RESPONSE DOCUMENT

INITIAL QUALIFICATION PROCEDURE –
SECOND LIST OF ISSUES SIMCYP SIMULATOR

June 26, 2024

CONTACT: **KAREN ROWLAND YEO**
SENIOR VICE-PRESIDENT
CERTARA UK LTD
LEVEL 2-ACERO
1 CONCOURSE WAY
SHEFFIELD S1 2BJ
UK

EMAIL: KAREN.YEO@CERTARA.COM

EMA Issue 1

The applicant stated that they only optimised kinact as based on their experience no optimisation is needed for KI. This could be understood if the plasma concentration is >> KI.

Ki and KI were however in the same order of magnitude, so this is unlikely to be the case. Please discuss.

Relevant excerpt from the scientific discussion: The Applicant stated that they only optimised kinact as based on their experience no optimisation is needed for KI. This may not always be the case. Therefore, two different scenarios in the clinical study (e.g. different dosing) may be required to optimise both parameters. For KI and kinact determination, standardised procedures (see e.g. Grimm et al., Drug Metabolism and Disposition, 2009, 1355) should be used. The Applicant should ensure that the KI and kinact values in the compound files are reliable. Thereby it should be demonstrated that no optimisation is needed for KI.

Response

Predicting the magnitude of time-dependent CYP-mediated DDIs from *in vitro* data requires estimates of inactivation parameters of the inhibitor (K_I, k_{inact}) obtained using a standardised approach and the turnover of the CYP enzyme (k_{deg}). The latter is discussed in the next section (Issue 2). Herein, we focus on the inactivation parameters. Before addressing the above question/comment relating to Issue 1, we feel that it is important to describe the experiment and subsequent data analysis used to derive the inactivation parameters as well as provide some background for context.

The rate of enzyme inactivation is proportional to the inactivator concentration and can be saturated at high concentrations according to the following equation:

$$k_{obs} = \frac{k_{inact} \times [I]}{K_I + [I]} \qquad (1)$$

where k_{obs} is the first-order rate constant of inactivation at inactivator concentration [I], k_{inact} is the maximum inactivation rate (a theoretical value that cannot be experimentally observed), and KI is the inactivator concentration when the rate of inactivation reaches half of k_{inact}. The kinetic parameters are typically generated using an experiment consisting of a preincubation step during which the perpetrator is incubated at varying concentrations for specific time periods. Following this, aliquots of the preincubations are diluted to quench the interaction and assessed for remaining CYP activity. At each of the inactivator concentrations, the natural logarithm of the remaining enzyme activity is plotted against the preincubation time. The apparent inactivation rate constants (k_{obs}) are determined from the slopes of the initial linear phase. Then the k_{obs} values are plotted against the inactivator concentrations ([I]) and the inactivation parameters are finally obtained by non-linear regression analysis. This standardised *in vitro* approach for derivation of the inactivation parameters has been described in detail previously by Grimm *et al.* [1]. It is important to note that in a recent complex but comprehensive numerical analysis of *in vitro* inactivation data and predictions of DDIs

involving MBI, Yadav *et al.* [2] concluded that as the conventional replot method described above [1] considers only the *initial* rate of inactivation, ignoring the curvature in the plots at later time points, and thus can lead to an overestimation of k_{inact}.

Given the complex study design as well as the subsequent data analysis, it is not surprising that inactivation parameters for the same inhibitor generated using the same systems (HLM or human hepatocytes (HH)) and a similar experimental approach can vary significantly across publications (Table 1). Typically, the inactivation parameters are expressed as a ratio to give an indication of the potency of the MBI compound. Thus, we cite this ratio as well as the inactivation parameters in the table below.

Table 1. Published inactivation parameters for CYP3A MBI compounds

Source	Hours	Hepatocytes			HLM			
		K_I	k_{inact}	k_{inact}/K_I	K_I	k_{inact}	k_{inact}/K_I	HLM/Hep
		uM	h^{-1}	h^{-1}/uM	uM	h^{-1}	h^{-1}/uM	
3	Diltiazem	8.90	1.37	0.15	1.53	1.44	0.94	6.12
3	Erythromycin	67.90	4.74	0.07	5.33	3.66	0.69	9.84
3	Verapamil	2.10	1.63	0.78	3.19	4.56	1.43	1.84
3	TAO	3.40	22.14	6.51	0.44	4.68	10.64	1.63
4	Crizotinib	0.89	0.78	0.88	0.37	6.90	18.65	21.28
5	Crizotinib	1.60	1.86	1.16	0.40	2.46	6.15	5.29
5	Erythromycin	1.60	3.72	2.33	0.40	1.92	4.80	2.06
6-Simcyp	**Clarithromycin**				**12.00**	**2.13**	**0.18**	
6-Simcyp	**Diltiazem**				**0.84**	**0.70**	**0.84**	
6-Simcyp	**NDM**				**4.75**	**1.09**	**0.23**	
6-Simcyp	**Erythromycin**				**23.20**	**3.18**	**0.14**	
6-Simcyp	**Verapamil**				**2.21**	**0.66**	**0.30**	1.
7	Clarithromycin	4.47	0.67	0.15	36.20	4.87	0.13	0.90
7	Diltiazem	8.96	1.30	0.15	1.32	0.65	0.50	3.41
7	NDM	2.96	0.76	0.26	0.96	0.57	0.60	2.31

7	Erythromycin	8.13	0.85	0.10	14.30	3.34	0.23	2.25
7	Verapamil	0.21	1.03	5.03	1.71	2.92	1.71	0.34

3. Chen Y, Liu L, Monshouwer M, Fretland AJ. Drug Metab Dispos. 2011 Nov;39(11):2085-92

4. Mao J, Johnson TR, Shen Z, Yamazaki S. Drug Metab Dispos. 2013 Feb;41(2):343-52.;

5. Mao J, Tay S, Khojasteh CS, Chen Y, Hop CE, Kenny JR. Pharm Res. 2016 May;33(5):1204-19.

6. Rowland Yeo K, Walsky RL, Jamei M, Rostami-Hodjegan A, Tucker GT. Eur J Pharm Sci. 2011 Jun 14;43(3):160-73.

7. Tseng E, Eng H, Lin J, Cerny MA, Tess DA, Goosen TC, Obach RS. Drug Metab Dispos. 2021 Oct;49(10):947-960.

Over the past decade, a number of key studies which aimed to assess whether inactivation parameters from HLM or HH led to more accurate predictions of clinical DDIs, have been published, albeit mainly for CYP3A4. In 2011, Chen *et al.* [3] reported that when comparing inactivation parameters in HLM, the KI values from hepatocytes were in general 4- to 13-fold higher than those in HLM, whereas the k_{inact} values in HH were similar or slightly higher or lower depending on the inhibitor. The inactivation potency (k_{inact}/K_I) for four CYP3A4 inactivators in HH was generally lower than that obtained in HLM due to either a lower affinity (KI) or a lower inactivation rate (k_{inact}) or both. Not surprisingly, when the data were applied in simulations using Simcyp, the prediction accuracy of the kinetic parameters from HH was better than that of the HLM data. However, there remained an overprediction even for hepatocytes. In these simulations the default CYP3A4 k_{deg} value of 0.0077 h^{-1} was applied.

Thereafter, we performed our own analysis in collaboration with Pfizer where we generated inactivation parameters for five mechanism-based inhibitors of CYP3A4 of variable potency (azithromycin, clarithromycin, diltiazem, erythromycin and verapamil) determined anew under standardised conditions in a single laboratory using HLM [6]. The inactivation parameters determined using the conventional experimental approach in HLM [1] in combination with the hepatic CYP3A4 k_{deg} value 0.0193 h^{-1} (see Issue 2) led to improved predictions of DDIs compared with Chen at al. [2].

More recently, Tseng *et al.* [7] published another CYP3A4 analysis whereby they identified 23 drugs with clinical DDI data. As part of the analysis, CYP3A4 inactivation parameters were generated in HLM and HH. Among the compounds evaluated, KI values ranged from 0.360 to 202 mM in HLM and 0.574 to 51.2 mM in HH. After correcting these values for nonspecific microsomal binding (NSMB – application of fumic), on average, the $K_{I,u}$, k_{inact}, and $k_{inact}/ K_{I,u}$ values in HLM were 1.2-, 3.5-, and 3.2-fold greater than in HH. Thus, it appears that applying a NSMB correction (via fu_{mic}), leads to minimal differences in KI values generated in HLM *versus* HH systems. Thereafter, the inactivation parameters in combination with a hepatic k_{deg}

value of 0.0193 h^{-1} were applied in simulations in Simcyp (V19); DDI predictions resulted in a GMFE of 1.7 and 1.6 for HLM and HH, respectively. Interestingly, the inactivation parameters generated in HLM in the study by Tseng *et al.* [6] were more potent than the ones reported in our publication, which led to the lower prediction accuracy.

Conclusion

Thus, based on the above information, for a new investigational drug that is a MBI compound, it is important to correct the KI for NSMB (apply a fumic value). Thereafter, if the clinical DDI is overpredicted, we recommend that the k_{inact} should be optimised initially (e.g crizotinib). However, we do recognise that in some cases the $K_{I,u}$ value may have to be optimised, especially to capture the non-linearity observed over multiple dosing of the compound due to auto-inhibition. In fact, of the 24 sets of inactivation parameters used for the moieties in the qualification DDI matrix, 3 k_{inact} values and 2 KI values were optimised.

References

1. Grimm SW, Einolf HJ, Hall SD, He K, Lim HK, Ling KH, Lu C, Nomeir AA, Seibert E, Skordos KW, Tonn GR, Van Horn R, Wang RW, Wong YN, Yang TJ, Obach RS. The conduct of in vitro studies to address time-dependent inhibition of drug-metabolizing enzymes: a perspective of the pharmaceutical research and manufacturers of America. Drug Metab Dispos. 2009 Jul;37(7):1355-70. doi: 10.1124/dmd.109.026716. Epub 2009 Apr 9. PMID: 19359406.
2. Yadav J, Paragas E, Korzekwa K, Nagar S. Time-dependent enzyme inactivation: Numerical analyses of in vitro data and prediction of drug-drug interactions. Pharmacol Ther. 2020 Feb;206:107449. doi: 10.1016/j.pharmthera.2019.107449. Epub 2019 Dec 11. PMID: 31836452; PMCID: PMC6995442.
3. Chen Y, Liu L, Monshouwer M, Fretland AJ. Determination of time-dependent inactivation of CYP3A4 in cryopreserved human hepatocytes and assessment of human drug-drug interactions. Drug Metab Dispos. 2011 Nov;39(11):2085-92. doi: 10.1124/dmd.111.040634. Epub 2011 Aug 11. PMID: 21835977.
4. Mao J, Johnson TR, Shen Z, Yamazaki S. Prediction of crizotinib-midazolam interaction using the Simcyp population-based simulator: comparison of CYP3A time-dependent inhibition between human liver microsomes versus hepatocytes. Drug Metab Dispos. 2013 Feb;41(2):343-52. doi: 10.1124/dmd.112.049114. Epub 2012 Nov 5. PMID: 23129213.
5. Mao J, Tay S, Khojasteh CS, Chen Y, Hop CE, Kenny JR. Evaluation of Time Dependent Inhibition Assays for Marketed Oncology Drugs: Comparison of Human Hepatocytes and Liver Microsomes in the Presence and Absence of Human Plasma. Pharm Res. 2016 May;33(5):1204-19. doi: 10.1007/s11095-016-1865-9. Epub 2016 Feb 11. PMID: 26869174.
6. Rowland Yeo K, Walsky RL, Jamei M, Rostami-Hodjegan A, Tucker GT. Prediction of time-dependent CYP3A4 drug-drug interactions by physiologically based pharmacokinetic modelling: impact of inactivation parameters and enzyme turnover. Eur J Pharm Sci. 2011 Jun 14;43(3):160-73. doi: 10.1016/j.ejps.2011.04.008. Epub 2011 Apr 20. PMID: 21540107.
7. Tseng E, Eng H, Lin J, Cerny MA, Tess DA, Goosen TC, Obach RS. Static and Dynamic Projections of Drug-Drug Interactions Caused by Cytochrome P450 3A Time-

Dependent Inhibitors Measured in Human Liver Microsomes and Hepatocytes. Drug Metab Dispos. 2021 Oct;49(10):947-960. doi: 10.1124/dmd.121.000497. Epub 2021 Jul 29. PMID: 34326140.

EMA Issue 2

Liver kdeg values are not identical between enzymes involved. The Applicant should discuss the robustness of kdeg values and the values that would be included in the public qualification report.

Relevant excerpt from the scientific discussion: *In addition, liver kdeg values play a role and are not identical between enzymes involved. It is unclear how robust the kdeg values are. For CYP2D6, for example, the Applicant provided several values, and it is unclear, which one was used in the model. The Applicant should ensure that kdeg values and associated uncertainty should be included in the public qualification report.*

Response

The k_{deg} values that will be specified in the public qualification report including the source references are indicated below in Table 1. We have also provided a comparison against the values cited in the final ICH M12 guidance.

Table 1. Values of k_{deg} used for individual CYP enzymes.

Enzyme	V19		Source references	ICH M12		Source references
	t1/2 (h)	kdeg (h^{-1})		t1/2 (h)	k_{deg} (h^{-1})	
CYP1A2	38.5	0.0183	1,2	38	0.018	5
CYP2C8	23	0.0301	2	22	0.0318	6
CYP2C9	104	0.0067	2	104	0.0066	2
CYP2C19	26	0.0267	2	26	0.0264	2
CYP2D6	70	0.0099	2	51	0.0138	7,8
CYP3A4	36	0.0193	3	36	0.0192	3
CYP3A4gut	24	0.03	4	24	0.0288	4

1 Diaz, D.; Fabre, I.; Daujat, M.; Saintaubert, B.; Bories, P.; Michel, H. and Maurel, P. (1990) Gastroenterology, 99(3), 737-747.

2 Renwick, A.B.; Watts, P.S.; Edwards, R.J.; Barton, P.T.; Guyonnet, I.; Price, R.J.; Tredger, J.M.; Pelkonen, O.; Boobis, A.R. and Lake, B.G. (2000) Drug Metab. Dispos., 28(10), 1202-1209.

3 Fromm MF; Busse D; Kroemer HK; Eichelbaum M. Hepatology. 1996 Oct;24(4):796-801.

4 Greenblatt DJ, von Moltke LL, Harmatz JS, Chen G, Weemhoff JL, Jen C, Kelley CJ, LeDuc BW, Zinny MA. Time course of recovery of cytochrome p450 3A function after single doses of grapefruit juice. Clin Pharmacol Ther. 2003 Aug;74(2):121-9. doi: 10.1016/S0009-9236(03)00118-8. PMID: 12891222.

5 Faber, M.S. and Fuhr, U. (2004) Clin. Pharmacol. Ther., 76(2), 178-184.

6 Backman JT, Honkalammi J, Neuvonen M, Kurkinen KJ, Tornio A, Niemi M, Neuvonen PJ. Drug Metab Dispos. 2009 Dec;37(12):2359-66.

7 Liston, H.L.; DeVane, C.L.; Boulton, D.W.; Risch, S.C.; Markowitz, J.S. and Goldman, J. (2002) J. Clin. Psychopharmacol., 22(2), 169-173.

8 Venkatakrishnan, K. and Obach, R.S. (2005) Drug Metab. Dispos., 33(6), 845-852.

A discussion on the derivation of the final values is given hereafter. *In vivo* enzyme levels are controlled by the rates of *de novo* enzyme synthesis and degradation. Over the past two decades, there has been much debate and discussion about appropriate values to use for the *in vivo* turnover half-lives of human CYP enzymes which can have a significant impact on accurate prediction of drug-drug interactions (DDIs) associated with mechanism (time)-based inhibition (MBI).

In 2008, we reviewed the available literature to provide an understanding of CYP regulation and discussed the pros and cons of various *in vitro* and *in vivo* approaches used to estimate the turnover of specific CYPs [9]. Ideally, the measurement of enzyme turnover requires specific labelling of the enzyme i.e. incorporation of [^{3}H]-leucine into the enzyme by preincubation, replacement with unlabelled leucine in the culture medium, and harvesting of cells for measurement of radioactivity in specific enzymes over time (*In vitro*: method 1). However, as this is difficult to do in humans, various *in vitro* and indirect *in vivo* methods have been used to estimate the turnover of CYP enzymes. *In vitro* studies have demonstrated that CYP apoprotein and enzyme activities decline in parallel over time in cultured human hepatocytes and liver slices. Assuming that these changes reflect solely endogenous enzyme degradation, turnover of CYP enzymes may be estimated based on the time-profile of enzyme level (*In vitro*: method 2). *In vivo* approaches involving the administration of multiple doses of an enzyme inducer until a new steady state enzyme level has been attained have also been applied (*In vivo*: method 1). After discontinuation of the inducer, the enzyme level returns to its basal level, which is monitored using a specific probe substrate. As the rate of recovery is a function of both enzyme turnover and the induction effect, a prerequisite of this approach is to use an inducer with a much shorter half-life than the target enzyme. This method can also be applied using mechanism-based inhibitors where recovery of the enzyme is monitored following inactivation (*In vivo*: method 2). Finally, chronic exposure to some drugs results in increased expression of the CYP enzyme(s) involved in their own metabolism (auto-induction). Modelling of the plasma concentration – time data of an auto-inducer after multiple doses allows characterisation of its time-dependent clearance, but also provides a means of determining the half-life of the induced enzyme (*In vivo*: method 3). Turnover of CYP enzymes estimated using the aforementioned approaches were collated and presented in the table shown below [9].

Table 2. Turnover half-lives of human hepatic CYP enzymes [9]

Enzyme	Method	n	$t_{1/2}$ (h) *
CYP1A2	*In vitro* Method 1	1	51
	In vitro Method 2	NC	43**
	In vitro Method 2	5	36 (8-58)
	In vivo Method 1	12	39 (27-54)
	In vivo Method 3	7	105
CYP2A6	*In vitro* Method 2	2	26 (19-37)
CYP2B6	*In vitro* Method 2	1	32
CYP2C8	*In vitro* Method 2	5	23 (8-41)
CYP2C9	*In vitro* Method 2	5	104
CYP2C19	*In vitro* Method 2	3	26 (7-50)
CYP2D6	*In vitro* Method 2	4	70
	In vivo Method 2	13	51
CYP2E1	*In vitro* Method 2	5	27 (7-40)
	In vivo Method 1	6	60
	In vivo Method 2	11	50 ± 19
CYP3A4	*In vitro* Method 1	1	44
	In vitro Method 2	NC	26**
	In vitro Method 2	4	79
	In vivo Method 1	15	72**
	In vivo Method 3	6	96 ± 38 (53-154)
	In vivo Method 1	7	72 (20-146)
	In vivo Method 1	3	(85-806)
	In vivo Method 1	8	(36-50)
	In vivo Method 3	13	10** (2-158)
	In vivo Method 3	35	94 (62-205)
	In vivo Method 3	7	70
	In vivo Method 3	16	85 ± 61
	In vivo Method 3	6	140 (48-284)
CYP3A5	*In vitro* Method 2	3	36 (15-70)

* $t_{1/2} = 0.693 / k_{deg}$. Values in brackets are ranges and ± indicates SD
** estimated from an analysis of the reported data

At the time (2008), given the absence of a consensus on values for the *in vivo* turnover half-lives of key CYPs, we recommended that a sensitivity analysis to assess the impact of the turnover half-lives on predicted DDIs should be an integral part of the modelling exercise, and the selective use of values should perhaps be avoided.

Since then, a number of key studies which aimed to address the above issue i.e. determine the most appropriate half-lives (or k_{deg} values) in combination with *in vitro* determined inactivation parameters have been published, albeit mainly for CYP3A4. In 2011, Chen *et al.* [10] demonstrated that when inactivation data generated in HLM and human hepatocytes (HH) for three CYP3A4 MBI compounds (erythromycin, verapamil, diltiazem) were applied in simulations using Simcyp, the prediction accuracy of the kinetic parameters from HH was better than that of the HLM data (Table 3). However, there remained an overprediction even for

hepatocytes as indicated by the GMFE values (Table 3). In these simulations the default CYP3A4 k_{deg} value of 0.0077 h^{-1} which reflected the turnover of the enzyme based on *in vivo* studies (see Table 1 above) was applied.

Table 3. Prediction accuracy of using inactivation parameters generated in HLM and HH [10]

	DDI Prediction Accuracy		
	HLM	HH: Add Method	HH: Dilution Method
Erythromycin (n = 11)			
MRS	7.73	0.317	−0.01
RMSE	10.8	1.10	1.06
GMFE	2.75	1.27	1.28
Verapamil (n = 7)			
MRS	1.62	−0.49	−0.15
RMSE	2.58	1.01	1.02
GMFE	1.50	1.31	1.28
Diltiazem (n = 12)			
MRS	2.64	−0.85	−1.01
RMSE	4.15	1.07	1.22
GMFE	1.71	1.44	1.54
All (n = 30)			
MRS	4.27	−0.34	−0.44
RMSE	7.17	1.07	1.11
GMFE	1.97	1.35	1.38

RMSE – root mean squared error; GMFE – geometric mean fold error.

Thereafter, we performed our own analysis in collaboration with Pfizer where we predicted the magnitude of DDIs observed in 29 *in vivo* studies involving six CYP3A4 probe substrates and five mechanism-based inhibitors of CYP3A4 of variable potency (azithromycin, clarithromycin, diltiazem, erythromycin and verapamil) [11]. Inactivation parameters determined anew in a single laboratory under standardised conditions together with data from substrate and inhibitor files within the Simcyp Simulator (Version 9.3) were used to determine a value of the hepatic k_{deg} (0.0193 or 0.0077 h^{-1}) most appropriate for the prediction of DDIs involving time-dependent inhibition of CYP3A4. The higher value resulted in decreased bias (GMFE – 1.05 *versus* 1.30) and increased precision (RMSE – 1.29 *versus* 2.30) of predictions of mean ratios of AUC in the absence and presence of inhibitor. Thus, based on the simulations presented within this publication the hepatic CYP3A4 k_{deg} value was changed from 0.0077 (turnover half-lie of 90 hours) to 0.0193 h^{-1} (turnover half-life of 36 hours reflective of *in vitro* data and Fromm et al. [3]) and can typically be used in conjunction with inactivation parameters determined by the conventional experimental approach.

More recently, Tseng *et al.* [12] published another CYP3A4 analysis whereby they identified 23 drugs with published clinical DDI. As part of the analysis, CYP3A4 inactivation parameters

were generated in HLM and HH, and subsequently included in simulations run using Simcyp (V19) and a hepatic k_{deg} value of 0.0193 h^{-1}. The authors state that accurate projections of DDI were obtained with GMFE of 1.7 and 1.6 for HLM and HH, respectively. Sensitivity and specificity were 100% and 67% when using HLM data. This is of particular note, because inactivation data appear to be more commonly generated using HLM rather than HH.

In a very recent publication, *in vitro* inactivation parameters were generated in pooled HLM for 19 drugs (and two metabolites) for which clinical DDIs were available; these involved CYP1A2, 2B6, 2C8, 2C9, 2C19 and 2D6 [13]. These data were incorporated into the Simcyp Simulator (V20) and the simulations results were compared against observed data; there was a bias of 1.13, the GMFE was 1.6 and 80% of the predicted DDIs were within 2-fold of observed data. The authors conclude that DDIs can be reliably predicted from *in vitro* inactivation data generated in HLM for the 6 enzymes. With the exception of the CYP2D6 k_{deg} where a higher value of 0.0217 h^{-1} (rather than the Simcyp default of 0.0099 $^{h-1}$) was applied, all k_{deg} values used in simulations were the default values of V19 (and V20) as indicated below in Table 4.

Table 4. Comparison of k_{deg} values used in MBI analyses

	k_{deg} values (h^{-1})		
Enzyme	Default V19	Tseng (2021) V19	Tseng (2024) V20
CYP1A2	0.0183		0.0183
CYP2C8	0.0301		0.0301
CYP2C9	0.0067		0.0067
CYP2C19	0.0267		0.0267
CYP2D6	0.0099		0.0217
CYP3A4	0.0193	0.0193	

Conclusion

In our view, the simulation results published by Tseng [12] and Tseng [13] along with our own analysis indicate that the k_{deg} values (Table 4) are reasonably robust when used in combination with inactivation parameters generated *in vitro* in HLM or HH using the standard dilution approach.

References

1. Diaz D, Fabre I, Daujat M, Saint Aubert B, Bories P, Michel H, Maurel P. Omeprazole is an aryl hydrocarbon-like inducer of human hepatic cytochrome P450.

Gastroenterology. 1990 Sep;99(3):737-47. doi: 10.1016/0016-5085(90)90963-2. PMID: 2136526.
2. Renwick AB, Watts PS, Edwards RJ, Barton PT, Guyonnet I, Price RJ, Tredger JM, Pelkonen O, Boobis AR, Lake BG. Differential maintenance of cytochrome P450 enzymes in cultured precision-cut human liver slices. Drug Metab Dispos. 2000 Oct;28(10):1202-9. PMID: 10997941.
3. Fromm MF, Busse D, Kroemer HK, Eichelbaum M. Differential induction of prehepatic and hepatic metabolism of verapamil by rifampin. Hepatology. 1996 Oct;24(4):796-801. doi: 10.1002/hep.510240407. PMID: 8855178.
4. Greenblatt DJ, von Moltke LL, Harmatz JS, Chen G, Weemhoff JL, Jen C, Kelley CJ, LeDuc BW, Zinny MA. Time course of recovery of cytochrome p450 3A function after single doses of grapefruit juice. Clin Pharmacol Ther. 2003 Aug;74(2):121-9. doi: 10.1016/S0009-9236(03)00118-8. PMID: 12891222.
5. Faber MS, Fuhr U. Time response of cytochrome P450 1A2 activity on cessation of heavy smoking. Clin Pharmacol Ther. 2004 Aug;76(2):178-84. doi: 10.1016/j.clpt.2004.04.003. PMID: 15289794.
6. Backman JT, Honkalammi J, Neuvonen M, Kurkinen KJ, Tornio A, Niemi M, Neuvonen PJ. CYP2C8 activity recovers within 96 hours after gemfibrozil dosing: estimation of CYP2C8 half-life using repaglinide as an in vivo probe. Drug Metab Dispos. 2009 Dec;37(12):2359-66. doi: 10.1124/dmd.109.029728. Epub 2009 Sep 22. PMID: 19773535.
7. Liston HL, DeVane CL, Boulton DW, Risch SC, Markowitz JS, Goldman J. Differential time course of cytochrome P450 2D6 enzyme inhibition by fluoxetine, sertraline, and paroxetine in healthy volunteers. J Clin Psychopharmacol. 2002 Apr;22(2):169-73. doi: 10.1097/00004714-200204000-00010. PMID: 11910262.
8. Venkatakrishnan K, Obach RS. In vitro-in vivo extrapolation of CYP2D6 inactivation by paroxetine: prediction of nonstationary pharmacokinetics and drug interaction magnitude. Drug Metab Dispos. 2005 Jun;33(6):845-52. doi: 10.1124/dmd.105.004077. Epub 2005 Mar 23. PMID: 15788540.
9. Yang J, Liao M, Shou M, Jamei M, Yeo KR, Tucker GT, Rostami-Hodjegan A. Cytochrome p450 turnover: regulation of synthesis and degradation, methods for determining rates, and implications for the prediction of drug interactions. Curr Drug Metab. 2008 Jun;9(5):384-94. doi: 10.2174/138920008784746382. PMID: 18537575.
10. Chen Y, Liu L, Monshouwer M, Fretland AJ. Determination of time-dependent inactivation of CYP3A4 in cryopreserved human hepatocytes and assessment of human drug-drug interactions. Drug Metab Dispos. 2011 Nov;39(11):2085-92. doi: 10.1124/dmd.111.040634. Epub 2011 Aug 11. PMID: 21835977.
11. Rowland Yeo K, Walsky RL, Jamei M, Rostami-Hodjegan A, Tucker GT. Prediction of time-dependent CYP3A4 drug-drug interactions by physiologically based pharmacokinetic modelling: impact of inactivation parameters and enzyme turnover. Eur J Pharm Sci. 2011 Jun 14;43(3):160-73. doi: 10.1016/j.ejps.2011.04.008. Epub 2011 Apr 20. PMID: 21540107.
12. Tseng E, Eng H, Lin J, Cerny MA, Tess DA, Goosen TC, Obach RS. Static and Dynamic Projections of Drug-Drug Interactions Caused by Cytochrome P450 3A Time-Dependent Inhibitors Measured in Human Liver Microsomes and Hepatocytes. Drug Metab Dispos. 2021 Oct;49(10):947-960. doi: 10.1124/dmd.121.000497. Epub 2021 Jul 29. PMID: 34326140.
13. Tseng E, Lin J, Strelevitz TJ, DaSilva E, Goosen TC, Obach RS. Projections of Drug-Drug Interactions Caused by Time-Dependent Inhibitors of Cytochrome P450 1A2, 2B6, 2C8, 2C9, 2C19, and 2D6 Using In Vitro Data in Static and Dynamic Models. Drug Metab Dispos. 2024 Feb 26:DMD-AR-2024-001660. doi: 10.1124/dmd.124.001660. Epub ahead of print. PMID: 38408867.

EMA Issue 3

Please explain how* Ki values lower than the lowest reported in vitro values have been selected for the *PBPK model without optimisation of these parameters. There is a high variability between the median reported in vitro values and the Ki values used in Simcyp: reported values are on average 40-fold higher with a range of 0.3-690-fold. Please, discuss how in vitro data can be used for competitive inhibition. Should a positive control be included in the in vitro assays?

Relevant excerpt from the scientific discussion: For application of PBPK model, it is preferred that the Ki values are determined using the same standardised in vitro method for a rank order approach of a new drug as reversible inhibitor of multiple enzymes. There is more uncertainty when the Ki values are obtained from literature from various sources. The Applicant admits that the uncertainty and its relevance will vary on a case-by-case basis. ***Therefore, the in vitro data for the compounds qualified should be included in the public qualification report.*** *Moreover, for a new drug a clinical study with a sensitive substrate / strong inhibitor will always be necessary to verify in vitro parameters.*

Response

Lower than reported Ki values

There are two possible reasons why the Simcyp Ki values could fall below the lowest reported *in vitro* values for perpetrators. Firstly, the Simcyp Ki values could have been optimised to capture clinical DDI data. Secondly, the Ki values could have been corrected for non-specific microsomal binding (NSMB) which may occur in incubations during experimental determination of the inhibitory potency of the perpetrator. It should be noted that Ki values reported in the CDID (formerly UOW database) are not always corrected for NSMB (by application of fu_{mic}) and may reflect Ki not Ki,u (fu_{mic} x Ki) values.

For each perpetrator within the Simcyp Simulator, a meta-analysis of Ki values derived from various experimental sources was conducted. Each Ki value was corrected for NSMB by applying either a measured or predicted fu_{mic} value [1] determined at the final protein concentration used in the incubation of the *in vitro* experiment. Thus, *in most cases*, the final Simcyp Ki value is actually Ki,u and fu_{mic}=1.

Example 1: For transparency, we demonstrate the approach herein using an example where the actual Ki and fu_{mic} values for paroxetine are derived and applied in the Simulator. For each Ki value from a source reference, the appropriate fu_{mic} value was applied (Table 1). As different microsomal protein concentrations were used in the *in vitro* experiments to determine the individual Ki values, fu_{mic} values were different. The fu_{mic} value is inversely proportional to the microsomal protein concentration [2]. Final fu_{mic} and CYP2D6 Ki values of 0.2 and 0.46 µM, respectively, appear in the paroxetine compound file (Table 1).

Table 1. Individual CYP2D6 Ki values determined for paroxetine

System	Enzyme	fu_{mic}	$K_{i,u}$ [uM]	K_i [uM]
HLM	2D6	0.25	0.0175	0.07
HLM	2D6	0.17	0.01105	0.065
HLM	2D6	0.17	0.0612	0.36
HLM	2D6	0.17	0.0255	0.15

HLM	2D6	0.51	0.2754	0.54
HLM	2D6	0.51	0.3111	0.61
HLM	2D6	0.51	0.2754	0.54
HLM	2D6	0.2	0.034	0.17
HLM	2D6	0.09	0.0405	0.45
HLM	2D6	0.09	0.0396	0.44
HLM	2D6	0.09	0.0252	0.28
HLM	2D6	0.09	0.0477	0.53
HLM	2D6	0.09	0.081	0.9
HLM	2D6	0.09	0.0396	0.44
HLM	2D6	0.09	0.117	1.3
HLM	2D6	0.09	0.09	1
HLM	2D6	0.09	0.117	1.3
HLM	2D6	0.09	0.0342	0.38
HLM	2D6	0.09	0.126	1.4
HLM	2D6	0.09	0.108	1.2
	mean=	0.2	0.0876	0.46

Not surprisingly, once the fu_{mic} corrections were applied to the individual Ki values, the fold variability associated with the inhibitory potency across different laboratories was significantly reduced (Table 2).

