

Automated modeling of population pharmacokinetic models using nlmixr2auto: a clinical trial and real-world data study

Zhonghui Huang

zhonghui.huang.20@ucl.ac.uk

University College London

Matthew Fidler

Novartis Pharmaceuticals Corporation

Bern-Thomas Nyang'wa

Médecins Sans Frontières

Joseph F Standing

University College London

Frank Kloprogge

University College London

Research Article

Keywords: population pharmacokinetics, automated modeling, metaheuristic algorithms, R package

Posted Date: July 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10421888/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: Competing interest reported. Matthew Fidler is an employee of Novartis and Bern-Thomas Nyang'wa declares employment at Médecins Sans Frontières (MSF). All other authors declared no competing interests for this work.

Abstract

Automated population pharmacokinetic (PopPK) modelling has emerged as a promising approach to reduce the time and expertise required for model development while enabling exploration of large model search spaces. However, existing studies have primarily evaluated automated methods using information criteria or by comparing them with expert-developed models, providing limited evidence of whether the selected model is truly optimal. A more rigorous assessment requires comparison against exhaustive search, which identifies the best model within a predefined search space. This study tested nlmixr2auto, an end-to-end automated PopPK modeling framework that integrates automated generation of initial parameter estimates, model evaluation, and model selection. The framework was evaluated using 22 real clinical datasets. To establish a ground truth model, exhaustive searches were performed across predefined search spaces containing up to 711 candidate models for intravenous datasets and 1,422 candidate models for oral datasets. The performance of metaheuristic algorithms implemented in nlmixr2auto was then compared with exhaustive and conventional stepwise model building. The results showed that metaheuristic algorithms outperformed the stepwise model building. Genetic algorithms, ant colony optimization, and tabu search successfully recovered the exhaustive-search optimum in 21, 20, and 19 of the 22 datasets, respectively, compared with only 7 datasets for the stepwise model building. In 63.6% (14 out of 22) cases, the metaheuristic algorithms completed the modeling task within 24 hours using a 4-CPU, 2-GB computing environment. These findings suggest that nlmixr2auto offers a transparent and efficient open-source framework with considerable potential to support PopPK modeling.

Introduction

Population pharmacokinetic (PopPK) modeling is a core tool in drug development and clinical practice, supporting dose selection in clinical trials [1] and enabling individualized therapy in routine care, especially for medicines with narrow therapeutic windows and special populations [2]. In this context, the overarching goal is to deliver the right medicine to the right patient, at the right time, with the right target and dose [3]. Conventional PopPK modeling remains a specialist task that demands substantial time and expertise. To reduce this burden, fully automated modeling approaches have emerged, particularly within the nonlinear mixed-effects (NLME) framework. These methods operate within predefined search spaces that encompass choices for structural, statistical, and covariate models. Recent examples include the automated modeling platforms pyDarwin and PharmPy [4–6], which have been explored in real-world data.

In Li et al.'s study [4], five pyDarwin search/ML algorithms were evaluated against an exhaustive search over 1.57 million candidate models. With two-bit local search, the methods recovered the ground-truth optimum identified by the exhaustive search on a single dataset. Richardson and colleagues [6] used pyDarwin's Bayesian optimization with local exhaustive search across five datasets, identifying structural and statistical base models that matched or outperformed expert-built models while evaluating fewer than 2.6% of candidate models. Yet, in the absence of a full exhaustive search, it is unclear whether the identified solutions represent the global optimum. Chen and colleagues [5] evaluated ten real datasets using PharmPy, which implemented a stepwise exhaustive search and obtained models from the automated tool that generally achieved lower BIC values than the expert-built models. They also tested different execution orders for the structural, IIV, and residual unexplained variability (RUV) modules, and structural consistency was observed across 9 of 10 datasets. However, the choice of IIV and RUV terms was strongly influenced by the order in which the modules were executed. Beyond the dataset-, evaluation-, and algorithm-specific limitations discussed above, if initialization strategies and stepwise decisions are not transparently recorded in an external, reviewable log, the workflow becomes difficult to interpret. Regulatory guidance requires that model development be documented to enable secondary evaluation, with clear justification for each model selection [7].

To address current limitations, this paper proposes an end-to-end automated modeling framework that optimizes the entire workflow from data input to model output. Unlike existing tools, the framework integrates automated generation

of initial estimates, fitness evaluation, and model optimization/selection. In parallel, an open-source R package *nlmixr2auto* is now available on CRAN to support the growth of community-supported tools, advancing transparency and accessibility in pharmacometrics (PK) modeling. Overall, we aimed to (1) test the automated framework *nlmixr2auto*, which combines automation of initial estimates, fitness function evaluation, and search algorithms into a complete workflow from data input to model output, and verify its feasibility on diverse clinical datasets (2) establish a comparative evaluation framework in which *nlmixr2auto* is compared with exhaustive search taken as the ground truth and further evaluated relative to stepwise methods across fitting performance, robustness, and computational efficiency; (3) provide an open-source implementation as an R package, thereby reducing reliance on commercial software, enhancing accessibility, and promoting objectivity and transparency in the model development.

Methods

Overview of *nlmixr2auto*

The *nlmixr2auto* package proposed in this study consisted of two core modules, *nlmixr2autoinit* and *nlmixr2auto*, which together form an end-to-end automated workflow for PopPK model development. The overall process (see Fig. 1) was divided into the following stages:

Initial parameter computation (*nlmixr2autoinit*). For fundamental pharmacokinetic parameters (CL , V_c , K_a), candidate values were derived by combining the adaptive one-point method, noncompartmental analysis (combined with the Wagner-Nelson method), and graphical methods. The best set of parameter estimates was selected based on the relative root mean squared error (rRMSE). For model-specific parameters in more complex structures, a parameter sweeping strategy was employed to search for plausible values within a predefined candidate range.

Model search and selection (*nlmixr2auto*). Following the determination of stable initial parameter estimates, *nlmixr2auto* explored and optimized the candidate model space using multiple algorithms, including GA, ACO, TS, and a stepwise model building algorithm. For its search space, structural models involved one-, two-, or three-compartment models, with either linear or Michaelis-Menten elimination. IIV was considered on parameters including CL , V_c , V_p , Q , K_m , and K_a (for oral administration), with options to model them as independent or correlated random effects. The RUV models consisted of additive, proportional, and combined error models.

Model assessment was based on a unified fitness function, primarily the BIC, complemented by RSE, shrinkage, the magnitudes of Ω (IIV variance), σ (RUV variance), and correlation penalties.

