











## ARTICLE

# Multi-model averaging improves the performance of model-guided infliximab dosing in patients with inflammatory bowel diseases

Wanee Kantasiripitak<sup>1</sup>  | An Outtier<sup>2,3</sup>  | Sebastian G. Wicha<sup>4</sup>  |  
 Alexander Kensert<sup>1,5</sup>  | Zhigang Wang<sup>1</sup>  | João Sabino<sup>2,3</sup>  | Séverine Vermeire<sup>2,3</sup>  |  
 Debby Thomas<sup>1</sup>  | Marc Ferrante<sup>2,3</sup>  | Erwin Dreesen<sup>1</sup> 

<sup>1</sup>Department of Pharmaceutical and Pharmacological Sciences, University of Leuven, Leuven, Belgium

<sup>2</sup>Department of Gastroenterology and Hepatology, University Hospitals Leuven, Leuven, Belgium

<sup>3</sup>Department of Chronic Diseases and Metabolism, University of Leuven, Leuven, Belgium

<sup>4</sup>Department of Clinical Pharmacy, Institute of Pharmacy, University of Hamburg, Hamburg, Germany

<sup>5</sup>Department of Chemical Engineering, Vrije Universiteit Brussels, Brussels, Belgium

## Correspondence

Erwin Dreesen, Department of Pharmaceutical and Pharmacological Sciences, University of Leuven, O&N II Herestraat 49 - Box 424, 3000 Leuven, Belgium.  
 Email: [erwin.dreesen@kuleuven.be](mailto:erwin.dreesen@kuleuven.be)

## Funding information

A.K. is funded by a joint initiative of the Research Foundation – Flanders (FWO), Belgium, and the Walloon Fund for Scientific Research (FNRS), Belgium (EOS research project “Chimic” [ID: 30897864]). Z.W. is a doctoral research fellow of the Research Foundation – Flanders (FWO), Belgium (grant number 1SF2922N). J.S. is a senior clinical investigator of the Research Foundation – Flanders

## Abstract

Infliximab dosage de-escalation without prior knowledge of drug concentrations may put patients at risk for underexposure and trigger the loss of response. A single-model approach for model-informed precision dosing during infliximab maintenance therapy has proven its clinical benefit in patients with inflammatory bowel diseases. We evaluated the predictive performances of two multi-model approaches, a model selection algorithm and a model averaging algorithm, using 18 published population pharmacokinetic models of infliximab for guiding dosage de-escalation. Data of 54 patients with Crohn’s disease and ulcerative colitis who underwent infliximab dosage de-escalation after an earlier escalation were used. A priori prediction (based solely on covariate data) and maximum a posteriori prediction (based on covariate data and trough concentrations) were compared using accuracy and precision metrics and the classification accuracy at the trough concentration target of 5.0 mg/L. A priori prediction was inaccurate and imprecise, with the lowest classification accuracies irrespective of the approach (median 59%, interquartile range 59%–63%). Using the maximum a posteriori prediction, the model averaging algorithm had systematically better predictive performance than the model selection algorithm or the single-model approach with any model, regardless of the number of concentration data. Only a single trough concentration (preferably at the point of care) sufficed for accurate and precise prediction. Predictive performance of both single- and multi-model approaches was robust to the lack of covariate data. Model averaging using four models demonstrated similar predictive performance with a five-fold shorter computation time. This model averaging algorithm was implemented in the TDMx software tool to guide infliximab dosage de-escalation in the forthcoming prospective MODIFI study (NCT04982172).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

(FWO), Belgium. M.F. is a senior clinical investigator of the Research Foundation – Flanders (FWO), Belgium. E.D. is a postdoctoral research fellow of the Research Foundation – Flanders (FWO), Belgium (grant number 12X9420N). Bioanalysis for this project was funded by TBM (Applied Biomedical Research with a Primary Social finality) project, the Research Foundation – Flanders (FWO), Belgium (grant number T003716N).

## Study Highlights

### WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

The benefits of model-informed precision dosing (MIPD) using a single infliximab population pharmacokinetic model showed clinical benefits over label dosing in patients with inflammatory bowel diseases. Model selection and external evaluation remain challenging.

### WHAT QUESTION DID THIS STUDY ADDRESS?

The performance of multi-model approaches for guiding infliximab dosage de-escalation was evaluated using a comprehensive set of diagnostic tools.

### WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

A model averaging algorithm performed better than a model selection algorithm and a single-model approach with any model. A posteriori prediction with a single, preferably most recently obtained, infliximab trough concentration has clinically acceptable accuracy and precision, even when covariate information is missing.

### HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

Utilizing model averaging during MIPD of infliximab may increase target attainment and thereby the probability of maintaining a therapeutic response while controlling the risk of relapse. The developed algorithm was implemented in the freely available TDMx software tool and will be used for prospective evaluation.

## INTRODUCTION

For over 2 decades, infliximab, an anti-tumor necrosis factor- $\alpha$  monoclonal antibody, has been approved for the treatment of several chronic immune-mediated diseases, including the inflammatory bowel diseases (IBDs) ulcerative colitis (UC) and Crohn's disease (CD).<sup>1,2</sup> The package label lists 5 mg/kg intravenous infusions at weeks 0, 2, and 6 (induction therapy) and every 8 weeks thereafter (maintenance therapy). However, ~20%–40% of the adult patients do not respond to standard induction therapy and up to half of the patients with a good initial response will lose response over time.<sup>3,4</sup>

Underexposure to infliximab is a common cause of loss of response in patients with IBD.<sup>5</sup> To boost infliximab trough concentrations (TCs) and subsequently regain the response, empirical dosage regimen escalations (i.e., shortening the dosing interval and/or increasing the dose) are widely practiced.<sup>6</sup> However, long-term maintenance of the escalated dosage regimen has financial, practical, and potential safety implications and is therefore not warranted.<sup>7–9</sup> Accordingly, many centers have attempted to de-escalate the infliximab dosing (i.e., extending the dosing interval and/or decreasing the dose) in patients who maintained response on an escalated infliximab dosage regimen.

Empirical de-escalation of infliximab dosing could put patients at risk for underexposure and trigger again the loss of response due to extensive interindividual pharmacokinetic (PK) variability.<sup>10</sup> Model-informed precision dosing (MIPD) has been proposed to ensure adequate exposure and maintained response compared to traditional therapeutic drug monitoring (TDM).<sup>11</sup> MIPD uses drug-specific population PK (PopPK) models, patient-specific monitoring data, and a Bayesian forecasting software tool to predict optimal doses for individual patients.<sup>12</sup> Selecting the most suitable PopPK model is challenging, especially when many models are available, as is the case for infliximab.<sup>13,14</sup> A single-model approach could potentially result in inappropriate dose recommendations, leading to suboptimal treatment outcomes or jeopardizing patients' safety. A multi-model selection algorithm (MSA) and a multi-model averaging algorithm (MAA) have previously been proposed by Uster et al.<sup>13</sup> to guarantee fit-for-purpose predictive performances during vancomycin MIPD.<sup>13</sup> The multi-model algorithms automate the MIPD procedure for selecting the prediction of either the best model (MSA) or the combination of models (MAA).

The aim of this study was to compare the predictive performance of published PopPK models and multi-model selection and averaging approaches for guiding infliximab dosage regimen de-escalation to ensure the attainment of a prespecified TC target.

## METHODS

### Clinical data

Adult patients with IBD who underwent infliximab dosage regimen de-escalation at the University Hospitals Leuven (Leuven, Belgium) between February 2017 and June 2020 were included. Dosage regimen de-escalation was defined as extending the dosing interval (with or without changing the dose) and/or decreasing the dose (with or without changing the dosing interval). The study was approved by the Ethics Committee (EC) Research UZ/KU Leuven (S63206). Serum samples were available from the CCare Biobank. All included patients have given written informed consent (B322201213950/S53684).

Patients with four consecutive trough samples, three before dosage regimen de-escalation (times  $T_{-2}$ ,  $T_{-1}$ , and  $T_0$ ) and one after de-escalation ( $T_{+1}$ ) were eligible for inclusion (Figure 1a). TCs were measured using the apDia Infliximab ELISA (apDia), with a lower limit of quantification of 0.3 mg/L.<sup>15</sup> Patients with unclassified IBD, with an ileal anal pouch anastomosis, with an ostomy, and who received infliximab prophylactically were excluded.

Sex, age, IBD type (UC or CD), disease duration, previous IBD surgery, previous biological use, and duration of infliximab treatment were recorded right before dosage regimen de-escalation (at  $T_0$ ) and were handled as time-invariant throughout the study follow-up.

Body weight, fat-free mass, serum albumin, C-reactive protein, fecal calprotectin, infliximab dose, concomitant medications use (i.e., systemic corticosteroids or the immunomodulator azathioprine), Partial Mayo score for patients with UC, and Crohn's Disease Activity Index (CDAI) and Harvey-Bradshaw Index (HBI) for patients with CD were handled as time-varying throughout the study follow-up. Single imputation with the last observation carried forward was used for handling missing covariate data.

### Candidate models

A systematic literature search of PubMed from January 1996 until June 2021 was performed to identify published PopPK models of infliximab in adult patients with IBD. The query was (infliximab) AND (model) AND (population) AND (pharmacokinetics). Articles were screened in full text for eligibility.

