DRUG–DRUG INTERACTION PREDICTION ASSESSMENT

Jihao Zhou, Zhaohui Qin, Quinney K. Sara, Seongho Kim, Zhiping Wang, Stephen D. Hall, and Lang Li
1Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA
2Department of Medicine, Division of Biostatistics, School of Medicine, Indiana University, Indianapolis, Indiana, USA
3Department of Medicine, Division of Clinical Pharmacology, School of Medicine, Indiana University, Indianapolis, Indiana, USA

Model-based drug–drug interaction (DDI) is an important in-silico tool to assess the in vivo consequences of in vitro DDI. Before its general application to new drug compounds, the DDI model is always established from known interaction data. For the first time, tests for difference and equivalent tests are implemented to compare reported and model-base simulated DDI (log AUCR) in the sample mean and variance. The biases and predictive confidence interval coverage probabilities are introduced to assess the DDI prediction performance. Sample size and power guidelines are developed for DDI model simulations. These issues have never been discussed in trial simulation studies to investigate DDI prediction. A ketoconazole (KETO)/midazolam (MDZ) example is employed to demonstrate these statistical methods. Based on published KETO and MDZ pharmacokinetics data and their in vitro inhibition rate constant data, current model-based DDI prediction underpredicts the area under concentration curve ratio (AUCR) and its between-subject variance compared to the reported study.

Key Words: Area under the concentration curve ratio (AUCR); Drug–drug interaction (DDI); Equivalence test; Pharmacokinetics; Simulation.

1. INTRODUCTION

Pharmacokinetic (PK) approaches designed to characterize drug absorption, distribution, and elimination are well established (Rowland and Tozer, 1995) and robust statistical methodologies have been developed (Davidian and Giltinan, 1995). Recently drug–drug pharmacokinetic interactions have received a great deal of attention because this phenomenon makes a significant contribution to adverse drug reaction profile of new drugs (Ito et al., 1998). The importance of drug–drug interactions (DDI) is exemplified by the interaction of ketoconazole and terfenadine, which caused potentially life-threatening ventricular arrhythmias (Monahan et al., 1990), and the interaction between sorivudine and fluorouracil, which resulted in fatal toxicity (Okuda et al., 1997; Watabe, 1996).

Received March 16, 2008; Accepted October 3, 2008
Address correspondence to Lang Li, Associate Professor, 410 West 10th Street, Suite 3000, Indianapolis, IN 46202, USA; E-mail: lali@iupui.edu

Downloaded By: [Wyeth Canada Inc] At: 01:49 5 August 2009
In the DDI research, one of the central questions is whether two individual drugs’ PK models and their in vitro DDI parameters can predict their in vivo DDI (Ito et al., 1998). There are extensive pharmacology literatures that discuss assumptions and challenges of in silico simulations to assess the in vivo consequences of in vitro DDI. One good review was provided by Rostami-Hodjegan and Tucker (2004). The strategy of establishing a general DDI model is to train the model based on known DDI data before predicting DDI for new compounds. Because DDI is usually metabolic enzyme based, most of the DDI models are developed from a single enzyme, with which the two drugs’ interaction is solely involved. For example, in understanding CY3A inhibition, pharmacologists often use ketoconazole (KETO) as the inhibitor and midazolam (MDZ) as the substrate. If a new compound is a CYP3A inhibitor, this model can predict its inhibition on MDZ; and if a new compound is a CYP3A substrate, this model can predict its inhibition by KETO.

In the most recent CYP3A DDI model development, Chien et al. (2006) promoted the importance of integrating sources of quantitative errors, such as between-subject variance among PK parameters for both substrate and inhibitor, into the DDI prediction. Compared to the early deterministic DDI work (Ito et al., 1998), this stochastic approach simulates the observed DDI at the clinical setting, in which there is a considerable amount of DDI variation between subjects. To validate their KETO/MDZ DDI model, Chien et al. (2006) graphically displayed the overlap between confidence intervals of simulated DDI distributions and published clinical DDI distributions.

This visual assessment of the difference or equivalence between stochastic model predicted DDI and the published DDI is a naïve, though practically appealing, demonstration. However, it is neither a probabilistic assessment nor a hypothesis test. The questions we would like to address include

1) How can we assess the accuracy of the predicted DDI related to the observed DDI through simulations?
2) Do we have enough evidences to infer that the observed DDI and the simulated DDI are equivalent, or different?
3) How many samples shall we simulate through the DDI model to assess their predictive performance?

To answer the first question, relative biases in mean and variances and confidence interval coverage probability between two DDI distributions can be utilized. They are well established and accepted statistical criteria.

- The second question is fundamentally a goodness/lack of fit problem. In the statistical literatures, there are a few strategies to solve it. One approach is cross-validation, in which the model is established from the training sample, and its prediction performance is assessed in the validation sample. Although training and validation samples may have different sizes, they share the same data structure, in other words they have both predictors and outcomes. However, in the DDI prediction scenario, training and validation data structures are totally different. The training sample includes the individual drug’s in vivo data and its in vitro interaction data. The validation sample is the in vivo interaction data, and the sample size is highly limited. In this paper, there is only one published in vivo
DDI mean and its SD serves as the validation sample. Therefore, cross-validation can’t be implemented. Certainly, if more real studies were published with data, the cross-validation would be ideal.

- Another approach is the Kolmogorov–Smirnov (KS) test, which can detect the difference between two distributions, i.e., the difference between predicted and published DDI distributions. This was one of the first methods we screened. The problem was that the KS method tests the inequality of overall distribution by comparing the empirical cumulative distribution functions. For instance, in the normal distribution case, it tests the overall discrepancy between the distributions resulting from the mean difference and/or variance difference, but the KS test can’t tell where the difference comes from, i.e., from mean or variance.

