DMD Fast Forward. Published on October 29, 2010 as doi:10.1124/dmd.110.036103
This article has not been copyedited and formatted. The final version may differ from this version.

Critique of the two-fold measure of prediction success for ratios: application for the assessment of drug-drug interactions

Eleanor J. Guest, Leon Aarons, J. Brian Houston, Amin Rostami-Hodjegan and Aleksandra Galetin

Centre for Applied Pharmacokinetic Research, School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom – E.J.G., L.A., J.B.H., A.R-H., A.G.

Simcyp Limited, Blades Enterprise Centre, John Street, Sheffield, S2 4SU, United Kingdom – A.R-H.

Running title: Critique of the two-fold measure of prediction success for ratios

Corresponding author: Dr A. Galetin

Centre for Applied Pharmacokinetic Research

School of Pharmacy and Pharmaceutical Sciences,

University of Manchester, Stopford Building

Oxford Road,

Manchester, M13 9PT, UK

Tel: (+) 44 161 275 6886

Fax: (+) 44 161 275 8349

Email: Aleksandra.Galetin@manchester.ac.uk

Number of text pages :	6
Tables:	2
Figures:	2
References:	22
Number of words in abstract:	249
Introduction:	561
Results and Discussion:	701

Abbreviations used: AUC, area under the concentration-time curve; CV, coefficient of variation; DDI, drug-drug-interactions; FDA, Food & Drug Administration; fu_p, fraction unbound in plasma; HVD, highly variable drug; K_i, inhibition constant

Abstract

Current assessment of drug-drug interaction (DDI) prediction success is based on whether predictions fall within a two-fold range of the observed data. This results in a potential bias towards successful prediction at lower interaction levels (ratio of the area under the concentration-time profile (AUC) in the presence of inhibitor/inducer compared to control is <2). This scenario can bias any assessment of different DDI prediction algorithms if databases contain large proportion of interactions in this lower range. Therefore, the current study proposes an alternative method to assess prediction success with a variable prediction margin dependent on the particular AUC ratio. The method is applicable for assessment of both induction and inhibition related algorithms. The inclusion of variability into this predictive measure is also considered using midazolam as a case study. Comparison of the traditional two-fold and the new predictive method was performed on a subset of midazolam DDIs collated from previous databases; in each case DDIs were predicted using dynamic model in Simcyp Simulator®. A 21% reduction in prediction accuracy was evident using the new predictive measure, in particular at the level of no/weak interaction (AUC ratio<2). However, inclusion of variability increased the prediction success at these levels by 2-fold. The trend of lower prediction accuracy at higher potency of DDIs reported in previous studies is no longer apparent when predictions are assessed via the new predictive measure. Thus, the study proposes a more logical method for the assessment of prediction success and its application for induction and inhibition DDIs.

Introduction

The current consensus for the in vitro-in vivo extrapolation of either clearance or drug-drug interactions (DDI) accepts prediction within a 2-fold (or occasionally 3-fold) range from the observed data as successful (Brown et al., 2006, Einolf, 2007, Galetin et al., 2006, Galetin et al., 2005, Teitelbaum et al., 2010, Wang, 2010). The commonly used metric to assess DDI is the ratio of the area under the plasma concentration-time curve (AUC) following multiple dosing of inhibitor or inducer in comparison to the control state (Fahmi et al., 2009, Houston and Galetin, 2008, Obach et al., 2006, Rostami-Hodjegan and Tucker, 2004). The assessment of different DDI algorithms involves retrospective prediction of *in vivo* studies and conclusions are often made following the separation of the predictions according to the *in vivo* DDI potency, analogous to the approach proposed by the FDA Guidelines for the classification of inhibitor potency (Bjornsson et al., 2003, Huang et al., 2007).

This study considers the importance of the two-fold criterion in the assessment of DDI prediction success. While a two-fold range may be appropriate for absolute values, the application of this to the prediction of a ratio has not been fully considered. Implications and importance of these considerations for DDI prediction success are discussed. This wide two-fold range at lower AUC ratio values can lead to false impression of high prediction accuracy and therefore a potential bias in prediction trends. For example, for an actual AUC ratio of 1 (classified as no interaction), the traditional two-fold measure accepts predicted ratios ranging from 0.5 (induction) to 2.0 (border of weak/moderate inhibition interaction), as successful. Many publications assessing DDI prediction accuracy have been based on databases containing almost half of studies with AUC ratios<2 (e.g. 42%: Einolf, 2007, 46%: Fahmi et al., 2009) and conclusions drawn may have been skewed by this proportion. This trend was evident in the analysis performed by Obach et al., where the inclusion of DDIs with <2-fold increase in AUC resulted in increased accuracy and precision of DDI prediction (Obach et al., 2006).

