, UK. E-mail: o.dellapasqua@ucl.ac.uk

Received
24 August 2014
Revised
8 February 2015
Accepted
15 April 2015

Tarjinder Sahota[1], Ian Sanderson[1], Meindert Danhof[1] and Oscar Della Pasqua[1,2,3]

[1]*Division of Pharmacology, Leiden Academic Centre for Drug Research, Leiden, The Netherlands,*
[2]*Clinical Pharmacology, Modelling and Simulation, GlaxoSmithKline, Uxbridge, UK, and*
[3]*Clinical Pharmacology & Therapeutics, University College London, London, UK*

BACKGROUND AND PURPOSE

Despite the increasing importance of biomarkers as predictors of drug effects, toxicology protocols continue to rely on the experimental evidence of adverse events (AEs) as a basis for establishing the link between indicators of safety and drug exposure. Furthermore, biomarkers may facilitate the translation of findings from animals to humans. Combined with a model-based approach, biomarker data have the potential to predict long-term effects arising from prolonged drug exposure. Here, we used naproxen as a paradigm to explore the feasibility of a biomarker-guided approach for the prediction of long-term AEs in humans.

EXPERIMENTAL APPROACH

An experimental toxicology protocol was set up for evaluating the effects of naproxen in rats, in which four active doses were tested (7.5, 15, 40 and 80 mg·kg^{-1}). In addition to AE monitoring and histology, a few blood samples were also collected for the assessment of drug exposure, TXB_2 and PGE_2 levels. Non-linear mixed effects modelling was used to analyse the data and identify covariate factors on the incidence and severity of AEs.

KEY RESULTS

Modelling results showed that besides drug exposure, maximum PGE_2 inhibition and treatment duration were also predictors of gastrointestinal ulceration. Although PGE_2 levels were clearly linked to the incidence rates, it appeared that ulceration severity is better predicted by measures of drug exposure.

CONCLUSIONS AND IMPLICATIONS

These results show that the use of a model-based approach provides the opportunity to integrate pharmacokinetics, pharmacodynamics and toxicity data, enabling optimization of the design, analysis and interpretation of toxicology experiments.

Abbreviations

AUC, area under the concentration versus time curve; CAOC, cumulative area over biomarker concentration versus time profile; C_{MAX}, maximum drug concentration over the period of 24 h; C_{MIN}, maximum biomarker inhibition over the period of 24 h; CWRES, conditional weighted residuals; I_{MAX}, maximum biomarker inhibition; IPRED, individual prediction; NOAEL, no-observed adverse-effect level; NSAID, nonsteroidal anti-inflammatory drug; OBS, observed concentration or levels; PD, pharmacodynamics; PK, pharmacokinetics; PKPD, pharmacokinetic-pharmacodynamic; PRED, population prediction; SCM, stepwise covariate method

Tables of Links

TARGETS
COX-1
COX-2

LIGANDS	
Arachidonic acid	PGE2
Naproxen	TXB2

These Tables list key protein targets and ligands in this article which are hyperlinked to corresponding entries in http://www.guidetopharmacology.org, the common portal for data from the IUPHAR/BPS Guide to PHARMACOLOGY (Pawson *et al.*, 2014) and are permanently archived in the Concise Guide to PHARMACOLOGY 2013/14 (Alexander *et al.*, 2013).

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. However, despite the availability of a mixture of acute, midterm and chronic toxicity data prior to drug approval, the identification of long-term risks often happens in phase IV post marketing surveillance.

Ideally, earlier identification of potential risks would enable the use of evidence-based risk mitigation strategies. In order to achieve this goal, well-designed, integrated experimental protocols are needed in which time-dependent physiological changes arising from repeated exposure to a drug are characterized. Such an objective may be hampered by the use of traditional, empirical toxicology 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 and Bridges, 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.

In this context, biomarkers can be of great relevance, as they offer the possibility to discriminate between acute and chronic treatment effects. 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 and Ruggeri, 2014). In the current investigation, we show therefore how 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 PGE_2 and TXB_2 and gastric ulceration in rats.

Non-selective nonsteroidal anti-inflammatory drugs (NSAIDs), such as naproxen, act by blocking COX, which catalyses the rate-limiting step in the formation of prostanoids from arachidonic acid (Chakraborti *et al.*, 2010). From a pharmacological perspective, various investigations have shown that both COX-1 and COX-2 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 activities of either isoform may be required for normal physiological function. Continuous COX-1 inhibition following prolonged administration of non-selective COX inhibitors is known to induce gastrointestinal (GI) adverse effects, especially ulceration and haemorrhagic bleeding (Loftin *et al.*, 2002; Whittle, 2004; Wallace, 2008; Takeuchi, 2012). 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 (Solomon *et al.*, 2004; Schneeweiss *et al.*, 2006).

In spite of known interspecies differences that exist in GI-related morbidity (Kargman *et al.*, 1996), we hypothesize that the characterization 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 (Jackson *et al.*, 2000; Halter *et al.*, 2001), which is then converted to macroscopic damage by the concomitant inhibition of COX, with decreased mucosal PGs, presumably because of their effect on the microvasculature (Fornai *et al.*, 2014).

In a previous investigation, we have shown how safety biomarkers can be used in conjunction with general toxicity protocols to predict the safety window in humans using an empirically derived safety threshold (Sahota *et al.*, 2014a). However, safety thresholds such as the no-observed adverse-effect level (NOAEL) have many statistical and experimental limitations (Dorato and Engelhardt, 2005; Sahota *et al.*, 2014b). Here we demonstrate that the use of a mechanism-based approach enables one to explore different parameterizations of biomarker response and drug exposure, facilitating the identification of potential (causal) factors associated with the overall treatment risk. In contrast to traditional

