IMI – Oral biopharmaceutics tools project – Evaluation of bottom-up PBPK prediction success part 3: Identifying gaps in system parameters by analysing In Silico performance across different compound classes

DOI:
10.1016/j.ejps.2016.09.037

Document Version
Accepted author manuscript

Published in:
European Journal of Pharmaceutical Sciences

IMI – Oral biopharmaceutics tools project – Evaluation of bottom-up PBPK prediction success part 3: Identifying gaps in system parameters by analysing *In Silico* performance across different compound classes

Adam S. Darwich, Alison Margolskee, Xavier Pepin, Leon Aarons, Aleksandra Galetin, Amin Rostami-Hodjegan, Sara Carlert, Maria Hammarberg, Constanze Hilgendorf, Pernilla Johansson, Eva Karlsson, Dónal Murphy, Christer Tannergren, Helena Thörn, Mohammed Yasin, Florent Mazuir, Olivier Nicolas, Sergej Ramusovic, Christine Xu, Shriram M. Pathak, Timo Korjamo, Johanna Laru, Jussi Malkki, Sari Pappinen, Johanna Tuunainen, Jennifer Dressman, Simone Hansmann, Edmund Kostewicz, Handan He, Tycho Heimbach, Fan Wu, Carolin Hoft, Yan Pang, Michael B. Bolger, Eva Huehn, Viera Lukacova, James M. Mullin, Ke X. Szeto, Chester Costales, Jian Lin, Mark McAllister, Sweta Modi, Charles Rotter, Manthena Varma, Mei Wong, Amitava Mitra, Jan Bevernage, Jeike Biewenga, Achiel Van Peer, Richard Lloyd, Carole Shardlow, Peter Langguth, Irina Mishenzon, Mai Anh Nguyen, Jonathan Brown, Hans Lennernäs, Bertil Abrahamsson

PII: S0928-0987(16)30420-1
DOI: doi:10.1016/j.ejps.2016.09.037
Reference: PHASCI 3743

To appear in:

Received date: 10 May 2016
Revised date: 23 August 2016
Accepted date: 26 September 2016

IMI – Oral Biopharmaceutics Tools project – Evaluation of Bottom-up PBPK Prediction Success Part 3: Identifying Gaps in System Parameters by Analysing *In Silico* Performance across Different Compound Classes

Adam S. Darwich[1], Alison Margolskee[1], Xavier Pepin[2,3], Leon Aarons[1], Aleksandra Galetin[1], Amin Rostami-Hodjegan[1,4], Sara Carlert[5], Maria Hammarberg[5], Constanze Hilgendorf[5], Pernilla Johansson[5], Eva Karlsson[5], Dónal Murphy[2], Christer Tannergren[5], Helena Thörn[5], Mohammed Yasin[2], Florent Mazuir[3], Olivier Nicolas[3], Sergej Ramusovic[6], Christine Xu[7], Shriram M. Pathak[4], Timo Korjamo[8], Johanna Laru[2,8], Jussi Malkki[8], Sari Pappinen[8], Johanna Tuunainen[8], Jennifer Dressman[9], Simone Hansmann[9], Edmund Kostewicz[9], Handan He[10], Tycho Heimbach[10], Fan Wu[10], Carolin Hoft[11], Yan Pang[11], Michael B. Bolger[12], Eva Huehn[12], Viera Lukacova[12], James M. Mullin[12], Ke X. Szeto[12], Chester Costales[13], Jian Lin[13], Mark McAllister[14], Sweta Modi[13], Charles Rotter[13], Manthena Varma[14], Mei Wong[14], Amitava Mitra[15], Jan Bevernage[16], Jeike Biewenga[16], Achiel Van Peer[16], Richard Lloyd[17], Carole Shardlow[17], Peter Langguth[18], Irina Mishenzon[18], Mai Anh Nguyen[18], Jonathan Brown[19], Hans Lennernäs[20], Bertil Abrahamsson[5]

[1]University of Manchester, United Kingdom

[2]AstraZeneca, United Kingdom

[3]Sanofi, France

[4]Simcyp Ltd, United Kingdom

[5]AstraZeneca, Sweden

[6]Sanofi, Germany

[7]Sanofi, United States

[8]Orion Pharma, Finland

[9]Goethe University Frankfurt am Main, Germany

[10]Novartis, United States

[11]AbbVie, Germany

[12]Simulations Plus, Inc., United States

[13]Pfizer, United States

[14]Pfizer, United Kingdom

[15]Merck Sharp & Dohme (MSD), United Kingdom

[16]Janssen, Belgium

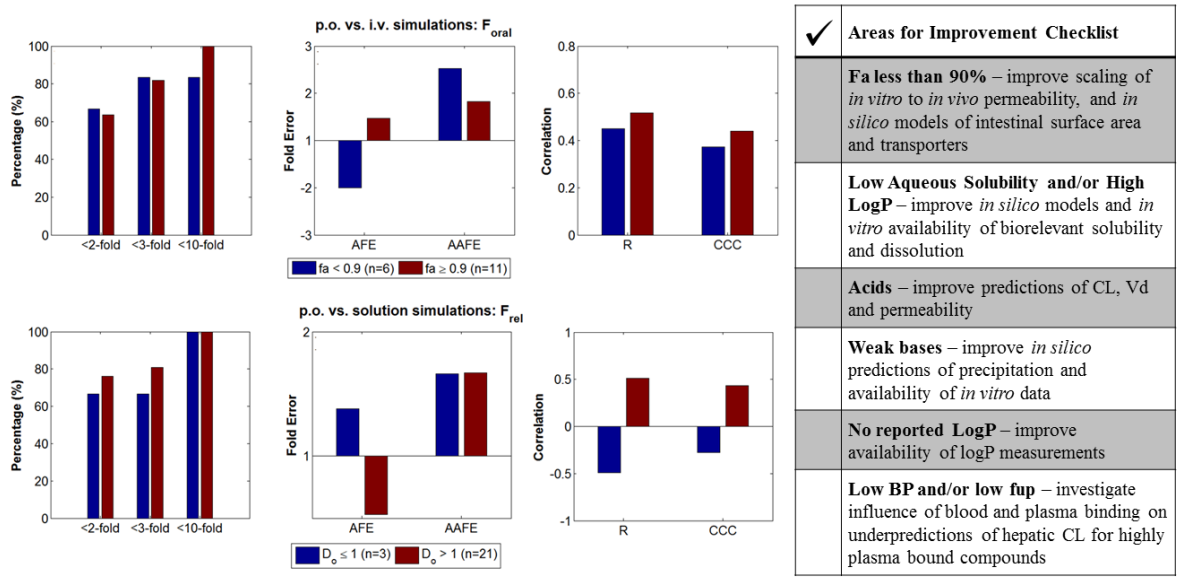
[17]GlaxoSmithKline, United Kingdom

[18]Johannes Gutenberg University of Mainz, Germany

[19]Bristol-Myers Squibb, United Kingdom

[20]Uppsala University, Sweden

Graphical Abstract



✓	Areas for Improvement Checklist
	Fa less than 90% – improve scaling of *in vitro* to *in vivo* permeability, and *in silico* models of intestinal surface area and transporters
	Low Aqueous Solubility and/or High LogP – improve *in silico* models and *in vitro* availability of biorelevant solubility and dissolution
	Acids – improve predictions of CL, Vd and permeability
	Weak bases – improve *in silico* predictions of precipitation and availability of *in vitro* data
	No reported LogP – improve availability of logP measurements
	Low BP and/or low fup – investigate influence of blood and plasma binding on underpredictions of hepatic CL for highly plasma bound compounds

Abstract

Three Physiologically Based Pharmacokinetic software packages (GI-Sim, Simcyp® Simulator, and GastroPlus™) were evaluated as part of the Innovative Medicine Initiative Oral Biopharmaceutics Tools project (OrBiTo) during a blinded “bottom-up” anticipation of

human pharmacokinetics. After data analysis of the predicted vs. measured pharmacokinetics parameters, it was found that oral bioavailability (F_{oral}) was underpredicted for compounds with low permeability, suggesting improper estimates of intestinal surface area, colonic absorption and/or lack of intestinal transporter information. F_{oral} was also underpredicted for acidic compounds, suggesting overestimation of impact of ionisation on permeation, lack of information on intestinal transporters, or underestimation of solubilisation of weak acids due to less than optimal intestinal model pH settings or underestimation of bile micelle contribution. F_{oral} was overpredicted for weak bases, suggesting inadequate models for precipitation or lack of *in vitro* precipitation information to build informed models. Relative bioavailability was underpredicted for both high logP compounds as well as poorly water-soluble compounds, suggesting inadequate models for solubility/dissolution, underperforming bile enhancement models and/or lack of biorelevant solubility measurements. These results indicate areas for improvement in model software, modelling approaches, and generation of applicable input data.

However, caution is required when interpreting the impact of drug-specific properties in this exercise, as the availability of input parameters was heterogeneous and highly variable, and the modellers generally used the data "as is" in this blinded bottom-up prediction approach.

Keywords:

Physiologically-based pharmacokinetics (PBPK); modelling and simulation (M&S); absorption; oral bioavailability (F_{oral}); biopharmaceutics; drug database

Abbreviations:

AFE = Average Fold Error

AAFE = Absolute Average Fold Error

API = Active pharmaceutical ingredient,

AUC = Area under the curve,

BCS = Biopharmaceutics classification system,

BE = Bioequivalence

BP = Blood-to-plasma ratio,

CCC = Concordance Correlation Coefficient

CL = Clearance,

CL/F = Apparent clearance

C_{max} = Maximum concentration,

D_o = Dose number,

f_a = Fraction absorbed,

FE = Fold Error

FIM = First In Man

F_{oral} = Absolute oral bioavailability,

F_{rel} = Relative bioavailability,

fu_p = fraction unbound in plasma,

GI = Gastro-Intestinal

IMI = Innovative Medicines Initiative,

i.v. = Intravenous,

IR = Immediate release,

LogP = Logarithm of the octanol/water partition coefficient,

$LogD_{pH}$ = Logarithm of the octanol/water partition coefficient at a given pH,

MAT = Mean Absorption Time

MTD = Maximal Tolerated Dose

MTT = Mean Transit Time)

MRT = Mean Residence Time

MW = Molecular weight,

<n-fold = % APIs within n-fold,

PAC = Post Approval Changes

PBPK = Physiologically-based pharmacokinetic,

PK = Pharmacokinetic,

p.o. = per oral,

QbD = Quality by Design

R = Pearson correlation coefficient

SME = Small or Medium Enterprise

t_{max} = Time at maximum concentration,

V_d = Volume of distribution,

V_d/F = Apparent volume of distribution,

1. Introduction

The oral route is the most favourable route for drug administration due to its ease of administration and minimal invasiveness. However, orally administered drug products are exposed to a number of potential barriers between administration and systemic exposure, such as dissolution of solid particles and potential precipitation in the gut lumen, permeation through the gut membrane, and intestinal and hepatic first pass metabolism. These processes can have a big impact on the ability to predict *in vivo* performance of drug products. The ability to anticipate the impact of these processes is of great importance in drug and formulation development.

The Innovative Medicines Initiative (IMI) Oral Biopharmaceutical Tools (OrBiTo) project aims to improve upon knowledge of oral drug absorption through the development of new methodologies and refinement of existing tools available in oral biopharmaceutical development. Through four workpackages (WP1-4), the OrBiTo project aims to improve on the tools for evaluating physico-chemical characterisations of active pharmaceutical ingredients (APIs), the development and characterisation of drug product formulations, better understanding of the human intestinal environment, and the *in silico* models for integrating these different aspects into quantitative and qualitative predictions of oral drug exposure (Lennernas et al., 2014). In our previous work we demonstrated the setup of the OrBiTo database of APIs and an overview of the results of the simulation exercise to evaluate the predictive performance of three established PBPK software packages (Margolskee et al. – Part 1 – Submitted; Margolskee et al. – Part 2 – Submitted).

Here we present an analysis of the prediction success with a focus on the impact of compound specific properties and other factors that may influence the outcome of predictions of oral drug exposure. For example, acid/base nature, as well as lipophilicity, are thought to play important roles in dissolution, absorption, and disposition, and are often used as input parameters of the PBPK model to dynamically calculate solubility in the different segments of the GI tract and account for the influence of prandial state on the drug solubility, dissolution rate and permeation rate due to the concentration of bile salts assuming different diffusion for free and micelle-bound drug (Miller et al., 2011). Weak bases may be subject to precipitation in the high pH of the intestinal environment after dissolving in the low pH of stomach acid. The highly ionised state of acids in the intestinal lumen may increase solubility, but also hinder permeation through the phospholipid membranes of the intestinal wall. Acids and bases have the potential to distribute differently into tissues in the body, depending upon the tissue composition and their affinity for different phospholipids; *in silico* predictions of volume of distribution account for tissue composition and these differences between acids and bases (Rodgers and Rowland 2006). Lipophilicity may also affect dissolution in aqueous media, and highly lipophilic compounds can be subject to enhanced solubilisation by bile salts (Mithani et al., 1996), an area which has the potential for error within the PBPK framework, especially if solubility in biorelevant media are not measured experimentally.

The Biopharmaceutics Classification System (BCS), as a classic categorisation of compounds into high and low solubility and high and low permeability, has been used extensively to qualitatively predict *in vivo* oral drug behaviour. In contrast, PBPK has the potential to quantitatively describe the qualitative dynamics indicated by the BCS classification and this can be tested by comparing the predictive abilities of PBPK for the different classes of compounds. For example, for PBPK to be at least as successful as the BCS benchmark, it

should be able to distinguish differences in bioavailability between high and low permeable compounds, and differences in relative bioavailability between high and low soluble compounds.

2. Methods

Inclusion criteria were employed in order to select APIs from the OrBiTo API database for the simulation exercise. The criteria were primarily based on the minimum set of parameters necessary to simulate a compound using the PBPK absorption model in the different software packages and included the availability of: molecular weight, LogP/D, fraction unbound in human plasma (fup), any clearance source scalable to human, *in vitro* permeability with reference compounds, at least one measure of solubility and available clinical data following per oral administration of the given drug. Of the 83 APIs in the OrBiTo database at the start of the exercise, 43 satisfied the inclusion criteria. For more details on the API selection process and comparison of the simulation set with the entire database, see companion paper (Margolskee et al. – Part 1 – Submitted).

A large scale evaluation of the predictive performance of existing *in silico* methods was undertaken. Three software packages, GastroPlus™ (Simulations Plus Inc., Lancaster, CA), Simcyp® (Certara, Sheffield, UK) and GI-Sim (AstraZeneca, London, UK), were employed to produce bottom-up predictions for all of the 43 APIs in the simulation set. Each participating institution generated predictions for all available clinical study arms for the APIs that they had been allocated. A certain degree of overlap in API allocation was allowed to test for user differences. Limited standard operating procedures were provided for Simcyp and GI-Sim, however, most decisions on parameter data selection and simulation setup were left to the individual modellers at each institution. For more details on the procedure for performing the simulations see companion paper (Margolskee et al. – Part 2 – Submitted).