Table 2. Variability across Ki measurements for paroxetine

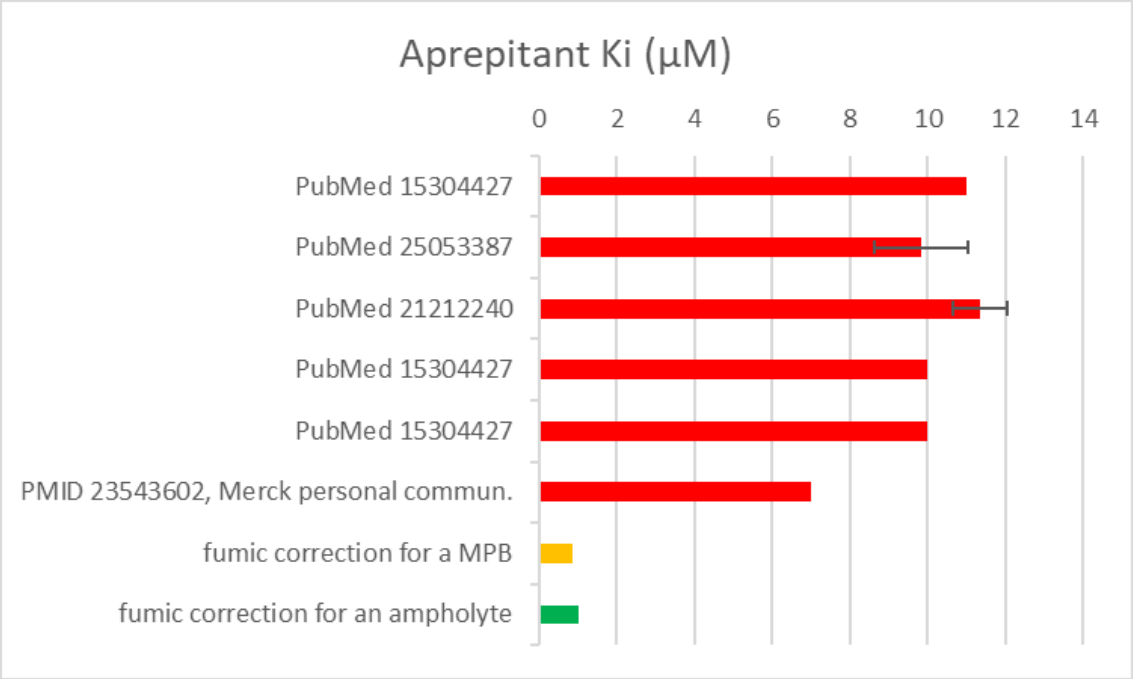
	Ki range before fu_{mic} correction	Ki range after fu_{mic} correction
Min [µM]	0.065	0.01105
Max [µM]	1.4	0.126
fold	21.54	11.4

For paroxetine, the final Ki,u value remains within the range of the CDID values; we mainly wanted to demonstrate the approach.

Example 2: For aprepitant, the average uncorrected Ki value from our meta-analysis falls in range of the CDID values. When corrected for fumic, $K_{i,u}$ is much lower than the lowest reported CDID value (Figure 1).

Interestingly, in this case, we contacted Merck directly who provided us with more precise data on the physicochemical properties of aprepitant allowing us to predict a more accurate fumic value; they indicated that the compound was an ampholyte not a monoprotic base (MPB) and this had an impact on the predicted fumic albeit small.

Figure 1. Range of cited values for CYP3A4 Ki values of aprepitant



The above examples confirm that it is important to correct *in vitro* Ki values for NSMB using a fu_{mic} value determined at the final microsomal protein concentration in the incubation. For

each of the 27 inhibitors (34 moieties including metabolites) used in our analysis, this was the initial step that was performed.

Thereafter, the *in vitro* derived $K_{i,u}$ values for the inhibitors were used in simulations of clinical DDIs involving probe/sensitive substrates. If the predicted magnitude of interaction was not captured within 1.25-fold of the observed value (typically), then the $K_{i,u}$ value was optimised.

Of the 59 *in vitro* derived $K_{i,u}$ values for the 34 moieties, 30.5% of them required optimisation using a clinical study with a probe/sensitive substrate.

Application of in vitro derived Ki values

Thus, for a new investigational drug, we expect the in vitro Ki values to be corrected for NSMB (apply fu_{mic}) to derive a $K_{i,u}$ value. In most cases, it is likely that the fu_{mic} will be predicted based on the drug's physicochemical properties rather than measured at least in early development [1].

The $K_{i,u}$ value will be used in initial simulations to assess the perpetrator potential of the investigational drug against sensitive/probe CYP substrates. If a significant interaction (AUC ratio >1.25-fold) is predicted, a clinical DDI study is likely to be required. The latter scenario would be covered by COU2- "The Simcyp Simulator (V19 R1) can be used to predict the CYP-mediated inhibitory effect of a drug on the exposure of other CYP substrates administered orally under fasted conditions or intravenously in healthy subjects when a clinical study with a sensitive CYP substrate has been conducted (and used to verify the competitive inhibition effect *in vivo*)."

Prior to conducting the clinical DDI study, the use of a positive control in the *in vitro* experiment could provide an indication of how robust the $K_{i,u}$ value of the investigation drug is. For example, if the latter is a CYP3A4 inhibitor, ketoconazole could be included as a positive control. The *in vitro* $K_{i,u}$ value for ketoconazole could be compared against the default value in the Simcyp Simulator which we know gives accurate predictions of CYP3A4-mediated DDIs. The fold difference (if there is one), could then be applied to the $K_{i,u}$ value of the investigational drug to investigate any potential uncertainty associated with the parameter.

References

1. Gardner I, Xu M, Han C, Wang Y, Jiao X, Jamei M, Khalidi H, Kilford P, Neuhoff S, Southall R, Turner DB, Musther H, Jones B, Taylor S. Non-specific binding of compounds in *in vitro* metabolism assays: a comparison of microsomal and hepatocyte binding in different species and an assessment of the accuracy of prediction models. Xenobiotica. 2022 Aug;52(8):943-956. doi: 10.1080/00498254.2022.2132426. PMID: 36222269.

EMA Issue 4

For CYP3A4, please present differential inhibition of intestinal or hepatic CYP3A4.

Response

Considerable effort has gone into identifying an appropriate *in vivo* probe for quantitative assessment of human hepatic and intestinal CYP3A activity. Various substrates including cortisol, lidocaine, dapsone, dextromethorphan, quinidine and erythromycin have been investigated but do not entirely fulfil the criteria laid out for an ideal probe [1; references within]. The most widely used CYP3A probe is midazolam whose IV and oral clearances reflect hepatic and first-pass CYP3A activity. Alfentanil is a newer CYP3A probe which is a low-extraction drug (compared to midazolam) cleared exclusively by hepatic metabolism. The systemic clearance of alfentanil was shown to be highly correlated with midazolam clearance and to accurately reflect hepatic CYP3A inhibition. Furthermore, it has been demonstrated that oral alfentanil is a sensitive probe for first-pass and intestinal CYP3A activity [1].

Our DDI qualification matrix, which was used to simulate 263 clinical studies, consisted of 25 studies involving CYP3A4 drugs where DDI data were available following IV and oral administration of the substrate; inhibitors were dosed orally. Amongst these 25 studies there were 6 substrate/inhibitor pairs. Importantly, the two substrates were midazolam and alfentanil. There were 9 studies involving IV administration of the two substrates and 16 studies involving oral administration.

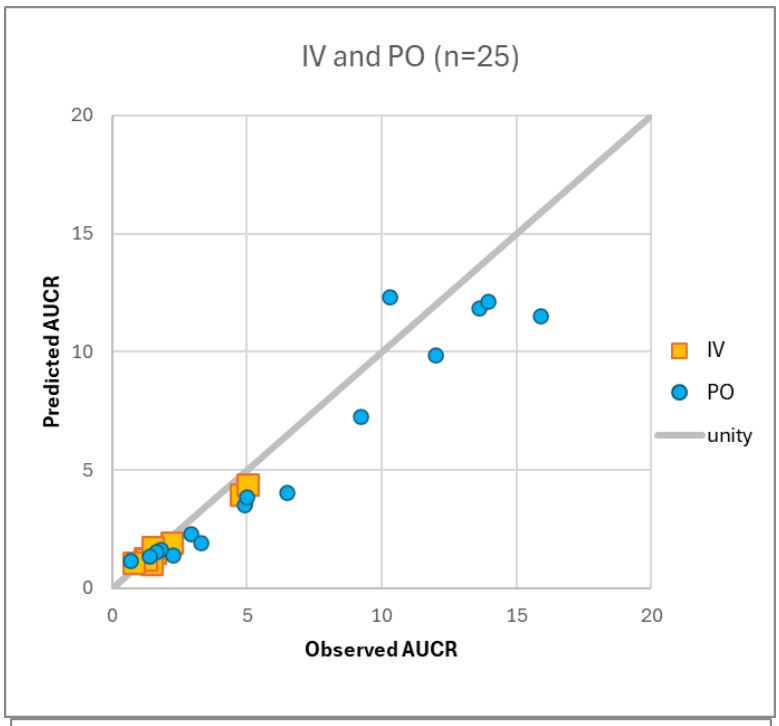
Table 1. Number of clinical studies for alfentanil and midazolam involving oral/IV administration

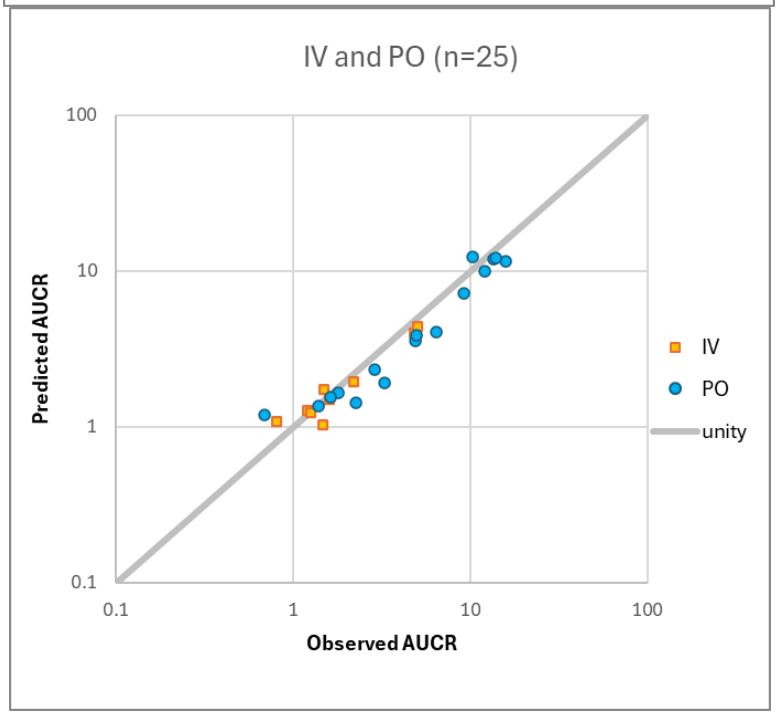
Substrate	Inhibitor	IV	PO	total
Alfentanil	Fluconazole	3	3	6
Alfentanil	Ketoconazole	1	3	4
Midazolam	Ketoconazole	1	5	6
Midazolam	Aprepitant	3	4	7
Midazolam	Fluvoxamine	1	1	2

We have provided compound file summaries for both midazolam and alfentanil to demonstrate that the predicted PK parameters are consistent with the observed data following both IV and

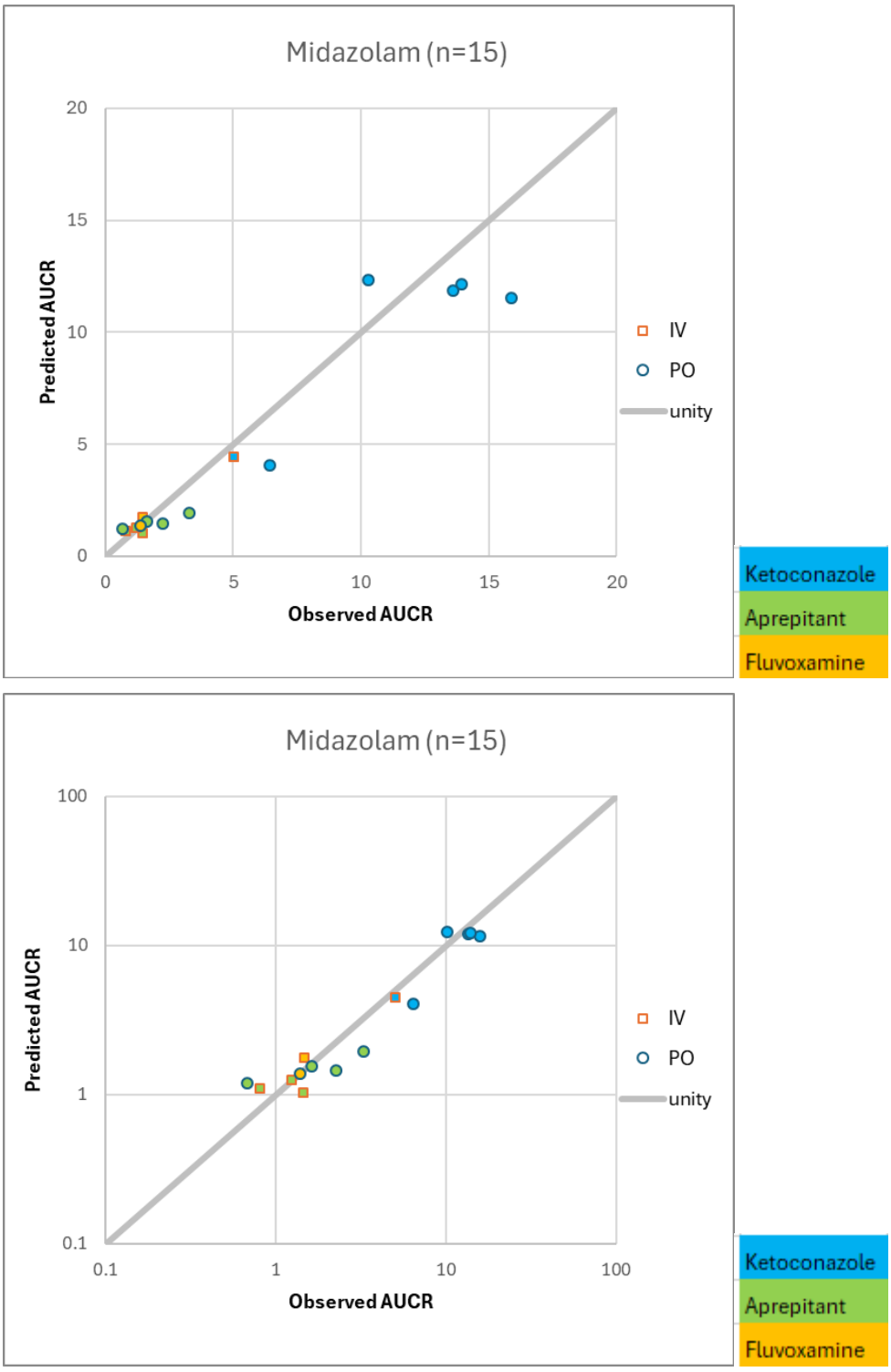
oral administration. The average fold error (AFE) for prediction of all 25 DDIs involving midazolam and alfentanil following oral and IV dosing was 0.90. We also provide the following figures:

1. A comparison of predicted *versus* observed AUC ratios for both alfentanil and midazolam (absolute scale and log scale)

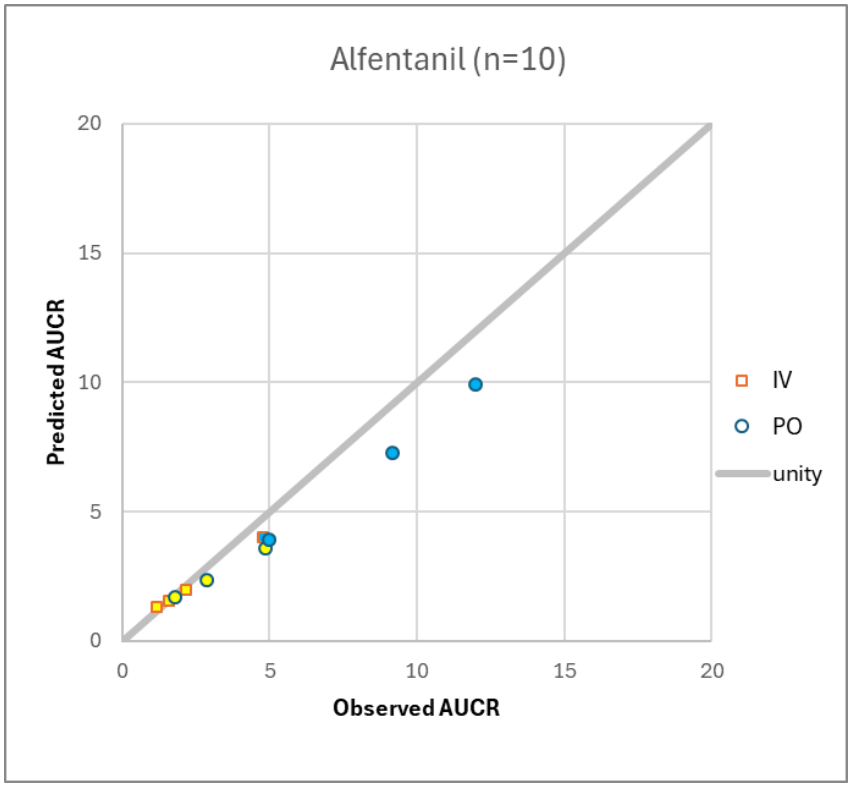




2. A comparison of predicted *versus* observed AUC ratios for midazolam (absolute scale and log scale)

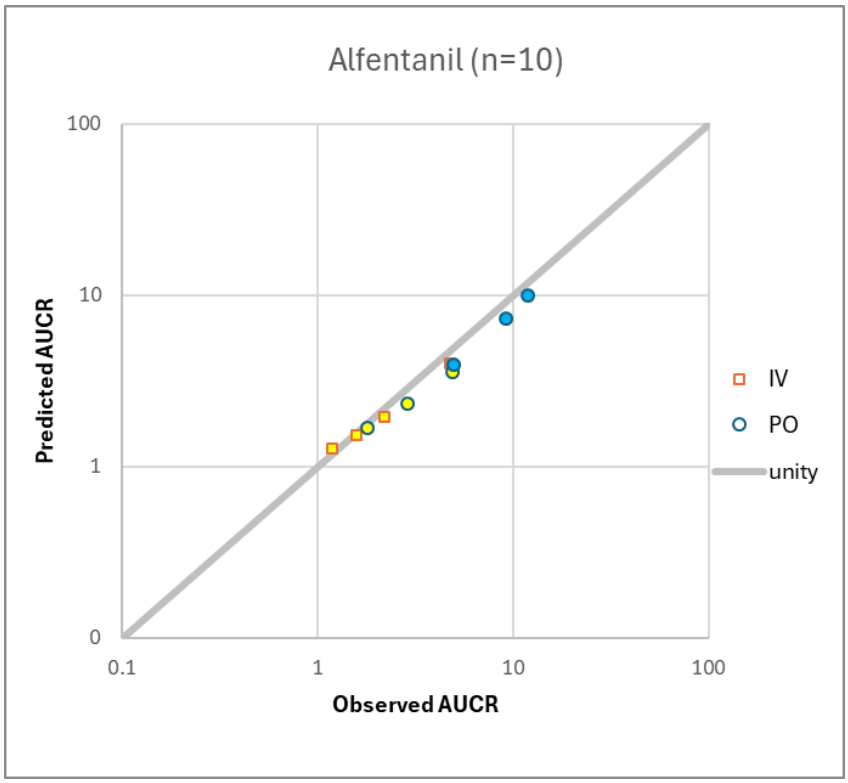


3. A comparison of predicted *versus* observed AUC ratios for alfentanil (absolute scale and log scale)



Ketoconazole

Fluconazole



Ketoconazole

Fluconazole

Conclusion

Predicted DDIs for both midazolam and alfentanil, two substrates identified as being good *in vivo* probes to assess differential inhibition in the liver (IV administration) and gut (oral administration assuming no change in fa) were compared against observed data. For both substrates, the simulations demonstrated that the differential inhibition of CYP3A in the gut and liver is captured reasonably well.

Reference

1. Kharasch ED, Walker A, Hoffer C, Sheffels P. Intravenous and oral alfentanil as in vivo probes for hepatic and first-pass cytochrome P450 3A activity: noninvasive assessment by use of pupillary miosis. Clin Pharmacol Ther. 2004 Nov;76(5):452-66. doi: 10.1016/j.clpt.2004.07.006. PMID: 15536460.

EMA Issue 5

The Applicant optimised some in vitro values in compound files based on clinical data. The Applicant should discuss, in which cases optimisation was regarded necessary, i.e. how the need for optimisation was determined.

Relevant excerpt from the scientific discussion: Some in vitro values in the compound files were optimised based on the clinical data, whereas the others were not. It remains unclear how the need for optimisation was determined. For a new drug, after a clinical study with a sensitive substrate / strong inhibitor has been conducted, a simulation should always be performed in order to verify in vitro parameters.

Response

The process of compound file development is described herein. In the initial stages, when developing a compound file, the aim is to accurately capture concentration-time profiles from clinical datasets based on *in vitro* data alone. Model development is performed initially using intravenous data (if available) with a focus on the distribution and elimination parameters. Thereafter, absorption related parameters are introduced into the PBPK models for each compound to predict plasma concentration-time profiles following oral administration. First order absorption is assumed initially followed by introduction of the ADAM model if the absorption is complex, transporters are involved, or formulation and food effect simulations are required. At each stage, optimisation of relevant parameters is performed using clinical data, if necessary, to ensure accurate recovery of observed parameters/profiles.

As part of the model development process, we would typically apply *in vitro-in vivo* extrapolation (IVIVE) to determine whether the volume of distribution (Vss) and CL (IV or oral) are adequately predicted (< 1.25-fold of the observed value), especially when the shape of the concentration-time profile is not captured. In addition, if observed data are available, predicted fa and F_G can be assessed accordingly.

Example 1: The compound file summary for midazolam demonstrates this clearly and indicates which parameters have been optimised in the development of the file. Accurate prediction of CL (following IV and oral administration) from *in vitro* data was demonstrated. However, the distribution was not accurately described using the predicted Vss value and the inputs (Vsac, k_{in} and k_{out}) had to be optimised to better capture the profile as seen in the extract below and the table of input parameters.

DISTRIBUTION

Following intravenous administration, midazolam is rapidly and widely distributed. In the V15 update to the Sim-Midazolam file, distribution was described using a SAC compartment within the minimal PBPK model. The parameters were fitted using IV data reported by Kupferschmidt et al. 1995. Values of kin = 0.2 1/h, kout = 0.25 1/h, Vsac 0.23 L/kg and Vss = 0.88 L/kg are used in the Sim-Midazolam file. Midazolam is significantly bound to plasma protein (fu 0.032).

Distribution Model	Minimal PBPk model	
Vss (L/kg)	0.88	(Kupferschmidt *et al.*, 1995)
Kin (1/h)	0.2	Derived from (Kupferschmidt *et al.*, 1995)
Kout (1/h)	0.25	Derived from (Kupferschmidt *et al.*, 1995)
V_{SAC} (L/kg)	0.23	Derived from (Kupferschmidt *et al.*, 1995)

Example 2: One of the more common *in vitro* parameters that requires optimisation is the metabolic intrinsic clearance obtained from *in vitro* data. Although IVIVE of investigational drugs metabolised by CYP enzymes has improved significantly over the past two decades, it remains an issue for the more metabolically stable drugs that are currently being developed. For a victim drug (substrate), it is important to characterise the clearance routes and demonstrate that when inhibited, the observed increase in exposure is accurately captured. Thus, if the *in vitro* metabolic data ($CLu_{H,int}$) underpredicts the observed CL (for either IV or oral), a retrograde model implemented within the Simcyp Simulator is used to estimate the hepatic intrinsic clearance ($CLu_{H,int}$) from an observed IV (reverse well stirred model) or oral clearance (Equation 1).

$$CLu_{H,int} = \frac{\frac{CL_{po}}{B:P} fa\, f_G - \frac{CL_R}{B:P}}{Uptake\, fu_b \left(1 + \frac{\frac{CL_R}{B:P}}{Q_H}\right)} \quad (1)$$

Where B:P is the concentration ratio of drug in blood to plasma; fu_b is fraction of unbound drug in blood (calculated from fu_p/B:P); Q_H is the blood flow in the hepatic vein; F_G is the fraction escaping first pass metabolism in the gut; CL_R is the renal clearance; fa is the fraction absorbed; Uptake is a factor that accounts for any active hepatic uptake). The final optimised $CLu_{H,int}$

value can then be apportioned to the different CYP enzymes based on the *in vitro* reaction phenotyping data and recombinant CYP enzymes.

To demonstrate this approach, we use metoprolol as an example. As described in the compound file summary, metoprolol is extensively metabolised in the liver. The major route of metabolism is O-demethylation to form O-Desmethylmetoprolol (ODM), which is then converted to a carboxylic acid as an end product (H117-04). Other routes are aliphatic hydroxylation to α-hydroxymetoprolol (αOHM) and oxidative deamination.

In vitro kinetic data describing the different routes of elimination are available; however, there was a significant underprediction of the CL when compared against the observed data following IV administration. Thus, a retrograde approach was used to estimate the $CL_{uint,H}$ from the IV clearance (57.4 L/h) after subtraction of the renal clearance (CL_R = 5.23 L/h).

Then the *in vitro* data were used to assign the fmCYP2D6 and this was able to capture the clinical DDIs. Therefore, no further optimisation was required. This is indicated in the text and the table of inputs extracted from the compound file summary and shown below.

Elimination

Metoprolol is extensively metabolised in the liver. The major route of metabolism is O-demethylation to form O-Desmethylmetoprolol, which is then converted to a carboxylic acid end product (H117-04). Other routes are aliphatic hydroxylation to *α*-hydroxymetoprolol and oxidative deamination. These routes account for 65, 10 and 10%, respectively of the dose recovered in urine (Borg *et al.* (1975)). Jordo *et al.* (1980) reported a bioavailability of 0.50 ± 0.11 (Mean ± SEM, n=6) after a **single oral dose** of 50 mg in healthy volunteers.

In the SV-Metoprolol file, the CL_{int} data were optimised using the *in vivo* clearance (CL_{iv}) and *in vitro* fm. Briefly, *in vitro* fm of 92% and 8% for CYP2D6 and CYP3A4 (in-house data), respectively, were used to back-calculate the CL_{int} data from CL_{iv} (57.4 L/h) and renal clearance (CL_R = 5.23 L/h (Regardh *et al.* (1974), Schaaf *et al.* (1987) and Hamelin *et al.* (2000))). Predicted mean oral clearance values (Mean ± SD) are 135.4 ± 109.3 L/h, 258.0 ± 220.2 L/h and 18.1 ± 6.5 L/h in CYP2D6 EM, UM and PMs, respectively, compared to the reported values of 144 ± 17 L/h, 369 ± 116 L/h and 25.4 ± 5.6 L/h, respectively (Blake *et al.* (2013)) (**Figure 8 A-C**). The reported values were based on a pooled analysis of CYP2D6 metaboliser phenotype and metoprolol pharmacokinetics combining 13 independent studies with a total sample size of 264 participants.

Enzyme	CYP2D6	
Pathway	Pathway 1	
CL_{int} (*μ*L/min/pmol)	3.61	Back calculated from meta analysis of CL_{iv} (see text) fm 2D6 from meta analysis of *in vitro* data including Madani *et al.* (1999), Otton *et al.* (1988)
Enzyme	CYP3A4	
Pathway	Pathway 1	
CL_{int} (*μ*L/min/pmol)	0.02	Back calculated from meta analysis of CL_{iv} (see text) fm 2D6 from meta analysis of *in vitro* data including Madani *et al.* (1999), Otton *et al.* (1988)
CL_R (L/h)	5.23	Regardh *et al.* (1974), Hamelin *et al.* (2000)

Example 3: For a perpetrator (inhibitor), it is necessary to ensure that after integration of the inhibitory parameters into the PBPK model, they lead to accurate prediction of clinical DDIs. Hence, after the PK profiles for the substrate and the inhibitor has been verified independently, the Ki values are evaluated.

For a bottom-up approach, an *in vitro* value that appropriately reflects the situation at the interacting binding site should be used. The assumption in the PBPK model is that the unbound drug in the intracellular water is the interacting concentration, hence the Ki value should reflect this and be corrected for the NSMB in the incubation of the system (i.e., fu_{mic} for the HLM assays). Unfortunately, fu_{mic} (more general the fu_{inc}) data are rarely determined and if they have been applied, they are seldom reported in the publication. The determination of such values is also still highly variable as we describe recently in a review and research paper [1]. The fu_{inc} corrected Ki values, $K_{i,u}$, can then be applied within the PBPK model.

However, it should be noted that while the fu_{mic} corrected Ki values often capture the relevant clinical DDIs, in some cases where the DDI is not predicted to within 1.25-fold, the $K_{i,u}$ may need to be optimised. In such a case, the value is optimised using a clinical DDI involving a sensitive probe for the relevant enzyme. Thereafter, verification of the optimised $K_{i,u}$ value is performed using another independent DDI study and if possible, a range of other probe substrates.

It should be noted that if there are no independent studies for verification, the compound becomes initially a 'Research' file. It is then made available with a compound summary on our curated members area so other researchers can build on the knowledge that we have already collated, and transparency is ensured.

For the purposes of the DDI qualification, we have generated several new compound files, most of which have been verified with several independent studies, However, we still call them 'Research' files as they are not integrated in the Simcyp Simulator V19. Several of these compounds have been integrated into later versions of the Simulator. Noteworthy is that several of those research files could only be verified with clinical studies that became available within the last three years, so after V19 release.

The verified $K_{i,u}$ and fu_{mic} values that have been used for the V19 qualification matrix for the competitive inhibition are listed together with their source are listed in Table 1.The verified $K_{app,u}$, fu_{mic}, and k_{inact} values that have been used for the V19 qualification matrix for the mechanism-based inhibition are listed together with their source in Tables 2a and b, respectively. For both sets of parameters, it is indicated when a value was optimised.

References

1. Gardner I, Xu M, Han C, Wang Y, Jiao X, Jamei M, Khalidi H, Kilford P, Neuhoff S, Southall R, Turner DB, Musther H, Jones B, Taylor S. Non-specific binding of compounds in *in vitro* metabolism assays: a comparison of microsomal and hepatocyte binding in different species and an assessment of the accuracy of prediction models. Xenobiotica. 2022 Aug;52(8):943-956. doi: 10.1080/00498254.2022.2132426. PMID: 36222269.

Table 1: Inhibitor information from Simcyp Simulator compound files (V19R1) showing classification, interaction parameters for competitive CYP inhibition.

