Received: 22 December 2023 | Revised: 5 July 2024 | Accepted: 14 July 2024

DOI: 10.1002/psp4.13213

ARTICLE

Simulation-based evaluation of the Pharmpy Automatic Model Development tool for population pharmacokinetic modeling in early clinical drug development

Zrinka Duvnjak[1,2,3] | **Franziska Schaedeli Stark**[1] | **Valérie Cosson**[1] | **Sylvie Retout**[1] | **Emilie Schindler**[1] | **João A. Abrantes**[1]

[1]Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Basel, Switzerland

[2]Department of Clinical Pharmacy and Biochemistry, Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany

[3]Graduate Research Training Program PharMetrX, Berlin/Potsdam, Germany

Correspondence
João A. Abrantes, Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Grenzacherstrasse 124, Basel, Switzerland.
Email: joao.abrantes@roche.com

Abstract

The Pharmpy Automatic Model Development (AMD) tool automates the building of population pharmacokinetic (popPK) models by utilizing a systematic stepwise process. In this study, the performance of the AMD tool was assessed using simulated datasets. Ten true models mimicking classical popPK models were created. From each true model, dataset replicates were simulated assuming a typical phase I study design—single and multiple ascending doses with/without dichotomous food effect, with rich PK sampling. For every dataset replicate, the AMD tool automatically built an AMD model utilizing NONMEM for parameter estimation. The AMD models were compared to the true and reference models (true model fitted to simulated datasets) based on their model components, predicted population and individual secondary PK parameters (SP) ($AUC_{0\text{-}24}$, c_{max}, c_{trough}), and model quality metrics (e.g., model convergence, parameter relative standard errors (RSEs), Bayesian Information Criterion (BIC)). The models selected by the AMD tool closely resembled the true models, particularly in terms of distribution and elimination, although differences were observed in absorption and inter-individual variability components. Bias associated with the derived SP was low. In general, discrepancies between AMD and true SP were also observed for reference models and therefore were attributed to the inherent stochasticity in simulations. In summary, the AMD tool was found to be a valuable asset in automating repetitive modeling tasks, yielding reliable PK models in the scenarios assessed. This tool has the potential to save time during early clinical drug development that can be invested in more complex modeling activities within model-informed drug development.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Developing population pharmacokinetic (popPK) models is a complex, often repetitive, and time-consuming task. The Pharmpy Automatic Model Development

(AMD) tool builds full classical popPK models automatically in a stepwise manner. To date, the AMD models have not been evaluated within a simulation framework by comparison with true models used for dataset simulation.

WHAT QUESTION DID THIS STUDY ADDRESS?

Within the context of early clinical drug development, this study investigated how well the AMD models describe simulated datasets compared to the true models.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The AMD models describe datasets well and are similar in structure to the true models. The discrepancies in derived secondary PK parameters are associated with stochasticity in simulations rather than differences in model structures.

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

Leveraging the AMD tool in early drug development can accelerate popPK model-building and free up resources for further modeling tasks, ultimately facilitating decision-making in a model-informed drug development (MIDD) framework.

INTRODUCTION

The use of nonlinear mixed-effects (NLME) modeling to characterize pharmacokinetic (PK) data has an important and well-established place in modern drug development settings.[1,2] It is commonly used for informing decision-making, clinical trial designs, and extrapolation to non-observed scenarios.[3,4]

Building population pharmacokinetic (popPK) models requires interdisciplinary expertise in both pharmaceutical sciences and statistics. However, it is also often a repetitive and time-consuming task, especially at the beginning of the modeling process. To address this problem, several approaches are being explored to accelerate model-building development. One approach is to implement machine learning (ML) algorithms.[5–7] Even though promising, this approach is complex and differs from the conventional modeling procedures. Other less complex non-ML-based tools of varying degrees of automation are also available, including those that solely generate and fit requested models with offering estimation information, as well as tools that optimize model components step-by-step and select the best model.[8–13]

The Pharmpy Automatic Model Development (AMD) tool is a tool that builds full classical popPK models utilizing a systematic stepwise process.[13,14] The AMD tool is implemented in Pharmpy, an open-source software package for pharmacometrics modeling in Python, with the wrapper for R, PharmR.[15,16] The AMD tool consists of different subtools that can be used together in a fully automated workflow or independently, each building a different part of the NLME PK model, such as modelsearch, for structural PK model development; IIVsearch, for inter-individual variability (IIV) model development; RUVsearch, for residual unexplained variability (RUV) model development; and COVsearch, for covariate model development (COV). Each subtool uses dedicated algorithms that create model candidates, fit them (under Pharmpy version 0.86.0, using NONMEM), and finally select the best model based on predefined criteria. The AMD tool builds model components step by step, using the best model from one subtool as input for the next. It only requires a dataset as input and provides details on all candidate models and estimation outputs, offering transparency in the underlying model-building process.

The AMD tool can play a key role in facilitating model-informed drug development (MIDD). However, an assessment of the tool and the respective automatically built models by comparison with true models used for the simulation of datasets is lacking. The aim of this study is to assess the performance of the AMD tool in a simulation framework, mimicking phase I clinical single ascending dose (SAD) and multiple ascending dose (MAD) study designs, with/without a dichotomous food effect, with rich PK sampling.

METHODS

An automated simulation and analysis framework was established for 10 scenarios with varying popPK model components (Figure 1). All scenarios were based on the same study design. For each scenario, the "true model" with known model components and parameter values was utilized to simulate 30 datasets that were used as input for

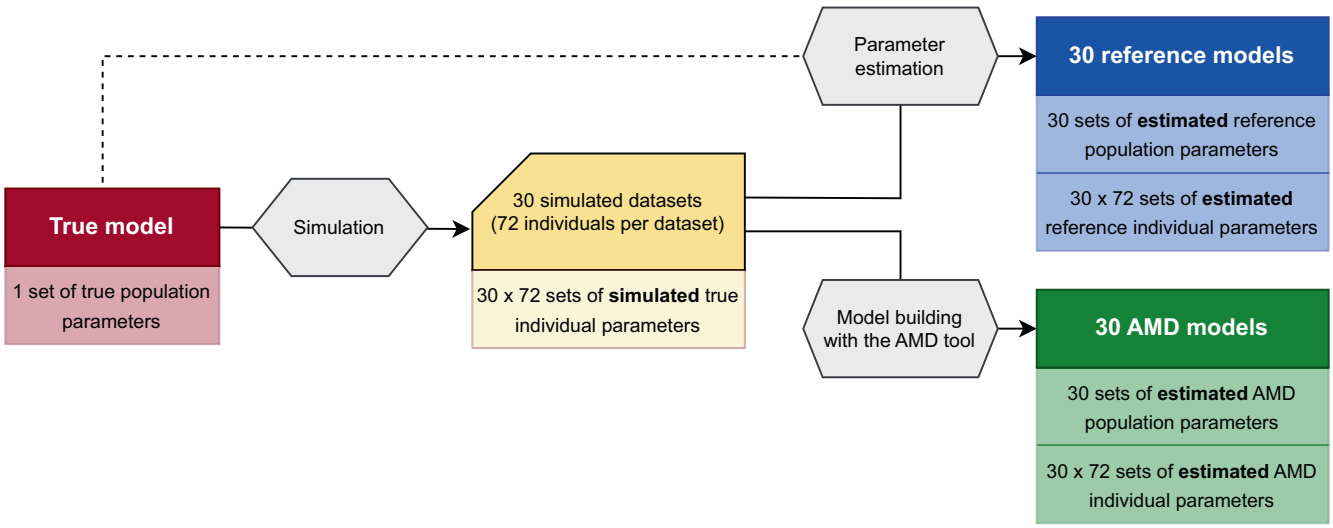

FIGURE 1 Automated simulation and analysis framework for a given scenario. The "true model" with known model components and model parameter values was used to simulate 30 datasets. During simulation, a set of individual model parameters was associated with each individual of each dataset replicate (true individual parameters). Each simulated dataset was used as input for one automatic model development (AMD) tool run. The final model selected by the AMD tool was referred to as the "AMD model" (30 AMD models per scenario, i.e., one for each dataset, each with its own population and individual model parameters estimates). In addition, the true model was fitted to each simulated dataset to account for stochasticity in simulations, resulting in 30 new sets of population and individual parameter estimates under the original model with the true model components, referred to as "reference models." Metrics related to the AMD models were later compared with metrics from the true and reference models. The gray hexagons represent the key workflow steps. Models are depicted as rectangles and datasets as card shapes. The dashed line indicates that the true model code was used for the estimation of the parameters, resulting in reference models. Lighter-color rectangles represent the nature and number of parameters derived from the corresponding elements.