### Single-model evaluations

The fit of the data to the candidate models was visually inspected with goodness-of-fit plots (measured

vs. individual predicted concentrations). In addition, simulation-based evaluations were performed, including prediction-corrected visual predictive checks (VPCs) and normalized prediction distribution errors (NPDEs). A normal distribution of NPDEs  $\mathcal{N}(0, 1)$  was tested using a Wilcoxon signed-rank test (to compare the median of the NPDE to zero), a Fisher variance test (to compare the variance of the NPDE to one), and a Shapiro–Wilk test (to compare the distribution of the NPDE to a normal distribution). An adjusted  $p$  value of all the three tests (a global test) was calculated to identify the best predictive model.<sup>16</sup>

### Multi-model approaches

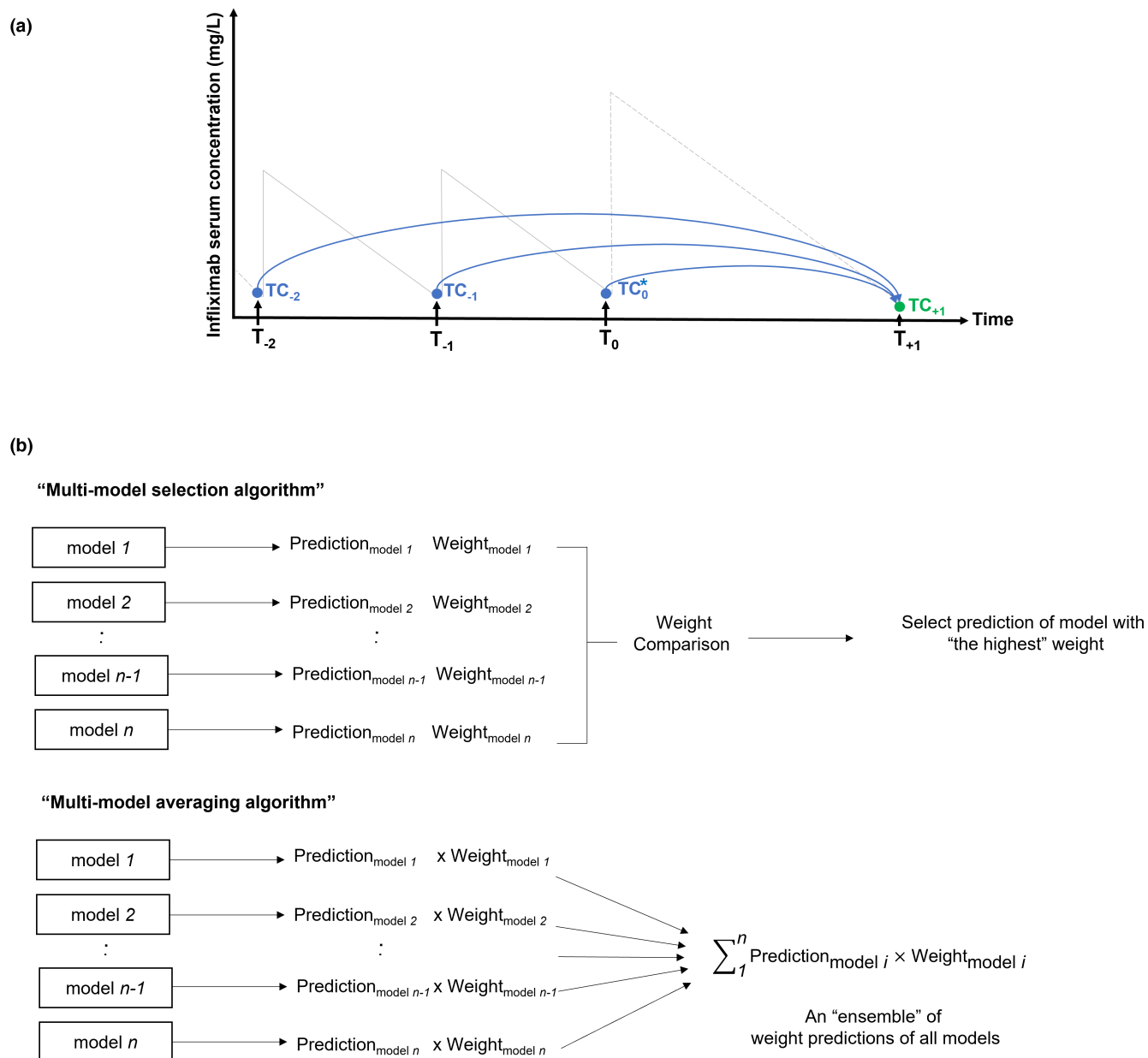
Two multi-model approaches were evaluated using all candidate models jointly; an MSA and an MAA.<sup>13</sup> The multi-model approaches used all of the candidate models simultaneously for predicting the infliximab concentration at  $T_{+1}$ . The prediction of the MSA was the prediction of the candidate model with the highest weight, whereas the prediction of the MAA was an ensemble of weighted predictions of all candidate models (Figure 1b). For each individual, predictions of the multi-model approaches were based on the weight ( $W$ ) calculated from the maximum likelihood estimate (MLE) of each candidate model  $i$  in relation to the sum of MLEs of all  $n$  candidate models, Equation 1.

$$W_{MLE_i} = \frac{MLE_i}{\sum_1^n MLE_n} = \frac{e^{(-0.5 \times OFV_i)}}{\sum_1^n e^{(-0.5 \times OFV_n)}} \quad (1)$$

### Predictive performance evaluations

The predictive performance was evaluated from the differences between the predicted and the measured TC at  $T_{+1}$  ( $TC_{+1}$ ) in two prediction modalities: a priori prediction (using only the patients' covariates) and maximum a posteriori prediction (MAP; including at least one previous TC in addition to covariates). Three a posteriori prediction settings were compared: prediction with one ( $TC_0$ ,  $TC_{-1}$ , or  $TC_{-2}$ ), two ( $TC_0$  and  $TC_{-1}$ ,  $TC_0$  and  $TC_{-2}$ , or  $TC_{-1}$  and  $TC_{-2}$ ), and three ( $TC_0$ ,  $TC_{-1}$ , and  $TC_{-2}$ ) previous TCs. The retrospective predictive performance of the models/algorithms was also evaluated by including the measured  $TC_{+1}$  in the a posteriori prediction in addition to the three previous TCs.

The model-predicted versus measured  $TC_{+1}$  in the different single-/multi-model approaches, prediction modalities, and evaluation settings were compared by calculating the relative bias (rBias, Equation 2) and the relative root mean square error (rRMSE, Equation 3) to determine accuracy and precision, respectively.



**FIGURE 1** Study diagram and workflows of multi-model approaches. (a) Study diagram of the prediction of the infliximab trough concentration (TC) at relative time + 1 ( $TC_{+1}$ ). In addition to covariate data, Bayesian forecasting was performed using one to three consecutive previously measured infliximab trough concentrations:  $TC_{-2}$ ,  $TC_{-1}$ , and  $TC_0$ . T, time; TCs, trough concentrations. \*Rapid assay needed to obtain  $TC_0$  for prospective implementation in model-informed precision dosing. (b) Flowcharts illustrate workflows of multi-model selection and multi-model averaging algorithms.

$$rBias = \frac{1}{n} \times \sum_1^i \left( \frac{TC_{+1, \text{predicted}, i} - TC_{+1, \text{measured}, i}}{TC_{+1, \text{measured}, i}} \right) \times 100\% \quad (2)$$

$$rRMSE = \sqrt{\frac{1}{n} \times \sum_1^i \left( \frac{(TC_{+1, \text{predicted}, i} - TC_{+1, \text{measured}, i})^2}{TC_{+1, \text{measured}, i}^2} \right)} \times 100\% \quad (3)$$

with  $n$  representing the total number of patients, and  $i$  each individual patient. An rBias between  $\pm 20\%$  with a 95%

confidence interval (CI) including zero was deemed clinically acceptable.<sup>13</sup> No rRMSE threshold for clinical acceptability was prespecified. Lower rRMSE values indicated more precise predictions.

## Robustness analysis and software implementation

A robustness analysis was performed to reduce the number of PopPK models without losing the predictive

performance of the multi-model approach algorithms.<sup>13</sup> The average computation time was compared between the multi-model approaches using all versus only the subset of models.

The subset of models was implemented in the TDMx software tool.<sup>17</sup> The performance of TDMx was cross-validated against nonlinear mixed-effect modeling (NONMEM).

## Bland–Altman analysis, classification accuracy, and sensitivity analysis

The MSA and MAA with the subset of models were evaluated using predictive performance metrics, Bland–Altman analysis,<sup>18</sup> classification accuracy, and a sensitivity analysis. The Bland–Altman plot was used to assess the agreement between the predicted and measured  $TC_{+1}$  across the range of measured  $TC_{+1}$ . The predicted and measured  $TC_{+1}$  were classified at the prespecified target TC of 5.0 mg/L.<sup>19</sup> The classification accuracy was calculated as

$$\text{Classification accuracy} = \frac{\text{TN} + \text{TP}}{n} \times 100\% \quad (4)$$

with TN and TP representing the numbers of true negative (predicted and measured  $TC_{+1} \geq 5.0$  mg/L) and true positive (predicted and measured  $TC_{+1} < 5.0$  mg/L) predictions, respectively, and  $n$  representing the total number of predictions. To note, outside the TDM context, a positive test result indicates the least desirable scenario which demands a clinical/pharmaceutical intervention. In the same way, we defined a positive TDM test result as a subtherapeutic concentration measurement ( $TC_{+1} < 5.0$  mg/L), warranting a dose optimization. A  $TC_{+1} \geq 5.0$  mg/L was thus defined as a negative test result, not needing any dose optimization. Consequently, a true or a false result was designated based on the correctness of the prediction with respect to the cut-off. The sensitivity of the predicted  $TC_{+1}$  to missing covariate data was evaluated using single imputation with the median value around which the covariate is centered in the model. McNemar's tests were performed to evaluate differences in classification performance between the MAA and the subset of models, or the MSA.

