RESEARCH PAPER

The Use of ROC Analysis for the Qualitative Prediction of Human Oral Bioavailability from Animal Data

Andrés Olivares-Morales • Oliver J. D. Hatley • David Turner • Aleksandra Galetin • Leon Aarons • Amin Rostami-Hodjegan

ABSTRACT

Purpose To develop and evaluate a tool for the qualitative prediction of human oral bioavailability (F_{human}) from animal oral bioavailability (F_{animal}) data employing ROC analysis and to identify the optimal thresholds for such predictions.
Methods A dataset of 184 compounds with known F_{human} and F_{animal} in at least one species (mouse, rat, dog and non-human primates (NHP)) was employed. A binary classification model for F_{human} was built by setting a threshold for high/low F_{human} at 50%. The thresholds for high/low F_{animal} were varied from 0 to 100 to generate the ROC curves. Optimal thresholds were derived from 'cost analysis' and the outcomes with respect to false negative and false positive predictions were analyzed against the BDDCS class distributions.
Results We successfully built ROC curves for the combined dataset and per individual species. Optimal F_{animal} thresholds were found to be 67% (mouse), 22% (rat), 58% (dog), 35% (NHP) and 47% (combined dataset). No significant trends were observed when sub-categorizing the outcomes by the BDDCS.
Conclusions F_{animal} can predict high/low F_{human} with adequate sensitivity and specificity. This methodology and associated thresholds can be employed as part of decisions related to planning necessary studies during development of new drug candidates and lead selection.

A. Olivares-Morales • O. J. D. Hatley • A. Galetin • L. Aarons •
A. Rostami-Hodjegan (✉)
Centre for Applied Pharmacokinetic Research
School of Pharmacy and Pharmaceutical Sciences
The University of Manchester, Stopford Building, Oxford Road
Manchester M13 9PT, UK
e-mail: amin.rostami@manchester.ac.uk

D. Turner • A. Rostami-Hodjegan
Simcyp Limited, Blades Enterprise Centre, Sheffield, UK

ABBREVIATIONS

AUC	Area under the ROC curve
BCS	Biopharmaceutical classification system
BDDCS	Biopharmaceutical drug disposition classification system
F	Oral bioavailability
fa	Fraction of the dose absorbed in the gastrointestinal tract
F_{animal}	Oral bioavailability in animals/species
F_G	Fraction of the dose absorbed that escapes gut wall first pass metabolism
F_H	Fraction of the dose absorbed that escapes hepatic first pass metabolism
F_{human}	Oral bioavailability in humans
FN	False negative
FP	False positive
GI	Gastrointestinal tract
iv	Intravenous
J	Youden's index
NHP	Non-human primates
NPV	Negative predictive value
PPV	Positive predictive value
QSAR	Quantitative structure-activity relationship
ROC	Receiver operating characteristics
t_A	Animal, high/low oral bioavailability threshold
t_H	Human, high/low oral bioavailability threshold
TN	True negative
TNR	True negative rate
TP	True positive
TPR	True positive rate
US-FDA	United States Food and Drug Administration

INTRODUCTION

Oral bioavailability (F) is considered a key parameter during drug development. It can be defined as the fraction of the dose administered orally that reaches systemic circulation, as its unchanged form, which becomes available at its intended site of action to produce the desired therapeutic effect. Even though the latter is difficult to measure, it is usually assumed that the amount of drug at the site of action is proportional to the amount of drug in plasma/blood. Oral bioavailability is dependent upon the fraction of dose that is absorbed in the gastrointestinal (GI) tract (f_a), as well as the fraction that escapes first pass metabolism in both the GI tract and the liver, F_G and F_H, respectively (Eq. 1)[1,2].

$$F = f_a \times F_G \times F_H \qquad (1)$$

Despite its importance, the information regarding oral drug bioavailability is not always available during the development stages, as it usually requires the administration of an intravenous (iv) dose as a reference. However, due to safety and solubility reasons the iv dose is not always available; therefore, drug oral bioavailability is generally unknown until later stages in the development process. In addition, low oral bioavailability is generally associated with higher inter individual variability [3].

It is a common practice during drug development to employ animal models for the *in vivo* determination of safety, efficacy and pharmacokinetic properties of a drug candidate [1]. The main goal of such studies is to predict drug's behaviour in humans/man based on animal data. The relationship between the animal models and human oral drug bioavailability has been studied on several occasions. One of the first attempts to investigate this relationship was performed by Sietsema in 1989, where a poor correlation was found between animal (rodents, dogs and non-human primates (NHP)) and human oral bioavailability [4]. The lack of correlation could be explained by the interspecies differences in the factors governing oral bioavailability such as morphophysiological differences in GI tract, abundance of transporters and metabolic enzymes and their regional distribution, given that physicochemical properties are inherent to the drug and/or formulation [1,4–11]. Nonetheless, correlations have been established for intestinal permeability and/or the fraction of dose absorbed between humans and preclinical species, in particular for rat and NHP [10,12–14], suggesting that for those particular species, bioavailability differences may rely on the metabolic component.

Based on Sietsema (1989) data, Grass and Sinko [15] plotted the relationship between animal and human oral bioavailability. The plot was similar to a scatter plot and therefore no trends were observed between animal and human bioavailability. However, issues were found with regards to the number of data points in the plot as compared to the original publication from Sietsema (1989), in addition species definitions and the fact of the publication was based on data from 1989 were suggestive that a new dataset was needed in order to perform a comprehensive analysis of the relationship between animal and human oral bioavailability. A recent study by Musther and co-workers [16], addressed those needs by introducing a more comprehensive and updated dataset and the correlations between animal and human oral bioavailability were investigated. Their results were in agreement with the analysis performed in the past by several groups [4,9, 13,17,18]. Amongst the species studied, NHP showed the strongest correlation with human, followed by poor correlations for dog, rat and mouse. However the large prediction intervals suggest that a pointwise correlation between human and preclinical species is not plausible. In terms of qualitative predictions, the study showed that mouse, rat and NHP underpredict human bioavailability. More interestingly, the median of the ratio between animal and human bioavailability was close to unity, however the large intervals for the median ratio suggest that these results should be treated carefully. Despite the suggested lack of predictability of human oral bioavailability from preclinical species, the models are still employed during drug development. The information gathered from oral bioavailability studies in animal models is employed as part of the decision making process of whether to continue or not with the development of any particular drug [19,20].