DMD Fast Forward. Published on October 29, 2010 as DOI: 10.1124/dmd.110.036103
This article has not been copyedited and formatted. The final version may differ from this version.

In addition, application of two-fold range at lower AUC ratio can lead to misclassification of DDI potential. Table 1 shows predicted AUC ratios for a range of midazolam DDIs (in all cases observed AUC ratio<2) obtained using the dynamic DDI prediction model in Simcyp Simulator®, as reported by Einolf (2007) and Fahmi et al., (2009). All DDIs were reported to be successfully predicted when assessed via the traditional two-fold measure. However, correct classification of the observed interaction (i.e., induction, no interaction or weak inhibition) was successfully predicted for less than 50% of the studies, often as a result of under-prediction of weak DDIs and subsequent classification as no interaction. The induction interaction with fluoxetine (AUC ratio 0.84) was predicted as weak inhibition (AUC ratio 1.28) and concluded as successful, despite this pertinent difference in classification.

Prediction of DDIs associated with highly variable drugs (HVDs) represents an additional concern. These victim drugs (e.g. chlorpromazine and cyclosporine (Shah et al., 1996)) have a high within-subject variability in either C_{max} and/or AUC (CV > 30%) (Davit et al., 2008, Tothfalusi et al., 2009) and a low observed AUC ratio in a DDI study could therefore be a result of either DDI or variability. The difference between the two is indistinguishable, emphasizing again that the prediction within traditional two-fold limits may be inadequate for this scenario. Therefore, this study proposes a new measure of prediction accuracy applicable for both induction and inhibition DDIs. In addition, this improved approach allows incorporation of the variability in pharmacokinetics of the victim drug when required.

Materials and Methods

The traditional two-fold predictive measure is bounded two-fold above and below the observed value: any prediction within these boundaries is classed as a successful prediction (see Figure 1). Therefore if the observed ratio, $AUC_{+inhibitor}/AUC_{control}$, is 1 the boundaries would be from 0.5 to 2.0. As was noted in the introduction, this is too wide for an interaction which in fact is not present. Consequently we propose new limits shown in Equations 1-3 below. The limits

coalesce when the observed ratio is one and approach the traditional two fold limits as the ratio becomes large (Figure 1).

Upper limit: Robs * Limit (1)

Lower limit: Robs / Limit (2)

$$\text{Limit}=\frac{1+2(\text{Robs}-1)}{\text{Robs}} \quad (3)$$

Where Robs represents $AUC_{+inhibitor}/AUC_{control}) \geq 1$, i.e. in the case of inhibition DDIs. The new predictive measure is also applicable for induction DDIs ($AUC_{+Inducer}/AUC_{control}<1$) if the reciprocal of the observed AUC ratio, $AUC_{control}/AUC_{+inducer}$, is used.

To allow for uncertainty in the observed ratio, the impact of variability was assessed by considering DDIs involving midazolam; a commonly used CYP3A4 victim drug (Bjornsson et al., 2003, Galetin et al., 2005). Upper and lower limits in this case are as defined in Equations 1 and 2, respectively, but the variability is now introduced into the limit as shown in Equation 4.

$$\text{Limit}=\frac{\delta+2(\text{Robs}-1)}{\text{Robs}} \quad (4)$$

Where δ is a parameter that accounts for variability. If $\delta=1$ there is no variability and limits revert to those defined by Equation 3. If $\delta=1.25$ and Robs=1 then the limits on R are between 0.80 and 1.25, corresponding to the conventional 20% limits used in bioequivalence testing (Food and Drug Administration (FDA), 2003). Note that these limits are symmetrical on the log scale. Assessment of the variability in the present study was based on approximately 20% CV reported for midazolam AUC (Kharasch et al., 1999, Kharasch et al., 2007) after i.v. dosing.

