



**HAL**  
open science

# Application of machine-learning models to predict the ganciclovir and valganciclovir exposure in children using a limited sampling strategy

Laure Ponthier, Bénédicte Franck, Julie Autmizguine, Marc Labriffe, Philippe Ovetchkine, Pierre Marquet, Anders Asberg, Jean-Baptiste Woillard

## ► To cite this version:

Laure Ponthier, Bénédicte Franck, Julie Autmizguine, Marc Labriffe, Philippe Ovetchkine, et al.. Application of machine-learning models to predict the ganciclovir and valganciclovir exposure in children using a limited sampling strategy. *Antimicrobial Agents and Chemotherapy*, 2024, pp.e0086024. 10.1128/aac.00860-24 . hal-04700415

**HAL Id: hal-04700415**

**<https://hal.science/hal-04700415v1>**

Submitted on 26 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# **Application of machine learning models to predict the ganciclovir and valganciclovir exposure in children using a limited sampling strategy.**

Laure Ponthier<sup>1,2</sup>, Bénédicte Franck<sup>3</sup>, Julie Autmizguine<sup>4,5,6</sup>, Marc Labriffe<sup>1,7</sup>, Philippe Ovetchkine<sup>4</sup>, Pierre Marquet<sup>1,7</sup>, Anders Åsberg<sup>8,9</sup>, Jean-Baptiste Woillard<sup>1,7</sup>

1 Pharmacology & Transplantation, INSERM U1248, Université de Limoges, Limoges, France

2 Department of Pediatrics, University Hospital of Limoges, Limoges, France

3 Department of Clinical and Biological Pharmacology and Pharmacovigilance, Clinical Investigation Center CIC-P 1414, Rennes, France.

4 Department of Pharmacology and Physiology, Université de Montréal, Montreal, Quebec, Canada

5 Research Center, Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada;

6 Department of Pediatrics, Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada;

7 Department of Pharmacology, Toxicology and Pharmacovigilance, University Hospital of Limoges, Limoges, France

8 Department of Transplantation Medicine, Oslo University Hospital – Rikshospitalet, Oslo, Norway.

9 Section of Pharmacology and Pharmaceutical Biosciences, Department of Pharmacy, University of Oslo, Oslo, Norway.

## **Corresponding author:**

Jean-Baptiste Woillard

Title: PharmD, PhD

ORCID: 0000-0003-1695-0695

Address:

Univ. Limoges, INSERM U1248 P&T, 2 rue du Pr Descottes, F-87000 Limoges, France.

Phone: +33 5 55 05 61 40

Fax: +33 5 55 05 61 62

Email: [jean-baptiste.woillard@unilim.fr](mailto:jean-baptiste.woillard@unilim.fr)

Keywords: Machine learning; pharmacokinetics; children; transplantation; ganciclovir; valganciclovir; Xgboost; artificial intelligence; pharmacometrics

Financials: the authors have no potential conflicts of interest to disclose.

## Abstract

**Introduction:** Intravenous ganciclovir (GCV) and oral valganciclovir (VGCV) display significant variability in GCV pharmacokinetics, particularly in children. Therapeutic drug monitoring currently relies on the area under the concentration-time (AUC). Machine Learning (ML) algorithms represent an interesting alternative to Maximum-a-Posteriori Bayesian-estimators (MAP-BE) for AUC estimation. The goal of our study was to develop and validate a ML-based limited sampling strategy (LSS) approach to determine GCV  $AUC_{0-24}$  after administration of either intravenous GCV or oral VGCV in children.

**Methods:** PK parameters from four published population pharmacokinetic models, in addition to the WHO growth curve for children, were used in the mrgsolve R package to simulate 10800 PK profiles of children. Different ML algorithms were trained to predict  $AUC_{0-24}$  based on different combinations of 2 or 3 samples. Performances were evaluated in a simulated test set and in an external dataset of real patients.

**Results:** The best estimation performances in the test set were obtained with the Xgboost algorithm using a 2 and 6 hours post dose LSS for oral VGCV (rMPE=0.4%, rRMSE=5.7%) and 0 and 2 hours post dose LSS for IV GCV (rMPE= 0.9%, rRMSE=12.4%). In the external dataset, the performance based on these 2 sample LSS were acceptable: rMPE=0.2%, rRMSE= 16.5% for VGCV and rMPE=-9.7% , rRMSE=17.2% for GCV.

**Conclusion:** The Xgboost algorithm developed resulted in a clinical relevant individual estimation using only 2 blood samples. This will improve the implementation of AUC-targeted GCV TDM in children.

## Introduction

Cytomegalovirus (CMV) is associated with significant morbidity in immunocompromised children, such as those having undergone stem cells or solid organ transplantation. CMV is highly prevalent with a seroprevalence estimated at 55% in pediatric donors or recipients (1). Despite preventive measures such as oral valganciclovir based primary prophylaxis or preemptive therapy including close monitoring of CMV replication, during a defined high-risk period after transplantation, CMV disease can result in significant morbidity and even mortality (2). The most frequently employed drugs for prevention and treatment of CMV disease are intravenous ganciclovir (GCV) and its oral prodrug valganciclovir (VGCV). Common drug-related toxicities include neutropenia, anemia, thrombocytopenia, diarrhea and fever (3).

GCV and VGCV display significant variability in pharmacokinetics, especially in children (4). Studies have identified that renal function, weight, age are predictors of ganciclovir clearance (5). A few pharmacokinetic studies were led in children, and the efficacy and safety targets determined for adults have been extrapolated for the preemptive strategy in children ( $AUC_{0-24} = 40-60 \text{ mg}\cdot\text{h/L}$ ). Pediatric GCV dosing recommendations are 5 mg/kg IV once a day for prophylaxis and the VGCV dosing recommendation are based on an algorithm;  $7 \cdot \text{BSA} \cdot \text{CrCL}$  once a day, where body surface area (BSA) is estimated using the Mosteller formula and creatinine clearance (CrCL) is estimated using the modified Schwartz formula. However, the probability of target attainment with GCV and VGCV is lower in children than in adults (5). Due to its large inter-individual variability, therapeutic drug monitoring (TDM) is recommended in children. As recently reviewed in Franck et al, the best exposure index for therapeutic drug monitoring is the area under the concentration-time (AUC) (5). However, AUC estimation requires multiple blood samples within a dose interval that can be invasive and distressing for children. Furthermore, repeated blood draws demand significant healthcare resources including personnel, laboratory availability and equipment. Numerous studies have been conducted to develop limited strategy samples (LSS) for ganciclovir using Bayesian estimators or multiple linear regression analyses. However, only a few have been proposed to predict ganciclovir  $AUC_{0-24}$  with heterogeneous populations (age, type of transplants) (6, 7).

Training Machine Learning (ML) algorithms requires large datasets. To address this challenge, we have recently demonstrated that ML algorithms could be efficiently trained on simulated data obtained using a population pharmacokinetic (POPPK) model from the literature (8, 9).

Therefore, the goal of our study was to develop and externally validate a ML-based limited sampling strategy (LSS) approach to determine GCV  $AUC_{0-24}$  after administration of either intravenous GCV or oral VGCV in children.

## Material and Methods

### Simulation of ganciclovir pharmacokinetic profiles in children administered IV GCV or oral VGCV

#### Creation of a covariate database

Ten thousand eight hundred simulated patients with different covariates used in previously published POPPK models of GCV and VGCV were simulated (corresponding to 300 profiles by sex and by age from 1 to 18 years old). In brief, age between 1 to 18 years old was simulated using a uniform distribution, weight and height were simulated based on a truncated normal distribution according to the World Health Organization (WHO) growth standard according to age group (1 year per year) (10). Creatinemia was simulated based on a truncated normal distribution between 20 and 250 (micromole/L) independently of the other covariates (7). Body surface area (BSA) was calculated with Mostellers formula ( $BSA = \sqrt{\frac{Weight \times height}{3600}}$ ). Sex and type of transplant (solid organ or stem cells) were simulated using a uniform distributions. The creatinine clearance (CrCL) was then calculated based on the simulated age and creatininemia using the modified Schwarz formula (11).

#### Monte Carlo Simulations

We performed Monte Carlo simulations in the mrgsolve R package (12) using previously published POPPK models (2 for GCV (7, 13) and 3 for VGCV (7, 13, 14)), the simulated individual covariates and dose recommendations for a prophylaxis treatment (Dose<sub>guidelines</sub> (4, 15, 16); 5 mg/kg once a day for GCV and  $7 \times BSA \times CrCL$  for VGCV, max 900mg). Individual predicted concentrations after IV GCV and VGCV administration at steady state (ss=1 option in mrgsolve) were simulated hourly. The additive error and proportional error were set close to 0 to only take into account the inter-individual variability and covariate effect as previously described (17). The steady state  $AUC_{0-24h,ref}$  was calculated using trapezoidal rule ( $AUC_{0-24,ref}$ ). Filters were applied to remove  $AUC_{0-24-ref}$  outliers (values outside the 1%-99% interval of simulated values) and  $AUC_{0-24-ref}$  higher than 120 mg\*h/L. All the values of covariates, the code used to simulate them and the supplemental data are available at [https://github.com/ponthL/LSS\\_ganciclovir](https://github.com/ponthL/LSS_ganciclovir). Methods are resumed in the supplemental Figure 1.

#### Machine learning analysis

All the pre-processing and machine learning analyses were performed using the tidymodels framework in R version 4.2.2 (18). Analyses were performed independently for GCV and VGCV. Data were split randomly into a training set (75%) and a test set (25%). Subsequently, the training set was further divided into an analysis set (80%) and an assessment set (20%). This latter split allowed us to compare ML algorithms efficiently and select only the one with the best performance while preserving the test set for the final validation. Our preprocessing steps involved normalization (centering and scaling) of numeric variables and one-hot encoding of categorical variables. We benchmarked different algorithms in the analysis set including Xgboost (eXtreme Gradient Boosting Training) (19), MARS (Multivariate Adaptive Regression Splines) (20), GLMNET (generalized linear model via penalized maximum likelihood) (21), Random Forests (22), SVM (support vector Machine) (23) for different LSS of 2 samples: 0/2 hours, 0/3 hours, 0/4 hours, 1/3 hours, 1/4 hours, 2/4 hours, 0/6 hours, 1/6 hours and 2/6 hours. We did not investigate combinations of concentrations 1 hour apart because we considered that this was too close and limited the last sampling point to six hours for clinical applicability. We also investigated combinations of 3 samples among the same time points (results presented in the supplemental table).

