

1 **A hybrid algorithm combining population pharmacokinetic and**  
2 **machine learning for isavuconazole exposure prediction**

3 Alexandre Destere<sup>1,2</sup>, Pierre Marquet<sup>1,3</sup>, Marc Labriffe<sup>1,3</sup>, Milou-Daniel Drici<sup>2</sup>, Jean-Baptiste  
4 Woillard<sup>1,3</sup>

5 <sup>1</sup> Pharmacology and Transplantation, INSERM U1248, Université de Limoges, 2 rue du Pr  
6 Descottes, 87000 Limoges, France

7 <sup>2</sup> Department of Pharmacology and Pharmacovigilance Center, Côte d’Azur University  
8 Medical Center, Nice, France

9 <sup>3</sup> Department of Pharmacology, Toxicology and Pharmacovigilance, CHU de Limoges,  
10 Limoges, France

11

12 Corresponding author:

13 **Jean-Baptiste Woillard**

14 Title: PharmD, PhD

15 ORCID: 0000-0003-1695-0695

16 Address:

17 Univ. Limoges,

18 INSERM U1248 IPPRITT,

19 2 rue du Pr Descottes,

20 F-87000 Limoges, France.

21 Phone: +33 5 55 05 61 40

22 Fax: +33 5 55 05 61 62

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

24

25 **Funding information:**

26 No funding was received for this study

27

28 **Conflicts of Interest/Competing Interests:**

29 None of the author have any conflicts of interest to declare in relation to this work.

30

31 **Consent for Publication:**

32 All the authors read and approved the final manuscript.

33

34 **Data availability:**

35 Data generated for this study are available upon request to the corresponding author.

36

37 **Authors contributions:**

38 AD, JBW contributed to the conception and design of the study, analysis and interpretation of data,

39 PM contributed to the conception and design of the study and interpretation of data, MLB and MD

40 contributed to the interpretation of data. All the authors participated in drafting the article and

41 approved the final version submitted.

42 **Acknowledgments**

43 We gratefully thank Miss Karen Poole for manuscript editing.

44 **Keywords:** isavuconazole; machine learning; population pharmacokinetics; fungal infections

## 45 Abstract

46

### 47 Objectives

48 Maximum a posteriori Bayesian estimation (MAP-BE) based on a limited sampling strategy  
49 and a population pharmacokinetic (POPPK) model is used to estimate individual  
50 pharmacokinetic parameters. Recently, we proposed a methodology that combined  
51 population pharmacokinetic and machine learning (ML) to decrease the bias and imprecision  
52 in individual iohexol clearance prediction. The aim of this study was to confirm the previous  
53 results by developing a hybrid algorithm combining POPPK, MAP-BE and ML that accurately  
54 predicts isavuconazole clearance.

### 55 Methods

56 A total of 1727 isavuconazole rich PK profiles were simulated using a POPPK model from the  
57 literature, and MAP-BE was used to estimate the clearance based on: (i) the full PK profiles  
58 (refCL); and (ii) C24h only (C24h-CL). Xgboost was trained to correct the error between refCL  
59 and C24h-CL in the training dataset (75%). C24h-CL as well as ML-corrected C24h-CL were  
60 evaluated in a testing dataset (25%) and then in a set of PK profiles simulated using another  
61 published POPPK model.

### 62 Results

63 A strong decrease in mean predictive error (MPE%), imprecision (RMSE%) and the number of  
64 profiles outside  $\pm 20\%$  MPE% (n-out20%) was observed with the hybrid algorithm (decreased  
65 in MPE% by 95.8% and 85.6%; RMSE% by 69.5% and 69.0%; n-out20% by 97.4% and 100% in  
66 the training and testing sets, respectively. In the external validation set, the hybrid algorithm  
67 decreased MPE% by 96%, RMSE% by 68% and n-out20% by 100%.

### 68 Conclusion

69 The hybrid model proposed significantly improved isavuconazole AUC estimation over MAP-  
70 BE based on the sole C24h and may improve dose adjustment.

71

## 72 Study Highlight

### 73 - **What question did this study address?**

74 This study evaluated the performances of a hybrid model combining population  
75 pharmacokinetics (POPPK) and machine learning to improve individual isavuconazole  
76 clearance estimation in comparison to POPPK alone.

### 77 - **What does this study add to our knowledge?**

78 A decreased by about 90% and 70% of the bias and imprecision was observed in comparison  
79 to the MAP-BE alone for prediction based on trough concentration.

### 80 - **How might this change drug discovery, development, and/or therapeutics?**

81  
82 The hybrid model developed may spearhead a new generation of tools for MIPD in routine  
83 practice.