Inhibitor	Enzyme	FDA Classification	$K_{i,u}$ (µM)	$K_{i,u}$ source	fu_{mic}
Cimetidine	CYP1A2	-	4.58	*In vitro-in vivo* scalar for CYP3A4 Ki value applied to CYP1A2 Ki (PMID: 10223772).	1
Ciprofloxacin	CYP1A2	Strong	2	A clinical DDI study with caffeine (PMID: 12908854) was used to derive an optimised CYP1A2 Ki value for ciprofloxacin.	0.92*
Fluvoxamine	CYP1A2	Strong	0.002	A clinical DDI study with caffeine using the fluvoxamine 100 mg BID dosage regimen (PMID: 16236038) was used to derive an optimised CYP1A2 Ki value for fluvoxamine.	1
Propranolol	CYP1A2	-	0.011	A clinical DDI study with theophylline using the propranolol 40 mg TID dosage regimen (PMID: 4041342) was used to derive an optimised CYP1A2 Ki value for propranolol.	1
Cimetidine	CYP2C19	-	8.2	**In vitro** value from (PMID: 11334262) with an applied *in vitro/in vivo* scalar of 19.2 (obtained from the CYP3A4 Ki *in vitro/in vivo* relationship).	1
Fluvoxamine	CYP2C19	**Strong**	0.006	Original value derived from **in vitro** HLM data. Value lowered ten-fold as ten-fold difference between in vitro and in vivo Ki for fluvoxamine was reported (PMID: 12695344; PMID: 11719727).	1
Fluoxetine	CYP2C19	Strong	9.39	**In vitro** Ki and fumic from (PMID: 23785064).	1
Nor-Fluoxetine	CYP2C19	-	2.99	**In vitro** Ki (PMID: 23785064, PMID: 19144769).	1
Voriconazole	CYP2C19	**Moderate**	0.1	**Optimised** using a DDI with Omeprazole in CYP2C19 EMs (PMID: 33236774); 10-fold reduction from in vitro Ki data reported by (PMID: 31853755).	1
Voriconazole N-oxide	CYP2C19	-	7.43	**In vitro** Ki data (PMID: 31853755).	1
Gemfibrozil	CYP2C8	**Strong**	24.1	Meta-analysis of **in vitro** 10 values.	1
Gemfibrozil 1-O-β Glucuronide	CYP2C8	-	4.88	**In vitro** HLM meta-analysis (PMID: 21911547; PMID: 16299161; PMID: 21697463).	1
Trimethoprim	CYP2C8	Weak	8.47	Lowest **in vitro** HLM value (PMID: 21697463).	1
Clopidogrel	CYP2C8	**Moderate**	23.6	**In vitro** measured in an IC50 shift experiment (PMID: 24971633). Reversible Ki was calculated as IC50/2 and corrected for fu_{inc} (Simcyp toolbox).	1
Acyl-Clopidogrel	CYP2C8	-	24.8	**In vitro** measured in an IC50 shift experiment (PMID: 24971633). Reversible Ki was calculated as IC50/2 and corrected for fu_{inc} (Simcyp toolbox).	1
Amiodarone	CYP2C9	Moderate	0.395	**In vitro** HLM data corrected for fu_{inc} (PMID: 22398967, PMID: 28679672).	1
Mono-desethyl Amiodarone	CYP2C9	-	0.126	**In vitro** HLM data corrected for fu_{inc} (PMID: 22398967, PMID: 28679672).	1
Fluconazole	CYP2C9	**Moderate**	20.4	Reported **in vivo Ki** (PMID: 12867493).	0.89~
Fluvoxamine	CYP2C9	Weak	0.126	Original value derived from meta-analysis of **in vitro** HLM data (PMID: 9384467, PMID: 10192756, PMID: 11719727). Value lowered ten-fold as ten-fold difference between in vitro and in vivo Ki for fluvoxamine reported (PMID: 12695344; PMID: 11719727).	1

Sulphaphenazole	CYP2C9	-	1.5	**Optimised** with the clinical study (PMID: 2311340).	1
Voriconazole	CYP2C9	-	2.4	fumic corrected **in vitro** Ki data (PMID: 19029318).	1
Duloxetine	CYP2C9	-	7.1	**In vitro** Ki data reported according to PMID: 31268632 by PMID: 21366359.	0.379*
Bupropion	CYP2D6	-	0.084	**Optimised** using DDI with Atomoxetine (PMID: 27518170)	1
OH-Bupropion	CYP2D6	-	0.0532	**Optimised** using DDI with Atomoxetine (PMID: 27518170)	1
Cimetidine	CYP2D6	-	3.5	**Optimised:** manual sensitivity analysis using a clinical DDI with Metoprolol as substrate (PMID: 7176462)	1
Cinacalcet	CYP2D6	**Moderate**	0.029	**In vitro** data Cyprotex 2020.	0.0206#
Fluoxetine	CYP2D6	**Strong**	0.012	**In vitro** HLM Ki and fumic (PMID: 8667235).	1
Norfluoxetine	CYP2D6	**-**	0.012	**Optimised t**o same **in vitro** value as fluoxetine (PMID: 11876575) shows that fluoxetine and Nor-fluoxetine are equal.	1
Fluvoxamine	CYP2D6	Weak	0.189	Original value derived from meta-analysis of **in vitro** HLM data (PMID: 1389951, PMID: 9681670, PMID: 9205822, PMID: 10192828, PMID: 8838442, PMID: 8477556, PMID: 8667235, PMID: 7782485, PMID: 11719727). Value lowered ten-fold as ten-fold difference between in vitro and in vivo Ki for fluvoxamine reported (PMID: 11719727).	1
Paroxetine	CYP2D6	**Strong**	0.46	**In vitro** meta-analysis of 20 HLM values (PMID: 8838442, Otton et al., CPT (1994)55:2 PI 71, PMID: 1389950, PMID: 1389951, PMID: 10917404, PMID: 10192828, PMID: 11302931, PMID: 9681670).	0.2**
Quinidine	CYP2D6	Strong	0.0119	**In vitro** meta-analysis of 9 values (PMID: 7575645, PMID: 2775613, PMID: 15039299, PMID: 10383919, PMID: 12695349).	1
Duloxetine	CYP2D6	Moderate	0.005	Optimised using (PMID: 12621382).	1
Mirabegron	CYP2D6	Moderate	2.4	**In vitro** IC50/2 is used as Ki (PMID: 31090092).	1
Amiodarone	CYP3A4	-	4.41	**In vitro** HLM data (PMID: 28679672).	1
Mono-desethyl Amiodarone	CYP3A4	-	0.511	**In vitro** HLM data (PMID: 28679672).	1
Aprepitant	CYP3A4	Moderate	1	**In vitro** - Incorporating fumic recalculated in Simcyp as ampholyte compound type (PMID: 23543602)	1
Cimetidine	CYP3A4	Weak	11	**Optimised** using a DDI with Midazolam (PMID: 10223772), PMID: 6137021 was then used to optimize the CYP3A4 Ki value to capture the DDI with triazolam and improve the predictions overall.	1
Clarithromycin	CYP3A4	**Strong**	10	**In vitro** HLM (PMID: 7895601).	0.87*
Erythromycin	CYP3A4	**Moderate**	82	**In vitro** - meta-analysis (n=5 values) (PMID: 7937537, PMID: 8450476, PMID: 9107550, PMID: 1815969).	0.909*
Fluconazole	CYP3A4	**Moderate**	10.7	**In vitro** (PMID: 9929500).	1
Fluoxetine	CYP3A4	-	31.42	**In vitro** Ki and fumic (PMID: 23785064).	1
Fluvoxamine	CYP3A4	Moderate	0.789	10-fold reduction on **in vitro** Ki obtained from a meta-analyse coming from 4 publications n=28 samples (PMID: 8846621, PMID: 8889898, PMID: 8690825, PMID: 9565774).	1
Ketoconazole	CYP3A4	Strong	0.0150	**In vitro** HLM (PMID: 9929500).	0.97*

Quinidine	CYP3A4	-	5.1	**In vitro** (PMID: 10383919, PMID: 12433827).	1
Norfluoxetine	CYP3A4		3.64	**In vitro** Ki and fumic (PMID: 23785064).	1
Clopidogrel	CYP3A4		51.4	Measured in an IC50 shift experiment (PMID: 24971633). Reversible Ki was calculated as IC50/2 and corrected for fu_{inc} (Simcyp toolbox).	1
Acyl-Clopidogrel	CYP3A4	-	165	Measured in an IC50 shift experiment (PMID: 24971633). Reversible Ki was calculated as IC50/2 and corrected for fu_{inc} (Simcyp toolbox).	1
Voriconazole	CYP3A4	Moderate	0.54	**In vitro** HLM value (PMID: 19029318).	1
Atazanavir	CYP3A4/5	-	2.35	**In vitro** value from CDER Clinical Pharmacology and Biopharmaceutics Reviews Application Number 21-567	1
Diltiazem	CYP3A4/5	Moderate	36.1	**In vitro** Ki value of 36.1 µM (PMID: 18854379).	1
Desmethyl-Diltiazem	CYP3A4/5	-	2.43	**In vitro** Ki (PMID: 18854379, PMID: 9223567, PMID: 17293381).	1
Ritonavir	CYP3A4/5	Strong	0.0019	**In vitro K**i calculation from IC50 (PMID: 28483778).	1
Fluconazole	CYP3A5	**Moderate**	84.6	**In vitro** rCYP (PMID: 9929500).	1
Fluvoxamine	CYP3A5	Moderate	5.82	10 -fold reduction on **in vitro** Ki (PMID: 11719727).	1
Ketoconazole	CYP3A5	Strong	0.1090	**In vitro** rCYP (PMID: 9929500).	0.96*

*Simcyp predicted

weighted mean of **in vitro fu mic

#A sensitivity analysis was conducted using the Dextromethorphan DDI (PMID: 17652181) to evaluate the effect on the DDI. Due to the data being below the LLOQ of the assay the fumic was increased by 10-fold.

~set equal to fup (PMID: 3004323)

Table 2a: Inhibitor information from Simcyp Simulator compound files (V19R1) showing classification, interaction parameters for mechanism-based CYP inhibition (K_{app}, fu_{mic}).

Inhibitor	Enzyme	FDA Classification	K_{app} (μM)	K_{app} source	fumic
Gemfibrozil 1-O-β Glucuronide	CYP2C8	-	27.1	Meta-analysis of in vitro HLM data (PMID: 16299161, PMID: 21697463) corrected for fumic for each study.	1
Amiodarone	CYP2C9	Moderate	0.0282	In vitro HLM data (PMID 28679672)	1
Fluoxetine	CYP2C19	Strong	2.69	In vitro (PMID 23785064)	1
Nor-Fluoxetine	CYP2C19	-	6	Manually optimised to Omeprazole DDI (PMID: 24688155, PMID: 23785064)	1
Ticlopidine	CYP2C19	Strong	0.432	In vitro (PMID: 19845434)	1
Paroxetine	CYP2D6	Strong	0.137	In vitro (PMID: 20007670, PMID: 27590024)	1
Amiodarone	CYP3A4	-	0.051	In vitro HLM data (PMID 28679672, PMID 26296708)	1
Aprepitant	CYP3A4	Moderate	2.66	In vitro value corrected for fumic back calculated from the reported fumic for an ampholyte compound type (PMID 23543602)	1
Atazanavir	CYP3A4/5	-	0.84	In vitro (CDER Clinical Pharmacology and Biopharmaceutics Reviews Application Number 21-567)	1
Clarithromycin	CYP3A4	Strong	12	In vitro (PMID: 21540107)	1
Clopidogrel	CYP3A4		53.8	Measured in vitro value (PMID 24971633) corrected for fumic (Simcyp toolbox)	1
Diltiazem	CYP3A4/5	Moderate	4.75	In vitro (PMID: 21540107)	1
Desmethyl-Diltiazem	CYP3A4/5	-	1.74	In vitro (PMID: 21540107)	1
Erythromycin	CYP3A4	Moderate	23.2	In vitro (PMID: 21540107)	1
Erythromycin	CYP3A5	Moderate	6.94	In vitro (PMID: 15377640)	0.97*
Fluoxetine	CYP3A4	-	6.9	Optimised as NOT corrected for fumic (PMID 10950845, PMID: 22490230); Verified with PMID: 14709940	1
Ritonavir	CYP3A4/5	Strong	0.18	In vitro (PMID: 21406602)	1
Verapamil	CYP3A4	Moderate	2.21	In vitro (PMID: 21540107)	1
Verapamil	CYP3A5	Moderate	3.99	In vitro (PMID: 21540107, PMID: 10640508, PMID: 17093004, PMID: 14744949)	1
Norverapamil	CYP3A4	Moderate	10.3	In vitro (PubMed 15689501)	1
Norverapamil	CYP3A5	Moderate	4.53	In vitro (PubMed 15689501)	1
Norfluoxetine	CYP3A4	-	16	Doubled from literature value as published value is for 1 enantiomer (PMID: 24688155; PMID: 23785064)	1
Acyl-Clopidogrel	CYP2C8	Moderate	6.09	Measured in vitro value (PMID 24971633) and corrected for fumic (Simcyp toolbox)	1
Mirabegron	CYP2D6	Moderate	8.1	In vitro (PMID 31090092)	1

*Simcyp predicted

Table 2b: Inhibitor information from Simcyp Simulator compound files (V19R1) showing classification, interaction parameters for mechanism-based CYP inhibition (k_{inact}).

Inhibitor	Enzyme	FDA Classification	k_{inact} 1/h)	k_{inact} source
Gemfibrozil 1-O-β Glucuronide	CYP2C8	-	6.5	In vitro HLM (PMID: 16299161; PMID: 21697463) using the average $K_{app,u}$ and multiplying that by the individual in vitro assay $k_{inact}/K_{app,u}$ ratio.
Amiodarone	CYP2C9	Moderate	0.6	In vitro (PMID 28679672)
Fluoxetine	CYP2C19	Strong	2.16	In vitro (PMID 23785064)
Nor-Fluoxetine	CYP2C19	-	3.27	In vitro (PMID: 23785064)
Ticlopidine	CYP2C19	Strong	4.434	In vitro (PMID: 19845434)
Paroxetine	CYP2D6	Strong	10.2	In vitro (PMID: 20007670, PMID: 27590024)
Amiodarone	CYP3A4	-	0.8	In vitro HLM data (PMID 28679672, PMID 26296708)
Aprepitant	CYP3A4	Moderate	6	In vitro (PMID 23543602)
Atazanavir	CYP3A4/5	-	3	In vitro (CDER Clinical Pharmacology and Biopharmaceutics Reviews Application Number 21-567)
Clarithromycin	CYP3A4	Strong	2.13	In vitro (PMID: 21540107)
Clopidogrel	CYP3A4		3.18	In vitro (PMID 24971633)
Diltiazem	CYP3A4/5	Moderate	0.702	In vitro (PMID: 21540107)
Desmethyl-Diltiazem	CYP3A4/5	-	1.09	In vitro (PMID: 21540107)
Erythromycin	CYP3A4	Moderate	2.25	In vitro (PMID: 21540107)
Erythromycin	CYP3A5	Moderate	0.66	In vitro (PMID: 15377640)
Fluoxetine	CYP3A4	-	0.66	In vitro (PMID: 10950845, PMID: 22490230, PMID: 14709940)
Ritonavir	CYP3A4/5	Strong	19.8	In vitro (PMID: 21406602)
Verapamil	CYP3A4	Moderate	2	In vitro (PMID: 21540107)
Verapamil	CYP3A5	Moderate	1.84	In vitro (PMID: 21540107, PMID: 10640508, PMID: 17093004, PMID: 14744949)
Norverapamil	CYP3A4	Moderate	18	In vitro (PubMed 15689501)
Norverapamil	CYP3A5	Moderate	4.2	In vitro (PubMed 15689501)
Norfluoxetine	CYP3A4	-	0.66	In vitro (PMID: 23785064)
Acyl-Clopidogrel	CYP2C8	Moderate	16	Optimised using a DDI between Pioglitazone and Clopidogrel (PMID 27260150). PMID 24971633, reported a value of 2.82 h-1.
Mirabegron	CYP2D6	Moderate	6.6	In vitro (PMID 31090092)

EMA Issue 6

The applicant is invited to present model diagnostics following the principles recommended by the qualification team in the scientific discussion above.

Relevant excerpt from the scientific discussion: [*The qualification team found the Bayesian approach presented by the applicant in response to issue 5 of the 1st list of issues very interesting. The QT team believe that a similar model, fitted to the DDI qualification matrix, could overcome some of the limitations of the simulation-based diagnostics. Different approaches and/or models may be suitable for the purpose of uncertainty quantification. For example, a Bayesian meta-regression model on the fold-error with parameters expressing the degree of bias and imprecision and different CYPs as random effects (and studies as nested random effects within CYPs) with the uncertainty of the literature data taken into account could be suitable for this purpose. A graphical evaluation of the posterior distribution of the random effects from this model could be helpful for the qualification team to better understand the accuracy and the associated uncertainty of the DDI predictions across different CYPs. Currently SAWP is of the opinion that further characterisation of model-based uncertainty is necessary for a qualification opinion.*]

Executive summary

To answer the EMA question, we performed a model-based analysis of the potential discrepancies between reported clinical data on DDIs and the corresponding Simcyp simulations. We used a hierarchical Bayesian meta-regression model to account for systematic differences between CYPs and the type of inhibition.

Our results are partly affected by the sometimes-poor reporting of clinical study data. Despite this, since we used 234 studies and standard modeling assumptions, we consider our analysis to be quite robust. Our conclusions are:

- GMR predictions' bias is globally of -4.5%, and ranges from 1% to -14% depending on the CYP and mode of inhibition considered. CYP2C19 GMRs (bias of -9% on average) and CYP2D6 GMRs (bias of -14% on average) are the most under-predicted.
- Simcyp V19 imprecision of GMR predictions is not directly accessible, as it is contaminated by between-study variability. We can therefore only estimate an upper bound on it. Overall, that upper bound is likely to be around 16%, and in the range of

5% to 25% after stratification by CYP or type of inhibition. Imprecision is highest for CYP2D6.

- Simcyp V19 tends to under-estimate total between subject variability (BSV) in GMRs by a factor 2.5. This result is conditioned by data quality and modeling assumptions, but can be explained by the fact that our Simcyp simulations did not consider within-subject variability nor concentration measurements' uncertainty.

Introduction

To go beyond the AFE and AAFE measures of predictive performance, we bridged, as suggested, published clinical observations of DDI effects and Simcyp predictions with a statistical model quantifying potential biases and imprecision in Simcyp DDI predictions.

The observed geometric mean ratios (GMR of AUC with inhibitor over AUC without inhibitor) and the Simcyp-predicted GMRs were bridged by a GMR bias (tendency to under- or over-predict the ratio) which was estimated (see Figure 1).

We also checked whether Simcyp-simulated between-subject variance of AUC ratios matches observed between-subject variance after taking study size and numerous other factors into account (Figure 2). The total variance of observed between-subject variance (BSV) ratios includes:

1. Between-study variance (a residual nuisance parameter that Simcyp has no way to simulate *a priori*).
2. Between-subject variance, divided by the known number of subjects in each study. Note that it stems both from within- and between-subject variability of physiological processes and that its between-subject component would be removed in cross-over design trials.

Observed and predicted variances were bridged by a between-subject variance bias (BVS bias), which we estimated.

Finally, the EMA pointed out that for a given CYP, studies using different victim drugs may differ, suggesting that a single K_i value may not be able to make perfect predictions for all victim drugs. The EMA called the corresponding component of between-study variance "imprecision". This concern is certainly valid, but with that definition, imprecision is not separately identifiable from between-study variance, and we may only access between-study variance as an upper bound to imprecision. Between-study variability was therefore also estimated.

Potential bias and imprecision of the Simcyp Simulator predictions of GMRs may depend on a number of factors including the individual CYP studied, the mechanism involved, the administration route, *etc*. Assessing bias for these different cases can be done by stratification of the data, or within a hierarchical (multilevel) framework. The former is simpler to implement because it just amounts to pre-processing the data and running several smaller analyses, but it can be limited by the small size of the stratified samples, leading to large uncertainties in the results. The latter tends to stabilize inference on smaller sub-groups of data and allows information pooling, when it makes sense; however, any hierarchical modelling is ultimately predicated on the breadth of data available for model development. Hence, we chose a mix of stratification (by inhibition mechanism and by CYP) and hierarchical modelling (for between-subject variability).

Hereafter, we describe the statistical bridging model, the methods and software used to estimate its parameters, and the results obtained. We then discuss these results and various limitations of our work.

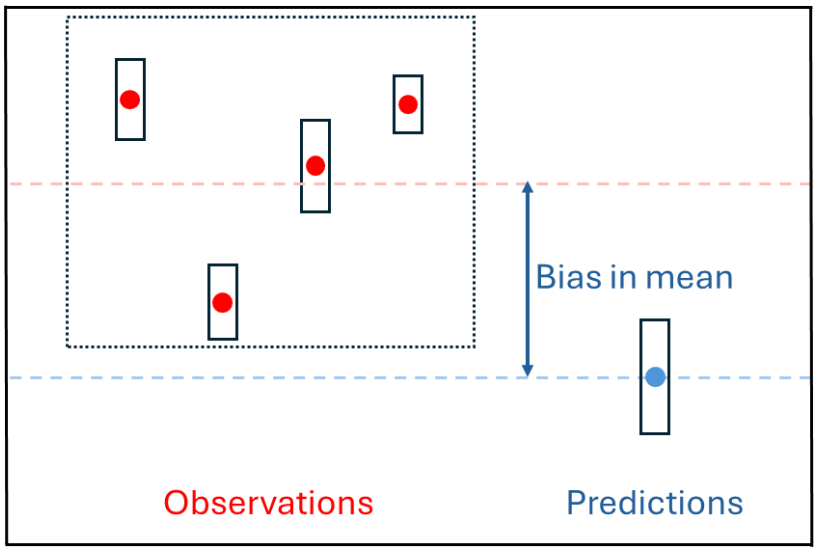


Figure 1: Estimation of difference (mean bias) between observed and predicted geometric mean ratios. Note that when there is one predicted ratio for each observation (study), match would be perfect after bias correction.

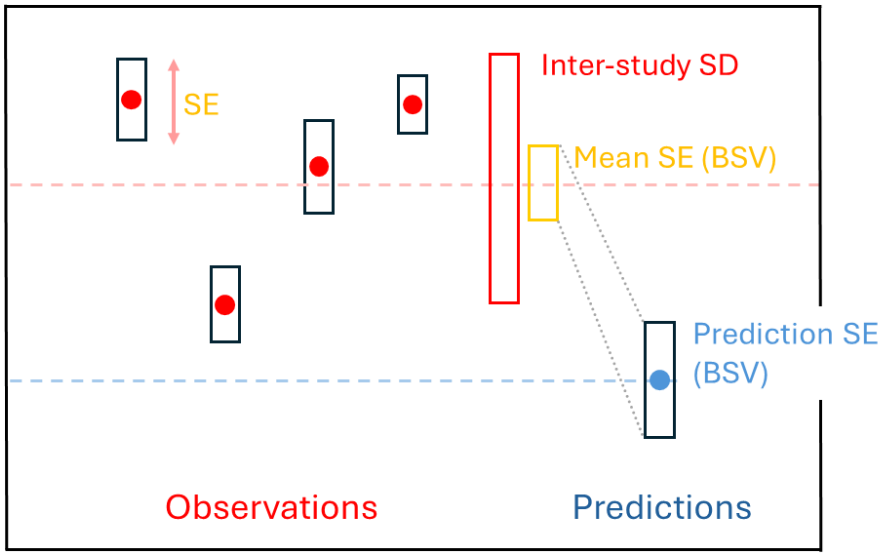


Figure 2: Checking between-subject variance (BSV) predictions and residual inter-study variability. BSV observed and predicted should match but may differ by a factor (BSV bias) we can estimate. The residual variance can also be estimated; it lumps inter-study variability and potential prediction imprecision.

Data

Two hundred and sixty-eight clinical DDI studies involving CYP1A2, 2D6, 2C8, 2C9, 2C19, and CYP3A4 were identified using the CDID (formerly UOW database, https://www.druginteractionsolutions.org/). Simulations were run using the actual clinical study design and the demographics of the individuals (n) recruited into the clinical study. Ten virtual trials of n subjects were run for each simulation (20 trials when a small number of subjects was used). Thus, two sets of data were considered in this analysis: the GMRs reported in the scientific literature, the associated population SDs, and the corresponding parameters generated as outputs by the Simcyp Simulator.

Published study results

The sources of the data have been reported in the updated analysis submitted in December, 2023. From the published reports involving six CYP enzymes, 46 victim drugs and 30 perpetrator drugs, we extracted the following information:

- The name of the victim drug and the route of administration (oral or IV).
- The name of the inhibitor drug, classification as a weak, moderate, or strong inhibitor and the route of administration (oral or IV).
- The type of inhibition: either competitive (CI) or mechanism-based inhibition (MBI).
- The total number of study subjects for which GMRs were reported in the clinical study.

- Whether the reported GMR was a ratio of the means (RoM) or the mean of individual ratios (MoR).
- One or several of the following: arithmetic mean of the ratios, geometric mean of the ratios (that is, GMR), upper and lower confidence interval bounds of the geometric mean, confidence interval coverage (90% or 95%), SD of the individual ratios, standard error of the arithmetic mean of the ratios.

Those data were preprocessed as follows (see script 1 in Appendix):

- When the standard deviation of the individual ratios was missing, we used the geometric mean of the ratios and its confidence limits to compute it. We also computed the geometric SD of the ratios and their arithmetic mean.
- We then removed the 34 studies for which the arithmetic mean of ratios or their SD were still missing at that stage.
- We computed, using arithmetic means and SDs, the geometric SDs of ratios still missing at that stage.
- Finally, we computed, using arithmetic means and geometric SDs, the GMR still missing at that stage.

There remained 234 studies for which the GMR and the geometric SD of individual ratios were available. These were used in the model, together with the numbers of subjects.

Simcyp predictions

For each of the above reported studies, there was a corresponding set of data from the Simcyp simulations: the arithmetic mean, the SD, and the geometric means of individual ratios, together with the number of simulated subjects. The Simcyp simulations were performed using a cross-over design and 10 or 20 times (when a small number of subjects was used in the clinical study) the number of subjects used in the reported studies (therefore we can consider the predictions to be much more precise than the reported values) (see the updated analysis submitted in December 2023. The simulation results were preprocessed as follows (see script 1 in the Appendix):

- We computed, using the arithmetic mean of individual ratios and their SD, the geometric SD of the predicted individual ratios.
- We checked that all the above computed geometric SDs were not null.

Statistical bias and imprecision model

We use a model that bridges the Simcyp Simulator's predictions of GMR and observed GMR with a set of free parameters characterizing biases and uncertainties. Those parameters can then be estimated from the observed GMR (taken as data) and the Simcyp Simulator predictions (taken an independent variables). That model is cast for inference in a Bayesian framework:

The GMR observed in study i (out of $N_{studies}$) is assumed to be lognormally distributed around a true GMR ($logGMR_i$, in log-space), with true standard deviation $\sigma_{GMR,i}$ (in log-space):

$$GMR_{i,obs} \sim \mathcal{LN}\left(logGMR_i, \sigma_{GMR,i}\right) \quad (1)$$

The true GMR is assumed equal to the Simcyp-predicted value, $GMR_{i,pred}$, corrected for potential bias, β_{GMR}:

$$logGMR_i = \log\left(GMR_{i,pred} \times (1 + \beta_{GMR})\right) \quad (2)$$

The true variance in log-space of $GMR_{i,obs}$, $\sigma^2_{GMR,i}$, is assumed to be equal to the sum of between-study variance, σ^2_{stu}, and a sampling variance equal to the true between-subject variance $\sigma^2_{sub,i}$, scaled by $N_{sub,i}$, the number of subject in study i:

$$\sigma^2_{GMR,i} = \sigma^2_{stu} + \frac{\sigma^2_{sub,i}}{N_{sub,i}} \quad (3)$$

Note that the between-study variance, σ^2_{stu}, may include a component due to Simcyp's imprecision in predictions, but this component is not separately estimable.

The true between-subject variance in log-space for study i is assumed to be equal to the Simcyp-predicted between-subject variance (in log-space), corrected for potential variance bias, $\beta_{BSV,i}$:

$$\sigma^2_{sub,i} = \sigma^2_{sub,i,pred} \times \beta_{BSV,i} \quad (4)$$

A hierarchical specification for $\beta_{BSV,i}$, which allows for information sharing in BSV predictions bias estimates, is used:

$$\beta_{BSV,i} \sim \mathcal{N}\left(\mu_{\beta BSV}, {\sigma_{\beta BSV}}^2\right) \quad (5)$$

Finally, the observed between-subject variance (in log-space) for study i is part of the data. According to the standard lognormal error model, it is assumed to be distributed according to the following gamma distribution, specified through shape and scale parameters k_i and θ_i respectively:

$$\sigma^2_{sub,i,obs} = \left[log\left(GSD_{i,obs}\right)\right]^2 \sim G\left(k_i, \theta_i\right) \quad (6)$$

$$k_i = N_{sub,i}/2 \quad (7)$$

$$\theta_i = \sigma^2_{sub,i}\ /k_i \quad (8)$$

The prior for the log GMR prediction bias, β_{GMR}, was set to a uniform distribution:

$$\beta_{BSV} \sim \mathcal{U}(-1, 10) \quad (9)$$

Logically, β_{GMR} cannot be lower than 1. We made sure that the upper bound was high enough to be never reached during MCMC sampling.

The priors for the parameters of the variance of prediction bias, μ_{BSV} and σ_{BSV} were respectively set to a uniform distribution (with upper bound high enough to be never reached during sampling) and to a normal distribution consistent with a degree of shrinkage:

$$\mu_{BSV} \sim \mathcal{U}(0, 50) \quad (10)$$

$$\sigma_{BSV} \sim \mathcal{N}(0{,}0.3)\ [0, 50]$$

The between-study variance, σ^2_{stu}, was assigned a vague truncated half-normal prior on the log scale:

$$\sigma^2_{stu} \sim \mathcal{N}(0, 1)\ [0, 5] \quad (11)$$

Methods and software

Inferences were performed on all the data, and stratified according to mechanism of inhibition, and according to CYP enzyme involved. A preliminary graphical analysis was performed to identify which covariate variables had a clear effect on study GMRs or GSDs (see script 2 in the Appendix). Inference itself was performed using MCMC simulations with the *R* package *Nimble* [1] (see script 1 in Appendix). Convergence of three chains of 105,000 iterations was assessed using Gelman and Rubin's $\hat{R}$ diagnostic [2].

Results

Preliminary visual analyses

The Appendix section "Preliminary covariate effects' graphical analyses", subsection "Effects on GMRs" shows plots of the Simcyp-predicted *vs*. observed *GMRs* for all data colored by:

- Type of inhibition (CI or MBI)
- CYP enzyme
- Number of subjects (low: less than 7, or high: more than 6)
- Type of GMR reported (MoR: mean of ratios, RoM: ratio of means)

- Inhibitor tested
- Strength of the inhibitor
- Substrate tested
- Route of administration of the substrate

The route of administration of the inhibitor was always oral and non-discriminant; it was therefore excluded from the analysis.

There is no clear effect of any of the above covariates on the deviations between observed and predicted GMRs.

Similar plots of Simcyp-predicted *vs*. observed between-subject *geometric SDs* are shown in the Appendix section "Preliminary covariate effects' graphical analyses", subsection "Effects on GSDs". Deviations can be quite large, but only the route of administration of the substrate seemed to have an effect on the deviations between observed and predicted geometric SDs (the intravenous route having smaller deviations). That is to be expected because oral administration is much more complex and variable.

Modeling results

Given the above results of covariate visual analyses, we decided to stratify our modeling analysis simply by type of inhibition and CYP involved.

Detailed results on MCMC convergence for the main parameters estimated (GMR bias, β_{GMR}; mean BSV bias, $\mu_{\beta BSV}$; SD of BSV bias, $\sigma_{\beta BSV}$; and between-study variance, σ^2_{stu}) are shown in Appendix section MCMC posterior plots. $\hat{R}$ convergence diagnostic values for all parameters were less than 1.02, indicating that convergence was always reached. The Appendix section of MCMC posterior summary statistics provides tables of the main parameter estimates, which are stratified. We review those results in the next paragraphs. A plot of the correlations between β_{GMR}, $\mu_{\beta BSV}$, and σ_{stu} samples is shown in the Appendix section MCMC posterior correlations between main parameters. In short, no particular correlations were observed and those estimates can be examined separately.

Biases in GMR predictions

The MCMC sampling estimates of Simcyp bias in GMR predictions are given in Table 1. Using all the data, without stratification, the mean of the bias parameter β_{GMR} was -0.045, with a 95% credibility interval (from the 2.5 percentile to the 97.5 percentile) ranging from -0.080 to -0.011. These (through equation 2) correspond to a small average Simcyp underprediction by a factor 0.955 (credibility interval from 0.92 to 0.989).

Models fitted to stratified subsets of data showed some variation in results. Figure 3 gives an overall view of those GMR bias estimates. The most significant biases are observed for CYP2C19 (by -9% on average) and CYP2D6 (by -14% on average). Predictions for MBI and CYP3A4 have the lowest biases (-0.6% and +1%, respectively). The uncertainty associated with bias estimates is driven primarily by the limited number of data points in each stratum and their own uncertainty. The small effect of bias correction on plots of observed *vs*. predicted GMRs can be examined per stratum in Appendix section MCMC posterior observed vs. predicted geometric mean ratios.

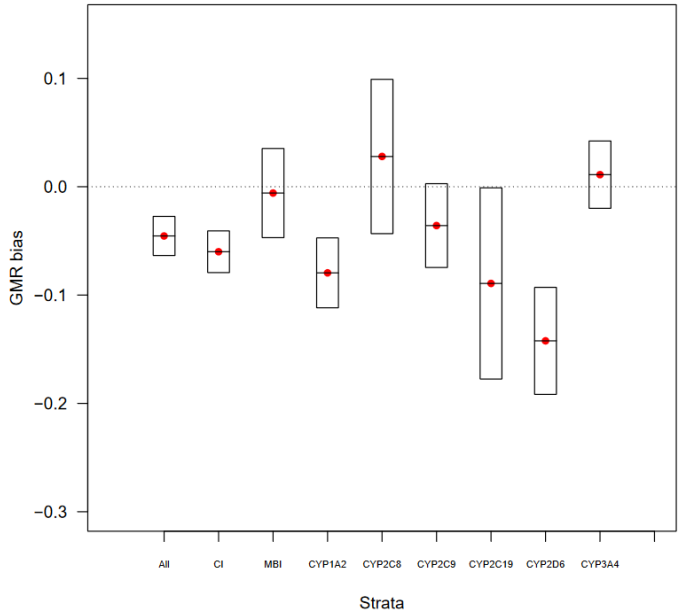


Figure 3: GMR biases stratified. Boxes show mean +/- SD. For each stratum, predicted GMRs have to be multiplied by 1+GMR bias to best match the observed GMRs.

Table 1: Statistical summaries of the posterior distributions of GMR bias in predictions.

Stratification	Mean	SD	2.5th pctile	50th pctile	97.5th pctile
All	-0.0454	0.0181	-0.0804	-0.0459	-0.0105
CI	-0.0600	0.0193	-0.0962	-0.0595	-0.0225
MBI	-0.00583	0.0412	-0.0828	-0.0068	0.0790
CYP1A2	-0.0795	0.0323	-0.141	-0.0804	-0.0138
CYP2C8	0.0279	0.0712	-0.0957	0.0217	0.190
CYP2C9	-0.0359	0.0387	-0.113	-0.0370	0.0415
CYP2C19	-0.0893	0.0883	-0.252	-0.0929	0.0967
CYP2D6	-0.142	0.0494	-0.239	-0.143	-0.0393
CYP3A4	0.0111	0.0311	-0.0501	0.0113	0.0721

Simcyp imprecision

After GM bias correction, the stratified values of residual between-study CVs of GMRs can be seen in Figure 4. Those CVs are upper bounds of Simcyp predictions' imprecision because they lump imprecision and actual (residual) between-study variability. We present CVs in natural space, rather than variances on the log-scale for ease of interpretation; CVs were obtained by the transformation $\sqrt{\exp(\sigma_{stu}^2) - 1}$. Their central estimates span values ranging from 13% to 31% with some uncertainty due to the limited database of studies. The stratified CV estimates are detailed in Table 2.

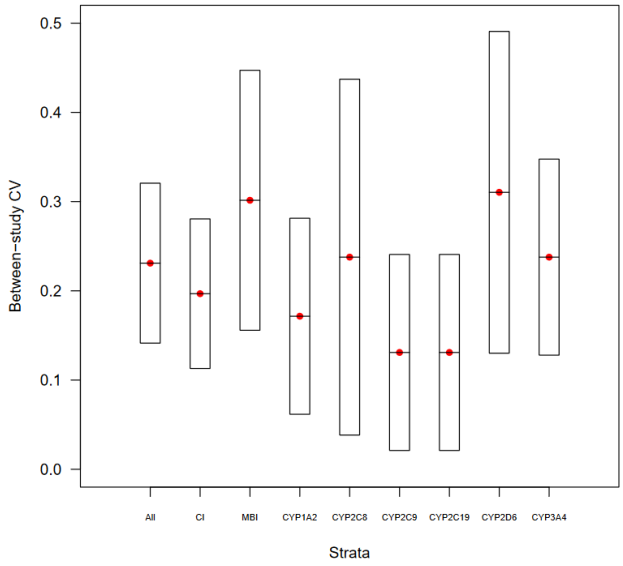


Figure 4: Between-study CVs of GMRs stratified. Boxes show mean +/- SD. Those CVs are an upper bound of Simcyp predictions' imprecision because they lump imprecision and actual between-study variability.