For each algorithm and LSS, hyperparameters were tuned using a ten-fold cross-validation in the analysis set based on RMSE and  $R^2$ . Then, optimized algorithms were evaluated in the assessment set and the one associated with the lowest mean predictive error (MPE) and the root mean square predictor error (RMSE) was selected. Finally, the algorithm selected was redeveloped by gathering the analysis and assessment sets and was evaluated in the test set. Variable importance plots were drawn using random permutations. The performances of the ML algorithm (MPE & RMSE) were compared to that of the Maximum a Posteriori Bayesian estimations using the same selected concentrations points and previously published POPPK models (2 for GCV (7, 13) and 3 for VGCV (7, 13, 14)). MAP-BE were performed with the mapbayR R package (24).

## External evaluation

The final ML algorithms based on 2 concentrations were also evaluated in an external set of 11 GCV  $AUC_{0-24}$  sampled in 9 patients treated with VGCV and 22 GCV  $AUC_{0-24}$  sampled in 17 patients for intravenous GCV administration (with an infusion duration of 1 hour). As in the test set, the performances (MPE & RMSE) were compared to that of the MAP-BE on the published population pharmacokinetic models. The external set were derived from the biological collection of the Sainte Justine Hospital. In brief, blood samples were routinely collected and 0; 0.5; 1; 1.5; 2; 4; 5; 6; and 12 h for VGCV and at 0; 1; 2; 3; 4; 5; 6; and 12 h for GCV. Patients were at steady-state. Patients received a preemptive treatment: 10 mg/kg/every 12 h for VGCV or 5 mg/kg/every

12 h for GCV. Parental informed consent was obtained for all infants. The study protocol was approved by institutional ethics committee (N°2018-1830). The authors confirmed that they have complied with the World Medical Association Declaration of Helsinki regarding ethical conduct of research involving human subjects.

## Results

### Patients

Ten thousand eight hundred GCV PK profiles were simulated for GCV and VGCV, respectively and 10341 were included for VGCV and 10644 for GCV after removing of the 1-99th percentiles outliers. For VGCV, 7753 were randomly assigned to the training set and 2588 in the test set (Table 1). Among the training set for VGCV, 6202 were randomly assigned in an analysis set and 1551 in assessment set. For GCV, 7980 were randomly assigned to the training set and 2664 in the test set (Table 2). Among the training set, 6384 were randomly assigned in an analysis set and 1596 in assessment set. No difference was observed for characteristics between the analysis, assessment sets, training test (analysis+ assessment sets) and the test set.

### Performances of the 2 samples algorithm

Spaghetti plots representing ganciclovir concentration vs time in the simulation dataset for oral VGCV or intravenous GCV are available in supplemental Figure 2. Results of the comparison of the 2 samples LSS-combinations with the 5 different ML algorithms (Glmnet, Xgboost, MARS, Random Forests, SVM) are shown in Table 3 for VGCV and Table 4 for GCV. The combinations of 2- and 6 hours for VGCV and 0- and 2 hours for GCV using the Xgboost showed the best performances in the assessment set.

In the test, the combinations of 2- and 6 hours for VGCV and 0- and 2 hours for GCV using the Xgboost showed the best performances compared to MAP-BE based on published POPPK models (table 5A for VGCV and table 6A for GCV). The scatter plot and Bland–Altman of the final Xgboost algorithm in the test set for predicted vs reference AUC is presented in Figure 1 for VGCV and in Figure 2 for GCV. Variable importance plot are presented in the Figure 3 for VGCV and 4 for GCV.

The performances when split by age are shown in table 5A for VGCV and Table 6A for GCV. The scatter plot and Bland–Altman of the final Xgboost algorithm in the test set for predicted vs reference AUC split by age classes is presented in Figure 5 for VGCV and in Figure 6 for GCV. The scatter plot and Bland–Altman of the MAP-BE

with the different POPPK models in the test set for predicted vs reference AUC are presented in Supplemental Figure 3 for VGCV and in Supplemental Figure 4 for GCV.

Results for comparison of 3 samples LSS-combinations are available in supplemental table: 0-, 1- and 6 hours for VGCV and 0-, 2- and 6 hours for GCV showed best performances in the analysis and assessment set. Adding one more sample did not significantly improve the AUC predictability.

## Evaluation in external patients

Characteristics of external real patients are detailed in Table 1 for VGCV and in Table 2 for GCV. For VGCV, the sampling times ranged between 1.8h and 2.3h for the 2h sample and between 5.9h and 6.3h for the 6h sample. For GCV, the sampling times in patient ranged between 1.9h and 2.4h for the 2h sample.

In external patients, the combinations of 2- and 6 hours for VGCV and 0- and 2 hours for GCV using the Xgboost showed the best performances compared to MAP-BE (table 5B for VGCV and table 6B for GCV). Xgboost algorithm yielded RMSE = 5.1 mg\*h/L and MPE = -0.17 mg\*h/L for VGCV and RMSE = 7.7 mg\*h/L and MPE = 9.7 mg\*h/L for GCV. The scatter plot of predicted AUC vs reference AUC for the Xgboost algorithm based on 2 samples in the external set is presented in Figure 1 for VGCV and in Figure 2 for GCV.

The performances when split by age are shown in table 5B for VGCV and Table 6B for GCV. The scatter plot and Bland–Altman of the Bayesian estimations with different POPPK models in external patients for predicted vs reference AUC are presented in Supplemental Figure 3 for VGCV and in Supplemental Figure 4 for GCV.

## Discussion

The Xgboost algorithm developed here allowed a good AUC estimation for VGCV using concentrations sampled at 2 and 6 hours and for GCV using concentrations sampled at 0 and 2 hours. The same samples included in combinations of 3 sample LSS but did not result in clinically significant better estimations. These samples schedule are easy to draw in a routine practice. The aim of our study was to simplify the management of AUC estimation in these young patients, by limiting the number of samples and the hardship of repeated blood sampling.

For VGCV, ML algorithms outperformed the MAP-BE based on published POPPK models, showing a decrease in bias and imprecision in both the test and external datasets. This suggests greater generalizability and robustness of the estimates provided by the ML algorithms compared to those from MAP-BE. For GCV, although the metrics



were the lowest in the test set, the bias observed in the external set is high (almost 10%). Nevertheless, the imprecision is low, and the bias value can be accounted for in the interpretation of the estimation.

A MAP-BE based on a POPPK model developed for children with solid organ or stem cells transplants has already shown that a sample drawn at maximum concentration and 6 hours were important in the limited sampling strategy (25). Another recent study developed a multilinear regression to predict  $AUC_{0-24h}$  with only 2 points (concentrations sampled at 1 and 8 hours ) for VGCV in renal transplants children (26). The authors observed good performances in the 2 validation datasets with RMSE = 7.45 and 6.38 mg\*h/L and MPE= 0.54 and 3.93 mg\*h/L . In our study, our 2-sample ML algorithm yielded RMSE = 5.1 mg\*h/L and MPE = -0.17 mg\*h/L for VGCV in the external dataset. Another advantage of our ML algorithm is also that the 6 hour concentration (or less) is easier to draw in routine care in comparison with an 8 hour concentration. In addition, 6 hour concentration versus 8 hour concentration did not improve the results in our dataset (data not shown). Another model developed with linear regression in adult renal transplants included the concentration at 0 hour and 4 hours (27). The model exhibited good performances in the validation dataset with rMAPE = 12.1 +/-9.5% and rMPE= 5.9+/-14.5%. In our study, the POPPK models used to simulate ganciclovir AUCs were built based on heterogeneous populations in terms of age, dose, height, weight and number of patient used to developed the models. This could be an explanation for the decreased performances in our external dataset in comparison to the isolated literature studies which evaluate their performance in a split or in a comparable validation population sampled from the same center. Nevertheless the performances are good, particularly for VGCV and would allow our algorithm to be usable in heterogeneous populations.

Due to high variability of GCV PK in children, AUC-targeted TDM is important even if the therapeutic window have been extrapolated from adults and have to be confirmed in children (5). Inadequate exposure could potentially result in breakthrough viremia, particularly in high risk transplants recipients, or adverse events (7). However in our study , we didn't access the exposure/response relationship and it would have been interesting. Nethertheless, repetitive blood tests can be a source of stress, pain and anxiety in children, and we must strive to develop algorithms that accurately predict exposure with limited sampling strategies.

In our study, we investigated 5 different ML algorithms: Glmnet which is a penalized regression, based on a linear relationship between predictors and AUC, the others being based on nonlinear relationships (Xgboost is an ensemble method that aggregate decision trees, MARS is a non-parametric derived regression that break a given distribution into small linear pieces, Random forrest use multiple decision trees and SVM finds a function that

predicts continuous target variables by maximizing the margin while minimizing the error between predicted and real values).

Monte Carlo simulations were drawn from 2 published POPPK models for IV ganciclovir (7, 13) and 3 for valganciclovir (7, 13, 14) to simulate AUC with dosing recommendations. We selected only POPPK models that included covariates to enhance the number of predictors. That lead us to remove the Pescovitz et al model from our simulations (28). Concerning the covariates, we simulated each weight and height independently using the World Health Organization growth data and based on gender. In order to prevent unrepresentative combinations, we used a small granularity (one year per year) and we removed outliers' values of  $AUC_{0-24}$ . One of the advantages of using several published models for the simulation is that the population used to develop the ML algorithm is not linked to one or another POPPK model and should theoretically be more representative of the target population. In addition, the dose used in the simulations was the one from dosing recommendations.

As expected, the two most important variables were the concentration of GCV, with only minor importance attributed to the remaining covariates. This is consistent with known findings in population pharmacokinetics. However, there are some limits to this approach, (i) the simulations can be very different leading to very different profiles and not very accurate estimation in an external set (as for GCV in the present study with a large variability of CL between the model used in relation to covariates), that problem was highlighted by the poor performances in the test sets of some of the POPPK models and showed that the ML algorithms were more generalisables ; (ii) dose recommendations for VGCV were overall higher in the simulations than the doses received by the patients in our external dataset, because we had fewer patients with impaired renal function and preemptive therapy were used for these patients leading to decrease the performances in the external dataset.

In additions to these limitations, the data concerning the external set were retrospective. The food intake was also not studied for valganciclovir administration, that could have an impact on exposure (29). Secondly, for GCV, the sampling range was large for the theoretical concentration at 6 hours and could explain the worst performances in comparison to the test set. In other words, deviating from the theoretically optimal sampling time could degrade the model's predictive accuracy. Nevertheless, we have chosen to take a wide interval of tolerance for sampling times because this corresponds to the real clinical practice, even if the deviation from the theoretical times was not considered in our algorithm. In previous works, we included a variable taking into account the deviation from theoretical sampling time (either in true profiles or simulated profiles) and we showed that this variable only minimally influenced the estimations (9, 30). Finally, some patients in the external base had reference AUC calculations using the trapezoidal rule based on a variable number of concentrations. Most of them had 8

concentrations but some had only 7 samples. The performances in the external population were poorer for GCV when compared to VGCV and when we looked at the performances according to the age class of the patient for GCV, the oldest children showed worse performance compared to younger children. On the contrary, the performances were similar according to the age classes for VGCV. The AUC showed large variability for the oldest group of patients receiving the same dosage (5 mg/kg). Franck et al showed for a similar dosing a high variation of GCV exposure and particularly for the oldest children (7).