- Another approach is the assessment of prediction errors. It is often used for outlier detection. In our DDI prediction problem, it is the difference between predicted and published DDI distributions that needs to be tested. In the outlier detection, we test the difference between a point and a distribution. The other approach to assess the prediction error is to evaluate the predictive biases and coverage probabilities. They are addressed in our response to the second comment.

In this paper, we will employ the conventional unequal variance \( t \)-test and \( F \)-test to test the equivalence or difference in mean and variance, respectively, between predicted and published DDI distributions. The test for difference provides evidence that the in vitro DDI model cannot predict in vivo DDI, i.e., lack of fit. The equivalence test looks for goodness of fit evidence. We use the posterior distributions of KETO and MDZ PK parameters and their between-subject variances derived from their Bayesian hierarchical population PK models (Li et al., 2007; Yu et al., 2008) to draw their subject-specific PK parameters, and we feed them into the interaction model to simulate their DDI. To assess this model-based DDI, its prediction is compared to a published DDI in both mean and variance. Our contributions include the maximum power derivation, minimum sample size calculation, tests of difference/equivalence implementation, and prediction performance assessment.

The DDI model-based prediction and its difference/equivalence test procedures and prediction performance assessment are developed in Section 2. The KETO/MDZ DDI example is demonstrated in Section 3. The discussion of the significance of our research and the application of model-based DDI simulation in clinical settings are provided in Section 4.

2. METHODS

2.1. Pharmacokinetics Model-Based KETO/MDZ DDI

The pharmacokinetics model for KETO/MDZ interaction was introduced in great detail in two recent publications (Li et al., 2007; Yu et al., 2008). In summary, KETO followed an oral administration, MDZ followed an IV route, both drugs were assumed to have two-compartment model kinetics (Eqs. (A1) and (A2) in Li et al., 2007), and their interaction was based on an inhibition assumption. The KETO-MDZ pair is also denoted as an inhibitor-substrate pair.
Denote $AUC_{S,W}(\beta, Ki)$ as the substrate (MDZ) area under concentration curve (AUC) with inhibitor, where $\beta$ is the PK parameter vector and $Ki$ is the inhibition constant. Denote $AUC_{S,WO}(\beta)$ as the substrate AUC without inhibitor (KETO). For any given set of PK parameters and their dose-combination, their $AUC_{S,W}(\beta, Ki)$ and $AUC_{S,WO}(\beta)$ can be calculated via the trapezoidal rule (Rowland and Tozer, 1995). A common criterion to evaluate the extent of interaction is the area under concentration ratio (AUCR) of substrate after and before inhibitor administration, $AUCR(\beta, Ki) = \frac{AUC_{S,W}(\beta, Ki)}{AUC_{S,WO}(\beta)}$.

### 2.2. Subject Specific AUCR Posterior Distributions

In the recent literature (Yu et al., 2008), a Bayesian hierarchical meta-analysis population PK model has been developed to estimate MDZ and KETO PK parameter vector, $\beta$, and their between-subject variances, $\Omega$. Posterior distributions were derived through Monte Carlo Markov chains (MCMC), and their 90% credit intervals were presented in Table 2 of Yu et al. (2008).

The key to estimate $AUCR(\beta, Ki)$ is to estimate $AUC_{S,W}(\beta, Ki)$ and $AUC_{S,WO}(\beta)$ of MDZ. When these two AUC are estimated with subject-specific parameter $\beta_i$, $AUCR(\beta_i, Ki)$ denotes the subject-specific DDI. We obtain samples of $AUCR(\beta_i, Ki)$ by first drawing $\beta$ and $\Omega$ from their posterior draws from Yu et al. (2008). That is, given any specific draw, $(\beta^{(\alpha)}, \Omega^{(\alpha)})$, we draw $\beta_i^{(\alpha)}$ from $[\beta_i \mid \beta^{(\alpha)}, \Omega^{(\alpha)}] = MVN(\beta^{(\alpha)}, \Omega^{(\alpha)})$. Corresponding to any draw of $\beta_i^{(\alpha)}$, a sample of predicted $AUCR(\beta_i^{(\alpha)}, Ki)$ can be obtained. The goal is to test the equivalence of predicted subject-specific AUCR to the reported subject-specific AUCR from clinical trials.

### 2.3. DDI Test for Difference in Sample Mean

#### 2.3.1. Test Statistic Specification.

A test for difference between the predicted log AUCR from DDI model and observed study log AUCR is formulated with the two-sample test. It is assumed that the $n_1$ samples of the observed study data (log AUCR$_{obs}$) are normally distributed with mean $\mu_1$ and variance $\sigma_1^2$, and $n_2$ samples of DDI model-simulated data (log AUCR$_{model}$) are normally distributed with mean $\mu_2$ and variance $\sigma_2^2$. The test for difference in mean, $\Delta = |\mu_1 - \mu_2|$, can be expressed as the following:

$$H : \Delta = 0 \quad \text{vs.} \quad H : \Delta \neq 0$$

An unequal variance $T$-statistic, Eq. (2.1) is chosen for this test

$$T = \frac{\log \overline{\text{AUCR}}_{\text{model}} - \log \overline{\text{AUCR}}_{\text{obs}}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