the AMD tool. The model finally selected by the AMD tool was referred to as the "AMD model" (30 AMD models per scenario, i.e., one for each dataset). In addition, the true model was fitted to each simulated dataset to account for stochasticity in simulations, resulting in 30 new sets of parameter estimates under the original model, referred to as "reference models." Metrics related to the AMD models were compared with metrics from the true and reference models. Additional details on the methodology are described in the following sections.

True models

A previously published popPK model was modified to create the true models for the 10 scenarios.[17] The overview of the model components and parameters for different scenarios can be found in Table 1 and the model code in Supporting Information S2. All models had two distribution compartments and differed in (i) absorption model complexity—first-order (FO) absorption with or without delay (three transit compartments), and slow or rapid absorption; (ii) elimination type—FO, Michaelis–Menten (MM), or mixed (MIX-FO-MM); (iii) complexity of IIV structures; (iv) magnitude of proportional RUV; and (v) presence/absence of food effect on absorption. Typical PK profiles and nominal sampling times after the first and the last dose are shown in Supporting Information S1 (Figure S1).

Simulated datasets

The study design for simulation was based on the phase I SAD and MAD clinical studies in a real drug development program for an orally administered new molecular entity (Figure S1a). The SAD study included six dose cohorts (0.25, 0.5, 0.75, 1.0, 1.5, and 2.0 mg) with six subjects per cohort. The MAD study included four dose cohorts (0.5, 1.0, 1.5, and 2.0 mg) with once-daily dosing over 14 days and nine subjects per cohort. For scenarios 8–10, where a food effect on mean absorption time (MAT) was included, SAD cohorts were assumed to be in fasted state and MAD cohorts in fed state. The PK sampling scheme was rich, with 24 samples per subject for the SAD study and 38 samples per subject for the MAD study. Deviation from nominal time in the simulated dosing and PK sampling times was introduced using a truncated normal distribution (mean = nominal time, standard deviation = 2.5 min, min/max = nominal time ± 5 min). The dataset with the above-described characteristics (including actual sampling times) was used as an input for the simulation of 30 replicates per scenario using NONMEM (Supporting Information S2).

TABLE 1 Model components of true (and reference) models for each scenario.

Scenario		Based on scenario	Absorption	Distribution	Elimination	IIV[a]	RUV	Covariates
1	"Base"[b]	–	FO	2 CMT	FO	[MAT]+[CL,VC,QP1,VP1]	PROP	–
2	"Base with simplified IIV"	1	FO	2 CMT	FO	[MAT]+[CL,VC]	PROP	–
3	"Delayed absorption"	1	3 transit CMT[c]+depot	2 CMT	FO	[MAT]+[CL,VC,QP1,VP1]	PROP	–
4	"Delayed and rapid absorption"	3	3 transit CMT[d]+depot[e]	2 CMT	FO	[MAT[f]]+[CL,VC,QP1,VP1]	PROP	–
5	"MM elimination"	2	FO	2 CMT	MM[g]	[CL]+[VC]	PROP	–
6	"MIX-FO-MM elimination"	2, 5	FO	2 CMT	MIX-FO-MM[h]	[CL]+[VC]	PROP	–
7	"Delayed absorption and MIX-FO-MM elimination"	3, 5, 6	3 transit CMT+depot	2 CMT	MIX-FO-MM[h]	[CL]+[VC]	PROP	–
8	"Delayed slow or rapid absorption with food effect"	3, 4[i]	3 transit CMT+depot	2 CMT	FO	[MAT]+[CL,VC,QP1,VP1]	PROP	Food on MAT
9	"Increased RUV"	8	3 transit CMT+depot	2 CMT	FO	[MAT]+[CL,VC,QP1,VP1]	PROP[j]	Food on MAT
10	"Decreased RUV"	8	3 transit CMT+depot	2 CMT	FO	[MAT]+[CL,VC,QP1,VP1]	PROP[k]	Food on MAT

Note: True models are models used for datasets simulation. Reference models are true models fitted to simulated datasets (model parameters were estimated)

Abbreviations: CL, clearance; CLMM, Michaelis–Menten clearance; CMT, compartment; CV, coefficient of variation; FO, first-order kinetics; IIV, inter-individual variability; KM, Michaelis constant; MAT, mean absorption time (MAT = $1/k_a$); MDT, mean delay time (MDT = $N_{transit}/k_{tr}$); MM, Michaelis–Menten kinetics; PROP, proportional; QP1, intercompartmental clearance for the first peripheral compartment; RUV, residual unexplained variability; VC, central apparent volume of distribution; VP1, apparent volume of distribution for the first peripheral compartment.

[a]IIV parameters in [] include off-diagonal elements.

[b]True parameters in "base" scenario and all other scenarios if not indicated differently: CL = 8.87 L/h, VC = 231 L, QP1 = 31.0 L/h, VP1 = 1140 L, MAT = 3.0 h, IIV_CL (CV) = 71.3%, IIV_VC (CV) = 25.9%, IIV_QP1 (CV) = 36.3%, IIV_VP1 (CV) = 37.4%, IIV_MAT (CV) = 53.8%, correlation between IIV_CL and IIV_VC = 0.651, correlation between IIV_CL and IIV_QP1 = 0.518, correlation between IIV_CL and IIV_VP1 = 0.248, correlation between IIV_VC and IIV_QP1 = 0.721, correlation between IIV_VC and IIV_VP1 = 0.350, correlation between IIV_VP1 and IIV_QP1 = 0.550, PROP_RUV (CV) = 26.4%.

[c]MDT = 0.120 h.

[d]MDT = 0.212 h.

[e]MAT = 0.664 h.

[f]IIV_MAT (CV) = 41.9%.

[g]CLMM = 8.87 L/h, KM = 1.00 ng/mL.

[h]CLMM = 5.92 L/h, KM = 1.00 ng/mL, CL = 2.96 L/h.

[i]SAD cohorts (36 patients) exhibiting PK profiles from scenario 4 (fasted state), MAD (36 patients) cohorts exhibiting PK profiles from scenario 3 (fed state).

[j]PROP_RUV (CV) = 50.0%.

[k]PROP_RUV (CV) = 10.0%.

Full adherence to dosing and sampling was assumed, and values below the LOQ were removed after the simulation. Only datasets for which reference models were estimated with ≥2.5 significant digits were kept for the analysis and used for model development with the AMD tool.