Table 2: Statistical summaries of the posterior distributions of the between-study CVs. Those are upper bounds on Simcyp predictions' imprecision.

Stratification	Mean	SD	2.5th pctile	50th pctile	97.5th pctile
All	0.231	0.090	0.197	0.229	0.265
CI	0.197	0.084	0.159	0.194	0.236
MBI	0.301	0.146	0.233	0.296	0.380
CYP1A2	0.172	0.110	0.114	0.165	0.244
CYP2C8	0.238	0.199	0.084	0.219	0.407
CYP2C9	0.131	0.110	0.063	0.123	0.219
CYP2C19	0.131	0.110	0.063	0.123	0.219
CYP2D6	0.310	0.180	0.215	0.301	0.429
CYP3A4	0.238	0.110	0.189	0.233	0.290

Bias in between-subject variability predictions

The discrepancies (biases) between Simcyp predictions of between-subject CVs of individual ratios and the observed CVs for all data are show in Figure 5. CVs are easier to interpret and were obtained from the model-estimated variances in log by the transformation $\sqrt{\exp(variance) - 1}$. The Simcyp predictions appear to underestimate true between-subject CVs by a factor 2.5 on average. This estimate is more pessimistic than a simple comparison of average observed between-subject GSD (equal to 1.7) to the average Simcyp predicted between-subject GSD (equal to 1.3) would suggest, as those two look quite similar. A similar Figure for the untransformed variances is shown in Appendix section MCMC posterior between-subject variance estimates. Note that the bias on the variances in log appears inflated because of the nonlinear relationship between variance in log and CVs in natural space. The bias on CV appears to be roughly equal to the square root of the bias on variance. This all goes to show that appraisal and comparisons of variance biases are difficult because they are scale-dependent.

The stratified mean between-subject *variance* bias estimates are detailed in Table 3, while the corresponding SDs (of biases, measuring their uncertainty) are given in Table 4. Uncertainty is large, because the SDs are only slightly lower than the means. Those biases differ and are probably conditioned by the victim drug route of administration, as we saw in the preliminary visual analyses presented above.

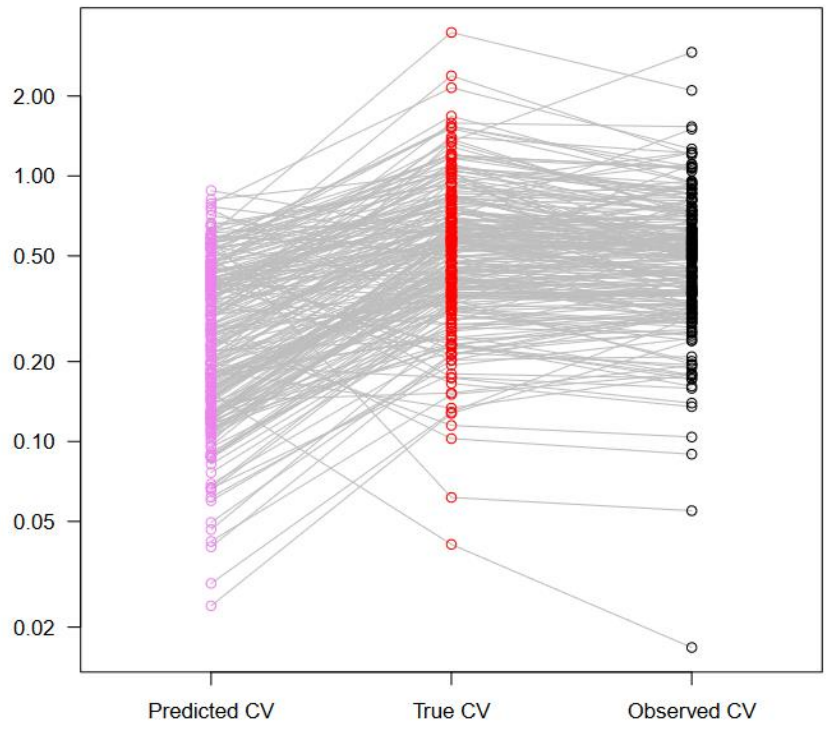


Figure 5: Simcyp-predicted, true (bias corrected) and observed between-subject CVs for the analysis including *all the data*. Bias corresponds on average to a factor 2.5.

Table 3: Statistical summaries of the posterior distributions of the average between-subject bias.

Stratification	Mean	SD	2.5th pctile	50th pctile	97.5th pctile
All	7.58	0.532	6.57	7.56	8.66
CI	8.94	0.685	7.65	8.91	10.3
MBI	4.67	0.629	3.46	4.67	5.93
CYP1A2	5.92	0.813	4.41	5.88	7.66
CYP2C8	5.53	1.21	3.36	5.47	8.08
CYP2C9	6.69	1.09	4.72	6.60	8.95
CYP2C19	2.19	0.364	1.55	2.17	2.95
CYP2D6	7.58	1.27	5.24	7.53	10.3
CYP3A4	9.71	0.959	7.9	9.69	11.7

Table 4: Statistical summaries of the posterior distributions of standard deviation of BSV bias.

Stratification	Mean	SD	2.5th pctile	50th pctile	97.5th pctile
All	6.97	0.450	6.14	6.95	7.89
CI	7.34	0.551	6.32	7.33	8.51
MBI	4.82	0.557	3.85	4.78	6.03
CYP1A2	3.99	0.614	2.92	3.95	5.35
CYP2C8	4.32	0.813	2.91	4.28	6.02
CYP2C9	3.72	0.723	2.51	3.66	5.31
CYP2C19	1.28	0.306	0.775	1.24	1.96
CYP2D6	7.17	0.824	5.70	7.12	8.83
CYP3A4	8.16	0.698	6.88	8.14	9.61

Discussion

The data

Data presented in this analysis consist of those reported in the scientific literature, and corresponding Simcyp Simulator predictions. In the former case, there are often limitations associated with the treatment and reporting of the data as discussed below.

Published study results

We feel it important to acknowledge up front that the reporting and published statistical treatment of the data in the clinical studies was in some cases quite poor. Indeed, it was not always clear whether arithmetic or geometric means had been reported. Furthermore, minimum and maximum values were sometimes the only measure of dispersion, and the basic standard assumption of lognormally distributed values was often disregarded.

Summary results were typically reported, frequently in the form of an arithmetic mean and SD of the AUC both with and without inhibition. Arithmetic moments are not stable and are indirect summaries for lognormally distributed data; the less data there are, the more that is true. Therefore, it was often necessary to estimate the GMR and geometric populations SDs using modelling assumptions, especially when the ratio of mean values was reported instead of the mean of the individual ratios. In that case, we used a normal approximation [3] to obtain the SD of the ratios of two arithmetic means, m_1 and m_2, given the corresponding SDs, s_1 and s_2:

$$SD_{R,2,1} = \frac{m_2}{m_1}\sqrt{\left(\frac{s_1}{m_1}\right)^2 + \left(\frac{s_2}{m_2}\right)^2} \quad (12)$$

The 'observed' data we used is therefore subject to its own, unquantified uncertainty and/or bias. This manifests in our results as extra and unwanted variability about our bias and uncertainty estimates. A small subset (38/263, ~14%) of the published crossover studies reported individual level data (either as tabulated values or, worse, graphical displays) on the AUC with and without inhibition – in further work, more detailed analysis of these measurements may allow us to improve our estimates and partially verify some model assumptions.

Simcyp predictions

Simcyp reporting of simulation is much more standardized. There were no parallel design studies in the database, and the Simcyp predictions represented simulations of cross-over studies with the same number of subjects used in the actual clinical study. Note also that Simcyp folds (at the level physiological parameters) within-subject variability and between-subject variability and that the two cannot be cleanly disentangled in DDI cross-over trial simulations. Analytical measurement noise (which would be typically conflated in data with within-subject variability) was not simulated. There were also no attempts to reproduce the published (even if incorrect) data treatment (*e.g.*, forming the minimum and maximum of the AUCs) because that would be of little use for modeling. All this at least partly explains why the Simcyp estimates of between-subject variability under-predict the observed variabilities by a factor 2.5 on average. However, this does not affect biases, as shown by the lack of correlation between biases and uncertainty estimates in our analyses.

Our statistical model

The model we used is of a special kind and bridges Simcyp and actual clinical data, in order to evaluate biases and imprecisions of predictions for different DDI scenarios. Such models are not completely new and an early reference in the statistical literature discusses "discrepancies" to bridge computer models and data [4]. The issue of bias has also been dealt with in the epidemiology literature [5,6] and later in pharmacometrics [7,8] although the focus has been on correcting for measurement bias as a consequence of measurement errors, confounding, non-representative samples *etc*. using a probabilistic approach which is implicitly Bayesian although not within a formal Bayesian framework. Since the seminal work of Kennedy and O'Hagan [9] there has been a growing literature on bias correction for deterministic computer models. Whilst in that work the model for bias correction is somewhat more complex than the one we considered, the conceptual approach and Bayesian workflow are certainly similar. In

any case, that's an interesting statistical question and could lead to useful tools for model evaluation in general.

Our model is a hierarchical Bayesian meta-regression model on the fold-error with a nested random effect structure to account for systematic differences between CYPs and type of inhibition. In our exploratory work, the effects of CYP, dosing route (intravenous, oral), mechanism (competitive inhibition, mechanism-based inhibition), substrate, inhibitor, and inhibitor type (weak, moderate, strong) were studied. One aspect of the model is that its dependent variables are summary statistics rather than primary measurements, which is a consequence of our reliance on published studies with incomplete data reporting. A further consequence is that the assumed lognormality of AUC ratios cannot be verified through posterior checks. The statistical model was fitted to data subsets obtained by stratification on individual CYPs and by mechanism of inhibition, to assess for systematic differences in results for these data sub-groups. A fully hierarchical model that would simultaneously adjusted for covariates would be preferable; however, the paucity of available data would probably have led to parameter identifiability problems. The analysis based on the full dataset, on subsets, and robust post-fitting checks was preferred. Retrospectively, it is unclear whether the introduction of a hierarchy for between-subject varance bias was much needed. The effect of shrinkage (regression toward the mean) is modest (see Figure 5) and forced us to introduce an *ad hoc* prior, which conflicts somewhat with the data likelihood. A fixed-effect analysis for between-subject variance bias might have been simpler. In any case, we think that the results would not differ markedly.

A more technical point is that all for the studies reporting ratios of means, we considered the variance of the ratios to be gamma-distributed likelihood (Eq. 6) like the others, because of the intractable conditional distribution of the variances and SDs computed with the normal approximation we used [3]. This could be further researched and improved upon.

Methods used

Inference was performed using MCMC simulations in a very standard setting with the *R* package *Nimble*. The model is quite compact and can be run in approximately one minute on an ordinary laptop computer, so that the results can be easily replicated.

Results obtained

Biases in GMR predictions

We found overall a very modest bias (4.5% underprediction) in Simcyp predictions of DDI GMRs. The most significant biases were observed for CYP2C19 (by -9% on average) and CYP2D6 (by -14% on average). The lowest biases were for predictions of MBI and for CYP3A4 (-0.6% and +1%, respectively). Since the input data are far from perfect (see discussion above), it would be interesting to see whether database inprovement would affect those estimates, but they seem quite robust to variants of the model (for example, the introduction of a hierarchical structure on between-subject variability did not affect them). This can also be deduced from the lack of posterior correlations between parameter estimates (as Figure 31 shows).

Simcyp imprecision

Simcyp imprecision, linked to residual variance after GMR bias adjustment, cannot be separated cleanly from between-study variability. We would need a separate estimate of between-study variability to constrain it. It might be possible to obtain it in further analyses and it could then be introduced as a prior. For now, the total between-study variance after mean GMR bias correction can be used as an upper bound on Simcyp imprecision.

That being said, we found upper bounds on precision ranging from 13% to 31%, depending on the CYP or type of inhibition considered. The unstratified estimate upper bounds on precision is 23%. We believe that true between-study variability accounts for a good part of the total, as it must be at least higher than the variability implied by the SE on the measured GMRs. That SE, which can be estimated empirically from the reported data, corresponds to 7% variability on average. If we subtract this lower bound estimate of true between-study variability from the total variability, we get a refined overall upper bound on imprecision equal to 16%. This could be refined but it would seem that Simcyp imprecision is in the range of 5% to 25% after stratification.

Bias in between-subject variability predictions

The headline results suggest an overall factor 2.5 underestimation by Simcyp of between subject variability, with considerable variation in the magnitude of this bias over the different CYPs and types of inhibition. However, the reported between-subject SDs we used are subject

to large uncertainty. The Simcyp predictions could also be improved by including variability stemming from concentration measurement errors. Also, for most studies, it was necessary to calculate the standard deviation of the AUC ratio through Eq. 12, with the GSD subsequently estimated based on an assumption that a log-normal distribution for between subject variability in the AUC ratio was reasonable. The model linking observations and Simcyp predictions was for variances and specified on the log-scale through Eq. 6. The propagation and amplification of error through various calculations is likely to be a significant factor in the large between study variability, quantified through σ_{BSV}. The parameters and μ_{BSV} and σ_{BSV} are thus (potentially very conservative) upper bounds on Simcyp bias in BSV. This could be examined and maybe improved in further analyses.

Although only a small subset of the studies contained individual level data, these would allow direct estimates of measures of between-subject variability (arithmetic and geometric standard deviations) and provide a small but more precise dataset for a) evaluating Simcyp estimates of between subject variability; and b) assessing the magnitude of error associated with estimates and transformations. Further work in this area would be useful.

Genotype effects

From the 263 clinical studies in the final dataset, only 38 reported inter-individual data. Not surprisingly these studies were predominantly for those enzymes known to carry functional genotypes (i.e. CYP2C9 (18% of the studies), CYP2C19 (18% of the studies), and CYP2D6 (25% of the studies)). However, even when the individual PK data with and without inhibition were reported, the relevant genotyping or phenotyping was not necessarily performed.

In the Simcyp Simulator V19, population frequencies and activities for ultra-rapid metabolizers (UM), extensive metabolizers (EM), intermediate metabolizers (IM), and poor metabolizers (PM) phenotypes can be accounted for in a simulation. System (population) data for CYP2C9 (EM, PM), CYP2C19 (EM, PM, UM), and CYP2D6 (EM, PM, UM) are supplied in the population libraries, and population differences are considered. In addition, the CYP2C9 genotypes (for *1/*1, *1/*2, *2/*2, *2/*3, and *3/*3) can also be simulated using population data within the Simcyp population libraries. For CYP1A2 and CYP3A4 all individuals are simulated as EMs, and population variability associated with these enzymes are propagated from the hepatic abundance and its variability (mean (CV)) of 52 pmol/mg microsomal protein (67%) and 137 pmol/mg microsomal protein (41%), respectively. Since CYP3A4 is also expressed in the intestine, the corresponding intestinal abundance (65.4 nmol/small intestine, 32%) is used.

As new system data become regularly available in the literature, we aim to update our population databases as soon as sufficient evidence is available. Hence in 2020/2021 (Simcyp V21), the phenotypes and genotypes for CYPs were evaluated for such data. For instance, the CYP2C19*17 allele had been identified in 2006. In our database 20 of the 28 clinical studies for CYP2C19 were published in papers before 2007. The CYP2C19*17 allele is associated with an increase in gene transcription and activity. It defines therefore a Rapid Metabolizer (RM) or Ultrarapid Metabolizer status. Thus, in genotype data specified before 2006, an over-estimation of EM frequency can be assumed. In combination with emerging LCMSMS data, the new evidence indicated that 5 phenotypes were required to address the population variability for CYP2C19, hence a RM phenotype was introduced in V21. A performance verification of the updated CYP2C19 phenotypes has been published using the Simcyp Simulator V21 [10]. As this functionality of a 5th phenotype is not available in V19, this information was not used for the submitted simulation. Nevertheless, the fact that a small negative bias for the GMR prediction for CYP2C19 of -0.09 (Figure 3, Table 1) was observed might be attributed to the above highlighted missing information from the clinical data, which leads to a less then possibly needed match in the simulated virtual population to recover the mean and variability for the CYP2C19 dataset (Figure 38).

Although the functionality of an IM phenotype was available in the Simcyp Simulator since 2006, until Version 21 no IM had been assigned for any CYP enzyme. For CYP2D6, the analysis from 2020/2021 indicated that more than 30% of Caucasians are in fact IMs with less than 33% of the protein abundance of the wild type (CYP2D6*1*1). Also, this information was added to the Simcyp Simulator in V21 and not considered in the current platform qualification for V19. While the V19 CYP2D6 frequency inputs were based on a meta-analysis solely of phenotype data, the V21 analysis was only possible since two large databases were available that linked the genotypes and phenotypes via an activity scoring system [11,12]. Hence for the CYP2D6 the GMR bias (Figures 3 and 39) as well as the large BSV bias (Table 4) has two sources that lead to the high imprecision for CYP2D6: The clinical data and the system data. Several clinical reports had not defined the genotype or phenotype and were underpowered to assess a phenotype specific population variability; eleven of the 44 CYP2D6 studies reported data for less than seven subjects for this highly variable enzyme. While the system data in V19 combined EMs and IMs, they were phenotype driven and clinical studies that reported CYP2D6 genotype data or even only allelic information needed to be translated to corresponding phenotypes.

Conclusions

Our conclusions are that

- GMR predictions' bias is globally of -4.5%, and ranges from 1% to -14% depending on the CYP and mode of inhibition considered. CYP2C19 GMRs (bias of -9% on average) and CYP2D6 GMRs (bias of -14% on average) are the most under-predicted. Those estimates should be quite robust to our input data and modelling assumptions.
- Simcyp imprecision of GMR predictions is not directly accessible, as it is contaminated by between-study variability. In the absence of better data, we can therefore only access an upper bound on it and, overall, that upper bound should be around 16%, and in the range of 5% to 25% after stratification by CYP or type of inhibition. Imprecision is highest for CYP2D6.
- According to the model, Simcyp tends to under-estimate between subject variability by a factor 2.5. However, the average *observed* between-subject GSD is 1.7, and the average Simcyp *predicted* between-subject GSD is 1.3. So, judging by the raw data, underestimation by Simcyp seems quite modest. This is an area where the model-based assessment seems strongly affected by data quality and modeling assumption. Both data and model should be improved to better estimate the actual magnitude of this bias. In any case, between-subject variability after oral administration seems to be more under-estimated than after intravenous administration.

References

1. NIMBLE Development Team. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling, https://cran.r-project.org, https://r-nimble.org. Zenodo, doi: 10.5281/zenodo.1211190; 2022.

2. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science. 1992;7:457–511.

3. Díaz-Francés E, Rubio FJ. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. Statistical Papers. 2013;54:309–23.

4. Strong M, Oakley JE. When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. SIAM/ASA Journal on Uncertainty Quantification. 2014;2:106–25.

5. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. International Journal of Epidemiology. 2014;43:1969–85.

6. Lash TL, Fox MP, Cooney D, Lu Y, Forshee RA. Quantitative bias analysis in regulatory settings. American Journal of Public Health. 2016;106:1227–30.

7. Ibrahim MMA, Nordgren R, Kjellsson MC, Karlsson MO. Model-based residual post-processing for residual model identification. The AAPS Journal. 2018;20:81.

8. Ibrahim MMA, Ueckert S, Freiberga S, Kjellsson MC, Karlsson MO. Model-based conditional weighted residuals analysis for structural model assessment. The AAPS Journal. 2019;21:34.

9. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2001;63:425–64.

10. Sychterz C, Gardner I, Chiang M, Rachumallu R, Neuhoff S, Perera V, et al. Performance verification of CYP2C19 enzyme abundance polymorphism settings within the Simcyp Simulator v21. Metabolites. 2022;12:1001.

11. Jukić MM, Smith RL, Molden E, Ingelman-Sundberg M. Evaluation of the CYP2D6 haplotype activity scores based on metabolic ratios of 4,700 patients treated with three different CYP2D6 substrates. Clinical Pharmacology and Therapeutics. 2021;110:750–8.

12. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. Genetics in Medicine. 2017;19:69–76.

Appendix

Simulation script 1 (data preprocessing and model v7)

```
## Simcyp DDI prediction validation exercise
## Estimate Simcyp predictions bias and uncertainty using predictions and
## observations.
## v1: Extend and adapt Pieter J. Colin model.
## v2: Assess also differences in between-subject variability.
## v3: Use an updated database; fix an error in total_var_of_ratio calculations
##     (true_BSV_of_ratio was not divided by number of subjects).
## v4: Use data version 2.
## v5: Use data version 3; new data has added rows and columns.
## v6: Use data version 5 (version 4 is partial and skipped, version 5 includes
##     small corrections made in version 6).
## v6b: Replace scaled chi-square distribution by the equivalent gamma
##      distribution. Use data version 6 (identical to version 5).
## v7:  Extend v6b to have one BSV_bias per study.
##=============================================================================

## Version tag
vtag = "_v7"

## Option flag to save results and plots
bsave = TRUE

## read the data
## =============
DDI.data = read.csv("./RawData/Full DB 06.csv") # check the rounding of values!
str(DDI.data)
dim(DDI.data)
head(DDI.data)
```

```
## process the data to complete missing information
## ================================================
## Add a column of observation geometric SD
Obs_GSD = rep(NA, dim(DDI.data)[1])
DDI.data = cbind(DDI.data, Obs_GSD)
head(DDI.data)

## Checking: number of simulated subjects is 10 or 20 times trial size...
plot(DDI.data$Obs_Nsub, DDI.data$Pred_Nsub)

## For the lines which have missing observation arithmetic SD, use the
## observed geometric mean and CI limits to compute it. Compute also GSD and
## arithmetic mean:
to.process = which(is.na(DDI.data$Obs_SD))
for (i in to.process) {
  if (!(is.na(DDI.data$Obs_GeoM[i]) |
        is.na(DDI.data$Obs_CI_lower[i]) |
        is.na(DDI.data$Obs_CI_upper[i]))) { # if we have what we need...
    log.mean = log(DDI.data$Obs_GeoM[i])
    log.SD   = sqrt(DDI.data$Obs_Nsubjects[i]) *
               (log(DDI.data$Obs_CI_upper[i]) -
                log(DDI.data$Obs_CI_lower[i])) /
                (2 * qnorm(1 - (1 - DDI.data$CI_coverage[i]/100)/2))
    DDI.data$Obs_GSD[i] = exp(log.SD)
    DDI.data$Obs_SD[i]  = sqrt((exp(log.SD^2) - 1) *
                               exp(2 * log.mean + log.SD^2))
    DDI.data$Obs_Mean[i]= exp(log.mean + (log.SD^2) / 2)
  }
}
## Check
par(mfrow = c(2,2))
mycol = rep("black", dim(DDI.data)[1])
mycol[to.process] = "red"
plot(DDI.data$Obs_SD,   las=1, log="y", col=mycol)
plot(DDI.data$Obs_Mean, las=1, log="y", col=mycol)
plot(DDI.data$Obs_SD/DDI.data$Obs_Mean, las=1, log="y", col=mycol,
     ylab="Obs CV")
plot(DDI.data$Obs_GSD, las=1, log="y", col=mycol)

## weed out lines without arithmetic mean or SD at this stage
to.remove = which(is.na(DDI.data$Obs_Mean) | is.na(DDI.data$Obs_SD))
DDI.data = DDI.data[-to.remove,] # all observed arithmetic SD available

## We need all observed geometric SDs; if missing, compute them
## from arithmetic means and SDs:
to.process = which(is.na(DDI.data$Obs_GSD))
for (i in to.process) {
  DDI.data$Obs_GSD[i] = exp(sqrt(log(DDI.data$Obs_SD[i]^2 /
                                 DDI.data$Obs_Mean[i]^2 + 1)))
}
## Check
par(mfrow = c(1,1))
mycol = rep("black", dim(DDI.data)[1])
mycol[to.process] = "red"
plot(DDI.data$Obs_GSD, las=1, log="y", col=mycol)

## We need all observed geometric means; if missing, compute them
## from arithmetic means and geometric SDs:
to.process = which(is.na(DDI.data$Obs_GeoM))
for (i in to.process) {
  DDI.data$Obs_GeoMean[i] = exp(log(DDI.data$Obs_Mean[i]) -
                                log(DDI.data$Obs_GSD[i])^2 / 2)
}
## Check
mycol = rep("black", dim(DDI.data)[1])
mycol[to.process] = "red"
plot(DDI.data$Obs_GeoM, las=1, log="y", col=mycol)
```

```
## We need prediction geometric SDs; compute them from
## arithmetic means and SDs:
Pred_GSD = exp(sqrt(log(DDI.data$Pred_SD^2 / DDI.data$Pred_Mean^2 + 1)))
DDI.data = cbind(DDI.data, Pred_GSD)
## Check
par(mfrow = c(1,2))
## plot(DDI.data$Obs_SD,  DDI.data$Pred_SD,  las=1, log="")
plot(DDI.data$Obs_GeoMean, DDI.data$Pred_GeoMean, las=1, log="xy")
abline(0,1)
plot(DDI.data$Obs_GSD, DDI.data$Pred_GSD, las=1, log="") # clearly two groups
abline(0,1)
## Remove the studies with null prediction variance
to.remove = which(DDI.data$Pred_GSD == 1 )
if (length(to.remove) != 0) DDI.data = DDI.data[-to.remove,]

## Save the database at this stage
## write.csv(DDI.data, file="Processed full DB 06.csv", row.names=F)
DDI.data = read.csv(file="Processed full DB 06.csv")

## How many studies do we have for each CYP?
table(DDI.data$CYP)
table(DDI.data$CYP[which(DDI.data$CI_MBI == "CI")])
table(DDI.data$CYP[which(DDI.data$CI_MBI != "CI")])


## Modeling and inference
## ======================

## stratify the data analysis, eventually
## --------------------------------------
my.stratum = "All"     ## All,
                       ## CI, MBI,
                       ## CYP1A2, CYP2C19, CYP2C8, CYP2C9, CYP2D6, CYP3A4
if (my.stratum != "All") { # stratify
  DDI.data = read.csv(file="Processed full DB 06.csv") # re-read
  if ((my.stratum == "CI") | (my.stratum == "MBI")) {
    to.remove = which(DDI.data$CI_MBI != my.stratum)
  } else {
    to.remove = which(DDI.data$CYP != my.stratum)
  }
  DDI.data = DDI.data[-to.remove,]
}
dim(DDI.data)

## Version tag revised to indicate stratification
vtag2 = paste0(vtag, "_", my.stratum)

## prepare for MCMC
## ----------------
library(coda)
library(MCMCvis)
library(corrplot)
library(nimble)

## Hierarchical core Nimble (BUGS) code
## ------------------------------------
myNimbleCode = nimbleCode({ ## BUGS (extended) code
  ##
  ## mean bias: difference between Simcyp predicted mean ratios and
  ## corresponding observed mean ratios
  mean_bias ~ dunif(-1, 10) # vague prior
  ##
  ## BSV bias: ratio of true between-subject variance over Simcyp-predicted
  ## between-subject variance. In v7, add hierarchy with vague priors
  mean_BSV_bias ~ dunif(0, 50)
  sd_BSV_bias   ~ T(dnorm(0, 0.3), 0, 50) # favor some shrinkage
  ##
  ## If each study was corrected with a specific mean bias, observations
```

```
  ## would just need to have the corrected between-subject variance.
  ## However, if mean bias is aggregated across studies one way or another,
  ## there would potentially be an extra-variance, in log, in the observed
  ## ratios, unless the mean biases were all the same for all
  ## studies (unlikely). The extra-variance aggregates inter-study
  ## variability and potential Simcyp imprecision.
  ## Prior of the extra-variance:
  extra_var_of_ratio ~ T(dnorm(0, 1), 0, 5) # vague prior on log scale
  ##
  for (i in 1:N_studies) {
    ## Simcyp predictions may be biased:
    log_corrected_ratio[i] <- log(simcyp_ratio[i] * (1 + mean_bias))
    ## Simcyp between-subject variance (in log) may also be biased:
    BSV_bias[i] ~ dnorm(mean=mean_BSV_bias, sd=sd_BSV_bias)
    true_BSV_of_ratio[i] <- BSV_simcyp_pred[i] * BSV_bias[i]
    ## The observed between subject variances in log are data:
    shape[i] <- N_subjects[i] / 2
    scale[i] <- true_BSV_of_ratio[i] / shape[i]
    obs_BSV_of_ratio[i] ~ dgamma(shape= shape[i], scale=scale[i])
    ## The total variance of the observed ratios, in log, is:
    total_var_of_ratio[i] <- true_BSV_of_ratio[i] / N_subjects[i] +
                             extra_var_of_ratio
    total_SD_of_ratio[i]  <- sqrt(total_var_of_ratio[i])
    ## Observed ratios are data:
    obs_ratio[i] ~ dlnorm(meanlog=log_corrected_ratio[i],
                          sdlog=total_SD_of_ratio[i])
  }
}) # End myNimbleCode

## Setup constants, inits and data
## -------------------------------
N_studies = dim(DDI.data)[1] # number of studies

constants = list(N_studies       = N_studies,
                 N_subjects      = DDI.data$Obs_Nsub,
                 simcyp_ratio    = DDI.data$Pred_GeoM,       # in natural space
                 BSV_simcyp_pred = log(DDI.data$Pred_GSD)^2) # in log space

data  = list(obs_ratio        = DDI.data$Obs_GeoM,       # in natural space
             obs_BSV_of_ratio = log(DDI.data$Obs_GSD)^2) # in log space

inits = list()

## Put the model together
## ----------------------
Rmodel = nimbleModel(myNimbleCode, constants, data, inits, calculate=F)

## Run the model to extract node anmes and check it
## ------------------------------------------------
## set.seed(1)
Node.names = Rmodel$getNodeNames(includeData=T)
## model$getNodeNames(stochOnly = T)
Rmodel$simulate(nodes = Node.names)
res = values(Rmodel, Node.names)
names(res) = Node.names
res

## Configure MCMC
## --------------
conf = configureMCMC(Rmodel, thin=1, # useConjugacy = FALSE,
                     monitors=c("mean_bias",
                                "mean_BSV_bias", "sd_BSV_bias", "BSV_bias",
                                "extra_var_of_ratio",
                                "log_corrected_ratio",
                                "true_BSV_of_ratio",
                                "total_SD_of_ratio",
                                ## "logProb_mean_bias",     # uniform prior
                                ## "logProb_BSV_bias",      # idem
```

```
                                  ## "logProb_mean_BSV_bias", # idem
                                  "logProb_sd_BSV_bias",
                                  "logProb_BSV_bias",
                                  "logProb_extra_var_of_ratio",
                                  "logProb_obs_BSV_of_ratio",
                                  "logProb_obs_ratio"))

Rmcmc  = buildMCMC(conf)

## Compile for speed
Cmodel = compileNimble(Rmodel, showCompilerOutput=F)
Cmcmc  = compileNimble(Rmcmc, project=Rmodel)

## Run MCMC sampling, takes seconds
## --------------------------------
N.iter = 105000
mysamples = runMCMC(Cmcmc, niter=N.iter, nburnin=5000,
                    nchains=3, setSeed=1, thin=100)

## Compute posterior summary statistics
## ------------------------------------
mcmc.sum = MCMCsummary(object = mysamples, round = 5)
mcmc.sum
range(mcmc.sum$Rhat)

## If converged, proceed...
## ------------------------

## Concatenate all chains
## ----------------------
dim(mysamples$chain1)
dim(mysamples$chain2)
dim(mysamples$chain3)
final.sample = rbind(mysamples$chain1, mysamples$chain2, mysamples$chain3)
dim(final.sample)

## Extract maximum posterior estimates
## -----------------------------------
myindex = grep("logProb_", colnames(final.sample))
logpost = apply(X=final.sample[,myindex], MAR=1, FUN=sum)
plot(logpost)
maxpost = max(logpost)
plot(exp(logpost - maxpost))
my.i = grep("mean_bias", rownames(mcmc.sum))
plot(final.sample[,my.i], exp(logpost - maxpost)) # nice
MAP.estimates = final.sample[which(logpost == maxpost),]

## Add maximum posterior estimates to the mcmc summary table and save it
## ----------------------------------------------------------------------
mcmc.sum = cbind(mcmc.sum, MAP.estimates)
if (bsave) {
  fname = paste0("mcmc.summary", vtag2, ".txt")
  write.table(mcmc.sum, file=fname, quote=F, sep="\t")
}

## Convergence diagnostic plots (always saved)
## -------------------------------------------
fname = paste0("Traces", vtag2, ".pdf")
MCMCtrace(object=mysamples,
          params = c("mean_bias", "mean_BSV_bias", "sd_BSV_bias",
                     "extra_var_of_ratio"),
          ISB=F,
          pdf=T, open_pdf=F,
          filename=fname,
          iter=7500,
          ind=TRUE, Rhat=TRUE, n.eff=TRUE)

## Plot observations vs. predictions
```

```
## ---------------------------------
if (bsave) {
  fname = paste0("Observations vs predictions", vtag2, ".pdf")
  pdf(fname)
}
## Extract the mean predictions from the summary stats
my.i = grep("log_corrected_ratio", rownames(mcmc.sum))
mypred = exp(mcmc.sum[my.i,"MAP.estimates"])
## Plot data vs. MAP predictions
mylims = c(0.2, 50)
plot(mypred, data$obs_ratio, log="xy", las=1, xlim=mylims, ylim=mylims,
     xlab="Prediction", ylab="Observation", col="black")
for (i in 1:N_studies) {
  ## Show were the predictions were before bias correction
  lines(x=c(DDI.data$Pred_GeoM[i], mypred[i]),
        y=rep(data$obs_ratio[i], 2), col="red")
  ## Show the BSV of each observation
  mySE = sqrt(data$obs_BSV[i] / constants$N_subjects[i])
  lines(x=rep(mypred[i], 2),
        y=c(exp(log(data$obs_ratio[i]) - mySE),
            exp(log(data$obs_ratio[i]) + mySE)), col="grey")
}
abline(0, 1, lwd=2)
abline(log10(0.50), 1, lty=5)
abline(log10(0.80), 1, lty=2)
abline(log10(1.25), 1, lty=2)
abline(log10(2.00), 1, lty=5)
if (bsave) dev.off()

## Plot correlations between posterior estimates of main parameters of interest
## ------------------------------------------------------------------------------
if (bsave) {
  fname = paste0("Posterior correlations", vtag2, ".pdf")
  pdf(fname)
}
my.transform = cbind(mysamples$chain1[,c("mean_bias","mean_BSV_bias")],
                     exp(sqrt(mysamples$chain1[,c("extra_var_of_ratio")])))
colnames(my.transform)[3] = "extra_GSD"
pairs(my.transform)
if (bsave) {
  dev.off()
}

## Plot variances to check
## Could be improved by showing the uncertainty about the data values.
## Could also plot SDs or GSD instead of variances in log...
## ------------------------------------------------------------------
if (bsave) {
  fname = paste0("Variances observed vs predicted", vtag2, ".pdf")
  pdf(fname)
}
my.i = grep("true_BSV", rownames(mcmc.sum))
x1 = 0.3
x2 = 1.0
x3 = 1.7
plot(rep(0.3, N_studies), data$obs_BSV_of_ratio, xaxt="n", log="y",
     xlab="", ylab="", las=1, type="n",
     xlim=c(0,2), ylim=range(c(data$obs_BSV_of_ratio,
                               mcmc.sum[my.i,"MAP.estimates"],
                               constants$BSV_simcyp_pred)))
for (i in 1:N_studies) {
  lines(c(x1, x2, x3), col="grey",
        c(constants$BSV_simcyp_pred[i],
          mcmc.sum[my.i,"MAP.estimates"][i],
          data$obs_BSV_of_ratio[i]))
}
points(rep(x1, N_studies), constants$BSV_simcyp_pred, col="violet")
points(rep(x2, N_studies), mcmc.sum[my.i,"MAP.estimates"], col="red")
```

```
points(rep(x3, N_studies), data$obs_BSV_of_ratio)
axis(1, at=c(x1, x2, x3), lab=c("Predicted BSV", "True BSV", "BSV Data"),
     cex=0.6)
if (bsave) dev.off()

## End.
## ===
```