84

## 85 1. Introduction

86 The use of model-based precision dosing (MIPD) to individualize the dose of drugs in  
87 addition to therapeutic drug monitoring (TDM) is used in many fields such as transplantation  
88 or infectious diseases, for example [1–6]. It relies on population pharmacokinetics (POPPK)  
89 models and Bayesian estimators (MAP-BE) [7,8] based on priors and a limited sampling  
90 strategy (LSS) and allows to estimate individual pharmacokinetics parameters (e.g.  
91 distribution volume or clearance...) and/or exposure indices as the area under the curve  
92 (AUC). However, the POPPK models and MAP-BE have some limitations including a  
93 systematic deviation toward the typical parameters in the case of a few individual  
94 information called eta-shrinkage. Since recently, the use of the machine learning (ML)  
95 algorithms in medicine and particularly in pharmacology has increased [9]. Different use of  
96 ML has been proposed including a direct estimation of the drug exposure [2], estimation of  
97 PK parameters [10] or of dose proposals [5] for several drugs. Despite the high performances  
98 of ML algorithms in terms of accuracy, they also exhibited limitations (“black box”, less  
99 flexible with respect to sampling times, no possibility of carrying out simulations...). Recently,  
100 studies have proposed to combine the 2 approaches (POPPK and ML) in different ways to  
101 overcome these limitations: (i) to select pertinent covariates [11] (ii) to flat prior in some  
102 patients [12] (iii) to improve antibiotics clearance estimation after POPPK estimation with  
103 covariates [10] (iv) to correct the bias in the tacrolimus trough concentration (C<sub>0</sub>) after MAP-  
104 BE estimation [13]. In a previous study, we propose an hybrid algorithm that use combine a  
105 POPPK model and a machine learning algorithm [14]. Briefly, based on simulations from a  
106 literature POPPK model of iohexol [4], a ML algorithm was trained to predict the bias  
107 between the CL obtained from simulated full PK profiles and the CL obtained after MAP-BE  
108 based on a three-points LSS. Then the CL obtained by the LSS in an external set were  
109 corrected by the ML algorithm leading to a decrease in bias and imprecision by 40.5% and  
110 23.8% respectively in comparison to the MAP-BE & LSS alone. To confirm the results  
111 previously obtained, we selected a new case study for isavuconazole. Isavuconazole (ISA) is a  
112 second-generation triazole antifungal agent, indicated for the treatment of invasive fungal  
113 infections such as aspergillosis and mucormycosis [15]. ISA is a substrate and moderate  
114 inhibitor of cytochrome 3A4/3A5 and is subject to numerous drug-drug interactions [16]. It is  
115 considered to exhibit stable pharmacokinetics [17,18] and to be equally effective as, but

116 safer than, other triazole antifungal agents (less hepatotoxicity, neurotoxicity and cardiac  
117 toxicity) [19]. However, the relationship between ISA exposure and toxicity is still the subject  
118 of debate and conflicting studies [20,21]. But in routine care, CO was used to monitor the  
119 isavuconazole exposition.

120 The aim of this work was to develop a hybrid model based on POPPK and ML algorithm to  
121 correct the CL estimation bias in the MAP-BE & LSS based on 1 sample (and the  
122 isavuconazole estimation of exposure).

## 123 2. Material & Methods

### 124 2.1. Population Pharmacokinetic

#### 125 2.1.1. Data simulation and estimation of clearance using MAP-BE

126

127 The population parameters and relevant covariates of the POPPK model developed by Wu et  
128 al. [22] using 471 samples from 79 solid organ transplant patients were used to simulate  
129 2000 rich PK profiles at steady-state for a typical dose of 200 mg once daily administered by  
130 intravenous infusion over one hour. Each profile was made up of 49 concentration-time  
131 points between 0 and 24h (one every 30 minutes) simulated using the 'mrgsolve' R package  
132 [23]. The sex ratio was 50/50 and the body mass index (BMI) was simulated based on a  
133 truncated normal distribution using the data of the original article (mean  $\pm$  SD [min to max] =  
134  $25.4 \pm 5.96$  [14.6 to 45.2] (as no information was provided in the original article about the  
135 correlation between sex and BMI, the same BMI range was used for both men and women).  
136 Briefly, the model consisted of a 2-compartment model with inter-individual variability on  
137 the apparent clearance (CL) and peripheral volume of distribution (V2), 2 covariates (sex and  
138 BMI) and a proportional residual error. ISA clearance (CL) was almost twice higher in men  
139 than in women and the V2 increased with BMI. Filters were applied to the simulated PK  
140 profiles by removing those with CL and V2 below the 1<sup>st</sup> and above the 99<sup>th</sup> percentiles, to  
141 eliminate extreme outliers.

142 The MAPBAYR package, was then used to estimate the individual PK parameters and  
143 individual concentrations using MAP-BE based on: (i) the simulated full PK profiles to derive  
144 the reference clearance values (refCL) and (ii) the simulated trough concentrations to derive  
145 C24h-CL (the degraded estimates to be corrected using the hybrid strategy). The simulated

146 full PK profiles that failed to be processed by MAPBAYR were removed (in these patient  
147 typical population parameters values are given by the program).

