



HAL
open science

Covariate Model Selection Approaches for Population Pharmacokinetics: A Systematic Review of Existing Methods, From SCM to AI

Mélanie Karlsen, Sonia Khier, David Fabre, David Marchionni, Jérôme Azé,
Sandra Bringay, Pascal Poncelet, Elisa Calvier

► **To cite this version:**

Mélanie Karlsen, Sonia Khier, David Fabre, David Marchionni, Jérôme Azé, et al.. Covariate Model Selection Approaches for Population Pharmacokinetics: A Systematic Review of Existing Methods, From SCM to AI. CPT: Pharmacometrics and Systems Pharmacology, In press, 10.1002/psp4.13306 . hal-04905888

HAL Id: hal-04905888

<https://hal.science/hal-04905888v1>

Submitted on 22 Jan 2025

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






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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

SYSTEMIC REVIEW OPEN ACCESS

Covariate Model Selection Approaches for Population Pharmacokinetics: A Systematic Review of Existing Methods, From SCM to AI

Mélanie Karlsen^{1,2}  | Sonia Khier^{3,4}  | David Fabre² | David Marchionni² | Jérôme Azé¹  | Sandra Bringay^{1,5}  | Pascal Poncelet¹  | Elisa Calvier² 

¹LIRMM, Laboratory of Computer Science, Robotics and Microelectronics in Montpellier, CNRS, Montpellier University, Montpellier, France | ²Pharmacokinetics Dynamics and Metabolism/Translational Medicine and Early Development, Sanofi R&D Montpellier, Montpellier, France | ³Pharmacokinetics and Pharmacometrics Department, Faculty of Pharmaceutical and Biological Sciences, Montpellier University, Montpellier, France | ⁴Institute of Mathematics Alexander Grothendieck (IMAG), CNRS UMR 5149, Montpellier University, Montpellier, France | ⁵Applied Mathematics, Computer Science and Statistics (AMIS), Montpellier 3 University, Montpellier, France

Correspondence: Mélanie Karlsen (melanie.karlsen@gmail.com)

Received: 25 September 2024 | **Revised:** 20 December 2024 | **Accepted:** 23 December 2024

Funding: This material is based upon work supported by the ANRT (Association nationale de la recherche et de la technologie) with a CIFRE fellowship granted to Mélanie Karlsen.

Keywords: artificial intelligence | covariate model building | covariate modeling | covariate screening | machine learning | pharmacometrics | population pharmacokinetic

ABSTRACT

A growing number of covariate modeling methods have been proposed in the field of popPK modeling, but limited information exists on how they all compare. The objective of this study was to perform a systematic review of all popPK covariate modeling methods, focusing on assessing the existing knowledge on their performances. For each method of each article included in this review, evaluation setting, performance metrics along with their associated values, and relative computational times were reported when available. Evaluation settings report was done for uncertainty assessment of communicated results. Results showed that EBES-based ML methods stood out as the best covariate selection methods. AALASSO, a hybrid genetic algorithm, FREM with a clinical significance criterion and SCM+ with stagewise filtering were the best covariate model selection techniques—AALASSO being the very best one. Results also showed a lack of consensus on how to benchmark simulated datasets of different scenarios when evaluating method performances, but also on which metrics to use for method evaluation. We propose to systematically report TPR (sensitivity), FPR (Type I error), FNR (Type II error), TNR (specificity), covariate parameter error bias (MPE) and precision (RMSE), clinical relevance, and model fitness by means of BIC, concentration prediction error bias (MPE), and precision (RMSE) of new proposed methods and compare them with SCM. We propose to systematically combine covariate selection techniques to SCM or FFEM to allow for comparison with SCM. We also highlight the need for an open-source benchmark of simulated datasets on a representative set of scenarios.