Fitness function

The fitness function explored from [8] was defined as the Bayesian Information Criterion (BIC) with penalty terms, including penalties for typical parameter values, RSE, shrinkage, IIV variance (ω), residual error models, and correlations among IIV terms that fell outside the ranges:

$$\begin{aligned}
Fitness = BIC &+ \sum_{i=1}^{n_{\theta}} (10 \cdot 1_{\{20\% < RSE_i \leq 40\%\}} + 10000 \cdot 1_{\{RSE_i > 40\%\}}) \\
&+ \sum_{j=1}^{n_{\omega}} (10 \cdot 1_{\{30\% < Shrinkage_j \leq 50\%\}} + 10000 \cdot 1_{\{Shrinkage_j > 50\%\}}) \\
&+ \sum_{j=1}^{n_{\omega}} (10000 \cdot 1_{\{BSV_j < 10\%\}}) + (10 \cdot 1_{\{0 < \sigma_{add} < 0.001\}} + 10 \cdot 1_{\{0 < \sigma_{prop} < 0.001\}}) \\
&+ \sum_{i < j} (10 \cdot 1_{\{0.05 \leq \rho_k < 0.1 \vee 0.8 < \rho_k \leq 0.95\}}) \\
&+ \sum_{i=1}^{n_{\theta}} 10000 \cdot 1_{\{\theta_i > \theta_i^{bound}\}} + 10000 \cdot 1_{\{covariance\ step\ failed\}}
\end{aligned}$$

where $1_{\{\cdot\}}$ denotes the indicator function (1 if the condition holds, 0 otherwise), RSE_i is the relative standard error of θ_i , $Shrinkage_j$ and BSV_j denote the shrinkage and between-subject variability (CV%) of the j_{th} IIV parameter, respectively, σ_{add} and σ_{prop} are additive and proportional residual error terms; ρ_k is the k th correlation estimate between two IIV parameters; and θ_i^{bound} is the pre-specified boundary value for θ_i

Search algorithms

The algorithms applied were ACO, GA, TS, and a stepwise algorithm. The methods for constructing solutions, detailed procedures, and designs of each algorithm have been described in the study [9]. Below is an overview of the core processes and key parameters of each algorithm. The detailed parameter settings of these algorithms are presented in Table 1.

Stepwise algorithm. The stepwise algorithm began with a base model including IIV on CL and a combined error structure. Model development proceeded through four modules comprising structural, IIV, correlation, and RUV models. At each stage, model performance was evaluated dynamically. The structural model stage focused on parameter estimates and their RSEs. During IIV model exploration, variance estimates and shrinkage were added to the previous criteria. In the IIV correlation stage, correlation estimates were evaluated, while in the residual error stage, error metrics were integrated. The best model from each module was used as the starting point for the next, until all modules had been completed.

Genetic algorithm. GA employed a binary encoding scheme to represent candidate PopPK models, mapping ten model components to twelve gene bits. For oral models, an additional bit was used to encode the presence or absence of IIV on K_a . Candidate models were evaluated through a predefined fitness function. Evolution in the GA was implemented through tournament selection, crossover, mutation, and 1-bit local neighborhood search, with an elitism strategy ensuring retention of the best model in each generation. The process was terminated after a fixed number of generations.

Ant colony optimization. ACO applied a simplified ant-colony framework in which each path through a construction graph corresponded to a PopPK model. Nodes encoded model features covering the four modules, forming a construction graph with 10 nodes and 22 edges. For oral models, an additional node representing IIV on K_a and two corresponding edges were included, resulting in 11 nodes and 24 edges. Candidate models were generated by decoding ant paths, followed by parameter estimation and fitness evaluation. Pheromone trails were updated based on the

historical ranking of candidate solutions, with the ranking position determining the pheromone increment assigned to each path. After each iteration, a fixed proportion of pheromone on all edges was evaporated to avoid unlimited accumulation. An elitism strategy retained the best-performing model at each iteration. The algorithm terminated after a predefined number of iterations.

Tabu search. TS began with an initial two-compartment model with IIV included only on clearance and a combined residual error structure and explored candidates through local neighborhood search. The tabu list stored recently executed moves (model component modifications) together with their tabu tenure, while the aspiration criterion was not applied. When no improving solution was found in a given iteration, a perturbation mechanism was applied by flipping two features in the model encoding to generate a new candidate solution. At each iteration, the best candidate from the neighborhood was selected as the new current solution. The algorithm continued until the maximum number of iterations was reached.

Table 1
Final parameter configurations for GA, ACO, and TS

Algorithm	Final parameter configuration
Genetic Algorithm (GA)	Population size: 20 Crossover probability: 0.7 Mutation probability: 0.1 Maximum iterations: 20 Local search: every 3 generations Elite preservation: best solution retained
Ant Colony Optimization (ACO)	Number of ants: 15 Maximum iterations: 20 Pheromone importance (α): 1 Evaporation coefficient (ρ): 0.5 Initial pheromone level (ϕ): 2 Elitism: retain one best solution per iteration Minimum path probability ($p_{\min}$): 0.2
Tabu Search (TS)	Maximum iterations: 20 Neighborhood definition: 1-bit flip Tabu tenure: 3 Aspiration criterion: not applied Perturbation mechanism: random flip of two features

Algorithm evaluation

The evaluation of algorithm performance was conducted from two perspectives: accuracy and computational efficiency. For accuracy, five dimensions were assessed including the structural model, the IIV model, the residual error model, correlations among IIV parameters, and overall consistency. Retrieval rate was defined as the proportion of components

correctly recovered relative to exhaustive-search solutions across all datasets, thereby capturing both agreement with exhaustive solutions and robustness of model recovery. For computational efficiency, performance was quantified by runtime and the number of unique models visited during execution for each case or dataset, with runtime reported as the cumulative time under a 4-core CPU allocation.

Data

Real-life data comprises four components: software-provided real-world example datasets, literature-derived datasets reconstructed from published studies, literature datasets obtained directly from publications, and datasets from a single clinical trial.

- Software-provided sample datasets. These included *theo_md* (theophylline, multiple doses) and *pheno_sd* (phenobarbital, single dose) from *nlmixr2data*, as well as the tobramycin dataset distributed with Monolix [10].
- Literature-derived datasets. A total of 1,350 papers were randomly sampled from 63,383 PK-related publications published between 1969 and 2023, and 10 eligible studies were subsequently retained. Pharmacokinetic information reported in these studies, together with the corresponding concentration-time data, was extracted to reconstruct ten real-life PopPK datasets.
- Literature-provided datasets. Using the PKPDAI keyword search term “NONMEM”, 2,750 publications were retrieved. Seven PopPK datasets were obtained from studies whose appendices provided complete dosing information and clearly defined sampling schedules. Studies with mixed dosing regimens or joint parent-metabolite models were excluded.
- Clinical trial datasets. Three datasets including bedaquiline, clofazimine, and pretomanid, were obtained from a single tuberculosis (TB) clinical trial [11, 12].

Therefore, this study collected 22 real-life clinical pharmacokinetic datasets. Among these, 10 datasets involved intravenous administration, while 12 datasets involved oral administration. With respect to sample size, 4 datasets included more than 100 subjects (imipenem, paclitaxel, vancomycin, and relebactam), 4 datasets involved 50–100 subjects (bedaquiline, pretomanid, phenobarbital, and linezolid), and the remaining 14 datasets each involved fewer than 50 subjects. Regarding sampling design, 13 datasets employed intensive sampling schemes (covering a complete dosing interval or longer than 24 h), whereas the other 9 relied on sparse sampling strategies or opportunistic plasma collection. These datasets covered diverse populations: 8 involved healthy volunteers, 9 involved adult patients, 3 involved pediatric patients, while 2 datasets included both healthy volunteers and patients. The information for these 22 real-life PopPK datasets, including the number of subjects and observations, route of administration, dosing regimen, and sampling schedule, was summarized in Table 2.