## 148 2.2. Machine Learning

### 149 2.2.1. Preparation of the Data and Feature Engineering

150 The raw estimation error between refCL and C24h-CL was calculated using the following  
151 formula:

$$Error = refCL - C24hCL$$

152 This error was considered as the outcome to be predicted by ML (regression problem). The  
153 variables used as predictors were: (i) the PK parameters (CL and V2) obtained using the  
154 MAP-BE on C24h, (ii) the simulated C24h ( $SC_{24h}$ ), (iii) the MAP-BE estimated C24h ( $EC_{24h}$ ), (iv)  
155 the simulated sex and BMI and (v) the  $\eta$  of MAP-BE CL and V2 (corresponding to the  
156 estimated deviation (in %) relative to the typical population CL and V2). Additionally, as in  
157 previous papers [14] some features were engineered to add information: (i) the differences  
158 between  $SC_{24h}$  and  $EC_{24h}$ ; (ii) the ratio between  $SC_{24h}$  and  $EC_{24h}$ ; (iii) the relative ratio  
159 between  $SC_{24h}$  and  $EC_{24h}$ . Finally, a total of 12 features were available for training the ML  
160 algorithm.

### 161 2.2.2. Training the machine learning algorithm

162 Simulations were split into training (75%) and testing (25%) datasets, which were compared  
163 using the Wilcoxon test ( $p < 0.05$ ). An Xgboost algorithm was trained to predict the errors of  
164 C24h-CL as compared to refCL, using the Tidymodels framework in R [25]. Briefly, a ten-fold  
165 cross-validation was applied to the training dataset to tune the hyperparameters and  
166 evaluate the model performances based on the  $r^2$ . A variable importance plot was drawn to  
167 evaluate the relative influence of the variables on the error. The predicted error was added  
168 to C24h-CL to obtain the corrected CL (corCL):

$$corCL = C24hCL + Xgboost\ predicted\ error$$

### 169 2.2.3. Xgboost performance in the training and testing datasets

170 The performances of the Xgboost algorithm developed to correct ISA CL in another 10-fold  
171 cross-validation set from the training set (to check out overfitting) and in the testing dataset  
172 were evaluated using the MPE%, the relative RMSE% and the percentage of profiles with

173 MPE% out of the  $\pm 20\%$  MPE interval between corCL and refCL. The unexplained error  
174 between the refCL and C24h-CL or corCL was plotted as a function of CL to check the  
175 dispersion over the CL range.

#### 176 2.2.4. Evaluation of the shrinkage before and after correction

177 The shrinkage of C24h-CL and hybrid model was estimated in the training and testing  
178 datasets using the formula proposed by Monolix [26]:

$$Shrinkage = 1 - \frac{var(\eta CL_x)}{\omega_{CL}^2}$$

179 Where  $x$  is the method used,  $\eta CL_x$  is the random effect of the parameter and  $\omega_{CL}^2$  is the  
180 variance of the parameter estimate.

#### 181 2.2.5. External evaluation with simulated patients using another POPPK model

182 For external validation of our hybrid algorithm, another set of simulations was obtained  
183 using another ISA POPPK model developed from pooled data of phase 1 and phase 3 studies,  
184 corresponding to 421 individuals and 6363 drug concentrations [27]. The same methodology  
185 as above (including simulations, filtering out of outliers and MAP-BE using MAPBAYR) was  
186 applied to draw 200 simulated PK profiles receiving 200 mg/24h by intravenous infusion  
187 over one hour. The simulations were performed for non-healthy & caucasian subjects, only  
188 using BMI as a covariate on V2 (simulated following the publication data (median[min to  
189 max] = 23.6[13.9 to 41.1])). The unexplained errors of each method (MAP-BE and hybrid  
190 model) were graphically compared.

## 191 3. Results

### 192 3.1. Simulated Data

193 After excluding the 1% extreme percentiles in CL and V2 (n=233) and thirty-five PK profiles  
194 that failed to be estimated by MAPBAYR, 1727 (out of the 2000) simulated profiles  
195 remained. The description of the simulated data and features tested as predictors is  
196 presented in Supplemental Table 1. No predictors were significantly different between the  
197 training and testing datasets. The correlation matrix between the features and predictors is  
198 presented in Supplemental Figure 1.

### 199 3.2. Bayesian estimation of ISA clearance based on full PK profiles or post-dose 200 trough plasma concentrations

201 The refCL and C24h-CL were (mean  $\pm$  SD [min-max])  $4.9 \pm 3.3$ [1.0-19.2] L/h and  $4.4 \pm 2.5$ [1.1-  
202 14.8], respectively. The density plot of the error between refCL and C24h-CL was skewed  
203 towards high values (log-normal distribution) (Figure 1). Actually, C24h-CL showed a strong  
204 deviation toward an underestimation for high values (CL > 5 L/h) as shown in the Figure 2A.

