Statistical Modeling in Preclinical Drug Proarrhythmic Assessment

Nan Xi (UCLA), Yu-Yi Hsu (CDER/OB/DBVI), Qianyu Dang (CDER/OB/DBVI), and Dalong Huang (CDER/OB/DBVI)

This poster reflects the views of the authors and should not be construed to FDA's views or policies.

Abstract

Torsades de pointes (TdP) is the irregular heart rhythm as a side effect of drugs and may cause sudden cardiac death. Much effort has been invested in understanding the mechanism of TdP in clinical and pre-clinical studies. However, a comprehensive statistical learning framework that can accurately predict the TdP risk of drugs from pre-clinical data is still lacking. Here, we proposed ordinal logistic regression and ordinal random forest models to predict low-, intermediate-, and high-risk drugs based on datasets generated from two experimental protocols. Leave-one-drug-out cross-validation, stratified bootstrap, and permutation predictor importance were applied to estimate and interpret the model performance under uncertainty. The potential outlier drugs identified by our models are consistent with their descriptions in the literature. Our method is accurate and interpretable and can be used as additional evidence in the assessment of drug safety.

Introduction

Torsades de pointes (TdP) is a rare but potentially fatal ventricular arrhythmia that can be induced as a side effect by cardiovascular and non-cardiovascular drugs. The identification of TdP risk is a critical step in the assessment of new drugs before reaching the market. Several pre-clinical paradigms have been promoted to identify the drug-induced TdP risk in vivo and in vitro. Among them, the comprehensive in vitro proarrhythmia assay (CiPA) and rabbit ventricular wedge assay (RVWA) are two cutting-edge methods. Although effort has been largely invested in the experimental development of CiPA and RVWA, a comprehensive statistical learning framework that provides accurate prediction of drug-induced TdP risks from experimental observations is still lacking.

In this study, we proposed two statistical models - ordinal logistic regression and ordinal random forest to accurately predict drug-induced TdP risks from CiPA and RVWA datasets. Our methods explored the ordinal information in low-, intermediate-, and high-risk levels instead of treating them as independent categories. The model performance on new drugs was evaluated by leave-one-drug-out cross-validation. In addition to the point estimate, the uncertainty of model performance was quantified by stratified bootstrap. We also examined the model interpretation through permutation predictor importance. Sensitivity analysis was conducted to investigate the impact of potential outlier drugs on the model performance. Motivated by the practice in experiments, we performed control analysis by selecting one high risk drug and placebo as positive and negative controls and provided further model validation.

The proposed modeling strategies were evaluated on hiPSC-CMs dataset (stem cell dataset) and isolated arterially-perfused rabbit ventricular wedge preparation dataset (wedge dataset). Both datasets contain the electrophysiological effects on biological models, which were induced by the same 28 drugs with low, intermediate, or high TdP. Our work is the first attempt to build multivariate statistical models with interpretability and uncertainty measurement that can predict the drug TdP risk accurately from in vitro experimental data. The proposed modeling and evaluation methods can be easily extended to new datasets and experimental protocols. The result of model prediction could be used as additional evidence in the assessment of drug safety.

Materials and Methods

Datasets

The first dataset in this study was generated from a multisite experiment of human-induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs). In this experiment, 28 drugs with known low, intermediate, and high risk of torsades de pointes (TdP) were applied to hiPSC-CMs. Electrophysiological responses to 28 drugs at different concentration were recorded, which generated 7 predictors in the dataset. For each drug, 15 observations were collected based on combinations of laboratories, hiPSC-CMs cell lines, and electrophysiological platforms. In total, there are 420 observations in this dataset.

The second dataset in this study was generated from a blinded validation of rabbit ventricular wedge assay (RVWA) for the assessment of drug-induced proarrhythmia. The same 28 drugs were applied to RVWA and 16 predictors were obtained from electrophysiological. For each drug, 4 observations were collected in one laboratory with the same biological model and electrophysiological platform. There are totally 112 observations in this dataset.

Utilizing ordinal information

The three risk categories are not independent – there is ordinal information in the risk categories, i.e., low < intermediate < high. We utilized this ordinal information by transferring the three-class classification problem into two binary classification problem. First, we trained a binary classifier $\hat{f}_1$ that differentiated between low-risk and not-low-risk (intermediate or high) drugs. Similarly, we trained binary classifier $\hat{f}_2$ that differentiated between high-risk and not-high-risk (low or intermediate) drugs. Second, for any new observation x, we predicted its probability of being low risk $p_x(L)$, not-low risk $p_x(MH)$, high risk $p_x(H)$, and not-high risk $p_x(LM)$ by

$$p_x(L) = \hat{f}_1(x),$$
$$p_x(MH) = 1 - p_x(L),$$
$$p_x(H) = \hat{f}_2(x),$$
$$p_x(LM) = 1 - p_x(H).$$

Third, the probability of being intermediate risk $p_x(M)$ was calculated by

$$p_x(M) = p_x(MH) - p_x(H).$$

Finally, the risk of observation x was predicted by

$$\widehat{risk}(x) = argmax_k\{p_x(k)\},$$

where $k \in \{L, M, H\}$. The prediction of drug risk is the majority vote of the observations belonging to that drug. We broke the tie in the majority vote in a conservative manner.