AMD tool settings

The AMD tool was run with the following subtools: Modelsearch, IIVsearch, and RUVsearch for all scenarios and, in addition, COVsearch for scenarios 8–10. The AMD subtools were run sequentially in the abovementioned order, where the model selected by one subtool was used as input for the next subtool. Default settings were used (additional information below), except for the initial estimates for the starting model, which were NCA-derived for each scenario based on the typical true profile following 1.0 mg drug administration. Specialized NONMEM ADVAN routines with explicit ordinary differential equations (ODEs) solutions were used, unless nonlinear elimination was present, in which ADVAN13 was utilized and ODEs were solved numerically. The first-order conditional estimation with interaction (FOCEI) was the estimation method used. A brief description of the AMD subtools is given below, and the reader is referred to the tool documentation for additional details.[13,14]

Modelsearch subtool

The default starting point for the model-building process is the creation and fitting of a one-compartment model with FO absorption and elimination, featuring IIV on MAT, clearance (CL), central volume of distribution (VC), assuming log-normal distribution of individual PK parameters, and CL-VC correlation to the input dataset. Afterward, additional structural components are tested. For the absorption, different processes are tested such as FO, zero-order (ZO), and sequential (SEQ-ZO-FO). Additionally, absorption delay models, for example, lag-time or the presence of 1, 3, or 10 transit compartments, with or without a depot compartment (k_a and k_{tr} estimated as two separate parameters or assumed equal), are considered. Absorption is parameterized in terms of MAT (MAT = $1/k_a$) and optional mean delay time (MDT = lag-time; MDT = $N_{tr\text{-}cmt}/k_{tr}$, where $N_{tr\text{-}cmt}$ is the number of transit compartments and k_{tr} the transit rate constant; or MDT = infusion_duration/2). The distribution is modeled using either one or two compartments, and for elimination the AMD tool tests FO, MM, and MIX-FO-MM processes. The default algorithm was used, which creates candidate models by adding one feature in each step in all possible orders and selects among identical model structures the best model before the inclusion of the next feature.[18] During structural model development, the starting IIV structure is preserved and IIV on MDT is added, if this parameter is present. A proportional RUV is used. The selection criterion for the final model within the Modelsearch subtool was the default Bayesian Information Criterion (BIC) type.[14]

IIVsearch subtool

The IIVsearch subtool employs a "brute force" algorithm and a full-block strategy to identify the best IIV model. In the first step, the algorithm adds a full covariance matrix with IIV on all parameters, removes IIV from one parameter at a time, and the best structure is selected. In the second step, the algorithm checks all possibilities of joint distributions between parameters. Individual PK parameters were assumed to be log-normally distributed. The selection criterion was the default BIC type.[14]

RUVsearch subtool

To shorten the time needed for RUV model development, RUVsearch is modeling RUV on the conditional weighted residuals of the fit of the input model to investigate which residual model to select.[19] Models tested included proportional, combined, and power models, IIV on RUV, and multiple time-varying RUV.[13,14] The input model is then updated with the selected RUV model and fitted to verify if the selected RUV model led to a better overall model using the likelihood ratio test (LRT) (at $\alpha = 0.05$). This iterative process is repeated multiple times to verify if the data support additional complexity in the RUV model.

COVsearch subtool

The COVsearch subtool implements the Stepwise Covariate Modeling (SCM) algorithm.[10] In this work, food status was tested as a categorical covariate on all parameters with IIV, one at a time, with $\alpha = 0.05$ and $\alpha = 0.01$ for forward and backward search, respectively.

Metrics used for AMD models evaluation

The AMD models were compared qualitatively and quantitatively to the true and reference models with respect

to model components, secondary PK parameters (SP), and model quality metrics, as detailed in the following sections.

Model components

The proportion of correctly selected and most commonly selected model components per scenario was evaluated for (i) absorption, (ii) distribution, (iii) elimination; (iv) IIV structure; (v) RUV structure, and, for scenarios 8–10, (vi) covariates.

Secondary PK parameters

To evaluate how the PK profiles derived from the AMD models compare with those from the true and the reference models, population and individual SP for the true, reference, and AMD models were obtained from a dense grid of both population-predicted and individual-predicted concentration–time points. Simultaneously with dataset simulation, the true dense predicted PK grids were recorded by requesting IPRED and PRED in the NONMEM output table. Reference and AMD dense PK grids were recorded during model estimation. Calculated SP were $AUC_{0\text{-}24}$ and c_{max} derived from after-the-first-dose data (both SAD and MAD cohorts) and steady-state parameter c_{trough} (MAD cohort).

The relative errors (RE) of AMD- and reference-derived (estimated models) population SP ($\widehat{SP_{pop}}$) compared with the true population SP (SP_{pop}) were calculated for each dth ($d=1, 2, ..., D$; $D=6$) investigated dose of kth replicate ($k=1, 2, ..., K$; $K=30$) as follows:

$$\%RE_{\widehat{SP_{pop,dk}}} = \frac{\widehat{SP_{pop,dk}} - SP_{pop,d}}{SP_{pop,d}} \cdot 100 \quad (1)$$

The RE of individual SP ($\widehat{SP_{ind}}$) for each ith ($i=1, 2, ..., N$; $N=72$) individual of kth replicate ($k=1, 2, ..., K$; $K=30$) were also calculated:

$$\%RE_{\widehat{SP_{ind,ik}}} = \frac{\widehat{SP_{ind,ik}} - SP_{ind,ik}}{SP_{ind,ik}} \cdot 100 \quad (2)$$

where $\widehat{SP}$ can be derived either from AMD model (for RE_{AMD}) or reference model (for RE_{ref}).

The possible deviation of reference SP from the true SP (i.e., RE_{ref}) is a consequence of stochasticity in the simulations given by the limited number of simulated subjects and sampling scheme. Hence, $RE_{AMD} > 0$ is also expected. To account for the impact of stochasticity in RE_{AMD}, the correlation between RE_{AMD} and RE_{ref} expressed as the coefficient of determination (R^2) was calculated for both SP_{pop} and SP_{ind}. Moreover, the impact of different model components selected by the AMD tool (absorption model, elimination model, IIV model, covariate model) on the magnitude of the RE_{AMD} was examined visually.

As an overall measure of accuracy and imprecision in AMD- and reference-derived population-predicted SP (in comparison to the true SP), the bias and the root mean squared error (RMSE) were used, respectively:

$$\%bias_{\widehat{SP_{pop}}} = \frac{1}{D}\frac{1}{K}\sum_{d=1}^{D}\sum_{k=1}^{K}\%RE_{\widehat{SP_{pop,dk}}} \quad (3)$$

$$\%RMSE_{\widehat{SP_{pop}}} = \sqrt{\frac{1}{D}\frac{1}{K}\sum_{d=1}^{D}\sum_{k=1}^{K}\%RE_{\widehat{SP_{pop,dk}}}^{2}} \quad (4)$$

Model quality metrics

The following quality metrics were summarized across replicates for each scenario: the number of (i) converged minimizations, (ii) successful covariance steps, (iii) condition number <1000, (iv) models with all relative standard errors (RSEs) <20% for fixed effect parameters, and <40% for random effect parameters, (v) difference in BIC ($BIC_{difference} = BIC_{AMD} - BIC_{reference}$), and (vi) difference in total number of parameters ($N_{par\_difference} = N_{par\_AMD} - N_{par\_reference}$).

Software and technical implementation

The true models were coded using the R interface to Pharmpy through the package pharmr (0.86.0) and assemblerr (0.1.2) in R (4.2.0).[20] Simulations were performed in NONMEM (7.5.1) using a study design dataset generated in R. For the AMD tool run and model fitting, pharmr was used, which utilized NONMEM. Ten models were estimated simultaneously on a 12-core high-performance computing (HPC) platform. Package qpNCA (1.1.6) was used for SP calculation.[21] All results were generated and analyzed within the validated environment Improve (2.9.4).[22,23]

RESULTS

The AMD tool successfully built models for the 10 simulated scenarios, with 30 AMD models developed per scenario. Replicates for which the starting AMD model failed to converge (3/300 replicates) were not included in the results and were replaced by new ones. In our settings, 3–8 h were needed to run the AMD tool and gather the results for one replicate. In this section, an overview of all results

is presented, with a graphical representation for scenarios 3 ("Delayed absorption") and 6 ("MIX-FO-MM elimination"). Details of the remaining scenarios are presented in Supporting Information S1.