## Software

All models were coded in NONMEM (version 7.5; Icon plc) and provided in [Supplementary Material](#). Predictions were performed using NONMEM with a GNU Fortran 95 compiler. Data were analyzed in R (version 4.0.3; R Foundation for Statistical Computing, R Core Team) with

the RStudio integrated development environment (version 1.2.5001; RStudio, Inc.).

## RESULTS

### Clinical data

Data were available from 54 patients with IBD (38 [70%] patients with CD and 16 [30%] patients with UC; [Table S1](#)). The majority of these patients (61%, 33/54) received 5 mg/kg infliximab every 6 weeks before changing the dosage regimen to 7.5 mg/kg infliximab every 8 weeks. The median interquartile range (IQR) of the measured  $TC_0$  and  $TC_{+1}$  were 7.0, IQR 5.3–9.4 mg/L and 5.0, IQR 3.8–6.7 mg/L, respectively. Only 52% (28/54) of the patients had  $TC_{+1}$  above or equal to 5.0 mg/L.

### Candidate models

Eighteen PopPK models were identified. They differed in structure, covariates, and parameter estimates, as well as population, dosing schedules, and sampling schemes that they were based on [Table 1](#), and [Tables S2](#) and [S3](#). Half of the models (50%, 9/18) were developed on data from mixed UC and CD populations. The majority of the models (67%, 12/18) were two-compartment models. Antibodies to infliximab, serum albumin, and body weight were the most frequently identified covariates on clearance. Body weight was the most commonly identified covariate on volumes of distribution.

### Single-model evaluations

Each patient contributed the same number of consecutive TC samples (i.e., four). The individual predicted concentrations of each model were in good agreement with their measured concentrations, except for the Edlund model predictions that showed a spread deviating from the identity line ([Figure S1](#)). The VPCs of the models differed markedly ([Figure S2](#)). The Buurman model displayed the best alignment of predicted and measured concentrations. In line with the VPC results, the Buurman model was identified as the best predictive model regarding the distribution of the NPDEs ([Figure S3](#), [Table S4](#)).

### Multi-model approaches

Using the MSA in the a posteriori prediction modality, the Buurman model was selected for 36% (32%–40%) of the

**TABLE 1** Overview of the 18 candidate infliximab PopPK models

Model	N	IBD type	Treatment phase	Sampling times	Number of compartments
Aubourg <sup>39</sup>	133	CD	Induction, maintenance	Peak, trough	2
Brandse_2016 <sup>40</sup>	20	UC	Induction	Peak, intermediate, trough	2
Brandse_2017 <sup>41</sup>	332	UC, CD	Induction, maintenance	Peak, intermediate, trough	2
Buurman <sup>42</sup>	42	UC, CD	Induction, maintenance	Trough	2
Dotan <sup>43</sup>	54	UC, CD, UI	Induction, maintenance	Peak, intermediate, trough	2
Dreesen_2019 <sup>44</sup>	204	UC	Induction	Trough	1
Dreesen_2021 <sup>45</sup>	116	CD	Induction, maintenance	Intermediate, trough	2
Edlund <sup>46</sup>	68	CD	Maintenance	Intermediate, trough	2
Fasanmade_2009 <sup>33</sup>	482	UC	Induction, maintenance	Peak, intermediate, trough	2
Fasanmade_2011 <sup>34</sup>	692	CD	Induction, maintenance	Peak, intermediate, trough	2
Grisic <sup>47</sup>	121	UC, CD, UI	Maintenance	Intermediate, trough	2
Matsuoka <sup>48</sup>	121	CD	Maintenance	Trough	1
Passot <sup>49</sup>	79	UC, CD	Induction, maintenance	Trough	1
Petitcollin <sup>36</sup>	91	UC, CD	Maintenance	Trough	1
Ternant_2008 <sup>50</sup>	33	UC, CD	Induction, maintenance	Peak, intermediate, trough	2
Ternant_2015 <sup>51</sup>	111	CD	Maintenance	Intermediate, trough	1
Ternant_2018 <sup>52</sup>	50	UC, CD	Induction, maintenance	Trough	1
Xu <sup>53</sup>	788	UC, CD	NS	NS	2

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel diseases; N, number of patients; NS, not specified; PopPK, population pharmacokinetic; UC, ulcerative colitis; UI, undetermined inflammatory bowel disease type.

patients, followed by the Ternant\_2008 model (25% [16%–34%] of patients) and the Dotan model (22% [22%–24%] of patients; Figure 2). Using the MAA in the a posteriori prediction modality with one previous TC, all models had nearly equal weights (Figure S4). By adding more previous TCs, some models started dominating the a posteriori predictions of the MAA.

## Predictive performances evaluations

A priori prediction of the  $TC_{+1}$  was clinically unacceptable with both single- and the multi-model approaches (rBias –75% to +483%, rRMSE 58% to 629%; Figure 3a), except for the prediction with the Edlund model (rBias +16%; 95% CI –5% to +36%, rRMSE 77%).

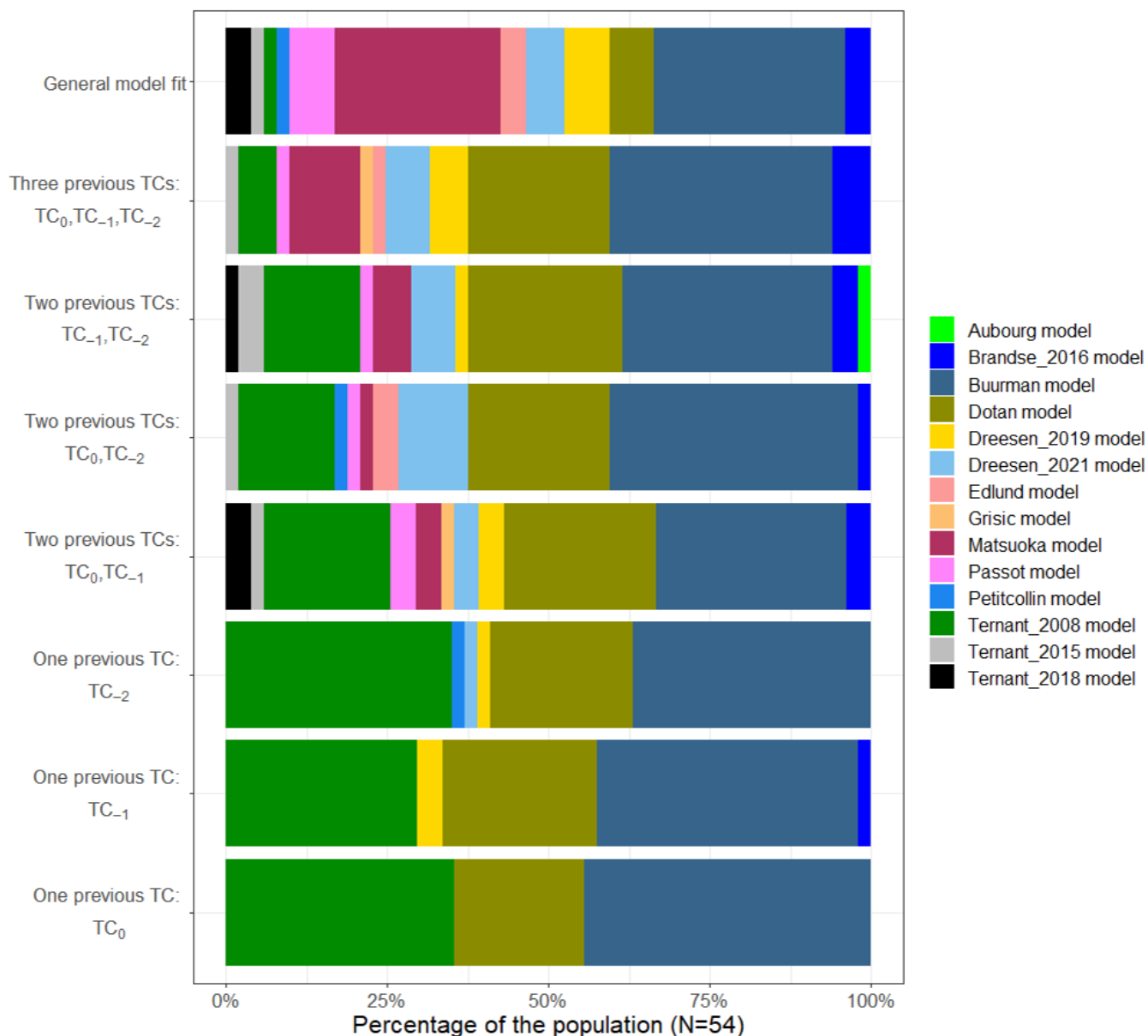
Providing one previous TC ( $TC_0$ ,  $TC_{-1}$ , or  $TC_{-2}$ ) greatly improved the predictive performance (rBias –27% to +38%, rRMSE 28% to 69%; Figure 3b). Providing more than one previous TC improved the predictive performances only marginally (Figure S5). Compared with the

single-model approach, the predictive performances of multi-model approaches were less sensitive to the number of provided TCs for MAP.