In order to assess the new predictive measure, DDI predictions were collated from 3 publications (Einolf, 2007, Fahmi et al., 2009, Guest et al., 2010) focusing on the prediction of DDIs involving midazolam as the victim drug. In all studies, predictions were obtained using the dynamic model in Simcyp Simulator® (n=89) and input parameters were as defined in the respective papers. Use of different parameter inputs (for example for K_i and fu_p) resulted in

DMD Fast Forward. Published on October 29, 2010 as DOI: 10.1124/dmd.110.036103
This article has not been copyedited and formatted. The final version may differ from this version.

different predictions even though around half of the DDIs overlapped between the three publications. Classification of the predicted DDI and the conclusions drawn in each study were compared using the conventional two-fold method and new measure of prediction accuracy. The impact of inclusion of the variability into the predictive measure was also assessed.

Results and Discussion

Figure 1 shows the differences in the limits of successful prediction for the traditional two-fold measure compared to the new predictive approach implemented using Equation 1-3; the corresponding observed data cover a 10-fold induction and inhibition range. The largest difference between methods is observed for AUC ratios ranging from 0.3-3, whereas the differences at 0.3 >AUC ratio> 3 are minimal (Figure 1). This is particularly important from a regulatory point of view, as it represents the distinction between a positive and negative DDI (AUC ratio ≥ 2) and therefore the decision on future follow-up clinical DDI trials will be based on the small scale studies and/or prediction from *in vitro* data using DDI models or prediction software such as Simcyp Simulator® (Hyland et al., 2008, Zhao et al., 2010).

The new proposed analysis in Figure 1 allows only a small deviation for successful prediction of AUC ratios at the level of no interaction (AUC ratio 1-1.25). However, this is the area where there may be deviation in the victim drug AUC as a result of variability. Variability reported for midazolam was incorporated as δ (1.25) into Equation 4; the limits obtained via this approach are shown in Figure 2A. Maximal impact of the variability is expected at the level of no interaction, whereas at higher AUC ratios the impact of variability is minimal in comparison to the increase in AUC ratio in the presence of an inhibitor and the limit approaches 2-fold.

Existing large DDI databases (Einolf, 2007, Fahmi et al., 2009, Guest et al., 2010) were used to assess the impact of this new predictive measure, focusing in particular on the analysis of the DDI prediction success involving midazolam as the victim drug (Figure 2B). Table 2 displays the prediction accuracy resulting from the traditional or the new predictive measure with or

without inclusion of the variability. Notable trends include the 21% reduction in the overall predictive accuracy using the new predictive measure compared to the traditional measure in all three studies; this was apparent in particular at the level of no or weak interactions (50-59% decrease in accuracy). The inclusion of variability into the new predictive measure resulted in a 2-fold increase in prediction accuracy for these particular studies. The overall difference for all studies was not as pronounced (12%) due to the low proportion of no and weak interactions considered in the subset (18/89).

The impact of the application of the new predictive measure on the conclusions previously made in the three publications was assessed. The overall conclusions on the performance of both static and dynamic models within the three publications did not change. However, all studies also reported the trend of reduced prediction accuracy and higher bias at higher potency/positive (AUC ratio ≥ 2) inhibition DDIs with both over- and under-predictions reported depending on the study. However, reanalysis with the new predictive measure shows a more consistent level of prediction accuracy across the different DDI potencies, with no clear relationship between DDI potency and prediction accuracy (Table 2). The initial trend of higher accuracy at the lower AUC ratios was likely to be due to the wide two-fold boundaries at this range based on the traditional DDI prediction measure.

The 20% value used here for the inclusion of variability was taken from the limits currently used in bioequivalence testing (FDA, 2003). This was in agreement with the reported variability in midazolam AUC (Kharasch et al., 1999, Kharasch et al., 2007). The CV used was based on i.v. dosing and would therefore exclude aspects of variability that may result after oral dosing - e.g., variability in intestinal first-pass (Galetin et al., 2008) and differences in GI tract physiology (e.g. gastric emptying) with the added impact of fasted/fed states in subjects (Shah et al., 1996). The use of 20% is proposed as a generic value when extending the methodology to other drugs in the absence of specific variability data.

Overall, this study critiques the traditional method used to assess predictive accuracy for ratios applied for drug-drug interactions. The proposed new methodology is appropriate for the assessment of ratios and allows tighter prediction boundaries for low AUC ratios, applicable across different interaction mechanisms (induction and inhibition). Importance of prediction accuracy and performance in the region below 2-fold change in AUC from a regulatory point of view has been addressed. In addition, this refined approach allows inclusion of variability into DDI predictions.