Simulation script 2 (exploratory covariate analysis v2)

```
## Simcyp DDI exploratory data analysis
## Look at effects of covariates on data and predictions.
## v1: basic, uses Processed full DB 03.csv data
## v2: uses Processed full DB 05.csv (corrected to include data
##     version 06 small corrections.
##============================================================================

## Version tag
vtag = "_v2"

## Option flag to save results and plots
bsave = FALSE

## read the data
## =============
DDI.data = read.csv("Processed full DB 05.csv", as.is=F)
str(DDI.data)
dim(DDI.data)
head(DDI.data)

## How many studies do we have for each CYP?
table(DDI.data$CYP)
table(DDI.data$CYP[which(DDI.data$CI_MBI == "CI")])
table(DDI.data$CYP[which(DDI.data$CI_MBI != "CI")])

## How many inhibitor route do we have?
table(DDI.data$Inhibitor_route)
if (length(table(DDI.data$Inhibitor_route)) == 1) { # remove if only 1 route
  DDI.data$Inhibitor_route = NULL
}

## Explore observed vs. pred geometric mean ratios (GMR)
## =====================================================

## Recode Obs_Nsubjects to use it as a category
DDI.data$Obs_Nsubjects = as.factor(ifelse(DDI.data$Obs_Nsubjects < 7,
                                          "Low", "High"))

## Plot, discriminating groups by colors
mylog = "xy"
for (i in 1:dim(DDI.data)[2]) {
  if (is.factor(DDI.data[,i])) {
    my.discrim = DDI.data[,i]
    myclasses = table(my.discrim)
    N.classes = length(myclasses)
    mypal = ifelse(N.classes < 9, 1, 15)
    mypalette = palette.colors(n=N.classes,
                               palette=palette.pals()[mypal], alpha=1)
    mycols = rep(NA, dim(DDI.data)[1])
    for (j in 1:N.classes) {
      my.class = names(myclasses)[j]
      my.index = which(my.discrim == my.class)
      mycols[my.index] = mypalette[j]
    }
    ylims = c(1, 20)
    pdf(paste0("Explore GMRs ", colnames(DDI.data)[i], ".pdf"))
    plot(DDI.data$Obs_GeoMean, DDI.data$Pred_GeoMean,
         xlab="Observed GMR",
```

```
         ylab="Predicted GMR",
         pch=16, col=mycols,las=1, log=mylog, ylim=ylims)
    abline(0, 1, lwd=2)
    abline(log10(0.50), 1, lty=5)
    abline(log10(0.80), 1, lty=2)
    abline(log10(1.25), 1, lty=2)
    abline(log10(2.00), 1, lty=5)
    legend("topleft", names(myclasses), col=mypalette,
           title=names(DDI.data)[i], title.adj=0.5,
           text.col=mypalette, lty=1, bty="n", cex=0.5)
    dev.off()
  } # end if
} # end for i


## Explore observed vs. pred GSDs
## ==============================

## Plot, discriminating groups by colors
mylog = ""
for (i in 1:dim(DDI.data)[2]) {
  if (is.factor(DDI.data[,i])) {
    my.discrim = DDI.data[,i]
    myclasses = table(my.discrim)
    N.classes = length(myclasses)
    mypal = ifelse(N.classes < 9, 1, 15)
    mypalette = palette.colors(n=N.classes,
                               palette=palette.pals()[mypal], alpha=1)
    mycols = rep(NA, dim(DDI.data)[1])
    for (j in 1:N.classes) {
      my.class = names(myclasses)[j]
      my.index = which(my.discrim == my.class)
      mycols[my.index] = mypalette[j]
    }
    ylims = c(1, 3)
    pdf(paste0("Explore GSDs ", colnames(DDI.data)[i], ".pdf"))
    plot(DDI.data$Obs_GSD, DDI.data$Pred_GSD,
         xlab="Observed between-subject GSD",
         ylab="Predicted between-subject GSD",
         pch=16, col=mycols,las=1, log=mylog, ylim=ylims)
    legend("topright", names(myclasses), col=mypalette,
           title=names(DDI.data)[i], title.adj=0,
           text.col=mypalette, lty=1, bty="n", cex=0.5)
    dev.off()
  } # end if
} # end for i


## End.
## ===
```

Preliminary covariate effects' graphical analyses

Effects on GMRs

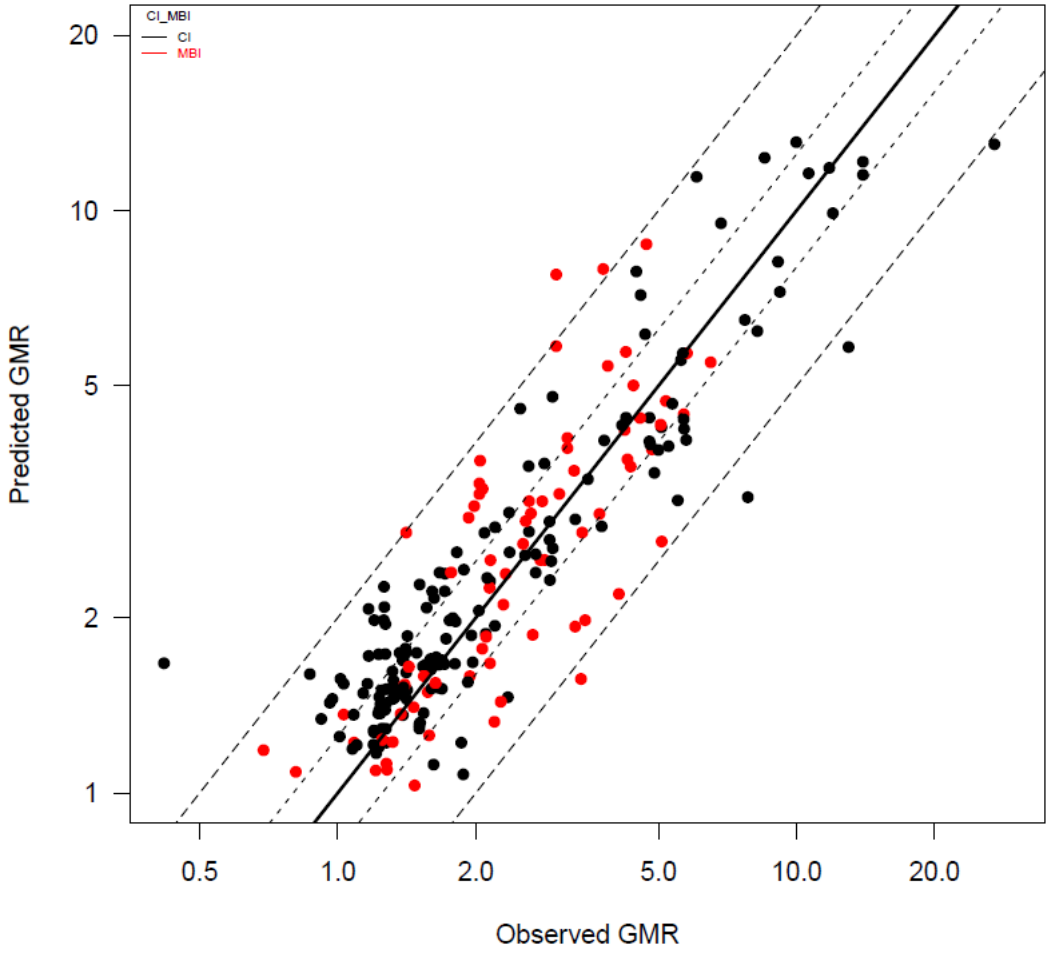


Figure 6: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on the type of inhibition assessed.

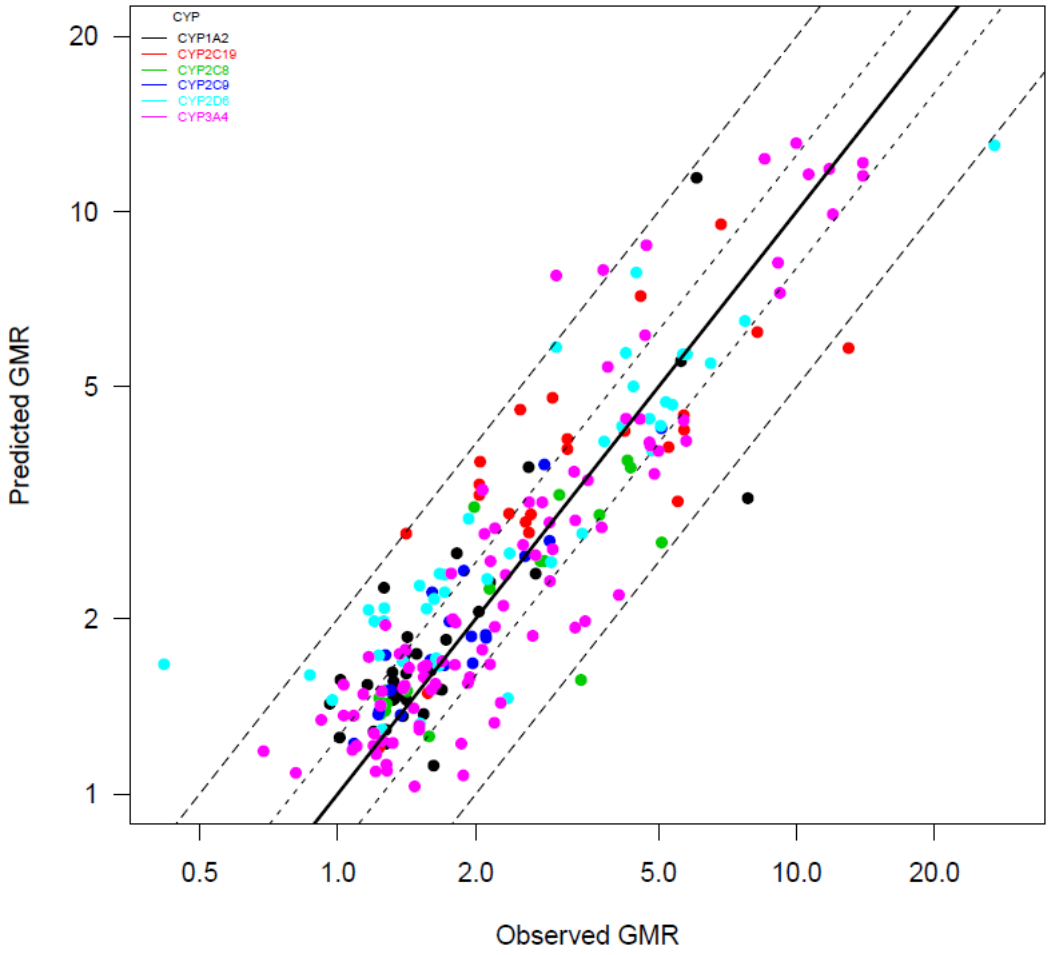


Figure 7: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on CYP studied.

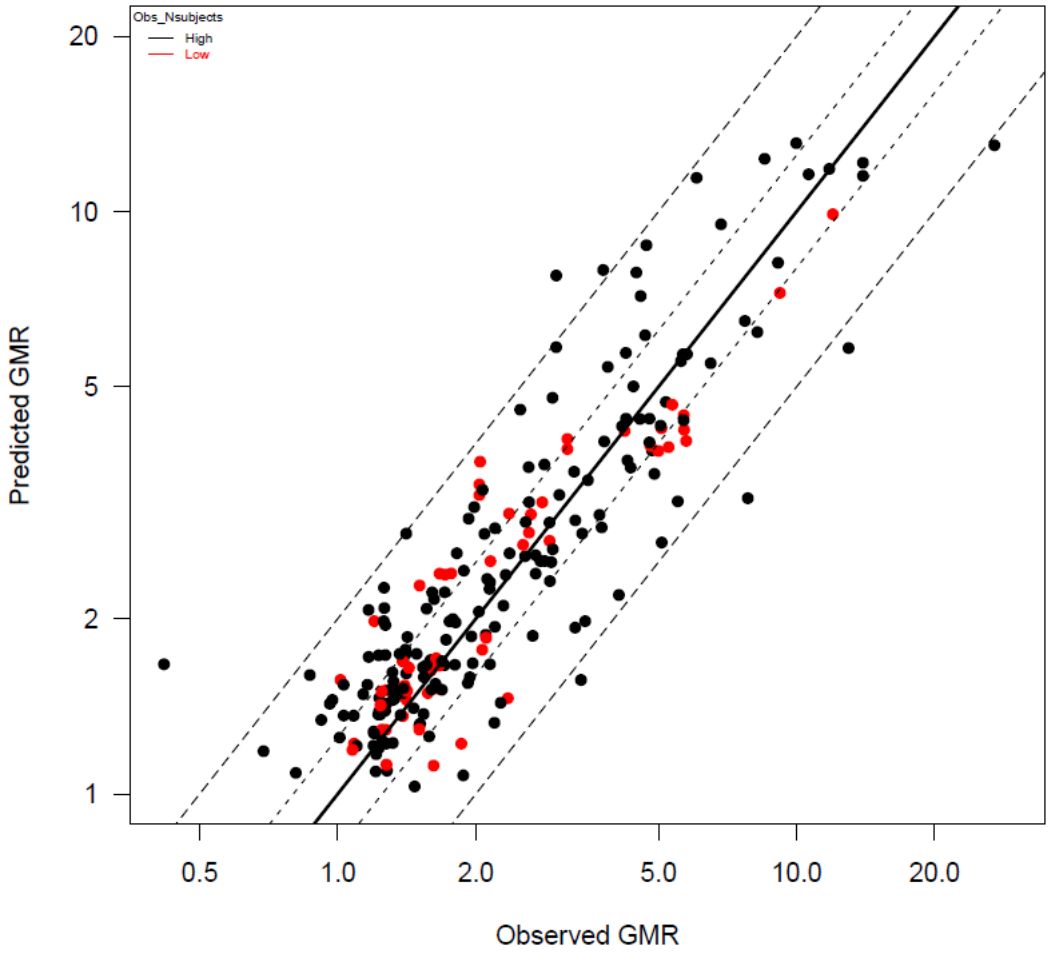


Figure 8: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on number of subjects in each trial (low: less than 7 subjects; high: more than 6 subjects).

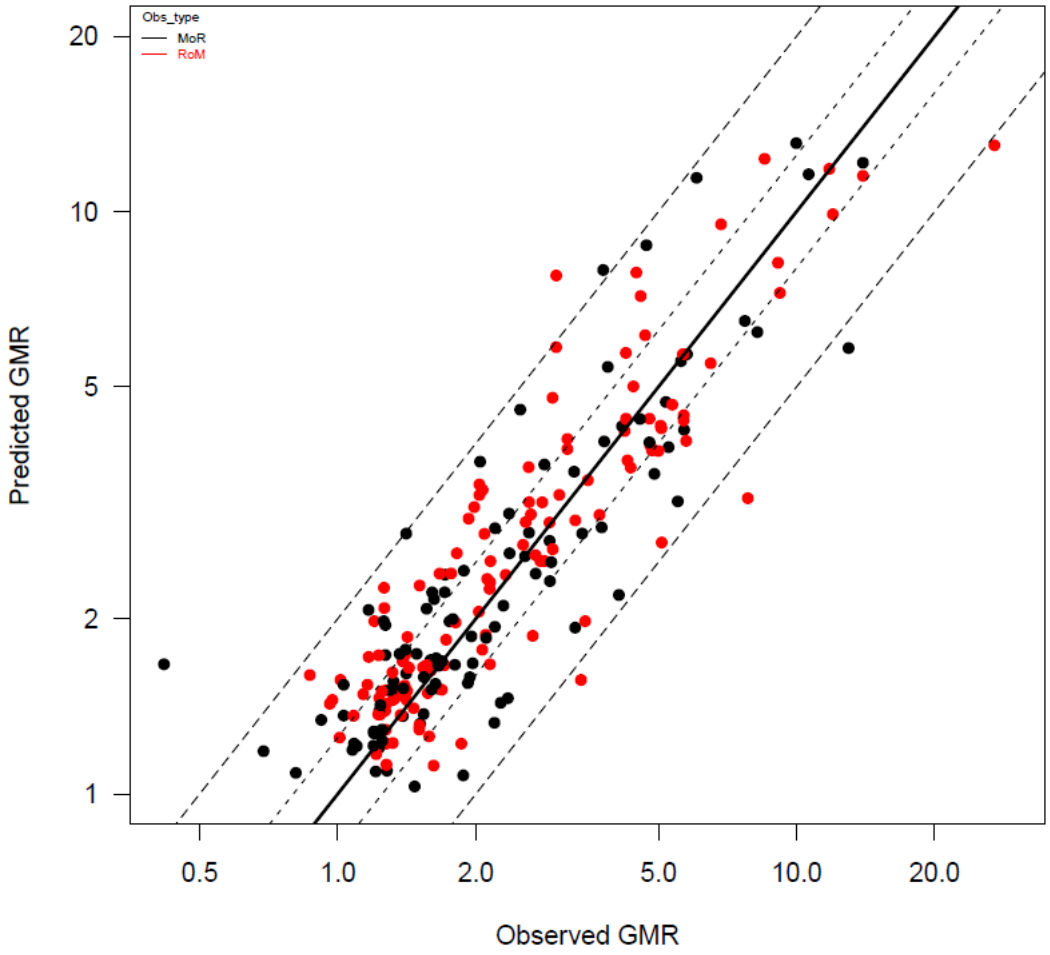


Figure 9: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on the type of GMR reported (MoR: mean of ratios, RoM: ratio of means).

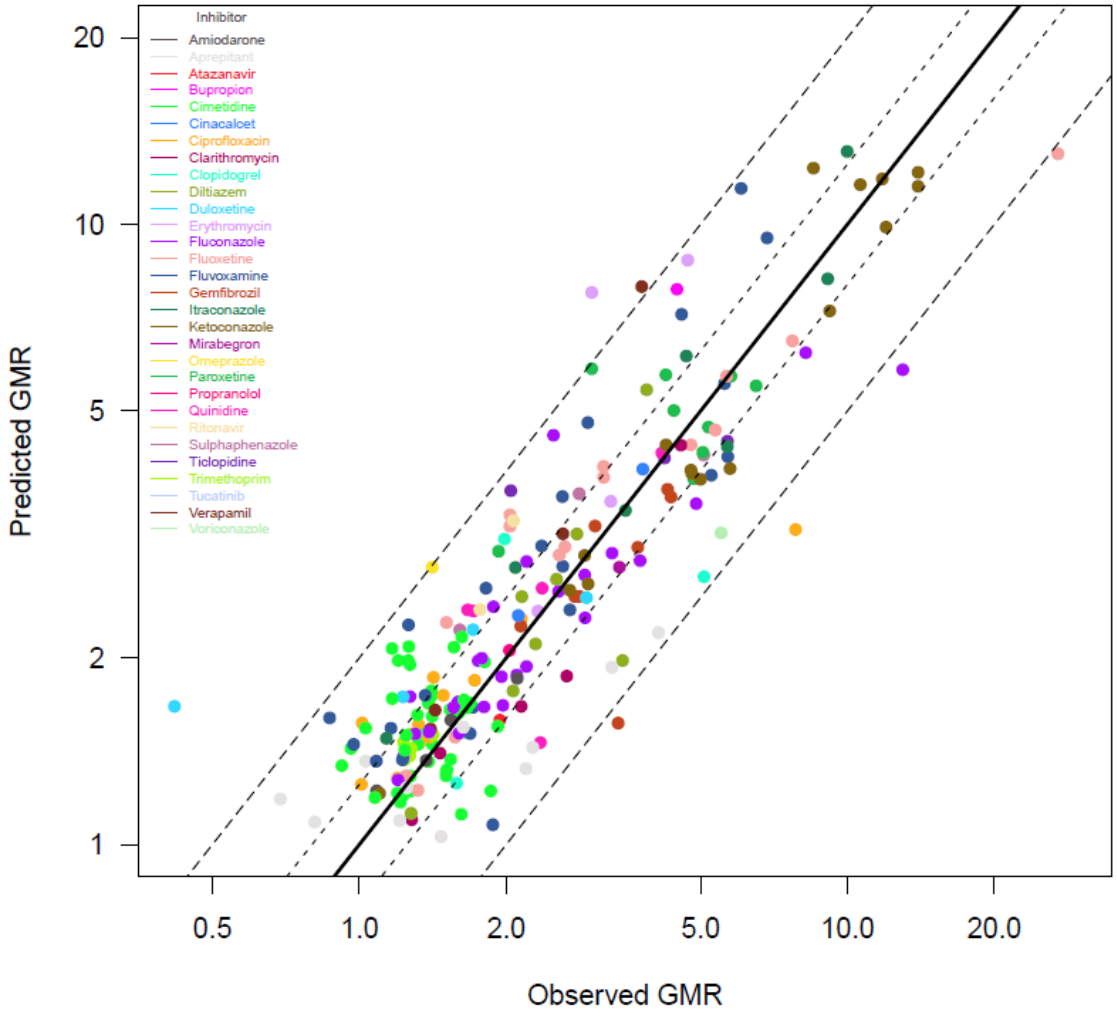


Figure 10: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on the inhibitor tested.

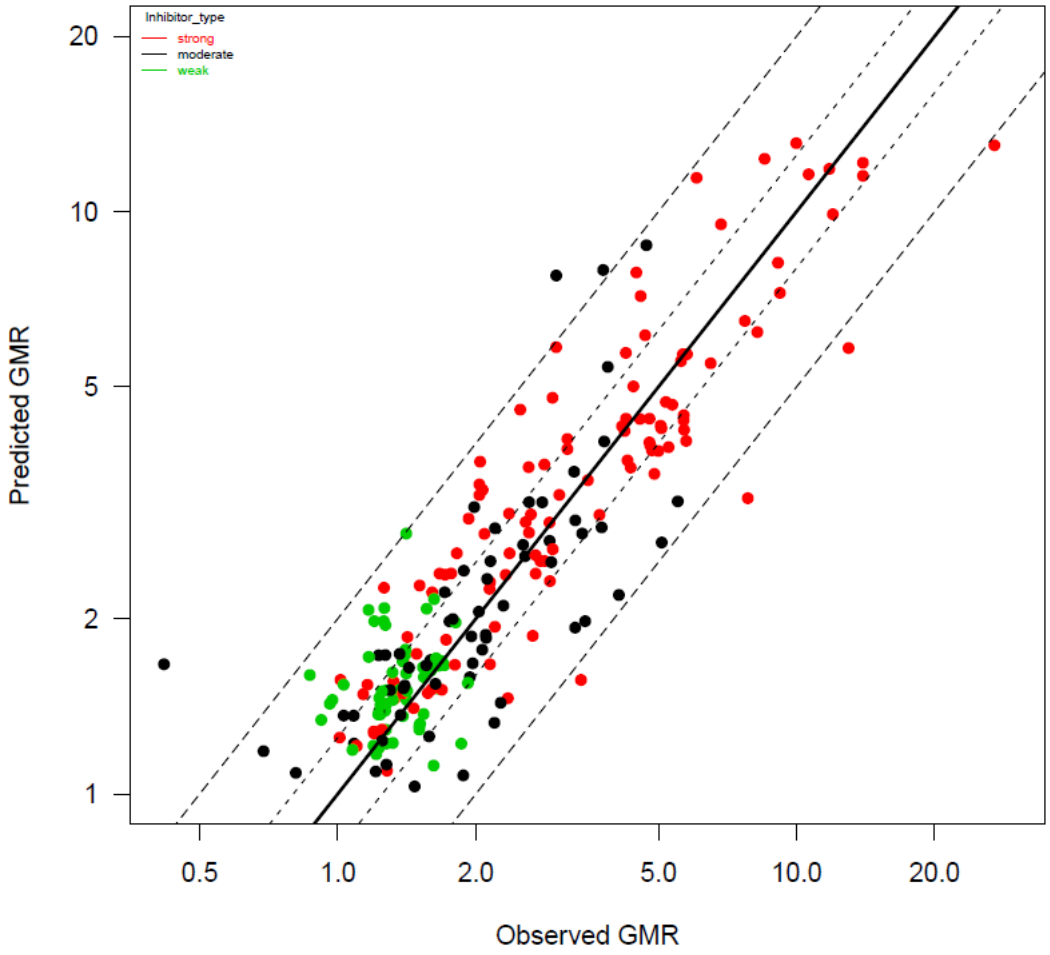


Figure 11: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on inhibitor type (weak, moderate, strong) tested.

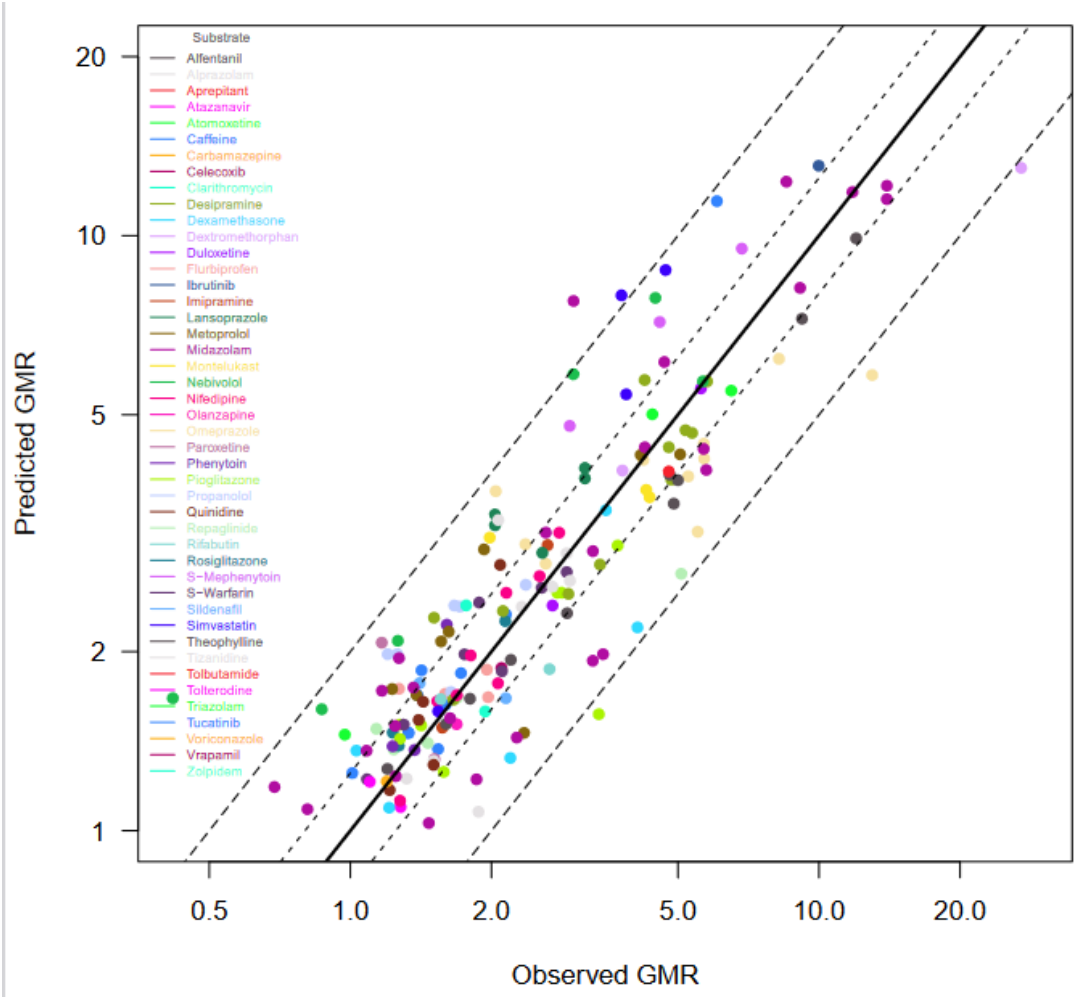


Figure 12: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on the substrate tested.

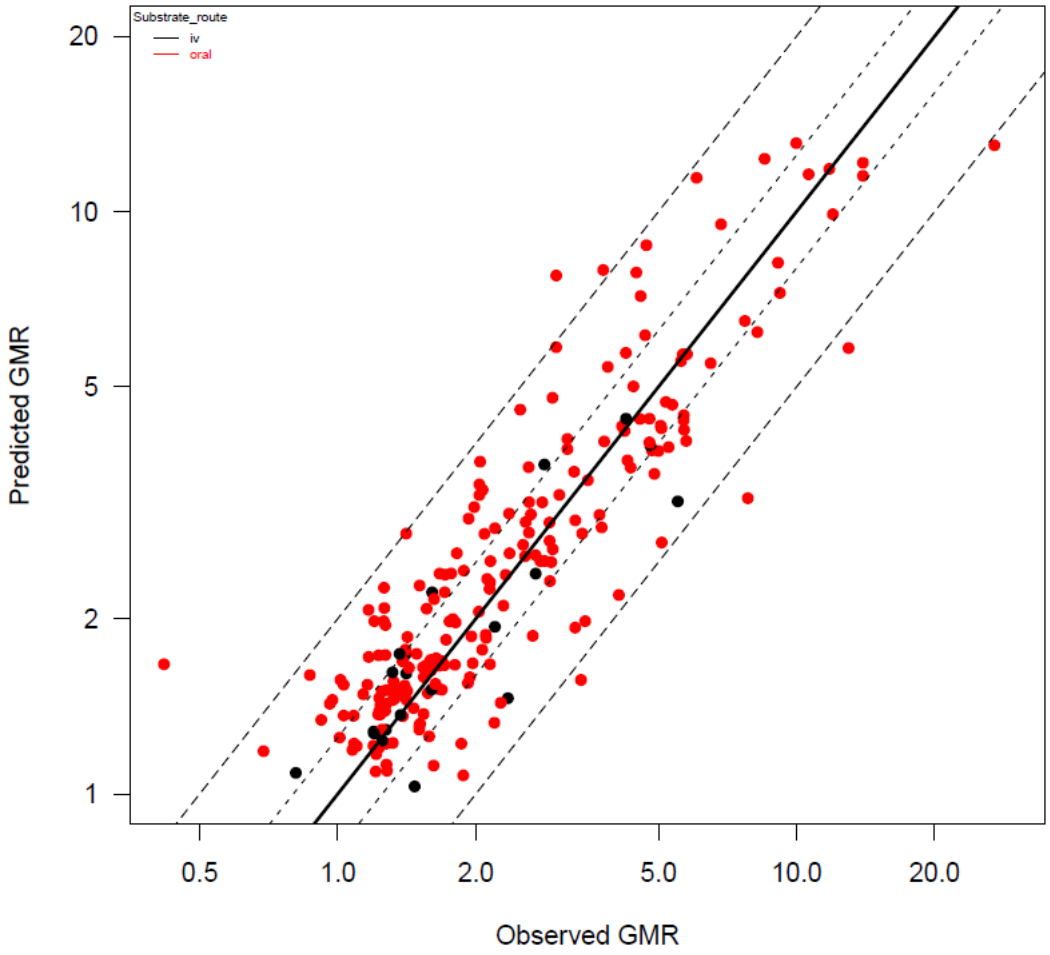


Figure 13: Simcyp-predicted *vs*. observed geometric mean ratios (with inhibitor over without inhibitor) for all data. The data points are colored according to the covariate data on the route of administration of the substrate tested.

