Short Communication

Critique of the Two-Fold Measure of Prediction Success for Ratios: Application for the Assessment of Drug-Drug Interactions

Received August 27, 2010; accepted October 25, 2010

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 strategy results in a potential bias toward successful prediction at lower interaction levels [ratio of the area under the concentration-time profile (AUC) in the presence of inhibitor/inducer compared with 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 the 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 two-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 two-fold (or occasionally three-fold) range from the observed data as successful (Galetin et al., 2005, 2006; Brown et al., 2006; Einolf, 2007; 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) after multiple dosing of inhibitor or inducer in comparison to the control state (Rostami-Hodjegan and Tucker, 2004; Obach et al., 2006; Houston and Galetin, 2008; Fahmi et al., 2009). The assessment of different DDI algorithms involves retrospective prediction of in vivo studies, and conclusions are often made after the separation of the predictions according to the in vivo DDI potency, analogous to the approach proposed by the United States Food and Drug Administration 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. Although a two-fold range may be appropriate for absolute values, the application of this method 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 AUC ratios approaching 1 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 the studies with AUC ratios <2 [e.g., 42% (Einolf, 2007) and 46% (Fahmi et al., 2009)], and the conclusions drawn may have been skewed by this proportion. This trend was evident in the analysis performed by Obach et al. (2006), where the inclusion of DDIs with <2-fold increase in AUC resulted in apparent increased accuracy and precision of DDI prediction.

In addition, application of two-fold range at AUC ratio around 1 can lead to misclassification of DDI potential. Table 1 shows predicted AUC ratios for a range of midazolam DDIs (in all cases, the observed AUC ratio was <2), which were obtained using the dynamic DDI prediction model in Simcyp simulator (Simcyp Ltd., Sheffield, UK), 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 underprediction of weak DDIs and subsequent classification as no interaction. The induction interaction with fluoxetine (AUC ratio, 0.84) was predicted as weak inhibition

ABBREVIATIONS: DDI, drug-drug interaction; AUC, area under the concentration-time curve; CV, coefficient of variation; K_i, inhibition constant; f_u, fraction unbound in plasma.
AUC ratio (1.28) and concluded as successful, despite this pertinent difference in classification.

Prediction of DDIs associated with highly variable drugs represents an additional concern. These victim drugs [e.g., chlorpromazine and cyclosporine (Shah et al., 1996)] have a high within-subject variability in either Cmax and/or AUC (CV >30%) (Davit et al., 2008; Tothfalusi et al., 2009), and a change in observed AUC 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 Fig. 1). Therefore, if the observed ratio, \( \frac{\text{AUC}_{\text{inducer}}}{\text{AUC}_{\text{control}}} \), is 1, the boundaries would be from 0.5 to 2.0. As noted in the Introduction, this range is too wide for an interaction, which is in fact not present. As a result, we propose new limits, as shown in eqs. 1 to 3.

\[
\text{Upper limit: } R_{\text{obs}}^* \text{ Limit} = \text{AUC}_{\text{inducer}} / \text{AUC}_{\text{control}} < 1
\]

\[
\text{Lower limit: } R_{\text{obs}} / \text{Limit} = 1 / \text{AUC}_{\text{inducer}} / \text{AUC}_{\text{control}} > 1
\]

\[
\text{Limit} = 1 + 2(R_{\text{obs}} - 1)
\]

where \( R_{\text{obs}} \) represents AUC\(_{\text{inducer}} / \text{AUC}_{\text{control}} \) ≥ 1, i.e., in the case of inhibition DDIs. The new predictive measure is also applicable for induction DDIs (AUC\(_{\text{inducer}} / \text{AUC}_{\text{control}} < 1 \) if the reciprocal of the observed AUC ratio, \( \text{AUC}_{\text{control}} / \text{AUC}_{\text{inducer}} \), 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). In this case, upper and lower limits are as defined in eqs. 1 and 2, respectively, but the variability is now introduced into the limit as shown in eq. 4.

\[
\text{Limit} = \frac{\delta + 2(R_{\text{obs}} - 1)}{R_{\text{obs}}}
\]

where \( \delta \) is a parameter that accounts for variability. If \( \delta = 1 \), there is no variability and limits revert to those defined by eq. 3. If \( \delta = 1.25 \) and \( R_{\text{obs}} = 1 \), then the limits on R are between 0.80 and 1.25, corresponding to the conventional 20% limits used in bioequivalence testing (United States Food and Drug Administration, 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, 2007) after intravenous dosing.

To assess the new predictive measure, DDI predictions were collated from three 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. The use of different parameter inputs (for example, for \( K_i \) and \( f_d \)) resulted in different predictions, even though approximately half of the DDIs overlapped among 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 eqs. 1 to 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 to 3, whereas the differences at 0.3 > AUC ratio >3 are minimal (Fig. 1). This result is particularly important from a regulatory point of view, because it represents the distinction between a positive and

---

**TABLE 1**

Accuracy in classification of midazolam DDIs (AUC ratio <2) based on predictions obtained using the dynamic model

All studies were reported as successfully predicted when assessed via the traditional two-fold measure approach.