(2.1)

where $S_1^2$ and $S_2^2$ are sample variances for observed and model simulated log AUCR, respectively. Under Eq. (2.1), a level $\alpha$ rejection region is

$$\{|T| > t_{1-\alpha/2, n_1+n_2-2}\}$$

(2.2)
where $t_{1-\alpha/2}$ is the $(1 - \alpha/2)*100\%$ percentile of the $t$-distribution function with $v$ df, and $v$ is defined in Eq. (2.3) (Snedecor and Cochran, 1980)

$$v = \frac{1}{C^2/(n_1 - 1) + (1 - C)^2/(n_2 - 1)}$$

$$C = (S_1^2/n_1)/(S_1^2/n_1 + S_2^2/n_2).$$

### 2.3.2. Maximum Power.

The sample size in the reported study, $n_1$, is always limited and fixed, whereas simulated sample size $n_2$ can be arbitrarily large in theory. Hence the equivalent test’s maximum power can be estimated in this sense. When $n_2$ goes to infinity, the $T$ statistic from Eq. (2.1) becomes

$$T = \frac{(\log AUCR_{obs} - \mu_{model})}{\sqrt{S_1^2/n_1}}$$

Its rejection region becomes

$$\{||T| > t_{1-\alpha/2,n_1-1}\}$$

(2.5)

If the true difference equals the observed $\Delta_{obs} = |\log AUCR_{model} - \log AUCR_{obs}|$, the maximum power the $t$-test is

$$\text{max power} = 1 - T_{n_1-1}\left[t_{1-\alpha/2,n_1-1} - \Delta_{obs}/\sqrt{S_1^2/n_1}\right] + T_{n_1-1}$$

$$-\left[t_{1-\alpha/2,n_1-1} - \Delta_{obs}/\sqrt{S_1^2/n_1}\right]$$

(2.6)

where $T_{n_1-1}$ is the $t$-distribution function with $n_1 - 1$ df. It can be seen from Eq. (2.6) that the max power depends only on $n_1$ for an observed $-\Delta_{obs}$.

### 2.3.3. Finite Sample and Power.

Given a finite simulated sample size, $n_2$, the power of the observed $-\Delta_{obs}$ becomes

$$\text{power} = 1 - T_i\left[t_{1-\alpha/2,v} - \Delta_{obs}/\sqrt{S_1^2/n_1 + S_2^2/n_2}\right]$$

$$+ T_i\left[-t_{1-\alpha/2,v} - \Delta_{obs}/\sqrt{S_1^2/n_1 + S_2^2/n_2}\right]$$

(2.7)

If the power is chosen to match 95% of the max power, Eq. (2.6), simulated sample size, $n_2$, can be calibrated from (2.7).

### 2.3.4. Test Procedure.

The goal of our proposed equivalence test is that the simulated sample size, $n_2$, needs to be high enough to reach its 95% max power when the type I error is set at a prespecified $\alpha$ level. Figure 1 and the following steps illustrate the simulation sample size and test procedure:

1. Calculate the max power of the $T$-statistic (2.4).
Figure 1 Flow-chart of DDI test for difference in either sample mean or variance.

(2) Calibrate simulated sample size, \( n_2 \), such that the power reaches 95% of max power.

(3) If \( \left| T \right| > t_{1-\alpha/2,r} \), reject \( H_0 \); otherwise, accept \( H_0 \).

### 2.4. DDI Equivalence Test in Sample Mean

#### 2.4.1. Equivalence Test Statistic Specification

Following the prescribed distribution assumptions for data, logAUCR\textsubscript{obs} and logAUCR\textsubscript{model}, the hypothesis can be expressed as following.

\[ H_0 : \Delta \geq \varepsilon \quad \text{vs.} \quad H_a : \Delta < \varepsilon. \]

The bound \( \varepsilon \) is chosen as 0.223, which corresponds to an interval, (0.80, 1.25), in AUCR-ratio. The same unequal variance \( T \)-statistic (2.1) is chosen for the equivalence test, which is an extension of an equal variance \( T \)-statistic by Wellek (2003, Chapter 6, Section 6.1, pp. 101–103). Under Eq. (2.1), following a similar derivation by Wellek (2003), a level \( \alpha \) rejection region is

\[ \left\{ \left| T \right| < \tilde{C}_{x, \alpha}(\varepsilon) \right\} \quad (2.8) \]
where \( \tilde{C}_{x,y}(e) \) is solved from

\[
T_i \left[ C - \varepsilon / \sqrt{S_1^2/n_1 + S_2^2/n_2} \right] - T_i \left[ - C - \varepsilon / \sqrt{S_1^2/n_1 + S_2^2/n_2} \right] = \alpha \tag{2.9}
\]

### 2.4.2. Maximum Power.

When \( n_2 \) goes to infinity, \( T \)-statistic in Eq. (2.1) becomes Eq. (2.4). Its rejection region is

\[
\{ |T| < \tilde{C}_{x,n_1-1}(e) \} \tag{2.10}
\]

where \( \tilde{C}_{x,n_1-1}(e) \) is solved from

\[
T_{n_1-1} \left[ C - \varepsilon / \sqrt{S_1^2/n_1} \right] - T_{n_1-1} \left[ - C - \varepsilon / \sqrt{S_1^2/n_1} \right] = \alpha \tag{2.11}
\]

If the true difference equals observed \( -\Delta_{\text{obs}} = |\log \text{UCR}_{\text{model}} - \log \text{UCR}_{\text{obs}}| \), the maximum power of the equivalence test is

\[
\text{max power} = T_{n_1-1} \left[ \tilde{C}_{x,n_1-1} - \Delta_{\text{obs}} / \sqrt{S_1^2/n_1} \right] - T_{n_1-1} \left[ -\tilde{C}_{x,n_1-1} - \Delta_{\text{obs}} / \sqrt{S_1^2/n_1} \right] \tag{2.12}
\]

This max power depends only on \( n_1 \) for an observed \( \Delta_{\text{obs}} \) in \( H_0, \Delta < e \).