Predictive performance of the PBPK software packages was evaluated through comparison of typical pharmacokinetic (PK) parameters between simulated and observed values. These PK parameters were calculated as described in Part 2 (Margolskee et al. – Part 2 – Submitted). The presented PK parameters included: $AUC_{0\text{-}t,last}$ (area under the curve from time zero to last measured time point), $AUC_{0\text{-}Inf}$ (AUC from time zero extrapolated to time infinity), C_{max} (maximum concentration in plasma), t_{max} (time of peak concentration), $t_{1/2}$ (terminal half-life), CL (clearance), CL/F (oral apparent CL), V_d (volume of distribution), V_d/F (oral apparent V_d), MTT (mean transit time), MRT (mean residence time), F_{oral} (bioavailability), MAT (mean absorption time), F_{rel} (relative bioavailability), relative C_{max}, and relative AUC.

Summary statistics for describing the overall performance of the simulations were decided upon through consensus between the involved institutions and calculated as specified in our companion paper (Margolskee et al. – Part 2 – Submitted). Statistical metrics included: % within two, three and ten-fold of observed, Average Fold Error (AFE), Absolute AFE (AAFE), Pearson regression coefficient (R) and Concordance Correlation Coefficient (CCC) (Lin, 1989; Poulin et al., 2011). The analysis presented in this manuscript focused on single dose and fasted state study arms only (excluding: multiple dose and fed state simulations).

2.1. Grouping based on drug-specific properties

APIs were separated based on drug-specific properties of interest to evaluate the potential impact on the performance of the models. Properties investigated include molecular weight (MW), acid/base nature, lipophilicity (logP and/or logD), BP, fu_p, BCS class, dose number (D_o) and estimated f_a from scaled human effective permeability (P_{eff}).

Groupings for acid/base nature included acid, ampholyte, neutral, weak base, and strong base categories, where strong bases had at least one pKa greater than 7. For each of the properties MW, logP, logD, BP and fu_p, the APIs were separated into four quartiles. Quartiles for each

of these properties are displayed in Table 1. The logD values used in the groupings were the reported logD values taken at the pH closest to 7.4 for each API. In the case for thirty five of the APIs this was pH 7.4, while for the remaining eight APIs the closest pH ranged from 6.5 to 8 (Margolskee et al. – Part 1 – Submitted).

Grouping based on D_o and f_a followed the BCS cut-offs of $D_o \leq 1$ and $D_o > 1$, and $f_a < 0.9$ and $f_a \geq 0.9$. Grouping based on BCS class was carried out according to the reported BCS class of the compounds in the database, or if this was not available, an estimated BCS classification was assigned from f_a based on scaled human Peff, and dose number based on maximum reported dose and aqueous solubility. For further details of these calculations, see companion paper (Margolskee et al. – Part 1 – Submitted).

Table 1: Quartiles for each of the properties of MW, logP, logD, BP, and fu_p for the simulated APIs

	Q1	**Q2**	**Q3**	**Q4**
MW	(150,365 g/mol]	(365,440 g/mol]	(440,505 g/mol]	(505,870 g/mol]
logP	(-0.72 , 2.545]	(2.545 , 3.3]	(3.3 , 4.49]	(4.49 , 7.75]
logD	(-1.45,1.29]	(1.29,2.55]	(2.55,3.17]	(3.17,5.8]
BP	(0.517,0.595]	(0.595,0.640]	(0.640,0.925]	(0.925,3.300]
fu_p	(0.0002,0.0125]	(0.0125,0.05]	(0.05,0.0855]	(0.0855,0.74]

Geometric mean FEs were calculated for each API, averaging over API specific study arms (to account for APIs with different numbers of simulated study arms). APIs were categorised according to properties of interest, and summary statistics of the PK parameters were calculated for each group.

In order to test for interdependencies in the API parameters that may impact the interpretation of the results, Pearson correlation coefficients were calculated for each pairwise combination

of relevant quantitative parameters. The parameters analysed were MW, logP, logD, highest basic pKa, lowest acidic pKa, logit(fu_p), log(BP-0.5), logit(f_a), and log(D_o), where logit(x) = log(x/(1-x)). The transformations for fu_p, BP, f_a and D_o were chosen so the transformed variables would be approximately normally distributed, allowing for more meaningful correlation estimates. Correlations greater than 0.7 in magnitude were considered strong, while correlations between 0.5 and 0.7 were considered moderate.

3. Results and Discussion

Inspection of correlations between API properties revealed a strong negative correlation between logP and logit(fu_p) and a strong positive correlation between lowest acidic pKa and log(BP – 0.5) (Figure 1). Moderate positive correlations included logP vs logD, logD vs lowest acidic pKa, highest basic pKa vs log(BP – 0.5), and logit(fu_p) vs log(BP – 0.5). Moderate negative correlations included highest basic pKa vs log(D_o) and logP vs log(BP – 0.5). Several of these correlations are not surprising, such as logP and logD which both relate to lipophilicity, and fu_p and BP which relate to plasma and blood binding properties. The high correlation between logP and plasma protein binding has also been well documented (Yamazaki and Kanaoka, 2004). These correlations should be taken into consideration when interpreting the results of this exercise.

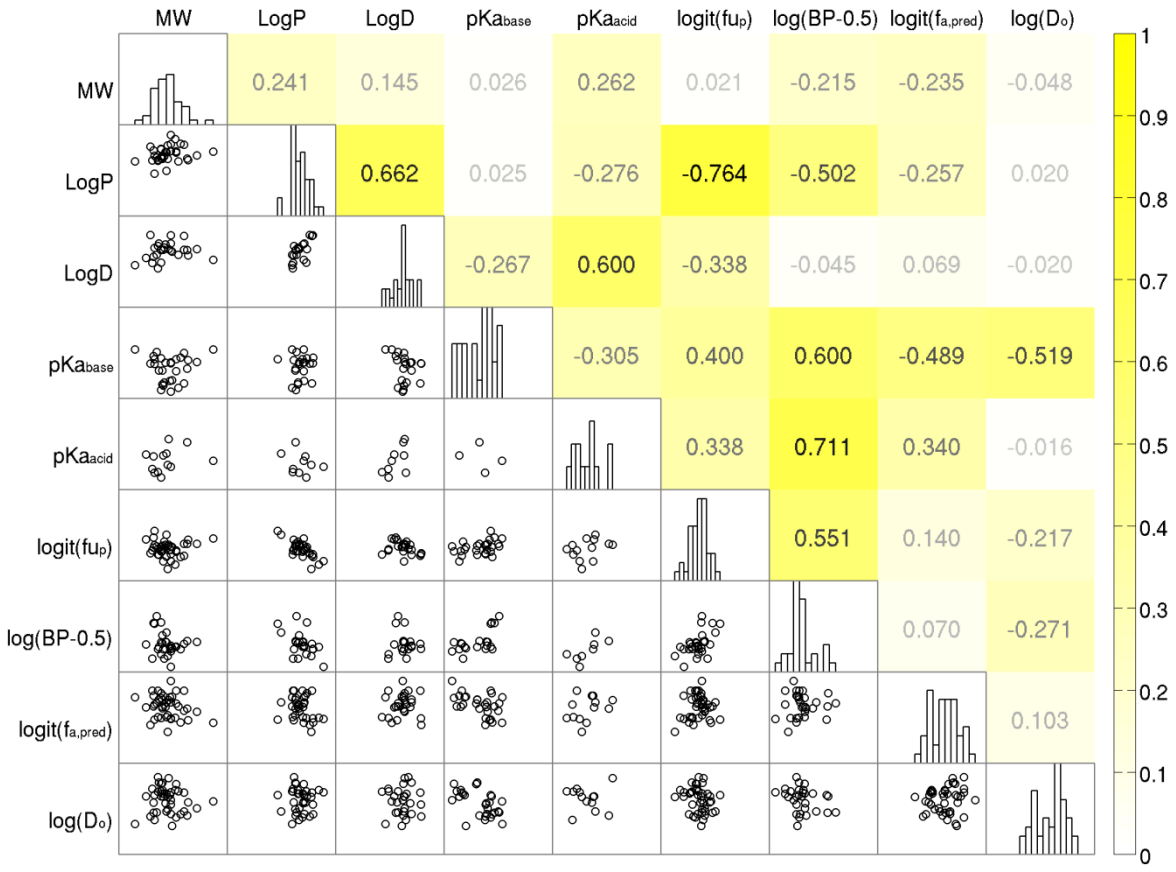


Figure 1: Pairwise comparison of API properties of interest detailing the Pearson correlation coefficients (above diagonal), pairwise scatterplots of property values (below diagonal), and individual histograms of property values (diagonal); MW = molecular weight, pKa_{base} = highest basic pKa, pka_{acid} = lowest acidic pKa, fu_p = fraction unbound in plasma, BP = blood-to-plasma ratio, $f_{a,pred}$ = fraction absorbed predicted from *in vitro* Caco-2 experiments, D_o = dose number, logit(x) = log(x/(1-x)).

Sections 3.1- 3.3 detail the influence of compound properties of interest on the predictive performance of different PK parameters including $AUC_{0\text{-}t,last}$, F_{oral}, F_{rel}, C_{max}, CL or CL/F and V_d or V_d/F. A summary of the findings is included in Table 2. Section 3.4 includes discussion around the interpretation of the results, and next steps. Analysis applied to the predictive performance of additional PK parameters can be found in the Supplementary Material.

Table 2. Summary of impact of drug-specific properties on predictive performance

Physicochemical properties	
Acid-base nature	**Weak bases:** Were poorly predicted where F_{oral} was slightly overpredicted and F_{rel} was underpredicted. **Acids:** Displayed negative bias for predicting F_{oral}.
LogP	**Low LogP:** APIs with low LogP gave better predictions of F_{oral} compared to compounds with higher LogP. **Midrange LogP:** Gave better predictions of F_{rel} compared to high and low LogP. **Calculated LogP:** Gave poor predictions compared to measured LogP.
LogD	**Low LogD:** Displayed better predictive performance of F_{rel} compared to high LogD.
Plasma and blood binding	
Fu_p	**Low fu_p:** Tended to underpredict of CL and CL/F.
BP	**High BP:** Tended to give better predictive performance of CL and CL/F.
BCS Classification	
BCS Class	**BCS III and IV:** F_{oral} predictions displayed negative bias as compared with BCS I and II (highly permeable compounds). **BCS II and IV:** F_{rel} predictions displayed negative bias as compared with overprediction for BCS I and III (freely soluble compounds).
f_a	**$f_a \geq 0.9$:** Displayed a lower bias, higher precision and better correlation metrics for predictions of F_{oral} as compared to the $f_a < 0.9$ group.
D_o	**$D_o > 1$:** Displayed an overall negative bias for predictions of F_{rel} and the $D_o \leq 1$ group displayed an overall positive bias.
F_{oral} = Oral bioavailability, F_{rel} = Relative bioavailability between oral formulations, fu_p = fraction unbound in plasma, BP = Blood-to-plasma ratio, BCS = Biopharmaceutics Classification System, f_a = fraction absorbed as predicted from *in vitro* experiments, D_o = dose number.	

3.1. Physicochemical properties

The correlations between physico-chemical properties and the success in predictions were investigated for different parameters including acid/base nature and lipophilicity, while MW was also investigated, no trends were apparent.

3.1.1. Acid-base Nature

Predictions of oral $AUC_{0\text{-}t,last}$ for neutral compounds and strong bases generally performed well, displaying 80.0% and 92.3% of predictions within three-fold of observed data. Further, low variability was observed for FEs for neutral APIs and strong bases, with calculated AAFEs of 1.92 and 1.63, respectively. Reasonable correlations between predicted and observed were noted, where neutrals and strong bases displayed R coefficients of 0.89 and 0.93, and CCC of 0.34 and 0.93 (Figure 2A-C).

Poor predictions of $AUC_{0\text{-}t,last}$ were observed for acids with 10.0% and 20.0% within two and three-fold for p.o. simulations and a high variability with AAFE of 6.33 (Figure 2A-C). This poor predictive performance may be due to issues in predicting disposition, as both F_{oral} and F_{rel} between p.o. formulations and solutions were generally well predicted (80% and 100% within two-fold) (Figure 2D,G). The $AUC_{0\text{-}t,last}$ predictions of p.o. simulations for acidic compounds displayed a low bias with AFE of 1.16 (Figure 2B), indicating the presence of both over- and underpredictions, while i.v. formulations displayed an overall overestimation as compared to observed data. This overprediction in i.v. exposure seems to be a combined underprediction of both CL (AFE of 0.346) and V_d (AFE of 0.564) for acids (Table A5 and Table A6).

A slight negative bias of F_{oral} was observed for predictions of acids with an AFE of 0.580 (1.72-fold underprediction) and poor correlation (R of -0.570 and CCC of -0.480). This could be related to a potential underestimation of permeability of acidic compounds, an overestimation of the impact of ionisation on permeation, or a lack of information on intestinal transporters. It could also be related to an underestimation of solubility of weak acids due to less than optimal intestinal model pH settings or an underestimation of contribution of bile micelles. However, it could also be due to a small number of poor predictions as the percent within two-fold was considered high at 80.0%.

Poor predictions of F_{oral} were seen for weak bases (25.0% within two-fold), showing a tendency towards overprediction (AFE 1.48), somewhat high source of variability (AAFE 2.46) and poor correlation between predicted and observed, with an R of -0.56 and CCC of -0.296 (Figure 2D-F). A slight overprediction of F_{oral} was observed, whereas F_{rel} between p.o. formulations and solution was underpredicted (Figure 2H). This trend in predictions for weak bases may be due to insufficient data to inform precipitation of the formulation, as few APIs had information regarding precipitation *in vitro* (Margolskee et al. – Part 1 - Submitted).

C_{max} was best predicted for neutral compounds, strong and weak bases with 60.0%, 53.8% and 45.5% predicted with two-fold of observed data, respectively. Neutral compounds and weak bases further displayed the strongest correlation metrics (R and CCC of 0.783 and 0.644, for weak bases; R and CCC of 0.922 and 0.749 for neutral APIs; Table A2).