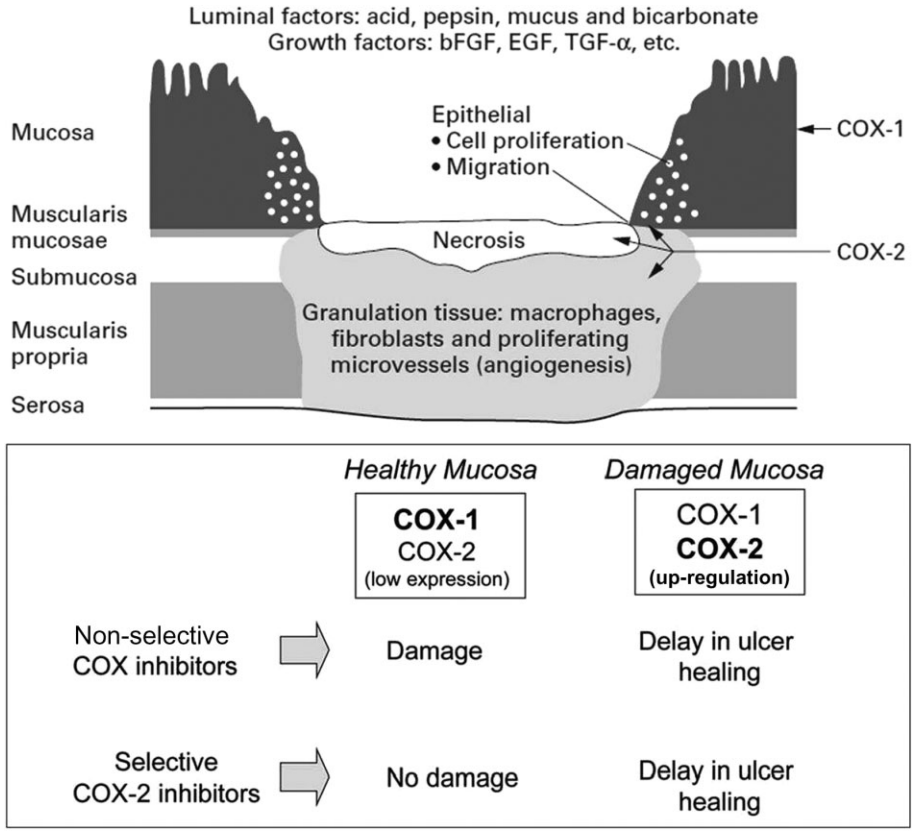

Figure 1

(Upper panel) Diagrammatic presentation of the processes underlying ulcer healing and factors affecting it. In the intact mucosa, COX-1 is the predominant COX isoform in the GI tract. In contrast, during wound healing, the expression of 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).

extrapolation methods such as allometric scaling, which relies primarily on the point estimates of safe exposures, the use of hierarchical models 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 are increasingly more important as chronic interventions are common across most therapeutic areas.

Methods

The present investigation is based on a previously published general toxicity study in rats by Sahota *et al.* (2014a), with the non-selective COX inhibitor naproxen. A detailed description of the study design, strain of rats, sample collection, and analysis and PKPD modelling can be found in Sahota *et al.* (2014a).

Summary of study design

Three different treatment durations were investigated (1, 2 and 4 weeks). Rats were given daily doses of naproxen by p.o. gavage. The initial protocol was aimed at evaluating four cohorts per treatment duration at dose levels of 0, 15, 40 and 80 $mg \cdot kg^{-1} \cdot day^{-1}$ naproxen. However, the proposed dose levels had to be modified during the experimental phase as all animals receiving 80 $mg \cdot kg^{-1}$ in the 1 week cohort suffered from moderate weight loss. Subsequent experiments for the 2 week and 1 month cohorts were therefore performed with 7.5, 15 and 40 $mg \cdot kg^{-1}$ doses.

Satellite animals received identical doses to the toxicology groups and were used to measure plasma drug concentration (PK) and biomarker (PD) data, namely TXB_2 and PGE_2. Briefly, the biomarker data were obtained from whole blood *ex vivo* assays using LPS and aspirin for the analysis of PGE_2, whereas as a stimulus clotted blood was used for the assessment of TXB_2. These protocols provide evidence of the pharmacological effects of COX inhibitors on COX-1 and COX-2 activity respectively. An optimized composite sampling scheme was used and sampling took place on days 1, 7, 14 and 28. Details regarding sample analysis can be found in Sahota *et al.* (2014a). Endpoints in toxicology groups included adverse events, GI histology and terminal PK and PD measurements. Animals in the toxicology groups were killed at the end of

their treatment period (1, 2 and 4 weeks). An overview of the study protocol is depicted in Figure 2. For each treatment duration, four groups were tested each of which received four different dose levels (n = 8 animals per dose level in the toxicology group and n = 24 animals per dose level in the satellite arm). This resulted in an experimental protocol with 96 animals in the toxicology group and 96 animals in the satellite group.

The experimental procotol was approved by the DEC (Dierexperimentencommissie), i.e., the Animal Experimentation Committee of the University of Leiden, The Netherlands. The protocol procedures were in agreement with the ARRIVE guidelines for reporting experiments involving animals (Kilkenny *et al.*, 2010; McGrath *et al.*, 2010) as well as with the Dutch regulations regarding the use of animals for scientific research, as defined by the Legislation on Animal Experiments (Wet op Dierproeven), currently in force in the Netherlands.

Histology

Histological evaluation of the stomach was performed to establish a correlation between acute and long-term adverse events. Immediately after the rats had been killed, the stomachs were removed and cut open along the greater curvature and then washed with warm saline. The inner surface was photographed to allow the quantification of the area covered by haemorrhagic ulceration, which was determined under a dissecting microscope. Gastric ulceration was measured in terms of incidence and severity (percentage stomach surface area affected by ulceration). The software Image J version 1.43 (Abramoff *et al.*, 2004) was used to calculate ulcer area and total stomach surface area. The person who performed the ulceration measurement was blinded as to animal identification and treatment group.

PKPD model

Naproxen concentrations and biomarker data were analysed by non-linear mixed effects modelling, as implemented in NONMEM version 7.2.0. Further details of the model building, diagnostics and results of the PK and PKPD analysis are reported elsewhere (Sahota *et al.*, 2014a). Briefly, the PK of naproxen in plasma was best described by a one-compartment model with first-order absorption, first-order elimination and non-linear dose-dependent bioavailability. Weight was included as a significant covariate on clearance and volume of distribution. The PKPD models for both biomarkers, PGE_2 and TXB_2 were characterized by direct sigmoid I_{MAX} models. Parameter values, precision estimates and goodness-of-fit diagnostics are shown in Supporting Information Table S1. Model-predicted biomarker levels were used as input (independent variables) in the analysis of adverse events.