Effects on GSDs

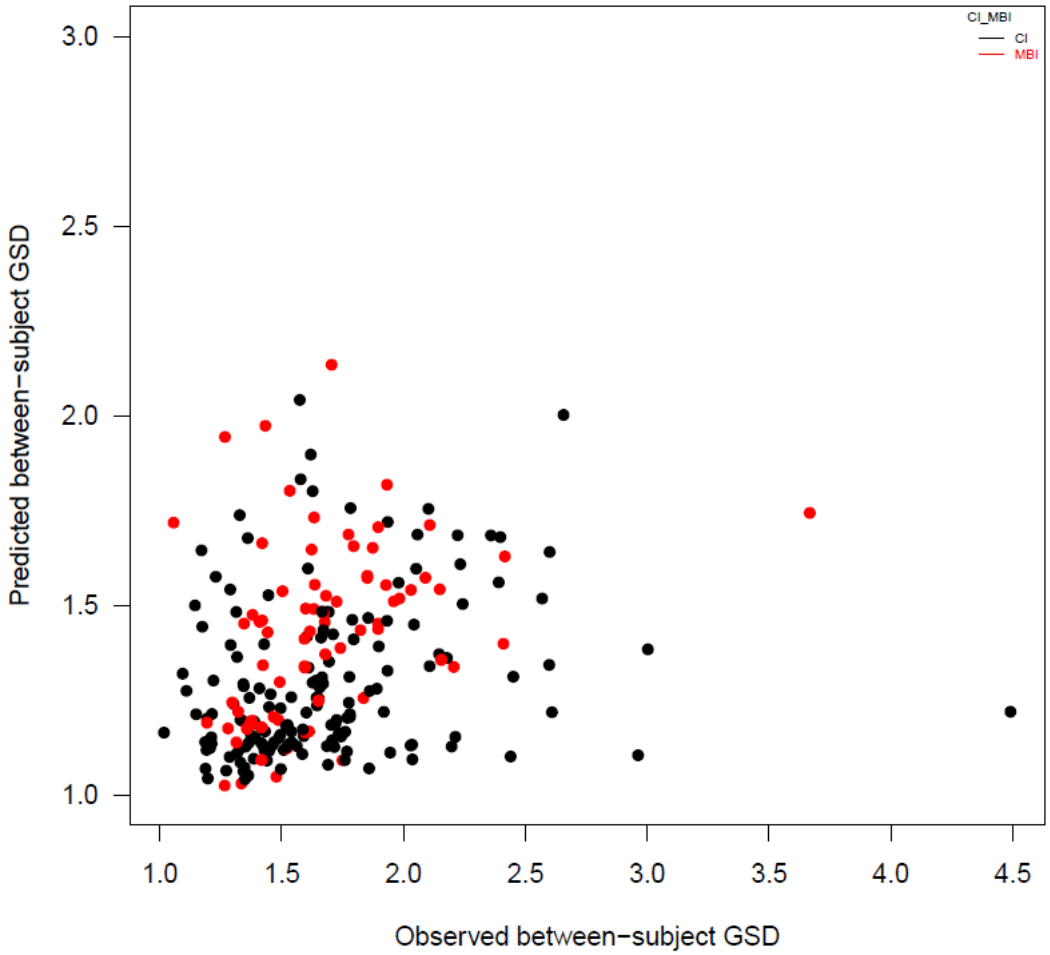


Figure 14: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on the type of inhibition assessed.

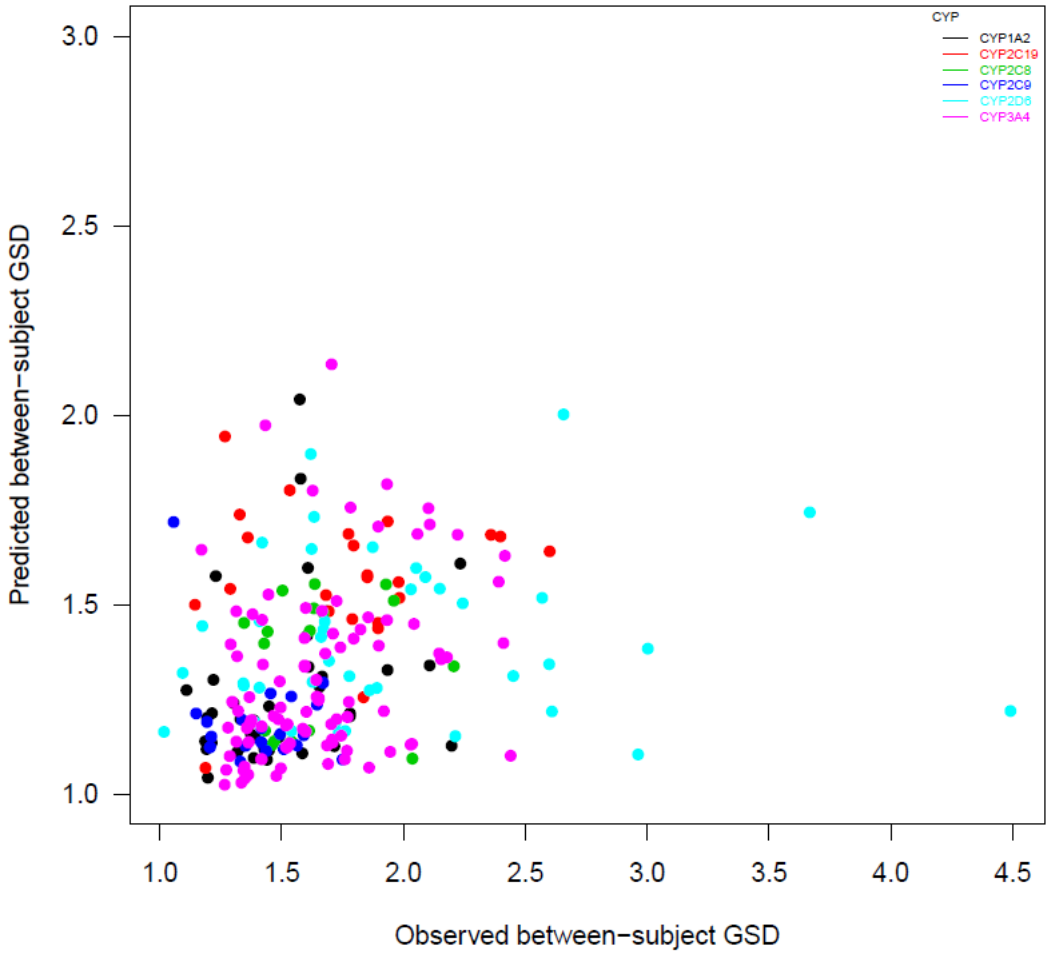


Figure 15: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on CYP studied.

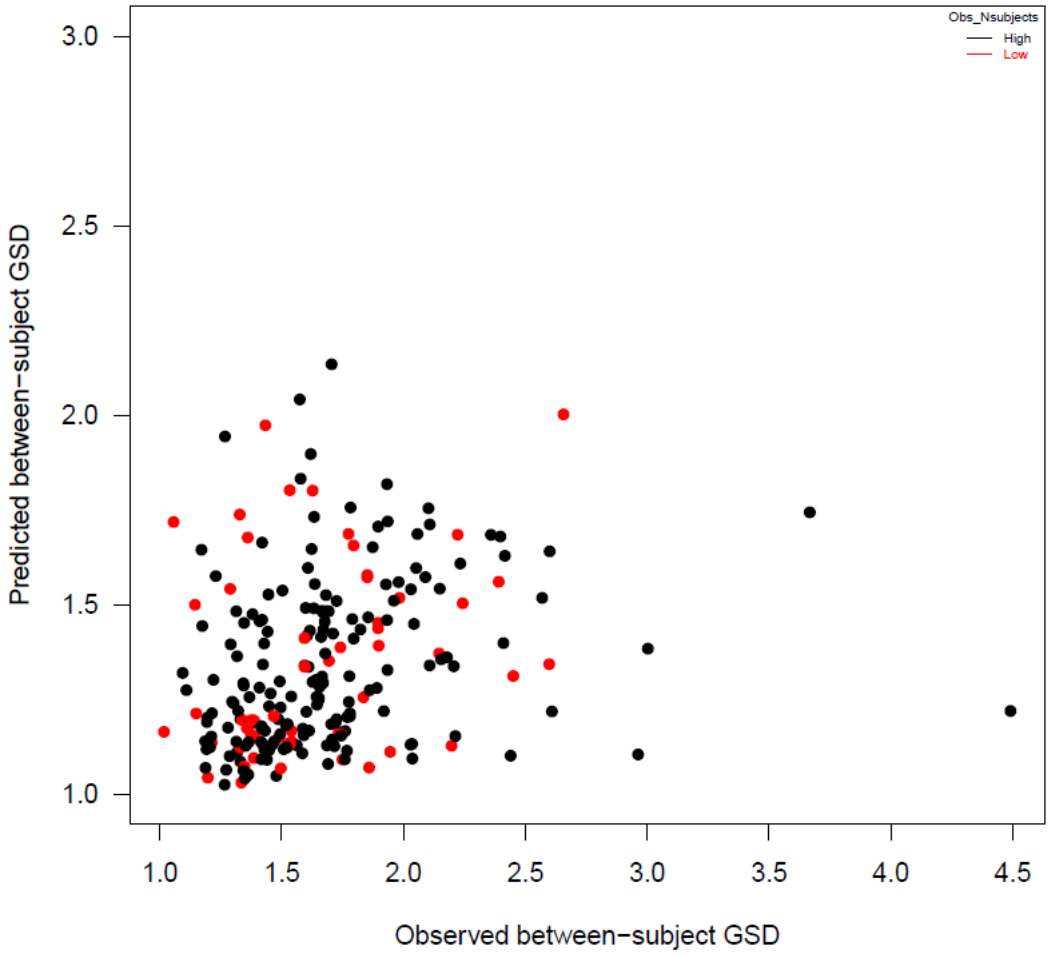


Figure 16: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on number of subjects in each trial (low: less than 7 subjects; high: more than 6 subjects).

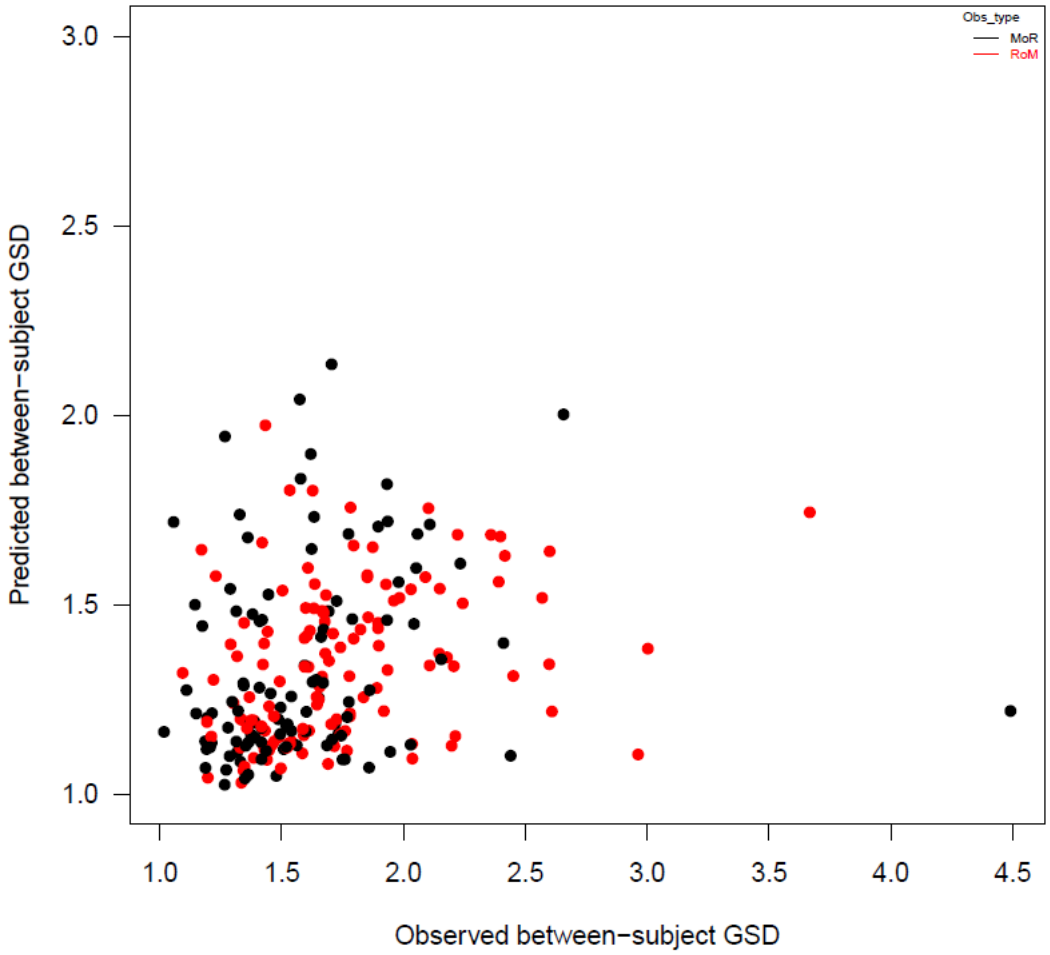


Figure 17: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on the type of GMR reported (MoR: mean of ratios, RoM: ratio of means).

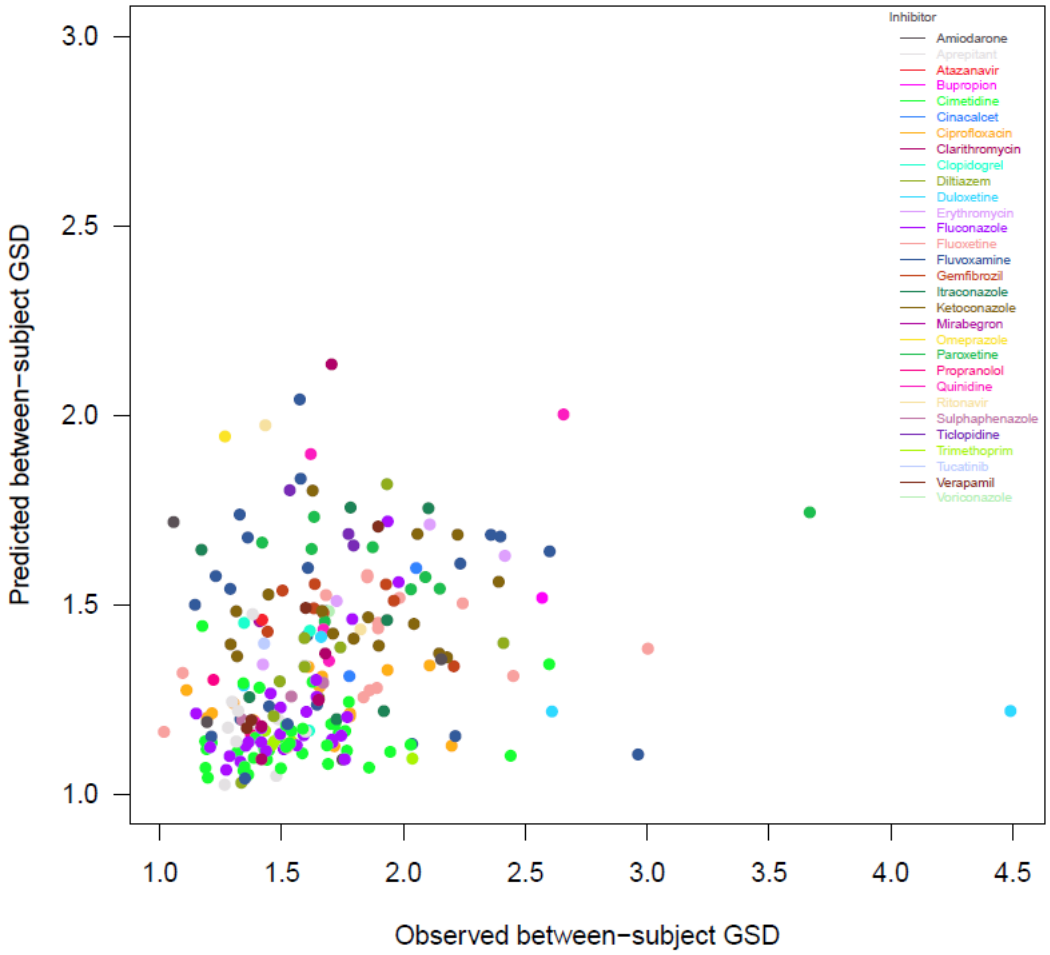


Figure 18: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on the inhibitor tested.

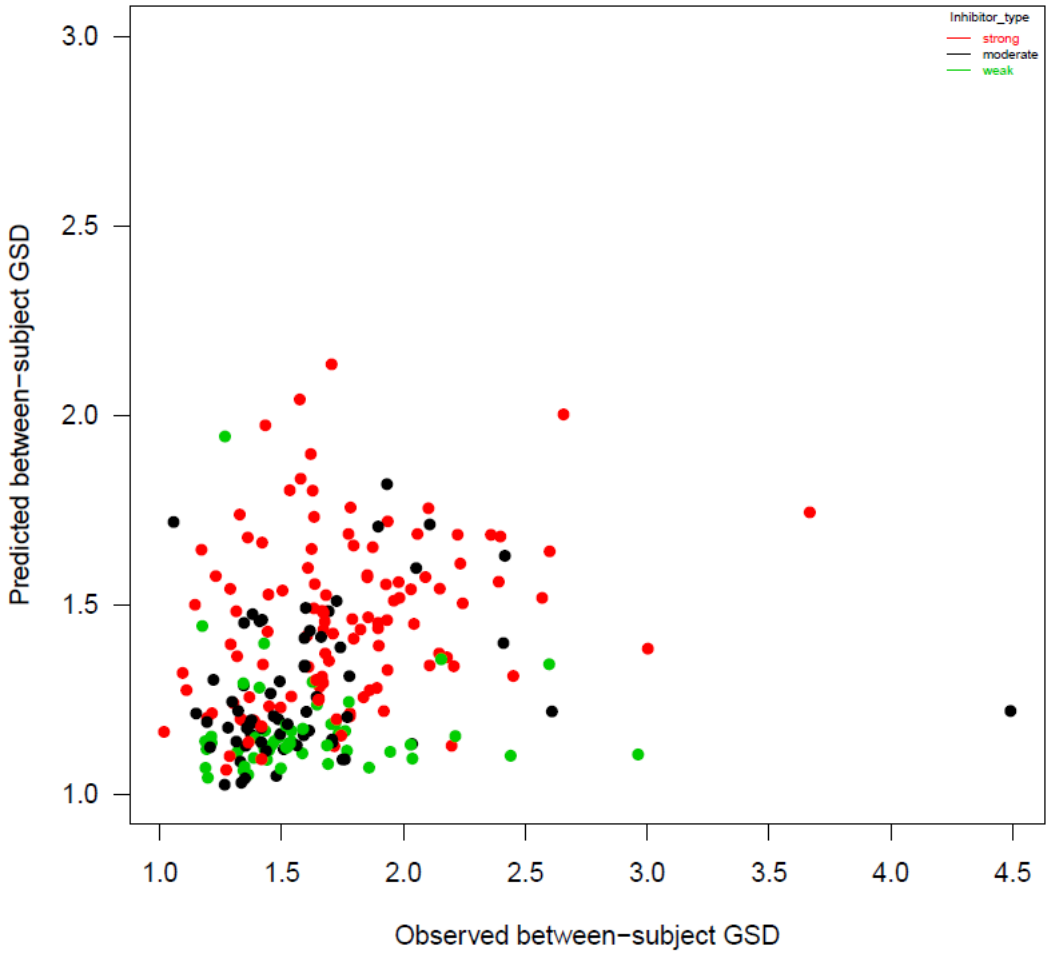


Figure 19: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on inhibitor type (weak, moderate, strong) tested.

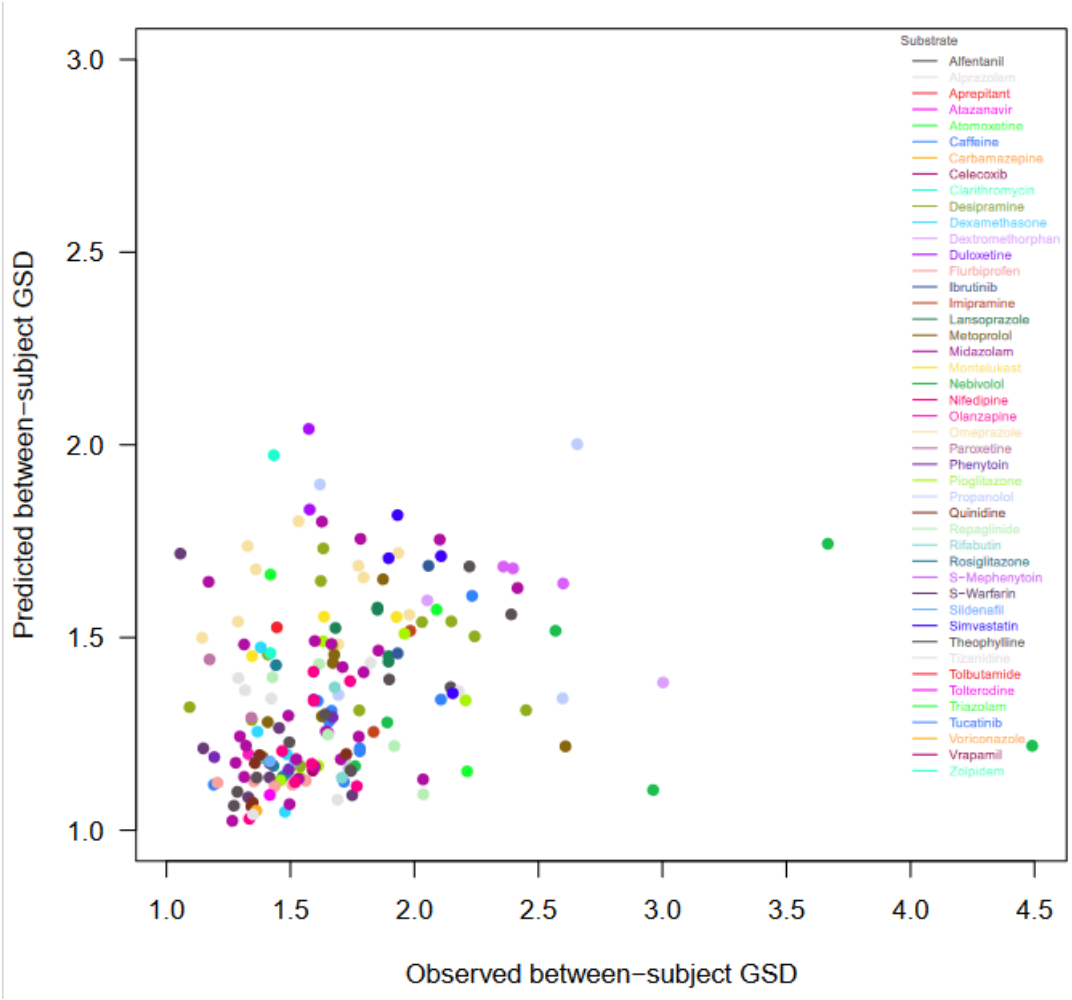


Figure 20: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on the substrate tested.

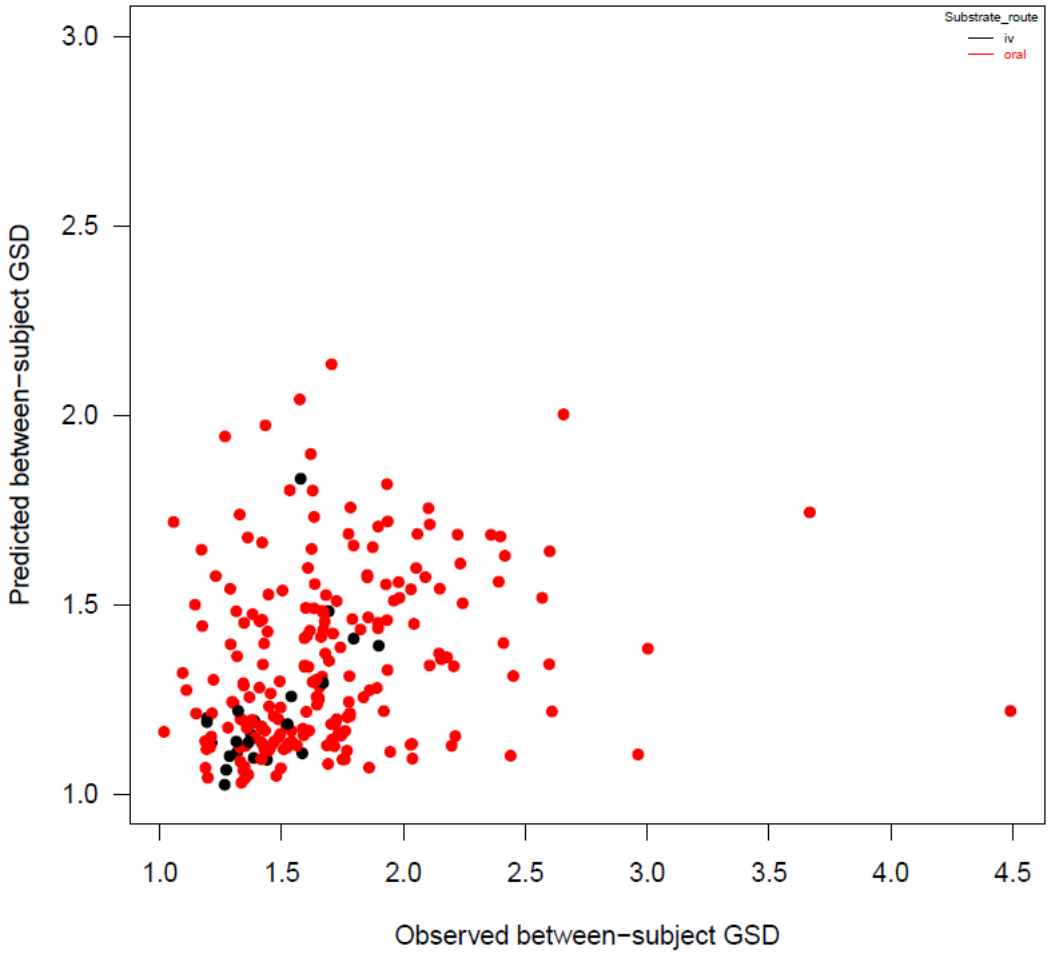


Figure 21: Simcyp-predicted *vs*. observed between-subject geometric standard deviations for all data. The data points are colored according to the covariate data on the route of administration of the substrate tested.

MCMC posterior plots

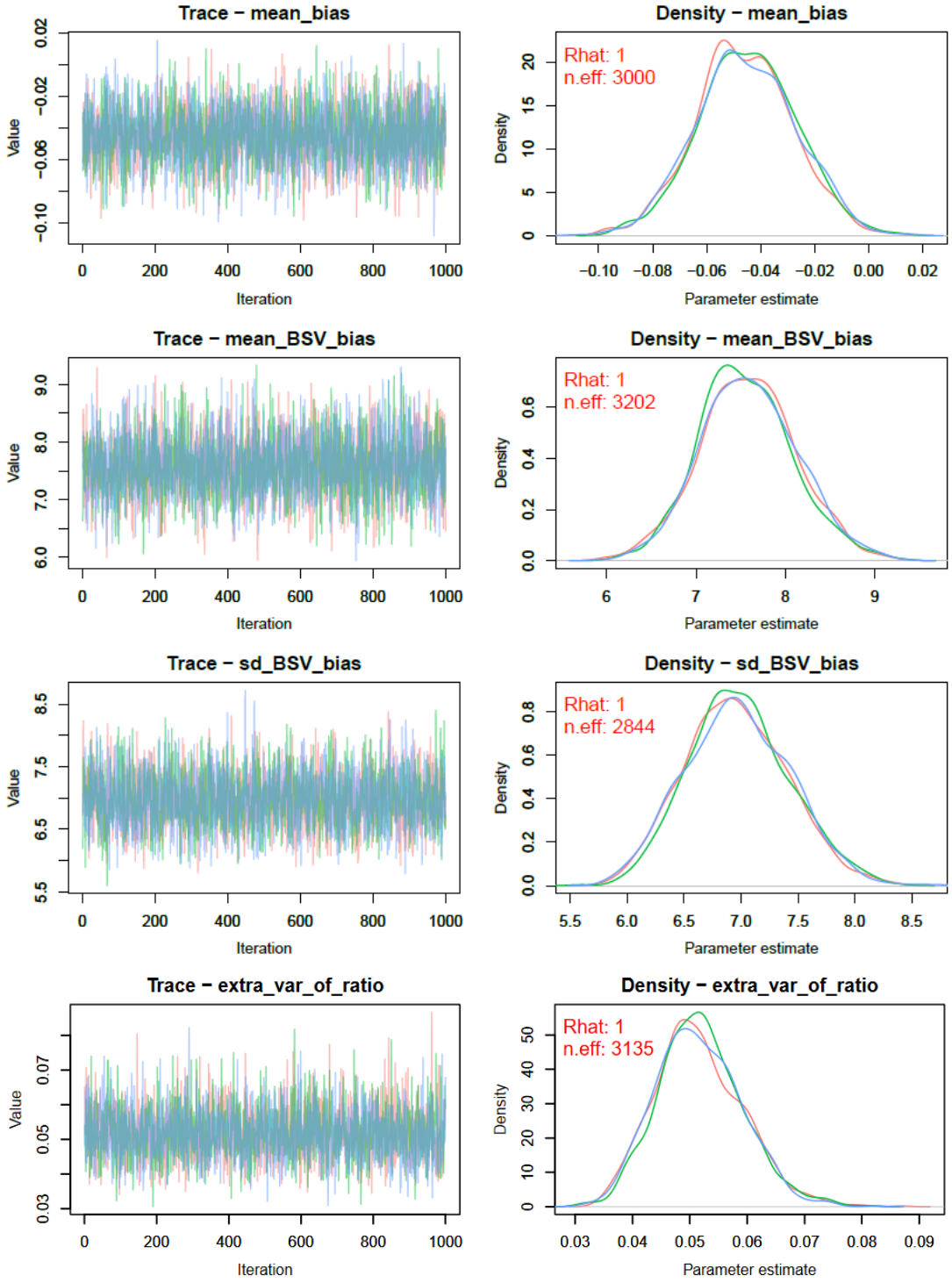


Figure 22: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *all the data*.

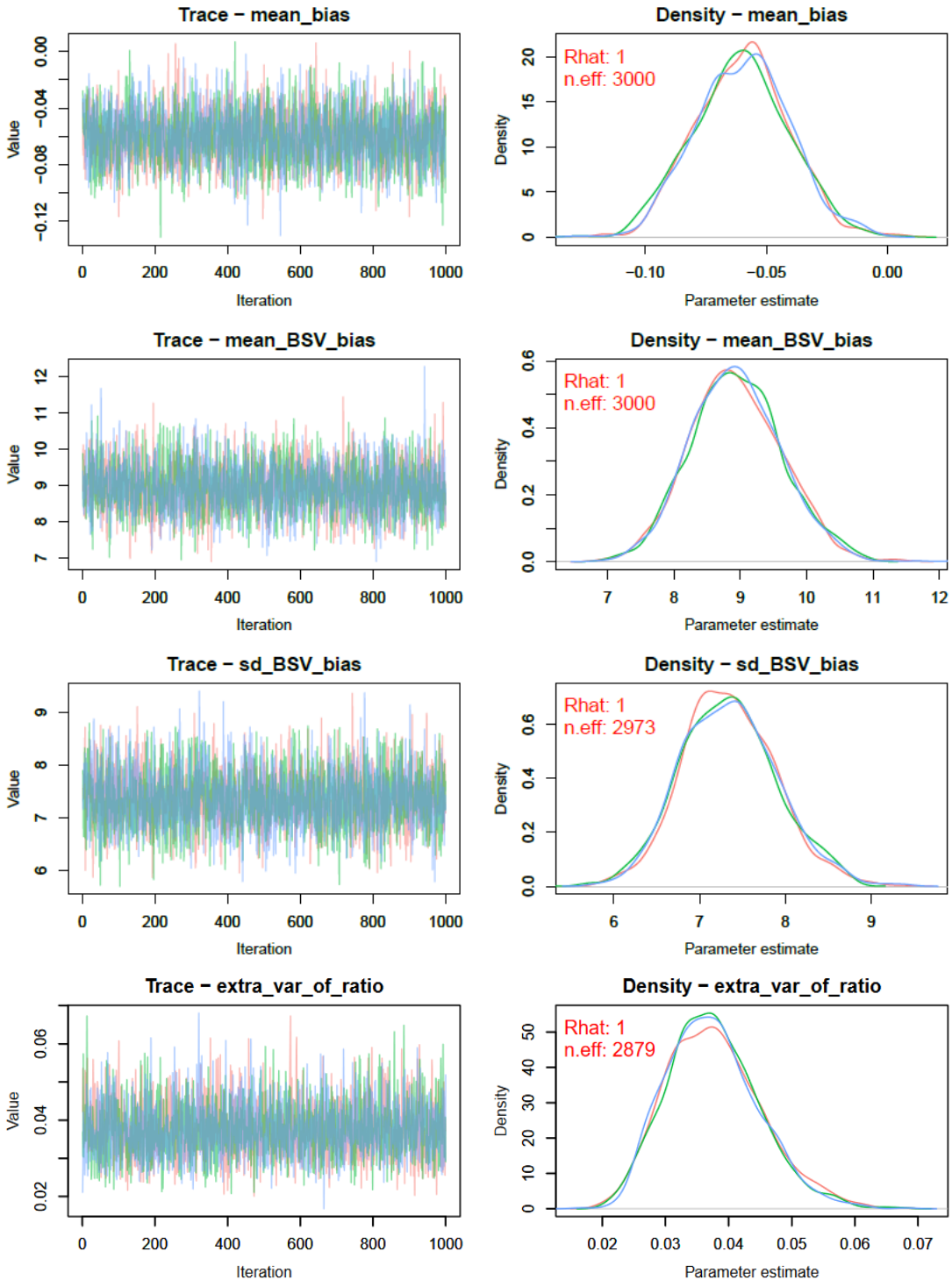


Figure 23: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for competitive inhibition*.