Multi-model averaging algorithm performed systematically better than MSA both in terms of accuracy and precision. MAA provided more precise predictions than MSA in all a posteriori prediction settings (one previous TC: rRMSE 33% to 41% for the MAA vs. rRMSE 50% to 57% for the MSA; 3 previous TCs: rRMSE 30% for MAA vs. rRMSE 46% for MSA; Figure 3b, Figure S5).

## Robustness analysis and software implementation

Four candidate models were selected considering their overall predictive performance metrics in the a posteriori prediction settings (Aubourg model, Dreesen\_2021 model, and Passot model; all with a negative rBias), and the best model with positive rBias (Ternant\_2008 model). The predictive performances of the multi-model



**FIGURE 2** The weight distribution of population pharmacokinetic models in the study population in seven a posteriori prediction settings and the general model fit setting. Four candidate models that were not selected by any patient in any of the evaluated settings are not shown in the legend (Brandse\_2017, Fasanmade\_2009, Fasanmade\_2011, and Xu model). TC, trough concentration.

approaches with only the four models were in good agreement with the multi-model approaches including all models (Figure 3a,b; Figure S5). In addition, by providing at minimum one previous TC, the predictive performances of both multi-model approaches were clinically acceptable even when only three or two instead of four models were used (rBias  $-4\%$  to  $+2\%$ ; Figure S6).

The average computation time of the multi-model approaches using only the four models decreased five-fold from the multi-model approaches using all 18 models (average 0.115 s vs. 0.576 s per patient).

An infliximab module was added to TDMx (<https://tdmx.shinyapps.io/infliximab/>). Results of the objective

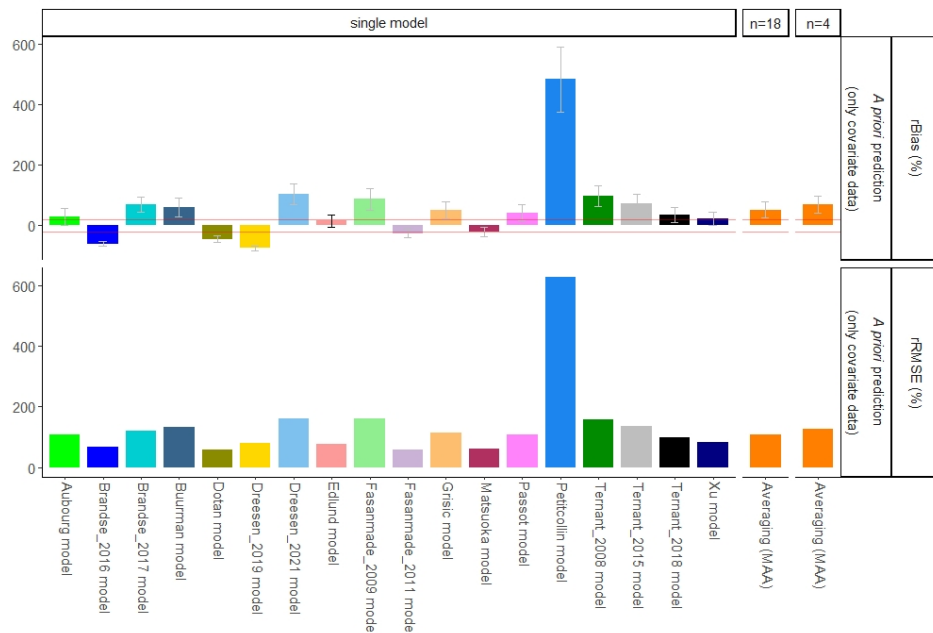
function values and model averaging predictions using TDMx were in good agreement with NONMEM (a posteriori prediction settings with  $TC_{-1}$ ; Figures S7 and S8).

### Bland–Altman analysis, classification accuracy, and sensitivity analysis

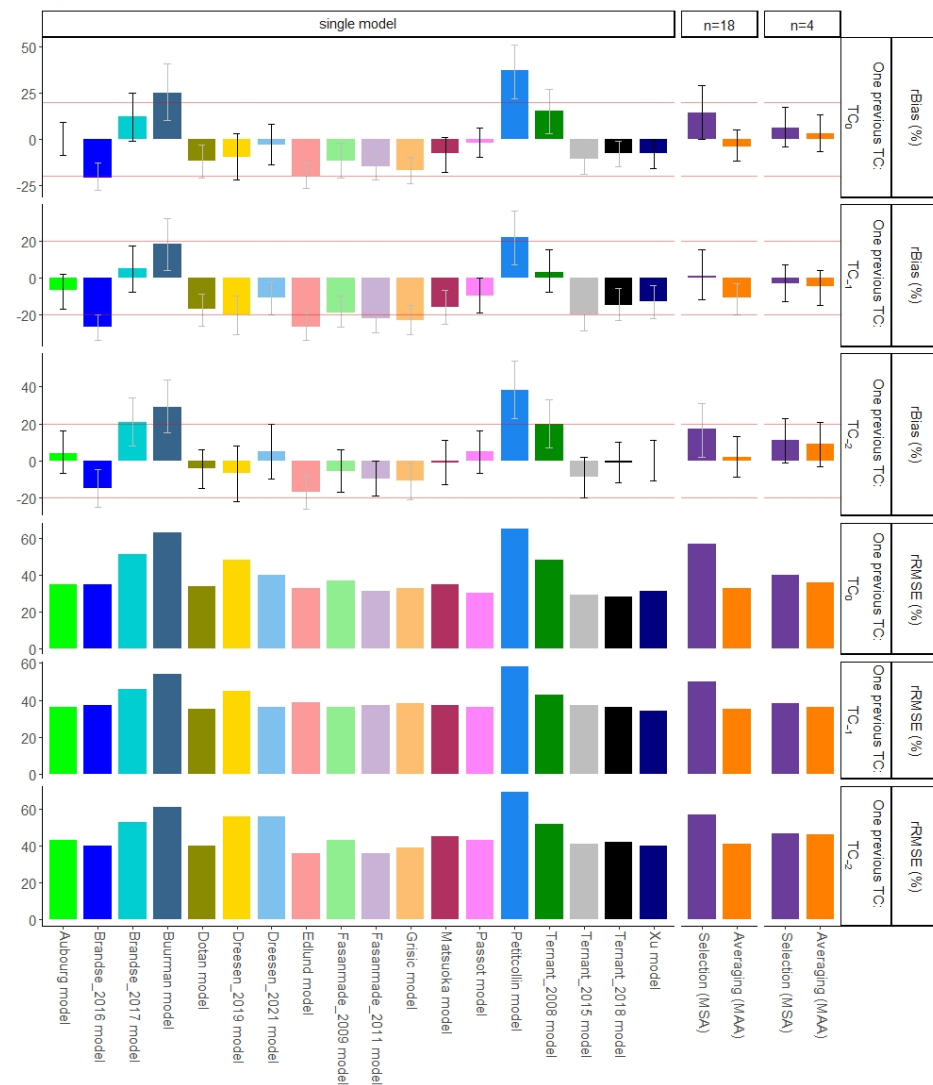
The tendency of prediction bias across the measured concentration range from 3.0 to 10.0 mg/L was the least by providing only  $TC_0$  in Bayesian forecasting (Figure 4).

A priori predictions of both single and multi-model approaches had the lowest classification accuracy

(a) *A priori* prediction



(b) *A posteriori* prediction





**FIGURE 3** The predictive performance of 18 single candidate population pharmacokinetic models versus the multi-model approaches using all 18 models versus the four models for predicting the infliximab trough concentration (TC) at time + 1 ( $TC_{+1}$ ). (a) a priori prediction (with only covariate data); (b) a posteriori prediction settings using covariate data and one previous TC ( $TC_0$ ,  $TC_{-1}$ , or  $TC_{-2}$ ). Whiskers indicate the 95% confidence interval (CI) of the relative bias calculated via the standard error (black whiskers indicate 95% CIs including 0). Horizontal red lines indicate  $\pm 20\%$  range of the relative that is deemed clinically acceptable. *Note:* Model weights during a priori prediction are equal ( $1/\text{number of models}$ ), precluding a model selection procedure in this setting. MAA, multi-model averaging algorithm; MSA, multi-model selection algorithm; rBias, relative bias; rRMSE, relative root mean square error.

(median 59%, IQR 59%–63%) and the highest percentage of falsely predicting the  $TC_{+1} \geq 5.0$  mg/L (false negative rate median 35%, IQR 30%–37%; **Figure 5**). In comparison with the a priori prediction, providing at least one previous TC significantly improved not only the classification accuracy (median 72%, IQR 71%–76%;  $p < 0.05$ ) but also significantly decreased the false negative prediction rate (median 8%, IQR 6%–15%;  $p < 0.05$ ). A posteriori prediction with the  $TC_0$  resulted in a significantly higher classification accuracy than with the  $TC_{-1}$  or the  $TC_{-2}$  ( $p < 0.01$ ). In addition, the availability of the  $TC_0$  significantly lowered the chance of falsely predicting the  $TC_{+1} < 5.0$  mg/L in comparison with only providing the  $TC_{-1}$  ( $p = 0.004$ ). Providing more than one previous TC did not improve the classification accuracy metrics (**Figure 5**). However, the classification performances of the MAA were not significantly different from the MSA and the other single models ( $p > 0.10$ ), except for the a posteriori predictions using the Ternant\_2008 model with the  $TC_{-1}$  ( $p = 0.023$ ; **Figure 5**).