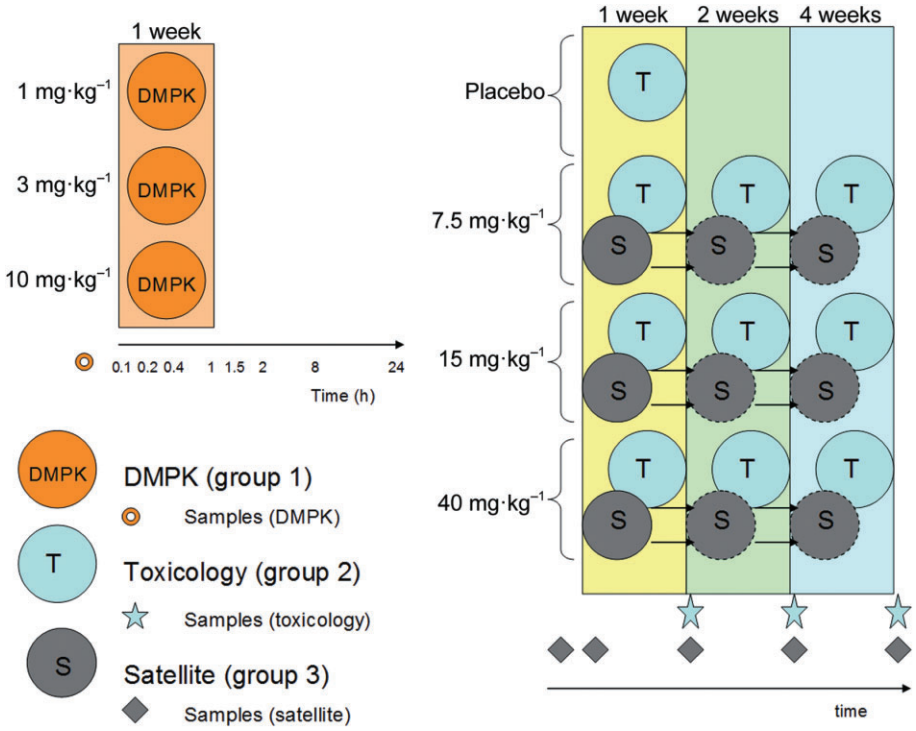

Figure 2

Schematic representation of the design of the experimental protocol. The use of an integrated approach allowed the combination of additional data from (standard) pharmacokinetic experiments (DMPK experiments). Animals were assigned to different groups according to treatment duration and dose level. Treatment duration varied between 1, 2 and 4 weeks. For each treatment duration, four groups were tested each of which received four different dose levels (n = 8 animals per dose level in the toxicology group and n = 24 animals per dose level in the satellite arm, i.e. 3 animals per sampling time point). DMPK = drug metabolism and pharmacokinetic experiments including serial blood sampling.

Data analysis

Final model parameters describing the incidence of gastric ulceration 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 minimization and completion of the 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, 2014). 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 (Hooker *et al.*, 2007).

Given the purpose of the study was to discriminate between acute and long-term effects of naproxen, different parameterizations were considered to describe the observed drug effects during the course of treatment. Predicted exposure and biomarker levels for each individual animal were calculated from the final PKPD model using *post hoc* empirical Bayes estimates (using MAXEVAL = 0). Details of the calculation methods are described in Table 1.

Ulceration model. Because 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 the severity of ulceration were considered during modelling. Incidence was modelled as the probability of occurrence of ulceration, $U_{i,j}$ at the time of killing, 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 equation 1:

$$P(U_{i,j}=1)=\frac{EXP\left(\theta_1+\sum_k \theta_k * COV_{i,j,k}\right)}{1+EXP\left(\theta_1+\sum_k \theta_k * COV_{i,j,k}\right)} \quad (1)$$

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 base model did not include any covariates on response. As shown in equations 2 and 3, two fixed effect model parameters were used, θ_1 and θ_2, which represent the typical values of the logit of the probability of ulceration and the severity of ulceration respectively.

$$P(U_{i,j}=1)=\frac{EXP(\theta_1)}{1+EXP(\theta_1)} \quad (2)$$

$$PER_{pred,i,j}=\begin{cases}0, \text{ if } U_{i,j}=0\\ \theta_2, \text{ if } U_{i,j}=1\end{cases} \quad (3)$$

$$\log(PER_{obs,i,j})=\log(PER_{pred,i,j})+\varepsilon_{i,j} \quad (4)$$

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

Table 1

Calculation of biomarker response and exposure variables

Parameter name	Symbol	Calculation
Area under drug concentration versus time profile	AUC	$\int_{t-24}^{t} C_p dt$
Area above biomarker concentration versus 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 versus time profile	CAUC	$\int_0^t C_p dt$
Cumulative area over biomarker concentration versus time profile	CAOC	$BC_p(0)-\int_0^t BC_p dt$
Maximum drug concentration over 24 h period	C_{MAX}	$\max(\{C_p(s) : t-24<s<t\})$
Maximum biomarker inhibition over 24 h period	C_{MIN}	$\min(\{BC_p(s) : t-24<s<t\})$

Individual-predicted naproxen concentrations and biomarker levels are denoted by $C_p(t)$ and $BC_p(t)$ respectively.

method (SCM) in PsN (Lindbom *et al.*, 2005). The variables defined in Table 1, in addition to body weight and age, were tested as potential influential factors on the incidence and severity of ulceration. Time measured in days (DAY) was also tested as a covariate and as a 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 base model. Linear, exponential and hyperbolic (sigmoid E_{max}) 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 is described by equation 5:

$$PER_{pred,i,j} = \begin{cases} 0, & \text{if } U_{i,j} = 0 \\ \theta_1 + (\theta_{SLOPE} * (COV - median(COV))), & \text{if } U_{i,j} = 1 \end{cases} \quad (5)$$

where θ_1 is population prediction and θ_{slpoe} is the slope of relationship between parameter and (centred) covariate.

Likewise, an exponential relationship was also evaluated as described by equation 6:

$$PER_{pred,i,j} = \begin{cases} 0, & \text{if } U_{i,j} = 0 \\ \theta_1 * \exp(\theta_{SLOPE} * (COV - median(COV))), & \text{if } U_{i,j} = 1 \end{cases} \quad (6)$$

Because data were sparse and maximum effect may not have been reached, the maximum effect was fixed to 100% during the evaluation of the sigmoid E_{max} function, as shown in equation 7:

$$PER_{pred,i,j} = \begin{cases} 0, & \text{if } U_{i,j} = 0 \\ \theta_2 * \dfrac{100 * COV^{\gamma}}{COV_{50}{}^{\gamma} + COV^{\gamma}}, & \text{if } U_{i,j} = 1 \end{cases} \quad (7)$$

where γ is an estimated Hill coefficient and COV_{50} is the covariate value required to achieve 50% of maximum effect.

Covariate relationships were centred by the median value of the covariate so that in this case $PER_{pred,i,j} = \theta_2$, as in the base model. To facilitate covariate testing, COV_{50} was reparameterized to be a factor θ of the median covariate value (equation 8).

$$COV_{50}{}^{\gamma} = \theta median * (COV) \quad (8)$$

As shown in equation 9, the hockey stick function was implemented according to the following function:

$$PER_{pred,i,j} = \begin{cases} \theta_1, \text{ if } U_{i,j} = 0 \text{ and } COV < \theta_{THRESH} \\ \theta_1 + (\theta_{SLOPE} * (COV - \theta_{THRESH})), \\ \quad \text{if } U_{i,j} = 1 \text{ and } COV \geq \theta_{THRESH} \end{cases} \quad (9)$$

where θ_1 is population prediction and θ_{slope} is the slope of relationship between parameter and (centred) covariate, and 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 at 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 using the bootscm option in PsN to estimate covariate inclusion probabilities (i.e. to estimate model specification uncertainty).