While this algorithm aims to improve target attainment, it's important to consider the relationship between target attainment and clinical effect (both desired efficacy and avoidance of adverse reactions). These targets are often established in adult studies (31). However, a recent literature review by Franck et al. highlights relationships between AUC and effect in children (5). An additional benefit lies in its ability to estimate AUC with only two data points, compared to other methods.

In order to make a demonstration and as a simple equation cannot be derived, we developed shiny.Apps : [https://pharmacophd.shinyapps.io/ganciclovir\\_LSS\\_PO/](https://pharmacophd.shinyapps.io/ganciclovir_LSS_PO/) for VGCV and [https://pharmacophd.shinyapps.io/Ganciclovir\\_LSS\\_IV/](https://pharmacophd.shinyapps.io/Ganciclovir_LSS_IV/) for GCV. We illustrate the application of this algorithm with an example. In practice, as the GCV half life is very short (~3-4h), the steady state is reached after the first administration. Consider a girl who starts treatment with a dose of 50 mg for stem cell transplantation at T0 and T24h, weighs 13 kg, height 95 cm, is 3 years old, and has a creatinine level of 50  $\mu\text{mol/L}$ . We measure the patient's drug concentration at two time points: at 26 and 30 hours after starting treatment (resulting in concentrations of 7.9 mg/L and 2.3 mg/L, respectively). Using a Shiny app, we calculate the GCV  $\text{AUC}_{0-24}$  which in this case is 82 mg\*h/L. Based on this value, we propose a new dose of 31 mg for the next administration, following the formula:  $\text{current dose} * 50 / \text{AUC (mg/L)}$ .

In conclusion, the Xgboost algorithm developed yielded clinically relevant individual GCV  $\text{AUC}_{0-24}$  estimations with only 2 blood samples in pediatric transplants for both oral VGCV and intravenous GCV. Further prospective investigations are needed to confirm its clinical benefit in this population.

## References

1. Behrendt CE, Rosenthal J, Bolotin E, Nakamura R, Zaia J, Forman SJ. 2009. Donor and recipient CMV serostatus and outcome of pediatric allogeneic HSCT for acute leukemia in the era of CMV-preemptive therapy. *Biol Blood Marrow Transplant J Am Soc Blood Marrow Transplant* 15:54–60.
2. Manuel O, Kralidis G, Mueller NJ, Hirsch HH, Garzoni C, van Delden C, Berger C, Boggian K, Cusini A, Koller MT, Weisser M, Pascual M, Meylan PR, Swiss Transplant Cohort Study. 2013.

- Impact of antiviral preventive strategies on the incidence and outcomes of cytomegalovirus disease in solid organ transplant recipients. *Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg* 13:2402–2410.
3. Bateman CM, Kesson A, Powys M, Wong M, Blyth E. 2021. Cytomegalovirus Infections in Children with Primary and Secondary Immune Deficiencies. *Viruses* 13:2001.
  4. Franck B, Autmizguine J, Marquet P, Ovetchkine P, Woillard J-B. 2022. Pharmacokinetics, Pharmacodynamics, and Therapeutic Drug Monitoring of Valganciclovir and Ganciclovir in Transplantation. *Clin Pharmacol Ther* 112:233–276.
  5. Franck B, Autmizguine J, Marquet P, Ovetchkine P, Woillard J-B. 2022. Pharmacokinetics, Pharmacodynamics, and Therapeutic Drug Monitoring of Valganciclovir and Ganciclovir in Transplantation. *Clin Pharmacol Ther* 112:233–276.
  6. Stockmann C, Roberts JK, Knackstedt ED, Spigarelli MG, Sherwin CM. 2015. Clinical pharmacokinetics and pharmacodynamics of ganciclovir and valganciclovir in children with cytomegalovirus infection. *Expert Opin Drug Metab Toxicol* 11:205–219.
  7. Franck B, Woillard J-B, Théorêt Y, Bittencourt H, Demers E, Briand A, Marquet P, Lapeyraque A-L, Ovetchkine P, Autmizguine J. 2021. Population pharmacokinetics of ganciclovir and valganciclovir in paediatric solid organ and stem cell transplant recipients. *Br J Clin Pharmacol* 87:3105–3114.
  8. Ponthier L, Ensueque P, Destere A, Marquet P, Labriffe M, Jacqz-Aigrain E, Woillard J-B. 2022. Optimization of Vancomycin Initial Dose in Term and Preterm Neonates by Machine Learning. *Pharm Res* 39:2497–2506.
  9. Woillard J-B, Labriffe M, Prémaud A, Marquet P. 2021. Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: The example of tacrolimus. *Pharmacol Res* 167:105578.
  10. <https://www.who.int/tools/child-growth-standards>.
  11. Schwartz GJ, Brion LP, Spitzer A. 1987. The use of plasma creatinine concentration for estimating glomerular filtration rate in infants, children, and adolescents. *Pediatr Clin North Am* 34:571–590.
  12. Elmokadem A, Riggs MM, Baron KT. 2019. Quantitative Systems Pharmacology and Physiologically-Based Pharmacokinetic Modeling With mrgsolve: A Hands-On Tutorial. *CPT Pharmacomet Syst Pharmacol* 8:883–893.
  13. Nguyen T, Oualha M, Briand C, Bendavid M, Béranger A, Benaboud S, Tréluyer J-M, Zheng Y, Foissac F, Winter S, Gana I, Boujaafar S, Lopez V, Berthaud R, Demir Z, Bouazza N, Hirt D. 2021. Population Pharmacokinetics of Intravenous Ganciclovir and Oral Valganciclovir in a Pediatric Population To Optimize Dosing Regimens. *Antimicrob Agents Chemother* 65:e02254-20.
  14. Facchin A, Elie V, Benyoub N, Magreault S, Maisin A, Storme T, Zhao W, Deschenes G, Jacqz-Aigrain E. 2019. Population pharmacokinetics of ganciclovir after valganciclovir in renal transplant children. *Antimicrob Agents Chemother* 63:e01192-19, AAC.01192-19.
  15. Kotton CN, Kumar D, Caliendo AM, Huprikar S, Chou S, Danziger-Isakov L, Humar A, The Transplantation Society International CMV Consensus Group. 2018. The Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation. *Transplantation* 102:900–931.
  16. Razonable RR, Humar A. 2019. Cytomegalovirus in solid organ transplant recipients-Guidelines of the American Society of Transplantation Infectious Diseases Community of Practice. *Clin Transplant* 33:e13512.
  17. Labriffe M, Woillard J-B, Debord J, Marquet P. 2022. Machine learning algorithms to estimate everolimus exposure trained on simulated and patient pharmacokinetic profiles. *CPT Pharmacomet Syst Pharmacol* <https://doi.org/10.1002/psp4.12810>.
  18. Kuhn M, Wickham H. tidymodels: Easily Install and Load the ‘Tidymodels’ Packages version 1.1.0 from CRAN. <https://rdrr.io/cran/tidymodels/>.
  19. Chen T, Guestrin C. 2016. XGBoost: A Scalable Tree Boosting System, p. 785–794. *In* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco California USA.
  20. Friedman JH, Roosen CB. 1995. An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 4:197–217.

21. Rohart F, Gautier B, Singh A, Lê Cao K-A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 13:e1005752.
22. Breiman L. 2001. Random Forests. *Mach Learn* 45:5–32.
23. Cortes C, Vapnik V. 1995. Support-vector networks. *Mach Learn* 20:273–297.
24. Le Louedec F, Puisset F, Thomas F, Chatelut É, White-Koning M. 2021. Easy and reliable maximum *a posteriori* Bayesian estimation of pharmacokinetic parameters with the open-source R package mapbayr. *CPT Pharmacomet Syst Pharmacol* 10:1208–1220.
25. Franck B, Autmizguine J, Åsberg A, Théorêt Y, Marquet P, Ovetchkine P, Woillard J-B. 2021. Thoroughly Validated Bayesian Estimator and Limited Sampling Strategy for Dose Individualization of Ganciclovir and Valganciclovir in Pediatric Transplant Recipients. *Clin Pharmacokinet* 60:1449–1462.
26. Facchin A, Benyoub N, Elie V, Magreault S, Jacqz-Aigrain E. 2023. Limited Sampling Strategies to Predict Ganciclovir Exposure after Valganciclovir Administration and to Reduce Monitoring Constraints in Renal Transplant Children. *Antimicrob Agents Chemother* 67:e0159722.
27. Rui W-B, An H-M, Shao K, Zhai X-H, Lu J-Q, Hu S-S, Chen B, Zhou P-J. 2019. Limited sampling strategy for the estimation of the area under the concentration-time curve for ganciclovir in Chinese adult renal allograft recipients. *Eur J Clin Pharmacol* 75:677–686.
28. Pescovitz MD, Ettenger RB, Strife CF, Sherbotie JR, Thomas SE, McDiarmid S, Bartosh S, Ives J, Bouw MR, Bucuvalas J. 2010. Pharmacokinetics of oral valganciclovir solution and intravenous ganciclovir in pediatric renal and liver transplant recipients. *Transpl Infect Dis Off J Transplant Soc* 12:195–203.
29. Märtson A-G, Edwina AE, Kim HY, Knoester M, Touw DJ, Sturkenboom MGG, Alffenaar J-WC. 2022. Therapeutic Drug Monitoring of Ganciclovir: Where Are We? *Ther Drug Monit* 44:138–147.
30. Woillard J-B, Labriffe M, Debord J, Marquet P. 2021. Tacrolimus Exposure Prediction Using Machine Learning. *Clin Pharmacol Ther* 110:361–369.
31. Wiltshire H, Paya CV, Pescovitz MD, Humar A, Dominguez E, Washburn K, Blumberg E, Alexander B, Freeman R, Heaton N, Zuideveld KP, Valganciclovir Solid Organ Transplant Study Group. 2005. Pharmacodynamics of oral ganciclovir and valganciclovir in solid organ transplant recipients. *Transplantation* 79:1477–1483.

Table 1: Characteristics of the simulated profiles and real-world patients for VGCV algorithm development.