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Actual AUC</th>
<th>Correct Classification of Predicted DDI Using Dynamic Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ratio and DDI Classification</td>
<td></td>
</tr>
<tr>
<td>Atomoxetine</td>
<td>1.0</td>
<td>NI</td>
</tr>
<tr>
<td>Atorvastatin</td>
<td>1.4</td>
<td>W</td>
</tr>
<tr>
<td>Chlorzoxanone</td>
<td>1.7</td>
<td>W</td>
</tr>
<tr>
<td>Cimetidine</td>
<td>1.4</td>
<td>W</td>
</tr>
<tr>
<td>Cimetidine</td>
<td>1.4</td>
<td>W</td>
</tr>
<tr>
<td>Cimetidine</td>
<td>1.5</td>
<td>W</td>
</tr>
<tr>
<td>Fluconazole</td>
<td>1.9</td>
<td>W</td>
</tr>
<tr>
<td>Flumazenil</td>
<td>0.97</td>
<td>I</td>
</tr>
<tr>
<td>Fluvoxamine</td>
<td>1.7</td>
<td>W</td>
</tr>
<tr>
<td>Gatifloxacin</td>
<td>1.1</td>
<td>NI</td>
</tr>
<tr>
<td>Parecoxib</td>
<td>1.1</td>
<td>NI</td>
</tr>
<tr>
<td>Ranitidine</td>
<td>1.2</td>
<td>NI</td>
</tr>
<tr>
<td>Ranitidine</td>
<td>1.3</td>
<td>W</td>
</tr>
<tr>
<td>Ranitidine</td>
<td>1.7</td>
<td>W</td>
</tr>
<tr>
<td>Azithromycin</td>
<td>1.3</td>
<td>W</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>0.04</td>
<td>I</td>
</tr>
<tr>
<td>Rifampin</td>
<td>0.05</td>
<td>I</td>
</tr>
<tr>
<td>Nitrendipine</td>
<td>0.93</td>
<td>I</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>1.1</td>
<td>W</td>
</tr>
<tr>
<td>Terbinafine</td>
<td>0.76</td>
<td>I</td>
</tr>
<tr>
<td>Valdecoxib</td>
<td>1.1</td>
<td>NI</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>0.84</td>
<td>I</td>
</tr>
<tr>
<td>Fluvoxamine</td>
<td>1.39</td>
<td>W</td>
</tr>
</tbody>
</table>

1. 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.
negative DDI (AUC ratio $\geq 2$), 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 Fig. 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 $\delta$ (1.25) into eq. 4; the limits obtained via this approach are shown in Fig. 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 two-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 (Fig. 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 result was apparent particularly 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 two-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 of 89).

The impact of the new predictive measure on 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 reported the trend of reduced prediction accuracy and higher bias at higher potency/positive (AUC ratio $\geq 2$) inhibition DDIs. 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 (United States Food and Drug Administration, 2003). This value was in agreement with the reported variability in midazolam AUC (Kharasch et al., 1999, 2007). The CV used was based on intravenous 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 gastrointestinal 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

**TABLE 2**

Prediction accuracy of 89 DDI studies involving midazolam as the victim drug, collated from three publications (Einolf, 2007; Fahmi et al., 2009; Guest et al., 2010)

Prediction accuracy is assessed by the following methods: the traditional two-fold measure, the new predictive measure, and the new predictive measure with the incorporation of limits to allow for the variability of midazolam.

<table>
<thead>
<tr>
<th>Predictive Measure (Number of Studies)</th>
<th>Induction (7)</th>
<th>No Interaction (6)</th>
<th>Weak (12)</th>
<th>Moderate (29)</th>
<th>Strong (35)</th>
<th>Total (89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional two-fold measure</td>
<td>86</td>
<td>100</td>
<td>92</td>
<td>83</td>
<td>80</td>
<td>84</td>
</tr>
<tr>
<td>New predictive measure</td>
<td>57</td>
<td>50</td>
<td>33</td>
<td>66</td>
<td>74</td>
<td>63</td>
</tr>
<tr>
<td>New predictive measure + variability</td>
<td>71</td>
<td>100</td>
<td>67</td>
<td>72</td>
<td>77</td>
<td>75</td>
</tr>
</tbody>
</table>
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). The importance of prediction accuracy and performance in the region below two-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, and Galetin.
- Conducted experiments: Guest.
- Performed data analysis: Guest.
- Wrote or contributed to the writing of the manuscript: Guest, Aarons, Houston, Rostami-Hodjegan, and Galetin.

Centre for Applied Pharmacokinetic Research,
School of Pharmacy and Pharmaceutical Sciences,
University of Manchester, Manchester, United Kingdom (E.J.G., L.A., J.B.H., A.R.-H., A.G.); and Simcyp Limited, Blades Enterprise Centre, Sheffield, United Kingdom (A.R.-H.)

References


Address correspondence to: 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. E-mail: aleksandra.galetin@manchester.ac.uk