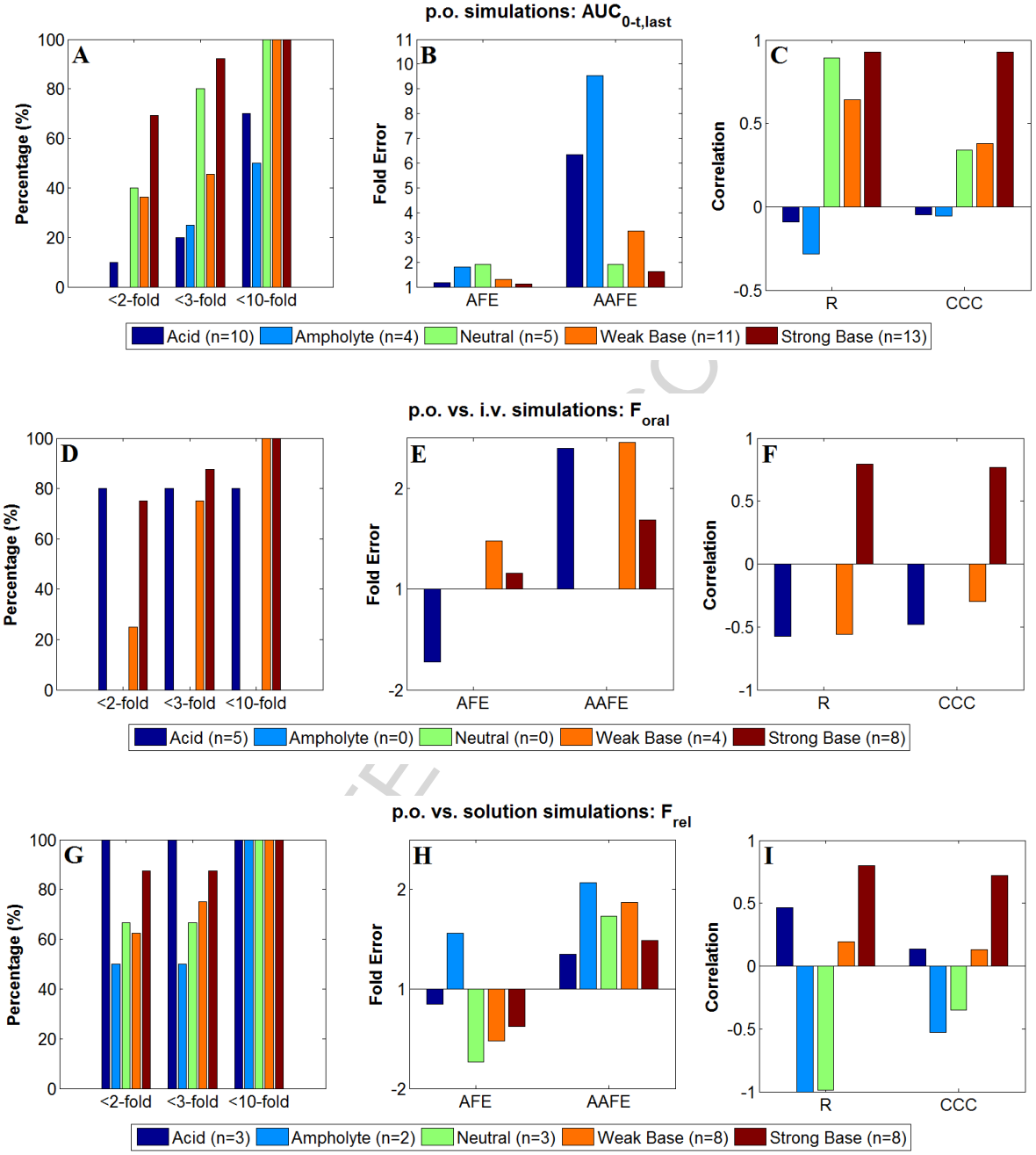


Figure 2. Prediction metrics for $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C), Foral for p.o. vs. i.v. simulations (D-F), and Frel for p.o. vs. solution simulations (G-I), grouped by acid-base nature.

3.1.2. Lipophilicity

3.1.2.1. LogP

Investigating the impact of logP on predictions of $AUC_{0\text{-}t,last}$ revealed predictions for APIs in the second quartile range of logP (Q2: 2.545-3.3) to show a tendency towards better performance as compared to either extremes, with 55.6% within two-fold error, AFE of 1.14, AAFE of 2.17, and R of 0.862 and CCC of 0.836 (Figure 3). Both low logP (Q1: -0.72-2.54) and high logP (Q4: 4.49-7.75) APIs performed poorly compared to the Q2, displaying a lack of correlations, with an R of 0.0411 and 0.0489 for Q1 and Q4, respectively, and a CCC of 0.0183 and 0.0481 (Figure 3).

APIs in the upper quartile of logP generally gave underpredictions of F_{oral} (AFE of 0.477 for Q4), while those in the third quartile gave overpredictions (AFE of 2.21). APIs with lower logP gave the least biased predictions (AFE of 0.819 and 0.920 for Q1 and Q2, respectively). Predictions of F_{oral} for APIs with measured logP were fairly consistent in their correlation with observed, with an R of 0.671, 0.464, and 0.588 for Q1, Q3, and Q4. APIs without measured logP gave highly inconsistent predictions when comparing to observed data with an R of -1 (Figure 3D-F).

Relative AUC between p.o. formulations and solution was predicted best for APIs in the third quartile of logP (Q3: 3.3-4.49), with 100% within two-fold (n=4), AFE of 1.06, low variability with AAFE of 1.19, and strong correlation with R of 0.99 and CCC of 0.87 (Figure 3G-I). Relative AUC showed underpredictions on average for both APIs with low logP (Q1) and high logP (Q4) compared to APIs with middle range logP (Q2 and Q3), displaying AFE

of 0.57 and 0.54 (1.75 and 1.85-fold underpredictions) for Q1 and Q4, respectively, compared to 0.74 (1.35-fold underprediction) and 1.06, for Q2 and Q3 (Figure 3H). Underpredictions of F_{rel} for compounds with logP values in Q4 suggest inadequate models for solubility and dissolution of highly lipophilic compounds, possibly underperforming bile enhancement models or lack of solubility data generated in biorelevant media for highly lipophilic compounds. Interestingly, Q4 showed high correlations between predicted and observed, with R of 0.985, but relatively low CCC of 0.510, suggesting the predictions were in the right direction but on the wrong scale (Figure 3I).

The average predictive performance for V_d or V_d/F for i.v. and p.o. simulations was relatively unbiased for low logP compounds with AFE ranging from 0.917 to 1.14 for Q1 through Q3, while a general trend of overprediction was observed for highly lipophilic compounds (AFE of 2.43 for Q4) (Figure 4 and Table A6). This may be related to the *in silico* methods utilised for predicting distribution into tissues, which varied between users and software. It is well known that *in silico* methods for predicting V_{ss}, such as the Poulin and Theil and Rodgers and Rowland model (Poulin and Theil, 2000; Rodgers et al., 2005; Rodgers and Rowland, 2006), overpredict the volume of distribution for highly lipophilic compounds. Alternative models have been developed to account for the overprediction in V_{ss} for highly lipophilic compounds (Berezhkovskiy, 2004; Poulin and Haddad, 2012) , and these were employed to some extent in the modelling of APIs for the OrBiTo simulation exercise at the modellers discretion. Further analysis will explore user differences such as this in an effort to test the performance of alternative methods.

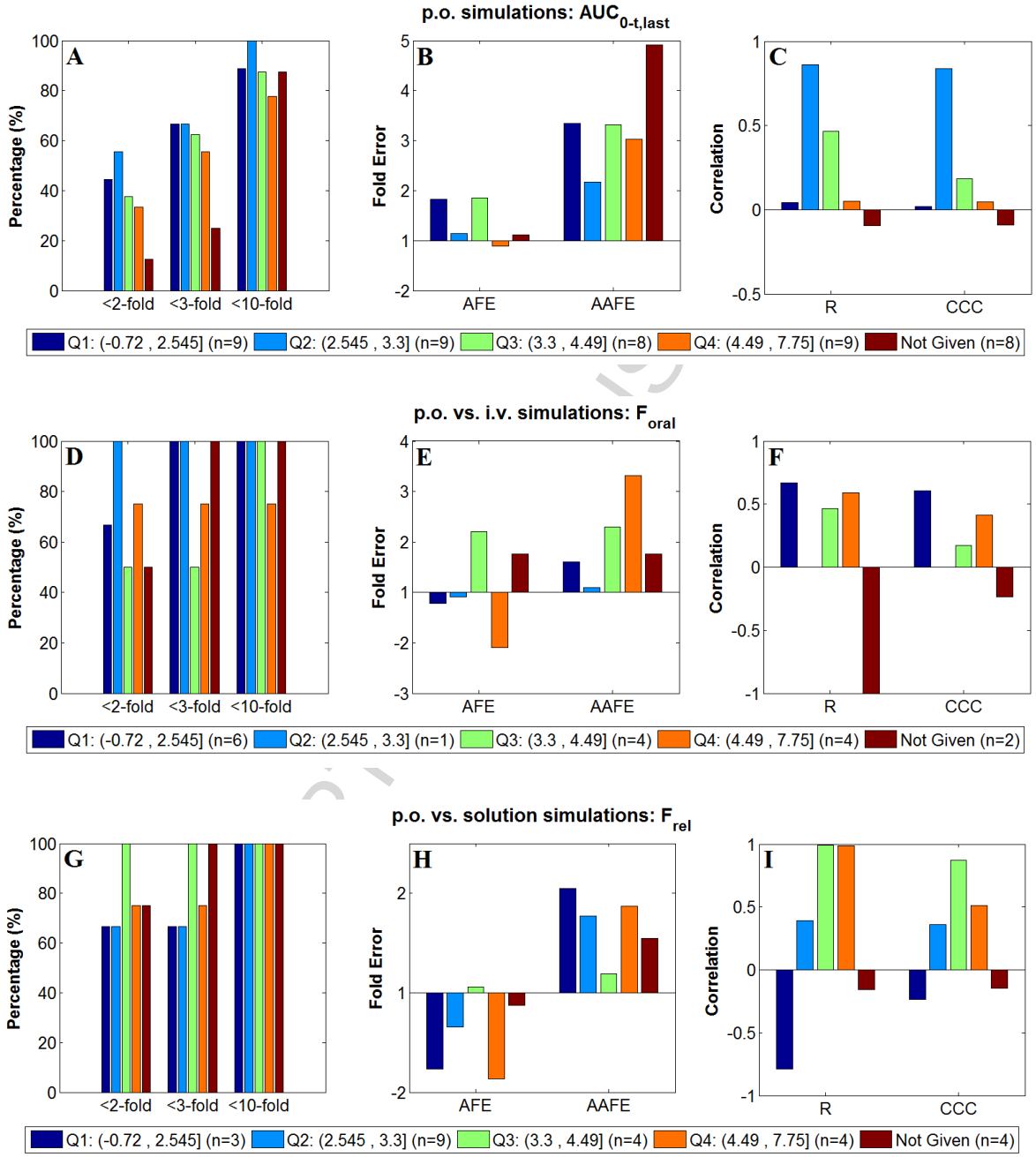


Figure 3: Prediction performance measures of $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C), F_{oral} for p.o. vs. i.v. simulations (D-F), and Frel for p.o. vs. solution simulations (G-I), grouped by quartiles of logP.

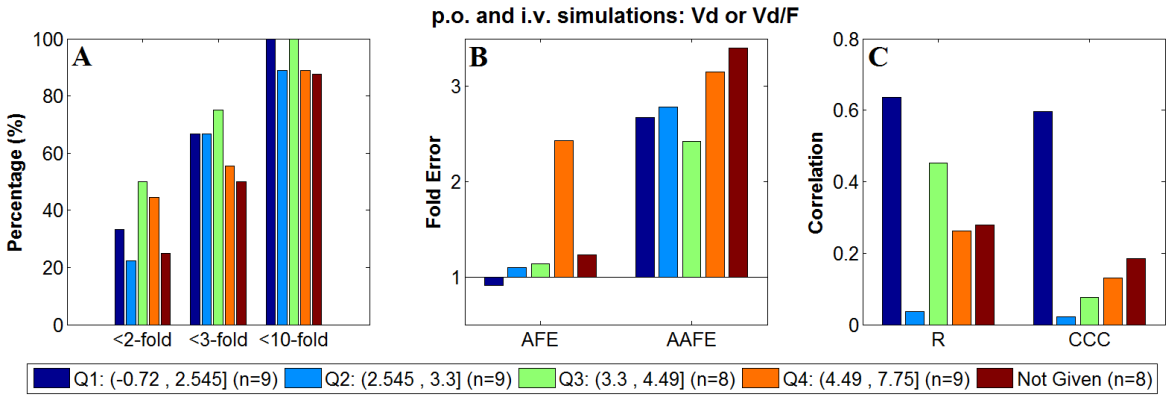


Figure 4. Prediction metrics of volume of distribution (V_d) for i.v. formulations or oral apparent V_d (V_d/F) for p.o. formulations, grouped by logP quartiles.

3.1.2.2. LogD

Prediction performance of $AUC_{0\text{-}t,last}$ for p.o. simulations for APIs in the first quartile of logD (Q1: -1.45-1.29, n=7) was poor, with 14.3%, 28.6% and 57.1% within two, three and ten-fold error, vs. 33.4%, 52.3%, and 90.7%, for the APIs in Q2-Q4 combined (Figure 5). $AUC_{0\text{-}t,last}$ was overpredicted on average for APIs in Q1, with AFE of 2.19 vs. 1.15 for Q2-Q4 combined. Further, Q1 showed high variability with AAFE of 7.60 vs. 3.29 for Q2-Q4, and low correlation between predicted and observed with R of -0.239 and CCC of -0.083 (Figure 5).

Predictions for $AUC_{0\text{-}t,last}$ F_{oral} were poor for APIs in Q1 of logD, displaying AFE of 0.418 (or 2.38-fold underprediction; n=3; vs. AFE of 1.14 for Q2-Q4; n=9), high variability with AAFE of 4.17 (vs. 1.87 for Q2-Q4), and poor correlation between predicted and observed with R of -0.855 and CCC of -0.706. However, more APIs in Q1 had predictions of F_{oral} within two-fold compared with Q2-Q4 (66.7% vs. 55.6%), and less within three and ten-fold,

66.7% and 66.7%, vs. 88.9% and 100%, for Q2-Q4, indicating the results may be skewed by a single outlier (Figure 5D-F).

Predictions for F_{rel} between p.o. and solutions seemed to perform better for lower quartiles of logD, with 100% within two-fold for Q1, 75.0% for each of Q2 and Q3, and 50.0% for Q4. On average, the bias for the different quartiles appeared comparable, with AFE of 0.868, 0.829, 1.09, and 0.606 for Q1, Q2, Q3, and Q4, respectively. The variability in the predictions appeared to increase with increasing logD, with AAFE of 1.35, 1.70, 1.66, and 1.88 for Q1, Q2, Q3 and Q4. Additionally, correlation between predicted and observed F_{rel} declined with increasing logD (R of 0.466, 0.136, 0.0683 and -0.392 and CCC of 0.136, 0.125, 0.0662, and -0.179 for Q1, Q2, Q3, and Q4; Figure 5G-I).