205 The shrinkage represents the percentage regression to the mean of a parameter  
206 distribution. The shrinkage of C24h-CL was 45.6% and 44.6% in the training and testing  
207 datasets, respectively.

### 208 3.3. Performance of ML to correct C24h-CL estimation errors in the training and 209 testing datasets

210 The Xgboost algorithm predicted the error with  $r^2 = 0.94$  in the training set and  $r^2 = 0.98$  in  
211 the testing set (Figure 3). The 3 features engineered were the most important according to  
212 the variable importance plot (Figure 4), followed by the CL-related variables (ETA CL and CL  
213 predicted). The corCL was (mean  $\pm$  SD [min-max])  $4.9 \pm 3.3$ [0.9-19.5] L/h. The  
214 hyperparameters of the best tuned Xgboost model are presented in the Supplemental Table  
215 2.

216 The MPE% represents the accuracy of the estimation, the RMSE% represents the imprecision  
217 of the estimation. Correction using Xgboost error estimates decreased drastically the MPE%  
218 (by 95.8% and 85.6%) and the RMSE% (by 69.5% and 69.0%) in the training and testing sets,  
219 respectively (Table 1). Additionally, less than 1% of corCL were outside the  $\pm 20\%$  MPE

220 interval in the training set and none in the testing set, corresponding to a decrease of 97.4%  
221 and 100% respectively in comparison to C24h-CL (Table 1).

222 The hybrid model decreased the shrinkage of ISA CL estimation by 54.0% and 58.5%  
223 respectively (training: 21.2% and testing: 18.5%).

### 224 3.4. Evaluation of the hybrid model on a dataset simulated using an independent 225 POPPK model

226 After applying filters, 5/200 full PK profiles were removed. The refCL and C24h-CL were  
227 (mean  $\pm$  SD [min-max])  $3.4 \pm 2.2$  [0.8-14.1] L/h and  $3.0 \pm 1.4$  [1.0-7.7], respectively. The  
228 performances of the hybrid model (corCL) in comparison to the MAP-BE applied to C24H  
229 (C24h-CL) are presented in Table 1. The hybrid model decreased the MPE% by approximately  
230 96% and the RMSE% by 68% as compared to MAP-BE alone. As for simulations from the first  
231 POPPK model, the unexplained error was underestimated (refCL-C24h-CL) for refCL larger for  
232 values  $> 5$  L/h (Figure 5). The hybrid model corrected this underestimation accurately for  
233 most of the cases but slightly overcorrected for some of them.

## 234 4. Discussion

235 In this study, we developed a hybrid algorithm combining MAP-BE and ML, able to estimate  
236 very accurately ISA CL based on the steady-state C24h. The performance obtained with this  
237 hybrid algorithm was better than that of MAP-BE alone.

238 All the data used in this study were obtained by simulation from two independent published  
239 POPPK models, one for the development [22] and the other for the validation [27] phases.  
240 The former was developed using data from solid organ transplant patients [22] and the  
241 latter using the data from the SECURE study [18], corresponding mainly to patients with  
242 hematologic malignancies, either active or at the stage of stem cell transplantation. This  
243 difference in patient populations may account for the slightly lower performance of our  
244 hybrid algorithm in the validation test set but it also suggests that it can be extrapolated to  
245 other types of patients on ISA.

246 We did not evaluate the performance of MAPBAYR in comparison to standard tool (such as  
247 NONMEM or Monolix) as it has been already done by Le Louedec et al in the original article

248 [24]. The authors reported, in comparison to NONMEM for the seven real models tested  
249 (carboplatin, ibrutinib, pazotinib, cabozantinib, sunitinib, methotrexate and voriconazole),  
250 satisfactory performances with overall 97.3% and 99.3% of accurate estimations for PK  
251 parameters and PK outcomes used for MIPD, respectively. Additionally, we acknowledge  
252 that the estimation of the true CL using MAPBAYR based on all the simulated samples could  
253 have been avoided by using the simulated CL directly. We did that to mimic real PK profiles  
254 in real patients in whom we would have had full PK profiles (but unknown true CL ) and  
255 evaluated a limited sampling strategy. Finally, the mean CL obtained with MAPBAYR applied  
256 to all the samples was identical to the simulated CL (data not shown).

257 MAP-BE fed only with C24h values strongly underestimated ISA CL for values >5 L/h, and the  
258 higher the C24h-CL, the stronger the correction made by the hybrid algorithm. Actually, with  
259 so little individual information (only 1 concentration), the estimated CL is largely shrunken  
260 toward the typical mean population value. Similarly, for values > 2.5 L/h, an overestimation  
261 is observed even if to a lower extent than the underpredicted values on Figures 4 and 5.  
262 Figure 5 also showed that in 10 cases with high CL values, the hybrid algorithm  
263 overcorrected the CL leading to an overestimation of refCL.