Marketed drugs and drug candidates can be classified according to extent of metabolism by means of the Biopharmaceutics Drug Disposition System (BDDCS) [21]. Similarly to the Biopharmaceutics Classification System (BCS) [22], the BDDCS divides the compounds into four classes based on their aqueous solubility and permeability. However, the main difference between the two systems is the permeability component. In the BCS, permeability relates to intestinal permeability rate and the extent of absorption, whereas with the BDDCS permeability relates to the permeability rate in the intestine and/or liver, which was found to be correlated with the extent of metabolism [21,23,24]. The BDDCS defines a highly soluble compound as a compound whose highest regulatory approved strength is able to dissolve in 250 mL (or less) of water throughout a pH range of 1 to 7.5 at 37°C. Likewise, the BDDCS defines highly permeable compounds as compounds where 70% or more of the administered oral dose is metabolized. A compound is considered poorly metabolized if 50% or more of the administered dose is excreted in the urine or bile in its unchanged form [25–28]. Thus, BDDCS Class 1 compounds are highly soluble and highly metabolized, Class 2 compounds are poorly soluble and highly metabolized, Class 3 compounds are highly soluble and poorly metabolized and Class 4 compounds are poorly

soluble and poorly metabolized. In addition, it has been suggested that BDDCS class can be useful to estimate the impact of intestinal transporters in drug absorption and metabolism as well as propose possible food effect and clinically relevant drug-drug interactions (21,25–29).

A model can be evaluated based upon its ability to correctly predict any particular outcome, where the prediction performance of a binary classification model can be evaluated by means of Receiver Operating Characteristic (ROC) analysis. The ROC space consists of a plot of the False Positive Rate (FPR) as a function of the true positive rate (TPR). A binary classification is represented by a single point in the ROC space, where a perfect classification will have a FPR of 0 and a TPR of 1. Likewise, a continuous system can be represented by curve in the ROC space and the area under the ROC curve (AUC) can be employed as a measure of the performance of the predictions made from the classification system or model. An AUC of 1 corresponds to a perfect classification/prediction and an AUC of 0.5 corresponds to a random classification/prediction (30–32).

In the present study, we introduce a new model for the categorical prediction of human oral bioavailability, stated as low or high, from animal data by employing a threshold decision tool based upon ROC analysis. Oral bioavailability for more than 180 compounds was analysed in different preclinical species to generate animal oral bioavailability thresholds that can be employed for the qualitative prediction of human oral bioavailability. Furthermore, the relationships between the resulting classifications were compared with the BDDCS classification of the compounds employed for the analysis.

MATERIALS AND METHODS

Dataset Employed

A total of 184 different compounds with reported oral bio-availability in both human and preclinical species, namely mouse, rat, dog and NHP, were employed in this study. The oral bioavailability values for the compounds were collated from the literature by Musther and co-workers, as described elsewhere (16).

Binary Classification and ROC Analysis

A binary classification model was implemented by establishing a threshold (t_H) for high and low human bioavailability at 50% (Eq. 2). A positive outcome (high F_{human}) was defined when F_{human} was greater or equal than 50%, while a negative outcome (low F_{human}) occurred when F_{human} was less than 50%.

$$F_{human} = \begin{cases} high, if \ F_{human} \geq t_H \\ low, if \ F_{human} < t_H \end{cases} \quad (2)$$

In a similar fashion, Eq. 2 was modified to classify high and low F_{animal}, by setting up a threshold for animal oral bioavailability (t_A) as per Eq. 3.

$$F_{animal, \ species} = \begin{cases} high, if \ F_{animal, \ species} \geq t_A \\ low, if \ F_{animal, \ species} < t_A \end{cases} \quad (3)$$

Thus for F_{animal} based predictions, false positives (FP) were compounds with high F_{animal} and low F_{human}, and false negatives (FN) were compounds with low F_{animal} and high F_{human} (Fig. 1). If both F_{animal} and F_{human} were high, the compound was classified as a true positive (TP), similarly if both F_{animal} and F_{human} were low, the compound was considered as a true negative (TN). The predictions based on the animal data analysis for bioavailability were evaluated by calculating its sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the determined animal thresholds (t_A), as shown in Table I.

All the statistical analysis and the ROC curve generation were implemented with Matlab® 2012a and its statistical toolbox (The MathWorks Inc., Natick, MA, USA). The construction of the ROC curves, for the overall dataset and by species, was achieved by varying t_A form 0 to 100 and recording the error rates (Table I) for each threshold. The overall (all the species combined) ROC curve was constructed by considering all the datapoints within the dataset, including the compounds with F_{animal} values in more than one preclinical
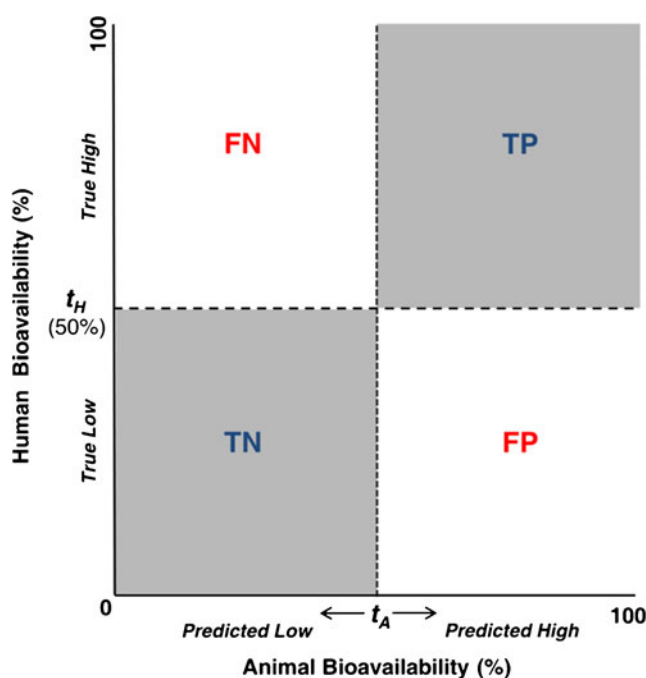

Fig. 1 Threshold based predictions of human oral bioavailability from animal data. FN, False negatives; TP, True positives; TN, True negatives; FP, False positive; t_A, Animal high/low bioavailability threshold; t_H, human high/low bioavailability threshold.