### 2.4.3. Finite Sample and Power.

Given a finite simulated sample size, \( n_2 \), the power of the observed \( -\Delta_{\text{obs}} \) becomes

\[
\text{power} = T_i \left[ \tilde{C}_{x,y} - \Delta_{\text{obs}} / \sqrt{S_1^2/n_1 + S_2^2/n_2} \right] - T_i \left[ -\tilde{C}_{x,y} - \Delta_{\text{obs}} / \sqrt{S_1^2/n_1 + S_2^2/n_2} \right] \tag{2.13}
\]

where \( \tilde{C}_{x,y} \) is solved from Eq. (2.11). If the power is chosen to match 95% of the max power, Eq. (2.12), simulated sample size, \( n_2 \), can be calibrated from Eq. (2.13).

### 2.4.4. Equivalence Test Procedure.

The goal of our proposed equivalence test is that the simulated sample size, \( n_2 \), needs to be high enough to reach its 95% max power when the type I error is set at a prespecified \( \alpha \) level. Figure 2 and the following steps illustrate the simulation sample size and test procedure:

1. If the observed difference \( -\Delta_{\text{obs}} > e \), accept the null hypothesis.
2. Otherwise, calculate the max power of the \( T \)-statistic, Eq. (2.4).
3. Calibrate simulated sample size, \( n_2 \), such that the power reaches 95% of max power.
4. If \( |T| > \tilde{C}_{x,\alpha} \), accept \( H_0 \); otherwise, reject \( H_0 \).
2.5. DDI Test for Difference in Variance

2.5.1. Test Statistic Specification. It is important to compare the sample variance between the DDI model simulated log AUCR and the observed study log AUCR. The hypothesis test in variance can be expressed as follows:

\[ H_0 : \sigma_1^2 / \sigma_2^2 = 1; \quad H_1 : \sigma_1^2 / \sigma_2^2 \neq 1 \]

An F-statistic is chosen to test this hypothesis

\[
F = \frac{S_1^2}{S_2^2} = \frac{\sum_{i=1}^{n_1} (\log AUCR_{\text{obs},i} - \log AUCR_{\text{obs}})^2 (n_1 - 1)}{\sum_{j=1}^{n_2} (\log AUCR_{\text{model},j} - \log AUCR_{\text{model}})^2 / (n_2 - 1)}
\]  

(2.14)

Under the null hypothesis, it follows an F distribution with \( v_1 = n_1 - 1 \) and \( v_2 = n_2 - 1 \). Under Eq. (2.14), its level-\( \alpha \) test has the following rejection region:

\[
\{ F < F_{\alpha/2,v_1,v_2} \ \| \ F > F_{1-\alpha/2,v_1,v_2} \}
\]  

(2.15)

where \( F_{\alpha/2,v_1,v_2} \) is the \( \alpha/2 \times 100\% \) percentile of the F-distribution.
2.5.2. Maximum Power. When \( n_2 \to \infty \), \( F \)-statistic, Eq. (2.14), becomes a chi-squared statistic.

\[
\lim_{n_2 \to \infty} F = \lim_{n_2 \to \infty} \frac{S_1^2}{S_2^2} = \frac{S_1^2}{\sigma_2^2} = G
\]  

(2.16)

Its rejection region becomes

\[
\{ G < \chi^2_{v_1}/v_1 \parallel G > \chi^2_{v_2}/v_2 \} = \{ G < C_1 \parallel G > C_2 \}
\]  

(2.17)

If the true variance ratio equals the observed \( G = \rho_{\text{obs}} = S_1^2/S_2^2 \), where \( \sigma_2^2 = S_2^2 \) as \( n_2 \to \infty \), the maximum power of equivalence test is

\[
\text{max power} = \chi^2_{v_1} \{ v_1 \times C_1/\rho_{\text{obs}} \} + 1 - \chi^2_{v_2} \{ v_1 \times C_2/\rho_{\text{obs}} \}
\]  

(2.18)

It can be seen from Eq. (2.18) that the maximum power depends only on \( v_1 = n_1 - 1 \) and observed \( \rho_{\text{obs}} \).

2.5.3. Finite Sample Size and Power. For a finite simulated sample size \( n_2 \), the power to detect the observed \( F = \rho_{\text{obs}} = S_1^2/S_2^2 \) is

\[
\text{power} = F_{v_1,v_2} \{ C_1/\rho_{\text{obs}} \} + 1 - F_{v_1,v_2} \{ C_2/\rho_{\text{obs}} \}
\]  

(2.19)

If the power is chosen to match 95% of the max power, Eq. (2.18), simulated sample size, \( n_2 \), can be calibrated from Eq. (2.19).

2.5.4. Test Procedure. The goal of our proposed equivalence test in sample variance is that the simulated sample size, \( n_2 \), needs to be high enough to reach its 95% max power when the type I error is set at a prespecified \( \alpha \) level. Our proposed test procedure is summarized in the following steps:

1. Calculate the max power.
2. Calibrate simulated sample size, \( n_2 \), such that the power reaches 95% of max power.
3. If \( F = \rho_{\text{obs}} \) falls outside of \( (F_{\chi^2_{v_1,v_2},F_{1-\alpha/2,v_1,v_2}}) \), reject \( H_0 \); otherwise, don’t reject \( H_0 \).