	All simulated patients N = 10341	Simulated patients in train set (Assessment+analysis) N= 7753	Simulated patients in test set N=2588	Real-world patients N=11
Weight (Kg)	32.3 [19.0;51.8]	32.1 [19.0;51.8]	33.0 [19.1;51.6]	17.5[9.4;36.4]
CrCL (mL/ min)	140 [95.8;207]	140 [95.8;207]	137 [95.6;208]	82.9[69.9;118.8]
Age (years)	9.00 [5.00;14.0]	9.00 [5.00;14.0]	9.00 [5.00;14.0]	5.8 [0.8;8.8]
Height (cm)	139 [115;161]	138 [115;161]	139 [116;161]	106[77;144]
Dose(mg)	900 [583;900]	900 [583;900]	900 [580;900]	250[110;393]
Dose (mg/kg)	20.1 [14.6;34.7]	20.5 [15.2;34.7]	20.2 [14.5;34.0]	12.9 [10.1;13.8]
Male Sex	5179 (50.1%)	3915 (50.5%)	1264 (48.8%)	10 (91%)
Type of transplant				
Transplant stem cells	5189 (50.2%)	3898 (50.3%)	1291 (49.9%)	1 (9%)
Solid organ	5152 (49.8%)	3855 (49.7%)	1297 (50.1%)	10 (91%)
AUC <sub>0-24,ref</sub> (mg*h/L)	55.9 [37.2;75.0]	55.9 [37.2;75.0]	55.9 [37.2;75.0]	26.06[16.3;26.44]
Concentration at 0 hour (mg/L)	0.11 [0.03;0.24]	0.11 [0.03;0.24]	0.11 [0.03;0.25]	0.01[0.0;0.10]
Concentration at 1 hour (mg/L)	7.77 [5.12;10.3]	7.80 [5.13;10.3]	7.69 [5.02;10.2]	4.23[3.07;4.89]
Concentration at 2 hours (mg/L)	9.28 [6.58;12.0]	9.26 [6.58;12.0]	9.33 [6.58;12.0]	4.37[3.37;5.80]
Concentration at 3 hours (mg/L)	7.70 [5.38;10.4]	7.71 [5.40;10.4]	7.70 [5.30;10.4]	3.21[2.5;5.23]
Concentration at 4 hours (mg/L)	6.09 [4.16;8.56]	6.09 [4.17;8.56]	6.07 [4.14;8.60]	2.21[1.57;2.97]
Concentration at 6 hours	3.75 [2.32;5.59]	3.75 [2.33;5.60]	3.76 [2.28;5.54]	0.665[0.37;0.795]

Continuous variables are presented as median [IQR]. CrCL : creatinine Clearance (using modified Schwartz formula). AUC<sub>0-24,ref</sub> : Area under the curve obtained using the trapezoidal rules ;

Table 2: Characteristics of the simulated profiles and external validation of actual patients for GCV algorithm development.

	All simulated patients N = 10644	Simulated patients in train set (Assessment+analysis) N= 7980	Simulated patients in test set N=2664	Real-world patients N=22
Weight (Kg)	33.0 [19.4;51.8]	33.1 [19.1;51.8]	33.1 [19.1;51.8]	21.2[11.25;32.5]
CrCL (mL/ min)	141 [96.8;209]	142 [97.0;209]	140 [95.7;211]	148[111;199]
Age (years)	9.00 [5.00;14.0]	9.00 [5.00;14.0]	10.0 [5.00;14.0]	6.7 [1.5;10.2]
Height (cm)	139 [116;161]	139 [116;161]	139 [116;161]	115[94;139]
Dose(mg)	165 [96.9;259]	164 [97.6;259]	165 [95.5;259]	185[103;390]
Dose (mg/kg)	5 [5;5]	5 [5;5]	5 [5;5]	5.1 [4.95;5.40]
Male Sex	5311 (49.9%)	4000 (50.1%)	1311 (49.2%)	17 (77.3%)
Type of transplant				
Transplant stem cells	5335 (50.1%)	4025 (50.4%)	1310 (49.2%)	3 (13.6%)
Solid organ	5309 (49.9%)	3955 (49.6%)	1354 (50.8%)	19 (86.4%)
AUC 0- 24,ref(mg*h/L)	34.0 [20.1;55.6]	34.0 [20.1;55.6]	34.0 [20.1;55.6]	32.5[24.1;63.7]
Concentration at 0 hour (mg/L)	0.03 [0.01;0.08]	0.03 [0.01;0.08]	0.03 [0.01;0.07]	0.24 [0.12;1.12]
Concentration at 1 hour (mg/L)	6.31 [3.75;10.5]	6.32 [3.76;10.5]	6.27 [3.72;10.5]	6.6[4.8;10.1]
Concentration at 2 hours (mg/L)	4.28 [2.52;7.30]	4.28 [2.52;7.30]	4.28 [2.52;7.34]	3.24 [2.1;4.3]
Concentration at 3 hours (mg/L)	3.02 [1.74;5.26]	3.01 [1.76;5.22]	3.02 [1.74;5.26]	1.6 [1.2;3.4]
Concentration at 4 hours (mg/L)	2.17 [1.21;3.87]	2.17 [1.21;3.87]	2.15 [1.22;3.88]	2.1 [1.7;3.8]
Concentration at 6 hours	1.15 [0.59;2.17]	1.15 [0.59;2.16]	1.15 [0.59;2.18]	0.94 [0.82;0.99]

Continuous variables are presented as median [IQR] and categorical variables are presented as number (%). CrCL : creatinine Clearance (using modified Schwartz formula). AUC<sub>0-24,ref</sub> : Area under the curve obtained using the trapezoidal rules ;

Table 3 : Results of the comparison of the 2 different samples combinations in the analysis et assessment set for VGCV (in bold is the model that exhibits the best performances)

Models	Relative MPE (%)	Relative RMSE (%)	Relative MPE (%)	Relative RMSE (%)
	Analysis set	Analysis set	Assesment set	Assesment set
Valganciclovir concentration at 0,2 hours with GLMNET algorithm	4.7	25.6	4.6	26.3
Vaganciclovir concentration at 0, 2 hours with Xgboost algorithm	1.9	14.4	4.6	25.3
Valganciclovir concentration at 0, 2 hours with MARS algorithm	2.9	18.9	2.9	18.5
Valganciclovir concentration at 0, 2 hours with Random Forrest algorithm	2.3	14.6	2.1	14.9
Valganciclovir concentration at 0, 2 hours with SVM algorithm	4.9	26.1	5.2	26.6
Valganciclovir concentration at 0, 3 hours with GLMNET algorithm	0.5	11.4	0.5	11.7
Valganciclovir concentration at 0, 3 hours with Xgboost algorithm	0.7	8.5	0.3	6.7
Valganciclovir concentration at 0, 3 hours with MARS algorithm	0.9	9.9	0.9	10.6
Valganciclovir concentration at 0, 3 hours with Random Forrest algorithm	0.8	8.4	0.9	8.8
Valganciclovir concentration at 0, 3 hours with SVM algorithm	0.3	11.8	0.5	12.6
Valganciclovir concentration at 0, 4 hours with GLMNET algorithm	0.7	8.5	0.8	8.8
Valganciclovir concentration at 0, 4 hours with Xgboost algorithm	0.8	8.7	0.3	4.9
Valganciclovir concentration at 0, 4 hours with MARS algorithm	0.7	8.8	0.8	8.1
Valganciclovir concentration at 0, 4 hours with Random Forrest algorithm	0.9	8.9	1.1	8.9
Valganciclovir concentration at 0, 4 hours with SVM algorithm	0.5	8.8	0.5	9.1
Valganciclovir concentration at 1,3 hours with GLMNET algorithm	3.1	17.1	3.0	16.7
Valganciclovir concentration at 1,3 hours with Xgboost algorithm	1.6	12.8	1.2	9.6
Valganciclovir concentration at 1,3 hours with MARS algorithm	2.1	16.3	1.9	15.8
Valganciclovir concentration at 1,3 hours with Random Forrest algorithm	1.8	12.4	2.6	12.6
Valganciclovir concentration at 1,3 hours with SVM algorithm	3.2	17.0	4.3	17.3
Valganciclovir concentration at 2,4 hours with GLMNET algorithm	2.2	13.6	2.4	13.8
Valganciclovir concentration at 2,4 hours with Xgboost algorithm	0.7	8.9	0.6	6.9
Valganciclovir concentration at 2,4 hours with MARS algorithm	0.9	10.9	0.8	10.9
Valganciclovir concentration at 2,4 hours with Random Forrest algorithm	0.7	8.2	1.4	8.2
Valganciclovir concentration at 2,4 hours with SVM algorithm	2.4	13.8	3.0	14.3
Valganciclovir concentration at 1,4 hours with GLMNET algorithm	2.1	12.7	2.2	12.9
Valganciclovir concentration at 1,4 hours with Xgboost algorithm	0.5	8.5	0.4	6.3
Valganciclovir concentration at 1,4 hours with MARS algorithm	0.6	11.1	0.7	10.4



Valganciclovir concentration at 1,4 hours with Random Forrest algorithm	0.9	8.3	1.3	8.1
Valganciclovir concentration at 1,4 hours with SVM algorithm	1.9	12.9	2.6	13.2
Valganciclovir concentration at 0,6 hours with GLMNET algorithm	3.2	17.0	3.4	18.4
Valganciclovir concentration at 0,6 hours with Xgboost algorithm	1.9	14.9	1.8	13.6
Valganciclovir concentration at 0,6 hours with MARS algorithm	1.9	15.6	2.0	16.2
Valganciclovir concentration at 0,6 hours with Random Forrest algorithm	2.2	15.0	2.4	15.2
Valganciclovir concentration at 0,6 hours with SVM algorithm	2.9	16.9	2.7	17.8
Valganciclovir concentration at 1,6 hours with GLMNET algorithm	1.3	9.0	1.6	10.1
Valganciclovir concentration at 1,6 hours with Xgboost algorithm	0.4	6.3	0.2	4.1
Valganciclovir concentration at 1,6 hours with MARS algorithm	0.6	7.3	0.7	7.6
Valganciclovir concentration at 1,6 hours with Random Forrest algorithm	0.5	6.3	0.7	6.6
Valganciclovir concentration at 1,6 hours with SVM algorithm	1.4	9.2	1.5	9.9
Valganciclovir concentration at 2,6 hours with GLMNET algorithm	0.9	8.4	1.1	8.7
<b>Valganciclovir concentration at 2,6 hours with Xgboost algorithm</b>	<b>0.4</b>	<b>5.7</b>	<b>0.2</b>	<b>3.7</b>
Valganciclovir concentration at 2,6 hours with MARS algorithm	0.6	7.5	0.4	7.2
Valganciclovir concentration at 2,6 hours with Random Forrest algorithm	0.4	5.4	0.8	5.8
Valganciclovir concentration at 2,6 hours with SVM algorithm	0.8	8.6	1.1	8.8
Valganciclovir concentration at 3,6 hours with GLMNET algorithm	1.8	11.8	1.9	12.5
Valganciclovir concentration at 3,6 hours with Xgboost algorithm	0.7	7.4	0.4	5.0
Valganciclovir concentration at 3,6 hours with MARS algorithm	0.7	8.6	0.7	8.7
Valganciclovir concentration at 3,6 hours with Random Forrest algorithm	0.7	7.2	1.2	7.4
Valganciclovir concentration at 3,6 hours with SVM algorithm	1.8	11.7	2.1	12.4
Valganciclovir concentration at 4,6 hours with GLMNET algorithm	1.6	11.5	1.6	11.8
Valganciclovir concentration at 4,6 hours with Xgboost algorithm	0.9	9.1	0.6	6.2
Valganciclovir concentration at 4,6 hours with MARS algorithm	1.3	10.6	1.2	10.2
Valganciclovir concentration at 4,6 hours with Random Forrest algorithm	0.8	8.7	1.5	8.9
Valganciclovir concentration at 4,6 hours with SVM algorithm	1.7	12.1	1.9	12.4