Table I Definitions and Formulae for the Evaluation of the Binary Classification System

Parameter	Formula	Probability
Sensitivity or TPR	$\frac{TP}{TP+FN}$	$P[\text{ high } F_{animal} \mid \text{high } F_{human}]$
Specificity or TNR	$\frac{TN}{TN+FP}$	$P[\text{ low } F_{animal} \mid \text{low } F_{human}]$
PPV	$\frac{TP}{TP+FP}$	$P[\text{ high } F_{human} \mid \text{high } F_{animal}]$
NPV	$\frac{TN}{TN+FN}$	$P[\text{low } F_{human} \mid \text{low } F_{animal}]$
FPR	$1-TNR$	$P[\text{ high } F_{animal} \mid \text{low } F_{human}]$

species. In addition, a bootstrap random resampling ($n = 10000$), with replacement, was performed for the determination of the average ROC curves and the bias corrected confidence intervals for the resulting AUC, sensitivity and specificity. The curves were averaged by threshold averaging function within Matlab®. The significance of the differences between the resulting AUCs (for each species and the overall dataset) and the random classification (AUC=0.5) were calculated by the non-parametric Mann-Whitney U test. Determinations of the optimal thresholds for the averaged ROC curves were calculated as the interception of a line of slope, S (Eq. 4), with the averaged ROC curve.

$$S = \frac{cost(FP)-cost(TN)}{cost(FN)-cost(TP)} \times \frac{N}{P} \qquad (4)$$

Where, *cost(FP)* and *cost(FN)* are the costs of FP and FN, respectively; *cost(TN)* and *cost(TP)* are the costs of TN and TP, respectively; N and P, are the number of positives and negatives values based on F_{human} data (32). The net costs for TN and TP were assumed to be 0, initially the net costs for FP and FN were assumed equal. Subsequently, the cost ratio between FP and FN was varied in order to evaluate the impact of the cost assumptions on the determination of the optimal threshold points.

For rat, dog and NHP, alternative thresholds were determined by visually comparing the relationship between animal threshold (t_A) *versus* its resulting sensitivity and specificity; the closest points to the intersection lines between the latter two were considered as alternative thresholds. New thresholds were compared against the optimal thresholds derived from cost analysis employing Youden's index (J) (Eq. 5), where the maximum value for J is 1 for a perfect classification, whereas the minimum value is 0 for a threshold with no predictive power (33).

$$J = sensitivity + specificity - 1 \qquad (5)$$

An additional ROC analysis was performed for the compounds in the dataset whose F_{human} values were ≤20% (very low oral bioavailability) in the overall dataset, and for these compounds the human threshold for high/low oral bioavailability was set up at 10%. The analysis was conducted using rat, dog and NHP data and the significance of the AUC and determination of optimal thresholds was performed, as described above (Eq. 4).

BDDCS Classification

The BDDCS classification for 155 of the compounds in the current study was obtained from the dataset published by Benet, and co-workers (34), while for the remaining compounds, a provisional BDDCS classification was given. All individual drugs used for the analysis and their allocated BDDCS classification are listed in Tables SI and SII of the Supplementary Material. For the classification, the extent of metabolism and/or percentage of the dose excreted unchanged in the urine and/or bile was collated from the literature. With regards to solubility determination, for US-FDA approved drugs, the maximum dose strength was taken from the label, when possible; otherwise the maximum dose strength was extracted from published data. Aqueous solubility values were taken from published data when possible, otherwise the "Mass Solubility" value within SciFinder® chemical properties was employed (35). High/low solubility was assigned using Eq. 6, where Dn is the dose number, D_{max} is the maximum dose strength (mg) and Cs is the aqueous solubility (mg/mL). High solubility was assigned to compounds with Dn equal or greater than 1, while low solubility was assigned for compounds with a Dn less than 1.

$$Dn = \frac{D_{max}/250 \ mL}{C_s} \qquad (6)$$

Class distribution was then compared within each of the threshold-based outcome groups (i.e. true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN)) for the rat, dog and NHP. Significance of the difference between the BDDCS class distribution within each of the outcome groups (i.e. TP, TN, FP, FN) were evaluated by visual inspection of the plots and by employing Fisher's exact test implemented in the R statistics Package (http://www.r-project.org/).

RESULTS

The analysed dataset consisted in total of 318 data points for animal and human oral bioavailability divided in 30, 122, 125 and 41 pairwise correlations between mouse, rat, dog and NHP with human oral bioavailability, respectively (Fig. 2a). For some of the compounds in the dataset oral bioavailability

values were available for more than one species as shown in Fig. 2b for rat, dog and NHP.

The summary of resulting ROC curves for the total dataset of 318 F_{animal} values and per individual species are shown in Table II. Combination of all F_{animal} values resulted in a smooth ROC curve with an AUC around 0.8 (Fig. 3). A similar outcome was observed employing individual dog and rat data (Fig. 4b, c) with ROC curve AUC for dog data slightly higher than for the rat (0.8 and 0.7, respectively). In contrast, ROC curves generated for mouse and NHP resulted in a step-like curve rather than a smooth curve (Fig. 4a, d), primarily attributed to the limited number of data points employed in their construction. However, the AUC values of the latter ROC curves were higher than for the overall dataset and the rest of the species. Interestingly, NHP showed an AUC value close to the unity (0.96). With regards to the significance tests for the above AUCs - all the species combined and by species- showed high significance levels compared to the random classification (AUC=0.5) yet for mouse data, *p-value* was slightly higher than for the rest of the species.

The 'cost analysis' determinations of the optimal F_{animal} thresholds are summarized in Tables III and IV For the overall dataset, threshold was 47%, in agreement with the 50% value employed for the human bioavailability threshold with resulting specificity and sensitivity of 0.82 and 0.66, respectively. A similar situation occurred for dog data, though the threshold was slightly higher than for the overall dataset, whereas specificity and sensitivity were balanced with values close to 0.75. The highest threshold was found when employing mouse data (t_A =67%), which gave rise to perfect specificity but this should be viewed with caution given the limited data available (n =30). Rat and NHP showed thresholds <50%, whilst the corresponding specificity for the rat was lower than for any species investigated. In agreement with the AUC values, NHP showed the highest specificity and sensitivity for its optimal threshold. Details of the drugs classified as
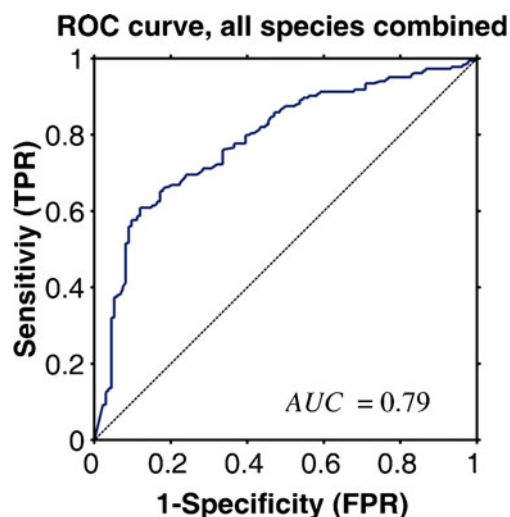

Fig. 3 Averaged ROC curve for the human *versus* animal bioavailability dataset for all the preclinical species (mouse, rat, dog and NHP) combined. The dashed line corresponds to the line for random classification, AUC = 0.79 for the overall dataset.