The predictive performance of the single-model approach with any model was maintained when applying median covariate imputation (**Figure 6**). In addition, there was no change of accuracy and precision of predictive performances for multi-model approaches in the a priori setting (MAA: rBias +68%, rRMSE 125% for true value vs. rBias +66%, rRMSE 125% for imputed value), and the a posteriori setting (MAA: rBias –5%, rRMSE 36% for both true value and imputed value; MSA: rBias –3%, rRMSE 38% for true value vs. rBias –1%, rRMSE 39% for imputed value).

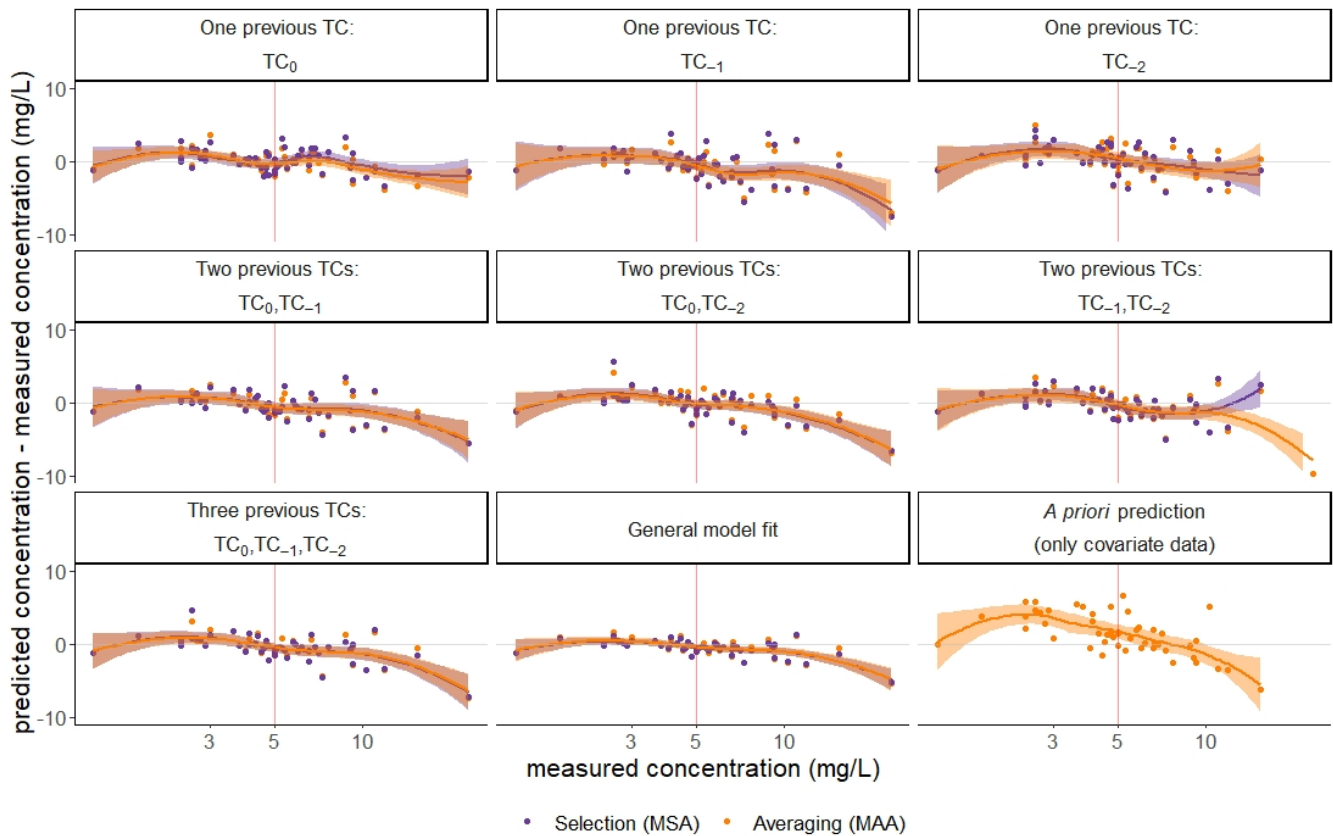
## DISCUSSION

The selection of a PopPK model for guiding individualized dose optimization is a crucial step in MIPD. For infliximab, 18 PopPK models have been developed to describe the PK characteristics of adult patients with IBD. To date, the benefits of MIPD with a single infliximab model in patients with IBD have been evidenced both retrospectively<sup>20</sup> and prospectively.<sup>21</sup> However, alternative approaches that integrate multiple PopPK models have not been investigated for infliximab. In our study, we found that an MAA resulted in the most

accurate and precise a posteriori predictions, regardless of the number of TCs provided, as compared to a single-model approach. A priori prediction using covariate data alone resulted in biased and imprecise predictions with either single- or multi-model approaches. The predictive performance of both single- and multi-model approaches was robust to the lack of covariate data.

PK variability of infliximab is challenging for traditional flowchart-guided TDM. No significant clinical benefits were shown for proactive TDM during infliximab induction therapy in patients with immune-mediated inflammatory diseases (i.e., NOR-DRUM A<sup>22</sup>). During infliximab maintenance therapy, the clinical benefit of proactive TDM could also not be addressed in patients with IBD (i.e., TAXIT<sup>23</sup> and TAILORIX<sup>24</sup>), yet it was addressed in a mixed population of patients with immune-mediated inflammatory diseases (i.e., NOR-DRUM B<sup>25</sup>). Recently, the PRECISION<sup>21</sup> trial using a single-model approach implemented in a Bayesian dashboard for infliximab dosing showed significant clinical benefit over label dosing during maintenance therapy. Due to the acknowledged benefits of MIPD in personalized medicine,<sup>11,26</sup> great efforts are being made to improve components of MIPD, such as methods for the selection of models<sup>13,27</sup> and methods for the estimation of parameters.<sup>28,29</sup> In this study, we investigated alternative approaches allowing the incorporation of multiple PopPK models simultaneously for MIPD. The MSA and MAA could provide more flexibility in PK parameter estimation and potentially increase generalizability to unseen data compared to MIPD using a single-model approach. In agreement with findings from Uster et al.<sup>13</sup> using vancomycin as a case study, the multi-model approaches had better predictive performance than any single-model approach. Furthermore, we found that the MAA outperformed the MSA and single-model approaches because the MAA systematically resulted in more precise predictions.

The predictive performance of infliximab PopPK models was previously externally evaluated in patients with inflammatory diseases,<sup>30</sup> including patients with IBD.<sup>31,32</sup> In the studies of Santacana et al.<sup>31</sup> and Schräpel et al.,<sup>32</sup> the two models developed by Fasanmade et al.<sup>33,34</sup> (using data from the phase III trials) demonstrated the best predictive

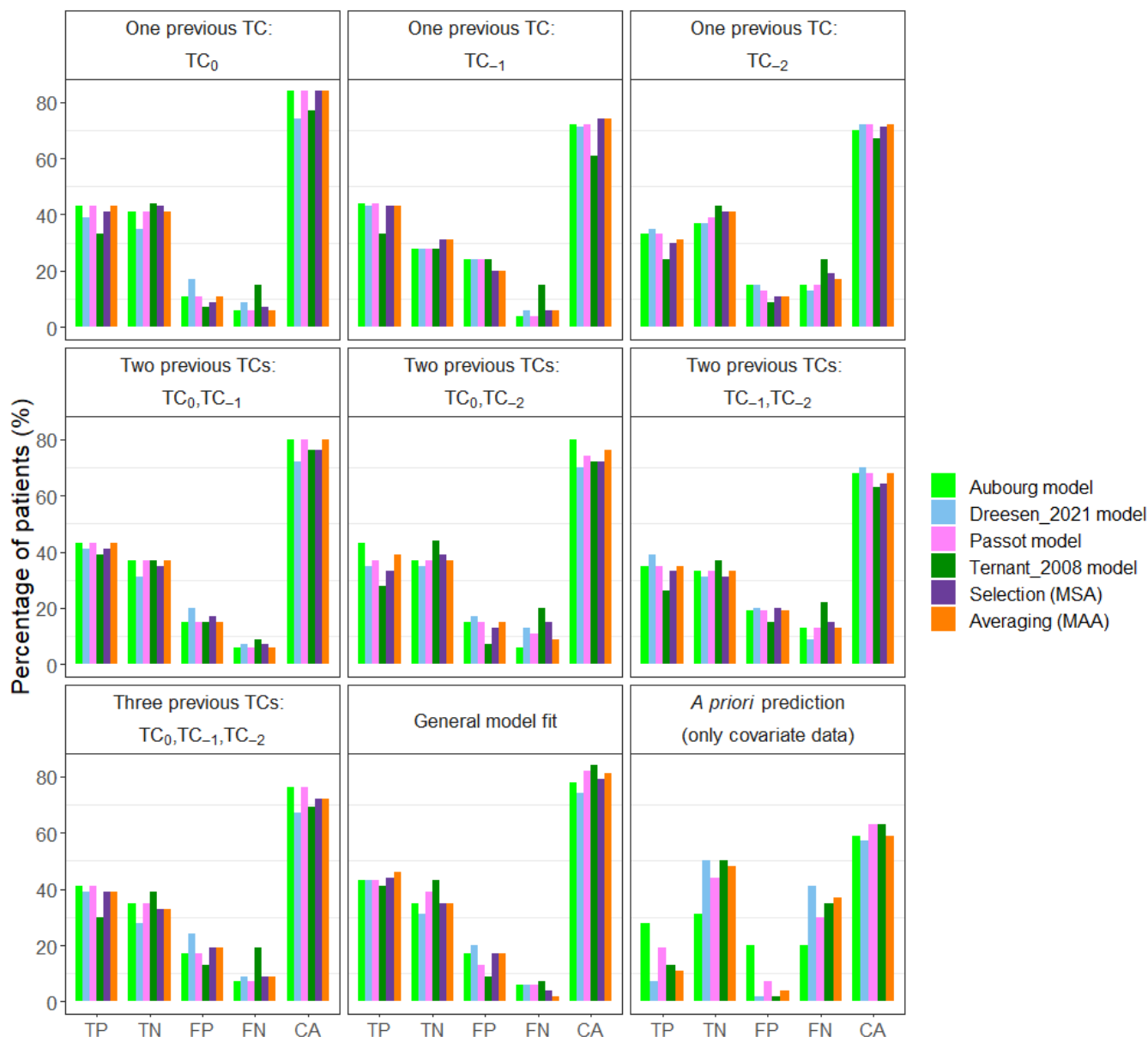


**FIGURE 4** Bland–Altman plots showing the agreement between the measured infliximab concentrations and the predicted infliximab concentrations across the range of measured infliximab concentrations in various prediction settings using the model averaging algorithm (MAA; orange) and the model selection algorithm (MSA; purple). The vertical red line indicates the 5.0 mg/L trough concentration (TC) target. The solid line with shaded area represents a locally weighted smoother with its 95% confidence interval based on the data (MAA in orange and MSA in purple).