2.6. DDI Sample Variance Equivalence Test

2.6.1. Equivalence Test Statistic Specification. The equivalence test in variance can be expressed as follows:

\[
H_0 : \rho \leq \delta_1 \quad \text{or} \quad \rho \geq \delta_2 \quad \text{vs.} \quad H_a : \delta_1 < \rho < \delta_2
\]  

(2.20)

Using the prescribed \( F \)-statistic, Eq. (2.14), a uniformly most powerful (UMP) level-\( \alpha \) test was derived by Wellek (2003, Chapter 6, Section 6.4, pp. 126–128), with the following rejection region

\[
\left\{ \tilde{C}^{(1)}_{x_1,v_2}(\delta_1, \delta_2) < F < \tilde{C}^{(2)}_{x_1,v_2}(\delta_1, \delta_2) \right\}
\]  

(2.21)
The critical boundaries are subject to the following equations:

\[
F_{1,2}(\bar{C}_2/\delta_1) - F_{1,2}(\bar{C}_1/\delta_1) = F_{1,2}(\bar{C}_2/\delta_2) - F_{1,2}(\bar{C}_1/\delta_2)
= \alpha \quad (0 < \bar{C}_1 < \bar{C}_2 < \infty),
\]

(2.22)

where \( F_{v_1,v_2}(\cdot) \) denotes the cdf of the standard central \( F \)-distribution with degrees of freedom \( v_1 \) and \( v_2 \). The choice for equivalence range is subjective and usually made as symmetric on the log-scale, such as \((0.50, 2.00)\) (Wellek, 2003).

2.6.2. Maximum Power. When \( v_2 = n_2 - 1 \) goes to infinity, \( F \)-statistic, Eq. (2.14) becomes a chi-squared statistic, Eq. (2.16), and its rejection region becomes

\[
\{ \bar{C}^{(1)}_{v_1,v_2}(\delta_1, \delta_2) < G < \bar{C}^{(2)}_{v_1,v_2}(\delta_1, \delta_2) \}, \quad \chi^2_{\alpha}(v_1 \times \bar{C}_2/\delta_1) - \chi^2_\alpha(v_1 \times \bar{C}_1/\delta_1) = \chi^2_{\alpha}(v_1 \times \bar{C}_2/\delta_1) - \chi^2_\alpha(v_1 \times \bar{C}_1/\delta_2) = \alpha \quad (0 < \bar{C}_1 < \bar{C}_2 < \infty).
\]

(2.23)

If the true variance ratio equals the observed \( G = \rho_{\text{obs}} = S_2^2/S_2^2 \), where \( \rho_1^2 = S_2^2 \) as \( n_2 \to \infty \), the maximum power of equivalence test is

\[
\max \text{ power} = \chi^2_{\alpha}(v_1 \times \bar{C}_2/\rho_{\text{obs}}) - \chi^2_\alpha(v_1 \times \bar{C}_1/\rho_{\text{obs}}).
\]

(2.24)

It can be seen from Eq. (2.24) that the maximum power depends only on \( v_1 = n_1 - 1 \) and observed \( \rho_{\text{obs}} \) in \( H_0, \delta_1 < \rho_{\text{obs}} < \delta_2 \).

2.6.3. Finite Sample Size and Power. For a finite simulated sample size \( n_2 \), the power to detect the observed \( F = \rho_{\text{obs}} = S_2^2/S_2^2 \) is

\[
\text{power} = F_{1,2}(\bar{C}_2/\rho_{\text{obs}}) - F_{1,2}(\bar{C}_1/\rho_{\text{obs}}).
\]

(2.25)

If the power is chosen to match 95% of the max power, simulated sample size, \( n_2 \), can be calibrated from Eq. (2.25).

2.6.4. Equivalence Test Procedure. The goal of our proposed equivalence test in sample variance is that the simulated sample size, \( n_2 \), needs to be high enough to reach its 95% max power when the type I error is set at a prespecified \( \alpha \) level. The test procedure is specified in Fig. 3 and the following steps:

1. If the observed difference \( \delta_1 > \rho_{\text{obs}} \) or \( \rho_{\text{obs}} > \delta_2 \), accept the null hypothesis.
2. Otherwise, calculate the max power.
3. Calibrate simulated sample size, \( n_2 \), such that the power reaches 95% of max power.
4. If \( F = \rho_{\text{obs}} \) falls outside of \((\bar{C}^{(1)}_{v_1,v_2}(\delta_1, \delta_2), \bar{C}^{(2)}_{v_1,v_2}(\delta_1, \delta_2))\), accept \( H_0 \); otherwise, reject \( H_0 \).
2.7. DDI Prediction Performance Assessment

2.7.1. Bias. The DDI prediction bias and its 95% CI interaction in sample mean are defined as

\[
\text{bias}_{\text{mean}} = \frac{AUCR_{\text{model}}}{AUCR_{\text{obs}}} - 1 \quad (2.26)
\]

\[
(1 - \alpha) \times 100\% \text{ CI} = \exp\left( t_{\alpha/2,v} + T \right) \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},
\]

\[
(1 - \alpha/2,v) \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} - 1 \quad (2.27)
\]

where \( T \) is defined in Eq. (2.1), and sample size \( n_2 \) is calculated in Eq. (2.7).
The DDI prediction bias and its 95% CI interaction in sample variance are defined as

$$bias_{var} = \frac{S_2^2}{S_1^2} - 1 = \frac{1}{\rho_{obs}} - 1$$  \hspace{1cm} (2.28)

$$\left(1 - \alpha\right) \times 100\% \text{ CI} = \left(1/F_{1-v_2/v_1,2}, 1/F_{2/v_1,2}\right)/\rho_{obs} - 1$$  \hspace{1cm} (2.29)

where sample size $$\nu_2 = n_2 - 1$$ is calculated in Eq. (2.19).