Model components

The model components most frequently selected by the AMD tool for each scenario are shown in Table 2. Detailed information on the selected components for the 30 replicates for scenarios 3 and 6 are illustrated using alluvial plots (Figure 2a,b, respectively). In general, the AMD tool selected models that were similar but not identical to the true models in all its components. The true two-compartment disposition model was identified for all scenarios and replicates. The true elimination was identified for all scenarios, except for a few replicates (9/60) in scenarios with MIX-FO-MM elimination, where MM elimination was selected instead (Figure 2b; Figure S6).

The most frequently selected absorption models did not always match the true absorption. In the scenario with delayed absorption (scenario 3), a 1-transit-compartment model was mostly selected (Figure 2a). In the scenario with delayed and rapid absorption (scenario 4), an absorption model with delay, but one less parameter (1-transit-compartment model with $k_a = k_{tr}$), was identified. In scenarios with nonlinear elimination (scenarios 5–7), the SEQ-ZO-FO absorption model was most frequently selected (Figure 2b). In the remaining scenarios, true absorption was mostly identified.

In all scenarios, a diversity of IIV models was selected across scenarios and replicates (Figure 2; Figures S2–S9). The true proportional RUV model was mostly identified, with a median of 22/30 replicates (range 20–26/30) for all scenarios (Figures S2–S9), except for the "increased RUV" scenario 9, where the power error model was mostly selected (26/30 replicates).

Where applicable (scenarios 8–10), the most commonly selected covariate effect was also the true effect (food effect on MAT). In all replicates that had IIV on MAT, a food effect on MAT was identified.

Secondary PK parameters

The bias in $SP_{AMD,pop}$ was found to be low across scenarios (median and maximum |bias| of 0.59% and 2.48% for $AUC_{0\text{-}24}$, 0.38% and 2.57% for c_{max} and 0.87% and 2.04% for c_{trough}), and the RMSE was found to be moderate (median and maximum of 3.91% and 6.65% for $AUC_{0\text{-}24}$, 4.27% and 8.18% for c_{max} and 7.02% and 11.5% for c_{trough}) (Table 3). The most extreme values of bias and RMSE were associated with increased RUV (scenario 9).

Both AMD- and reference-derived SP_{pop} and SP_{ind} tended to differ similarly from true SP for most scenarios, replicates, and subjects, as shown by the assessment of the correlation between RE_{AMD} and RE_{ref} (Figure 3; Table S1; Figures S11–S20). The median coefficient of determination between $RE_{AMD,ind}$ and $RE_{ref,ind}$ across all scenarios and SP was 0.926. The RE_{AMD} was notably higher than RE_{ref} only for a subset of subjects in 14/300 replicates and was attributed to either oversimplified IIV (11/14), wrong absorption (1/14), or elimination model (2/14).

Model quality metrics

The summary of model quality metrics is shown in Table 4. In FO elimination scenarios (1–4, 8–10), successful minimization was achieved for most models (95% and 96% for AMD and reference models, respectively). The covariance step was obtained for 82% AMD and 74% reference models. When obtained, the condition number was <1000 for all AMD models, but multiple had high RSEs (>20% for fixed effects or >40% for random effects) for at least one parameter (26% and 49% AMD models for fixed and random effects, respectively). In nonlinear elimination scenarios (5–7), both reference and AMD models less frequently achieved successful minimization and estimated covariance step.

The BIC of AMD models was mostly similar to the BIC of reference models (median difference between −3 and 8). Few AMD replicates with several hundred points BIC increase had an oversimplified IIV structure. The difference in the total number of parameters between AMD and reference models ranged between −8 and 9 (mostly random effects parameters).

DISCUSSION

The AMD tool automates popPK model building, aiming to streamline and facilitate a process that involves repetitive steps, minimizing the time needed to develop models. In this simulation-based study, the descriptive performance of models generated by the AMD tool using NONMEM was systematically evaluated across PK scenarios often found in an early clinical drug development context. This multifaceted evaluation included an assessment of AMD models, compared with the true and reference models, in terms of (i) model components, (ii) SP as model-independent quantitative measures of exposure, and (iii) model quality metrics for parameter identifiability, precision, and goodness of fit. In general, the AMD tool

TABLE 2 Model components most commonly selected by the automatic model development (AMD) tool for each scenario.

Scenario		Absorption	Distribution	Elimination	IIV[a]	RUV	Covariates
1	"Base"	FO (19/30)	2 CMT (30/30)	FO (30/30)	[VC] + [CL,QP1,VP1] (8/30)	PROP (20/30)	–
2	"Base with simplified IIV"	FO (17/30)	2 CMT (30/30)	FO (30/30)	[MAT] + [CL,VC] (29/30)	PROP (23/30)	–
3	"Delayed absorption"	1 transit CMT + depot (27/30)	2 CMT (30/30)	FO (30/30)	[VC] + [CL,QP1,VP1] (10/30)	PROP (24/30)	–
4	"Delayed and rapid absorption"	1 transit CMT without depot (30/30)	2 CMT (30/30)	FO (30/30)	[MAT] + [CL,VC,QP1,VP1] (22/30)	PROP (24/30)	–
5	"MM elimination"	1 transit CMT + depot/SEQ-ZO-FO (7/30)	2 CMT (30/30)	MM (30/30)	[CLMM,VC]/ [CLMM,VC] + [MAT] + [MDT] (6/30)	PROP (26/30)	–
6	"MIX-FO-MM elimination"	SEQ-ZO-FO (13/30)	2 CMT (30/30)	MIX-FO-MM (23/30)	[CL] + [VC] (14/30)	PROP (22/30)	–
7	"Delayed absorption and MIX-FO-MM elimination"	SEQ-ZO-FO (16/30)	2 CMT (30/30)	MIX-FO-MM (28/30)	[CL] + [VC] (21/30)	PROP (22/30)	–
8	"Delayed slow or rapid absorption with food effect"	3 transit CMT + depot (18/30)	2 CMT (30/30)	FO (30/30)	[MAT] + [CL,VC,QP1,VP1] (18/30)	PROP (22/30)	Food on MAT (11/30)
9	"Increased RUV"	3 transit CMT + depot (16/30)	2 CMT (30/30)	FO (30/30)	[VC] + [CL, QP1,VP1]/ [MAT] + [CL, QP1,VP1] (8/30)	power (26/30)	Food on MAT (8/30)
10	"Decreased RUV"	1 transit CMT + depot (12/30)	2 CMT (30/30)	FO (30/30)	[MAT]+ [CL,VC,QP1,VP1] (21/30)	PROP (20/30)	Food on MAT (26/30)

Note: The cells in gray indicate that the model component most commonly selected by the AMD tool was identical to the true model component. The exact number of replicates with the corresponding model component is shown in parentheses.

Abbreviations: CL, clearance; CLMM, Michaelis–Menten clearance; CMT, compartment; CV, coefficient of variation; FO, first-order kinetics; IIV, inter-individual variability; MAT, mean absorption time; MDT, mean delay time; MM, Michaelis–Menten kinetics; PROP, proportional; QP1, intercompartmental clearance; RUV, residual unexplained variability; SEQ-ZO-FO, sequential zero-order-first-order kinetics; VC, central apparent volume of distribution; VP1, peripheral apparent volume of distribution.