Table 4 : Results of the comparison of the 2 different samples combinations in the analysis et assessment set for GCV (in bold is the model that exhibits the best performances)

Models	Relative MPE (%)	Relative RMSE (%)	Relative MPE (%)	Relative RMSE (%)
	Analysis set	Analysis set	Assesment set	Assesment set
Ganciclovir concentration at 0,2 hours with GLMNET algorithm	2.4	13.8	2.2	13.7
<b>Ganciclovir concentration at 0, 2 hours with Xgboost algorithm</b>	<b>1.7</b>	<b>13.1</b>	<b>1.0</b>	<b>8.4</b>
Ganciclovir concentration at 0, 2 hours with MARS algorithm	1.9	13.2	1.5	13.0
Ganciclovir concentration at 0, 2 hours with Random Forrest algorithm	1.9	13.3	1.9	13.1
Ganciclovir concentration at 0, 2 hours with SVM algorithm	2.1	13.9	2.1	13.6
Ganciclovir concentration at 0, 3 hours with GLMNET algorithm	2.13	13.8	1.99	13.7
Ganciclovir concentration at 0, 3 hours with Xgboost algorithm	1.8	13.8	0.9	8.4
Ganciclovir concentration at 0, 3 hours with MARS algorithm	1.8	13.4	1.6	13.5
Ganciclovir concentration at 0, 3 hours with Random Forrest algorithm	1.8	13.8	1.4	13.6
Ganciclovir concentration at 0, 3 hours with SVM algorithm	1.8	13.9	1.6	13.6
Ganciclovir concentration at 0, 4 hours with GLMNET algorithm	2.5	15.6	2.3	15.5
Ganciclovir concentration at 0, 4 hours with Xgboost algorithm	2.3	15.8	1.3	9.9
Ganciclovir concentration at 0, 4 hours with MARS algorithm	2.5	15.3	2.3	15.1
Ganciclovir concentration at 0, 4 hours with Random Forrest algorithm	2.4	15.9	1.7	15.6
Ganciclovir concentration at 0, 4 hours with SVM algorithm	2.1	15.6	1.5	15.4
Ganciclovir concentration at 1,3 hours with GLMNET algorithm	2.2	14.4	2.2	14.2
Ganciclovir concentration at 1,3 hours with Xgboost algorithm	1.3	11.8	0.9	9.1
Ganciclovir concentration at 1,3 hours with MARS algorithm	1.9	13.9	1.9	14.2
Ganciclovir concentration at 1,3 hours with Random Forrest algorithm	1.7	11.6	1.6	14.5
Ganciclovir concentration at 1,3 hours with SVM algorithm	1.9	14.1	1.7	14.2
Ganciclovir concentration at 2,4 hours with GLMNET algorithm	2.2	14.0	2.2	14.2
Ganciclovir concentration at 2,4 hours with Xgboost algorithm	1.2	11.6	0.9	8.9
Ganciclovir concentration at 2,4 hours with MARS algorithm	1.9	13.2	1.5	13.6
Ganciclovir concentration at 2,4 hours with Random Forrest algorithm	1.6	11.2	1.5	11.2
Ganciclovir concentration at 2,4 hours with SVM algorithm	2.2	14.0	2.0	14.1
Ganciclovir concentration at 1,4 hours with GLMNET algorithm	2.1	13.7	2.1	13.9
Ganciclovir concentration at 1,4 hours with Xgboost algorithm	1.1	11.3	0.8	8.7

Ganciclovir concentration at 1,4 hours with MARS algorithm	1.8	13.2	1.8	13.3
Ganciclovir concentration at 1,4 hours with Random Forrest algorithm	1.6	11.2	1.5	10.9
Ganciclovir concentration at 1,4 hours with SVM algorithm	1.7	13.4	1.8	13.7
Ganciclovir concentration at 0,6 hours with GLMNET algorithm	3.5	20.6	3.3	19.6
Ganciclovir concentration at 0,6 hours with Xgboost algorithm	2.9	19.5	1.9	13.3
Ganciclovir concentration at 0,6 hours with MARS algorithm	3.3	20.2	2.8	19.8
Ganciclovir concentration at 0,6 hours with Random Forrest algorithm	3.3	19.7	2.5	18.9
Ganciclovir concentration at 0,6 hours with SVM algorithm	3.4	20.4	2.6	19.4
Ganciclovir concentration at 1,6 hours with GLMNET algorithm	1.9	13.2	1.9	13.3
<b>Ganciclovir concentration at 1,6 hours with Xgboost algorithm</b>	<b>1.1</b>	<b>10.9</b>	<b>0.8</b>	<b>8.4</b>
Ganciclovir concentration at 1,6 hours with MARS algorithm	1.6	12.6	1.7	12.7
Ganciclovir concentration at 1,6 hours with Random Forrest algorithm	1.5	10.5	1.3	10.7
Ganciclovir concentration at 1,6 hours with SVM algorithm	2.1	13.7	1.9	13.7
Ganciclovir concentration at 2,6 hours with GLMNET algorithm	2.1	13.1	2.0	13.8
Ganciclovir concentration at 2,6 hours with Xgboost algorithm	1.1	11.1	0.8	8.7
Ganciclovir concentration at 2,6 hours with MARS algorithm	1.6	12.7	1.6	12.8
Ganciclovir concentration at 2,6 hours with Random Forrest algorithm	1.6	10.7	1.3	10.7
Ganciclovir concentration at 2,6 hours with SVM algorithm	1.7	13.9	1.5	14.0
Ganciclovir concentration at 3,6 hours with GLMNET algorithm	2.2	14.2	2.1	14.3
Ganciclovir concentration at 3,6 hours with Xgboost algorithm	1.3	12.2	0.9	9.0
Ganciclovir concentration at 3,6 hours with MARS algorithm	1.6	13.1	1.8	13.4
Ganciclovir concentration at 3,6 hours with Random Forrest algorithm	1.5	11.9	1.4	11.8
Ganciclovir concentration at 3,6 hours with SVM algorithm	2.2	14.8	1.8	14.8
Ganciclovir concentration at 4,6 hours with GLMNET algorithm	1.9	13.7	1.2	19.2
Ganciclovir concentration at 4,6 hours with Xgboost algorithm	0.9	10.4	1.2	8.6
Ganciclovir concentration at 4,6 hours with MARS algorithm	1.7	13.5	2.35	14.9
Ganciclovir concentration at 4,6 hours with Random Forrest algorithm	2.0	13.6	1.7	13.1
Ganciclovir concentration at 4,6 hours with SVM algorithm	2.3	14.1	2.0	14.1

Table 5 : Results of the comparison of different algorithms/models in the test set and external patients for VGCV

Data	Model	Relative MPE (%)	Relative RMSE (%)	Out of the $\pm 20\%$ interval number(%)
Test set	Xgboost algorithm based on 2 and 6 hours	0.4	5.7	20 (0.8%)
	Bayesian estimations based on 2 and 6 hours With Franck model	-1.9	46.6	1107 (43.9%)
	Bayesian estimations based on 2 and 6 hours With Facchin model	1.1	12.9	178 (7%)
	Bayesian estimations based on 2 and 6 hours With Nguyen model	28.9	97.7	917 (36.4%)
	Xgboost algorithm based on 2 and 6 hours age < 5 years old	0.4	5.8	6 (1.0%)
	Xgboost algorithm based on 2 and 6 hours age between 5 and 10 years old	0.3	5.7	6 (0.8%)
	Xgboost algorithm based on 2 and 6 hours age > 10 years old	0.5	5.7	8 (0.6%)
External patients	Xgboost algorithm based on 2 and 6 hours	0.2	16.5	4 (36.3%)
	Bayesian estimations based on 2 and 6 hours with Franck model	37.3	56.4	8 (72.7%)
	Bayesian estimations based on 2 and 6 hours with Facchin model	-7.7	35.2	8 (72.7%)
	Bayesian estimations based on 2 and 6 hours with Nguyen model	108.0	123.0	10 (90.9%)

Table 6: Results of the comparison of different algorithms/models in the test set and external patients for GCV

Data	Model	Relative MPE (%)	Relative RMSE (%)	Out of the $\pm 20\%$ interval number(%)
Test set	Xgboost algorithm based on 0 and 2 hours	0.9	12.4	278 (10.4%)
	Bayesian estimations based on 0 and 2 hours With Franck model	-5.6	44.4	1634 (61.3%)
	Bayesian estimations based on 0 and 2 hours With Nguyen model	4.4	61.8	2016 (75.7%)
	Xgboost algorithm based on 0 and 2 hours age < 5 years old	1.1	12.5	64 (10.6%)
	Xgboost algorithm based on 0 and 2 hours age between 5 and 10 years old	5.6	12.8	78 (10.7%)
	Xgboost algorithm based on 0 and 2 hours age > 10 years old	10.6	12.2	136 (10.2%)
External patients	Xgboost algorithm based on 0 and 2 hours	- 9.7	17.2	6 (28.6%)
	Bayesian estimations based on 0 and 2 hours with Franck model	4.1	56.4	21 (95.5%)
	Bayesian estimations based on 0 and 2 hours with Nguyen model	105.7	125.8	22 (100%)

Figures:

Figure 1: Scatter plot of AUC0-24s estimated using the Xgboost algorithm based on 2 points at 2 and 6 h vs reference trapezoidal AUC0-24 in the test set (A) and in the external data set (C) for VGCV, and corresponding Bland–Altman plots (B) and (D). Difference is the difference between the Reference and the Xgboost AUC0-24s, and mean is the average of both.