Model validation. The performance of the ulceration models was 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 and displayed for the incidence and percentage gastric area affected. Validation procedures also included normalized prediction distribution errors, which are based on the assumption that the normalized (de-correlated) prediction distribution errors (discrepancies) are normally distributed (Comets *et al.*, 2008). One hundred data sets 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 toxicology group animals. These revealed gastric ulceration at all dose levels except for placebo, where no adverse events were observed. As a consequence, no NOAEL could be obtained for any of the treatment durations (Figure 3).

It should be noted that despite an initial protocol design with dose levels up to 80 mg·kg^{-1}, dose levels of naproxen had to be reduced for the 2 week and 1 month cohorts for ethical reasons. This was due to moderate weight loss observed in all animals receiving 80 mg·kg^{-1} in the 1 week cohort. The animals in this cohort were immediately killed. The final protocol was therefore modified for the 2 week and 1 month cohorts, which received lower doses (7.5, 15 and 40 mg·kg^{-1}). Histological examinations were performed on these animals and terminal blood samples taken. No other adverse events were observed.

Logistic models for gastric ulcerations

Empirical analysis of these data revealed some peculiarities in the data, in that there was no significant dose-response until week 4. In fact, due to the significantly negative parameter Θ_2 (in Table 2), 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. Exploratory evaluation of the relationship between naproxen exposure and biomarker levels instead of dose did not provide further evidence of an apparent relationship.

After an initial attempt to describe the data without the use of covariates, a clear model misspecification was observed. As shown in Figure 3, the apparent negative dose–toxicity relationship for week 1 and week 2 was not replicated by the final model predictions. An overview of the parameter estimates for the final model is summarized in Table 2, whereas Figure 4 shows the model specification uncertainty, as determined by the results of the bootstrap procedure.

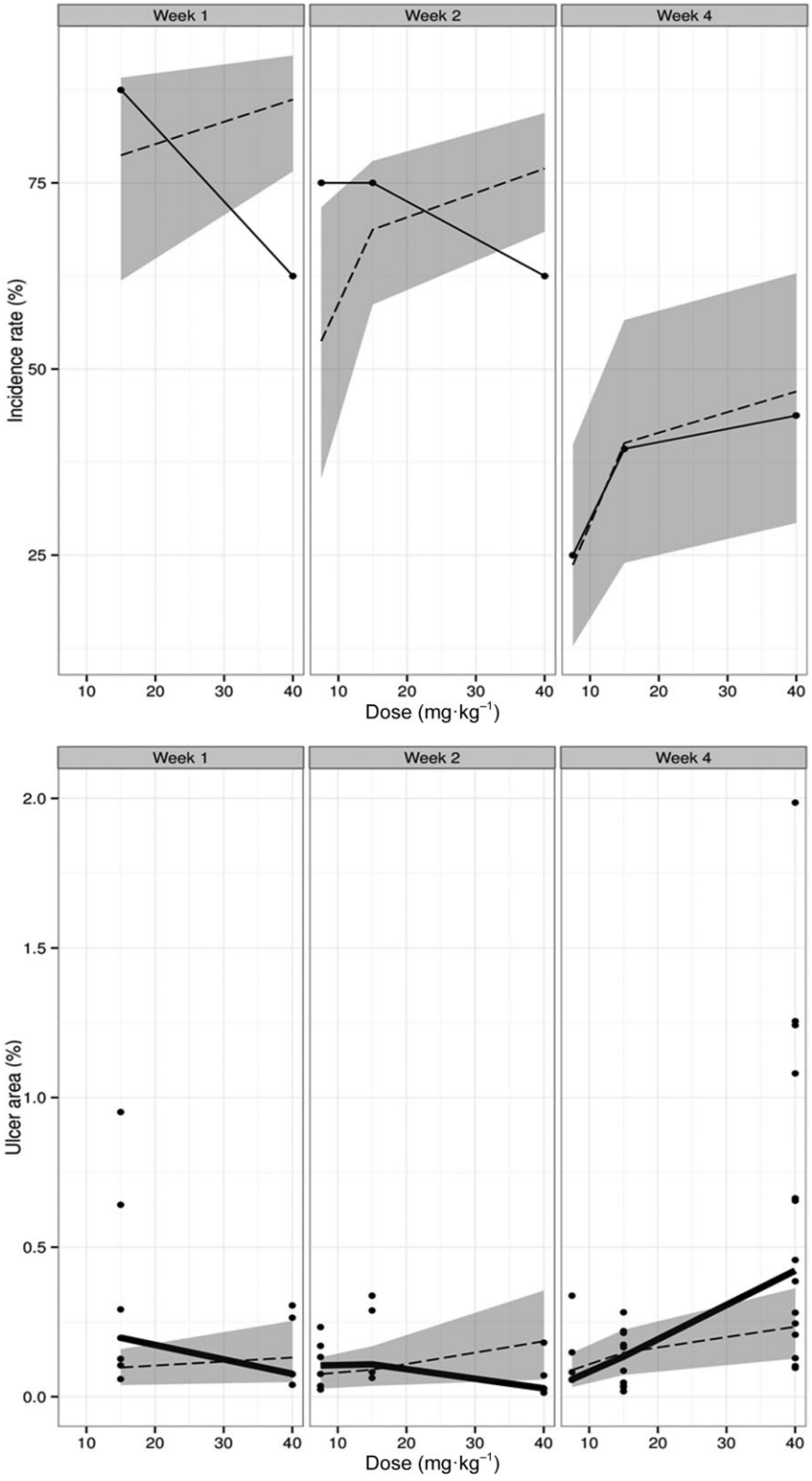

Figure 3

Plots of the observed and predicted ulcer incidence (top) and severity (bottom). Dots in the upper panel show observed percentage of total animals in each cohort manifesting GI toxicity. Ulcer severity was measured as % of stomach area affected by ulceration. The shaded area depicts the 95% uncertainty in population prediction of the model (dashed lines depict the 50th percentile). The model was 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.