Table 2
Study characteristics of 22 real clinical datasets

Drug	Route	N (n)	Population	Regimen & Sampling
Aprindine [13]	Oral	21 (251)	Adult Patients	MD; intensive sampling (0.5–48 h post-dose)
Bedaquiline [11]	Oral	94 (952)	Adult Patients	MD; Day 1 (pre-dose, 2, 23h), Week 8 (pre-dose, 6.5, 23h), Weeks 12–24, 32,72 (pre-dose)
Cefaclor [14]	Oral	20 (199)	HV	MD; intensive sampling (0.5–76 h)
Ceftriaxone [15]	Infusion	20 (64)	HV	SD; 1, 12, 24 h post-infusion
Cephalexin [14]	Oral	20 (264)	HV	MD; intensive sampling (0.5–76 h)
Clofazimine [16]	Oral	30 (286)	Adult Patients	MD; Day 1 (pre-dose, 2, 23h), Week 8 (pre-dose, 6.5, 23h), Weeks 12–24, 32,72 (pre-dose)
Diazepam [17, 18]	Injection	14 (149)	HV	SD; intensive sampling (pre-dose to 48 h post-dose)
Fluorouracil [19, 20]	Injection	13 (66)	Adult Patients	SD; intensive sampling (5–120 min)
Imipenem [21]	Infusion	815 (4454)	HV + patients	MD; intensive PK sampling (0–24 h post-dose) or sparse PK sampling
Linezolid (1002789)[22]	Oral	48 (196)	Children patients (4.6 years (0.5–15.3))	MD; Sampling at pre-dose, 1, 2, 4, 8, 6/11 h; or pre-dose, 1, 4, 10 h (children)
Linezolid (844567) [23]	Infusion	83 (127)	Adult Patients	MD, opportunistic sampling
Oxprenolol (i.v.) [24]	Infusion	6 (102)	HV	SD; intensive sampling (10 min–8 h)
Oxprenolol (oral)	Oral	6 (203)	HV	SD; intensive sampling (0.25–8 h)
Paclitaxel [25]	Oral	197 (4322)	Adult Patients	MD; intensive sampling (pre-dose to 168–337 h, about 14–41 samples/subject)
Phenobarbital [26]	Oral	59 (155)	Preterm infant patients (8.2 days (1–16))	MD; samples from routine clinical care
Pindolol [27]	Oral	12 (104)	HV	SD 5mg; intensive sampling (0.5–24 h)
Polymyxin B [28]	Infusion	22 (64)	Adult Patients	MD; sampled at troughs, peaks, and 6–8 h post-infusion
Pretomanid [29]	Oral	94 (952)	Adult Patients	MD; Day 1 (pre-dose, 2, 23h), Week 8 (pre-dose, 6.5, 23h), Weeks 12–24, 32,72 (pre-dose)
Relebactam [21]	Infusion	649 (4814)	Adult HV + patients	MD; intensive PK sampling (0–24 h post-dose) or sparse PK sampling

Note: N = number of subjects; n = number of samples. SD = single dose; MD= multiple doses; HV=healthy volunteers. The theophylline dataset was based on the day 1 theophylline data [30], with concentrations for days 2–7 simulated under a once-daily regimen. For the three pediatric datasets, age was reported as the median (range) for the linezolid (1002789) and vancomycin datasets, and for the phenobarbital dataset the time from birth to sampling for the last phenobarbital concentration was reported as mean (range).

Drug	Route	N (n)	Population	Regimen & Sampling
Theophylline [30]	Oral	12 (264)	HV	MD; intensive sampling 0–24 h post-dose
Tobramycin [31]	Short infusion	97 (322)	Adult Patients	MD; sparse sampling at varied time points
Vancomycin [32]	Infusion	207 (900)	Neonate patients (2.3 weeks (0–16.4))	MD; peaks collected 1 h after end of infusion
Note: N = number of subjects; n = number of samples. SD = single dose; MD= multiple doses; HV=healthy volunteers. The theophylline dataset was based on the day 1 theophylline data [30], with concentrations for days 2–7 simulated under a once-daily regimen. For the three pediatric datasets, age was reported as the median (range) for the linezolid (1002789) and vancomycin datasets, and for the phenobarbital dataset the time from birth to sampling for the last phenobarbital concentration was reported as mean (range).				

Software

All analyses were performed in R [33], and model parameter estimation was carried out using the *nlmixr2* package [34], with parameter initialization by the *nlmixr2autoinit* developed previously [35]. The automated modeling framework *nlmixr2auto* introduced in this paper is provided as an open-source R package [36]. All model runs were executed on the UCL Myriad High Performance Computing Facility (Myriad@UCL), with each job allocated 4 CPU cores and 2 GB of memory.

Results

Algorithm comparison

Table 3 shows comparative performance across 22 datasets. GA, ACO, and tabu all demonstrated substantially superior performance relative to the stepwise method. GA produced results identical to those obtained via exhaustive search across 21 datasets, with deviation observed only in the fluorouracil case, yielding an overall concordance rate of 95.5%. ACO successfully identified the correct model in 20 of 22 datasets, with discrepancies confined to the clofazimine and paclitaxel datasets, yielding an overall concordance rate of 90.9%. Tabu matched the ground truth in 19 datasets but diverged in the fluorouracil, paclitaxel, and relebactam datasets, corresponding to an overall concordance rate of 86.4%. In sharp contrast, the stepwise method attained results consistent with exhaustive search in only 7 datasets, with an overall concordance rate of 31.8%.

Further analysis of the consistency of model components showed that the structural model retrieval rates for GA, ACO, and Tabu all exceeded 90% (95.5%, 90.9%, and 90.9%, respectively). In contrast, the stepwise method performed considerably worse, with a structural model retrieval rate of only 68.2% (15/22) and the lowest IIV retrieval rate, matching the exhaustive search in just 8 datasets (36.4%). Its retrieval rates were 68.2% (15/22) for the IIV correlation and 77.3% (17/22) for the RUV model.

Table 3

Final models for all drugs obtained using ACO, GA, tabu, and stepwise algorithms, compared with exhaustive search

Drug	Selection method	Parameter	Fitness	Structural model	IIV model	IIV correlation	RUV model
aprindine	Exhaustive search	6	370	oral_1cmpt	CLKA	No	additive
aprindine	ACO	6	370	oral_1cmpt	CLKA	No	additive
aprindine	GA	6	370	oral_1cmpt	CLKA	No	additive
aprindine	Tabu	6	370	oral_1cmpt	CLKA	No	additive
aprindine	Stepwise	7	10366	oral_1cmpt	KAVmax	No	additive
bedaquiline	Exhaustive search	14	14244	oral_2cmpt	CLVPQKA	Yes	combined
bedaquiline	ACO	14	14244	oral_2cmpt	CLVPQKA	Yes	combined
bedaquiline	GA	14	14244	oral_2cmpt	CLVPQKA	Yes	combined
bedaquiline	Tabu	14	14244	oral_2cmpt	CLVPQKA	Yes	combined
bedaquiline	Stepwise	10	14685	oral_3cmpt	CL	No	combined
cefaclor	Exhaustive search	7	1143	oral_1cmpt	CLVC	No	combined
cefaclor	ACO	7	1143	oral_1cmpt	CLVC	No	combined
cefaclor	GA	7	1143	oral_1cmpt	CLVC	No	combined
cefaclor	Tabu	7	1143	oral_1cmpt	CLVC	No	combined
cefaclor	SF	7	11137	oral_1cmpt	CLKA	No	combined
ceftriaxone	Exhaustive search	6	726	iv_1cmpt	CLVC	Yes	additive
ceftriaxone	ACO	6	726	iv_1cmpt	CLVC	Yes	additive
ceftriaxone	GA	6	726	iv_1cmpt	CLVC	Yes	additive
ceftriaxone	Tabu	6	726	iv_1cmpt	CLVC	Yes	additive
ceftriaxone	Stepwise	4	761	iv_1cmpt	CL	No	additive
cephalexin	Exhaustive search	6	11118	oral_1cmpt	CLVC	No	proportional
cephalexin	ACO	6	11118	oral_1cmpt	CLVC	No	proportional
cephalexin	GA	6	11118	oral_1cmpt	CLVC	No	proportional
cephalexin	Tabu	6	11118	oral_1cmpt	CLVC	No	proportional
cephalexin	Stepwise	6	11122	oral_1cmpt	CLKA	No	proportional
clofazimine	Exhaustive search	15	3356	oral_3cmpt	CLVCQ	Yes	combined
Note: 1cmpt = one-compartment model, 2cmpt = two-compartment model, 3cmpt = three-compartment model, additive = additive residual error model, proportional = proportional residual error model, combined = additive and proportional additive residual error model.							