Figure 2: Scatter plot of AUC0-24s estimated using the Xgboost algorithm based on 2 points at 1 and 6 h vs reference trapezoidal AUC0-24 in the test set (A) and in the external data set (C) for GCV, and corresponding Bland–Altman plots (B) and (D). Difference is the difference between the Reference and the Xgboost AUC0-24s, and mean is the average of both.

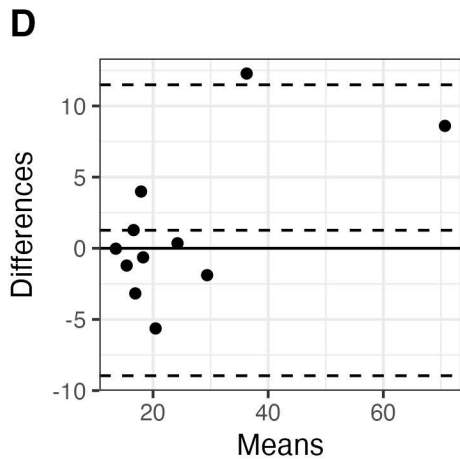
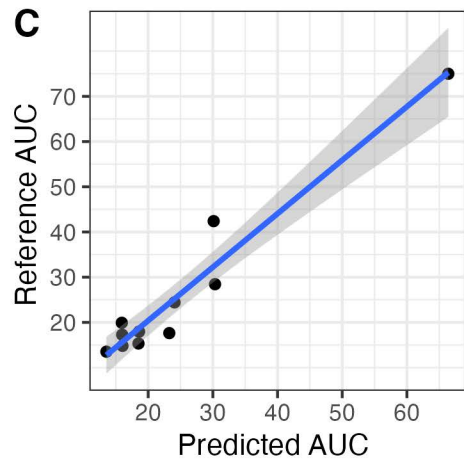
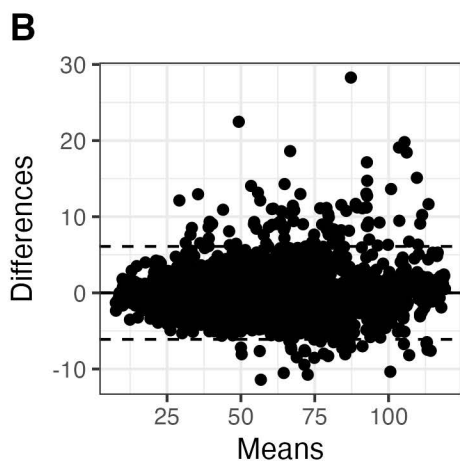
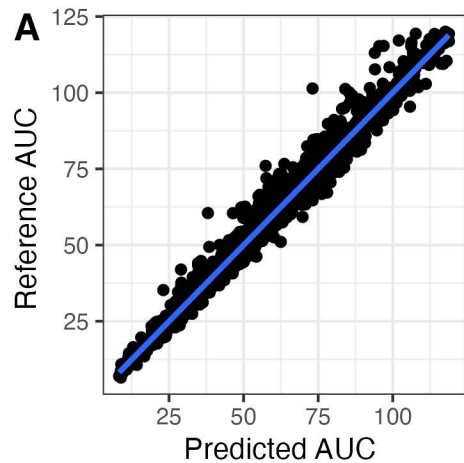
Figure 3: Variable importance plot obtained by random permutation for VGCV. The score obtained quantifies the impact of each variable on the model's predictions

Figure 4: Variable importance plot obtained by random permutation for GCV. The score obtained quantifies the impact of each variable on the model's predictions

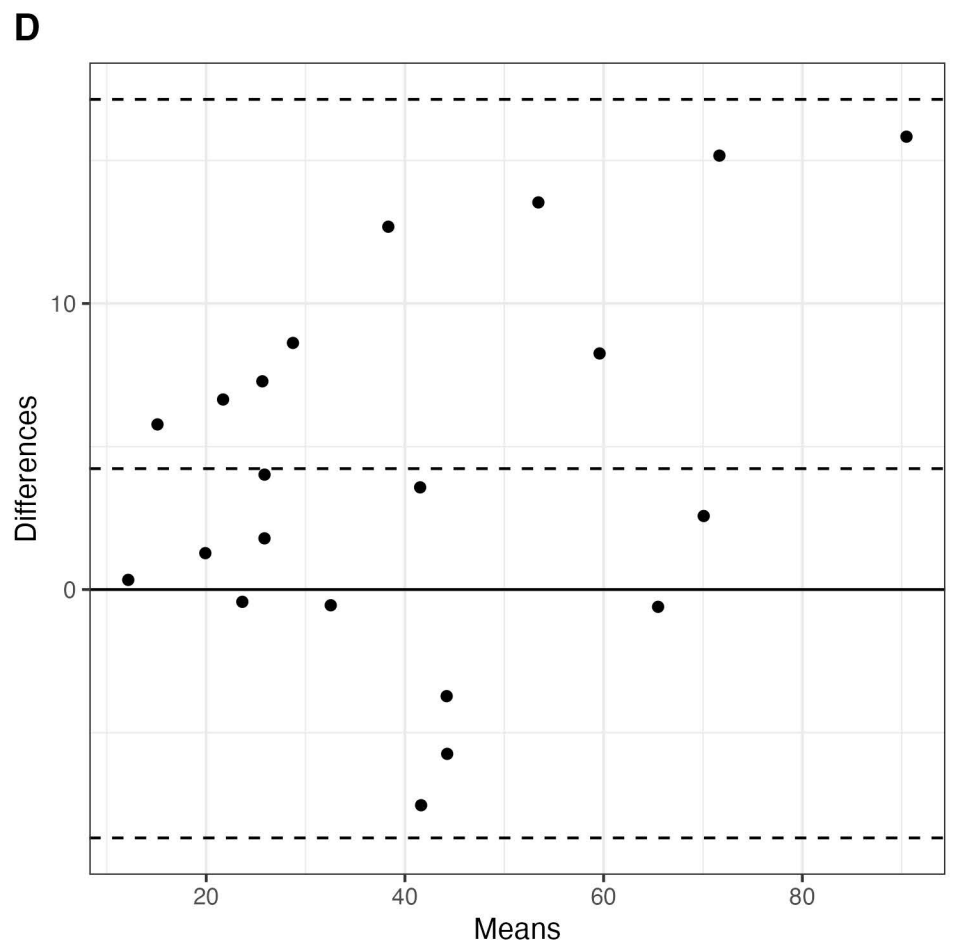
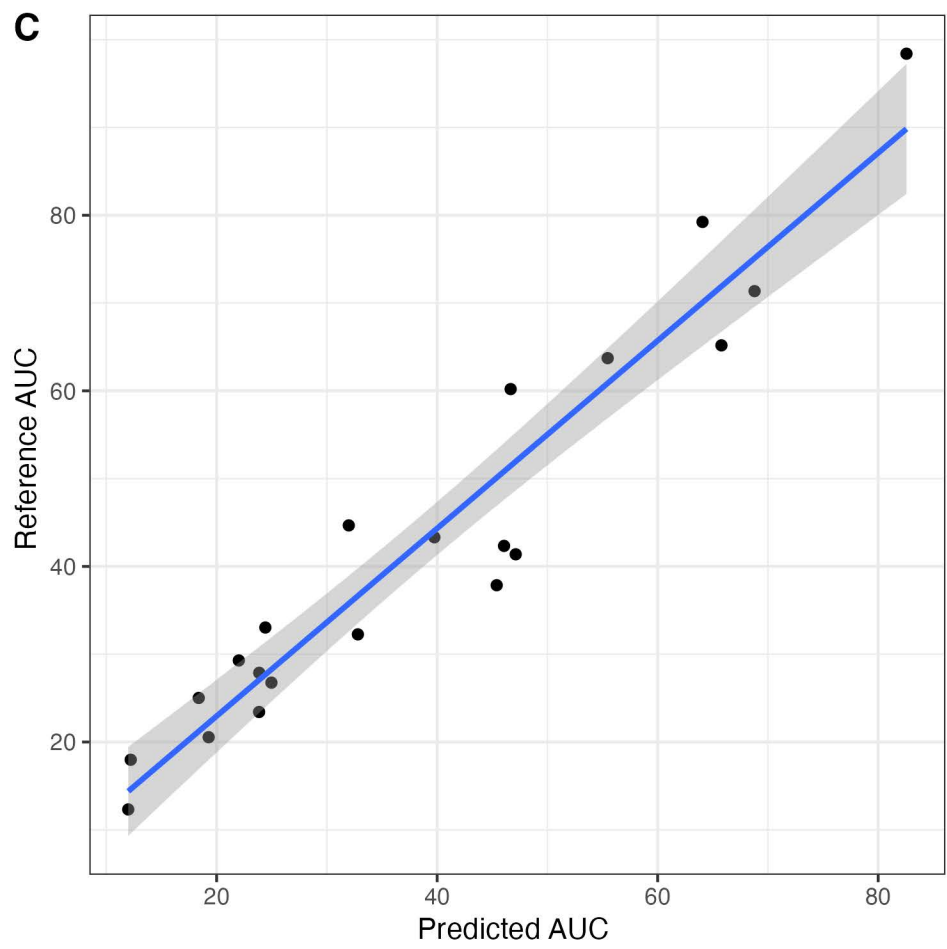
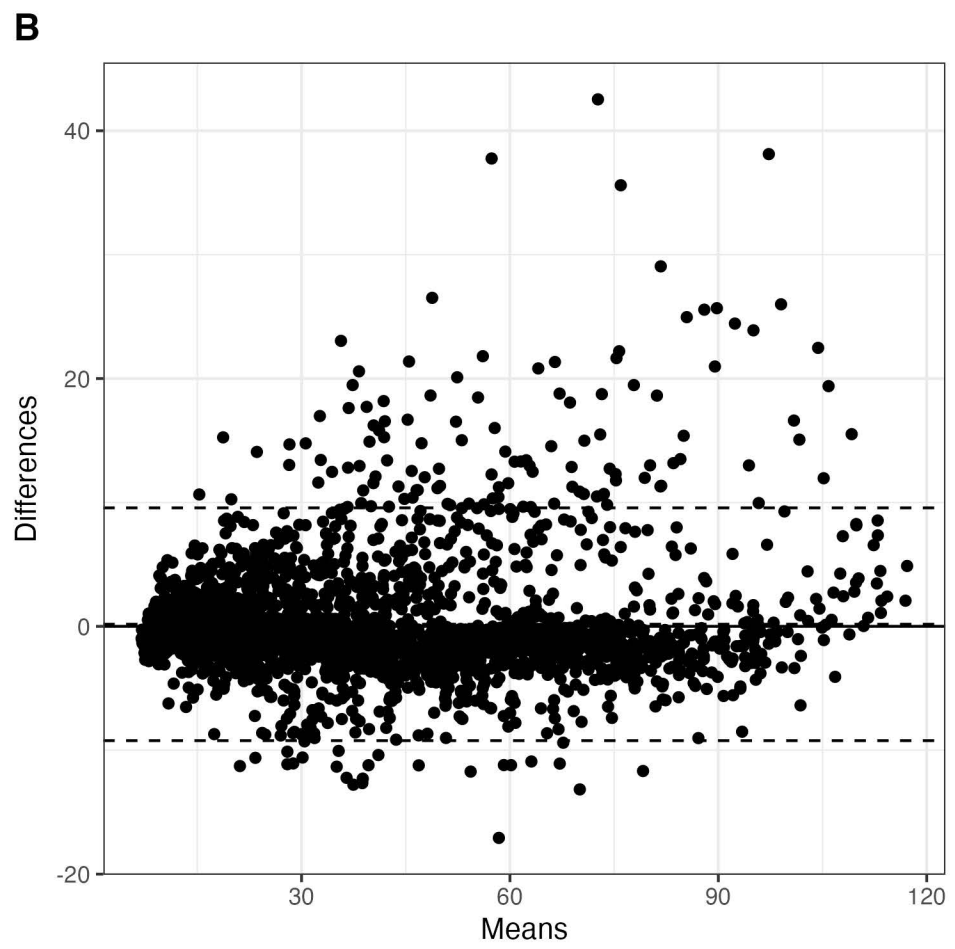
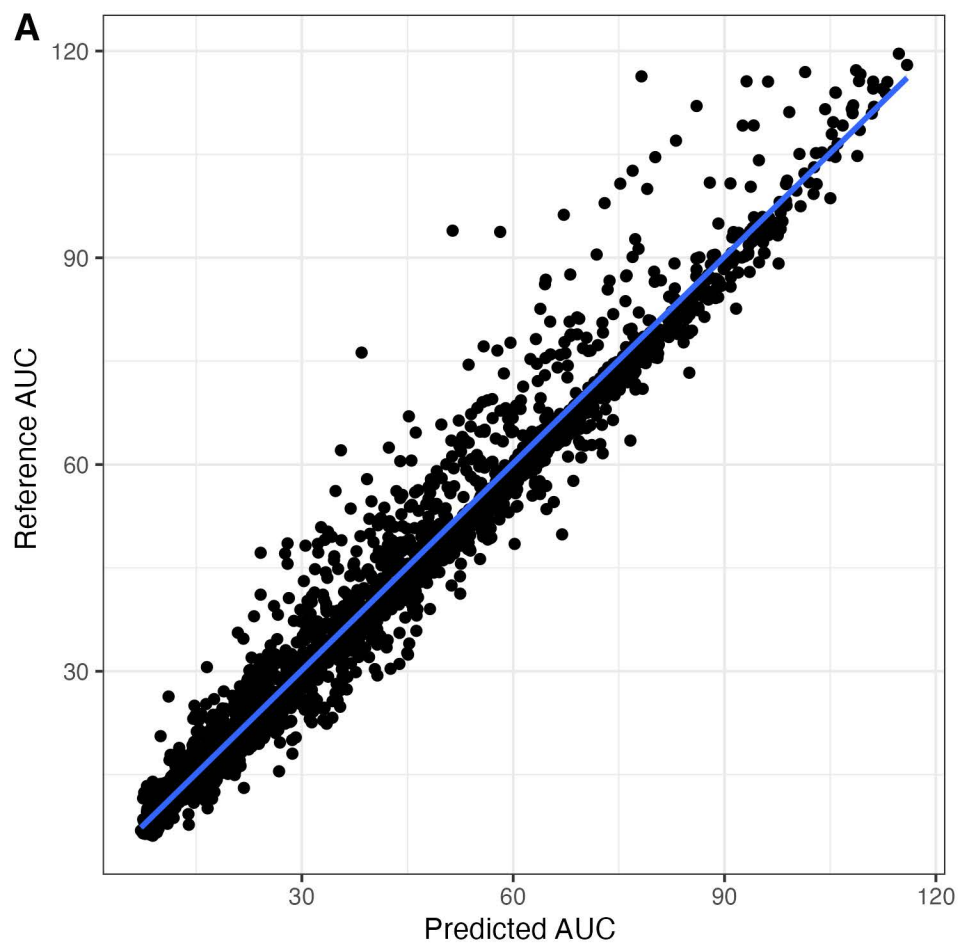
Figure 5: Scatter plot of AUC0-24s estimated using the Xgboost algorithm based on 2 points at 2 and 6 h vs reference trapezoidal AUC0-24 in the test set for VGCV, and corresponding Bland–Altman plots for different classes of age (<5, 5-10 and >10 years old). Difference is the difference between the Reference and the Xgboost AUC0-24s, and mean is the average of both.