performance in patients with IBD. In our study, both Fasanmade models gave inaccurate a posteriori predictions in most of the evaluated settings. In addition, these two models were not selected for any of the patients in the MSA. The differences in predictive performances of candidate models between studies could potentially be caused by differences in the approaches used to assess the model's predictive performance (e.g., provided measured concentration for MAP, the estimation of empirical Bayes PK parameter approach, and predictive performance metrics). The observed differences emphasize the importance of site-specific external validation prior to clinical implementation. In our study, we evaluated predictive performance of candidate models for predicting TCs. Infliximab clearance is the PK parameter that mainly drives the TC. Our case is different from, for example, vancomycin, where the exposure target is an area under the curve, which is driven by all PopPK parameters. Therefore, as expected, we did not observe any difference in the predictive performance of one- and two-compartment models (data not shown). Although we reduced the number of models participating in the multi-model approach to gain computation time

without losing predictive performance, this action may not be as innocent as it appears and may show to be a sacrifice in a more extensive external validation/application. Therefore, external validation with all identified 18 models may be suggested prior to using our developed software in other settings.

In our study, we used a comprehensive set of model qualification tools, ranging from closeness of study population and goodness-of-fit plots over predictive performance and classification accuracy assessments to Bland–Altman analysis and sensitivity analysis. Nevertheless, to control the inherent risks associated with PK prediction as much as possible, a wider set of diagnostic tools for model qualification for MIPD may still be needed.<sup>26</sup> Apparent contradictory findings between model qualification tools are common. A model that conforms to various model evaluation standards may not perform well in the prediction evaluations. For example, the Edlund model fitted the data worst, but it was the only model with clinically acceptable a priori predictions. Furthermore, whereas the Petitcollin model was developed using data from a clinical setting closest to the one that we are studying, the

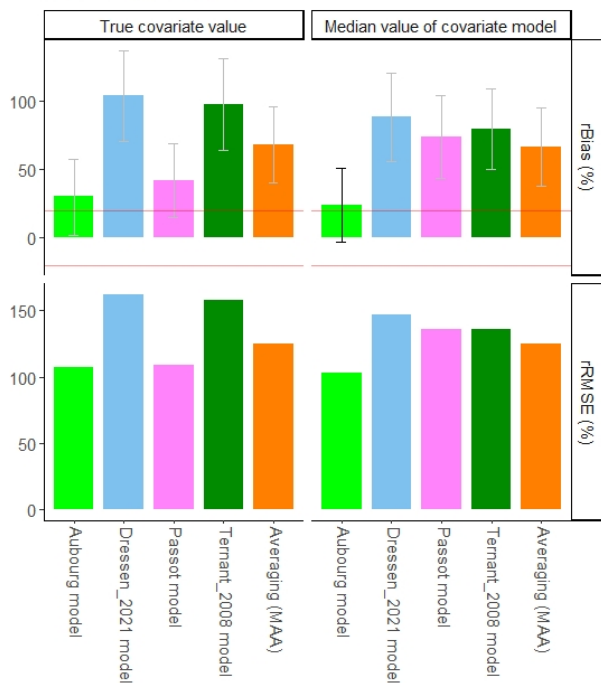
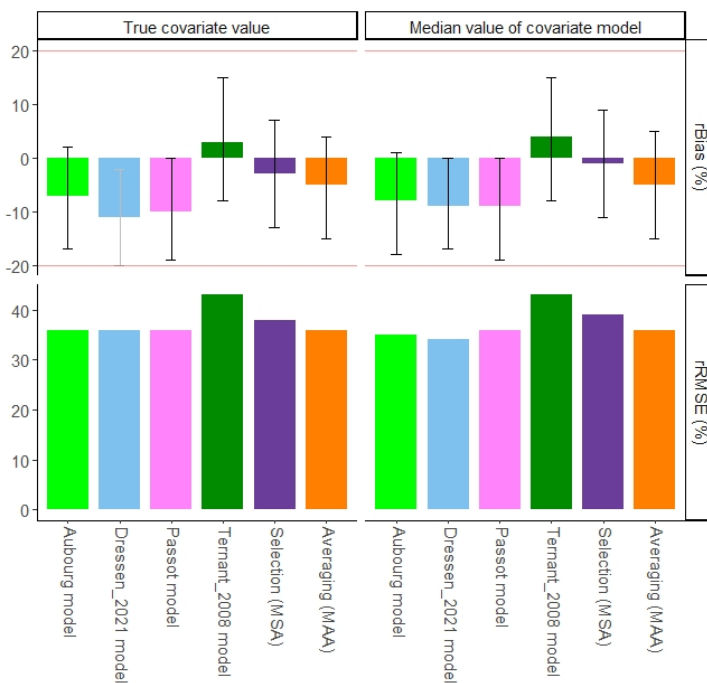


**FIGURE 5** The percentage of patients ( $N = 54$ ) in four classes based on the predicted and measured  $TC_{+1}$  according to the prespecified trough concentration (TC) target of 5.0 mg/L in various prediction settings: (i) true positive (TP): both measured and predicted  $<5.0$  mg/L; (ii) true negative (TN): both measured and predicted  $\geq 5.0$  mg/L; (iii) false positive (FP): measured  $\geq 5.0$  mg/L, but predicted  $<5.0$  mg/L; (iv) false negative (FN): measured  $<5.0$  mg/L, but predicted  $\geq 5.0$  mg/L. (v) Classification accuracy (CA): the number of correct predictions (TP and TN) divided by the total number of predictions ( $n = 54$ ).

Buurman model was the best model based on VPC and NPDE. Nevertheless, both models did not perform well in a priori and a posteriori predictions. The a priori prediction is a population prediction based solely on covariate data, whereas VPC and NPDE take into account both covariate and concentration in the evaluations. Therefore, an a priori prediction performance may not be indicated via VPC and NPDE. The complementary use of a comprehensive set of model qualification tools should be considered during model selection. In addition, standard goodness-of-fit evaluations are not appropriate for evaluating the

suitability and predictive performance of models for MIPD. Yet, because the multi-model approaches rely on the calculation of model weights based on a goodness-of-fit measure, the standard model evaluation toolkit should not just yet be discarded, and the relation between the descriptive and predictive ability of models requires further investigation.

A single TC suffices to allow accurate and precise MIPD. Based on our findings, the acceptable timeframe of TC monitoring to predict the  $TC_{+1}$  accurately was TC from previously consecutive dosing that was not further

(a) *A priori* prediction(b) *A posteriori* prediction with TC<sub>-1</sub>

**FIGURE 6** Comparison of the predictive performance between scenarios with and without covariate data available. The scenario of missing covariate information used a single imputation strategy with the median covariate value around which the covariate effect was centered in the respective model. (a) *A priori* prediction (with only covariate data); (b) the *a posteriori* prediction settings using covariate data and one previous TC (TC<sub>-1</sub>). Whiskers indicate the 95% confidence interval (CI) of the relative bias calculated via the standard error (black whiskers indicate 95% CIs including 0). Horizontal red lines indicate  $\pm 20\%$  range of the relative that is deemed clinically acceptable. MAA, multi-model averaging algorithm; MSA, multi-model selection algorithm; rBias, relative bias; rRMSE, relative root mean square error.

than three dosing intervals before dosage regimen de-escalation. Due to interoccasion variability, an “old” concentration may have lost the ability to predict future exposure. Therefore, the predictive performance of MIPD using TCs from the later timepoints may require further investigation. Moreover, providing only one TC adequately informs about PK parameters and subsequently makes the covariate data become relatively unimportant for the predictive performance. Median imputation of missing covariate data is, therefore, a safe strategy. This finding is intuitive, knowing that covariates generally only explain a small part of the interindividual variability (up to 6% for clearance<sup>33,34</sup>), whereas Bayesian forecasting can identify the remaining, often high “unexplained” interindividual variability (median of 32.7%, IQR 28.0–36.0% on clearance<sup>33,34</sup>).