Drug	Selection method	Parameter	Fitness	Structural model	IIV model	IIV correlation	RUV model
clofazimine	ACO	11	3494	oral_2cmpt	VCQVmax	Yes	additive
clofazimine	GA	15	3356	oral_3cmpt	CLVCQ	Yes	combined
clofazimine	Tabu	15	3356	oral_3cmpt	CLVCQ	Yes	combined
clofazimine	Stepwise	7	13455	oral_1cmpt	CLVC	No	combined
diazepam	Exhaustive search	9	-412	iv_2cmpt	CLVCVPQ	No	proportional
diazepam	ACO	9	-412	iv_2cmpt	CLVCVPQ	No	proportional
diazepam	GA	9	-412	iv_2cmpt	CLVCVPQ	No	proportional
diazepam	Tabu	9	-412	iv_2cmpt	CLVCVPQ	No	proportional
diazepam	Stepwise	6	-266	iv_1cmpt	CLVC	No	combined
fluorouracil	Exhaustive search	11	30591	iv_2cmpt	CLVPQ	Yes	proportional
fluorouracil	ACO	11	30591	iv_2cmpt	CLVPQ	Yes	proportional
fluorouracil	GA	9	30726	iv_3cmpt	CLVC	No	proportional
fluorouracil	Tabu	7	30750	iv_2cmpt	CL	No	combined
fluorouracil	Stepwise	7	30750	iv_2cmpt	CL	No	combined
imipenem	Exhaustive search	11	89641	iv_2cmpt	CLVCVP	Yes	proportional
imipenem	ACO	11	89641	iv_2cmpt	CLVCVP	Yes	proportional
imipenem	GA	11	89641	iv_2cmpt	CLVCVP	Yes	proportional
imipenem	Tabu	11	89641	iv_2cmpt	CLVCVP	Yes	proportional
imipenem	Stepwise	6	92883	iv_1cmpt	CLVC	Yes	proportional
linezolid	Exhaustive search	7	1096	oral_1cmpt	CLVCKA	No	additive
linezolid	ACO	7	1096	oral_1cmpt	CLVCKA	No	additive
linezolid	GA	7	1096	oral_1cmpt	CLVCKA	No	additive
linezolid	Tabu	7	1096	oral_1cmpt	CLVCKA	No	additive
linezolid	Stepwise	7	11057	oral_1cmpt	CLVC	Yes	proportional
linezolid_844567	Exhaustive search	4	655	iv_1cmpt	CL	No	additive
linezolid_844567	ACO	4	655	iv_1cmpt	CL	No	additive
linezolid_844567	GA	4	655	iv_1cmpt	CL	No	additive
Note: 1cmpt = one-compartment model, 2cmpt = two-compartment model, 3cmpt = three-compartment model, additive = additive residual error model, proportional = proportional residual error model, combined = additive and proportional additive residual error model.							

Drug	Selection method	Parameter	Fitness	Structural model	IIV model	IIV correlation	RUV model
linezolid_844567	Tabu	4	655	iv_1cmpt	CL	No	additive
linezolid_844567	Stepwise	4	655	iv_1cmpt	CL	No	additive
oxprenolol	Exhaustive search	6	1298	iv_1cmpt	CLVC	Yes	proportional
oxprenolol	ACO	6	1298	iv_1cmpt	CLVC	Yes	proportional
oxprenolol	GA	6	1298	iv_1cmpt	CLVC	Yes	proportional
oxprenolol	Tabu	6	1298	iv_1cmpt	CLVC	Yes	proportional
oxprenolol	Stepwise	6	1298	iv_1cmpt	CLVC	Yes	proportional
oxprenolol_oral	Exhaustive search	5	2312	oral_1cmpt	CL	No	proportional
oxprenolol_oral	ACO	5	2312	oral_1cmpt	CL	No	proportional
oxprenolol_oral	GA	5	2312	oral_1cmpt	CL	No	proportional
oxprenolol_oral	Tabu	5	2312	oral_1cmpt	CL	No	proportional
oxprenolol_oral	Stepwise	5	2312	oral_1cmpt	CL	No	proportional
paclitaxel	Exhaustive search	13	55532	oral_3cmpt	CLVCQQ2	No	combined
paclitaxel	ACO	10	55724	oral_2cmpt	CLVCVPQ	No	proportional
paclitaxel	GA	13	55532	oral_3cmpt	CLVCQQ2	No	combined
paclitaxel	Tabu	10	55724	oral_2cmpt	CLVCVPQ	No	proportional
paclitaxel	Stepwise	9	58725	oral_1cmpt	CLVCKA	Yes	combined
pheno_sd	Exhaustive search	6	1015	iv_1cmpt	CLVC	Yes	additive
pheno_sd	ACO	6	1015	iv_1cmpt	CLVC	Yes	additive
pheno_sd	GA	6	1015	iv_1cmpt	CLVC	Yes	additive
pheno_sd	Tabu	6	1015	iv_1cmpt	CLVC	Yes	additive
pheno_sd	Stepwise	6	1015	iv_1cmpt	CLVC	Yes	additive
pindolol	Exhaustive search	7	691	oral_1cmpt	CLVC	Yes	proportional
pindolol	ACO	7	691	oral_1cmpt	CLVC	Yes	proportional
pindolol	GA	7	691	oral_1cmpt	CLVC	Yes	proportional
pindolol	Tabu	7	691	oral_1cmpt	CLVC	Yes	proportional
pindolol	Stepwise	8	10677	oral_1cmpt	CLVCKA	No	combined
Note: 1cmpt = one-compartment model, 2cmpt = two-compartment model, 3cmpt = three-compartment model, additive = additive residual error model, proportional = proportional residual error model, combined = additive and proportional additive residual error model.							