Table 2

Logistic model estimates describing ulcer incidence and severity

Parameter	Value	SEM
$\text{logit}(P(U_{i,j}=1)) = \theta_1 + \theta_2 * \text{IMAX}_{PGE,i,j} + \theta_3 * \text{DAY}_{i,j}$		
θ_1	−0.226	0.305
θ_2	0.042	64.8%
θ_3	−0.066	0.022
$PER_{pred,i,j} = \begin{cases} \theta_1, & \text{if } U_{i,j}=0 \text{ and } COV < \theta_{THRESH} \\ \theta_1 + (\theta_{SLOPE} * (CAUC_{TXB} - \theta_{THRESH})), & \text{if } U_{i,j}=1 \text{ and } COV \geq \theta_{THRESH} \end{cases}$		
θ_4	0.21%	17.2%
θ_5	149.9	74%
θ_6	94.3%	13%
Residual variability (proportional error)	78%	16%

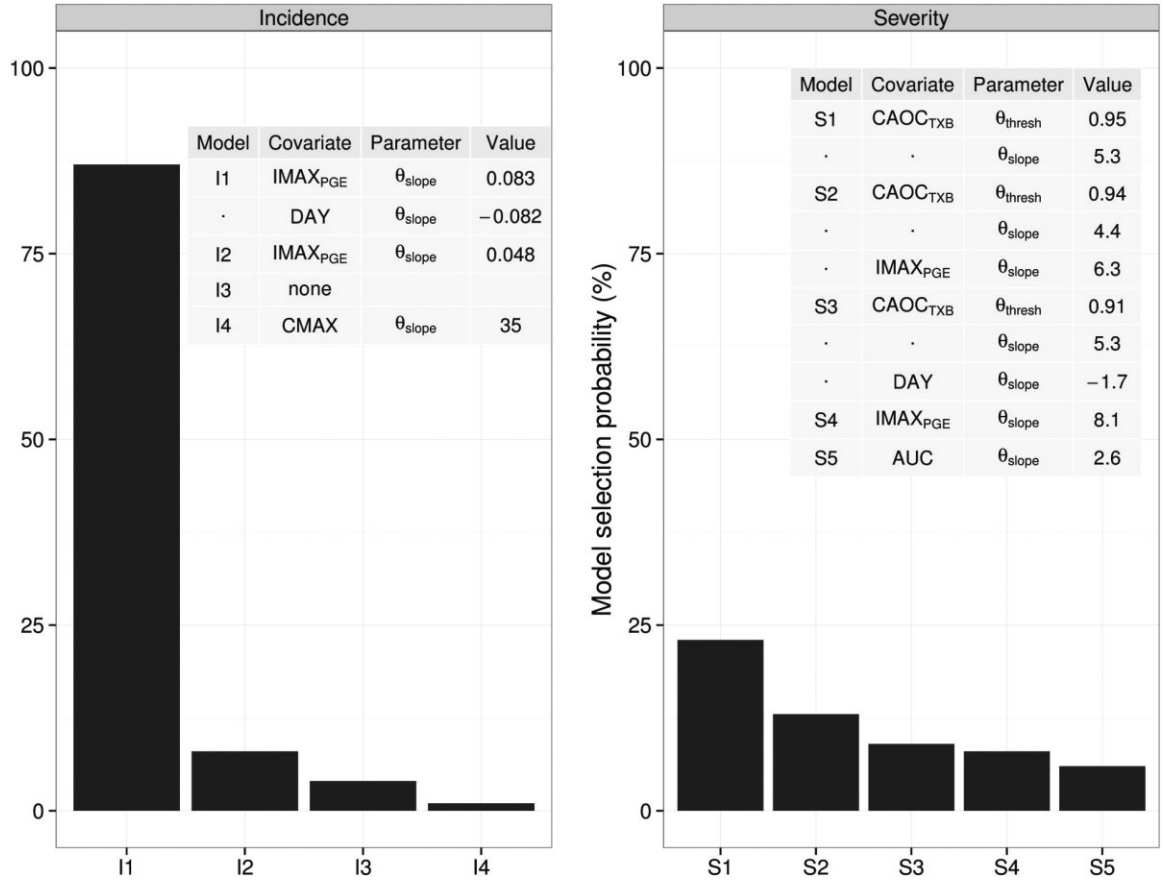

Figure 4

Model specification uncertainty. Bars indicate model selection probability as determined by the bootstrap SCM ordered from most to least probable. Only the top five most probable models are 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 inset shows numerical details of model selection probability.

The wide, flat distribution reflects high model specification uncertainty (i.e. two or more different models may be indistinguishable). In addition, 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 (Figure 5). By contrast, upon incorporation of this time-dependent effect into the final model, the overall fit