264 It is of note that the typical CL in the external validation population was 2.36 L/h (theta 1)  
265 while in the results, a mean value of 3.4 L/h was observed. The discrepancy is probably  
266 stochastic and linked to the number of simulations as increasing the number of simulations  
267 to 1000 led to a mean CL =2.8 and median = 2.38 L/h.

268 The bimodal distribution (as function of sex) observed in the unexplained error of MAB-BE  
269 C0 was removed after correction (Figure 2). The correction of the shrinkage of CL by the  
270 hybrid model is probably mediated by a decrease in the strong influence of the covariate  
271 'Sex' carried by the POPPK model (as this is the only covariate that explain the interindividual  
272 variability of CL). In contrast, the covariate 'Sex' provides the least important information for  
273 the ML algorithm to predict the unexplained error as shown in the Figure 4.

274 The correlation matrix between features and predictors shows a strong correlation between  
275 some of the predictors, preventing the use of them in a regression or mixed effect analysis.  
276 For example, the simulated and predicted concentrations exhibited  $r = 0.997$ , CL eta and CL  
277 MAP-BE a  $r = 0.865$  or BMI and V2 a  $r = 1.0$ . Despite this, all the features were used with a

278 different importance in the VIP plot. This is probably linked to the principle of the Xgboost  
279 algorithm that randomly selects part of the features to build each tree during the iterative  
280 development of the ensemble algorithm.

281 The PK/PD exposure marker initially used in the preclinical model for ISA was the AUC [28–  
282 30], but its use in routine practice is complicated since good Bayesian estimation would  
283 require more than one concentration-time point. In the present work, the algorithm  
284 developed yields very accurate estimation of ISA CL, hence of its AUC (CL = dose/AUC).  
285 Thereby, an individualized dose after intravenous administration of ISA can be proposed to  
286 reach an individual AUC target derived from the ECOFF or MIC [31]. This hybrid algorithm  
287 has not been developed for oral ISA and its development will be performed in further work,  
288 even if the relevance of dose adjustment is less pronounced due to the pill dosages (multiple  
289 of 100 mg) that do not allow precise dose individualization.

290 We previously developed a similar hybrid algorithm for the estimation of the glomerular  
291 filtration rate (GFR) after iohexol administration and found that it decreased the shrinkage  
292 and improved the individual predictions [14]. The present study confirms that this  
293 methodology is generalizable. It is of note that in the previous work, the glmnet algorithm  
294 was also evaluated and yielded slightly better results than the XGboost. However, we did not  
295 evaluate it here and chose directly the Xgboost algorithm as (i) when we evaluated both  
296 algorithms on different LSS, we observed that the Xgboost obtained a more constant  
297 improvement in comparison to the glmnet and (ii) as discussed above, the Xgboost algorithm  
298 is less sensitive to the high collinearity between the predictors as compared to linear  
299 regression. The principle of Xgboost, an algorithm based on boosting, is to iteratively build a  
300 simple regression tree by finding split values among part of the predictors randomly selected  
301 that minimize the loss function (here the root mean square error). The iterative process  
302 constructs additional regression trees of the same structure including other randomly  
303 selected predictors, but minimizes the residual errors of the previous aggregated regression  
304 trees.

305 In this work, we did not try to develop a ML algorithm to directly predict the CL of  
306 isavuconazole based on features and available concentrations as previously done by our  
307 group for other drugs [2,32]. Even if a comparison of the hybrid algorithm with the direct use

308 of a ML algorithm is of interest, the problem with the ML algorithm alone is a loss in  
309 interpretability (black box). The aim of the hybrid algorithm is both to improve the  
310 performance of existing population pharmacokinetics models while preserving the  
311 pharmacological interpretability.

312 The filtering of extreme simulations prevented Xgboost from learning from unrealistic  
313 profiles, even if this algorithm in particular makes nonlinear prediction and discretizes the  
314 continuous values into quantiles, which is less sensible to outliers data [33]. Supplemental  
315 Figure 2 shows a typical case illustrating that the patients filtered out were characterized by  
316 a typical value of CL assigned by MAPBAYR.

317 Interestingly, the most important variables for predicting the error between refCL and C24h-  
318 CL were those engineered and those related to CL. Indeed, these variables inform about the  
319 importance of deviation in respect to typical CL in the population.

320 Since this combined algorithm may be complicated to reproduce, we have developed a shiny  
321 interface to allow testing it, for research purposes only ([https://hybrid-  
322 models.shinyapps.io/isavuconazole/](https://hybrid-models.shinyapps.io/isavuconazole/)).

323 In this work we only evaluated the improvement by using the hybrid algorithm for a single  
324 POPPK model. It could be of interest to compare the improvement to recent and accurate  
325 model averaging approaches [34]. Indeed in a previous collaborative work, we showed that  
326 these approaches had similar (or slightly better) performances in comparison to the ML  
327 algorithm that directly estimated the CL of vancomycin [32]. We could also try to develop  
328 hybrid algorithms for model averaging approaches but that would require some additional  
329 works. Additionally, previous works by other groups have combined POPPK model and ML  
330 algorithms. For example, Tang et al combined ML and POPPK to improve antibiotic CL  
331 estimation based on covariates while we used both covariates but in addition, the estimated  
332 PK parameters, observations and the POPPK model predictions[10]. Another example is the  
333 works of Hughes et al who developed a ML algorithm which learned from the error between  
334 observations and predictions and was able to select patients in whom it would be better to  
335 use flat priors rather than conventional MAP-BE for vancomycin [12].