### 2.7.2. Coverage Probability for Predictive Confidence Interval.

For a new subject, the DDI, $$\log \text{AUCR}_{\text{model,new}}$$, can be predicted by our model-based approach, which will have mean $$\log \text{AUCR}_{\text{model}}$$ and variance $$S^2_2$$, and
two = \left(\log \text{AUCR}_{\text{model,new}} - \log \text{AUCR}_{\text{model}}\right)/S_2 \sim t_{n_2-1}$. Its \((1 - \alpha)100\%\) predictive confidence interval is

$$(C^*_1, C^*_2) = \left(S_2 t_2/v_2, n_2 - 1 + \log \text{AUCR}_{\text{model}}\right)$$  \hspace{1cm} (2.30)

If a new DDI, $$\log \text{AUCR}_{\text{obs,new}}$$, is observed from a clinical study, it will have mean $$\log \text{AUCR}_{\text{obs}}$$ and variance $$S^2_1$$, and $$t_1 = (\log \text{AUCR}_{\text{obs,new}} - \log \text{AUCR}_{\text{obs}})/S_1 \sim t_{n_1-1}$. Its coverage probability of \((C^*_1, C^*_2)\) will be assessed as

$$\text{Coverage Probability} = \frac{T_{n_1-1} \left(C^*_2 - \log \text{AUCR}_{\text{obs}}\right)}{S_1} - T_{n_1-1} \left(C^*_1 - \log \text{AUCR}_{\text{obs}}\right)$$  \hspace{1cm} (2.31)

If observed and model-based DDI follows the same predictive distributions, the coverage probability will be close to its nominal level, \((1 - \alpha)100\%\).

### 3. DIFFERENCE/EQUIVALENCE TESTS TO ASSESS KETO/MDZ MODEL BASED DDI

A published KETO/MDZ DDI study of intravenous MDZ administration with or without oral KETO is available to assess our model-based DDI prediction. In this published paper, it showed that the observed sample mean log AUCR was 1.629 and its $$S_1^2 = 0.1391$$ (Tsunoda et al., 1999). The study contained nine subjects, and dose combination was KETO/MDZ = 200/2mg.

With the same dose combination and a prespecified inhibition mechanism represented by $$K_i$$, KETO/MDZ DDI is predicted through methods described in Sections 2.1 and 2.2. Different inhibition constant, $$K_i$$, were measured from different in vitro experiments with different inhibition mechanism assumptions (von Moltke et al., 1996; Wang et al., 1999). It has been a long debate in the research, and each publication contains a research group’s own theory. The $$K_i$$ values from different studies can’t be simply summarized with a meta-analysis scheme. In this paper, we test both an upper bound and a lower bound of published $$K_i$$. The upper bound $$K_i = 0.18 \mu M$$, was measured through a noncompetitive inhibition model assumption (Wang et al., 1999); whereas the lower bound, $$K_i = 0.0037 \mu M$$, was calculated through a competitive inhibition model assumption (von Moltke et al., 1996). In addition, we also tried two more hypothetical $$K_i$$ values, 0.0005 and 0.00005 \mu M.
3.1. DDI Sample Mean Difference/Equivalence Tests

Table 1 presents the DDI sample mean difference/equivalence tests. When $K_i$ is fixed at 0.18 and 0.0037 ($\mu$M), the predicted log AUCR sample means are 0.179 and 1.075, respectively. Compared to the observed study log AUCR sample mean, 1.629, there is strong evidence that the DDI model underpredicts observed AUCR ($p$-value = 0.000006 and 0.002, respectively). If $K_i$ is reduced to 0.0005 and 0.00005 $\mu$M, there is not enough evidence to conclude that predicted and observed DDI are different ($p$-value = 15 and 0.98, respectively). It is interesting to see that the minimal simulation sample size to reach 95% max power is less than 158. As $K_i$ decreases, the power to test the difference between observed and predicted DDI reduces from 95% to 5.04%.

In testing the equivalence, when $K_i$ is fixed at 0.18 and 0.0037 ($\mu$M), the differences in mean are $\Delta_{\text{obs}} = 1.450$ and 0.555, respectively. These two $\Delta_{\text{obs}}$ are outside of the equivalence bound ($-0.223, 0.223$), and we don’t have enough evidence to support that predicted and observed DDI are the same. When $K_i$ is reduced to 0.0005, though, $\Delta_{\text{obs}} = 0.197$ falls into the equivalence bound ($-0.223, 0.223$), and the $T$-statistic is 1.30, which is still outside of the rejection boundary, ($-0.184, 0.184$). Only when $K_i$ is reduced to 0.00005 do we have enough evidence to conclude that predicted and observed DDI are equivalent, i.e., $T$-statistic $= 0.024$ falls into the rejection boundary, ($-0.273, 0.273$). As indicated in Table 1, the minimal simulation sample size to reach 95% max power is less than 125.

Therefore, together with the KETO/MDZ DDI model, neither non-competitive inhibition $K_i(0.18 \mu$M) nor competitive inhibition $K_i(0.0037 \mu$M) are good enough to predict observed AUCR at the sample mean level, and they both