[a]IIV parameters in [] include off-diagonal elements.

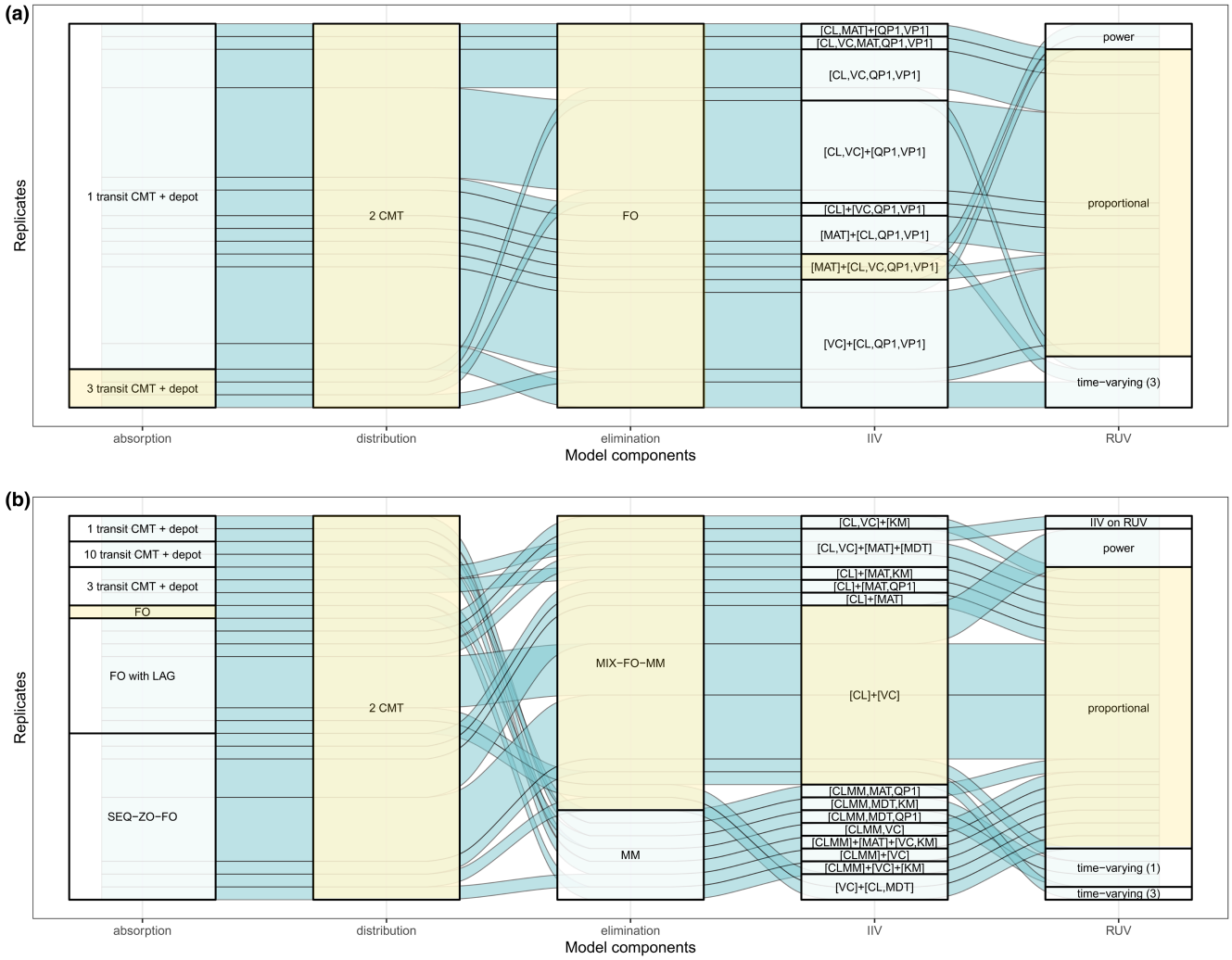

FIGURE 2 Alluvial plots for the components of the final AMD models (models automatically developed by the automatic model development (AMD) tool) for (a) scenario 3 ("Delayed absorption"); (b) scenario 6 ("MIX-FO-MM elimination"). Each longitudinal band represents one AMD model (30 replicates in total). Bands from AMD models that have identical model components are merged together. Each band passes through different transversal rectangles representing different model components selected by the AMD tool. Highlighted yellow boxes represent the true model components; IIV parameters in [] include off-diagonal elements; the number in parentheses next to the "time-varying" RUV component indicates which, out of three time-varying models tested, was selected; IIV, inter-individual variability; RUV, residual unexplained variability; CMT, compartment; FO, first-order kinetics; MM, Michaelis–Menten kinetics; CL, clearance; VC, central apparent volume of distribution; QP1, intercompartmental clearance for the first peripheral compartment; VP1, apparent volume of distribution for the first peripheral compartment; MAT, mean absorption time (MAT = $1/k_a$); MDT, mean delay time (MDT = $N_{transit}/k_{tr}$); CLMM, Michaelis–Menten clearance; KM, Michaelis constant.

selected models that describe the datasets well within the tested space. Considerations related to the results and limitations of the tool are discussed in the following sections.

In most scenarios and replicates, models selected by the AMD tool closely resembled but were not identical to the true models in all its components. Structural model components for distribution and elimination (both linear and nonlinear) were correctly identified in most replicates. However, notable variations in absorption models were observed. In some cases, the absorption model was simplified (e.g., delayed and rapid absorption scenario (scenario 4) with fewer transit compartments (1 vs. 3) and simplified absorption parameters ($k_a = k_{tr}$ vs. $k_a \neq k_{tr}$)), which could be explained by the design with rather short true delay and absorption times (for scenario 4, MDT = 0.2 h, MAT = 0.7 h, time_of_the_first_measurement = 0.5 h). In other cases, the absorption models were complexified in comparison to the true (FO) absorption model (e.g., scenarios 1–2, 5–6). However, estimates of the introduced MDT parameter had a negligible impact on population PK profiles. Finally, when nonlinear elimination was present (scenarios 5–7), the AMD tool had more difficulty in selecting the true absorption model and the SEQ-ZO-FO absorption model was often selected. In all cases, the bias associated with the c_{max} was low (<3%).

TABLE 3 Accuracy and precision in AMD and reference population-predicted secondary PK parameters for each scenario, calculated compared to the true values.

Scenario		Secondary PK parameter	% bias$_{\widehat{SP_{pop}}}$		% RMSE$_{\widehat{SP_{pop}}}$	
			AMD	Reference	AMD	Reference
1	"Base"	$AUC_{0\text{-}24}$	0.5	−1.25	3.96	3.97
		c_{max}	−0.05	−1.37	4.41	4.35
		c_{trough}	1.06	0.81	7.33	7.14
2	"Base with simplified IIV"	$AUC_{0\text{-}24}$	−1.45	−1.58	2.49	2.57
		c_{max}	−2.57	−2.88	4.25	4.39
		c_{trough}	0.89	0.74	6.52	6.53
3	"Delayed absorption"	$AUC_{0\text{-}24}$	−0.17	−1.47	3.85	4.13
		c_{max}	0.33	−1.46	4.29	4.40
		c_{trough}	−0.03	−0.03	7.04	6.82
4	"Delayed and rapid absorption"	$AUC_{0\text{-}24}$	−0.81	−1.41	4.06	3.95
		c_{max}	−0.13	−1.34	3.65	3.34
		c_{trough}	0.11	0.06	6.71	6.66
5	"MM elimination"	$AUC_{0\text{-}24}$	−0.27	−0.36	1.75	1.66
		c_{max}	0.4	−0.23	3.65	2.17
		c_{trough}	−0.99	−1.03	2.88	2.84
6	"MIX-FO-MM elimination"	$AUC_{0\text{-}24}$	−0.68	−0.31	1.78	1.50
		c_{max}	−0.18	−0.20	2.14	2.12
		c_{trough}	−1.88	−0.27	6.99	4.86
7	"Delayed absorption and MIX-FO-MM elimination"	$AUC_{0\text{-}24}$	−0.23	−0.17	1.49	1.45
		c_{max}	−0.35	−0.28	1.80	1.94
		c_{trough}	−0.55	−0.09	5.18	4.87
8	"Delayed slow or rapid absorption with food effect"	$AUC_{0\text{-}24}$	−1.00	−1.46	4.83	4.18
		c_{max}	−1.27	−2.14	5.51	5.08
		c_{trough}	−0.71	0.19	9.00	7.13
9	"Increased RUV"	$AUC_{0\text{-}24}$	2.48	1.05	6.65	4.41
		c_{max}	2.04	0.24	8.18	5.14
		c_{trough}	2.04	4.19	11.5	9.24
10	"Decreased RUV"	$AUC_{0\text{-}24}$	0.24	−0.32	5.00	4.47
		c_{max}	−0.67	−0.94	5.02	4.56
		c_{trough}	0.85	0.40	7.76	7.86