336 This study has some limitations. First, the algorithm development and validation are only  
337 based on simulations, meaning that it still requires validation in “real” patients before it can  
338 be implemented in the clinics. Simulations were performed with a single dosage (200 mg)  
339 which corresponds to the recommended dose. The impact of ISA dose on pharmacokinetic  
340 parameters or as a predictor for the hybrid model was therefore not explored and PK  
341 linearity was assumed (i.e. CL is independent of the dose and a dose increase will result in a  
342 proportional increase in AUC). The inclusion of the dose would require providing the  
343 algorithm with initial simulations for different dose levels and would allow to confirm the  
344 hypothesis of dose proportionality. However, the goal of such an algorithm is to propose  
345 different doses in the population so that each patient can reach a predefined target and that  
346 should be investigated in future works. The absence of clear AUC targets is another  
347 limitation and when the strain sensitivity is unknown, the mean AUC target of the SECURE  
348 study (100 h.mg/L) can be proposed [27]. Finally, the relevance of TDM has not been  
349 evaluated in this work which only proposes tools to perform it.

## 350 5. Conclusions

351 In conclusion, the hybrid model associating a POPPK model, MAP-BE with a limited sampling  
352 strategy and ML developed here may spearhead a new generation of tools for model-  
353 informed precision dosing in routine practice.

354

## 355 References

- 356 1. Marquet P, Destère A, Monchaud C, Rérolle J-P, Buchler M, Mazouz H, et al. Clinical  
357 Pharmacokinetics and Bayesian Estimators for the Individual Dose Adjustment of a Generic  
358 Formulation of Tacrolimus in Adult Kidney Transplant Recipients. *Clin Pharmacokinet.*  
359 2021;60:611–22.
- 360 2. Labriffe M, Woillard J, Debord J, Marquet P. Machine learning algorithms to estimate  
361 everolimus exposure trained on simulated and patient pharmacokinetic profiles. *CPT*  
362 *Pharmacom & Syst Pharma.* 2022;psp4.12810.
- 363 3. Woillard J-B, Saint-Marcoux F, Debord J, Åsberg A. Pharmacokinetic models to assist the  
364 prescriber in choosing the best tacrolimus dose. *Pharmacol Res.* 2018;130:316–21.
- 365 4. Destere A, Gandonnière CS, Åsberg A, Loustaud-Ratti V, Carrier P, Ehrmann S, et al. A  
366 single Bayesian estimator for iohexol clearance estimation in ICU, liver failure and renal  
367 transplant patients. *Brit J Clinical Pharma.* 2021;bcj.15197.
- 368 5. Ponthier L, Ensuque P, Destere A, Marquet P, Labriffe M, Jacqz-Aigrain E, et al.  
369 Optimization of Vancomycin Initial Dose in Term and Preterm Neonates by Machine  
370 Learning. *Pharm Res [Internet].* 2022 [cited 2022 Aug 11]; Available from:  
371 <https://link.springer.com/10.1007/s11095-022-03351-6>
- 372 6. Franck B, Autmizguine J, Åsberg A, Théorêt Y, Marquet P, Ovetchkine P, et al. Thoroughly  
373 Validated Bayesian Estimator and Limited Sampling Strategy for Dose Individualization of  
374 Ganciclovir and Valganciclovir in Pediatric Transplant Recipients. *Clin Pharmacokinet.*  
375 2021;60:1449–62.
- 376 7. Woillard J-B, Debord J, Benz-de-Bretagne I, Saint-Marcoux F, Turlure P, Girault S, et al. A  
377 Time-Dependent Model Describes Methotrexate Elimination and Supports Dynamic  
378 Modification of MRP2/ABCC2 Activity. *Ther Drug Monit.* 2017;39:12.
- 379 8. Benkali K, Rostaing L, Premaud A, Woillard J-B, Saint-Marcoux F, Urien S, et al. Population  
380 Pharmacokinetics and Bayesian Estimation of Tacrolimus Exposure in Renal Transplant  
381 Recipients on a New Once-Daily Formulation: Clinical Pharmacokinetics. 2010;49:683–92.
- 382 9. Goecks J, Jalili V, Heiser LM, Gray JW. How Machine Learning Will Transform Biomedicine.  
383 *Cell.* 2020;181:92–101.
- 384 10. Tang B-H, Guan Z, Allegaert K, Wu Y-E, Manolis E, Leroux S, et al. Drug Clearance in  
385 Neonates: A Combination of Population Pharmacokinetic Modelling and Machine Learning  
386 Approaches to Improve Individual Prediction. *Clin Pharmacokinet.* 2021;60:1435–48.
- 387 11. Koch G, Pfister M, Daunhawer I, Wilboux M, Wellmann S, Vogt JE. Pharmacometrics and  
388 Machine Learning Partner to Advance Clinical Data Analysis. *Clin Pharmacol Ther.*  
389 2020;107:926–33.
- 390 12. Hughes JH, Keizer RJ. A hybrid machine learning/pharmacokinetic approach outperforms  
391 maximum a posteriori Bayesian estimation by selectively flattening model priors. *CPT*  
392 *Pharmacometrics Syst Pharmacol.* 2021;10:1150–60.
- 393 13. Li Z, Li R, Niu W, Zheng X, Wang Z, Zhong M, et al. Population Pharmacokinetic Modelling  
394 Combined with Machine Learning Approach Improved Tacrolimus Trough Concentrations  
395 Prediction in Chinese Adult Liver Transplant Recipients. *The Journal of Clinical Pharma.*  
396 2022;jcph.2156.
- 397 14. Destere A, Marquet P, Gandonnière CS, Åsberg A, Loustaud-Ratti V, Carrier P, et al. A  
398 Hybrid Model Associating Population Pharmacokinetics with Machine Learning: A Case Study  
399 with Iohexol Clearance Estimation. *Clin Pharmacokinet [Internet].* 2022 [cited 2022 Jun 9];  
400 Available from: <https://link.springer.com/10.1007/s40262-022-01138-x>