1 | Introduction

Population pharmacokinetic models are mathematical models describing the time course of drug concentrations, usually in

plasma. These models are used to describe processes (i.e., absorption, distribution, and elimination) and assess pharmacokinetic (PK) parameters and their variability within a population. Part of this variability can be explained through the inclusion of

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

© 2025 The Author(s). *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

Summary

- What is the current knowledge on the topic?
 - Efficiency of SCM is compromised under scenarios with complex models or high number of tested covariates. Alternative covariate modeling approaches have recently emerged. Most of them claim better performances than SCM. Meanwhile, ML-based methods have the potential to run faster.
- What question did this study address?
 - What are the existing covariate modeling methods, and which one could outperform SCM in terms of predictivity and computational time?
- What does this study add to our knowledge?
 - We have listed 22 covariate modeling methods. All methods showed unreported or better computational times than SCM, except for Genetic Algorithms. EBES-based ML methods were the best covariate selection methods, and AALASSO, H-GA-ML, FREM with a clinical significance criterion and SCM+ with SF were the best covariate model selection methods. The different settings used for method evaluation and comparison, however, do not allow for systematic validation of the superiority of these methods over SCM.
- How might this change drug discovery, development, and/or therapeutics?
 - Guiding pharmacometricians towards more performing and more efficient covariate modeling methods can help reducing the time required for this task and ultimately speed up the population pharmacokinetic model development.

covariates in the model, which are usually demographic, biological, physiological, or genotypic variables.

During population pharmacokinetic (popPK) model development, covariate modeling is of utmost importance as it explains exposure variability (e.g., area under the curve, AUC) and effect variability across patients. It also serves as the basis of dose tailoring, thereby increasing the benefit–risk ratio. The field of pharmacometrics is facing the need for new methods that can handle a larger covariate scope. The amount of data collected during drug development keeps increasing in an attempt to better explain and predict drug pharmacology. An increase in covariate scope leads to an exponential increase in computational time of standard methods such as stepwise covariate modeling (SCM). Computational time is highly relevant from an industry perspective, which faces tight timelines for model development and models with long run time due to the use of large datasets of pooled clinical studies and often complex models. In this context, new methods may help reduce model building time but should not add significant computational burden.

This review aims at listing all existing covariate model building methods and assessing existing knowledge on their performances both in terms of predictivity and computational time. To do so, performance metrics used for method evaluation together

with their values, relative computational time when available, and setting used for assessing all these elements were collected and compared.

2 | Methods

This systematic review was conducted following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [1].

For an article to be included in this review, several eligibility criteria were used. Inclusion criteria ensured that each article in the database either introduced, evaluated, or applied a covariate modeling method, or compared two or more covariate modeling methods. This implies that reviews and articles performing within-method comparisons or evaluations were excluded from this systematic review. Exclusion criteria consisted in excluding articles that performed a simple application of a state-of-the-art method such as: SCM, generalized additive models (GAM), least absolute shrinkage and selection operator (LASSO) or full model-based methods.

Search for articles was conducted through a query (see Figure S1) launched in PubMed (PubMed: <https://pubmed.ncbi.nlm.nih.gov/>). A single reviewer screened each record manually for eligibility assessment and no automation tool was used for this process.

Methods identified in the articles included in this systematic review were grouped in categories, namely: Stepwise procedures (SP), full model-based approaches (FMA), genetic algorithms (GA), empirical Bayes estimates (EBEs) based ML methods (ML-EBE), and Non-EBEs-based ML methods (ML-nonEBE). For each method of each category, four elements were systematically reported when available: (i) method objective, (ii) evaluation setting used for performance assessment, (iii) metrics used for performance assessment and their associated values, and finally (iv) relative computational times of the methods. These elements were summarized in tables and figures.

Regarding (ii) the evaluation setting used for performance assessment, the number of real datasets and the number of criteria used for the setup of simulated scenarios were reported. A criterion is an element of the underlying true model that the authors decided to vary. A scenario is defined as an ensemble of criteria and their values used simultaneously to build one or more datasets. The number of built-up scenarios and the number of simulated datasets per scenario were also reported.

Methods were ultimately ranked based on combined comparative performance and computational time data provided in the article database. Ranking based on computational time was straightforward since time is a single metric. However, ranking methods based on their comparative performance was more complex due to the use of multiple evaluation metrics. To address this, we relied on the rankings provided by the authors in each article and verified that their conclusions aligned with the actual metrics. In cases where two or more articles presented conflicting rankings, all versions of the rankings were displayed.