Authorship contributions

Participated in research design: Guest, Aarons, Houston, Rostami-Hodjegan, Galetin

Conducted experiments: Guest

Contributed new reagents or analytic tools:

Performed data analysis: Guest

Wrote or contributed to the writing of the manuscript: Guest. Aarons, Houston, Rostami-Hodjegan, Galetin.

References

Bjornsson TD, Callaghan JT, Einolf HJ, Fischer V, Gan L, Grimm S, Kao J, King SP, Miwa G, Ni L, Kumar G, McLeod J, Obach RS, Roberts S, Roe A, Shah A, Snikeris F, Sullivan JT, Tweedie D, Vega JM, Walsh J and Wrighton SA (2003) The Conduct of In Vitro and In Vivo Drug-Drug Interaction Studies: A Pharmaceutical Research and Manufacturers of America (PhRMA) Perspective. *Drug Metab Dispos* **31**: 815-832.

Brown HS, Galetin A, Hallifax D and Houston JB (2006) Prediction of in vivo drug-drug interactions from in vitro data. *Clin Pharmacokinet* **45**: 1035-1050.

Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PR, Suh K, Thompson CL and Yu LX (2008) Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications. *AAPS J* **10**: 148-56.

Einolf HJ (2007) Comparison of different approaches to predict metabolic drug-drug interactions. *Xenobiotica* **37**: 1257-94.

Fahmi OA, Hurst S, Plowchalk D, Cook J, Guo F, Youdim K, Dickins M, Phipps A, Darekar A, Hyland R and Obach RS (2009) Comparison of different algorithms for predicting clinical drug-drug interactions, based on the use of CYP3A4 in vitro data; predictions of compounds as precipitants of interaction. *Drug Metab Dispos* **37**: 1658-66.

Food and Drug Administration: Centre for Drug Evaluation and Research (March 2003) Guidance for Industry: Bioavailability and bioequivalence studies for orally administered products - General considerations.

Galetin A, Burt H, Gibbons L and Houston JB (2006) Prediction of time-dependent CYP3A4 drug-drug interactions: impact of enzyme degradation, parallel elimination pathways, and intestinal inhibition. *Drug Metab Dispos* **34**: 166-175.

Galetin A, Gertz M and Houston JB (2008) Potential role of intestinal first-pass metabolism in the prediction of drug-drug interactions. *Expert Opin Drug Metab Toxicol* **4**: 909-22.

Galetin A, Ito K, Hallifax D and Houston JB (2005) CYP3A4 substrate selection and substitution in the prediction of potential drug-drug interactions. *J Pharmacol Exp Ther* **314**: 180-190.

Guest EJ, Rowland Yeo K, Rostami-Hodjegan A, Tucker GT, Houston B and Galetin A (2010) Assessment of algorithms for predicting drug-drug interactions via inhibition mechanisms: Comparison of dynamic and static models. . *Br J Clin Pharmacol* **(in press)**.

Houston JB and Galetin A (2008) Methods for predicting in vivo pharmacokinetics using data from in vitro assays. *Curr Drug Metab* **9**: 940-51.

Huang SM, Temple R, Throckmorton DC and Lesko LJ (2007) Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther* **81**: 298-304.

Hyland R, Dickins M, Collins C, Jones H and Jones B (2008) Maraviroc: in vitro assessment of drug-drug interaction potential. *Br J Clin Pharmacol* **66**: 498-507.

Kharasch ED, Jubert C, Senn T, Bowdle TA and Thummel KE (1999) Intraindividual variability in male hepatic CYP3A4 activity assessed by alfentanil and midazolam clearance. *J Clin Pharmacol* **39**: 664-9.

Kharasch ED, Walker A, Isoherranen N, Hoffer C, Sheffels P, Thummel K, Whittington D and Ensign D (2007) Influence of CYP3A5 Genotype on the Pharmacokinetics and Pharmacodynamics of the Cytochrome P4503A Probes Alfentanil and Midazolam. *Clin Pharmacol Ther* **82**: 410.