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$K_i = 0.18$ (Noncomp. inhibition)</th>
<th>$K_i = 0.0037$ (Comp. inhibition)</th>
<th>$K_i = 0.0005$ (Hypothetical)</th>
<th>$K_i = 0.00005$ (Hypothetical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUCR</td>
<td>1.196</td>
<td>2.928</td>
<td>4.187</td>
<td>5.116</td>
</tr>
<tr>
<td>log AUCR</td>
<td>0.179</td>
<td>1.075</td>
<td>1.43</td>
<td>1.632</td>
</tr>
<tr>
<td>$S^2_1$</td>
<td>0.006</td>
<td>0.075</td>
<td>0.100</td>
<td>0.103</td>
</tr>
<tr>
<td>Test for difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max power $\times$ 95%</td>
<td>95%</td>
<td>95%</td>
<td>30%</td>
<td>5.04%</td>
</tr>
<tr>
<td>Minimal $n_2$</td>
<td>2</td>
<td>9</td>
<td>141</td>
<td>158</td>
</tr>
<tr>
<td>$P$-value</td>
<td>0.0000006</td>
<td>0.002</td>
<td>0.15</td>
<td>0.98</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Rej $H_0$</td>
<td>Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
</tr>
<tr>
<td>Equivalence Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta_{\text{obs}}$</td>
<td>1.450</td>
<td>0.555</td>
<td>0.197</td>
<td>0.003</td>
</tr>
<tr>
<td>max power $\times$ 95%</td>
<td>N/A</td>
<td>N/A</td>
<td>6.2%</td>
<td>20.9%</td>
</tr>
<tr>
<td>Minimal $n_2$</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>125</td>
</tr>
<tr>
<td>$(\tilde{C}_d^{(1)}, \tilde{C}_d^{(2)})$</td>
<td>N/A</td>
<td>N/A</td>
<td>$(-0.184, 0.184)$</td>
<td>$(-0.273, 0.273)$</td>
</tr>
<tr>
<td>$T$</td>
<td>N/A</td>
<td>N/A</td>
<td>1.30</td>
<td>0.024</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Rej $H_0$</td>
</tr>
</tbody>
</table>

Note: The reported study, AUCR $= 5.10$, logAUCR $= 1.629$, variance $S^2_1 = 0.139$; the equivalence test bound is $(\varepsilon_1, \varepsilon_2) = (-0.223, 0.223)$. 
underpredicted log AUCR. Only when $K_i$ is reduced to 0.00005 is there enough evidence to support the equivalence between predicted and reported DDI.

3.1.1. DDI Variance Difference/Equivalence Tests. Table 2 presents the DDI variance difference/equivalence test. The sample variance of simulated log AUCR, $S^2$, depends on interaction constant $K_i$. When $K_i$ is chosen as a noncompetitive inhibition mechanism, 0.18 (μM), $S^2 = 0.006$ and $\rho_{obs} = 23.045$. The $F$-test suggests that predicted and observed DDI have different variances ($p = 0.002$), with 95% power through only four simulated samples. However, when $K_i$ is chosen as other values, there is not enough evidence to conclude that two DDI distributions have different variances. The simulation sample size is less than 40 in order to achieve the 95% of the maximum power, and the maximum power is less than 16%.

When testing the equivalence, the $F$-test is either outside of the equivalence bound (0.5, 2) or outside of rejection regions. Therefore, there exists no evidence to conclude that two DDI distributions have the same variances. The simulation sample size is less than 70 in order to achieve 95% of the maximum power, and all of the maximum powers are below 10%.

As a result, noncompetitive inhibition $K_i/(0.18 \mu M)$ underpredicts the DDI variance ($p = 0.002$). There is not enough evidence to support that the competitive inhibition $K_i/(0.0037 \mu M)$ predicts either equivalent or different variances in AUCR.

3.1.2. DDI Prediction Performance Assessment. The largest relative bias in predicting DDI mean, $-0.76$ with 95% CI = (--0.83, --0.68), is with $K_i = 0.18$. This bias reduces to 0.003 when $K_i$ becomes 0.00005. Similarly, $K_i = 0.18$ leads to the largest relative bias in predicting DDI variance, $-0.96$ with 95% CI = (--0.99, --0.80); it reduces to $-0.26$ when $K_i$ becomes 0.00005. Following a similar trend, the coverage probability of predictive confidence interval goes up when $K_i$ becomes smaller.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$K_i = 0.18$ (Noncomp. inhibition)</th>
<th>$K_i = 0.0037$ (Comp. inhibition)</th>
<th>$K_i = 0.0005$ (Hypothetical)</th>
<th>$K_i = 0.00005$ (Hypothetical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^2$</td>
<td>0.006</td>
<td>0.075</td>
<td>0.100</td>
<td>0.103</td>
</tr>
<tr>
<td>Test for difference max power × 95%</td>
<td>95%</td>
<td>15.8%</td>
<td>7.1%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Minimal $n_2$</td>
<td>4</td>
<td>40</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>$P$-value</td>
<td>0.002</td>
<td>0.32</td>
<td>0.63</td>
<td>0.66</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
</tr>
<tr>
<td>Equivalence Test</td>
<td>$F = \rho_{obs}$</td>
<td>23.045</td>
<td>1.848</td>
<td>1.39</td>
</tr>
<tr>
<td>max power × 95%</td>
<td>N/A</td>
<td>5.6%</td>
<td>9.0%</td>
<td>9.4%</td>
</tr>
<tr>
<td>Minimal $n_2$</td>
<td>(C_{21(1)}, C_{21(2)})</td>
<td>N/A</td>
<td>(0.930, 1.076)</td>
<td>(0.881, 1.024)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
<td>Not Rej $H_0$</td>
</tr>
</tbody>
</table>

Note: The reported study log AUCR variance $S^2_{1} = 0.139$; the equivalence test bound is $(\delta_1, \delta_2) = (0.5, 2.0)$. 