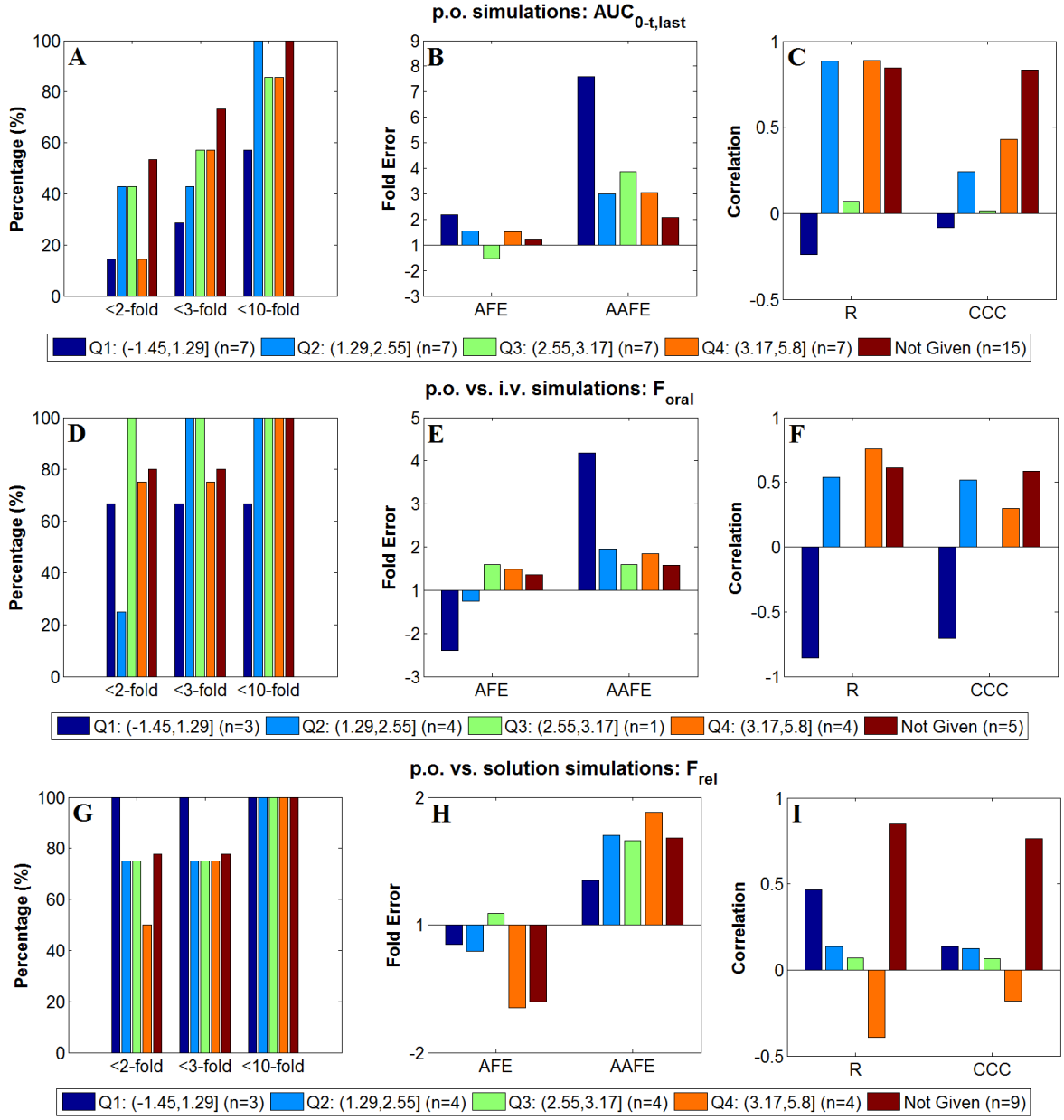


Figure 5: Prediction performance of $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C), Foral for p.o. vs. i.v. simulations (D-F), and Frel for p.o. vs. solution simulations (G-I), grouped by logD quartiles.

Predictions of p.o. $AUC_{0\text{-}t,last}$ performed poorly for APIs for which logP was not given (n=8) compared to those for which logP was reported (n=35). While the overall average performance of the APIs without logP was very good at predicting $AUC_{0\text{-}t,last}$ (AFE of 1.11),

the variability was high, with an AAFE of 4.91 (vs. 2.90 for APIs providing logP) and 12.5% fell within 2 fold (vs. 42.9%), and additionally, there was poor correlation between predicted and observed $AUC_{0\text{-}t,last}$ with R of -0.0936 and CCC of -0.0913 (Figure 3). In contrast, the group of APIs for which logD was not reported performed very well in predicting $AUC_{0\text{-}t,last}$, with 53.3% and 73.3% within two and three-fold error, vs. 28.7% and 46.4%, for APIs where logD was provided (Q1-Q4 combined). APIs not reporting logD also had relatively unbiased estimates and reasonable variability compared with APIs reporting logD. Correlation between predicted and observed was also very high with R of 0.846 and CCC of 0.834 (Figure 5).

3.2. Plasma and blood binding values

In this section, the connection between blood and plasma binding properties and prediction performance was investigated for different pharmacokinetic parameters. Summary statistics were investigated for APIs divided into BP quartiles, and divided into fu_p quartiles (Table 1).

Q1 and Q2 for BP showed poor predictions of $AUC_{0\text{-}t,last}$, with a low percentage falling within n-fold, high variability, and poor correlations between predicted and observed (Figure 6). Clearance predictions also showed high variability and poor correlation for lower values of BP (Figure 6D-F, and Table A5). A similar trend was seen for fu_p, where Q1 and Q2 generally underperformed compared with Q3 and Q4 with low correlations between predicted and observed (Figure 7C). There was a moderate correlation between fu_p and BP for our dataset (Figure 1), and a great overlap between APIs of the lowest BP quartiles and the lowest fu_p group ($fu_p < 0.01$). Low BP indicates lack of uptake into red blood cells, which may reflect the high plasma protein binding in these cases. The underestimations of clearance observed for lower fu_p (highly protein bound) compounds could be explained by unknown

mechanisms for the transfer of highly lipophilic drugs into cells from proteins, as a modest positive relationship between logP and percent bound to protein in plasma (or negative relationship between logP and fu_p) has been documented in the literature (Yamazaki and Kanaoka, 2004), similarly a strong correlation was observed between logP and fu_p for our dataset (Figure 1).

De Buck 2007 investigated the prediction of hepatic clearance using two different methods, method 1 using the traditional formula involving fu_p/BP and fu_{inc}, and method 2 where the effects of fu_p/BP and fu_{inc} were assumed to cancel each other out. The second method performed considerably better at predicting *in vivo* hepatic clearance from *in vitro* CL_{int}, potentially through unintentionally compensating for the inherent underprediction in CLint when scaling from HLMs or human hepatocytes (Hallifax et al., 2012). The results of De Buck 2007 and the relationship we have observed here between low BP and high variability, and low fu_p and high variability in the prediction of clearance may be related.

3.2.1. Blood to plasma ratio (BP)

Simulated APIs were divided into four quartiles (Q1 to Q4; Table 1) based on average simulated blood-to-plasma ratios (BPs). The percent within n-fold increased with increasing BP quartile, where 0.00, 33.3, 50.0 and 63.6% fell within two-fold for Q1 through Q4, respectively, 27.3, 44.4, 66.7 and 81.8% within three-fold, and 72.7, 77.8, 100, and 100% within ten-fold. Similarly, the quartiles showed increasing precision with increasing BP with AAFE of 5.91, 4.46, 2.12, and 2.09 for Q1 through Q4. While Q1 showed a relatively low bias with AFE of 0.951, Q2 through Q4 continued the trend of better performance with increasing BP, with AFE of 2.13, 1.57, and 0.98. Q1 and Q2 also showed poor correlation

between predicted and observed with R of 0.0147 and -0.112 and CCC of 0.0136 and -0.101. In contrast, Q3 and Q4 displayed good correlation between predicted and observed data with R of 0.825 and 0.921 and CCC of 0.727 and 0.908 (Figure 6A-C).

When examining the prediction of CL and CL/F for p.o. and i.v. simulations, there was a tendency of better performance for the higher quartiles of BP. Q1 displayed 0.00, 18.2 and 81.8% of predictions within two, three and ten-fold with similar results for Q2. Q3 and Q4 showed higher frequency within n-fold, with 50.0, 58.3 and 100% for Q3, and 63.6, 81.8, and 100% for Q4 falling within two, three and ten-fold. Q2 displayed the largest bias (AFE: 0.298, or a 3.36-fold underprediction) and poorest precision (AAFE: 7.00), whereas Q1 displayed the lowest AFE at 1.02 and Q4 the lowest AAFE at 2.05. Correlations between predicted and observed CL or CL/F progressively improved with increasing BP quartiles (R: 0.197, 0.189, 0.537, 0.793 for Q1 through Q4 and CCC: 0.0464, 0.123, 0.243, 0.790 for Q1 through Q4; Figure 6D-F, and Table A5).

Investigating the impact of BP on the performance of predicting C_{max} revealed an overall trend of improvement towards compounds with higher BP, for instance CCC ranged from 0.224 for Q1 to 0.636 for Q4. Similarly, improvements were seen in the bias (AFE Q1-Q4: 0.247-0.75) and precision (AAFE Q1-Q4: 7.36-2.73) with increasing BP (Table A2).

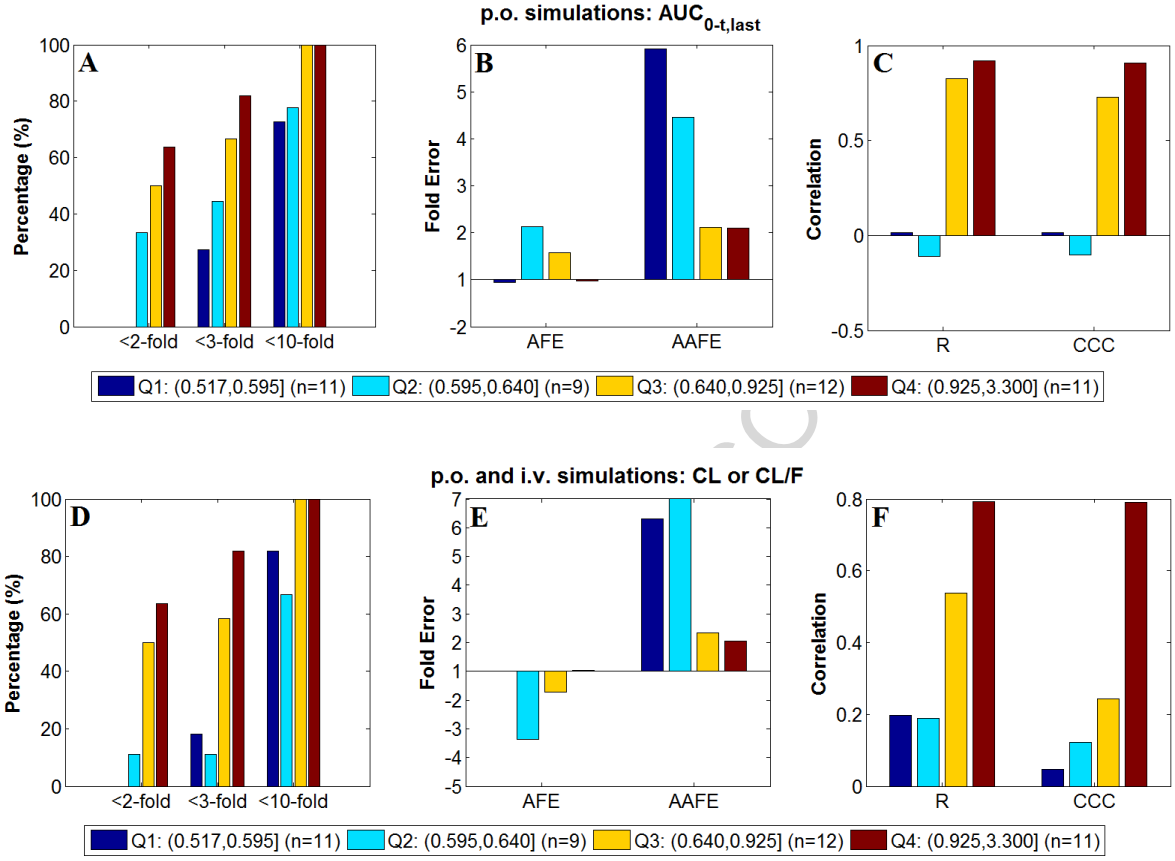


Figure 6: Prediction metrics for $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C) and CL or CL/F for i.v. and p.o. simulations (D-F) as compared to blood-to-plasma ratio value as divided into the 1st to 4th quartile (Q1-4).

3.2.2. Fraction unbound in plasma (fu_p)

Examining the predictive performance of simulated p.o. $AUC_{0\text{-}t,last}$ in relation to simulated average fu_p neglected to show any apparent trends. However, there was an increase in performance in terms of R and CCC with increasing percentiles, with R values of -0.0658, 0.00985, 0.527, and 0.866 for Q1 through Q4, and CCC of -0.0614, 0.00851, 0.400, and 0.810 for Q1 through Q4, respectively (Figure 7).

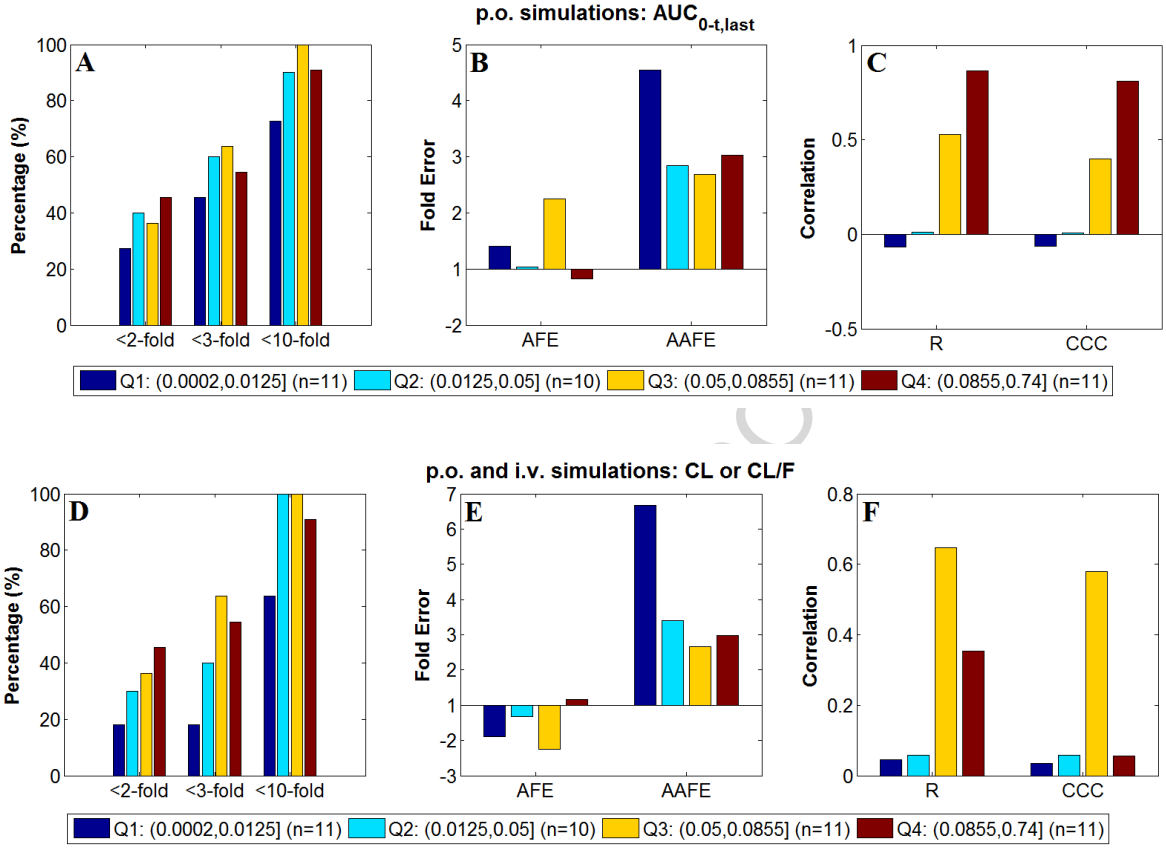


Figure 7: Prediction metrics for $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C) and CL or CL/F for i.v. and p.o. simulations (D-F) as compared to as compared to fu_p value as divided into the 1st to 4th quartile (Q1-4).