Note: True models are models used for datasets simulation. Reference models are true models fitted to simulated datasets (model parameters were estimated). AMD models are models automatically developed by the automatic model development (AMD) tool.

Abbreviations: AUC, area under the curve; FO, first-order kinetics; IIV, inter-individual variability; MM, Michaelis–Menten kinetics; PK, pharmacokinetic; RMSE, root mean squared error; RUV, residual unexplained variability; SP_{pop}, population secondary PK parameter.

Alternative ways of including IIV in model parameters during structural model development are available but not set as default in the AMD tool. Upon review of the results, a modeler may decide to rerun modelsearch subtool as standalone, using other IIV strategies, to evaluate the impact of more or less extensive IIV structure on the final structural model. Moreover, Chen et al. showed that running the AMD tool with different order of subtools (structural/IIV/RUV, structural/RUV/IIV, and RUV/structural/IIV) yielded generally consistent structural models, but the resulting IIV and RUV models tended to vary, without a clear superiority of a given order.[13]

The highest diversity in AMD model components was observed in IIV structures. Different IIV structures with similar complexity (similar number of parameters) did not affect the SP (Figure 3a). However, oversimplified IIV components resulted in biased SP (Figure S11a, Figure S14a). They were also associated to be the main driver of large differences in BIC between AMD and reference models ($BIC_{AMD} > BIC_{reference}$) found in a few outlying replicates. The oversimplified IIV

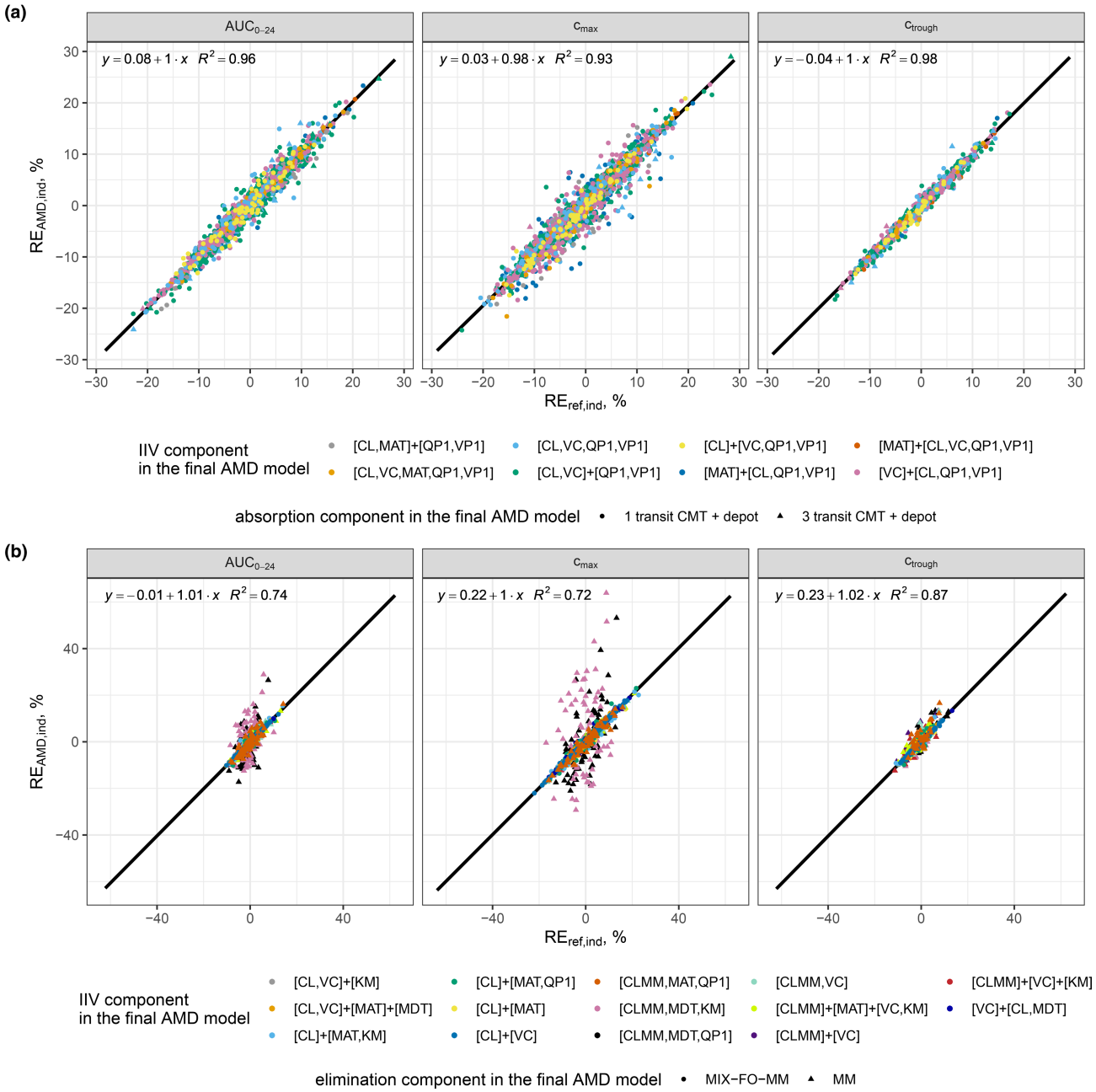

FIGURE 3 Correlation between relative errors (RE) in individual-predicted secondary pharmacokinetic (PK) parameters of estimated models ($RE_{\widehat{SP}_{ind,ik}}$ i.e., $RE_{AMD,ind}$ and $RE_{ref,ind}$) with corresponding regression lines and R^2 values. Each point represents one RE of individual-predicted secondary PK parameter derived from one subject of one replicate. Secondary PK parameters derived from all 72 subjects from all 30 replicates are pooled to calculate the correlation. Stratification by model components allows for identification of the source of discrepancies between $RE_{AMD,ind}$ and $RE_{ref,ind}$ (IIV, or absorption/elimination component); (a) scenario 3 ("Delayed absorption")—points following the regression line show that all replicates, no matter the absorption or IIV model selected by the AMD tool, have accurate individual secondary PK parameters; (b) scenario 6 ("MIX-FO-MM elimination")—data points that are not aligned with the regression line belong to a subset of subjects in two out of seven replicates for which MM elimination instead of true MIX-FO-MM elimination was selected by the AMD tool. IIV parameters in [] include off-diagonal elements; Reference models are true models fitted to simulated datasets (model parameters were estimated). AMD models are models automatically developed by the automatic model development (AMD) tool. RE, relative error; ind, individual-predicted; ref, reference; PK, pharmacokinetic; CL, clearance; VC, central apparent volume of distribution; QP1, intercompartmental clearance for the first peripheral compartment; VP1, apparent volume of distribution for the first peripheral compartment; MAT, mean absorption time ($MAT = 1/k_a$); MDT, mean delay time ($MDT = N_{transit}/k_{tr}$); CLMM, Michaelis–Menten clearance; KM, Michaelis constant; CMT, compartment; IIV inter-individual variability.