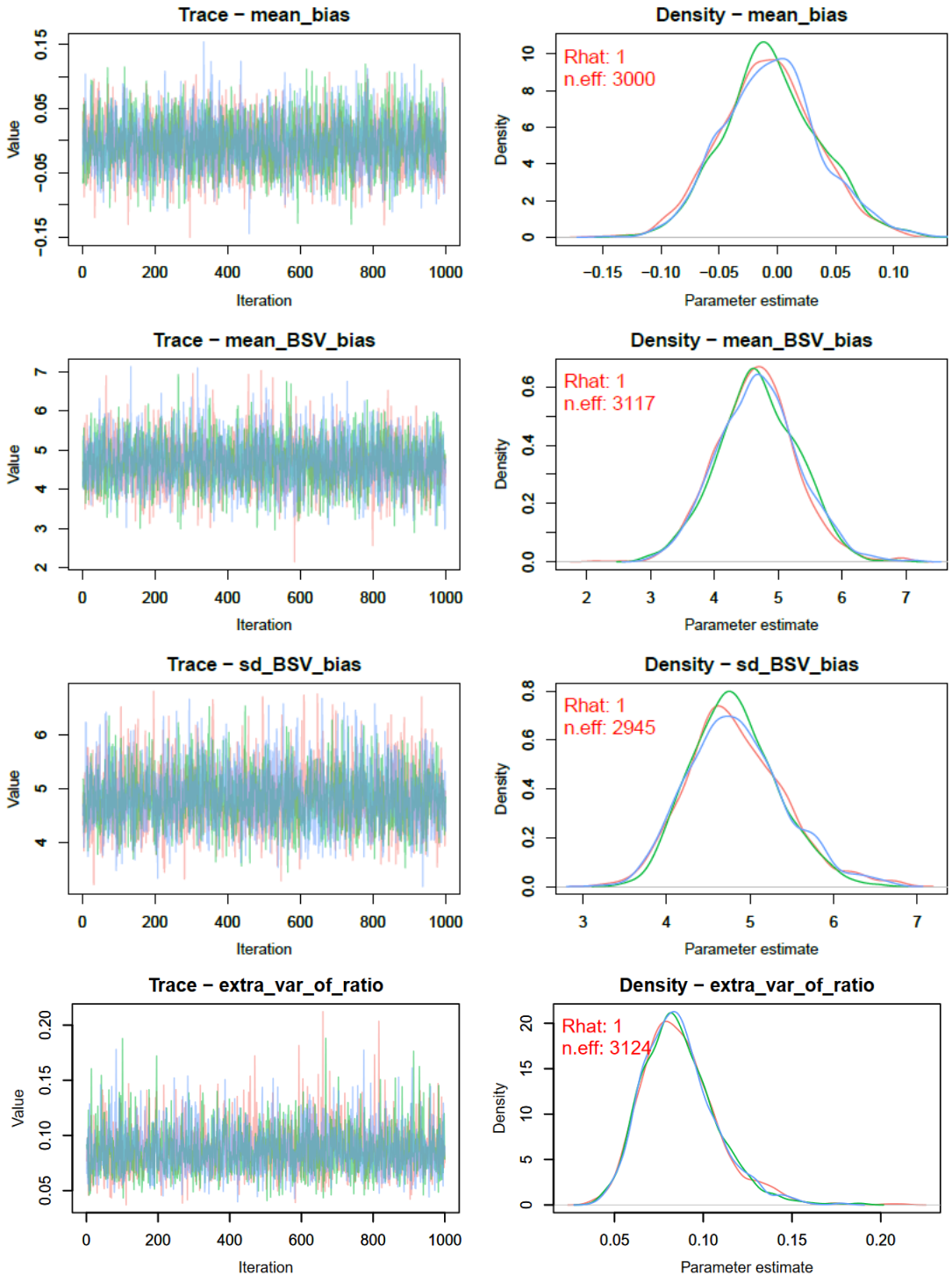


Figure 24: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for mechanism-based inhibition*.

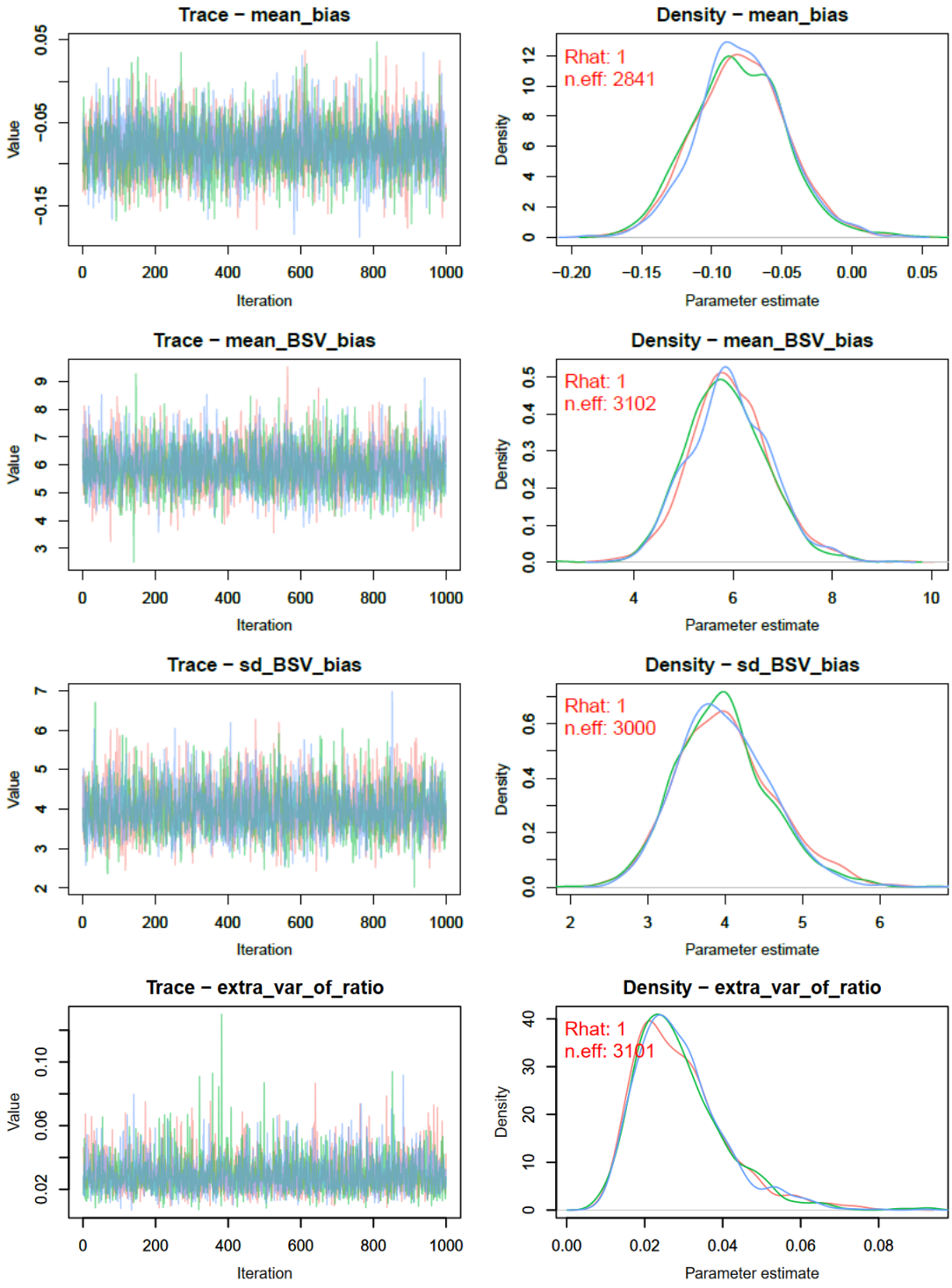


Figure 25: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP1A2*.

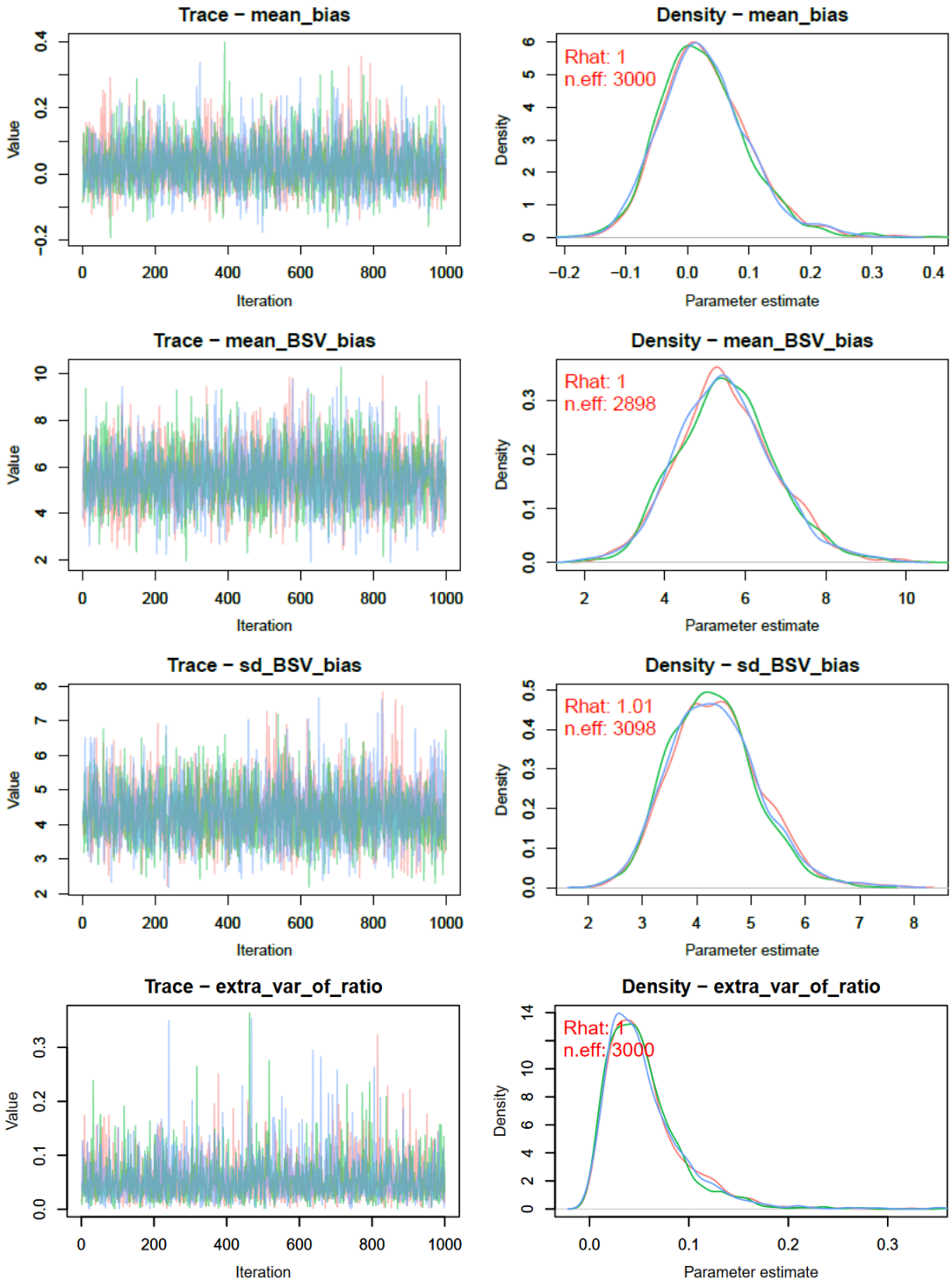


Figure 26: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP2C8*.

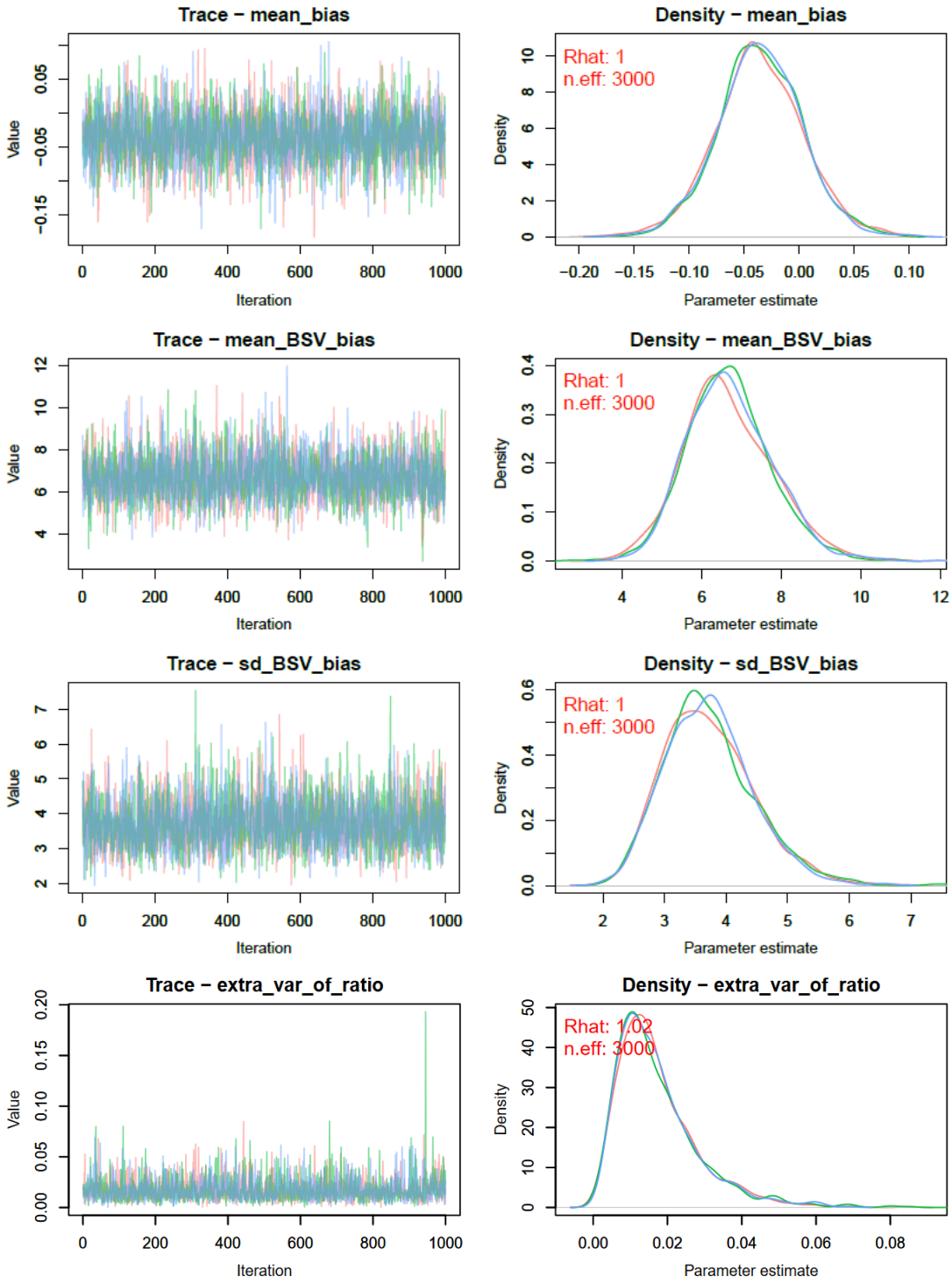


Figure 27: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP2C9*.

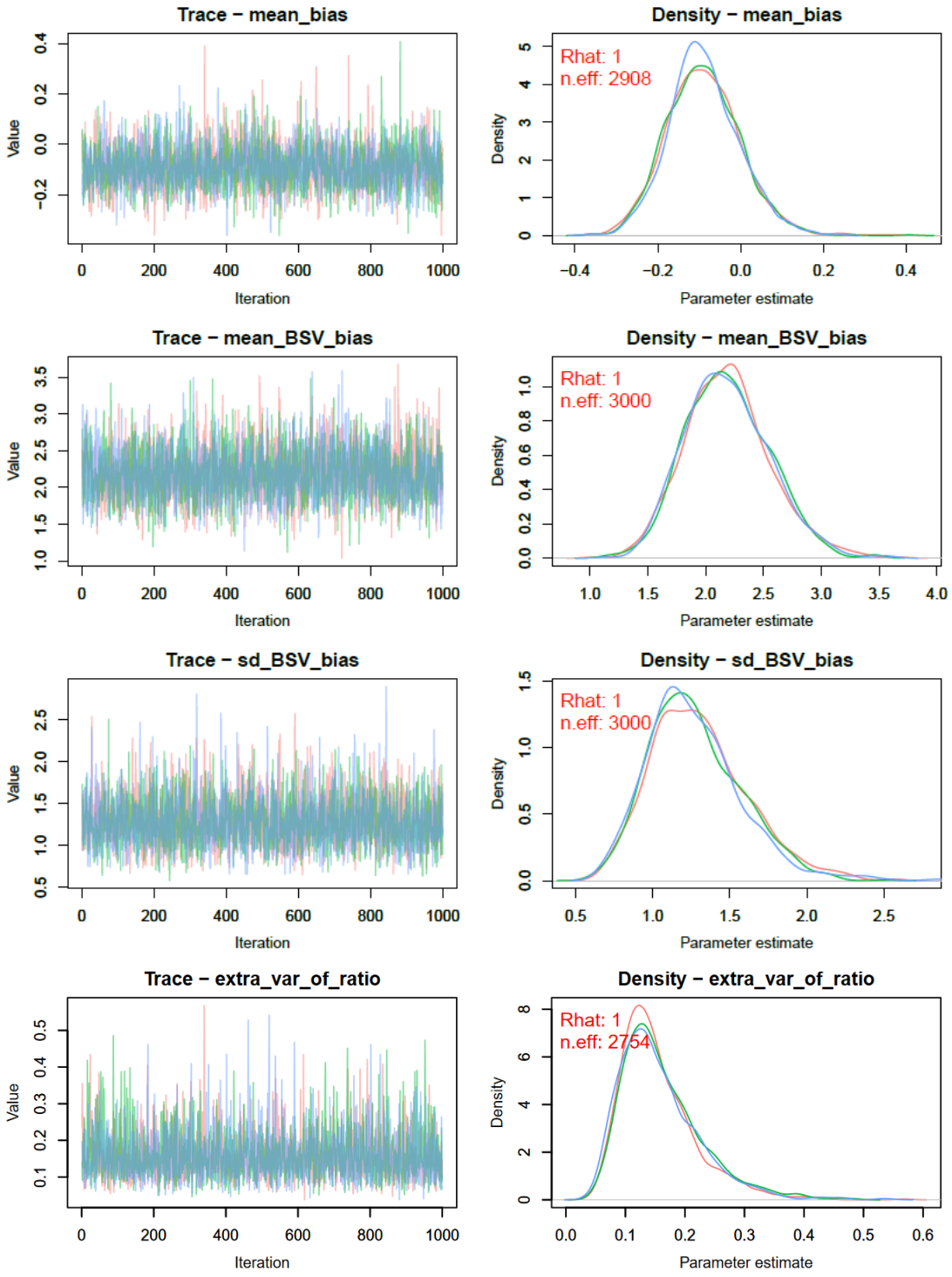


Figure 28: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP2C19*.

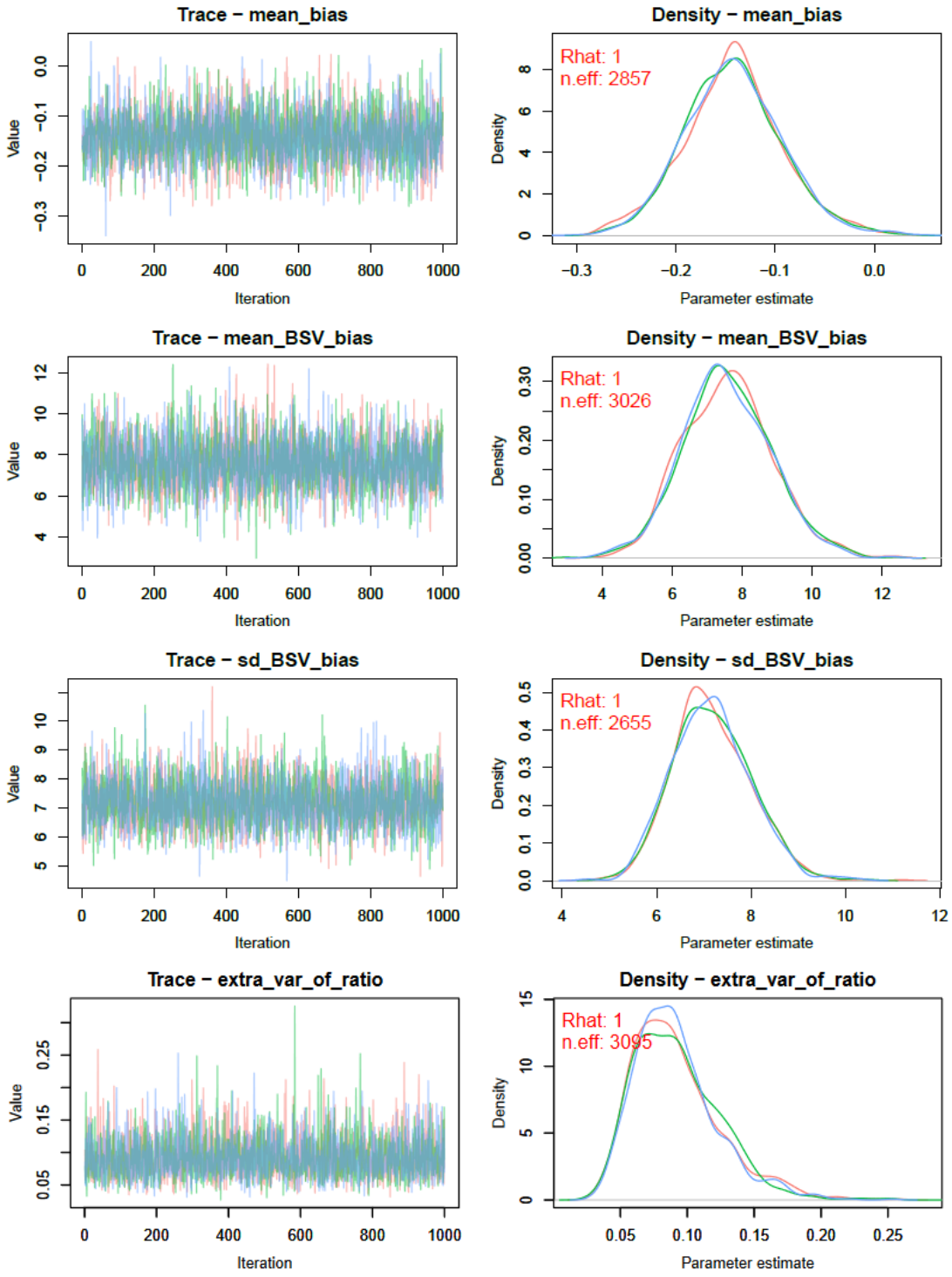


Figure 29: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP2D6*.

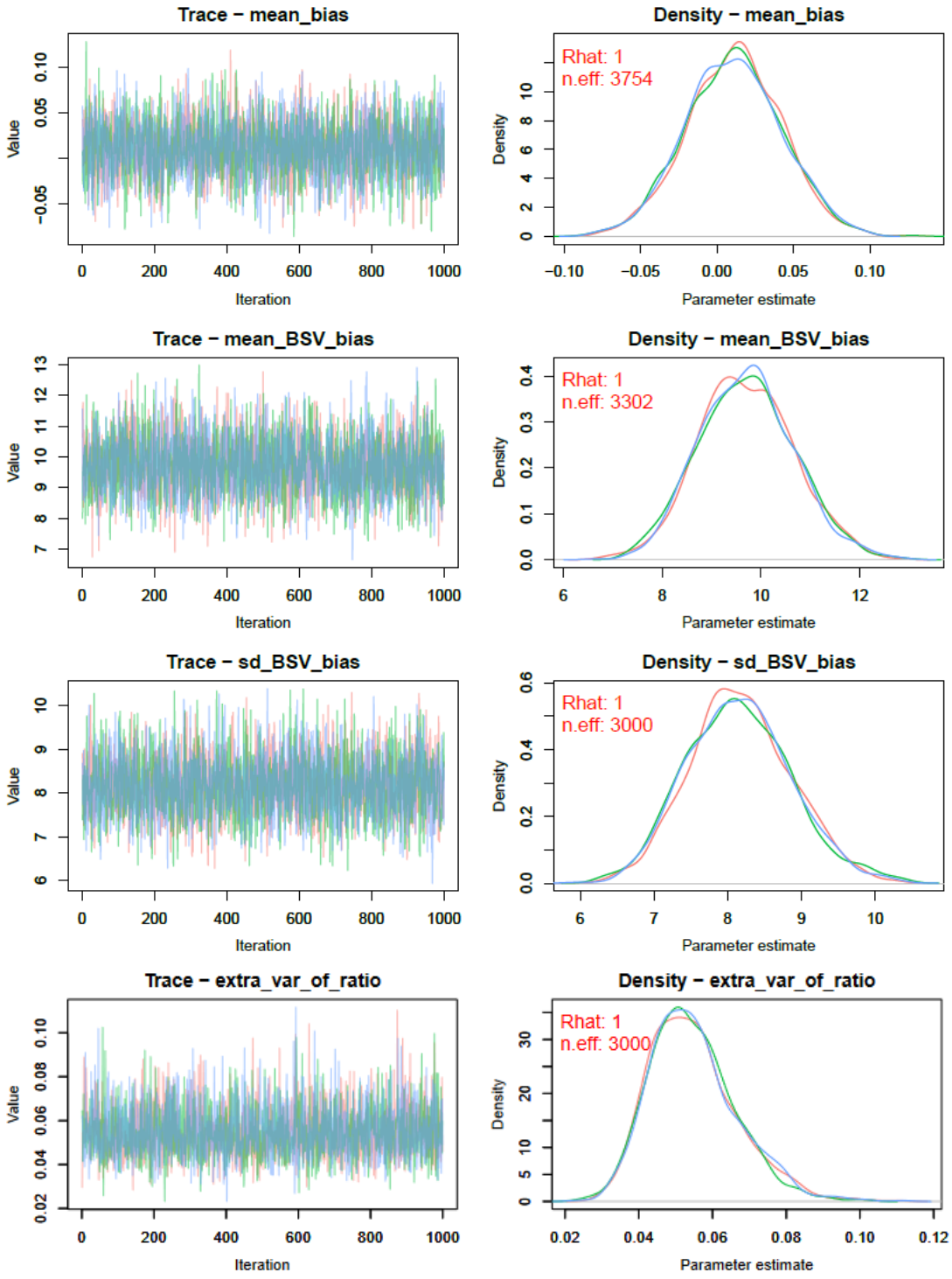


Figure 30: MCMC chains (left side) and estimated posterior distribution (one per chain, right side) for the analysis including *only the data for CYP3A4*.

MCMC posterior summary statistics

Table 5: Statistical summaries of the posterior distributions of the main parameters for the analysis including *all the data*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	**Mean**	**SD**	**2.5%tile**	**Median**	**97.5%tile**	$\hat{R}$
mean bias	-0.045	0.018	-0.08	-0.046	-0.011	1
mean BSV bias	7.579	0.532	6.572	7.561	8.655	1
SD of BSV bias	6.971	0.45	6.136	6.952	7.888	1
between-study variance	0.052	0.008	0.038	0.051	0.068	1

Table 6: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for competitive inhibition*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	**Mean**	**SD**	**2.5%tile**	**Median**	**97.5%tile**	$\hat{R}$
mean bias	-0.06	0.019	-0.096	-0.059	-0.022	1
mean BSV bias	8.938	0.685	7.651	8.906	10.304	1
SD of BSV bias	7.344	0.551	6.318	7.331	8.511	1
between-study variance	0.038	0.007	0.025	0.037	0.054	1

Table 7: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for mechanism-based inhibition*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	**Mean**	**SD**	**2.5%tile**	**Median**	**97.5%tile**	$\hat{R}$
mean bias	-0.006	0.041	-0.083	-0.007	0.079	1
mean BSV bias	4.674	0.629	3.462	4.667	5.927	1
SD of BSV bias	4.823	0.557	3.849	4.782	6.029	1
between-study variance	0.087	0.021	0.053	0.084	0.135	1

Table 8: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP1A2*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	**Mean**	**SD**	**2.5%tile**	**Median**	**97.5%tile**	$\hat{R}$
mean bias	-0.079	0.032	-0.141	-0.08	-0.014	1
mean BSV bias	5.917	0.813	4.405	5.884	7.658	1
SD of BSV bias	3.994	0.614	2.921	3.952	5.348	1
between-study variance	0.029	0.012	0.013	0.027	0.058	1

Table 9: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP2C8*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	Mean	SD	2.5%tile	Median	97.5%tile	$\hat{R}$
mean bias	0.028	0.071	-0.096	0.022	0.19	1
mean BSV bias	5.53	1.209	3.361	5.465	8.083	1
SD of BSV bias	4.323	0.813	2.908	4.281	6.024	1
between-study variance	0.055	0.039	0.007	0.047	0.153	1

Table 10: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP2C9*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	Mean	SD	2.5%tile	Median	97.5%tile	$\hat{R}$
mean bias	-0.036	0.039	-0.113	-0.037	0.042	1
mean BSV bias	6.685	1.087	4.719	6.6	8.945	1
SD of BSV bias	3.72	0.723	2.508	3.656	5.308	1
between-study variance	0.017	0.012	0.004	0.015	0.047	1.02

Table 11: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP2C19*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	Mean	SD	2.5%tile	Median	97.5%tile	$\hat{R}$
mean bias	-0.089	0.088	-0.252	-0.093	0.097	1
mean BSV bias	2.19	0.364	1.552	2.167	2.954	1
SD of BSV bias	1.281	0.306	0.775	1.243	1.962	1
between-study variance	0.017	0.012	0.004	0.015	0.047	1.02

Table 12: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP2D6*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	Mean	SD	2.5%tile	Median	97.5%tile	$\hat{R}$
mean bias	-0.142	0.049	-0.239	-0.143	-0.039	1
mean BSV bias	7.575	1.268	5.241	7.529	10.26	1
SD of BSV bias	7.17	0.824	5.697	7.121	8.828	1
between-study variance	0.092	0.032	0.045	0.087	0.169	1

Table 13: Statistical summaries of the posterior distributions of the main parameters for the analysis including *only the data for CYP3A4*. $\hat{R}$ is a convergence diagnostic [2], which should be close to 1.

Parameter	Mean	SD	2.5%tile	Median	97.5%tile	$\hat{R}$
mean bias	0.011	0.031	-0.05	0.011	0.072	1
mean BSV bias	9.707	0.959	7.899	9.691	11.70	1
SD of BSV bias	8.157	0.698	6.875	8.136	9.610	1
between-study variance	0.055	0.012	0.035	0.053	0.081	1

MCMC posterior correlations between main parameters

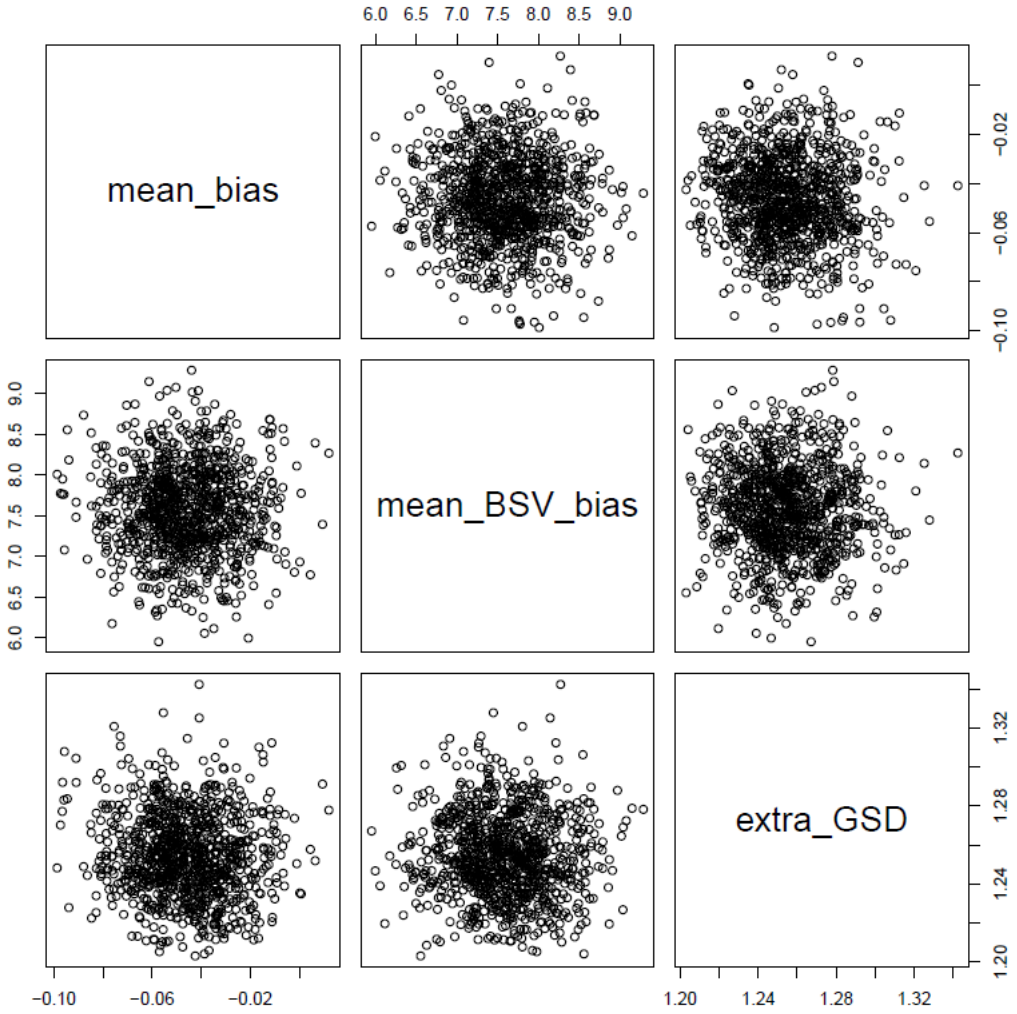


Figure 31: Correlations between the posterior samples of the three main parameters for the analysis including *all the data*. Note that extra GSD is the exponential of the square root of between-study variability variance. The same pattern of very low correlations is obtained for the other analyses (data not shown).