Table 2 DDI variance equivalence test
Table 3  DDI prediction assessment

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$K_i = 0.18$ (Noncomp. inhibition)</th>
<th>$K_i = 0.0037$ (Comp. inhibition)</th>
<th>$K_i = 0.0005$ (Hypothetical)</th>
<th>$K_i = 0.00005$ (Hypothetical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUCR</td>
<td>1.196</td>
<td>2.928</td>
<td>4.187</td>
<td>5.116</td>
</tr>
<tr>
<td>log AUCR</td>
<td>0.179</td>
<td>1.075</td>
<td>1.43</td>
<td>1.632</td>
</tr>
<tr>
<td>$S_i^2$</td>
<td>0.006</td>
<td>0.075</td>
<td>1.10</td>
<td>0.103</td>
</tr>
<tr>
<td>Relative bias in mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>−0.76</td>
<td>−0.42</td>
<td>−0.18</td>
<td>0.003</td>
</tr>
<tr>
<td>95% CI</td>
<td>(−0.83, −0.68)</td>
<td>(−0.58, −0.20)</td>
<td>(−0.38, 0.09)</td>
<td>(−0.242, 0.32)</td>
</tr>
<tr>
<td>Relative bias in variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>−0.96</td>
<td>−0.42</td>
<td>−0.28</td>
<td>−0.26</td>
</tr>
<tr>
<td>95% CI</td>
<td>(−0.99, −0.80)</td>
<td>(−0.78, 0.89)</td>
<td>(−0.74, 1.63)</td>
<td>(−0.94, 1.72)</td>
</tr>
<tr>
<td>Coverage probability of predictive confidence interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.0016</td>
<td>0.23</td>
<td>0.54</td>
<td>0.60</td>
</tr>
<tr>
<td>80%</td>
<td>0.0023</td>
<td>0.31</td>
<td>0.63</td>
<td>0.70</td>
</tr>
<tr>
<td>90%</td>
<td>0.0040</td>
<td>0.43</td>
<td>0.75</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Note: The reported study, AUCR = 5.10, log AUCR = 1.629, variance $S_i^2 = 0.139$; the equivalence test bound is $(e_1, e_2) = (-0.223, 0.223)$.

4. DISCUSSION

In this paper, for the first time, we implement difference/equivalence tests to assess the mean and variance of model-based simulated DDI with published DDI data. Most importantly, we provide sample size and power guidelines for model-based DDI simulation and difference/equivalence tests and DDI prediction performance assessment statistics. These critical issues have never been discussed in trial simulation studies to investigate DDI (Chien et al., 2006; Yang et al., 2003).

In the KETO/MDZ example, the sample mean difference test suggests that both $K_i = 0.18$ (noncompetitive inhibition mechanism) and 0.0037 μM (competitive inhibition mechanism) lead to underpredicted AUCR compared to the observed AUCR ($p = 0.0000006$ and 0.002, respectively; and the relative prediction biases are −0.76 and −0.42, respectively). At the mean time, $K_i = 0.18$ (noncompetitive inhibition mechanism) underpredicts the observed DDI variance ($p = 0.002$), and relative bias in prediction is −0.96. Also, $K_i = 0.0037$ (competitive inhibition mechanism) underpredicts the variance without enough evidence from either difference or equivalence tests.

Both $K_i$ fail to predict observed DD, which implies that the inhibition mechanism alone is not sufficient to predict observed DDI. One possible alternative is that the active concentration of KETO (inhibitor) in liver is higher than the plasma concentration described by its two-compartment model. Therefore, a physiological model with a liver component might be a better strategy (Chien et al., 2006). The current KETO/MDZ DDI model underpredicts reported log AUCR between-subject variances, under either noncompetitive or competitive inhibition mechanisms. It suggests that there exists between-subject variance for $K_i$, whose information is not available yet. It calls for additional PK studies to assess $K_i$'s variation among the population.
Both equivalence and difference tests suggest that simulation sample size does not have to be large (less than 300) to reach 95% of the maximum power. It makes this model-based simulation approach a highly promising prediction tool in DDI research. The maximum power of these tests can become small if the differences of predicted and observed distributions are neither small nor large, and the maximum power is limited by the sample size of observed DDI study.

The purpose of the equivalence tests is to establish the validity of model-based DDI prediction. Once the equivalences are reached from a training example, for example MDZ/KETO, the DDI model can be modified to predict investigational drugs’ DDI. In terms of drug development, its application is at the transition period between nonclinical and clinical stages. It fits well with newly developed FDA guidelines for exploratory IND studies (http://www.fda.gov/cder/guidance/6695dft.pdf). This document highly recommends a PK study of an investigational drug at a very low dosing level and a short duration. Its goal is to obtain this drug’s PK instead of investigating its pharmacologic effects. If this investigational drug is an inhibitor, its PK data and model can be incorporated with PK data that is available for many common substrates and the $K_i$ values determined in vitro for the substrate-inhibitor combination. The interaction potential of this new agent can thus be predicted through established DDI models. This same strategy can be applied to a new compound that is a substrate by nature, utilizing PK data for the enzyme inhibitor (e.g., KETO).

Another application of an established DDI model to the drugs on the market is the investigation of the worst DDI scenario through in silico trial simulations. For example, different dose combinations, lagging dosing time, and multiple dose regimens were investigated for KETO/MDZ, and their strongest DDI were studied (Yang et al., 2003). Since these worst scenario DDI studies are usually unethical to be conducted using human subjects, DDI model-based trial simulation provides a perfect platform. The other application of an established DDI model is to design a clinical DDI study for a new investigational drug. Most of the DDI models are based on pharmacology mechanisms, therefore the DDI data simulation will be much more relevant. Unfortunately, the statistical literatures on the link between the model-based DDI trial simulations and clinical DDI trial designs are lacking.

ACKNOWLEDGMENT

Drs. Lang Li and Stephen Hall researches are supported by NIH grants, R01 GM74217 (LL), R01 GM67308 (SH), and FD-T-001756(SH).

REFERENCES

DRUG-DRUG INTERACTION PREDICTION ASSESSMENT