3.3. BCS Classification

For BCS class I compounds, 37.5% of APIs displayed a predicted average oral $AUC_{0\text{-}t,last}$ within two-fold, with 87.5% of predictions falling within ten-fold, BCS class II compound displayed a similar level of percentage within two, three and ten-fold. BCS IV APIs displayed a lower degree of predictability with 33.33% falling within two-fold and 77.8% within ten-fold; whereas analysis of BCS class III was limited to two APIs and could therefore be considered undetermined (Figure 5A). BCS class I, II and IV displayed a low AFE, 0.796, 1.55, and 1.47, respectively, and AAFE increased with increasing BCS class, 2.81, 3.30 and

4.00 (Figure 5B). With respect to correlation parameters, BCS class I compounds outperformed the remaining classes, displaying an R value of 0.850 and a CCC value of 0.837 (Figure 8).

F_{oral} predictions for BCS classes III and IV (low permeable compounds) displayed negative bias as compared with BCS I and II (high permeable compounds), displaying AFE of 0.357 for BCS III and IV combined (n=5) and 1.54 for BCS I and II combined (n=12; Figure 8E). Similarly, F_{oral} predictions for estimated $f_a < 0.9$ displayed negative bias with AFE of 0.499 (n=6) vs. 1.47 for APIs with estimated $f_a \geq 0.9$ (Figure 9E). This underprediction of F_{oral} for low permeable compounds could potentially indicate an oversensitivity of the models to *in vitro* permeability measurements, improper intestinal surface area estimates, underestimates of colonic absorption and/or lack of intestinal transporter information. Simulations for highly permeable compounds ($f_a \geq 0.9$) in general performed better than those for low permeable compounds in terms of predictions of p.o. $AUC_{0\text{-}t,last}$, C_{max} and F_{oral}. This is possibly not surprising, as for highly permeable compounds f_a will saturate at a value close to1.0, thus providing limited information on prediction performance of the model. Looking at lower permeable compounds gives a better indication of the (lack of) precision and accuracy of scaling and modelling of permeability.

F_{rel} predictions for BCS classes II and IV (low soluble compounds) displayed negative bias, with AFE of 0.677 (a 1.48-fold underprediction) as compared with 1.20 overprediction for BCS I and III (high soluble compounds) (Figure 5H). Similarly, compounds with $D_o > 1$ (n=21) displayed AFE of 0.681 vs. 1.38 for compounds with $D_o \leq 1$ (Figure 10E). While the number of F_{rel} predictions for compounds in the higher soluble category was small (4 based

on reported BCS and 3 based on calculated D_o), the underpredictions of F_{rel} for low soluble compounds suggest over sensitivity of the dissolution models to aqueous solubility. However, this could also be due to a lack of available measurements of solubility in biorelevant media, as 72.1% of the simulation set were missing this data (Margolskee et al. – Part 1 - Submitted).

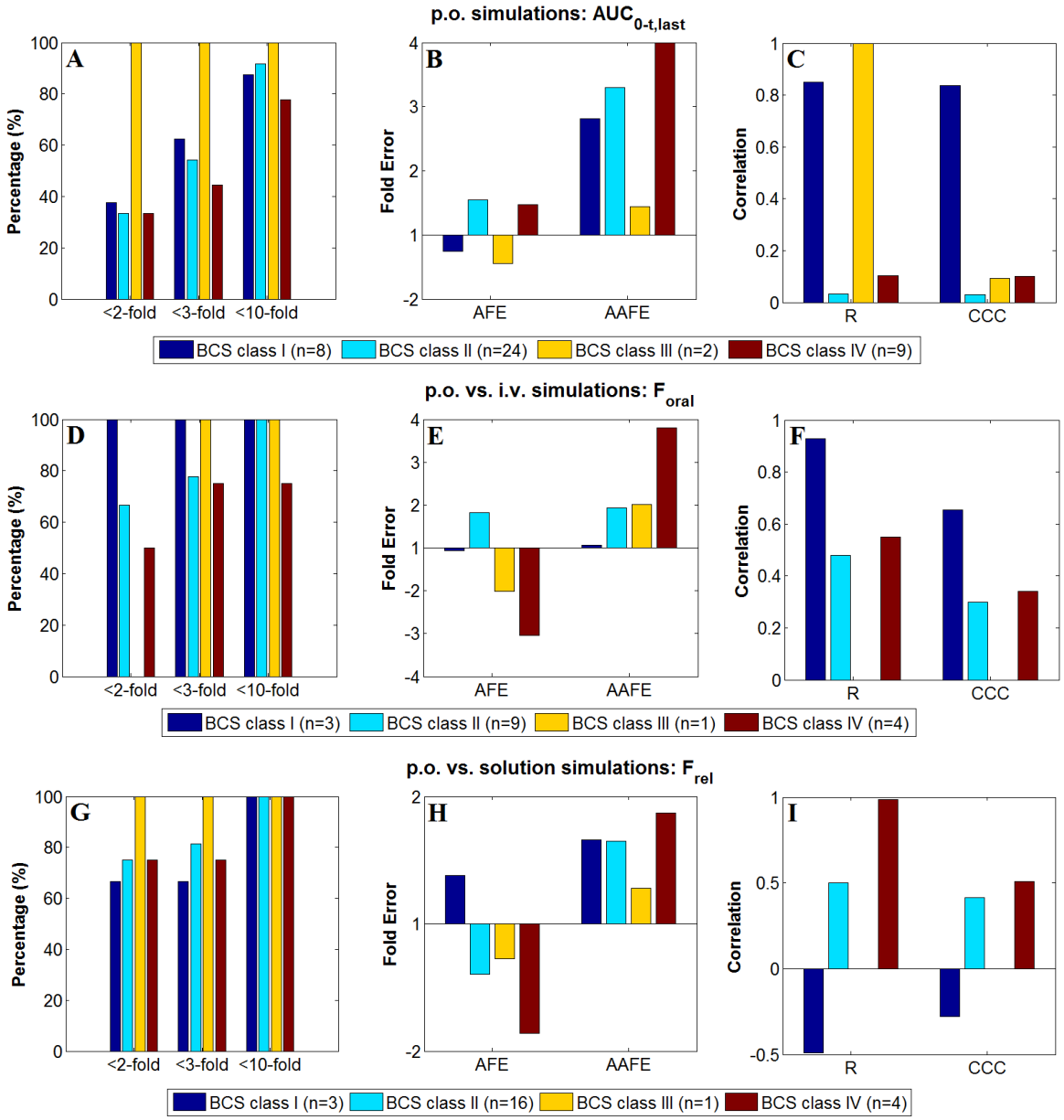


Figure 8: Statistical metrics of prediction success of oral $AUC_{0\text{-}t,last}$ for p.o. formulations (A-C), F_{oral} for p.o. vs. i.v. simulations (D-F), and F_{rel} for p.o. vs. solution simulations (G-I), as compared to the biopharmaceutics classification system (BCS) classes I-IV.

3.3.1. Estimated fraction absorbed from in vitro permeability assay

Permeability-limited APIs ($f_a < 0.9$) displayed an improved prediction of p.o. $AUC_{0\text{-}t,last}$ as compared to highly permeable APIs ($f_a \geq 0.9$) with regards to within two-fold with calculated values of 46.7 and 32.1%, respectively; whereas highly permeable compounds displayed a higher frequency within ten-fold as compared to permeability limited APIs with 96.4 and 73.3% (Figure 9A). Oral predictions of AUC for highly permeable APIs displayed slightly higher accuracy and precision compared to permeability-limited APIs with AFEs of 1.22 and 1.47, AAFEs of 3.08 and 3.47 (Figure 9B). Further, highly permeable APIs displayed a better correlation between predicted and observed AUC compared with low permeable compounds (R: 0.199 and -0.0389 and CCC: 0.195 and -0.0110; Figure 9C).

Predictions of F_{oral} in relation to estimated f_a gave comparable percent within two and three-fold whereas APIs with higher estimated $f_a \geq 0.9$ displayed a larger percentage within ten-fold as compared to the $f_a < 0.9$ group (100% vs. 83.3%; Figure 9D). The $f_a \geq 0.9$ group displayed a lower bias and higher precision as compared to the $f_a < 0.9$ group, with AFE of 1.47 and 0.499, respectively and AAFE of 1.83 and 2.53 (Figure 9E). Correlation coefficients R and CCC suggested similar correlations between predicted and observed F_{oral} for the two groups with a slight favour towards $f_a \geq 0.9$ (R: 0.517, CCC: 0.439) as compared to $f_a < 0.9$ (R: 0.451, CCC: 0.373; Figure 9F).

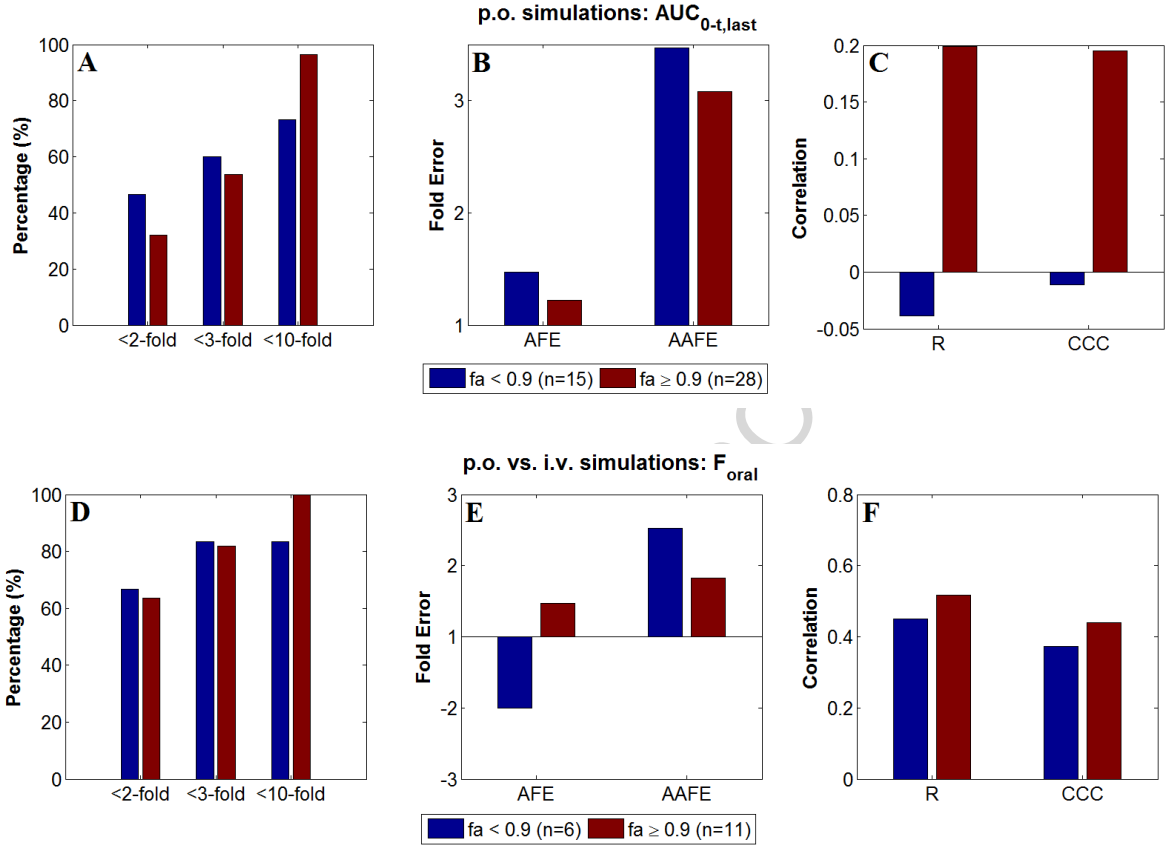


Figure 9: Statistical metrics of prediction success of oral $AUC_{0\text{-}t,last}$ (A-C) and F_{oral} (D-F) as compared to the biopharmaceutics classification system (BCS) cut-off point for permeability (f_a=0.9) divided into permeability-limited ($f_a < 0.9$) and highly permeable ($f_a \geq 0.9$) active pharmaceutical ingredients (APIs).

3.3.2. Dose number (D_o)

Examining the predictive success of oral $AUC_{0\text{-}t,last}$ in relation to dose number grouping revealed a comparable percentage within two, three and ten-fold. Freely soluble APIs displayed 37.5, 62.5 and 87.5% within two, three and ten-fold and solubility-limited compounds displayed 37.1, 54.3 and 88.6% within two, three and ten-fold, respectively. Freely soluble drugs displayed improved AFE and AAFE as compared to solubility-limited compounds, with an AFE of 0.796 (or 1.26-fold underprediction) compared to 1.46 for $D_o \leq 1$

and $D_o > 1$ and AAFE of 2.81 and 3.31. There was an apparent difference in R and CCC for predicted vs. observed oral data for APIs divided based on the D_o cut-off point, with calculated R coefficients of 0.850 and 0.0684 for $D_o \leq 1$ and $D_o > 1$, , and CCC of 0.837 and 0.0640 (Figure 10).

There were only 86 simulations and 3 APIs in the $D_o \leq 1$ group for which F_{rel} was obtainable, whereas there were 594 simulations and 21 APIs in the $D_o > 1$ enabling calculation of F_{rel}. Thus, there were a very limited number of comparators in the $D_o \leq 1$ group. However, the two groups displayed comparable percent within two, three, and ten-fold (66.7%, 66.7% and 100%, respectively for $D_o \leq 1$ and 76.2%, 81% and 100% for $D_o > 1$), and comparable precision with AAFE of 1.66 and 1.67 for $D_o \leq 1$ and $D_o > 1$. One area of noticeable difference was in the overall bias of the F_{rel} predictions, for which the $D_o > 1$ group displayed an overall negative bias with AFE of 0.681 (a 1.47-fold underprediction) and the $D_o \leq 1$ group displayed a positive bias with an AFE of 1.38 (Figure 10).

Freely soluble APIs (compounds with $D_o \leq 1$) displayed slightly poorer performance in C_{max} prediction with respect to percent within N-fold, AFE and AAFE, while displaying a minor improvement in correlation between predicted and observed C_{max}, (R and CCC of 0.627 and 0.533, respectively, vs. 0.442 and 0.438 for solubility limited compounds; Table A2).