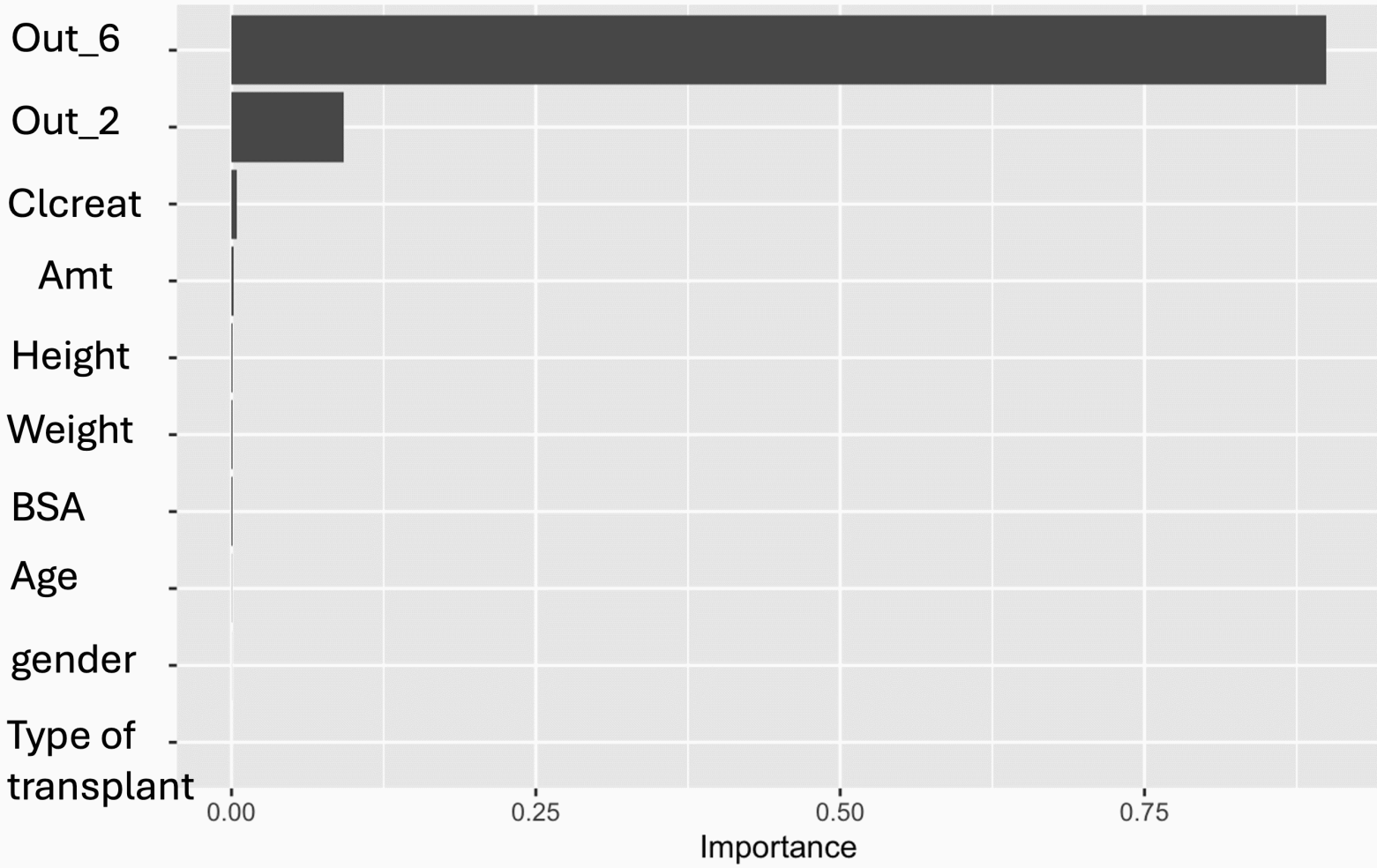
Figure 6: Scatter plot of AUC0-24s estimated using the Xgboost algorithm based on 2 points at 1 and 6 h vs reference trapezoidal AUC0-24 in the test set for GCV, and corresponding Bland–Altman plots for different classes of age (<5, 5-10 and >10 years old). Difference is the difference between the Reference and the Xgboost AUC0-24s, and mean is the average of both.