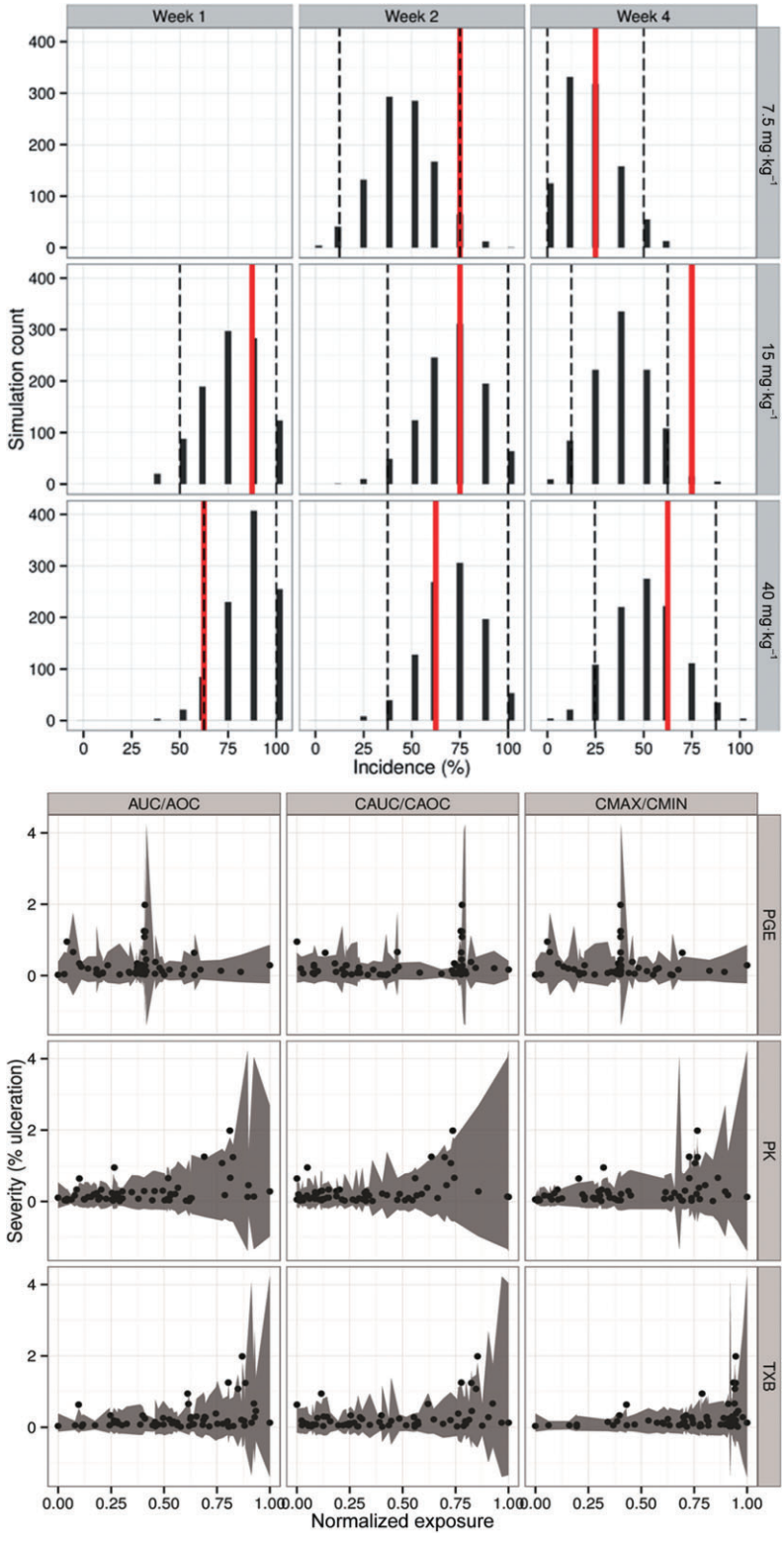

Figure 5

Visual predictive checks. The upper panel depicts the prediction distribution for the incidence of ulcers by dose level. Observed incidence is highlighted by the red lines whereas dashed lines represent the 90% prediction intervals (PIs). Of the eight groups, one would expect on average one group to fall beyond the 90% PIs, which is in line with the observed results. Ulcer incidence is determined as the observed percentage of total animals manifesting GI toxicity. The lower panel shows ulceration severity against predicted normalized exposure for different variables of interest (AOC = area above biomarker concentration vs. time profile; CAUC = cumulative area under the concentration vs. time profile). The shaded area represents the 95% PI; dots depict the actual data. PK = naproxen concentrations.

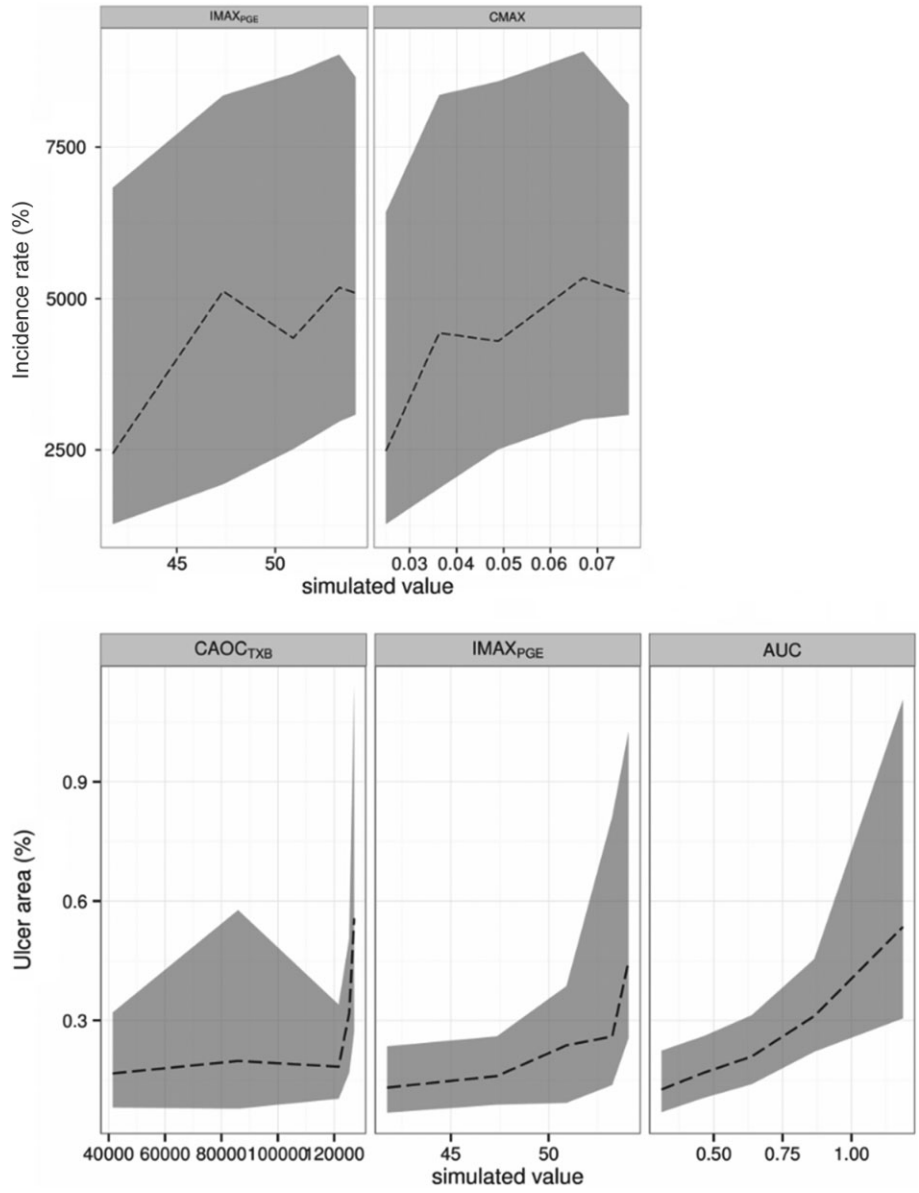

Figure 6

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 represent the median profile of simulated values using the final model, whereas shaded area represent the 95% prediction intervals. 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. However, cumulative TXB_2 inhibition was found to be the best explanatory variable for the severity of ulceration. Given the degree of model uncertainty, a note of caution is required before assigning clinical meaning to these covariates.

improved (Figure 6). C_{MAX} and AUC were shown not be the primary drivers of toxicity, although these parameters may be considered indirectly correlated with risk.

Discussion and conclusions

Model-based evaluation of safety pharmacology and toxicology data

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 characterize the relationship between drug exposure and toxicity, but rather to explore the boundary between therapeutic and toxic exposures (Josa *et al.*, 2001). 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; 2008; Bai *et al.*, 2013).

From a methodological perspective, general toxicology studies represent a challenge for model-based analysis because sparse pharmacokinetic data 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 data sets, as compared with studies aimed at the characterization of PKPD relationships, which are now commonly used in early drug development (Knight, 2007).