TABLE 4 Model quality metrics of AMD and reference models for each scenario.

Scenario		Minimization successful		Successful NONMEM covariance step		Condition number <1000		RSE for all fixed effect parameters <20%		RSE for all random effect parameters <40%		BIC difference (AMD–reference), median [range]	Difference in number of parameters (AMD–reference), median [range]
		AMD	Reference	AMD	Reference	AMD	Reference	AMD	Reference	AMD	Reference		
1	"Base"	29/30	25/30	22/30	13/30	22/22	13/13	15/22	13/13	13/22	0/13	4 [−12, 422]	−4 [−6, 2]
2	"Base with simplified IIV"	29/30	30/30	18/30	25/30	18/18	25 /25	15/18	25/25	10/18	12 /25	0 [−16, 9]	0 [−2, 4]
3	"Delayed absorption"	28/30	28/30	28/30	13/30	28/28	12/13	20/28	10/13	18/28	0/13	−2 [−10, 10]	−4 [−5, 4]
4	"Delayed and rapid absorption"	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30	12/30	9/30	−3 [−11, 697]	−1 [−8, 0]
5	"MM elimination"	21/30	26/30	0/30	11/30	0/0	7/11	0/0	6/11	0/0	8/11	8 [5, 147]	2 [1, 5]
6	"MIX-FO-MM elimination"	23/30	23/30	0/30	7/30	0/0	7/7	0/0	7/7	0/0	6/7	8 [−642, 31]	3 [1, 6]
7	"Delayed absorption and MIX-FO-MM elimination"	18/30	22/30	2/30	0/30	1/2	0/0	2/2	0/0	2/2	0/0	0 [−10, 26]	1 [0, 5]
8	"Delayed slow or rapid absorption with food effect"	29/30	30/30	24/30	29/30	24/24	28/29	18/24	29/29	11/24	3/29	−1 [−19, 14]	0 [−4, 1]
9	"Increased RUV"	26/30	28/30	24/30	16/30	24 /24	16/16	5/24	15/16	14/24	0/16	−2 [−24, 75]	−3 [−5, 1]
10	"Decreased RUV"	29/30	30/30	26/30	30/30	26/26	30/30	25/26	30/30	10/26	12/30	1 [−7, 1774]	0 [−4, 4]

Note: Reference models are true models fitted to simulated datasets (model parameters were estimated). AMD models are models automatically developed by the automatic model development (AMD) tool.

Abbreviations: BIC, Bayesian information criterion; FO, first-order kinetics; IIV, inter-individual variability; MM, Michaelis–Menten kinetics; RSE, relative standard error; RUV, residual unexplained variability.

structures were selected during IIVsearch since models with IIV on correct parameters, but with full covariance matrix, did not converge, for example, due to numerical issues. This issue could be potentially solved by using different estimation algorithms, such as the stochastic approximation expectation maximization (SAEM) method.[24]

The true proportional RUV model was identified in all scenarios and most replicates, except for scenario 9 with increased RUV, where the power model, which provides more flexibility with an additional parameter, was more frequently selected instead.

The COVsearch identified the food effect in scenarios 8–10 in all replicates. However, the parameter associated with the effect depended on which parameters included IIV; if IIV on MAT was present, the true parameter–covariate relationship (food on MAT) was always identified; if IIV on MAT was not present, then the food effect was not tested on MAT, and the covariate effect was found to be associated with one of the parameters with IIV. This behavior illustrates a limitation of an end-to-end automatic model-building approach, where the knowledge of which model will be selected, and therefore which parameters may include unexplained variability is unknown at the start. With default AMD tool settings (v.0.86.0), critical covariates could be missed; nonetheless, this could be prevented. Possible workarounds are to apply the COVsearch subtool in a subsequent independent step after the structural, IIV, or RUV models have been optimized, to include the covariate in the base model, or to define the search space. This way the user has increased flexibility on which parameter–covariate relationships to test, informed by model structure and mechanistic knowledge, which is especially important when multiple, potentially correlated covariates are available and causal interpretation is wanted. The covariate model was not the focus of this study; therefore, more complex scenarios (e.g., multiple covariates) were not explored. Developing an appropriate covariate model is critical for dose adjustments, identifying subpopulation, bridging to different populations or other objectives of the analysis. The currently implemented SCM is a widely used method that is considered suitable for general covariate screening.[10] However, the covariate selection methodology should be guided by the objective of the covariate analysis.[11,12,25–27]

In scenarios involving FO elimination, AMD and reference models had similar model quality metrics. In nonlinear elimination scenarios (5–7), the number of AMD models with minimization and covariance step successful was lower than for the remaining scenarios, mainly due to rounding errors. However, this feature was also observed in reference models, reflecting lower numerical stability within NONMEM when describing more complex PK based on the generated data. When an AMD model component had more structural parameters than the true model, it was noticed that these extra parameters tended to be estimated with low precision, and their values appeared uninformative suggesting that the corresponding parts of the model were not fully supported by the data. Nevertheless, such over-parameterized models were selected based on the lowest BIC, possibly driven by influential individuals. This indicates that model selection based solely on BIC, which was the case in the Phampy tool version tested, may at times lead to over-parameterized models which may cause numerical instability of subsequent analysis steps, for example, IIVsearch for an over-parameterized structural model. Refining model selection strictness criteria, based on criteria such as parameter precision or parameter estimates, can further improve the AMD tool robustness. Moreover, the inclusion of structural elements in a model informed by mechanistic or prior knowledge could help avoid the selection of statistically significant models that are not mechanistically interpretable. However, it must be pointed out that in most cases where AMD structural models were different from the true model, the effect on derived SP was negligible.

The methodology followed by the AMD tool was kept unchanged across and within scenarios. However, it was observed that the selected AMD model components varied for the replicates within scenarios despite the underlying datasets coming from the same true model. It was also observed that reference SP tended to deviate from true parameters as also observed for AMD SP (high R^2 values). The cause can be attributed to the stochasticity associated with simulations used to generate datasets carrying limited information about the true structural and stochastic model components. This is often the case when characterizing the PK of compounds in early clinical development based on data from a limited number of subjects.

In this work, classical PK scenarios with rich PK sampling were examined, mimicking only typical phase I clinical trial designs. In future research, it will be of value to further assess the impact of the study design (e.g., variation in the number of individuals, or use of sparse sampling), alternative combinations of true model components and more complex PK models, the AMD models' predictive performance, as well as identifiability. It is also important to mention that this study covers only the evaluation of AMD models selected by using default tool settings that can be adjusted in multiple ways, for instance, by changing the order of the subtools, or by defining the search space for structural and covariate model, which may influence the final model structure.[13] Furthermore, novel algorithms have been

proposed for the building of certain model components (i.e., IIV or covariate model) and the full NLME models. It would be valuable to compare these algorithms in terms of their performance, time, and computational power needed but also complexity and time required for setting them up.[5–8]

This study showcases the usefulness of the AMD tool in an early drug development setting, where a full popPK model that describes a given dataset adequately can be developed in a short period of time. It was shown that even though components of the AMD models were sometimes different from the true model, both individual and population-predicted SP, which are crucial in a MIDD context, were accurate in most of the replicates. Nevertheless, including mechanistic knowledge through defining the algorithm search space, along with a critical review of all model-building steps is recommended to check for the plausibility and robustness of the selected models. In addition, the automated generation of model code in a standardized manner was found to be particularly useful, as well as the exhaustive permutation of predefined model components. Other than developing models end-to-end, the tool is a valuable resource for model screening when running a popPK analysis on a new dataset, for generating a starting point for more complex models, or for model refinement, as the subtools can be used independently, which increases the applicability. In conclusion, the AMD tool was found to be a valuable asset which yields reliable models in the scenarios assessed, offering the opportunity to advance MIDD.