FN, FP, TN and TP according to the aforementioned thresholds are shown in Tables SIII and SIV of the Supplementary Material.

The NPV and PPV values showed an opposite relationship to their corresponding sensitivities and specificities; this trend was evident for all of the species and the overall dataset. All of the corresponding PPV values were above 0.8, with the exception of the rat (0.72). Interestingly, NPV values were similar for almost all of the species and the overall dataset; however, corresponding values were higher for NHP. The analysis of the impact of the ratio between the cost of FP and FN on the resulting thresholds showed a sigmoid like increase on the threshold value when varying the ratio from 0.0025 to 20 (Fig. 5). However, in some regions the thresholds remained invariant to changes in the FP/FN ratios.
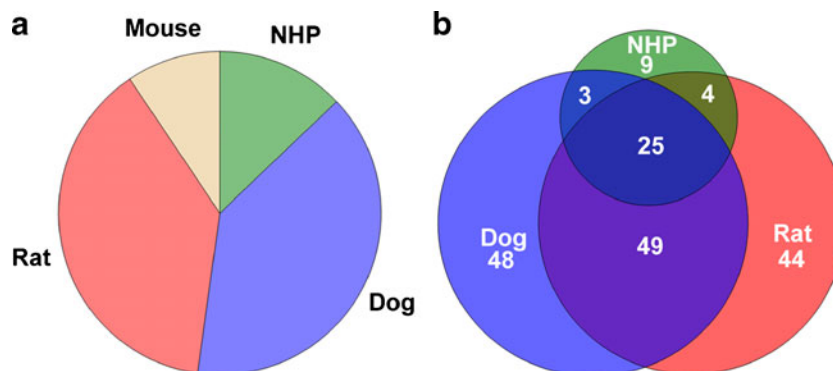

Fig. 2 (**a**) Pie chart of the distribution of the oral bioavailability data points employed for the analysis by species, mouse (n = 30), rat (n = 122), dog (n = 125) and non-human primates (NHP) (n = 41). (**b**) Venn diagram of the relationship between oral bioavailability data points for rat, dog and NHP. The area of the circles represents the number of compounds with oral bioavailability data for both animal species and humans, the areas of the interception represents the number of compounds with bioavailability data for more than one species.
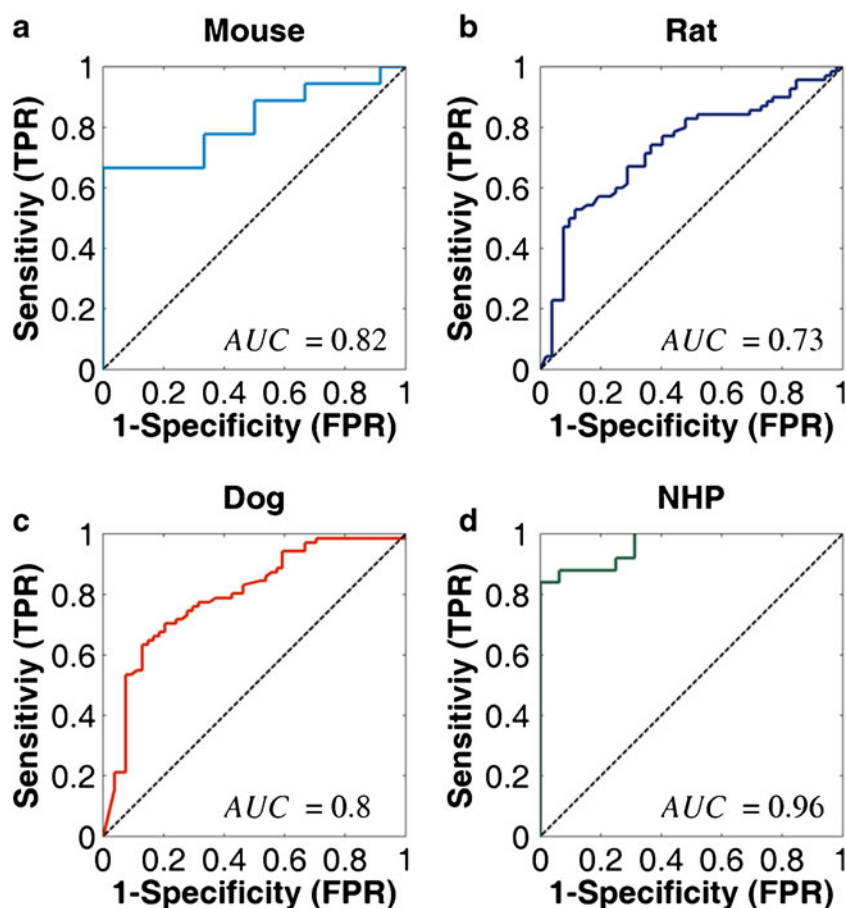
Fig. 4 Averaged ROC curves for the human *versus* animal bioavailability dataset by preclinical species. (**a**) Mouse ROC curve, AUC = 0.82; (**b**) Rat ROC curve, AUC = 0.73; (**c**) Dog ROC curve, AUC = 0.80; (**d**) NHP ROC curve, AUC = 0.96; Dashed line corresponds to the line for random classification.

Visual inspection of the plots in Fig. 6 led to alternative animal thresholds at 28%, 54%, and 31% for rat, dog and NHP, respectively (Table III). Even though specificity and sensitivity were balanced for all the species, NPV and PPV showed similar relationships to the previous analysis. Yet, the alternative thresholds showed no improvement of the overall predictability of animal data for the particular species as compared with the former thresholds derived from the cost analysis. Youden's index (J) values for alternative thresholds

were lower than the values achieved by the aforementioned thresholds (Tables III and V).