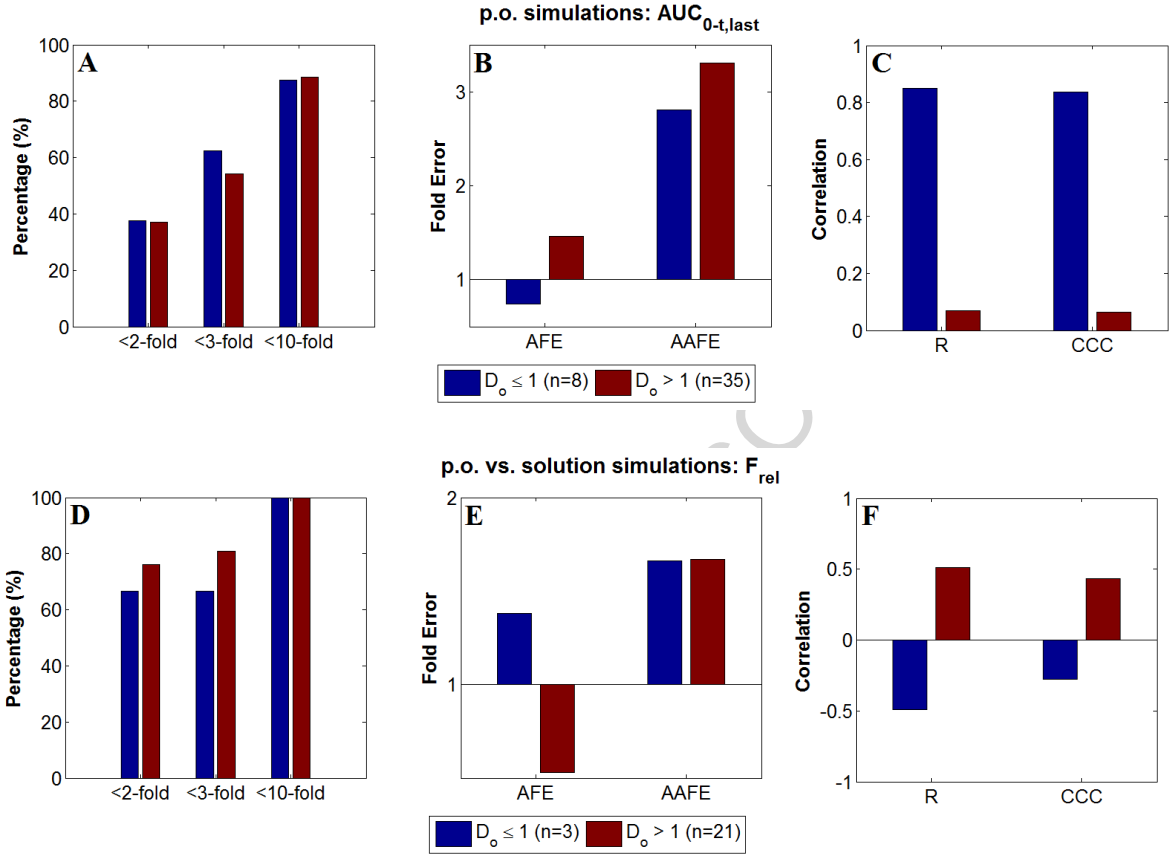


Figure 10: Statistical metrics of prediction success of oral $AUC_{0\text{-}t,last}$ (A-C) and F_{rel} (D-F) as compared to the biopharmaceutics classification system (BCS) dose number (D_o) divided into freely soluble ($D_o \leq 1$) and solubility limited ($D_o > 1$) drug substances.

3.4. Interpretation

The purpose of this large-scale simulation exercise was to evaluate and identify areas for improvement in the current PBPK modelling approach to predicting oral exposure, bioavailability and biopharmaceutics effects. The analysis of the simulation exercise managed to highlight both cases where the PBPK absorption modelling approach performed in line with clinical data and cases within the drug-specific parameter space where simulations deviated from the expected.

There are challenges in interpreting the results of this analysis, as performance is a function of data, model and modeller. Data to inform parameters may be of varied quality and in many cases was lacking, and clinical data may be misrepresentative due to low sample sizes or high variability. Models may fail to appropriately describe gastrointestinal physiology, morphology and the underlying processes governing F_{oral}. The modellers' interpretation and selection of input parameters can also significantly impact performance.

One should be cautious when interpreting the impact of drug-specific properties on the success of PBPK predictions of oral exposure in the current study, due to the heterogeneity and variable nature of the analysed dataset. There was lack of uniformity in reported API parameter data, with data sources and availability of preclinical and clinical data differing widely between APIs. There was also an intention to examine the impact of user differences on the prediction success, thus modellers were relatively unrestricted in their selections of input data and modelling approaches. Some of these decisions included: Selection of clearance sources for extrapolation, methods for estimating volume of distribution, permeability assays used to inform P_{eff}, selection of solubility and/or dissolution sources and formulation properties. Further, simulations were carried out only for APIs which fulfilled minimal criteria for available data with a degree of missingness allowed for certain parameters (BP and logP) which were replaced with estimates.

The high degree of missingness together with a lack of information regarding experimental protocols could be attributed to historical compounds for which key information was not generated, including for example information on the contribution of metabolic pathways, main route of elimination and biorelevant solubility. However, the prevalence of historic

compounds within the OrBiTo database could not be confirmed due to its blinded nature. Certain elements of missing information may also be due to lack of data or the inability to disclose data to outside parties. Other reasons for missing information may be due to unclear standards for the information required and/or desired for prospective PBPK modelling, e.g. lack of pre-clinical data for a number of APIs. The key missing data which may influence model performance will be addressed throughout the OrBiTo project through EFPIA effort in generating the data and updating their dataset.

Limitations of the simulation exercise put into question whether a true evaluation of PBPK absorption model performance was in fact successful. One can argue that without a full dataset of input parameters to inform the model the boundaries cannot fully be tested. The utilisation of a minimum set of input parameters in most cases will result in an advanced model collapsing down to a simpler one, e.g. a lack of particle size distributional parameters for a given formulation will collapse down to modelling a single uniform particle size (measured or assumed).

However, the broad spectrum analysis of this exercise was able to identify several areas for future model improvement and key inputs needed for building a robust model, which may not have been possible with a smaller scale evaluation of a more data-rich compound set (Table 3). Cautious interpretation of the current findings can be used to inform future directions for improvement of *in silico* models and available inputs, which should then be evaluated with more targeted test sets. For example, a set of data-rich compounds of high lipophilicity and low aqueous solubility, but relatively high *in vivo* relative bioavailability could be used to test improvements in *in silico* biorelevant solubility predictions. A set of data-rich BCS III

compounds with high solubility and low fraction absorbed could be used to evaluate improvements of intestinal surface area estimates and the contribution of colonic absorption.

Table 3: Areas for Improvement Checklist

✓	**Areas for Improvement Checklist**
	Fa less than 90% – improve scaling of *in vitro* to *in vivo* permeability, and *in silico* models of intestinal surface area and transporters
	Low Aqueous Solubility and/or High LogP – improve *in silico* models and *in vitro* availability of biorelevant solubility and dissolution
	Acids – improve predictions of CL, Vd and permeability
	Weak bases – improve *in silico* predictions of precipitation and availability of *in vitro* data
	No reported LogP – improve availability of logP measurements
	Low BP and/or low fu_p – investigate influence of blood and plasma binding on underpredictions of hepatic CL for highly plasma bound compounds

Several tasks are identified in OrBiTo WP4 and across the whole consortium to address the points highlighted in this paper. One task addresses the dynamic calculation of bile salt concentration in the GI tract, while another is making changes to the GI tract physiology, and proposing a more biorelevant model for gastric emptying, lumen and mucosal liquid volumes, together with a gallbladder emptying model. Other projects include proposing new models for supersaturation and precipitation, new algorithms for passive and active permeability and finally a new model for lymphatic absorption.

In parallel to these model improvements, the learning of where models perform well and where they need to improve will also guide the future data collection work or data completion

work needed from the EFPIA companies. The database will be complemented with missing information on existing drug records such as rCYP *in vitro* characterisation of Vmax and Km of the main enzymes involved in the drug metabolism when pre-systemic extraction is suspected, measurements of log P, pKa, aqueous solubility or solubility in biorelevant media. This work also highlights the need for more EFPIA examples of drugs that are relevant for intestinal absorption and oral formulation modelling such as poorly absorbed compounds and weak bases with their precipitation characterisation. All of these examples if they exist could be fed into the OrBiTo database in order to enrich the dataset with more relevant examples.

4. Conclusion

The results of this exercise suggest that PBPK modelling of oral bioavailability generally performs well for well-behaved compounds (e.g. neutral or strong base, mid-range logP (2.545 to 3.3), high permeability, high solubility, BP >0.64 and fu_p > 0.05). However, as shown in this study, an increasing level of complexity, e.g. solubility and permeability limitations, and increasing complexity of delivery system, were met with decreased prediction performance. In such scenarios, modelling efforts may rely more heavily on quality of input data, model assumptions, and modeller experience. It would therefore be advisable to take into account the increasing degree of uncertainty on prediction success.

For interpreting the results of this study, one must take into perspective the level of availability, detail and quality of data that was used to generate simulations as well as the limitation in contact that could occur between modellers and API owners which would normally occur in the pharmaceutical industry. As such, this approach therefore can be regarded as an opportunistic blinded modelling exercise driven by availability of parameter data. It is the opinion of the authors' that future similar simulation exercises should strive for a more synergistic approach between data gathering and model building in order to ensure the

exercise to produce relevant results. Further analysis of the simulation output is required to explore the performance of formulation and food effects.

5. Acknowledgements

This work was performed under the OrBiTo Project, which has received support from the Innovative Medicines Joint Undertaking (http://www.imi.europa.eu) under Grant Agreement No. 115369. The authors would also like to acknowledge the contributions of all participants in the OrBiTo simulation exercise, especially those who could not be named in the author list. The extensive list of participants and their affiliations follows below, organised in decreasing order of workload by institution, and then alphabetical order by last name:

Name	Affiliation(s)	Contributions(s)
Leon Aarons	University of Manchester	PI for University of Manchester
Adam S. Darwich	University of Manchester	Performed gap analysis on database, performed simulations, analysed results
Aleksandra Galetin	University of Manchester	PI for University of Manchester
Alison Margolskee	University of Manchester	Performed gap analysis on database, performed simulations, analysed results
Amin Rostami-Hodjegan	University of Manchester/SimCYP	Work package co-leader, PI for University of Manchester, PI for SimCYP
Sara Carlert	AstraZeneca	Performed simulations, analysed results
Maria Hammarberg	AstraZeneca	Performed simulations
Constanze Hilgendorf	AstraZeneca	Performed simulations
Pernilla Johansson	AstraZeneca	Performed simulations
Eva Karlsson	AstraZeneca	Performed simulations
Donal Murphy	AstraZeneca	Performed simulations
Christer Tannergren	AstraZeneca	Performed simulations, PI for AstraZeneca
Helena Thörn	AstraZeneca	Performed simulations
Mohammed Yasin	AstraZeneca	Performed simulations
Florent Mazuir	Sanofi	Performed simulations
Olivier Nicolas	Sanofi	Performed simulations, analysed results
Xavier Pepin	Sanofi/AstraZeneca	Work package co-leader, PI for Sanofi until March 2015, performed gap analysis on database, performed simulations, analysed results
Sergej Ramusovic	Sanofi	Performed simulations
Christine Xu	Sanofi	Performed simulations
Shriram Pathak	SimCYP	Performed gap analysis on database, performed simulations, analysed results

Timo Korjamo	Orion Pharma	Performed simulations, analysed results
Johanna Laru	Orion Pharma	Performed simulations
Jussi Malkki	Orion Pharma	Performed simulations, analysed results
Sari Pappinen	Orion Pharma	Analysed results
Johanna Tuunainen	Orion Pharma	Analysed results
Jennifer Dressman	Goethe University	PI for Goethe University
Carmen Gött	Goethe University	Analysed results
Simone Hansmann	Goethe University	Performed simulations, analysed results
Edmund Kostewicz	Goethe University	PI for Goethe University
Handan He	Novartis	Performed simulations, analysed results
Tycho Heimbach	Novartis	Performed simulations, analysed results
Fan Wu	Novartis	Performed simulations, analysed results
Carolin Hoft	AbbVie	Performed simulations
Yan Pang	AbbVie	Performed simulations
Michael B. Bolger	Simulations Plus	PI for Simulations Plus, lead for analysis of impact of solubility and dissolution
John DiBella	Simulations Plus	Financial and time accounting for Simulations Plus
Eva Huehn	Simulations Plus	Performed gap analysis on database, performed simulations
Viera Lukacova	Simulations Plus	Co-PI for Simulations Plus
James M. Mullin	Simulations Plus	Performed gap analysis on database, performed simulations
Ke X. Szeto	Simulations Plus	Performed gap analysis on database, performed simulations
Joanne Bennett	Pfizer	Collation of data for database
Chester Costales	Pfizer	Performed simulations
Jian Lin	Pfizer	Performed simulations
Mark McAllister	Pfizer	Performed simulations
Sweta Modi	Pfizer	Performed simulations
Charles Rotter	Pfizer	Performed simulations
Manthena Varma	Pfizer	Performed simulations
Mei Wong	Pfizer	Performed simulations
Amitava Mitra	Merck Sharp & Dohme (MSD)	Performed simulations, analysed results
Jan Bevernage	Janssen	Performed simulations
Jeike Biewenga	Janssen	Performed simulations
Achiel Van Peer	Janssen	Performed simulations
Richard Lloyd	GlaxoSmithKline	Performed simulations, analysed results
Carole Shardlow	GlaxoSmithKline	Performed simulations, analysed results
Peter Langguth	University of Mainz	PI for University of Mainz
Irina Mishenzon	University of Mainz	Performed simulations
Mai Anh Nguyen	University of Mainz	Performed simulations
Jonathan Brown	Bristol-Myers Squibb	Performed simulations

References

Berezhkovskiy LM (2004) Volume of distribution at steady state for a linear pharmacokinetic system with peripheral elimination. *J Pharm Sci* 93:1628-1640.

Hallifax D and Houston JB (2006) Binding of drugs to hepatic microsomes: comment and assessment of current prediction methodology with recommendation for improvement. *Drug Metab Dispos* 34:724-726; author reply 727.

Hallifax D, Turlizzi E, Zanelli U, and Houston JB (2012) Clearance-dependent underprediction of in vivo intrinsic clearance from human hepatocytes: comparison with permeabilities from artificial membrane (PAMPA) assay, in silico and caco-2 assay, for 65 drugs. *Eur J Pharm Sci* 45:570-574.

Lennernas H, Aarons L, Augustijns P, Beato S, Bolger M, Box K, Brewster M, Butler J, Dressman J, Holm R, Julia Frank K, Kendall R, Langguth P, Sydor J, Lindahl A, McAllister M, Muenster U, Mullertz A, Ojala K, Pepin X, Reppas C, Rostami-Hodjegan A, Verwei M, Weitschies W, Wilson C, Karlsson C, and Abrahamsson B (2014) Oral biopharmaceutics tools - Time for a new initiative - An introduction to the IMI project OrBiTo. *Eur J Pharm Sci*. 57:292-299.

Lin L (1989) A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45:255-268.