- 401 15. Rybak JM, Marx KR, Nishimoto AT, Rogers PD. Isavuconazole: Pharmacology,  
402 Pharmacodynamics, and Current Clinical Experience with a New Triazole Antifungal Agent.  
403 Pharmacotherapy. 2015;35:1037–51.
- 404 16. Townsend R, Dietz A, Hale C, Akhtar S, Kowalski D, Lademacher C, et al. Pharmacokinetic  
405 Evaluation of CYP3A4-Mediated Drug-Drug Interactions of Isavuconazole With Rifampin,  
406 Ketoconazole, Midazolam, and Ethinyl Estradiol/Norethindrone in Healthy Adults. Clin  
407 Pharmacol Drug Dev. 2017;6:44–53.
- 408 17. Sivasubramanian G, Chandrasekar PH. Efficacy and safety of Isavuconazole for the  
409 treatment of invasive *Aspergillus* infection - an update of the literature. Expert Opinion on  
410 Pharmacotherapy. 2022;23:543–9.
- 411 18. Maertens JA, Raad II, Marr KA, Patterson TF, Kontoyiannis DP, Cornely OA, et al.  
412 Isavuconazole versus voriconazole for primary treatment of invasive mould disease caused  
413 by *Aspergillus* and other filamentous fungi (SECURE): a phase 3, randomised-controlled, non-  
414 inferiority trial. The Lancet. 2016;387:760–9.
- 415 19. Kaindl T, Andes D, Engelhardt M, Saulay M, Larger P, Groll AH. Variability and exposure–  
416 response relationships of isavuconazole plasma concentrations in the Phase 3 SECURE trial  
417 of patients with invasive mould diseases. Journal of Antimicrobial Chemotherapy.  
418 2019;74:761–7.
- 419 20. Furfaro E, Signori A, Di Grazia C, Dominietto A, Raiola AM, Aquino S, et al. Serial  
420 monitoring of isavuconazole blood levels during prolonged antifungal therapy. Journal of  
421 Antimicrobial Chemotherapy. 2019;74:2341–6.
- 422 21. Kosmidis C, Otu A, Moore CB, Richardson MD, Rautemaa-Richardson R. Isavuconazole  
423 Therapeutic Drug Monitoring during Long-Term Treatment for Chronic Pulmonary  
424 Aspergillosis. Antimicrob Agents Chemother. 2020;65:e01511-20.
- 425 22. Wu X, Venkataramanan R, Rivosecchi RM, Tang C, Marini RV, Shields RK, et al. Population  
426 Pharmacokinetics of Intravenous Isavuconazole in Solid-Organ Transplant Recipients.  
427 Antimicrob Agents Chemother. 2020;64:e01728-19.
- 428 23. Baron KT. mrgsolve: Simulate from ODE-Based Models [Internet]. 2022. Available from:  
429 <https://CRAN.R-project.org/package=mrgsolve>
- 430 24. Le Louedec F, Puisse F, Thomas F, Chatelut É, White-Koning M. Easy and reliable  
431 maximum *a posteriori* Bayesian estimation of pharmacokinetic parameters with the  
432 open-source R package mapbayr. CPT Pharmacometrics Syst Pharmacol. 2021;10:1208–20.
- 433 25. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine  
434 learning using tidyverse principles. [Internet]. 2020. Available from:  
435 <https://www.tidymodels.org>
- 436 26. Monolix version 2018R1. Antony, France: Lixoft SAS, 2018.  
437 <http://lixoft.com/products/monolix/>.
- 438 27. Desai A, Kovanda L, Kowalski D, Lu Q, Townsend R, Bonate PL. Population  
439 Pharmacokinetics of Isavuconazole from Phase 1 and Phase 3 (SECURE) Trials in Adults and  
440 Target Attainment in Patients with Invasive Infections Due to *Aspergillus* and Other  
441 Filamentous Fungi. Antimicrob Agents Chemother. 2016;60:5483–91.
- 442 28. Warn PA, Sharp A, Parmar A, Majithiya J, Denning DW, Hope WW. Pharmacokinetics and  
443 Pharmacodynamics of a Novel Triazole, Isavuconazole: Mathematical Modeling, Importance  
444 of Tissue Concentrations, and Impact of Immune Status on Antifungal Effect. Antimicrob  
445 Agents Chemother. 2009;53:3453–61.
- 446 29. Seyedmousavi S, Brüggemann RJM, Meis JF, Melchers WJG, Verweij PE, Mouton JW.  
447 Pharmacodynamics of Isavuconazole in an *Aspergillus fumigatus* Mouse Infection Model.