ROC Analysis of the reduced dataset of compounds with very low F_{human} was not possible for NHP due to the lack of compounds. Analysis of dog data ($n = 11$) showed that this preclinical animal model cannot be applied to the categorical prediction of very low bioavailability compounds since AUC was not significantly different from the random classification (data not shown). Nevertheless, for rat data ($n = 34$) ROC

Table II Area Under the ROC Curve for Animal Models

Species	n^a	AUC	Lower 95% confidence interval [b]	Upper 95% confidence interval [b]	p-value
All	318	0.786	0.734	0.835	<0.0001
Mouse	30	0.819	0.613	0.936	<0.005
Rat	122	0.731	0.634	0.818	<0.0001
Dog	125	0.796	0.708	0.870	<0.0001
NHP[c]	41	0.963	0.871	0.992	<0.0001

[a] n, number of data points employed for the determination of the ROC curve

[b] determined by bootstrap ($n = 10000$)

[c] NHP, non-human primates

Table III Cost Analysis Derived Optimal Thresholds for F_{Animal} and its Corresponding Evaluation Metrics

Species	Opt. t_A (%)	Specificity (95% CI)[a]	Sensitivity (95% CI)[a]	J
All	47	0.82 (0.75, 0.88)	0.66 (0.59, 0.73)	0.48
Mouse	67	1.00 (1.00, 1.00)	0.67 (0.41, 0.86)	0.67
Rat	22	0.60 (0.46, 0.72)	0.77 (0.67, 0.86)	0.37
Dog	58	0.80 (0.67, 0.89)	0.70 (0.59, 0.80)	0.50
NHP	35	1.00 (1.00, 1.00)	0.84 (0.65, 0.96)	0.84

Opt. t_A, optimal threshold for F_{animal}; 95% CI, 95% confidence interval; J, Youden's Index as per Eq. 5

[a] determined by bootstrap ($n = 10000$)

Table IV Cost Analysis Derived Optimal Thresholds for F_{Animal} and its Corresponding PPV and NPV

Species	Opt. t_A (%)	NPV	PPV
All	47	0.64	0.84
Mouse	67	0.67	1.00
Rat	22	0.66	0.72
Dog	58	0.67	0.82
NHP	35	0.80	1.00

Opt. t_A, optimal threshold for F_{animal}; NPV, negative predictive value; PPV, positive predictive value

analysis was possible (Fig. 7) and the results showed that the rat model provided a significantly improvement on the predictions as compared to the random classification (AUC=0.77; 0.52–0.91, 95% confidence interval (CI)). In this case, the threshold for high/low F_{animal} was 7.4% with a corresponding sensitivity of 0.95(0.75–1.00, 95% CI) and specificity of 0.67(0.40–0.88, 95% CI), indicating that rat can be used for predictions of very low values of human bioavailability predictions.

Provisional BDDCS classification was assigned to 29 of the compounds in the dataset (Table SI in the Supplementary Material). The majority of the compounds analysed were classified as Class 1 (47%), followed by Classes 3 (24%), 2 (22%) and 4 (7%). The proportion of compounds classified as BDDCS Class 1 to 4 for every category after the application of the classification model (i.e., TP, FN, TN and FP) is represented in Fig. 8. The results from Fisher's test showed no significant difference (*p value* > 0.05) between the overall BDDCS class distribution ("Ini." In Fig. 8) and the
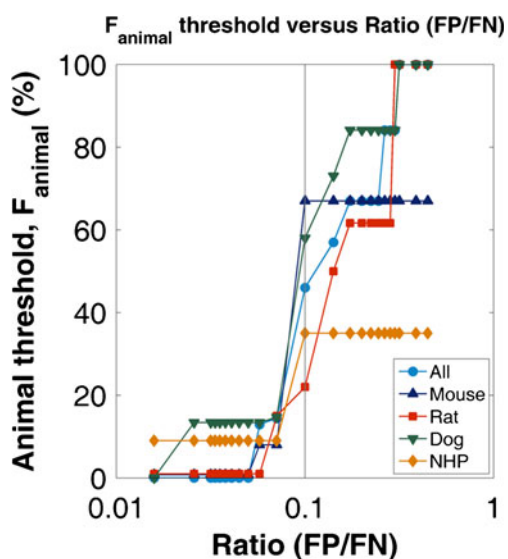

Fig. 5 Impact of the FP/FN cost ratio on the determination of the optimal thresholds by Eq. 4. Sky-blue line and circles, thresholds for the combined dataset; Blue line and upper triangles, mouse thresholds; Red line and squares, rat thresholds; Green line and lower triangles, dog thresholds; Yellow line and diamonds, NHP thresholds.

distribution within each threshold model classes (FN, TP, TN and FP) for any of the preclinical species.

DISCUSSION

Several groups have suggested that pointwise correlation of oral bioavailability between preclinical species and human is almost non-existent and, therefore, data from such studies in preclinical species should be treated with caution if the intention is to predict human oral bioavailability quantitatively (4,15,16,36). The current study performed a systematic analysis of F_{animal} data from four different species, including 184 compounds across all BDDCS classes. Rather than attempting to predict a particular value for F_{human} employing F_{animal} data, ROC analysis performed here showed that the animal data can be employed for a categorical prediction of high or low human bioavailability. The area under the ROC curve is representative of the probability of correctly classifying F_{human} as high and low when employing F_{animal} data: in our analysis the probability was around 80% which can be considered as a high value (30). Even though mouse data showed similar performance, the reduced number of data points employed in the analysis limit us from making any conclusions about the utility of this preclinical model for making this type of prediction. On the contrary, the higher abundance of rat and dog data during preclinical development stages was reflected in the smother ROC curves built for those particular species and the resulting high probabilities for a correct classification employing those species. This finding is in contrast to previous attempts to establish correlations between animal and human oral bioavailability (10,13,15,16,37).

Albeit with a much smaller sample size than rat and dog, NHP showed the best performance of all the species with a probability for a correct classification around 100%. The results are consistent with previous findings for this particular species from pointwise correlation analyses (13,16). However, the cost and ethical implications on the use of NHP during preclinical development will limit their use to later stages of the drug development process.