Drug	Selection method	Parameter	Fitness	Structural model	IIV model	IIV correlation	RUV model
polymyxinB	Exhaustive search	4	231	iv_1cmpt	CL	No	proportional
polymyxinB	ACO	4	231	iv_1cmpt	CL	No	proportional
polymyxinB	GA	4	231	iv_1cmpt	CL	No	proportional
polymyxinB	Tabu	4	231	iv_1cmpt	CL	No	proportional
polymyxinB	Stepwise	6	259	iv_2cmpt	CL	No	additive
pretomanid	Exhaustive search	6	13285	oral_1cmpt	CL	No	combined
pretomanid	ACO	6	13285	oral_1cmpt	CL	No	combined
pretomanid	GA	6	13285	oral_1cmpt	CL	No	combined
pretomanid	Tabu	6	13285	oral_1cmpt	CL	No	combined
pretomanid	Stepwise	6	13285	oral_1cmpt	CL	No	combined
relebactam	Exhaustive search	17	92835	iv_3cmpt	CLVCVPVP2	Yes	proportional
relebactam	ACO	17	92835	iv_3cmpt	CLVCVPVP2	Yes	proportional
relebactam	GA	17	92835	iv_3cmpt	CLVCVPVP2	Yes	proportional
relebactam	Tabu	11	92848	iv_2cmpt	CLVCVP	Yes	proportional
relebactam	Stepwise	8	93373	iv_2cmpt	CLVC	Yes	proportional
theo_md	Exhaustive search	6	918	oral_1cmpt	CLVC	No	additive
theo_md	ACO	6	918	oral_1cmpt	CLVC	No	additive
theo_md	GA	6	918	oral_1cmpt	CLVC	No	additive
theo_md	Tabu	6	918	oral_1cmpt	CLVC	No	additive
theo_md	Stepwise	5	936	oral_1cmpt	CL	No	additive
tobramycin	Exhaustive search	5	807	iv_1cmpt	CL	No	combined
tobramycin	ACO	5	807	iv_1cmpt	CL	No	combined
tobramycin	GA	5	807	iv_1cmpt	CL	No	combined
tobramycin	Tabu	5	807	iv_1cmpt	CL	No	combined
tobramycin	Stepwise	5	807	iv_1cmpt	CL	No	combined
vancomycin	Exhaustive search	5	7336	iv_1cmpt	CL	No	combined
vancomycin	ACO	5	7336	iv_1cmpt	CL	No	combined
Note: 1cmpt = one-compartment model, 2cmpt = two-compartment model, 3cmpt = three-compartment model, additive = additive residual error model, proportional = proportional residual error model, combined = additive and proportional additive residual error model.							

Drug	Selection method	Parameter	Fitness	Structural model	IIV model	IIV correlation	RUV model
vancomycin	GA	5	7336	iv_1cmpt	CL	No	combined
vancomycin	Tabu	5	7336	iv_1cmpt	CL	No	combined
vancomycin	Stepwise	5	7336	iv_1cmpt	CL	No	combined
Note: 1cmpt = one-compartment model, 2cmpt = two-compartment model, 3cmpt = three-compartment model, additive = additive residual error model, proportional = proportional residual error model, combined = additive and proportional additive residual error model.							

Figure 2 shows the convergence of fitness values for three metaheuristic algorithms (ACO, GA, and TS) in real-world applications. For the imipenem dataset, all three algorithms successfully converged to the same global minimum fitness value of 89,641.23. Based on the aggregated results from all 22 datasets, tabu search displayed a median of 3 iterations (range: 1–7) for the first identification of the best model across the 22 datasets, while GA and ACO both showed a median of 3.5 iterations with ranges of 1–17 and 1–12, respectively. The fitness convergence curves for all 22 drug datasets are provided in **Supplementary Figs. 1–22**.

Computational efficiency comparison

The computational cost comparison results indicated that the actual execution time of exhaustive search increased substantially with the scale of the dataset (Fig. 3). Among the 22 datasets, when calculated as cumulative runtime under a 4-CPU, 2-GB memory configuration, 6 (27%) were completed within one day, 9 (41%) required 1–10 days, 6 (27%) fell within the range of 10–100 days, 1 (4.5%) required more than 100 days. In contrast, the actual runtime of GA, ACO, and tabu test cases was considerably lower than that of exhaustive search: 14 out of 22 tasks (63.6%) were completed within one day, and none exceeded 100 days. The stepwise method demonstrated the shortest runtimes, with all 22 tasks (100%) completed within one day.

Discussion

An automated modeling framework was developed based on the nlmixr2 platform, and the performance of three metaheuristic algorithms, GA, ACO, and TS, was compared with the stepwise model building method across 22 real clinical datasets. The results demonstrated that the metaheuristic approaches outperformed the stepwise algorithm in both model recovery accuracy and robustness. GA successfully recovered the reference model across all datasets except one, while ACO and TS also achieved comparably good performance in most cases. These findings suggest that, in contrast to the locally driven decision rules of the stepwise method, metaheuristic algorithms employing global search strategies offered a more reliable solution for the automated selection of PopPK models.

From the results, GA successfully recovered the ground-truth model in 21 of 22 (95.5%) datasets, yielding an overall retrieval rate of 95.5%. The retrieval rates of ACO and TS were 90.9% and 86.4%, respectively, which were largely consistent with the results (90%) of the previously presented simulation study [9]. Within the relatively limited search space defined in this study, the retrieval robustness of metaheuristic algorithms was close to that of exhaustive search. By contrast, although the stepwise method demonstrated the fastest computational speed, its retrieval rate was only 31.8%, suggesting that efficiency was achieved at the expense of performance. This discrepancy implied that, in complex data environments, the stepwise approach was prone to inconsistencies in structural model identification, partly due to its dependence on the search order [5]. Furthermore, previous research suggested that the superior performance of the GA can be derived from its combination of global and local search mechanisms [4]. It is worth noting that this study employed only the 1-bit GA rather than the more sophisticated, computationally demanding 2-bit GA, suggesting that, in relatively small search spaces, the 1-bit GA was sufficient. Importantly, the ACO algorithm

applied in this study did not incorporate local search, indicating considerable potential for further methodological development and performance improvement.

From the perspective of computational efficiency, the findings of this study clearly revealed the substantial differences in feasibility among the methods. Although exhaustive search could, in theory, guarantee a globally optimal solution, its computational cost was prohibitively high. In the 4-core cluster environment used in this study, the cumulative completion time for some cases even exceeded 100 days, despite the search space being limited to approximately 1,422 candidates. This demonstrated that as model complexity and candidate space grew exponentially, exhaustive search became virtually infeasible for practical research and applications. By contrast, the three metaheuristic algorithms exhibited marked advantages in computational efficiency: their runtime was often only 1/10 to 1/100 that of exhaustive search, while their retrieval accuracy and robustness remained close to, or even equivalent to, the exhaustive results. This reflected the fundamental purpose of developing and applying global optimization algorithms to substantially reduce computational burden while ensuring high-quality solutions—thereby making automated model selection a truly feasible approach in practical pharmacokinetic research.