Obach RS, Walsky RL, Venkatakrishnan K, Gaman EA, Houston JB and Tremaine LM (2006) The utility of in vitro cytochrome P450 inhibition data in the prediction of drug-drug interactions. *J Pharmacol Exp Ther* **316**: 336-348.

Rostami-Hodjegan A and Tucker G (2004) `In silico' simulations to assess the `in vivo' consequences of `in vitro' metabolic drug-drug interactions. *Drug Discov Today* **1**: 441-448.

Shah VP, Yacobi A, Barr WH, Benet LZ, Breimer D, Dobrinska MR, Endrenyi L, Fairweather W, Gillespie W, Gonzalez MA, Hooper J, Jackson A, Lesko LJ, Midha KK, Noonan PK, Patnaik R and Williams RL (1996) Evaluation of Orally Administered Highly Variable Drugs and Drug Formulations. *Pharm Res* **13**: 1590-1594.

Teitelbaum Z, Lave T, Freijer J and Cohen AF (2010) Risk assessment in extrapolation of pharmacokinetics from preclinical data to humans. *Clin Pharmacokinet* **49**: 619-32.

Tothfalusi L, Endrenyi L and Arieta AG (2009) Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. *Clin Pharmacokinet* **48**: 725-43.

Wang YH (2010) Confidence assessment of the Simcyp time-based approach and a static mathematical model in predicting clinical drug-drug interactions for mechanism-based CYP3A inhibitors. *Drug Metab Dispos* **38**: 1094-104.

Zhao P, Zhang L, Lesko LJ and Huang SM (2010) Utility of Physiologically-Based Pharmacokinetic Modeling and Simulation in Drug Development and Challenges in Regulatory Reviews. *Clin Pharmacol Ther* **87**: S68-S95.

Footnotes

Financial Support Credit

The work was funded by a consortium of pharmaceutical companies (GlaxoSmithKline, Lilly, Novartis, Pfizer and Servier) within the Centre for Applied Pharmacokinetic Research at the University of Manchester. E.J.G was financially supported by a Simcyp studentship.

Address correspondence to:

Dr A. Galetin, School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Oxford Rd, Manchester, M13 9PT, UK

Figure Legends

Figure 1: Schematic graph displaying the limits of the different predictive measures; the traditional 2-fold predictive measure (dashed lines) and the proposed new predictive measure (dotted lines). Observed AUC ratio include both induction and inhibition DDIs

Figure 2: A. Limits of DDI prediction with dashed lines representing the new predictive measure with inclusion of intra-individual variability, calculated via Equation 1 and 2, with the limits defined in Equation 4. **B.** Prediction of DDIs involving midazolam as the victim drug, taken from 3 publications, where ■ is Einolf et al., 2007, ▲ is Fahmi et al., 2009 and ○ is Guest et al., 2010. The new predictive measure and inclusion of intra-individual variability is utilised. Two induction DDIs are not shown (AUC ratio 0.04 and 0.05); both were successfully predicted with all methods. The vertical lines represent the limits between potency classifications, where I, NI, W, M and S represents induction, no interaction, weak, moderate and strong inhibition interaction, respectively.

DMD Fast Forward. Published on October 29, 2010 as DOI: 10.1124/dmd.110.036103
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 1

Accuracy in classification of midazolam DDIs (AUC ratio<2) based on predictions obtained using dynamic model. All studies were reported as successfully predicted when assessed via the traditional two-fold measure approach

Inhibitor	Actual AUC ratio and DDI classification		Correct classification of predicted DDI using dynamic model
Atomoxetine	1.0	NI	✓
Atorvastatin	1.4	W	× (NI)
Chlorzoxazone	1.7	W	× (NI)
Cimetidine	1.4	W	× (NI)
Cimetidine	1.4	W	× (NI)
Cimetidine	1.5	W	× (NI)
Fluconazole	1.9	W	✓
Flumazenil	0.97	I	× (NI)
Fluvoxamine	1.7	W	× (NI)
Gatifloxacin	1.1	NI	✓
Parecoxib	1.1	NI	× (I)
Ranitidine	1.2	NI	✓
Ranitidine	1.3	W	× (NI)
Ranitidine	1.7	W	× (NI)
Azithromycin	1.3	W	✓
Carbamazepine	0.04	I	✓
Rifampin	0.05	I	✓
Nitrendipine	0.93	I	× (NI)
Simvastatin	1.1	NI	✓
Terbinafine	0.76	I	× (NI)
Valdecoxib	1.1	NI	✓
Fluoxetine	0.84	I	× (W)
Fluvoxamine	1.39	W	× (NI)

I, NI and W represent induction, no interaction and weak inhibition interactions, respectively. ✓ represents the correct classification predicted and × represents incorrect prediction, with the incorrect classification in parenthesis.