The resulting thresholds for the cost analysis in rat, dog and NHP are consistent with the correlation analysis performed by Musther and co-workers (16). Rat and NHP proved to be underpredictive of human oral bioavailability and in the same fashion the thresholds for high/low F_{Animal} were lower than the threshold for high/low F_{human}. In contrast, threshold for dog (58%) was relatively close to the human value, which is consistent with the observed trend in the dog to neither underpredict nor overpredict human oral bioavailability. The model achieved high TPR (>0.70) and acceptable FPR (≤0.40) employing the aforementioned thresholds for all the preclinical species. The former suggests that it is unlikely to have low oral bioavailability values in human when having high oral bioavailability in any of the preclinical species.
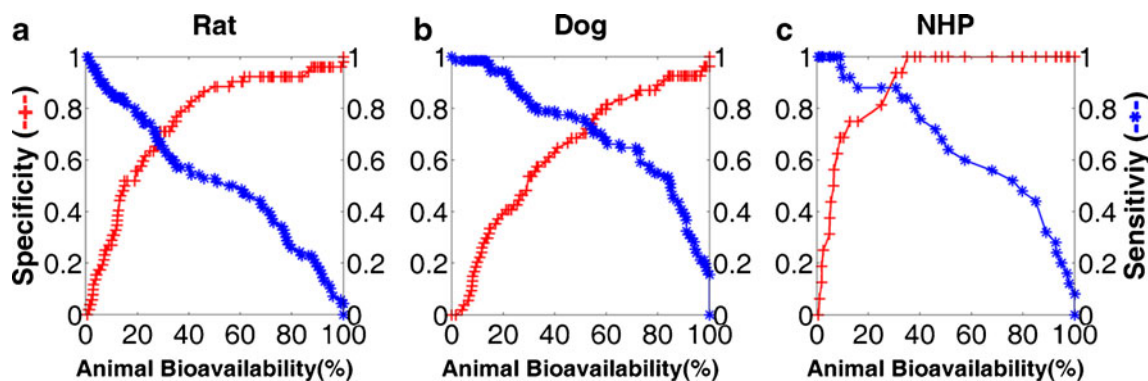
Fig. 6 Sensitivity and specificity as a function of F_{animal} thresholds for rat (**a**), dog (**b**) and NHP (**c**). An increase on the thresholds will increase the specificity but at the same time will decrease its sensitivity, the thresholds closer to the intercept between the two lines were chosen as alternative thresholds. Red line and cross (-+-), specificity; Blue line and asterisk (-*-), sensitivity.

When we determined the optimal thresholds by an alternative method, the results were similar to the ones based on cost analysis for rat, dog and NHP. In addition, the new thresholds did not provide any significant improvement of the ability to predict human high/low bioavailability from animal data. The optimal thresholds may also be determined by other alternative methods, such as the Euclidean distance between the ROC curve and the hypothetical perfect classification (e.g., point 0, 1 in the ROC space) [38]. However, this particular method does not take into account the nature of the dataset unlike the 'cost analysis' determination based on Eq. 4.

Determination of the optimal thresholds according to Eq. 4 will depend upon the assigned net cost of FP and FN. In a conservative fashion, a higher cost can be assigned to the FP and the thresholds will move to a higher value and the opposite will occur when assigning a higher net cost to FN. However, as shown in Fig. 5, the increase of the ratio follows a step like increase in the thresholds (showing insensitive regions to the assigned net cost ratios). For example a two-fold increase on the net cost of a FP compared to the net cost of FN will increase the threshold for both rat and dog, whereas a three-fold increase will have the same impact on the resulting thresholds as a tenfold increase. However, moving the thresholds will affect the corresponding sensitivity and specificity

(Fig. 6) and thus the thresholds represent the best estimator based on a balanced situation.

Even though rat predictions showed higher FPR than the rest of the preclinical species, its convenience and availability during preclinical development stages make it one the commonest species employed to generate bioavailability data for new drug candidates. Our threshold for rat predictions of high/low F_{human} was 22%, consistent with similar thresholds ($F \geq 20$–30%) for the evaluation of rat F_{animal} during drug development published by different groups in the industrial setting [19,20]. In contrast to the current analysis, previously proposed thresholds fail to provide any indication of the expected range for human bioavailability and also the corresponding TPR and FPR for the rat based predictions. From the analysis of the reduced dataset for very low human oral bioavailability (<20%), rat data showed the best performance across species for prediction of F_{human} in this range, with the probability for a correct prediction of high/low F_{human} of 77%.

Table V Alternative Thresholds for F_{animal} and its Corresponding Evaluation Metrics

Species	Alt. t_A (%)	Specificity (95% CI)[a]	Sensitivity (95% CI)[a]	NPV	PPV	Youden's Index (J)
Rat	28	0.65 (0.51, 0.77)	0.67 (0.55, 0.77)	0.6	0.72	0.33
Dog	54	0.72 (0.58, 0.83)	0.73 (0.62, 0.83)	0.67	0.78	0.46
NHP	31	0.94 (0.67, 1.00)	0.88 (0.69, 0.96)	0.83	0.96	0.82

Alt. t_A, alternative threshold for F_{animal}; 95% CI, 95% confidence interval; NPV, negative predictive value; PPV, positive predictive value; J, Youden's Index as per Eq. 5

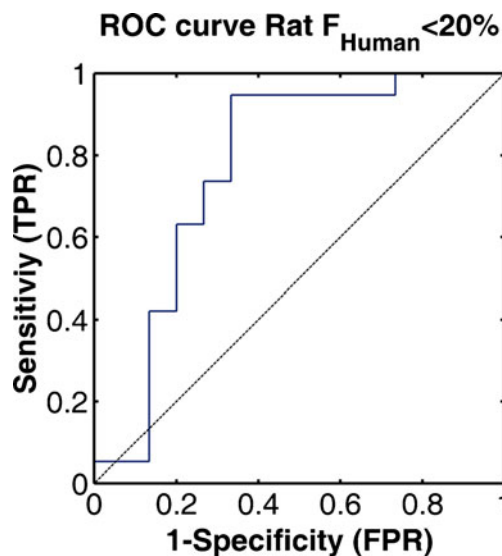[a] determined by bootstrap ($n = 10000$)