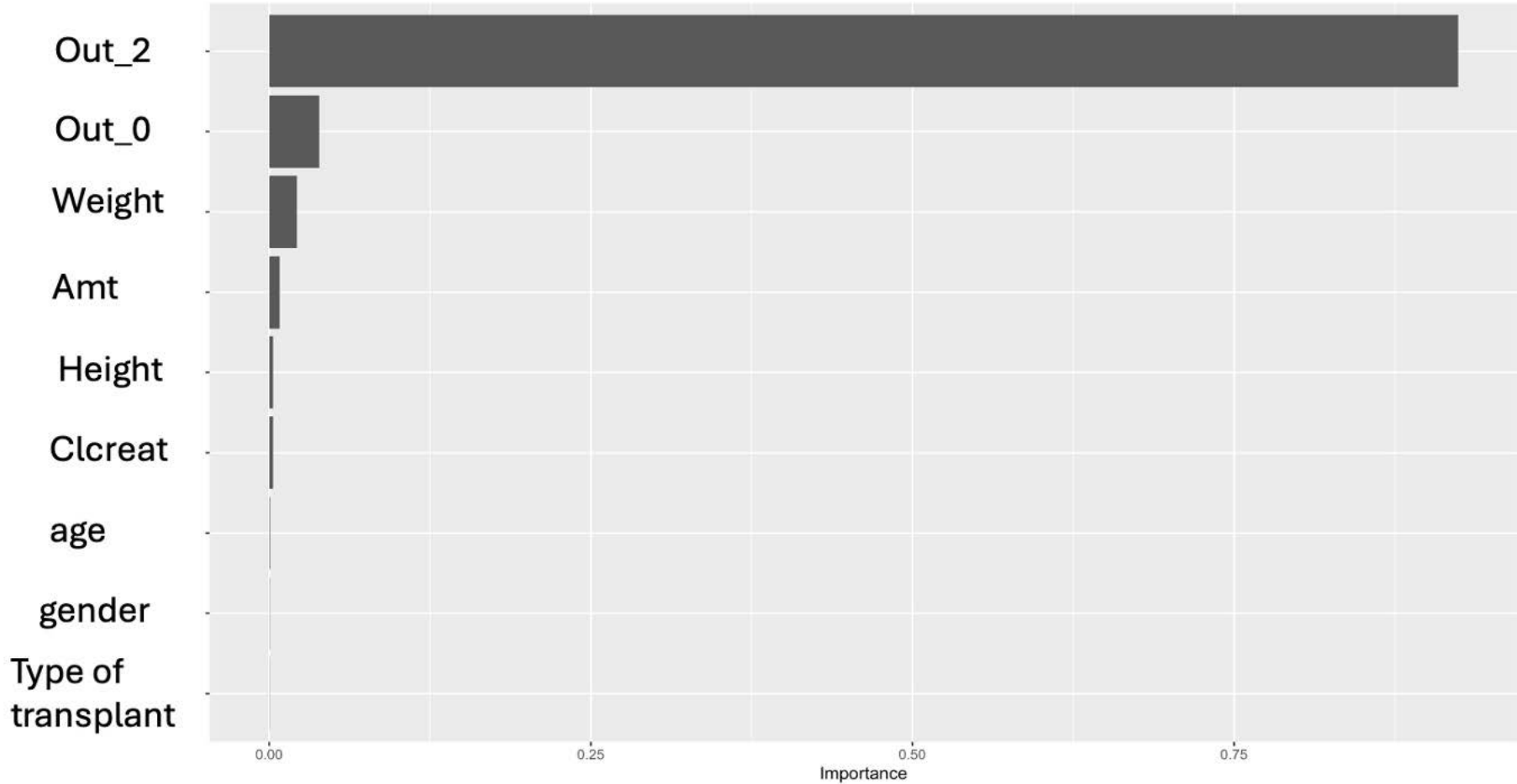




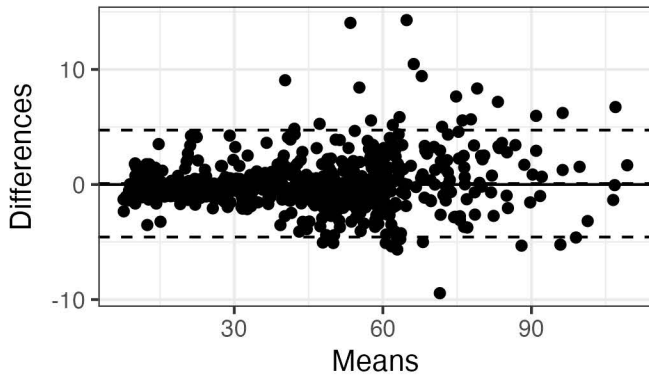
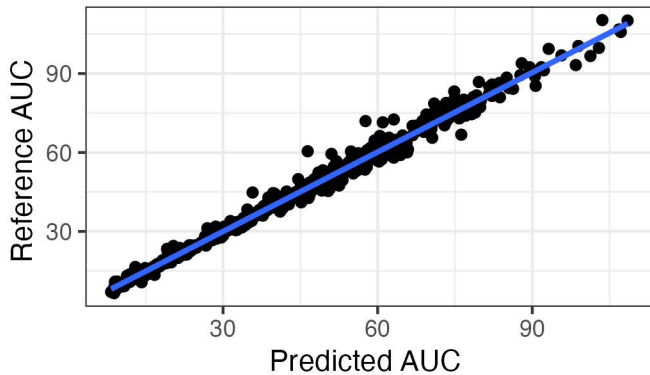




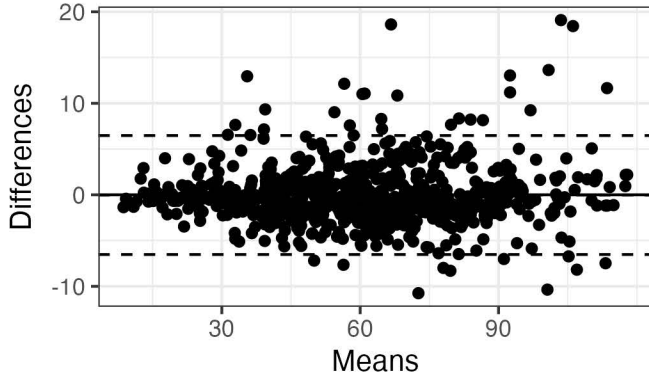
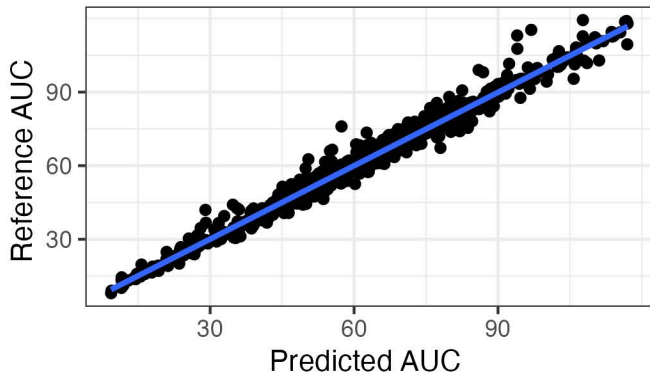




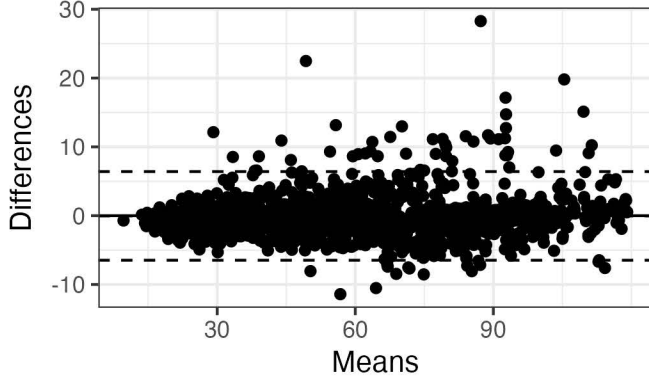
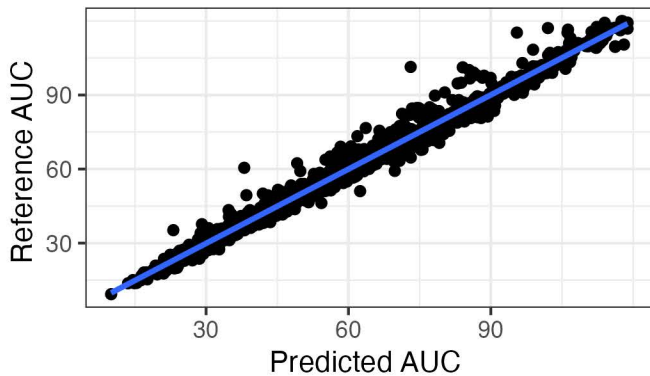
< 5 years



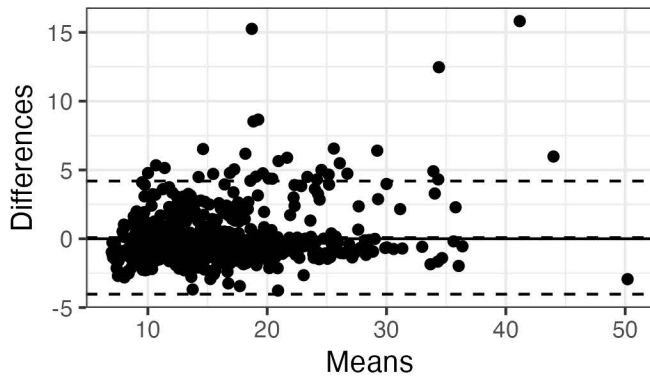
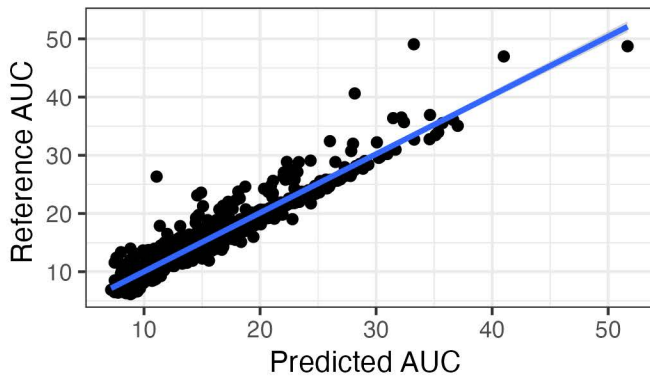
5-10 years



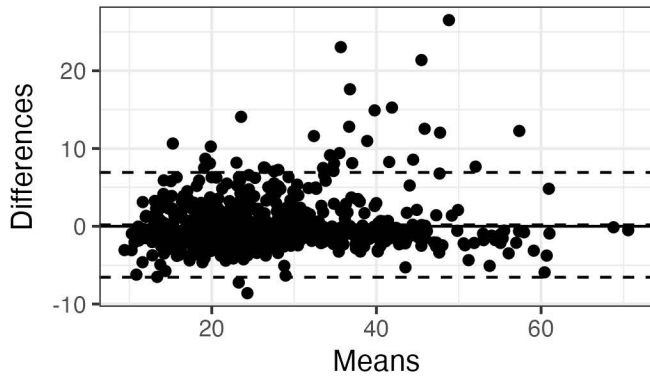
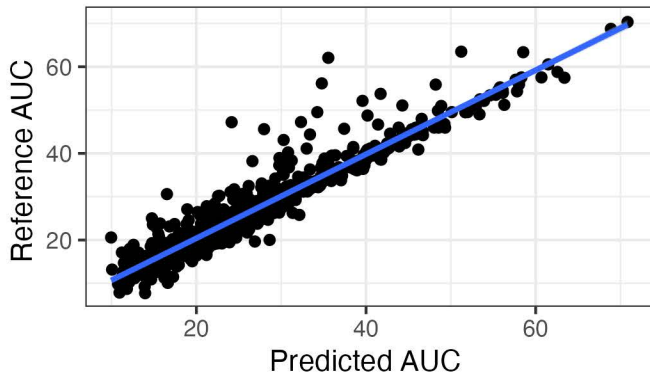
> 10 years



< 5 years



5-10 years



> 10 years