TABLE 2

Prediction accuracy of 89 DDI studies involving midazolam as the victim drug, collated from 3 publications (Einolf, 2007, Fahmi et al., 2009, Guest et al., 2010). Prediction accuracy is assessed by the following methods: the traditional 2-fold measure, the new predictive measure and the new predictive measure with the incorporation of limits to allow for the variability of midazolam

Predictive Measure (Number of studies)	Induction (7)	No interaction (6)	Weak (12)	Moderate (29)	Strong (35)	Total (89)
Traditional 2-fold measure	86%	100%	92%	83%	80%	**84%**
New predictive measure	57%	50%	33%	66%	74%	**63%**
New predictive measure + variability	71%	100%	67%	72%	77%	**75%**

Figure 1

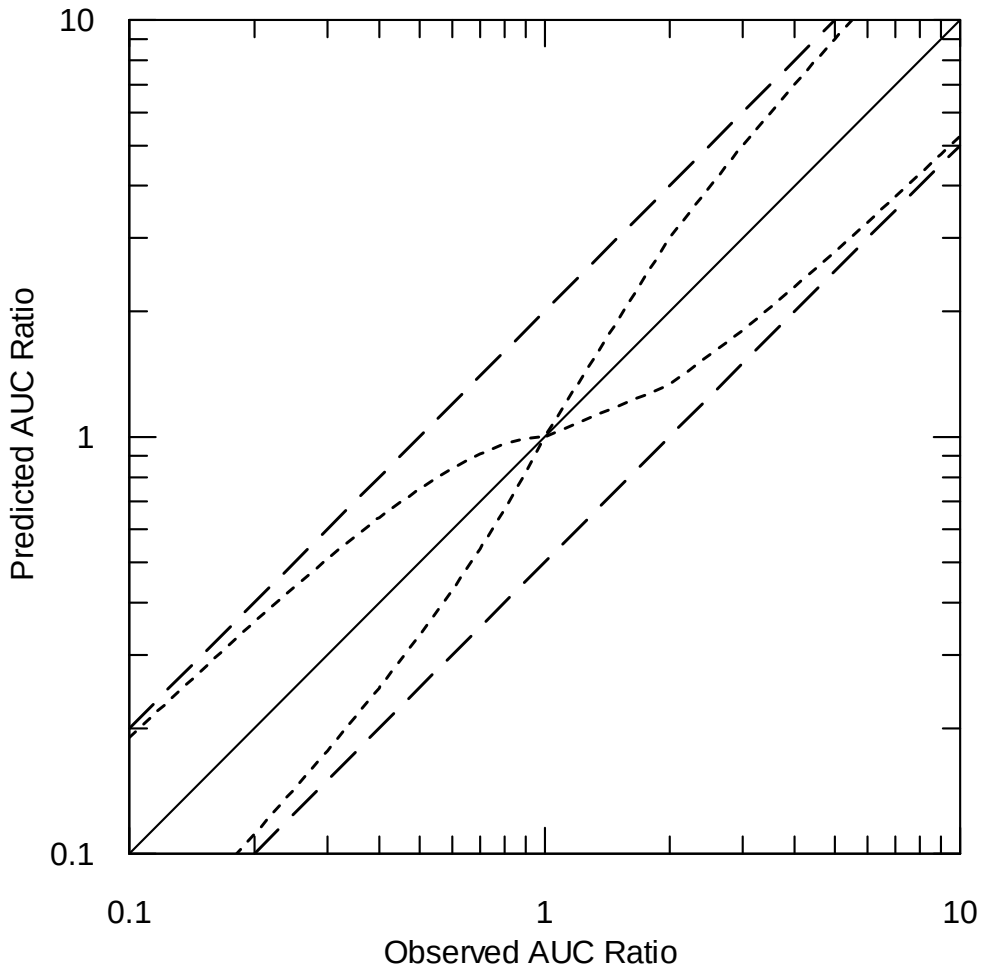

Figure 2