Fig. 7 ROC curve for rat predictions of very low human bioavailability ($F_{human} \leq 20\%$) dataset. The dashed line corresponds to the line of random classification.
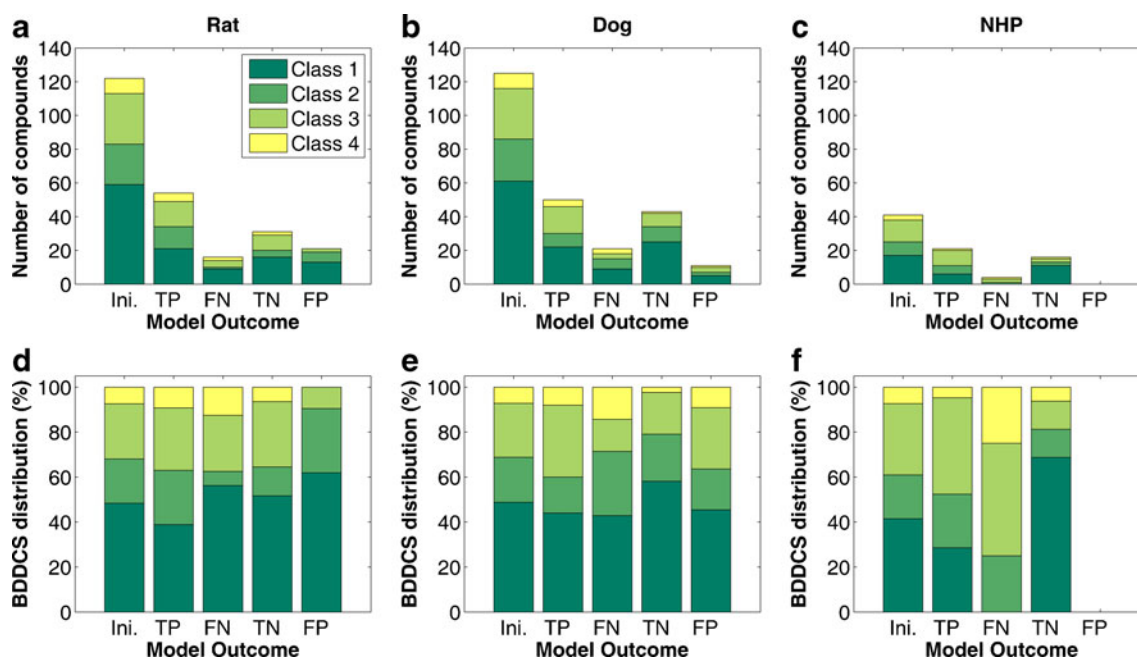
Fig. 8 Number of compounds and BDDCS class distribution for rat (**a**), dog (**b**) and NHP (**c**) as a function of the outcome of the threshold based model. (**d**, **e** and **f**) BDDCS class distribution for the outcome (in percentage of the outcome groups) for rat, dog and NHP, respectively. Ini., initial number of compounds; TP, compounds classified as true positives, FN, compounds classified as false negatives; TN, compounds classified as true negatives; FP, compounds classified as false positives.

The threshold (7.4%), in this case was close the human threshold (10%), with a high TPR of 0.95 and with a small decrease in the FPR to 0.33, confirming further the suitability of the rat data for qualitative predictions of F_{human} during drug development. Nevertheless, the shape of the ROC curve (Fig. 7) suggests that it is necessary to obtain more data to validate this conclusion.

The dataset employed for the generation of this predictive model was based upon successful candidates and hence a relatively high prevalence (58%) of compounds with high F_{human} values was observed. It would be of interest to test this model with a dataset reflective of the "true" drug development process. However, we do not expect a high variation on the rat threshold nor the resulting TNR and TPR , as our threshold was in line with the ones employed in the development setting and the latter are insensitive to the prevalence of high F_{human} values (33).

ROC analysis for the evaluation of a predictive model of human bioavailability has been applied before. Langdon and Barret (39) developed a model for the prediction of oral bioavailability in both rat and human based on QSAR and genetic programing (GP). The evaluation of such a model was based on the ability of the model to predict high ($F \geq 33\%$) and low ($F < 33\%$) bioavailability in both human and rat. Interestingly TNR and FNR were in agreement with results of our study.

BDDCS distributions of the drugs investigated in our analysis were in agreement with the observed distribution for marketed drugs, where Classes 1 and 3 are the predominant

classes. Nevertheless, a different distribution can be expected for new drug candidates, with a higher tendency for Class 2 and 3 drugs (23,25). However, our analysis on the outcome of the classification model showed no clear tendency for any particular BDDCS Class. One reason for this observation can be attributed to the fact that BDDCS is a categorical tool with no clear distinction between overall metabolism and first pass metabolism. In addition, the binary nature of both BCS and BDDCS does not account for the continuity of the properties employed within them- e.g., solubility, intestinal permeability and/or extent of metabolism- which might be an issue for the middle range of those properties and therefore affecting the actual class distribution. Finally, interspecies differences in metabolic activity are not accounted for by the existing BDDCS, and therefore the class distribution for a particular drug could be different in a particular preclinical species.

Even though our methodology was applied on animal oral bioavailability data, a similar approach could be employed to its constituents-*fa*, F_G and F_H-which might be of more interest for some groups involved in preclinical development. However, we suggest that a more mechanistic approach should be attempted if the intention is to quantitatively predict drug's behaviour in man, taking in consideration the physiological differences between the preclinical species and humans in addition to drug's intrinsic characteristic that can impact on drug's oral bioavailability and its individual components.

CONCLUSION

A new method for categorical prediction of high and low human bioavailability from animal bioavailability data was developed by employing ROC analysis. Oral bioavailability of more than 50% in animals can successfully predict high oral bioavailability in human, with high TPR and low FPR. A similar scenario can be expected for bioavailability values equal or greater than 22%, 58% and 35% in rat, dog and NHP. Even though NHP was the best predictor of F_{Human}, rat was shown to be the best predictor for low human oral bioavailability, supporting the use of this animal model for the F_{human} predictions in mid to early stages of drug development. Thresholds proposed in the current study can be employed in the pharmaceutical industry as part of the tool box of methods for making decisions related to planning necessary studies during the development of new drug candidates and lead selection.