Among the data sources analyzed in this study, five had previously published expert PopPK models available as references (imipenem, linezolid (oral), linezolid (iv), paclitaxel, and relebactam). For imipenem, the models selected by exhaustive search, ACO, GA, and tabu were fully consistent with the published model in terms of structure (iv_2cmpt), IIV (CL, V_c , V_p), and RUV (proportional), whereas the stepwise method yielded a one-compartment model with IIV on CL and V_c . For linezolid (oral), all methods identified the oral_1cmpt structure, but the models from exhaustive search, ACO, GA, and tabu included an additional IIV on K_a and adopted an additive RUV model, while the published model specified IIV on CL and V_c with a combined RUV model. This discrepancy may have been related to the absence of fixed-effect covariates such as allometric scaling (as a pediatric study), which consequently led to an additional introduction of variability in K_a to account for this part of the unexplained variability. For linezolid (iv), all methods (including the stepwise method) agreed with the published model in structure (one-compartment model) and IIV (CL) but applied an additive RUV model instead of the proportional RUV reported in the literature. For paclitaxel, the published model was a three-compartment model. Both exhaustive search and GA also identified the same structure, but with an additional IIV on Q_2 and a combined RUV model, whereas the published model specified IIV on CL, V_c , Q with a log-additive RUV. This difference was due to the fact that the search space did not include log-additive error structures. For relebactam, tabu search produced results fully consistent with the published model in structure (two-compartment model), IIV (CL, V_c , V_p), and RUV (proportional). In contrast, exhaustive search, ACO, and GA tended towards a three-compartment model with an additional IIV on V_{p2} . These highlighted the importance of global metaheuristic algorithms, given that the stepwise method failed to reproduce the published expert-level models, whereas global search strategies were capable of approaching or even fully recovering them.

Despite its promising performance across multiple real-world datasets, the nlmixr2auto framework had several limitations. Algorithmic performance may have depended on parameter settings and dataset characteristics, underscoring the need for systematic evaluation of parameter sensitivity and robustness in more complex scenarios. From a design perspective, improvements such as multi-start strategies for TS or incorporating prior pharmacokinetic knowledge into the pheromone updating of ACO could have enhanced the ability to escape local optima and improve efficiency. Moreover, the current study focused on structural and statistical models without addressing automated covariate selection. Future work should therefore extend the framework to integrate covariate selection within global optimization to improve both interpretability and clinical applicability.

In summary, the nlmixr2auto framework demonstrated strong potential for automated model selection and optimization in real-world PopPK applications. By combining automation with heuristic search strategies based on iterative learning

and feedback, the framework substantially reduced computational time compared with exhaustive search while maintaining accuracy and robustness. With its open-source nature and methodological advantages, nlmixr2auto had the potential to provide researchers and the broader community with an efficient, reliable, and reproducible tool that could support both population pharmacokinetic research and clinical applications.

Declarations

Competing Interests

Matthew Fidler is an employee of Novartis and Bern-Thomas Nyang'wa declares employment at Médecins Sans Frontières (MSF). All other authors declared no competing interests for this work.

Author Contribution

ZHH, FK and JS: design and conceptualization; ZHH: algorithm programming and data analysis; BTN: provided the clinical data; ZHH and MF: data interpretation and methodological best practices; ZHH: first draft of manuscript; All authors contributed to writing, editing and reviewing the manuscript.

Acknowledgement

F.K. is recipient of a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 220587/Z/20/Z).

Data Availability

Simulated datasets are provided within the supplementary material. Real-life datasets are obtained from published papers.

References

1. Lalonde RL, Kowalski KG, Hutmacher MM et al (2007) Model-based Drug Development. *Clin Pharmacol Ther* 82:21–32. <https://doi.org/10.1038/sj.clpt.6100235>
2. Berezowska M, Hayden IS, Brandon AM et al (2025) Recommended approaches for integration of population pharmacokinetic modelling with precision dosing in clinical practice. *Br J Clin Pharmacol* 91:1064–1079. <https://doi.org/10.1111/bcp.16335>
3. Kruse M, Stankeviciute S, Perry S (2025) Clinical pharmacology—how it shapes the drug development journey. *Eur J Clin Pharmacol* 81:597–604. <https://doi.org/10.1007/s00228-025-03811-z>
4. Li X, Sale M, Nieforth K et al (2024) pyDarwin machine learning algorithms application and comparison in nonlinear mixed-effect model selection and optimization. *J Pharmacokinet Pharmacodyn* 51:785–796. <https://doi.org/10.1007/s10928-024-09932-9>
5. Chen X, Nordgren R, Belin S et al (2024) A fully automatic tool for development of population pharmacokinetic models. *CPT: Pharmacometrics Syst Pharmacol* 13:1784–1797. <https://doi.org/10.1002/psp4.13222>
6. Richardson S, Irurzun Arana I, Nowojewski A et al (2025) A machine learning approach to population pharmacokinetic modelling automation. *Commun Med* 5:327. <https://doi.org/10.1038/s43856-025-01054-8>

7. Committee for Medicinal Products for Human Use (CHMP) (2007) Guideline on Reporting the Results of Population Pharmacokinetic Analyses. London, UK
8. Huang Z, Fidler M, Standing JF, Klopogge F (2026) Fitness Functions for Population Pharmacokinetic Modeling. *Journal of Pharmacokinetics and Pharmacodynamics*
9. Huang Z, Fidler M, Standing JF, Klopogge F (2026) Automated Model Selection in Population Pharmacokinetics Using Metaheuristic Algorithms. *Journal of Pharmacokinetics and Pharmacodynamics*
10. Lixoft (2024) Tobramycin Data Set
11. Nyang'wa B-T, Motta I, Moodliar R et al (2026) Pharmacokinetics, bactericidal activity and toxicity of short oral regimens for rifampicin-resistant tuberculosis treatment. *Nat Commun.* <https://doi.org/10.1038/s41467-026-74335-y>
12. Berry C, du Cros P, Fielding K et al (2022) TB-PRACTECAL: study protocol for a randomised, controlled, open-label, phase II–III trial to evaluate the safety and efficacy of regimens containing bedaquiline and pretomanid for the treatment of adult patients with pulmonary multidrug-resistant tuberculosis. *Trials* 23:484. <https://doi.org/10.1186/s13063-022-06331-8>
13. Hagemijer F (1975) Absorption, half-life, and toxicity of oral aprindine in patients with acute myocardial infarction. *Eur J Clin Pharmacol* 9:21–25. <https://doi.org/10.1007/BF00613425>
14. Korzeniowski OM, Scheld WM, Sande MA (1977) Comparative Pharmacology of Cefaclor and Cephalexin. *Antimicrob Agents Chemother* 12:157–162. <https://doi.org/10.1128/aac.12.2.157>
15. Martin E, Koup JR, Paravicini U, Stoeckel K (1984) Pharmacokinetics of ceftriaxone in neonates and infants with meningitis. *J Pediatr* 105:475–481. [https://doi.org/10.1016/S0022-3476\(84\)80032-3](https://doi.org/10.1016/S0022-3476(84)80032-3)
16. Nyang'wa B-T, Motta I, Moodliar R et al (2025) Clofazimine population pharmacokinetics and target attainment in rifampicin-resistant Tuberculosis patients. 2025.12.04.25340990
17. Ghoneim MM, Mewaldt SP, Ambre J (1975) Plasma Levels of Diazepam and Mood Ratings. *Anesth Analgesia* 54:173
18. Kaplan SA, Jack ML, Alexander K, Weinfeld RE (1973) Pharmacokinetic profile of diazepam in man following single intravenous and oral and chronic oral administrations. *J Pharm Sci* 62:1789–1796. <https://doi.org/10.1002/jps.2600621111>
19. MacMillan WE, Wolberg WH, Welling PG (1978) Pharmacokinetics of Fluorouracil in Humans¹. *Cancer Res* 38:3479–3482
20. Phillips TA, Howell A, Grieve RJ, Welling PG (1980) Pharmacokinetics of oral and intravenous fluorouracil in humans. *J Pharm Sci* 69:1428–1431. <https://doi.org/10.1002/jps.2600691220>
21. Bhagunde P, Patel P, Lala M et al (2019) Population Pharmacokinetic Analysis for Imipenem–Relebactam in Healthy Volunteers and Patients With Bacterial Infections. *CPT: Pharmacometrics & Systems Pharmacology* 8:748–758. <https://doi.org/10.1002/psp4.12462>
22. Garcia-Prats AJ, Schaaf HS, Draper HR et al (2019) Pharmacokinetics, optimal dosing, and safety of linezolid in children with multidrug-resistant tuberculosis: Combined data from two prospective observational studies. *PLoS Med* 16:e1002789. <https://doi.org/10.1371/journal.pmed.1002789>
23. Wu F, Zhang X-S, Dai Y et al (2022) Dosage Strategy of Linezolid According to the Trough Concentration Target and Renal Function in Chinese Critically Ill Patients. *Front Pharmacol* 13. <https://doi.org/10.3389/fphar.2022.844567>
24. Mason WD, Winer N (1976) Pharmacokinetics of oxprenolol in normal subjects. *Clin Pharmacol Ther* 20:401–412. <https://doi.org/10.1002/cpt1976204401>
25. He J, Jackson CGCA, Deva S et al (2022) Population pharmacokinetics for oral paclitaxel in patients with advanced/metastatic solid tumors. *CPT: Pharmacometrics Syst Pharmacol* 11:867–879.