An important aspect of our analysis was therefore 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 highlighted that the lack of NOAEL (due to the presence of adverse events at all tested dose levels) has not prevented us from further characterizing the exposure–adverse event relationships. The main modifications consisted of the additional collection of biomarker data from animals and the choice to treat histological observations as a continuous data type. The incorporation of biomarkers into the assessment of long-term toxicity enabled us to further understand time dependencies and nonlinearities in downstream effects related to the primary pharmacological target (Huntjens *et al.*, 2010).

Mechanism-based analysis of long-term safety and toxicology data

The present analysis suggested that the ulcerative effect is acute and diminishes with sustained long-term exposure. In addition, C_{MAX} and AUC were not selected features of the final model, 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 ($IMAX_{PGE}$) was the best predictor of adverse event incidence, with the bootstrap SCM showing low model uncertainty. On the other hand, cumulative TXB_2 inhibition ($CAOC_{TXB}$) was found to be the best explanatory variable for the severity of ulceration (Figure 6). Other physiologically plausible explanatory factors such as maximum PGE_2 inhibition or DAY (treatment duration) were found to be fraught with significant model uncertainty. Physiologically, it is plausible that the ulcerative effect of naproxen primarily reflects the blockade of tissue damage repair by COX-2 (Brzozowski *et al.*, 2001), whereas the severity of ulcerations may potentially be linked to the antiplatelet (i.e. COX-1 dependent) prostanoid inhibition in the GI tract (Capone *et al.*, 2007).

PKPD relationships as a translational factor for the evaluation of risk in humans

It is important to emphasize that for the GI toxicity, animal data are not easily translated to humans. In humans, it is still unclear what clinical end point is best related to the inhibition of prostanoid biosynthesis in the GI tract and which isozyme is involved. The results of randomized clinical trials together with the knowledge of the impact of the different drugs on COX-1 and COX-2 activity at therapeutic doses (using human whole blood assays *ex vivo*) suggest therefore that gastroduodenal lesions develop as a consequence of persistent, moderate inhibition of mucosal COX-1 activity in a large proportion of exposed patients, while bleeding complications occur as a result of transient, high-grade inhibition of platelet COX-1 in a very small percentage of exposed patients.

Randomized clinical trials with coxibs versus traditional NSAIDs have focused mainly on three classes of clinical end points (Rostom *et al.*, 2007; Higuchi *et al.*, 2009). First, GI symptoms such as dyspepsia, which have a poor signal-to-noise ratio and for which the COX-1 dependence of the signal is uncertain. In addition, these symptoms also fail to correlate with the presence of detectable GI lesions. Second, endoscopically detectable lesions, which are largely COX-1 dependent and have a favourable signal-to-noise ratio. However, it remains uncertain whether they are reliable predictors of serious GI complications. Third, serious GI complications (i.e. perforation, ulcer and bleeding; perforation, obstruction and bleeding). These measures have an uncertain signal-to-noise ratio because of lack of adequately sized placebo-controlled studies. Moreover, the haemorrhagic nature of the most prevalent component of these combined endpoints makes it likely that they reflect primarily GI complications related to inhibition of platelet by COX-1.

Taking into account the points mentioned above, it becomes clear that the relationship between chronic exposure and the incidence and severity of adverse events should be considered a critical but not sufficient requirement to predict safety and toxicity in humans. Therefore, PKPD models need to be parameterized 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, is sine qua non conditions to translate and accurately interpret safety findings (Zuideveld *et al.*, 2007; Chain *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. In fact, the role of COX inhibition in

the characterization of NSAID-induced adverse events was recently investigated by Massó González *et al.* (2010), who describe the relationship between the degree of inhibition of whole blood COX-1 and COX-2, the maximum plasma concentration (Cmax) of individual NSAIDs, and relative risk of upper GI bleeding/perforation in humans. Drugs that have a long half-life or slow-release formulation and/or correlate with profound and coincident inhibition of both COX isozymes were found to be associated with a greater risk of upper GI bleeding/perforation.

Whereas the experimental results for naproxen seem to concur with the conclusions drawn by Massó González *et al.* (2010), 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, 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, in a study performed by 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 of PGE_2 and TXB_2.

Potential limitations

Unfortunately, our investigation is just a first step into the use of a pharmacologically based approach to the evaluation of adverse events. This work does not address differences between the occurrence of adverse events in animals versus humans. This is a weakness that applies to any analysis of these data regardless of whether it is based on PKPD modelling or traditional safety thresholds (e.g. NOAEL). We anticipate that to address this limitation, systematic, integrated analysis of drug effects *in vitro*, in animals and humans is required. Such an analysis implies access to data from different doses and compounds, so that one can discriminate between drug- and system-specific properties.

We also understand that establishing correlations between clinical and preclinical data would be useful to support claims about the predictive value of the approach. However, this would also imply establishing the sensitivity and specificity of the endpoints, including a range of compounds, rather than a single drug. Our initial intent was therefore to show that any attempt to translate or predict drug effects in humans requires careful consideration of the PKPD relationships of the adverse events of interest. In addition, we realize that the logistic models proposed here are empirical in nature, but we believe that such a parameterization may not necessarily undermine its use prospectively. Fully mechanistic models (i.e. systems biology type of models) will be possible only for known targets and molecules which have already been widely tested. For novel chemical and biological entities, understanding of the mechanisms underlying adverse events is very limited, and in most cases unknown at the time safety pharmacology and toxicology protocols are implemented. This issue will remain even when modelling techniques are applied (Nyman *et al.*, 2012). The difference, however, is the possibility to identify drug- (CL, Vd) or system-specific properties (e.g. IC_{50}, I_{MAX}) that can be generalized and/or scaled across species. The merit of this work is therefore to show that markers of pharmacological activity can be treated as explanatory variables for adverse events. These findings can be used in conjunction with allometrically scaled pharmacokinetic data to explore the implications of such effects in humans.

The current manuscript is a first step towards a more pharmacologically based approach for the evaluation of toxicity. The limitations of the modelling exercise presented here are mostly due to the nature of the data package available during (early) preclinical evaluation (i.e. the types of experimental protocols currently used in safety pharmacology and general toxicology). We acknowledge that the absence of ulcerations in vehicle-treated animals and the lack of additional cohorts 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. We also recognize that adopting a different modelling approach to link the probability of adverse effect with the time course of the selected biomarkers and unobserved tissue repair mechanisms may provide further insight into the drivers of toxicity, more specifically ulceration and bleeding (Motsko *et al.*, 2006). However, this approach would require a much more labour-intensive protocol design, including destructive sampling at different intervals to ensure sufficient data are available on the time course of biomarkers and tissue repair status.