Margolskee A, Darwich AS, Pepin X, Pathak SM, Bolger MB, Aarons L, Rostami-Hodjegan A, Angstenberger J, Graf F, Laplanche L, Müller T, Carlert S, Daga P, Murphy D, Tannergren C, Yasin M, Greschat-Schade S, Mück W, Muenster U, van der Mey D, Frank KJ, Lloyd R, Adriaenssen L, Bevernage J, De Zwart L, Swerts D, Tistaert C, Van Den Bergh A, Van Peer A, Beato S, Nguyen-Trung AT, Bennett J, McAllister M, Wong M, Zane P, Ollier C, Vicat P, Kolhmann M, Marker A, Brun P, Mazuir F, Beilles S, Venczel M, Boulenc X, Loos P, Lennernäs H, Abrahamsson B (Submitted) IMI – Oral Biopharmaceutics Tools project – Evaluation of Bottom-up PBPK Prediction Success Part 1: Characterisation of the OrBiTo Database of Compounds. *Eur J Pharm Sci.*

Margolskee A, Darwich A, Pepin X, Aarons L, Galetin A, Rostami-Hodjegan A, Carlert S, Hammarberg M, Hilgendorf C, Johansson P, Karlsson E, Murphy D, Tannergren C, Thorn H, Yasin M, Mazuir F, Nicolas O, Ramusovic S, Xu C, Pathak SM, Korjamo T, Laur J, Malkki J, Pappinen S, Tuunainen J, Dressman J, Hansmann S, Kostewicz E, He H, Heimbach T, Wu F, Hoft C, Laplanche L, Pang Y, Bolger MB, Huehn E, Lukacova V, Mullin JM, Szeto KX, Costales C, Lin J, McAllister M, Modi S, Rotter C, Varma M, Wong M, Mitra A, Bevernage J, Biewenga J, Van Peer A, Lloyd R, Shardlow C, Langguth P, Mishenzon I, Nguyen MA, Brown J, Lennernas H, and Abrahamsson B (Submitted) IMI – Oral Biopharmaceutics Tools project – Evaluation of Bottom-up PBPK Prediction Success Part 2: An Introduction to the Simulation Exercise and Overview of Results. *Eur J Pharm Sci.*

Miller JM, Beig A, Krieg BJ, Carr RA, Borchardt TB, Amidon GE, Amidon GL, and Dahan A (2011) The Solubility–Permeability Interplay: Mechanistic Modeling and

Predictive Application of the Impact of Micellar Solubilization on Intestinal Permeation. *Molecular Pharmaceutics* 8(5): 1848-1856

Mithani SD, Bakatselou V, TenHoor CN, and Dressman JB (1996) Estimation of the increase in solubility of drugs as a function of bile salt concentration. *Pharm Res* 13:163-167.

Poulin P and Haddad S (2012) Advancing prediction of tissue distribution and volume of distribution of highly lipophilic compounds from a simplified tissue-composition-based model as a mechanistic animal alternative method. *J Pharm Sci* 101:2250-2261.

Poulin P, Jones HM, Jones RD, Yates JW, Gibson CR, Chien JY, Ring BJ, Adkison KK, He H, Vuppugalla R, Marathe P, Fischer V, Dutta S, Sinha VK, Bjornsson T, Lave T, and Ku MS (2011) PhRMA CPCDC initiative on predictive models of human pharmacokinetics, part 1: Goals, properties of the PhRMA dataset, and comparison with literature datasets. *J Pharm Sci*. 100(10):4050-4073.

Poulin P and Theil FP (2000) A priori prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. *J Pharm Sci* 89:16-35.

Rodgers T, Leahy D, and Rowland M (2005) Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci* 94:1259-1276.

Rodgers T and Rowland M (2006) Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci* 95:1238-1257.

Yamazaki K and Kanaoka M (2004) Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *J Pharm Sci* 93:1480-1494.

Appendix

Table A1: Summary statistics for $AUC_{0\text{-}t,last}$ predictions for p.o. simulations, grouped by different compound specific properties.

	no. APIs	% within 2 fold	% within 3 fold	% within 10 fold	AFE	AAFE	R	CCC	R (of log data)	CCC (of log data)
Molecular Weight										
Q1: (150,365 g/mol]	11	27.3	54.5	81.8	1.25	3.91	-0.129	-0.0427	0.486	0.453
Q2: (365,440 g/mol]	10	40.0	50.0	80.0	0.604	3.36	0.0851	0.0784	0.734	0.712
Q3: (440,505 g/mol]	11	36.4	54.5	100	2.32	2.75	0.557	0.439	0.952	0.884
Q4: (505,870 g/mol]	11	45.5	63.6	90.9	1.53	2.94	0.372	0.186	0.744	0.728
Acid/Base Nature										
Acid	10	10.0	20.0	70.0	1.16	6.33	-0.0921	-0.0480	0.713	0.700
Ampholyte	4	0.00	25.0	50.0	1.81	9.54	-0.284	-0.0559	0.723	0.516
Neutral	5	40.0	80.0	100	1.92	1.92	0.891	0.340	0.836	0.541
Weak Base	11	36.4	45.5	100	1.30	3.26	0.641	0.380	0.736	0.679
Strong Base	13	69.2	92.3	100	1.11	1.63	0.929	0.928	0.949	0.947
LogP										
Q1: (-0.72 , 2.545]	9	44.4	66.7	88.9	1.83	3.34	0.0411	0.0183	0.495	0.444
Q2: (2.545 , 3.3]	9	55.6	66.7	100	1.14	2.17	0.862	0.836	0.803	0.784
Q3: (3.3 , 4.49]	8	37.5	62.5	87.5	1.86	3.32	0.465	0.183	0.461	0.423
Q4: (4.49 , 7.75]	9	33.3	55.6	77.8	0.899	3.03	0.0489	0.0481	0.878	0.877
Not Given	8	12.5	25.0	87.5	1.11	4.91	-0.0936	-0.0913	0.483	0.483
LogD										
Q1: (-1.45,1.29]	7	14.3	28.6	57.1	2.19	7.59	-0.239	-0.083	0.457	0.425
Q2: (1.29,2.55]	7	42.9	42.9	100	1.54	3.01	0.885	0.241	0.712	0.600
Q3: (2.55,3.17]	7	42.9	57.1	85.7	0.648	3.86	0.0692	0.016	0.541	0.515
Q4: (3.17,5.8]	7	14.3	57.1	85.7	1.53	3.04	0.887	0.432	0.717	0.695
Not Given	15	53.3	73.3	100	1.22	2.08	0.846	0.834	0.944	0.937
BCS Classification										
BCS class I	8	37.5	62.5	87.5	0.796	2.81	0.850	0.837	0.765	0.740
BCS class II	24	33.3	54.2	91.7	1.55	3.30	0.0327	0.0304	0.780	0.759
BCS class III	2	100	100	100	0.694	1.44	1.00	0.0923	1.00	0.132
BCS class IV	9	33.3	44.4	77.8	1.47	4.00	0.103	0.0995	0.751	0.727
Dose number										
$D_o \leq 1$	8	37.5	62.5	87.5	0.796	2.81	0.85	0.837	0.765	0.74
$D_o > 1$	35	37.1	54.3	88.6	1.46	3.31	0.0684	0.064	0.77	0.759
Estimated fa										
$f_a < 0.9$	15	46.7	60	73.3	1.47	3.47	-0.0389	-0.0110	0.776	0.768
$f_a \geq 0.9$	28	32.1	53.6	96.4	1.22	3.08	0.199	0.195	0.728	0.72
BP										
Q1: (0.517,0.595]	11	0.00	27.3	72.7	0.951	5.91	0.0147	0.0136	0.702	0.701
Q2: (0.595,0.640]	9	33.3	44.4	77.8	2.13	4.46	-0.112	-0.101	0.787	0.761
Q3: (0.640,0.925]	12	50.0	66.7	100	1.57	2.12	0.825	0.727	0.831	0.786
Q4: (0.925,3.300]	11	63.6	81.8	100	0.98	2.09	0.921	0.908	0.807	0.802
fu_p										
Q1: (0.0002,0.0125]	11	27.3	45.5	72.7	1.41	4.54	-0.0658	-0.0614	0.811	0.804
Q2: (0.0125,0.05]	10	40.0	60.0	90.0	1.04	2.84	0.00985	0.00851	0.587	0.577

Q3: (0.05,0.0855]	11	36.4	63.6	100	2.25	2.68	0.527	0.400	0.894	0.808
Q4: (0.0855,0.74]	11	45.5	54.5	90.9	0.854	3.03	0.866	0.810	0.699	0.686

Table A2: Summary statistics for C_{max} predictions for p.o. simulations, grouped by different compound specific properties.

	no. APIs	**% within 2 fold**	**% within 3 fold**	**% within 10 fold**	**AFE**	**AAFE**	**R**	**CCC**	**R (of log data)**	**CCC (of log data)**
Molecular Weight										
Q1: (150,365 g/mol]	11	27.3	45.5	81.8	0.612	4.07	0.178	0.176	0.596	0.558
Q2: (365,440 g/mol]	10	50.0	60.0	80.0	0.416	3.20	0.478	0.388	0.761	0.692
Q3: (440,505 g/mol]	11	54.5	81.8	90.9	0.819	2.36	0.969	0.631	0.952	0.867
Q4: (505,870 g/mol]	11	27.3	63.6	90.9	0.667	3.38	0.651	0.632	0.716	0.696
Acid/Base Nature										
Acid	10	10.0	30.0	60.0	0.329	7.22	0.146	0.144	0.800	0.677
Ampholyte	4	25.0	25.0	75.0	2.00	4.21	0.668	0.228	0.932	0.706
Neutral	5	60.0	100	100	0.668	1.73	0.922	0.749	0.816	0.744
Weak Base	11	45.5	72.7	90.9	0.806	2.76	0.783	0.644	0.741	0.725
Strong Base	13	53.8	76.9	100	0.536	2.24	0.457	0.362	0.886	0.830
LogP										
Q1: (-0.72 , 2.545]	9	11.1	66.7	100	0.923	2.88	0.730	0.516	0.787	0.725
Q2: (2.545 , 3.3]	9	44.4	66.7	88.9	0.679	2.66	0.318	0.274	0.615	0.575
Q3: (3.3 , 4.49]	8	75.0	75.0	75.0	0.925	2.72	0.610	0.530	0.105	0.105
Q4: (4.49 , 7.75]	9	33.3	55.6	66.7	0.234	5.36	0.294	0.032	0.798	0.694
Not Given	8	37.5	50.0	100	0.694	2.88	0.258	0.255	0.743	0.722
LogD										
Q1: (-1.45,1.29]	7	14.3	42.9	71.4	1.12	4.73	-0.0136	-0.0131	0.632	0.61
Q2: (1.29,2.55]	7	42.9	85.7	100	0.903	2.16	0.962	0.557	0.865	0.805
Q3: (2.55,3.17]	7	14.3	28.6	85.7	0.512	4.38	0.668	0.309	0.558	0.518
Q4: (3.17,5.8]	7	71.4	85.7	85.7	0.645	2.48	0.778	0.751	0.683	0.665
Not Given	15	46.7	66.7	86.7	0.416	3.09	0.686	0.64	0.903	0.819
BCS Classification										
BCS class I	8	37.5	62.5	75.0	0.605	4.01	0.627	0.533	0.597	0.529
BCS class II	24	45.8	66.7	91.7	0.759	2.66	0.443	0.443	0.857	0.832
BCS class III	2	0.00	50.0	100	0.222	4.51	1.00	0.00868	1.00	0.0278
BCS class IV	9	33.3	55.6	77.8	0.45	3.9	0.44	0.0493	0.708	0.646
Dose number										
$D_o \leq 1$	8	37.5	62.5	75	0.605	4.01	0.627	0.533	0.597	0.529
$D_o > 1$	35	40	62.9	88.6	0.619	3.03	0.442	0.438	0.818	0.791
Estimated fa										
$f_a < 0.9$	15	20	46.7	73.3	0.337	5.05	0.0876	0.0781	0.759	0.688
$f_a \geq 0.9$	28	50	71.4	92.9	0.851	2.5	0.512	0.51	0.805	0.792
BP										
Q1: (0.517,0.595]	11	27.3	45.5	45.5	0.247	7.36	0.242	0.224	0.698	0.604
Q2: (0.595,0.640]	9	44.4	66.7	100	0.889	2.58	0.425	0.421	0.946	0.927
Q3: (0.640,0.925]	12	50.0	75.0	100	0.903	2.01	0.678	0.340	0.821	0.812
Q4: (0.925,3.300]	11	36.4	63.6	100	0.75	2.73	0.760	0.636	0.746	0.722
fu_p										
Q1: (0.0002,0.0125]	11	36.4	54.5	81.8	0.413	3.99	0.458	0.436	0.883	0.819
Q2: (0.0125,0.05]	10	60.0	80.0	80.0	0.661	2.88	0.315	0.311	0.629	0.610

Q3: (0.05,0.0855]	11	45.5	63.6	100	0.839	2.40	0.645	0.491	0.786	0.772
Q4: (0.0855,0.74]	11	18.2	54.5	81.8	0.633	3.73	0.499	0.495	0.665	0.646

Table A3: Summary statistics for F_{oral} predictions between p.o. and i.v. simulations, grouped by different compound specific properties.

	no. APIs	**% within 2 fold**	**% within 3 fold**	**% within 10 fold**	**AFE**	**AAFE**	**R**	**CCC**	**R (of log data)**	**CCC (of log data)**
Molecular Weight										
Q1: (150,365 g/mol]	5	100	100	100	1.08	1.29	0.215	0.198	0.123	0.111
Q2: (365,440 g/mol]	4	50.0	50.0	75.0	0.701	3.93	-0.243	-0.226	-0.150	-0.144
Q3: (440,505 g/mol]	4	75.0	100	100	1.25	1.58	-0.325	-0.119	-0.314	-0.0894
Q4: (505,870 g/mol]	4	25.0	75.0	100	1.05	2.47	-0.156	-0.127	0.222	0.221
Acid/Base Nature										
Acid	5	80.0	80.0	80.0	0.581	2.40	-0.574	-0.478	-0.512	-0.189
Ampholyte	0									
Neutral	0									
Weak Base	4	25.0	75.0	100	1.48	2.46	-0.557	-0.296	-0.485	-0.363
Strong Base	8	75.0	87.5	100	1.16	1.69	0.794	0.767	0.839	0.815
LogP										
Q1: (-0.72 , 2.545]	6	66.7	100	100	0.819	1.61	0.671	0.607	0.729	0.548
Q2: (2.545 , 3.3]	1	100	100	100	0.92	1.09				
Q3: (3.3 , 4.49]	4	50.0	50.0	100	2.21	2.29	0.464	0.170	0.554	0.125
Q4: (4.49 , 7.75]	4	75.0	75.0	75.0	0.477	3.32	0.588	0.412	0.257	0.223
Not Given	2	50.0	100	100	1.76	1.76	-1.00	-0.236	-1.00	-0.231
LogD										
Q1: (-1.45,1.29]	3	66.7	66.7	66.7	0.418	4.17	-0.855	-0.706	-0.823	-0.283
Q2: (1.29,2.55]	4	25.0	100	100	0.797	1.96	0.540	0.515	0.530	0.422
Q3: (2.55,3.17]	1	100	100	100	1.59	1.59				
Q4: (3.17,5.8]	4	75.0	75.0	100	1.49	1.85	0.755	0.297	0.907	0.245
Not Given	5	80.0	80.0	100	1.36	1.58	0.61	0.584	0.803	0.75
BCS Classification										
BCS class I	3	100	100	100	0.94	1.06	0.928	0.653	0.928	0.66
BCS class II	9	66.7	77.8	100	1.82	1.94	0.478	0.3	0.768	0.591
BCS class III	1	0.00	100	100	0.497	2.01				
BCS class IV	4	50.0	75.0	75.0	0.329	3.81	0.55	0.34	0.209	0.155
Dose number										
$D_o \leq 1$	3	100	100	100	0.94	1.06	0.928	0.653	0.928	0.66
$D_o > 1$	14	57.1	78.6	92.9	1.02	2.36	0.18	0.179	0.286	0.283
Estimated fa										
$f_a < 0.9$	6	66.7	83.3	83.3	0.499	2.53	0.451	0.373	0.326	0.276
$f_a \geq 0.9$	11	63.6	81.8	100	1.47	1.83	0.517	0.439	0.643	0.575
BP										
Q1: (0.517,0.595]	3	66.7	66.7	66.7	0.377	3.94	-0.728	-0.566	-0.665	-0.232
Q2: (0.595,0.640]	5	80.0	100	100	1.32	1.57	0.602	0.568	0.906	0.862
Q3: (0.640,0.925]	2	50.0	50.0	100	2.45	2.45	1.00	0.329	1.00	0.228