MCMC posterior observed vs. predicted geometric mean ratios

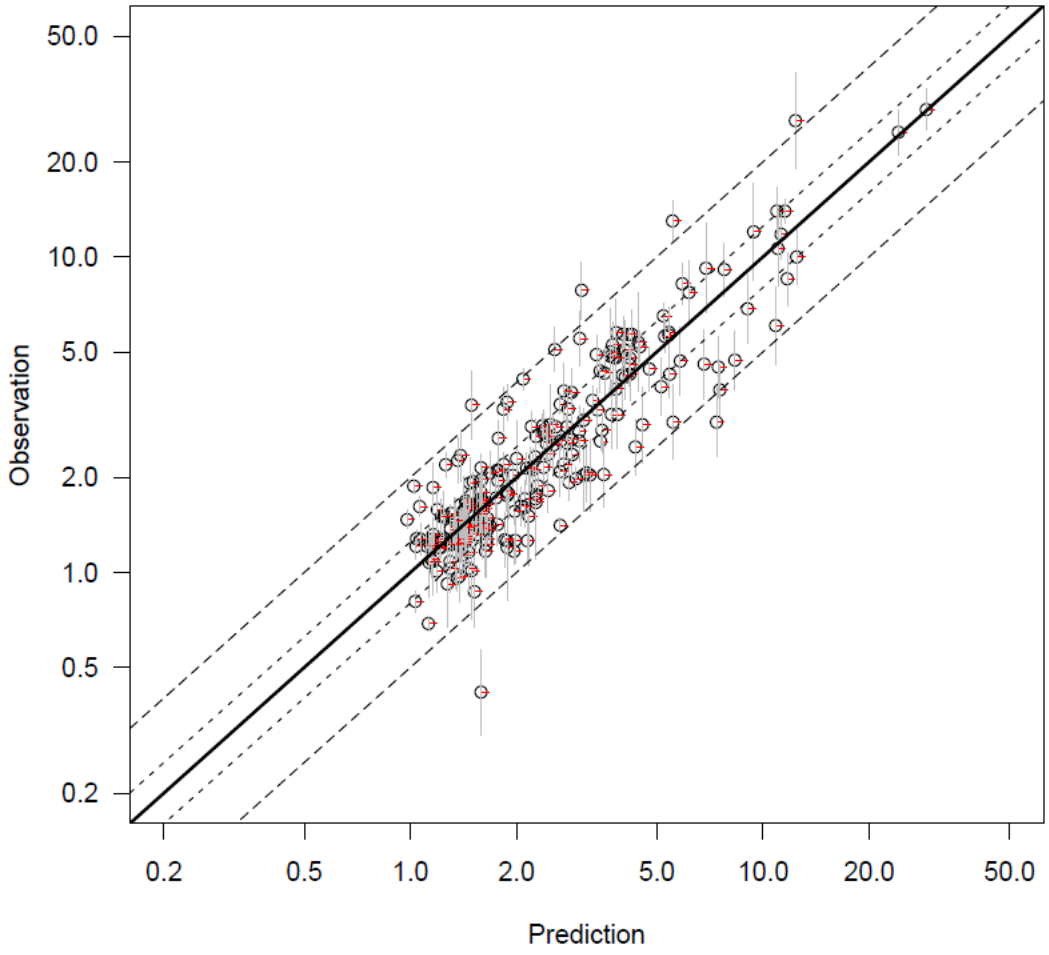


Figure 32: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *all the data*. The red line segments indicate the shift induced by bias correction.

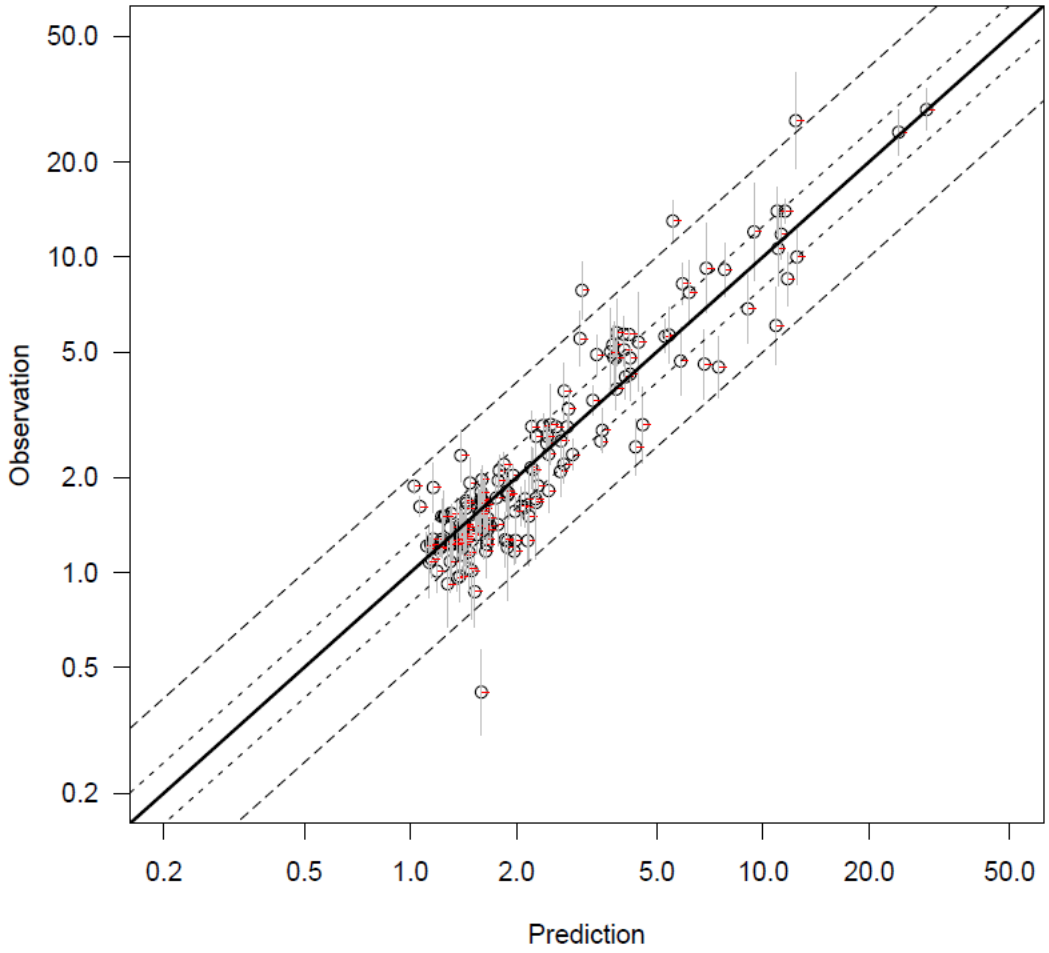


Figure 33: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for competitive inhibition*. The red line segments indicate the shift induced by bias correction.

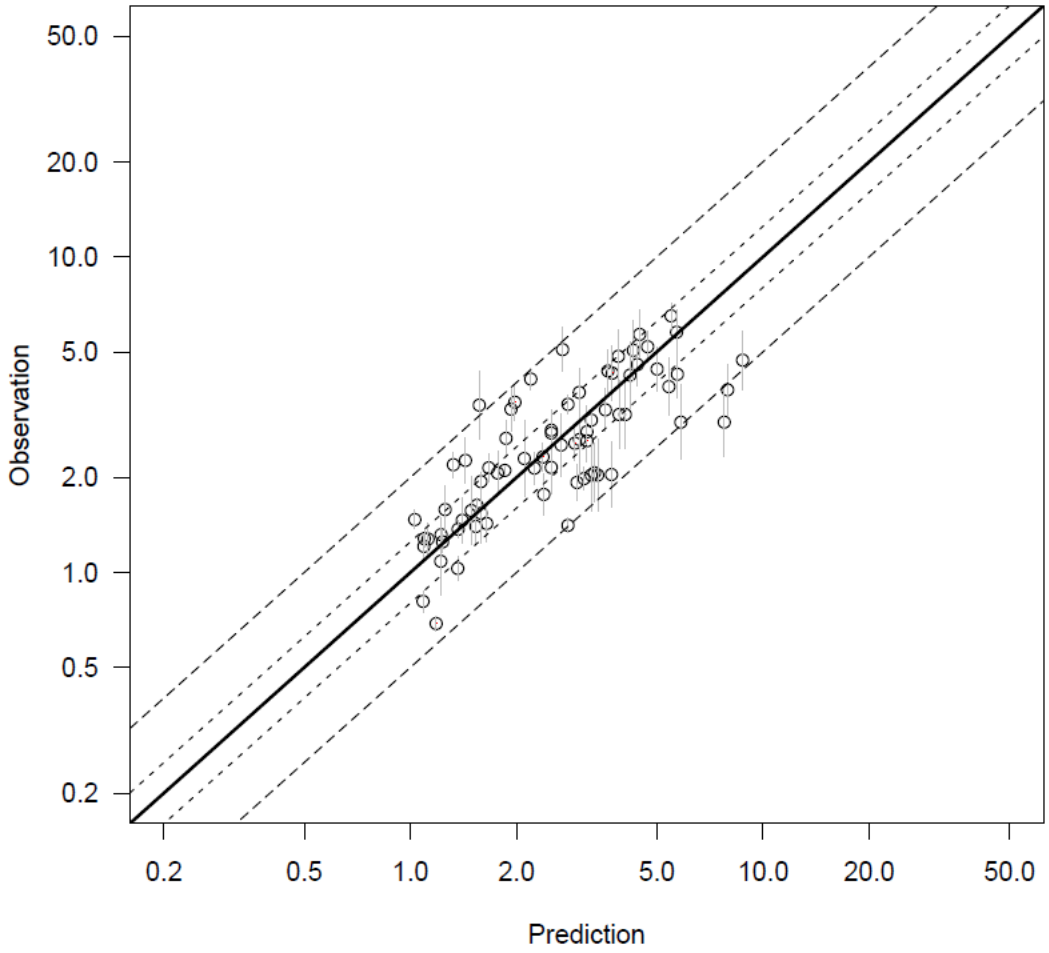


Figure 34: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for mechanism-based inhibition*. The red line segments indicate the shift induced by bias correction.

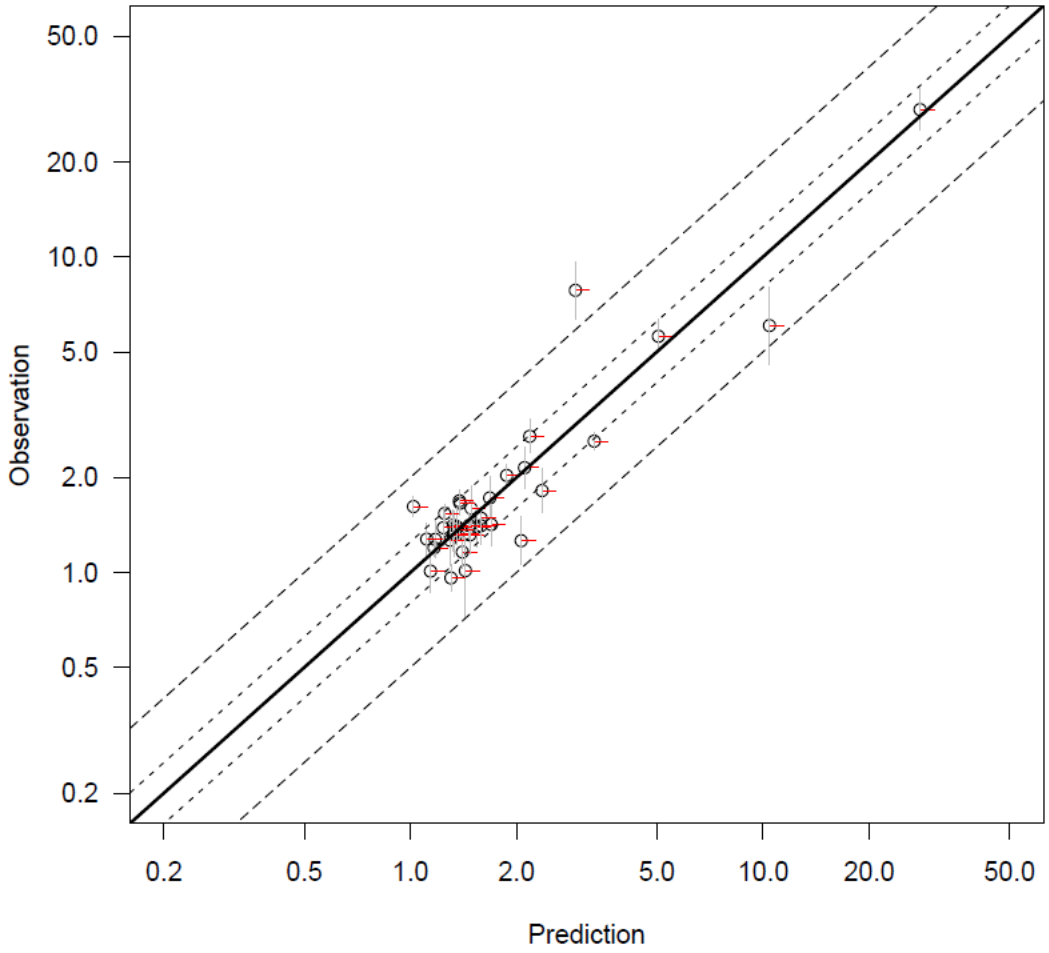


Figure 35: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP1A2*. The red line segments indicate the shift induced by bias correction.

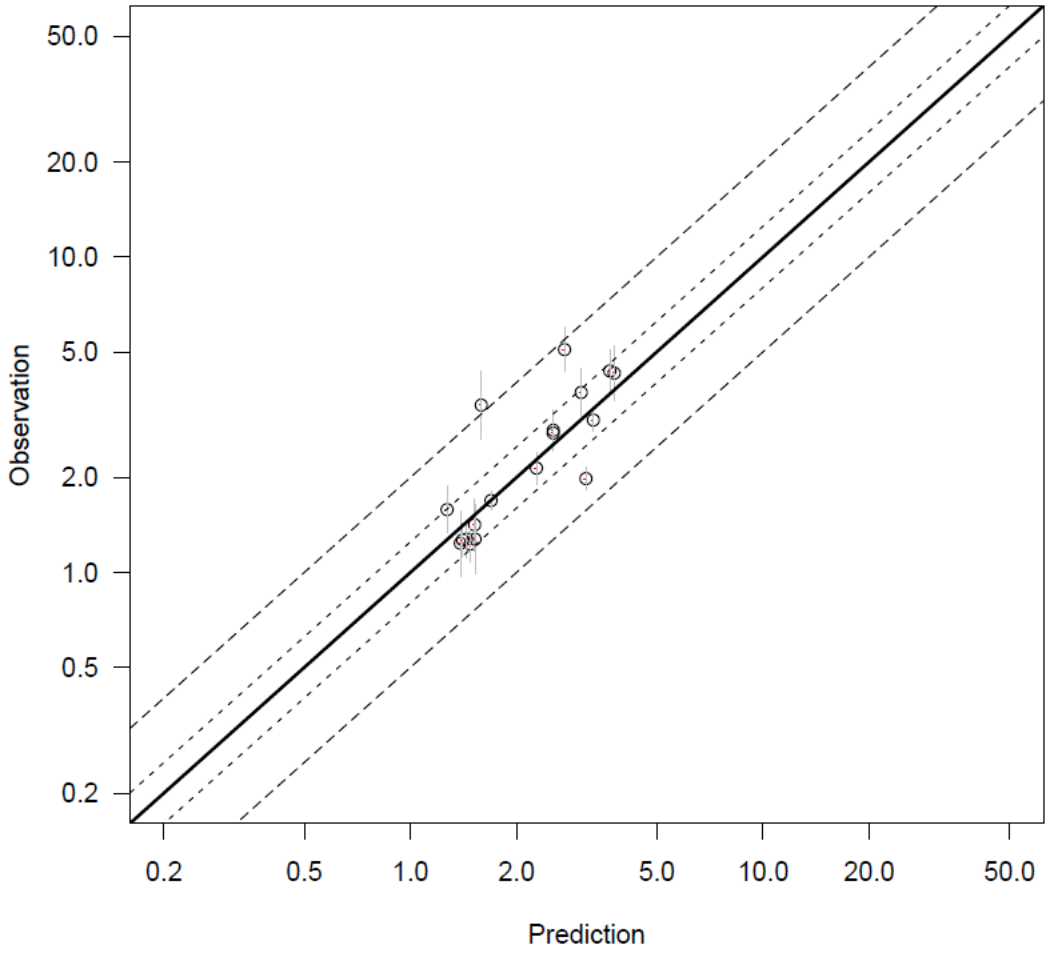


Figure 36: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP2C8*. The red line segments indicate the shift induced by bias correction.

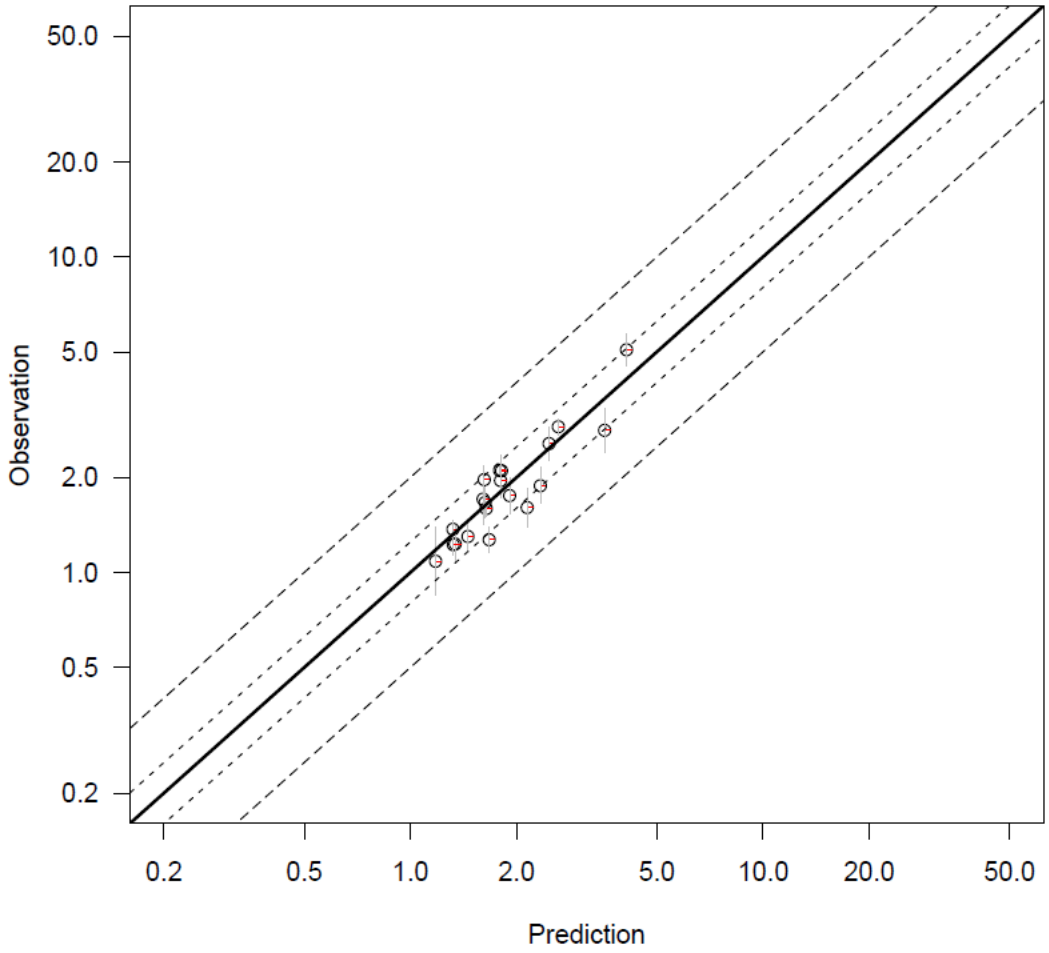


Figure 37: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP2C9*. The red line segments indicate the shift induced by bias correction.

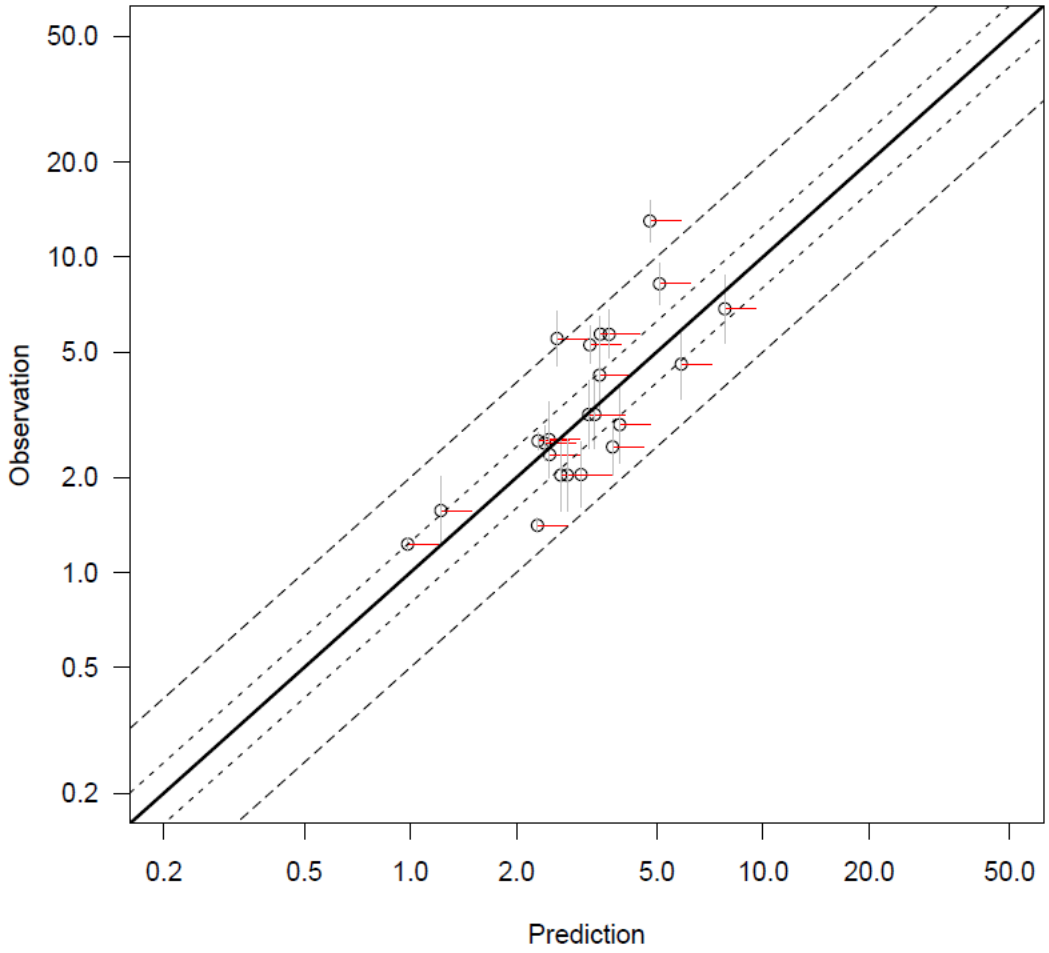


Figure 38: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP2C19*. The red line segments indicate the shift induced by bias correction.

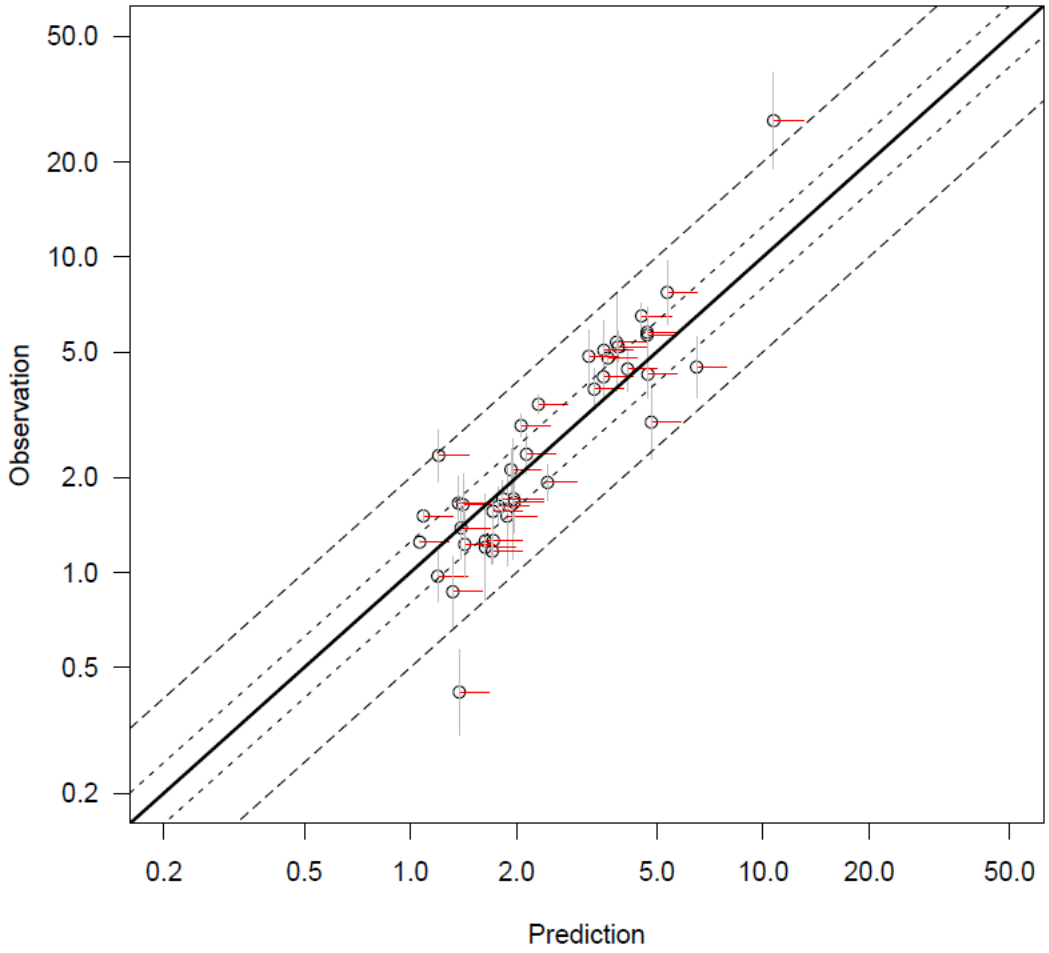


Figure 39: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP2D6*. The red line segments indicate the shift induced by bias correction.

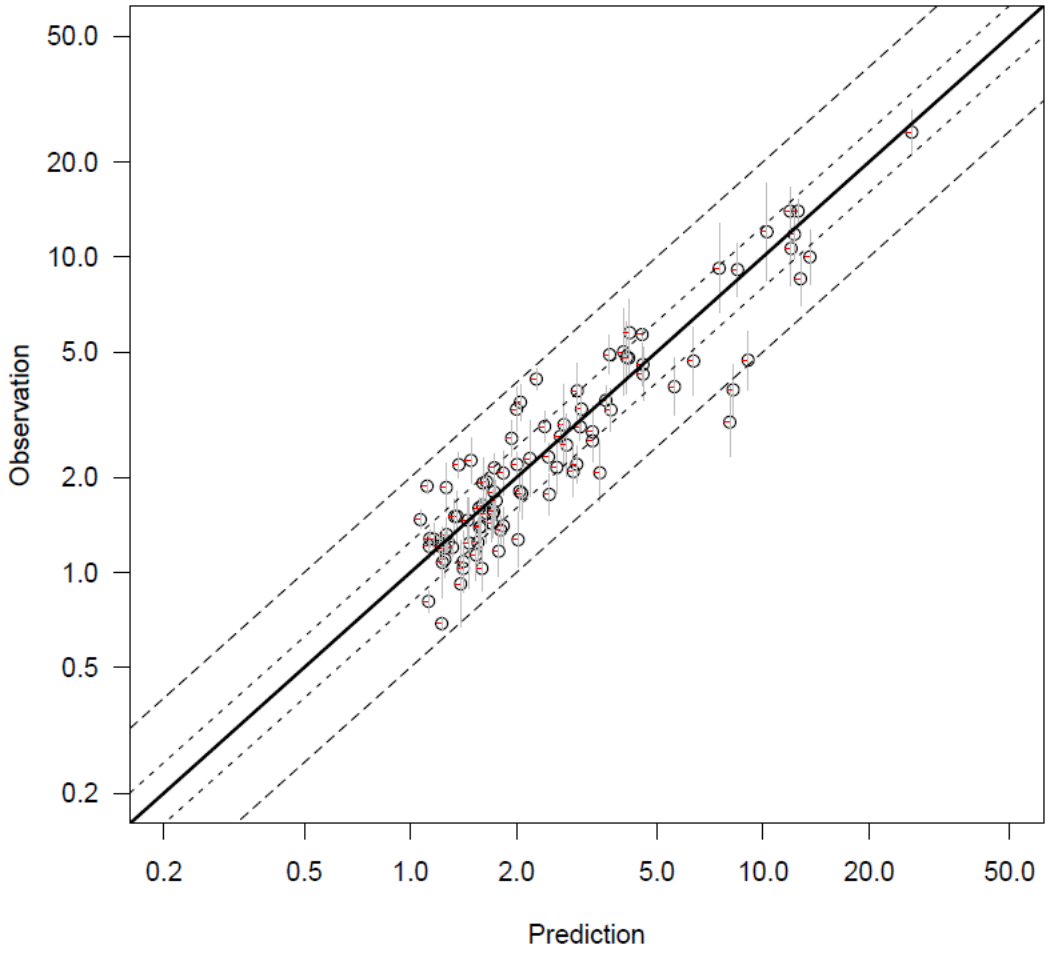


Figure 40: Observed *vs*. maximum posterior probability bias-corrected predictions of geometric mean ratios (with inhibitor over without inhibitor) for the analysis including *only the data for CYP3A4*. The red line segments indicate the shift induced by bias correction.

MCMC posterior between-subject variance estimates

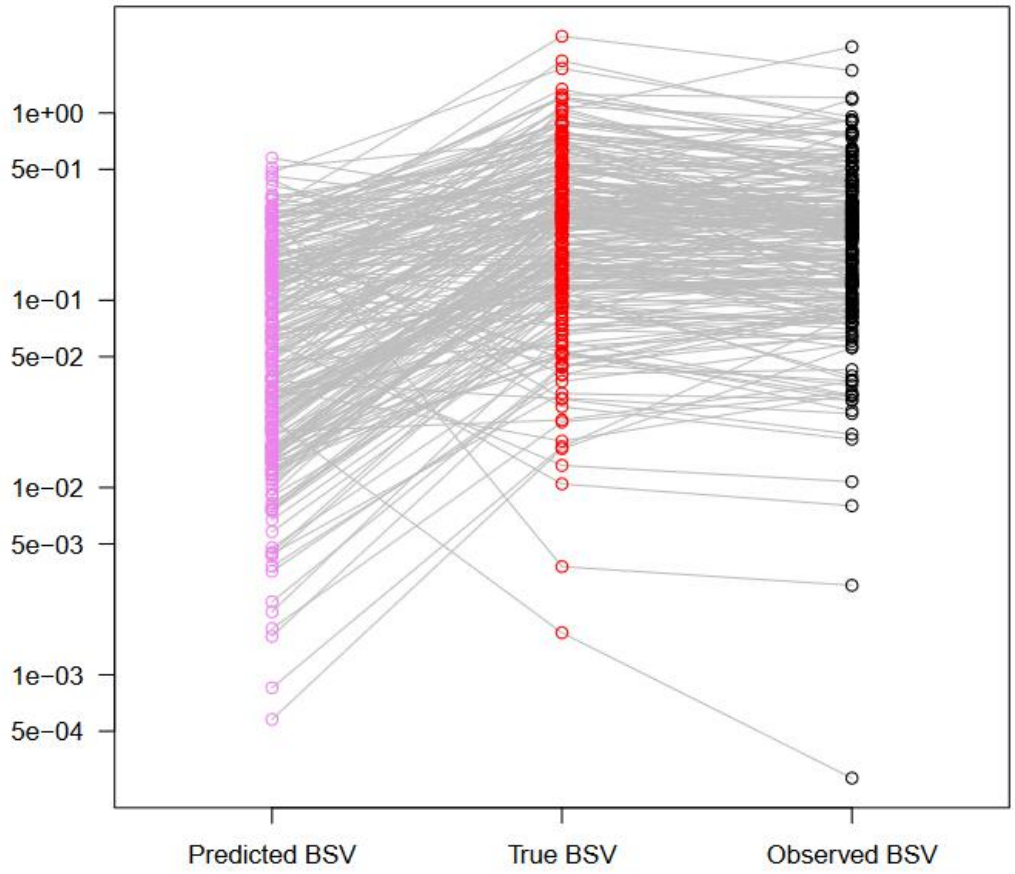


Figure 41: Simcyp-predicted ($\sigma^2_{sub,i,pred}$), true ($\sigma^2_{sub,i}$, bias corrected) and observed between-subject variances ($\sigma^2_{sub,i,obs}$, on the log scale) for the analysis including *all the data*. Bias $\mu_{\beta BSV}$ is on average by a factor 7.58.