In summary, the findings of this study demonstrate the feasibility and potential benefits of a model-based approach for the evaluation of chronic safety pharmacology and toxicity. The present work has shown that even for relatively frequent adverse events, model uncertainty can be significant and therefore one should quantify it. This probably arises from the fact that toxicity studies are generally designed to find safety windows and not to explore the entire exposure-adverse event profile. Moreover, the availability of PKPD relationships may allow us to make inferences about untested doses and dosing regimens, providing an opportunity for risk mitigation, even before obtaining experimental clinical data.

Author contributions

O. D. P. and T. S. defined the research questions, planned the experimental protocol and selected the analysis methods. T. S. and I. S. performed the *in vivo* experiments and data analysis. T. S. wrote the first draft of the paper that has been finalized with the contributions of M. D. and O. D. P.

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 Int 11: 36–42.

Alexander SPH, Benson HE, Faccenda E, Pawson AJ, Sharman JL, Spedding M *et al.* (2013). The Concise Guide to PHARMACOLOGY 2013/14: enzymes. Br J Pharmacol. 170: 1797–1867.

Bai JP, Fontana RJ, Price ND, Sangar V (2013). Systems pharmacology modelling: 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–S6.

Brzozowski T, Konturek PC, Konturek SJ, Sliwowski Z, Pajdo R, Drozdowicz D *et al.* (2001). Classic NSAID and selective cyclooxygenase (COX)-1 and COX-2 inhibitors in healing of chronic gastric ulcers. Microsc Res Tech 53: 343–353.

Capone ML, Tacconelli S, Sciulli MG, Anzellotti P, Di Francesco L, Merciaro G *et al.* (2007). Human pharmacology of naproxen sodium. J Pharmacol Exp Ther 322: 453–460.

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: 708–724.

Chakraborti AK, Garg SK, Kumar R, Motiwala HF, Jadhavar PS (2010). Progress in COX-2 inhibitors: a journey so far. Curr Med Chem 17: 1563–1593.

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: Cohen GM (ed.). Target Organ Toxicity. CRC Press: Boca Raton, FL, pp. 89–115.

Coruzzi G, Venturi N, Spaggiari S (2007). Gastrointestinal safety of novel nonsteroidal anti-inflammatory 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: 1432–1437.

Danhof M, de Lange EC, Della Pasqua OE, Ploeger BA, Voskuyl RA (2008). Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. Trends Pharmacol Sci 29: 186–191.

Della Pasqua O (2013). Translational pharmacology: from animal to man and back. Drug Discov Today Technol 10: e315–e317.

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

Fornai M, Antonioli L, Colucci R, Pellegrini C, Giustarini G, Testai L *et al.* (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: 443–453.

Higuchi K, Umegaki E, Watanabe T, Yoda Y, Morita E, Murano M *et al.* (2009). Present status and strategy of NSAIDs-induced small bowel injury. J Gastroenterol 44: 879–888.

Hooker AC, Staatz CE, Karlsson MO (2007). Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. Pharm Res 24: 2187–2197.

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.

Huntjens DR, Spalding DJ, Danhof M, Della Pasqua OE (2010). Impact of chronic inflammation on the pharmacokinetic-pharmacodynamic relationship of naproxen. Eur J Pain 14: 227, e1–10.

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 *et al.* (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 *et al.* (1996). Characterization of prostaglandin G/H synthase 1 and 2 in rat, dog, monkey, and human gastrointestinal tracts. Gastroenterology 111: 445–454.

Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG (2010). Animal research: Reporting in vivo experiments: the ARRIVE guidelines. Br J Pharmacol 160: 1577–1579.

Knight A (2007). Animal experiments scrutinised: systematic reviews demonstrate poor human clinical and toxicological utility. ALTEX 24: 320–325.

Lanza FL, Royer GL Jr, 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. Dig 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: 241–257.

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: 177–185.

Massó González EL, Patrignani P, Tacconelli S, García Rodríguez LA (2010). Variability among nonsteroidal anti-inflammatory drugs in risk of upper gastrointestinal bleeding. Arthritis Rheum 62: 1592–1601.

McGonigle P, Ruggeri B (2014). Animal models of human disease: challenges in enabling translation. Biochem Pharmacol 87: 162–171.

McGrath J, Drummond G, McLachlan E, Kilkenny C, Wainwright C (2010). Guidelines for reporting experiments involving animals: the ARRIVE guidelines. Br J Pharmacol 160: 1573–1576.

Morita I (2002). Distinct functions of COX-1 and COX-2. Prostaglandins Other Lipid Mediat 68–69: 165–175.

Motsko S, Rascati K, Busti A, Wilson J, Barner J, Lawson K (2006). Temporal relationship between use of NSAIDs, including selective COX-2 inhibitors, and cardiovascular risk. Drug Saf 29: 621–632.

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: 206–218.

Pawson AJ, Sharman JL, Benson HE, Faccenda E, Alexander SP, Buneman OP *et al.*; NC-IUPHAR (2014). The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. Nucl. Acids Res. 42 (Database Issue): D1098–106.

R Development Core Team (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available at http://www.R-project.org/ (accessed 15/10/2014).

Rostom A, Muir K, Dubé C, Jolicoeur E, Boucher M, Joyce J *et al.* (2007). Gastrointestinal safety of cyclooxygenase-2 inhibitors: a Cochrane collaboration systematic review. Clin Gastroenterol Hepatol 5: 818–828.

Sahota T, Sanderson I, Danhof M, Della Pasqua O (2014a). Model-based analysis of thromboxane B_2 and prostaglandin E_2 as biomarkers in the safety evaluation of naproxen. Toxicol Appl Pharmacol 278: 209–219.

Sahota T, Danhof M, Della Pasqua O (2014b). Utility of model based approaches to predict the risk of adverse events from preclinical toxicology protocols. In: Pharmacology-based toxicity assessment – towards quantitative risk prediction in humans. PhD Thesis, University of Leiden, 2014.

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, Kiyota Y, Levin R, Mogun H *et al.* (2004). Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults Circulation 109: 2068–2073.

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.

Wallace JL (2008). Prostaglandins, NSAIDs, and gastric mucosal protection: why doesn't the stomach digest itself? Physiol Rev 88: 1547–1565.

Whittle BJ (2004). Mechanisms underlying intestinal injury induced by anti-inflammatory COX inhibitors. Eur J Pharmacol 500: 427–439.

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: 100–107.

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: 2031–2039.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

http://dx.doi.org/10.1111/bph.13167

Table S1 Summary of naproxen population PK and PKPD model parameters (Sahota *et al.*, 2014).