Improvement of simulated PK data-based machine learning model with real-world data to predict infliximab pharmacokinetics in pediatric patients with Crohn's disease

Irie K[1], Minar P[2, 3], Reifenberg J[4], Mizuno T[1, 3]

1 Division of Translational and Clinical Pharmacology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.
2 Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA;
3 Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA.
4 University of Cincinnati School of Medicine, Cincinnati, Ohio, USA.

Background: Rich real-world data is not always available to develop a robust machine-learning (ML) model. Recent studies revealed that simulated PK data generated using a population PK model can facilitate ML modeling (Labriffe et al. CPT PSP. 2022). In this study, we developed a simulated PK data-based ML model for predicting infliximab (IFX) concentrations and further improved the predictive performance using an ensemble modeling approach with real-world data from children with Crohn's disease.

Methods: The simulated PK data-based XGBoost model (1st XGBoost model) was developed with simulated PK data (n=580,000) generated using the published pediatric IFX population PK model (Xiong et al. CPT. 2021). Residuals between ML-predicted and observed IFX concentrations were computed using real-world data from pediatric patients with Crohn's disease (292 concentrations from 93 patients). To correct the gap between predictions and real-world observations, another XGBoost model (2nd XGBoost model) was developed by considering additional features identified based on the 1st modeling. The features for the 2nd model were selected based on the Pearson correlation coefficient (r) between the residuals and real-world data such as lab data and dosing history. The predictive performance of the ML models was evaluated with the root mean square error (RMSE) and mean prediction error (MPE).

Results: For the 1st XGBoost model, the RMSE and MPE were 6.4 µg/mL and 1.6 µg/mL, respectively. There were associations between the residuals and the 1st XGBoost-predicted IFX concentrations ($r = 0.59$, $p<0.001$), average cumulative IFX dose ($r = 0.28$, $p<0.001$), and the dosing interval ($r = -0.255$, $p<0.001$). The 2nd XGBoost model was developed with these features and then combined with the 1st XGBoost model. The ensemble model provided improved predictive performance, with RMSE and MPE of 5.3 µg/mL and 0.3 µg/mL for the train data set (n=219) and 4.6 µg/mL and 0.2 µg/mL for the test data set (n=73), respectively.

Conclusions: The simulated PK data-based ML model was further improved with real-world data through ensemble modeling. The ensemble modeling approach holds promise for continuously improving the ML model predictions by adaptively learning new clinical data for dose optimization.

Key Words: Machine learning, Pharmacokinetics, XGboost, Infliximab, Crohn's disease