Raltegravir plasma exposure prediction using a machine learning–based limited sampling strategy

Matheus L Thomaz[1], Kathley L R de Azevedo[2], Tiago A Paz[1], Glauco H B Nardotto[3], Fernanda L Moreira[4], Geraldo Duarte[2], Jean-Baptiste Woillard[5], Vera L Lanchote[1], Joao P B Ximenez[1]. School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, SP, Brazil[1], Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, SP, Brazil[2], Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, US[3], Faculty of Pharmacy, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil[4], Department of Pharmacology, Toxicology and Pharmacovigilance, CHU de Limoges, Limoges, France[5].

Introduction. Raltegravir (RAL) displays high pharmacokinetic variability, and trough concentrations are poor surrogates of total drug exposure (Blonk et al., 2015). Estimation of the area under the plasma concentration–time curve (AUC0–12) is therefore desirable but impractical in routine care due to the need for intensive sampling.
Aims. To develop and externally validate a machine learning (ML)–based limited sampling strategy (LSS) to predict steady-state RAL AUC0–12 from a small number of plasma samples.
Methods. ML models were trained on 4,400 simulated pharmacokinetic profiles derived from a published population model of RAL 400 mg twice daily. Four algorithms (XGBoost, random forest, GLMNet, and SVM) were evaluated using all possible combinations of two or three samples collected within 4 h post-dose. Performance was assessed by cross-validation, testing set, independent simulations from another population model, and real clinical data.
Results. XGBoost using concentrations at 0.5, 2, and 4 h post-dose provided the best performance. In the test set, bias and relative RMSE were 0.8% and 8.7% (R^2 = 0.987). Performance remained good in independent simulated data (bias 1.9%, relative RMSE 14.3%) and acceptable in real patients (bias 5.0%, relative RMSE 24.1%).
Discussion. This study demonstrates that ML can reliably estimate RAL AUC0–12 using only three early post-dose samples, overcoming key limitations of traditional LSS and Bayesian approaches. The robustness observed across two independent simulation frameworks and real patient data supports the generalizability of the method. By reducing the need for intensive sampling, this approach facilitates exposure-based analyses in clinical pharmacokinetic and PK/PD studies and may enable broader application of individualized RAL dosing, particularly in special populations.

Blonk MI et al (2015) Clin Infect Dis. 2015;61(5):809-816.