AUTHOR CONTRIBUTIONS

Z.D., F.S.S., V.C., S.R., E.S., and J.A.A. wrote the manuscript. Z.D., F.S.S., and J.A.A. designed the research. Z.D., F.S.S., V.C., S.R., E.S., and J.A.A. performed the research. Z.D. analyzed the data.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Simon Buatois and Dr. Nicolas Frey for their contributions to the development of the AMD tool in collaboration with Uppsala University, the Uppsala Pharmacometrics Research Group, and Scinteco for their assistance during the technical implementation of the framework described in this manuscript.

FUNDING INFORMATION

This work was funded by F. Hoffmann-La Roche Ltd.

CONFLICT OF INTEREST STATEMENT

F.S.S., V.C., S.R., E.S., and J.A.A. are employees of F. Hoffmann-La Roche Ltd. Z.D. was an employee of F. Hoffmann-La Roche Ltd. at the time of performing the research and currently is a doctoral student at Freie Universität Berlin, enrolled in PharMetrX Graduate Research Training Program. F. Hoffmann-La Roche Ltd. funded the development of the AMD tool.

ORCID

Zrinka Duvnjak https://orcid.org/0009-0004-7080-8985
Franziska Schaedeli Stark https://orcid.org/0009-0002-5692-3448
Sylvie Retout https://orcid.org/0000-0003-4867-8438
Emilie Schindler https://orcid.org/0000-0002-4654-1131
João A. Abrantes https://orcid.org/0000-0002-7752-0229

REFERENCES

1. The Food and Drug Administration website. Population Pharmacokinetics Guidance for Industry. Accessed August 13, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/population-pharmacokinetics
2. European Medicines Agency website. Guideline on reporting the results of population pharmacokinetic analyses. Accessed August 13, 2023. https://www.ema.europa.eu/en/reporting-results-population-pharmacokinetic-analyses-scientific-guideline#current-effective-version-section
3. Milligan PA, Brown MJ, Marchant B, et al. Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin Pharmacol Ther*. 2013;93:502-514.
4. Madabushi R, Seo P, Zhao L, Tegenge M, Zhu H. Model-informed drug development approaches in the lifecycle of drug development and regulatory decision-making. *Pharm Res*. 2022;39:1669-1680.
5. Sale M, Nieforth K, Craig J. Machine Learning Based Model Selection with pyDarwin in Pirana [abstract]. Page 31. 2023; 10417. Accessed August 4, 2023. www.page-meeting.org/?abstract=10417
6. Sibieude E, Khandelwal A, Girard P, Hesthaven JS, Terranova N. Population pharmacokinetic model selection assisted by machine learning. *J Pharmacokinet Pharmacodyn*. 2022;49:257-270.
7. Ismail M, Sale M, Yu Y, et al. Development of a genetic algorithm and NONMEM workbench for automating and improving population pharmacokinetic/pharmacodynamic model selection. *J Pharmacokinet Pharmacodyn*. 2022;49(2):243-256.
8. Prague M, Lavielle M. SAMBA: a novel method for fast automatic model building in nonlinear mixed-effects models. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:161-172.
9. Bodak B, Kuritz K, Kuemmel A. Facilitating structural population PK model building for pharmacometricians and beyond [abstract]. Page 31 2023 10462. Accessed August 4, 2023. www.page-meeting.org/?abstract=10462
10. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. *Pharm Res*. 1998;15:1463-1468.
11. Svensson RJ, Jonsson EN. Efficient and relevant stepwise covariate model building for pharmacometrics. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1210-1222.

12. Ayral G, Si Abdallah J, Magnard C, Chauvin J. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: the COSSAC approach. *CPT Pharmacometrics Syst Pharmacol*. 2021;10(4):318-329.
13. Chen X, Nordgren R, Belin S, et al. A fully automatic tool for development of population pharmacokinetic models. *CPT Pharmacometrics Syst Pharmacol*. 2024. doi: 10.1002/psp4.13222
14. AMD tool website. Accessed November 5, 2023. https://pharmpy.github.io/latest/amd.html
15. pharmpy/pharmr github website. Accessed August 13, 2023. https://github.com/pharmpy/pharmr#readme
16. Nordgren R, Belin S, Yngman G, et al. Pharmpy: a versatile open-source library for pharmacometrics [abstract]. *Page 30* 2022 10096. Accessed August 4, 2023. www.page-meeting.org/?abstract=10096
17. Cosson V, Schaedeli-Stark F, Arab-Alameddine M, et al. Population pharmacokinetic and exposure-dizziness modeling for a metabotropic glutamate receptor subtype 5 negative allosteric modulator in major depressive disorder patients. *Clin Transl Sci*. 2018;11(5):523-531.
18. Hamdan A, Chen X, Belin S, et al. Development of Pharmacokinetic Structural Models—Pharmpy Model Search Too l[abstract]. *Page 30*. 2022 10020. Accessed August 4, 2023. www.page-meeting.org/?abstract=10020
19. Ibrahim M, Nordgren R, Kjellsson MC, Karlsson MO. Model-based residual post-processing for residual model identification. *AAPS J*. 2018;20:1-9.
20. assemblerr website. Accessed March 7, 2023. https://github.com/UUPharmacometrics/assemblerr
21. qpNCA website. Accessed March 7, 2023. https://github.com/qPharmetra/qpNCA
22. Abrantes JA, Smart K, Petryet C, et al. ADaMO: End-to-end automation of Pharmacometric modelling in drug development, from dataset building to output generation [abstract]. *Page 30*. 2022 10051. Accessed August 4, 2023. www.page-meeting.org/?abstract=10051
23. Hackl M, Schwarzenbrunner W, Retout S, Abrantes JA, Flandorfer C. Visualising and integrating automated model development within a validated environment [abstract]. Page 31. 2023 10586. Accessed August 4, 2023. www.page-meeting.org/?abstract=10586
24. Bauer RJ. NONMEM tutorial part II: estimation methods and advanced examples. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(8):538-556.
25. Ribbing J, Nyberg J, Caster O, Jonsson EN. The lasso—a novel method for predictive covariate model building in nonlinear mixed effects models. *J Pharmacokinet Pharmacodyn*. 2007;34(4):485-517.
26. Yngman G, Bjugård Nyberg H, Nyberg J, Jonsson EN, Karlsson MO. An introduction to the full random effects model. *CPT Pharmacometrics Syst Pharmacol*. 2022;11(2):149-160.
27. Sanghavi K, Ribbing J, Rogers JA, et al. Covariate modeling in pharmacometrics: general points for consideration. *CPT Pharmacometrics Syst Pharmacol*. 2024;13(5):710-728.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Duvnjak Z, Schaedeli Stark F, Cosson V, Retout S, Schindler E, Abrantes JA. Simulation-based evaluation of the Pharmpy Automatic Model Development tool for population pharmacokinetic modeling in early clinical drug development. *CPT Pharmacometrics Syst Pharmacol*. 2024;13:1707-1721. doi:10.1002/psp4.13213