Ordinal logistic regression

We built an ordinal logistic regression model to predict the drug-risk category by substituting logistic regression for the binary classifier in the previous framework. Formally, for any observation x, the two binary classifiers $\hat{f}_1$ and $\hat{f}_2$ are defined as

$$\hat{f}_1 : log\frac{p_x(L)}{1-p_x(L)} = X\hat{\beta}_1$$
$$\hat{f}_2 : log\frac{p_x(H)}{1-p_x(H)} = X\hat{\beta}_2 ,$$

where X is the predictor vector of observation x while $\hat{\beta}_1$ and $\hat{\beta}_2$ are the vectors of model parameters. $X\hat{\beta}_1$ and $X\hat{\beta}_2$ are inner products between vectors of predictors and parameters. $\hat{f}_1$ and $\hat{f}_2$ were fitted by maximum likelihood estimation. After obtaining $\hat{f}_1$ and $\hat{f}_2$, the drug risk and the probability of being each risk category can be predicted as describe in the previous section.

Ordinal random forest

Random forest is the ensemble of multiple decision trees that can capture the nonlinearity in the dataset. To build a random forest from multiple decision trees, we generated $B(B = 500$ in our implementation) bootstrap samples from the training data. We applied the previous splitting rule to build one decision tree per each bootstrap resample. Instead of searching all p predictors, we randomly selected $\sqrt{p}$ predictors as candidates in each split to reduce the correlation among different decision trees. The final random forest $\{T_b\}_1^B$ contains B decision trees and the prediction of risk is the majority vote of all decision trees. As random forest is a combination of low-correlated decision trees, it reduces the variance of prediction result from single decision tree and usually exhibits high prediction accuracy in real-world applications.

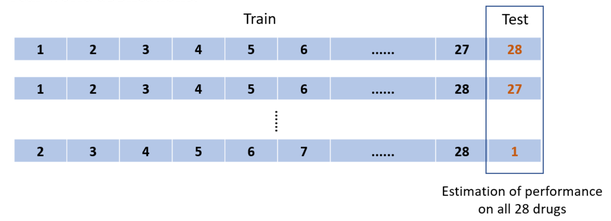

Figure 1. Leave-one-drug-out cross-validation. In each iteration, the predictive model was trained on 27 drugs and predicted on the one left-out drug. The process was repeated until each drug was predicted and the model performance was calculated by combining the prediction result of each drug.

Results and Discussion

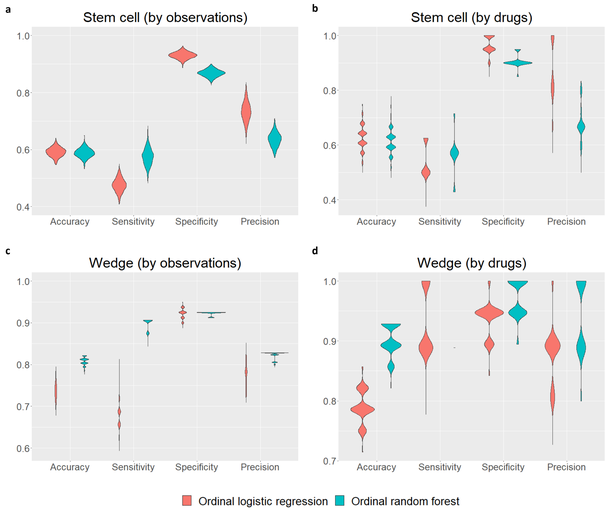

Figure 2. The empirical distribution of four performance measurements generated by stratified bootstrap.

Table 1. The model performance on the stem cell dataset. Sensitivity, specificity, and precision are shown for high-risk category. The better performance between two models are underscored.

	Model	Accuracy	High risk		
			Sensitivity	Specificity	Precision
By observations	Ordinal logistic regression	<u>0.605</u>	0.475	<u>0.953</u>	<u>0.802</u>
	Ordinal random forest	0.600	<u>0.600</u>	0.866	0.642
By drugs	Ordinal logistic regression	<u>0.678</u>	0.500	<u>1.000</u>	<u>1.000</u>
	Ordinal random forest	0.642	<u>0.625</u>	0.900	0.714

Table 2. The model performance on the wedge cell dataset. Sensitivity, specificity, and precision are shown for high-risk category. The better performance between two models are underscored.

	Model	Accuracy	High risk		
			Sensitivity	Specificity	Precision
By observations	Ordinal logistic regression	0.741	0.593	<u>0.937</u>	0.791
	Ordinal random forest	<u>0.803</u>	<u>0.906</u>	0.925	<u>0.828</u>
By drugs	Ordinal logistic regression	0.821	0.875	0.900	<u>0.947</u>
	Ordinal random forest	<u>0.892</u>	<u>1.000</u>	<u>0.950</u>	0.888

Conclusion

In this study, we proposed a comprehensive statistical modeling framework to predict drug-induced TdP risk from two pre-clinical experimental protocols. Ordinal logistic regression and ordinal random forest were trained on the human-induced pluripotent stem cell-derived cardiomyocytes dataset (stem cell dataset) and isolated arterially-perfused rabbit ventricular wedge preparation dataset (wedge dataset). The unbiased estimate of model performance was measured by leave-one-drug-out cross-validation. The uncertainty of model performance was evaluated by stratified bootstrap. Our proposed method provided interpretability consistent with domain knowledge through permutation prediction importance. Sensitivity analysis identified potential outlier drugs that were also described in the literature. Motivated by the practice in experiments, we conducted control analysis and further validated the model performance. Overall, our statistical models exhibited high accuracy in predicting drug-induced TdP risks. The best performance was achieved by ordinal random forest on the wedge dataset, with 0.803 accuracy by observations and 0.892 accuracy by drugs. One supplement of current study is to validate the model performance by cross-dataset validation. The predictive model trained on one dataset would be evaluated on another dataset generated by the same experimental protocol. Such evaluation will provide a more realistic estimation of model performance.