Theoretically, utilizing point-of-care testing may improve the clinical and economic benefits of MIPD. In this study, we found that a single most recent TC (at T<sub>0</sub>) resulted in the highest classification accuracy with not only a low chance of falsely predicting the TC<sub>+1</sub>  $\geq 5.0$  mg/L (i.e., risk of losing therapeutic response) but also a low chance of falsely predicting the TC<sub>+1</sub>  $< 5.0$  mg/L (i.e., risk of unnecessary dose escalation). However, a recently published

prospective study using a rapid assay during traditional flowchart-guided proactive TDM (i.e., a decision making flowchart designed to maintain infliximab concentration within the desirable therapeutic range) could not show clinical benefits in patients with IBD during infliximab maintenance therapy.<sup>35</sup> Nevertheless, a rapid assay may show its full potential when used in combination with an MIPD software tool. Yet, a prospective evaluation is warranted.

An MIPD approach could potentially improve the treatment efficacy in patients undergoing dosage regimen de-escalation. Petitcollin et al.<sup>36</sup> reported that the clearance of infliximab in these patients was not only a factor in patient selection but also a predictor for disease relapse after treatment de-escalation. In addition, the infliximab clearance gradually increased over time in association with body weight variations. Therefore, the MAA as implemented in the TDMx Bayesian forecasting software tool will be used to guide infliximab dosing in the forthcoming prospective MODIFI study (NCT04982172). In the MODIFI study, we aim to deliver proof-of-concept of the superiority of MIPD over empirical dosage regimen de-escalation. The primary end point is the proportion of patients maintaining steroid-free, combined clinical and

biological remission during 1 year after the start of infliximab de-escalation.

This study had several strengths. First, we evaluated the predictive performance of multi-model approaches for MIPD in a very different context (i.e., biological drug and chronic disease) from Uster et al.<sup>13</sup> (i.e., vancomycin and infectious diseases). Second, additional analysis tools (i.e., Bland–Altman analysis and classification accuracy) for evaluating the fit-for-purpose of PopPK models for use in MIPD were introduced. Currently, there is no well-established target of classification accuracy for MIPD approach. To the best of our knowledge, classification accuracy has only been included for model predictive performance evaluation for infliximab in Schräpel et al.<sup>32</sup> Therefore, defining clinical relevance of these additional analysis tools still requires further investigation to facilitate the translation and appropriate use of the MIPD approaches in clinical care. Third, we also scrutinized the importance of utilizing point-of-care testing and the availability of covariate data on the predictive performances.

Still, our study had some limitations. First, incomplete reporting of information on error models, median values for centering covariate effects, and variance–covariance matrices limited the reproducibility of the published PopPK models. Therefore, assumptions had to be made for the missing information (Table S2). In recent years, the importance of an “Open” approach to science and the accessibility to mathematical models has become well-recognized as a crucial step in maintaining reproducibility, rigor, and integrity in published pharmacometrics models.<sup>37</sup> Second, a limited number of patient data from a single clinical center was used in this analysis. This study was an exploratory study and so was not powered to obtain statistical significance. Therefore, interpretation of our results should be done with care and we recognize the importance of continued validation of our MIPD algorithm in patients with IBD in other clinical centers.<sup>38</sup> In addition, the need for center-specific external validation of our algorithm will be required before broader clinical implementation. The differences between clinical centers include level of health care (e.g., primary care, secondary care, and tertiary care), bioanalysis method, clinical workflows, etc. To allow us and others to do so, we provide the weblink to the MIPD tool in this paper. Third, due to the retrospective nature of our study, a potential selection bias cannot be ruled out. We only collect data of patients who have given written informed consent for collecting their data and serum samples. Therefore, future prospective confirmation of our findings will be needed. Last, the generalizability of our work beyond the studied clinical context will require further investigation to rule out potential bias. We studied the value of MIPD specifically for guiding dose de-escalation, but the value of our work may be of

interest in other clinical scenarios as well (e.g., dose intensification, proactive TDM, and reactive TDM). In addition, concentration data used in this analysis were measured using only one commercially available assay. Therefore, external validations with larger and different cohorts in other clinical centers using other bioanalysis assays are needed to confirm the generalizability of our work.

To conclude, we developed a robust and precise MAA for guiding infliximab MIPD using a single recently measured TC. The algorithm is implemented in the freely available TDMx software tool and will be evaluated in the prospective MODIFI study (NCT04982172).

## AUTHOR CONTRIBUTIONS

W.K. wrote the manuscript. W.K. and E.D. designed the research. W.K., S.G.W., A.K., Z.W., and E.D. performed the research. W.K. and E.D. analyzed the data. A.O., D.T., J.S., S.V., and M.F. contributed new reagents/analytical tools.

## ACKNOWLEDGEMENTS

The authors thank Melissa Nigro and Julie Thijs for collecting the clinical data, Vera Ballet for maintaining the VLECC database, Sophie Tops and Griet Compernelle for performing the infliximab and anti-infliximab antibody assays, Omar Elkayal for helping with the creation of the infliximab model repository, and Professor Deirdre Cabooter for the guidance and management of PhD student Alexander Kensert. The authors acknowledge that parts of this analysis were submitted to the 2021 meeting of the Population Approach Group in Europe (PAGE; PAGE 29 [2021] Abstr 9757) and the 2022 congress of the European Crohn’s and Colitis Organization (ECCO; Poster P333).

## CONFLICT OF INTEREST

J.S. received financial support for research from Galapagos, speaker fees from Abbvie, Falk, Takeda, Janssen, and Fresenius, and consultancy fees from Janssen and Ferring. S.V. received financial support for research from AbbVie, Johnson and Johnson, Pfizer, Galapagos, and Takeda, and consultancy and/or speaker fees from AbbVie, Abivax, Agomab, Arena Pharmaceuticals, Avaxia, Bristol Myers Squibb, Boehringer Ingelheim, Celgene, Dr. Falk Pharma, Ferring, Galapagos, Genentech-Roche, Gilead, GSK, Hospira, Janssen, Mundipharma, MSD, Pfizer, Prodigest, Progenity, Prometheus, Robarts Clinical Trials, Second Genome, Shire, Surrozen, Takeda, Theravance, and Tillots Pharma AG. M.F. received financial support for research from AbbVie, Amgen, Biogen, Janssen, Pfizer, Takeda, and Viatrix, speaker fees from AbbVie, Amgen, Biogen, Boehringer Ingelheim, Falk, Ferring,

Janssen, Lamepro, MSD, Mylan, Pfizer, Sandoz, Takeda, and Truvion Healthcare, and consultancy fees from AbbVie, Boehringer Ingelheim, Celltrion, Janssen, Lilly, Medtronic, MSD, Pfizer, Sandoz, Takeda, and Thermo Fisher. All other authors declare no conflicts of interest.