<https://doi.org/10.1002/psp4.12799>

26. Grasela TH Jr, Donn SM (1985) Neonatal Population Pharmacokinetics of Phénobarbital Derived from Routine Clinical Data. *Dev Pharmacol Ther* 8:374–383. <https://doi.org/10.1159/000457062>
27. Gugler R, Herold W, Dengler HJ (1974) Pharmacokinetics of pindolol in man. *Eur J Clin Pharmacol* 7:17–24. <https://doi.org/10.1007/BF00614385>
28. Liang D, Liang Z, Deng G et al (2023) Population pharmacokinetic analysis and dosing optimization of polymyxin B in critically ill patients. *Front Pharmacol* 14. <https://doi.org/10.3389/fphar.2023.1122310>
29. Nyang'wa B-T, Motta I, Moodliar R et al (2026) Population pharmacokinetics and target attainment of pretomanid in rifampicin-resistant tuberculosis patients. *Sci Rep* 16:15255. <https://doi.org/10.1038/s41598-026-46217-2>
30. Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: An overview and update. *JABES* 8:387. <https://doi.org/10.1198/1085711032697>
31. Aarons L, Vozeh S, Wenk M et al (1989) Population pharmacokinetics of tobramycin. *Br J Clin Pharmacol* 28:305–314. <https://doi.org/10.1111/j.1365-2125.1989.tb05431.x>
32. Lee SM, Yang S, Kang S, Chang MJ (2021) Population pharmacokinetics and dose optimization of vancomycin in neonates. *Sci Rep* 11:6168. <https://doi.org/10.1038/s41598-021-85529-3>
33. R Core Team (2026) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
34. Fidler M, Wilkins JJ, Hooijmaijers R et al (2019) Nonlinear Mixed-Effects Model Development and Simulation Using nlmixr and Related R Open-Source Packages. *CPT: Pharmacometrics Syst Pharmacol* 8:621–633. <https://doi.org/10.1002/psp4.12445>
35. Huang Z, Fidler M, Lan M et al (2025) An automated pipeline to generate initial estimates for population Pharmacokinetic base models. *J Pharmacokinet Pharmacodyn* 52:60. <https://doi.org/10.1007/s10928-025-10000-z>
36. Huang Z, Standing J, Fidler M, Klopogge F (2026) nlmixr2auto. Automated Population Pharmacokinetic Modeling

Figures

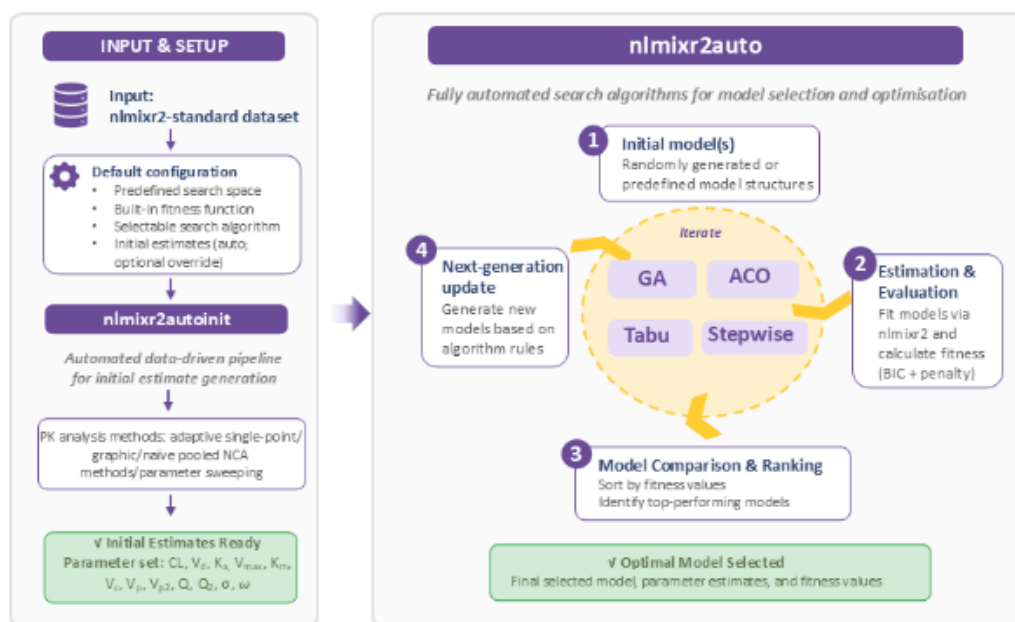


Figure 1

Workflow of the nlmixr2auto family pipeline. The pipeline takes a nlmixr2 standard dataset as input. Parameter initialization is performed by nlmixr2autoinit, which combines adaptive single-point methods, naïve pooled non-compartmental analysis, graphic methods, parameter sweeping, and pragmatic random-effect initialization to generate robust starting values. These estimates are then passed to nlmixr2auto, where model development is conducted using metaheuristic search algorithms (genetic algorithm, ant colony optimization, tabu search, and stepwise algorithm). Model estimation and evaluation are guided by a unified fitness function, and the system iteratively learns and updates to output a final optimized model