Q4: (0.925,3.300]	7	57.1	85.7	100	0.973	1.78	0.641	0.641	0.675	0.674
fu_p										
Q1: (0.0002,0.0125]	4	100	100	100	1.13	1.41	0.797	0.771	0.968	0.940
Q2: (0.0125,0.05]	5	60.0	60.0	80.0	0.819	3.35	-0.719	-0.712	-0.595	-0.442
Q3: (0.05,0.0855]	4	50.0	75.0	100	1.91	1.99	0.744	0.644	0.874	0.754
Q4: (0.0855,0.74]	4	50.0	100	100	0.602	1.66	0.956	0.800	0.938	0.633

Table A4: Summary statistics for F_{rel} predictions between p.o. and solution simulations, grouped by different compound specific properties.

	no. APIs	**% within 2 fold**	**% within 3 fold**	**% within 10 fold**	**AFE**	**AAFE**	**R**	**CCC**	**R (of log data)**	**CCC (of log data)**
Molecular Weight										
Q1: (150,365 g/mol]	6	66.7	66.7	100	0.928	1.77	-0.743	-0.304	-0.626	-0.362
Q2: (365,440 g/mol]	7	57.1	71.4	100	0.482	2.07	0.827	0.427	0.704	0.242
Q3: (440,505 g/mol]	4	100	100	100	1.29	1.29	0.774	0.287	0.779	0.224
Q4: (505,870 g/mol]	7	85.7	85.7	100	0.693	1.47	0.894	0.82	0.726	0.465
Acid/Base Nature										
Acid	3	100	100	100	0.868	1.35	0.466	0.136	0.523	0.113
Ampholyte	2	50.0	50.0	100	1.56	2.07	-1.00	-0.527	-1.00	-0.351
Neutral	3	66.7	66.7	100	0.578	1.73	-0.986	-0.352	-0.968	-0.462
Weak Base	8	62.5	75.0	100	0.656	1.87	0.195	0.130	0.148	0.0671
Strong Base	8	87.5	87.5	100	0.727	1.49	0.798	0.719	0.631	0.448
LogP										
Q1: (-0.72 , 2.545]	3	66.7	66.7	100	0.567	2.05	-0.788	-0.234	-0.704	-0.332
Q2: (2.545 , 3.3]	9	66.7	66.7	100	0.745	1.77	0.393	0.359	0.236	0.181
Q3: (3.3 , 4.49]	4	100	100	100	1.06	1.19	0.991	0.871	0.982	0.885
Q4: (4.49 , 7.75]	4	75.0	75.0	100	0.537	1.87	0.985	0.510	0.946	0.250
Not Given	4	75.0	100	100	0.888	1.55	-0.157	-0.146	-0.231	-0.220
LogD										
Q1: (-1.45,1.29]	3	100	100	100	0.868	1.35	0.466	0.136	0.523	0.113
Q2: (1.29,2.55]	4	75.0	75.0	100	0.829	1.70	0.136	0.125	0.0478	0.0368
Q3: (2.55,3.17]	4	75.0	75.0	100	1.09	1.66	0.0683	0.0662	-0.0243	-0.0236
Q4: (3.17,5.8]	4	50.0	75.0	100	0.606	1.88	-0.392	-0.179	-0.341	-0.214
Not Given	9	77.8	77.8	100	0.624	1.68	0.853	0.762	0.744	0.485
BCS Classification										
BCS class I	3	66.7	66.7	100	1.38	1.66	-0.49	-0.278	-0.476	-0.185
BCS class II	16	75	81.3	100	0.717	1.65	0.499	0.416	0.435	0.351
BCS class III	1	100	100	100	0.784	1.28				
BCS class IV	4	75	75	100	0.537	1.87	0.985	0.51	0.946	0.25
Dose number										
$D_o \leq 1$	3	66.7	66.7	100	1.38	1.66	-0.49	-0.278	-0.476	-0.185
$D_o > 1$	21	76.2	81	100	0.681	1.67	0.514	0.434	0.464	0.331
Estimated fa										
$f_a < 0.9$	8	87.5	87.5	100	0.681	1.56	0.555	0.457	0.61	0.324
$f_a \geq 0.9$	16	68.8	75	100	0.778	1.72	0.425	0.375	0.275	0.244
BP										
Q1: (0.517,0.595]	3	100	100	100	0.615	1.63	0.999	0.912	0.996	0.814
Q2: (0.595,0.640]	8	62.5	75.0	100	0.678	1.89	-0.109	-0.0845	0.0833	0.0609

Q3: (0.640,0.925]	8	75.0	75.0	100	0.681	1.57	0.683	0.540	0.664	0.337
Q4: (0.925,3.300]	5	80.0	80.0	100	1.11	1.52	-0.0527	-0.0156	-0.174	-0.0427
fu_p										
Q1: (0.0002,0.0125]	5	60.0	80.0	100	0.611	1.80	0.576	0.303	0.689	0.187
Q2: (0.0125,0.05]	8	75.0	75.0	100	0.674	1.66	0.515	0.416	0.454	0.363
Q3: (0.05,0.0855]	6	66.7	66.7	100	0.988	1.89	0.275	0.271	0.161	0.151
Q4: (0.0855,0.74]	5	100	100	100	0.755	1.33	0.709	0.493	0.757	0.589

Table A5: Summary statistics for CL or CL/F predictions for i.v. and p.o. simulations, grouped by different compound specific properties.

	no. APIs	**% within 2 fold**	**% within 3 fold**	**% within 10 fold**	**AFE**	**AAFE**	**R**	**CCC**	**R (of log data)**	**C (** **d**
Molecular Weight										
Q1: (150,365 g/mol]	11	27.3	45.5	81.8	0.736	4.02	0.261	0.0852	0.382	
Q2: (365,440 g/mol]	10	40	50	80	1.45	3.70	0.668	0.607	0.524	
Q3: (440,505 g/mol]	11	27.3	36.4	100	0.375	3.51	0.448	0.181	0.626	
Q4: (505,870 g/mol]	11	36.4	45.5	90.9	0.553	3.48	0.613	0.0676	0.538	
Acid/Base Nature										
Acid	10	10.0	10.0	80.0	0.832	6.84	-0.0331	-0.0329	0.238	
Ampholyte	4	25.0	25.0	50.0	0.642	8.17	0.919	0.0712	0.468	
Neutral	5	40.0	60.0	100	0.430	2.33	0.856	0.116	0.791	
Weak Base	11	27.3	27.3	90.9	0.611	4.06	0.0307	0.0235	0.496	
Strong Base	13	53.8	84.6	100	0.754	1.95	0.756	0.754	0.801	
LogP										
Q1: (-0.72 , 2.545]	9	44.4	55.6	88.9	0.502	3.56	0.532	0.499	0.497	
Q2: (2.545 , 3.3]	9	55.6	66.7	100	0.827	2.31	0.211	0.153	0.500	
Q3: (3.3 , 4.49]	8	25.0	50.0	87.5	0.481	3.62	0.812	0.081	0.786	
Q4: (4.49 , 7.75]	9	22.2	22.2	88.9	0.878	4.16	0.408	0.372	0.524	
Not Given	8	12.5	25.0	75.0	0.78	5.65	0.148	0.0884	0.422	
LogD										
Q1: (-1.45,1.29]	7	14.3	28.6	71.4	0.459	7.08	0.721	0.0724	0.488	
Q2: (1.29,2.55]	7	28.6	42.9	100	0.610	3.27	0.448	0.430	0.554	
Q3: (2.55,3.17]	7	57.1	57.1	85.7	1.65	3.64	0.216	0.212	0.315	
Q4: (3.17,5.8]	7	14.3	28.6	71.4	0.455	4.41	0.185	0.0734	0.795	
Not Given	15	40.0	53.3	100	0.672	2.63	0.203	0.169	0.516	
BCS Classification										
BCS class I	8	37.5	62.5	87.5	1.23	2.83	0.177	0.0636	0.458	
BCS class II	24	29.2	37.5	87.5	0.567	3.91	0.428	0.423	0.528	
BCS class III	2	100	100	100	1.45	1.45	1	0.947	1	
BCS class IV	9	22.2	33.3	88.9	0.529	4.82	0.733	0.0794	0.596	
Dose number										
$D_o \leq 1$	8	37.5	62.5	87.5	1.23	2.83	0.177	0.0636	0.458	
$D_o > 1$	35	31.4	40	88.6	0.588	3.9	0.347	0.0885	0.545	
Estimated fa										
$f_a < 0.9$	15	40	46.7	80	0.604	4	0.3	0.0661	0.585	

$f_a \geq 0.9$	28	28.6	42.9	92.9	0.716	3.51	0.404	0.402	0.458	
BP										
Q1: (0.517,0.595]	11	0.00	18.2	81.8	1.02	6.31	0.197	0.0464	0.435	
Q2: (0.595,0.640]	9	11.1	11.1	66.7	0.298	7.00	0.189	0.123	0.231	
Q3: (0.640,0.925]	12	50.0	58.3	100	0.577	2.35	0.537	0.243	0.730	
Q4: (0.925,3.300]	11	63.6	81.8	100	1.03	2.05	0.793	0.790	0.624	
fu_p										
Q1: (0.0002,0.0125]	11	18.2	18.2	63.6	0.53	6.68	0.0460	0.0356	0.362	
Q2: (0.0125,0.05]	10	30.0	40.0	100	0.758	3.40	0.0589	0.0587	0.209	
Q3: (0.05,0.0855]	11	36.4	63.6	100	0.445	2.66	0.648	0.580	0.824	
Q4: (0.0855,0.74]	11	45.5	54.5	90.9	1.17	2.98	0.353	0.0553	0.523	

Table A6: Summary statistics for V_d or V_d/F predictions for i.v. and p.o. simulations, grouped by different compound specific properties.

	no. APIs	**% within 2 fold**	**% within 3 fold**	**% within 10 fold**	**AFE**	**AAFE**	**R**	**CCC**	**R (of log data)**	**CC (of dat**
Molecular Weight										
Q1: (150,365 g/mol]	11	27.3	63.6	100	1.33	2.94	0.895	0.189	0.804	
Q2: (365,440 g/mol]	10	30.0	60.0	80.0	2.06	3.23	0.448	0.340	0.601	
Q3: (440,505 g/mol]	11	36.4	72.7	90.9	0.968	2.6	0.308	0.270	0.637	
Q4: (505,870 g/mol]	11	45.5	54.5	100	1.08	2.74	0.0893	0.0374	0.442	
Acid/Base Nature										
Acid	10	10.0	40.0	90.0	1.52	4.87	0.306	0.141	0.616	
Ampholyte	4	0.00	25.0	75.0	0.659	5.53	0.848	0.116	0.325	
Neutral	5	60.0	80.0	100	1.76	1.90	0.920	0.378	0.924	
Weak Base	11	36.4	63.6	90.9	1.33	2.62	0.146	0.046	0.707	
Strong Base	13	53.8	84.6	100	1.2	1.96	0.694	0.687	0.670	
LogP										
Q1: (-0.72 , 2.545]	9	33.3	66.7	100	0.917	2.67	0.637	0.596	0.802	
Q2: (2.545 , 3.3]	9	22.2	66.7	88.9	1.10	2.78	0.0359	0.0227	0.262	
Q3: (3.3 , 4.49]	8	50.0	75.0	100	1.14	2.42	0.452	0.0755	0.533	
Q4: (4.49 , 7.75]	9	44.4	55.6	88.9	2.43	3.15	0.263	0.131	0.511	
Not Given	8	25.0	50.0	87.5	1.23	3.40	0.279	0.185	0.534	
LogD										
Q1: (-1.45,1.29]	7	28.6	57.1	100	0.725	3.47	0.376	0.113	0.705	
Q2: (1.29,2.55]	7	42.9	57.1	100	0.698	2.35	0.738	0.639	0.736	
Q3: (2.55,3.17]	7	14.3	42.9	71.4	1.73	4.67	0.0981	0.0884	0.251	
Q4: (3.17,5.8]	7	28.6	71.4	100	1.37	2.44	0.425	0.122	0.817	
Not Given	15	46.7	73.3	93.3	1.89	2.45	0.437	0.358	0.601	
BCS Classification										
BCS class I	8	37.5	62.5	100	1.12	2.75	0.909	0.172	0.622	
BCS class II	24	29.2	66.7	87.5	1.31	2.86	0.347	0.319	0.65	
BCS class III	2	50	50	100	1.71	3.28	1	0.463	1	
BCS class IV	9	44.4	55.6	100	1.29	2.88	0.0934	0.0442	0.671	
Dose number										
$D_o \leq 1$	8	37.5	62.5	100	1.12	2.75	0.909	0.172	0.622	
$D_o > 1$	35	34.3	62.9	91.4	1.33	2.89	0.194	0.185	0.644	
Estimated fa										
$f_a < 0.9$	15	33.3	53.3	93.3	1.37	3.53	0.0441	0.0419	0.634	

$f_a \geq 0.9$	28	35.7	67.9	92.9	1.24	2.56	0.385	0.342	0.606	
BP										
Q1: (0.517,0.595]	11	36.4	45.5	81.8	2.34	4.25	0.0059	0.00519	0.635	
Q2: (0.595,0.640]	9	11.1	44.4	88.9	1.18	3.74	0.338	0.314	0.686	
Q3: (0.640,0.925]	12	33.3	75.0	100	0.907	2.17	0.612	0.378	0.664	
Q4: (0.925,3.300]	11	54.5	81.8	100	1.11	2.10	0.888	0.886	0.642	
fu_p										
Q1: (0.0002,0.0125]	11	36.4	54.5	81.8	1.66	3.63	0.415	0.156	0.687	
Q2: (0.0125,0.05]	10	30.0	70.0	100	1.61	2.55	0.311	0.224	0.550	
Q3: (0.05,0.0855]	11	45.5	63.6	100	0.977	2.2	0.789	0.775	0.771	
Q4: (0.0855,0.74]	11	27.3	63.6	90.9	1.07	3.25	-0.0327	-0.0301	0.472	