ACKNOWLEDGMENTS AND DISCLOSURES

REFERENCES

1. Lin JH. Species similarities and differences in pharmacokinetics. Drug Metab Dispos. 1995;23:1008–21.
2. Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A. The Simcyp® population-based ADME simulator. Expert Opin Drug Metab Toxicol. 2009;5:211–23.
3. Hellriegel ET, Bjornsson TD, Hauck WW. Interpatient variability in bioavailability is related to the extent of absorption: implications for bioavailability and bioequivalence studies. Clin Pharmacol Ther. 1996;60:601–7.
4. Sietsema WK. The absolute oral bioavailability of selected drugs. Int J Clin Pharmacol Ther Toxicol. 1989;27:179–211.
5. Chhabra RS, Pohl RJ, Fouts JR. A comparative study of xenobiotic-metabolizing enzymes in liver and intestine of various animal species. Drug Metab Dispos. 1974;2:443–7.
6. Nishimuta H, Nakagawa T, Nomura N, Yabuki M. Species differences in hepatic and intestinal metabolic activities for 43 human cytochrome P450 substrates between humans and rats or dogs. Xenobiotica. 2013. doi:10.3109/00498254.2013.787155.
7. Nishimuta H, Sato K, Mizuki Y, Yabuki M, Komuro S. Species differences in intestinal metabolic activities of cytochrome P450 isoforms between cynomolgus monkeys and humans. Drug Metab Pharmacokinet. 2011;26:300–6.
8. Zamek-Gliszczynski MJ, Lee CA, Poirier A, Bentz J, Chu X, Ellens H, et al. ITC recommendations for transporter kinetic parameter estimation and translational modeling of transport-mediated PK and DDIs in humans. Clin Pharmacol Ther. 2013;94:64–79.
9. Akabane T, Tabata K, Kadono K, Sakuda S, Terashita S, Teramura T. A comparison of pharmacokinetics between humans and monkeys. Drug Metab Dispos. 2010;38:308–16.
10. Cao X, Gibbs S, Fang L, Miller H, Landowski C, Shin H-C, et al. Why is it challenging to predict intestinal drug absorption and oral bioavailability in human using rat model. Pharm Res. 2006;23:1675–86.
11. Bueters T, Juric S, Sohlenius-Sternbeck AK, Hu Y, Bylund J. Rat poorly predicts the combined non-absorbed and presystemically metabolized fractions in the human. Xenobiotica. 2013;43:607–16.
12. Chiou WL, Barve A. Linear correlation of the fraction of oral dose absorbed of 64 drugs between humans and rats. Pharm Res. 1998;15:1792–5.
13. Chiou WL, Buehler PW. Comparison of oral absorption and bioavailability of drugs between monkey and human. Pharm Res. 2002;19:868–74.
14. Zhao YH, Abraham MH, Le J, Hersey A, Luscombe CN, Beck G, et al. Evaluation of rat intestinal absorption data and correlation with human intestinal absorption. Eur J Med Chem. 2003;38:233–43.
15. Grass GM, Sinko PJ. Physiologically-based pharmacokinetic simulation modelling. Adv Drug Deliv Rev. 2002;54:433–51.
16. Musther H, Olivares-Morales A, Hatley OJD, Liu B, and Rostami-Hodjegan A. Animal versus human oral drug bioavailability: do they correlate? Eur J Pharm Sci. 2013. doi:10.1016/j.ejps.2013.08.018.
17. Takahashi M, Washio T, Suzuki N, Igeta K, Yamashita S. The species differences of intestinal drug absorption and first-pass metabolism between cynomolgus monkeys and humans. J Pharm Sci. 2009;98:4343–53.
18. Ward KW, Nagilla R, Jolivette LJ. Comparative evaluation of oral systemic exposure of 56 xenobiotics in rat, dog, monkey and human. Xenobiotica. 2005;35:191–210.
19. Caldwell GW, Ritchie DM, Masucci JA, Hageman W, Yan Z. The new pre-preclinical paradigm: compound optimization in early and late phase drug discovery. Curr Top Med Chem. 2001;1:353–66.
20. Thomas VH, Bhattachar S, Hitchingham L, Zocharski P, Naath M, Surendran N, et al. The road map to oral bioavailability: an industrial perspective. Expert Opin Drug Metab Toxicol. 2006;2:591–608.
21. Wu C-Y, Benet LZ. Predicting drug disposition via application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. Pharm Res. 2005;22:11–23.
22. Amidon GL, Lennernas H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm Res. 1995;12:413–20.
23. Broccatelli F, Cruciani G, Benet LZ, Oprea TI. BDDCS Class prediction for new molecular entities. Mol Pharm. 2012;9:570–80.
24. Chen ML, Amidon GL, Benet LZ, Lennernas H, Yu LX, The BCS. BDDCS, and regulatory guidances. Pharm Res. 2011;28:1774–8.
25. Benet LZ. The role of BCS (biopharmaceutics classification system) and BDDCS (biopharmaceutics drug disposition classification system) in drug development. J Pharm Sci. 2013;102:34–42.
26. Benet LZ. Predicting drug disposition via application of a biopharmaceutics drug disposition classification system. Basic Clin Pharmacol Toxicol. 2010;106:162–7.
27. Custodio JM, Wu CY, Benet LZ. Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. Adv Drug Deliv Rev. 2008;60:717–33.

28. Benet L, Amidon G, Barends D, Lennernäs H, Polli J, Shah V, *et al*. The use of BDDCS in classifying the permeability of marketed drugs. Pharm Res. 2008;25:483–8.

29. Shugarts S, Benet LZ. The role of transporters in the pharmacokinetics of orally administered drugs. Pharm Res. 2009;26:2039–54.

30. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.

31. Fawcett T. An introduction to ROC analysis. Patttern Recogn Lett. 2006;27:861–74.

32. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283–98.

33. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care. 2004;8:508–12.

34. Benet LZ, Broccatelli F, Oprea TI. BDDCS applied to over 900 drugs. AAPS J. 2011;13:519–47.

35. Chemical Abstracts Service. SciFinder®, version 2013, American chemical society. https://scifinder.cas.org/scifinder/view/scifinder/scifinderExplore.jsf (accessed March 30, 2013). Calculated using Advanced Chemistry Development (ACD/Labs) Software V11.02 (© 1994–2013 ACD/Labs).

36. Shanks N, Greek R, Greek J. Are animal models predictive for humans? Philosophy, Ethics, and Humanities in Medicine. 2009;4:2.

37. Chiou WL, Jeong HY, Chung SM, Wu TC. Evaluation of using dog as an animal model to study the fraction of oral dose absorbed of 43 drugs in humans. Pharm Res. 2000;17:135–40.

38. Pérez-Garrido A, Helguera AM, Borges F, Cordeiro MNDS, Rivero V, Escudero AG. Two new parameters based on distances in a receiver operating characteristic chart for the selection of classification models. J Chem Inf Model. 2011;51:2746–59.

39. Langdon WB, Barrett SJ. Genetic programming in data mining for drug discovery. In Ghoshand A, Jain L (eds), Evolutionary computation in data mining, Vol. 163, Springer Berlin Heidelberg; 2005. p. 211–235.