448 Antimicrob Agents Chemother. 2015;59:2855–66.  
449 30. Lepak AJ, Marchillo K, VanHecker J, Diekema D, Andes DR. Isavuconazole  
450 Pharmacodynamic Target Determination for Candida Species in an *In Vivo* Murine  
451 Disseminated Candidiasis Model. Antimicrob Agents Chemother. 2013;57:5642–8.  
452 31. Arendrup MC, Meletiadis J, Mouton JW, Guinea J, Cuenca-Estrella M, Lagrou K, et al.  
453 EUCAST technical note on isavuconazole breakpoints for Aspergillus, itraconazole  
454 breakpoints for Candida and updates for the antifungal susceptibility testing method  
455 documents. Clinical Microbiology and Infection. 2016;22:571.e1-571.e4.  
456 32. Bououda M, Uster DW, Sidorov E, Labriffe M, Marquet P, Wicha SG, et al. A Machine  
457 Learning Approach to Predict Interdose Vancomycin Exposure. Pharm Res. 2022;39:721–31.  
458 33. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd  
459 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet].  
460 San Francisco California USA: ACM; 2016 [cited 2021 Dec 24]. p. 785–94. Available from:  
461 <https://dl.acm.org/doi/10.1145/2939672.2939785>  
462 34. Uster DW, Wicha SG. Optimized sampling to estimate vancomycin drug exposure:  
463 Comparison of pharmacometric and equation-based approaches in a simulation-estimation  
464 study. CPT Pharmacom & Syst Pharma. 2022;11:711–20.  
465

466  
467

Table 1 Performances of different method (MAP-BE and hybrid model) in the training and testing datasets and in the external set obtained by simulation from Desai A. et al. POPPK model

	Training set		Testing set		External validation [27]	
	C24h-CL	Hybrid model	C24h-CL	Hybrid model	C24h-CL	Hybrid model
MPE (%)	3.4	- 0.15	3.6	- 0.5	7.8	-0.3
RMSE (%)	11.2	3.4	11.4	3.5	16.3	5.2
Out of $\pm$ 20% error interval (%)	7.6	0.2	5.6	0	22.4	0

468 *MPE is mean predictive error, RMSE is root mean square error and MAP-BE is maximum a posteriori Bayesian estimator*

469

## 470 [Figure Legends](#)

471 Figure 1: Density plot of C24h-CL error in the overall simulations.

472 Figure 2: Unexplained error (true CL – MAP-BE LSS C24H CL or true CL – hybrid CL) before  
473 and after correction of clearances in the training (A) and testing (B) dataset. Males were  
474 represented by blue points and females by red points.

475 Figure 3: Correlation between C24h-CL error and Xgboost estimation error in the training (A)  
476 and testing (B) dataset.

477 Figure 4: Variable Importance Plot for the prediction of the error (reference CL based on  
478 MAP-BE applied on one simulated sample every 30min between 0 and 24h – CL obtained  
479 using MAP-BE applied on C24H only) by the Xgboost algorithm.

480 Figure 5: Unexplained error in CL prediction (ref CL –C24h-CL (grey round) or ref CL – hybrid  
481 CL (black square)) as function of refCL in the external validation set.

## 482 [Supplementary Materials](#)

483 Supplemental Table 1: Comparison of predictors in the training and testing datasets.

484 Supplemental Table 2: hyperparameters selected for the best Xgboost algorithm

485 Supplemental figure 1: Correlation plot between the different evaluated features used to  
486 predict the difference between the reference clearance and the C24h-CL.

487 Supplemental figure 2: Observed and individual predicted concentrations in a removed  
488 simulation. black points: "simulated concentrations"; black line: "individual predicted  
489 concentrations".

490