## ORCID

Wanee Kantasiripitak  <https://orcid.org/0000-0001-9285-799X>

An Outtier  <https://orcid.org/0000-0002-7351-0023>

Sebastian G. Wicha  <https://orcid.org/0000-0002-8773-4845>

Alexander Kensert  <https://orcid.org/0000-0002-5295-010X>

Zhigang Wang  <https://orcid.org/0000-0001-6415-2044>

João Sabino  <https://orcid.org/0000-0002-8892-7075>

Séverine Vermeire  <https://orcid.org/0000-0001-9942-3019>

Debby Thomas  <https://orcid.org/0000-0003-3075-9846>

Marc Ferrante  <https://orcid.org/0000-0003-1492-0716>

Erwin Dreesen  <https://orcid.org/0000-0002-0785-2930>

## REFERENCES

- Torres J, Mehandru S, Colombel J-F, Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389:1741-1755.
- Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel J-F. Ulcerative colitis. *Lancet*. 2017;389:1756-1770.
- Ben-Horin S, Chowers Y. Review article: loss of response to anti-TNF treatments in Crohn's disease. *Aliment Pharmacol Ther*. 2011;33:987-995.
- Melsheimer R, Geldhof A, Apaolaza I, Schaible T. Remicade® (infliximab): 20 years of contributions to science and medicine. *Biol Theory*. 2019;13:139-178.
- Seow CH, Newman A, Irwin SP, Steinhart AH, Silverberg MS, Greenberg GR. Trough serum infliximab: a predictive factor of clinical outcome for infliximab treatment in acute ulcerative colitis. *Gut*. 2010;59:49-54.
- Einarson TR, Bereza BG, Ying Lee X, Lelli F. Dose escalation of biologics in Crohn's disease: critical review of observational studies. *Curr Med Res Opin*. 2017;33:1433-1449.
- Bodger K. Cost effectiveness of treatments for inflammatory bowel disease. *Pharmacoeconomics*. 2011;29:387-401.
- Williams CJM, Peyrin-Biroulet L, Ford AC. Systematic review with meta-analysis: malignancies with anti-tumour necrosis factor- $\alpha$  therapy in inflammatory bowel disease. *Aliment Pharmacol Ther*. 2014;39:447-458.
- Ford AC, Peyrin-Biroulet L. Opportunistic infections with anti-tumor necrosis factor- $\alpha$  therapy in inflammatory bowel disease: meta-analysis of randomized controlled trials. *Am J Gastroenterol*. 2013;108:1268-1276.
- Lucidarme C, Petitcollin A, Brochard C, et al. Predictors of relapse following infliximab de-escalation in patients with inflammatory bowel disease: the value of a strategy based on therapeutic drug monitoring. *Aliment Pharmacol Ther*. 2019;49:147-154.
- Wang Z, Dreesen E. Therapeutic drug monitoring of anti-tumor necrosis factor agents: lessons learned and remaining issues. *Curr Opin Pharmacol*. 2020;55:53-59.
- Kantasiripitak W, van Daele R, Gijsen M, Ferrante M, Spriet I, Dreesen E. Software tools for model-informed precision dosing: how well do they satisfy the needs? *Front Pharmacol*. 2020;11:620.
- Uster DW, Stocker SL, Carland JE, et al. A model averaging/selection approach improves the predictive performance of model-informed precision dosing: vancomycin as a case study. *Clin Pharmacol Ther*. 2021;109:175-183.
- Broeker A, Nardecchia M, Klinker KP, et al. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. *Clin Microbiol Infect*. 2019;25:1286.e1-1286.e7. doi:10.1016/j.cmi.2019.02.029
- Vande Castele N, Buurman DJ, Sturkenboom MGG, et al. Detection of infliximab levels and anti-infliximab antibodies: a comparison of three different assays. *Aliment Pharmacol Ther*. 2012;36:765-771.
- Brendel K, Comets E, Laffont C, Laveille C, Mentré F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm Res*. 2006;23:2036-2049.
- Wicha SG, Kees MG, Solms A, Minichmayr IK, Kratzer A, Kloft C. TDMx: a novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine. *Int J Antimicrob Agents*. 2015;45:442-444.
- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J Roy Stat Soc D*. 1983;32:307-317.
- Feuerstein JD, Nguyen GC, Kupfer SS, et al. American Gastroenterological Association Institute guideline on therapeutic drug monitoring in inflammatory bowel disease. *Gastroenterology*. 2017;153:827-834.
- Faelens R, Wang Z, Bouillon T, et al. Model-informed precision dosing during infliximab induction therapy reduces variability in exposure and endoscopic improvement between patients with ulcerative colitis. *Pharmaceutics*. 2021;13:1623.
- Strik AS, Löwenberg M, Mould DR, et al. Efficacy of dashboard driven dosing of infliximab in inflammatory bowel disease patients; a randomized controlled trial. *Scand J Gastroenterol*. 2021;56:145-154.
- Syversen SW, Goll GL, Jørgensen KK, et al. Effect of therapeutic drug monitoring vs standard therapy during infliximab induction on disease remission in patients with chronic immune-mediated inflammatory diseases: a randomized clinical trial. *JAMA*. 2021;325:1744-1754.
- Vande Castele N, Ferrante M, van Assche G, et al. Trough concentrations of infliximab guide dosing for patients with inflammatory bowel disease. *Gastroenterology*. 2015;148:1320-1329.e3.
- D'Haens G, Vermeire S, Lambrecht G, et al. Increasing infliximab dose based on symptoms, biomarkers, and serum drug concentrations does not increase clinical, endoscopic, and corticosteroid-free remission in patients with active luminal Crohn's disease. *Gastroenterology*. 2018;154:1343-1351.e1.
- Syversen SW, Jørgensen KK, Goll GL, et al. Effect of therapeutic drug monitoring vs standard therapy during maintenance infliximab therapy on disease control in patients with immune-mediated inflammatory diseases: a randomized clinical trial. *JAMA*. 2021;326:2375-2384.
- Keizer RJ, ter Heine R, Frymoyer A, Lesko LJ, Mangat R, Goswami S. Model-informed precision dosing at the bedside: scientific challenges and opportunities. *CPT Pharmacometrics Syst Pharmacol*. 2018;7:785-787.

27. Maier C, de Wiljes J, Hartung N, Kloft C, Huisinga W. A continued learning approach for model-informed precision dosing: updating models in clinical practice. *CPT Pharmacometrics Syst Pharmacol*. 2021;11:185-198. doi:10.1002/psp4.12745
28. Maier C, Hartung N, Kloft C, Huisinga W, de Wiljes J. Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:241-254.
29. Hughes JH, Keizer RJ. A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1150-1160.
30. Konecki C, Feliu C, Cazaubon Y, et al. External evaluation of population pharmacokinetic models and Bayes-based dosing of infliximab. *Pharmaceutics*. 2021;13:1191.
31. Santacana E, Rodríguez-Alonso L, Padullés A, et al. External evaluation of population pharmacokinetic models of infliximab in patients with inflammatory bowel disease. *Ther Drug Monit*. 2018;40:120-129.
32. Schräpel C, Kovar L, Selzer D, et al. External model performance evaluation of twelve infliximab population pharmacokinetic models in patients with inflammatory bowel disease. *Pharmaceutics*. 2021;13:1368.
33. Fasanmade AA, Adedokun OJ, Ford J, et al. Population pharmacokinetic analysis of infliximab in patients with ulcerative colitis. *Eur J Clin Pharmacol*. 2009;65:1211-1228.
34. Fasanmade AA, Adedokun OJ, Blank M, Zhou H, Davis HM. Pharmacokinetic properties of infliximab in children and adults with Crohn's disease: a retrospective analysis of data from 2 phase III clinical trials. *Clin Ther*. 2011;33:946-964.
35. Bossuyt P, Pouillon L, Claeys S, et al. Ultra-proactive therapeutic drug monitoring of infliximab based on point-of-care-testing in inflammatory bowel disease: results of a pragmatic trial. *J Crohns Colitis*. 2021;16:199-206. doi:10.1093/ecco-jcc/jjab127
36. Petitcollin A, Brochard C, Siproudhis L, et al. Pharmacokinetic parameters of infliximab influence the rate of relapse after de-escalation in adults with inflammatory bowel diseases. *Clin Pharmacol Ther*. 2019;106:605-615.
37. Rowland Yeo K, Hennig S, Krishnaswami S, et al. CPT: Pharmacometrics & Systems Pharmacology – inception, maturation, and future vision. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:649-657.
38. Keizer R. Experiences in applied clinical pharmacometrics: challenges, recommendations, and research opportunities. 2018;27. [www.page-meeting.org/?abstract=8758](http://www.page-meeting.org/?abstract=8758). Accessed February 7, 2022.
39. Aubourg A, Picon L, Lecomte T, Bejan-Angoulvant T, Paintaud G, Ternant D. A robust estimation of infliximab pharmacokinetic parameters in Crohn's disease. *Eur J Clin Pharmacol*. 2015;71:1541-1542.
40. Brandse JF, Mathôt RA, van der Kleij D, et al. Pharmacokinetic features and presence of antidrug antibodies associate with response to infliximab induction therapy in patients with moderate to severe ulcerative colitis. *Clin Gastroenterol Hepatol*. 2016;14:251-258.e1-2.
41. Brandse JF, Mould D, Smeekees O, et al. A real-life population pharmacokinetic study reveals factors associated with clearance and immunogenicity of infliximab in inflammatory bowel disease. *Inflamm Bowel Dis*. 2017;23:650-660.
42. Buurman DJ, Maurer JM, Keizer RJ, Kosterink JGW, Dijkstra G. Population pharmacokinetics of infliximab in patients with inflammatory bowel disease: potential implications for dosing in clinical practice. *Aliment Pharmacol Ther*. 2015;42:529-539.
43. Dotan I, Ron Y, Yanai H, et al. Patient factors that increase infliximab clearance and shorten half-life in inflammatory bowel disease: a population pharmacokinetic study. *Inflamm Bowel Dis*. 2014;20:2247-2259.
44. Dreesen E, Faelens R, van Assche G, et al. Optimising infliximab induction dosing for patients with ulcerative colitis. *Br J Clin Pharmacol*. 2019;85:782-795. doi:10.1111/bcp.13859
45. Dreesen E, Berends S, Laharie D, et al. Modelling of the relationship between infliximab exposure, faecal calprotectin and endoscopic remission in patients with Crohn's disease. *Br J Clin Pharmacol*. 2021;87:106-118.
46. Edlund H, Steenholdt C, Ainsworth MA, et al. Magnitude of increased infliximab clearance imposed by anti-infliximab antibodies in Crohn's disease is determined by their concentration. *AAPS J*. 2017;19:223-233.
47. Griscic A-M, Eser A, Huisinga W, Reinisch W, Kloft C. Quantitative relationship between infliximab exposure and inhibition of C-reactive protein synthesis to support inflammatory bowel disease management. *Br J Clin Pharmacol*. 2021;87:2374-2384.
48. Matsuoka K, Hamada S, Shimizu M, et al. Factors contributing to the systemic clearance of infliximab with long-term administration in Japanese patients with Crohn's disease: analysis using population pharmacokinetics. *Int J Clin Pharmacol Ther*. 2020;58:89-102.
49. Passot C, Mulleman D, Bejan-Angoulvant T, et al. The underlying inflammatory chronic disease influences infliximab pharmacokinetics. *MAbs*. 2016;8:1407-1416.
50. Ternant D, Aubourg A, Magdelaine-Beuzelin C, et al. Infliximab pharmacokinetics in inflammatory bowel disease patients. *Ther Drug Monit*. 2008;30:523-529.
51. Ternant D, Berkane Z, Picon L, et al. Assessment of the influence of inflammation and FCGR3A genotype on infliximab pharmacokinetics and time to relapse in patients with Crohn's disease. *Clin Pharmacokinet*. 2015;54:551-562.
52. Ternant D, Passot C, Aubourg A, et al. Model-based therapeutic drug monitoring of infliximab using a single serum trough concentration. *Clin Pharmacokinet*. 2018;57:1173-1184.
53. Wojciechowski J, Upton RN, Mould DR, Wiese MD, Foster DJR. Infliximab maintenance dosing in inflammatory bowel disease: an example for in silico assessment of adaptive dosing strategies. *AAPS J*. 2017;19:1136-1147.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Kantasiripitak W, Outtier A, Wicha SG, et al. Multi-model averaging improves the performance of model-guided infliximab dosing in patients with inflammatory bowel diseases. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1045-1059. doi:10.1002/psp4.12813