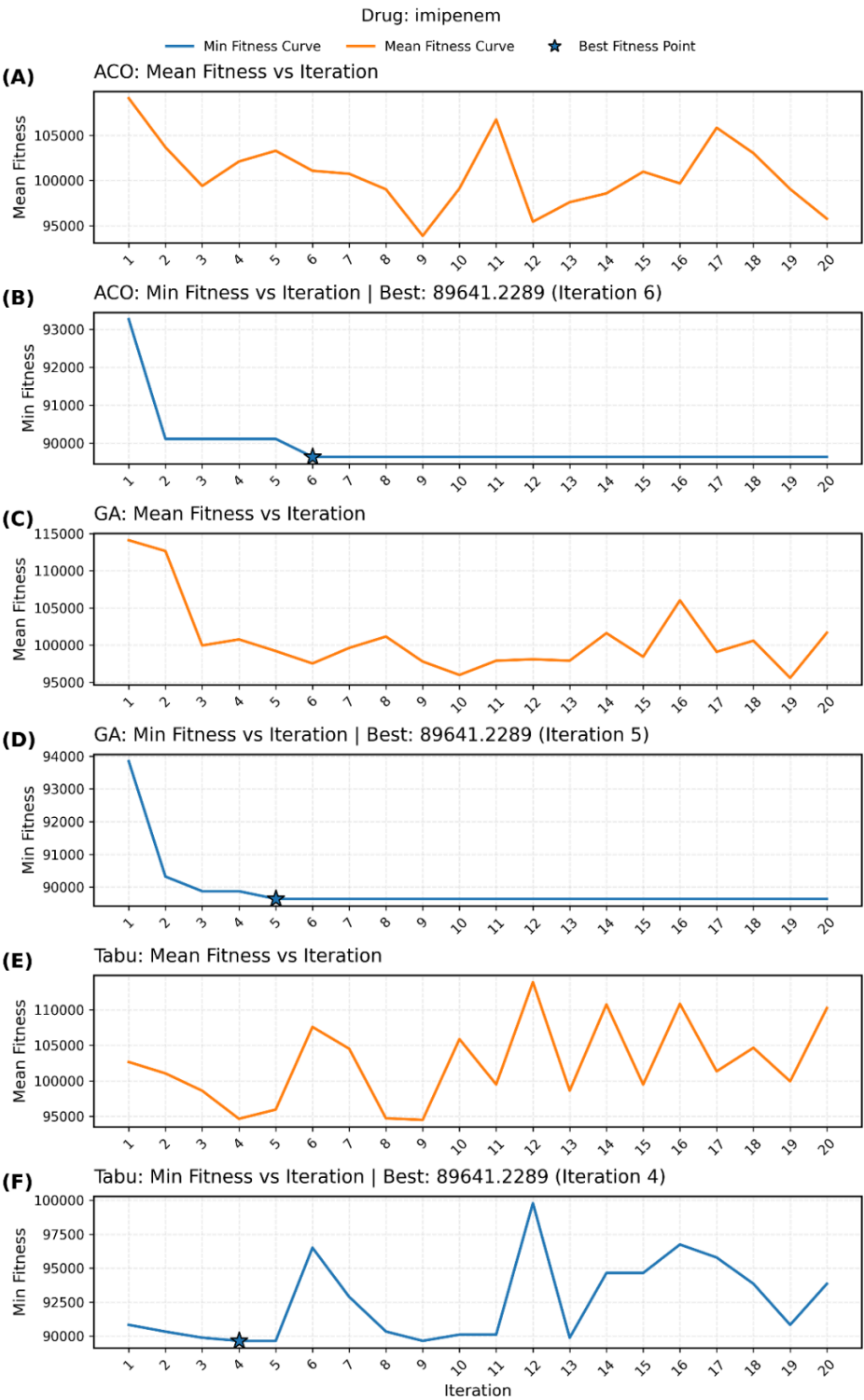


Figure 2

Convergence curves of three metaheuristic algorithms (ACO, GA, and TS) on the imipenem dataset. Stars indicate the best solution with fitness value and iteration number annotated. Mean fitness is calculated with infinite values excluded

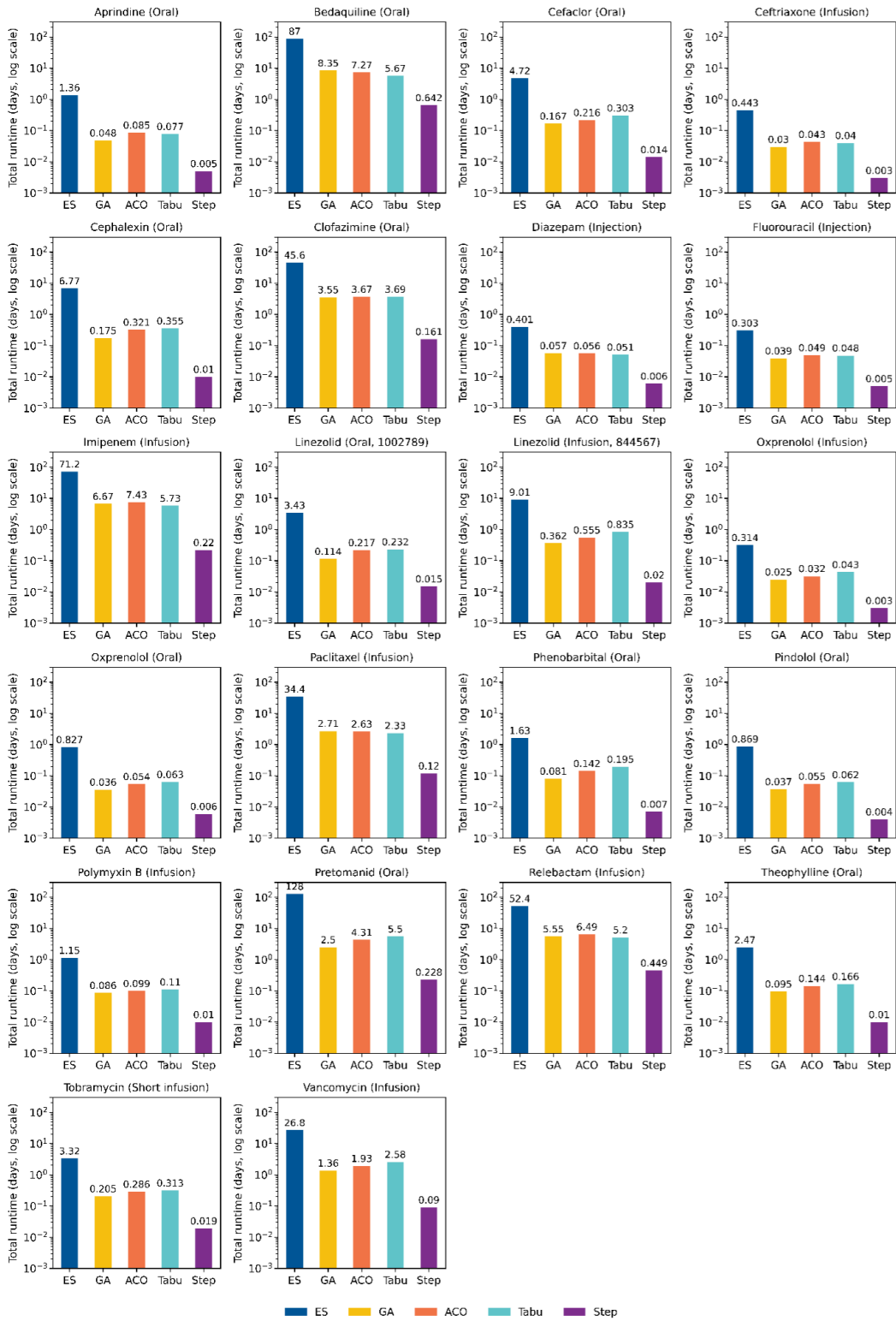


Figure 3

Runtime comparison of model selection algorithms across 22 datasets. The reported runtime represents the cumulative execution time under a 4-core CPU allocation, shown in days on a logarithmic scale on the y-axis. The x-axis indicates

the model selection algorithm: exhaustive search (ES), ant colony optimization (ACO), genetic algorithm (GA), tabu search (TS), and stepwise model building algorithm (Stepwise)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.docx](#)