3 | Results

3.1 | Database of Articles

The PubMed query was launched on July 31, 2024, and returned 226 articles, from which 106 were immediately excluded based on title and abstract screening. After screening of the actual article content, 94 additional articles were excluded based on the exclusion criteria, which resulted in 26 articles. Finally, two articles were manually added, leading to a total of 28 articles in the database of the systematic review. Indeed, articles were manually sought after to ensure that the review contained the introductory article of all methods mentioned, however, articles using a nonpharmacokinetical setting to introduce a method were not included. All these steps are summarized in Figure 1.

Among the 28 articles in the database, a total of 22 different covariate modeling methods were identified. These methods had different objectives: (i) covariate selection, (ii) covariate model (covariate and relationship to PK parameters) selection, (iii) covariate model and other popPK model components selection, and (iv) alternative. The overview of methods objectives according to their category can be seen in Figure 2.

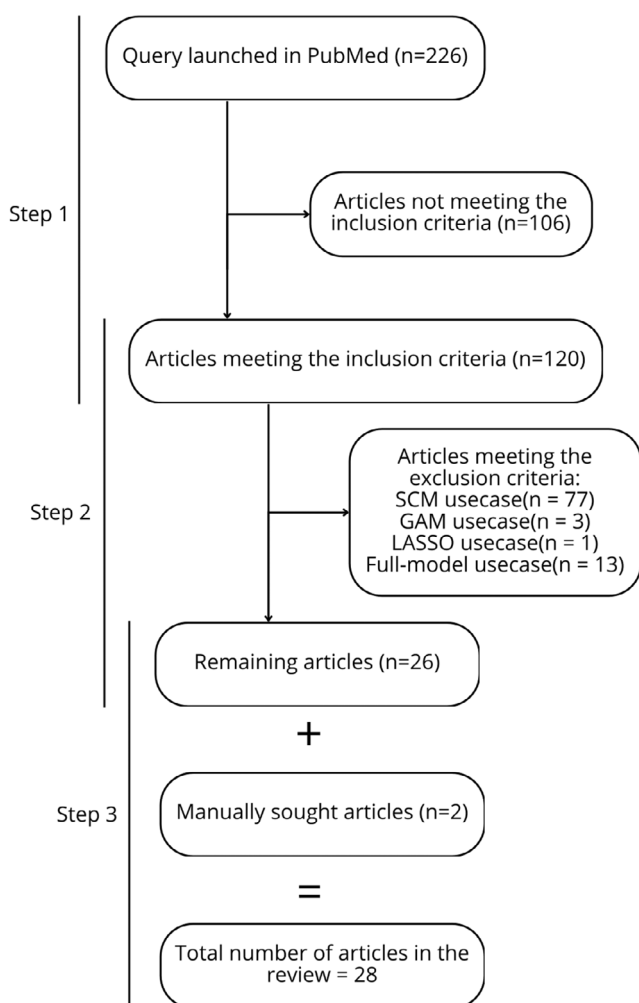


FIGURE 1 | Procedure for building the database of articles. n refers to the number of articles when relevant.

3.2 | Evaluation Settings and Metrics

Settings and performance metrics used for method evaluation and comparison greatly varied across the articles. Settings for method evaluation using simulated data were mostly restricted to few criteria being investigated, offering a robust evaluation on the investigated criteria, yet in some cases, no possible generalizability on communicated results. We therefore invite the readers to consider details of the evaluation settings as a mean for assessing the uncertainty to be put on the communicated performance results.

Tables 1 and 2 display the different settings that were used for method evaluation in the reviewed articles. Table 1 informs on the criteria used for creating simulated scenarios, when applicable. The eight criteria listed in this table represent an exhaustive list of the simulation criteria encountered in the articles, namely: number of covariates (N-Cov), extent of covariate effect (E-Cov), correlations between covariates (C-Cov), number of individuals (N-Ind), sampling design (Design), correlations between PK parameters (PK-Cor), presence of interindividual variability (IIV) and residual variabilities (σ). The corresponding total number of criteria used is reported in Table 2 together with the number of scenarios investigated and the number of datasets simulated per scenario. For each article, the number of real datasets used is also reported in Table 2. Note that a single article might appear several times in the tables if that article handled methods sharing different objectives. Tables were built that way to easily visualize groups of methods sharing the same objective. In the case of the paper from Ribbing et al. [2], the simulated scenarios actually correspond to a bootstrap resampling of a real clinical dataset.

Overall, there was a median of 1 criterion used for simulations across all articles, showing how constrained most of the investigated scenarios were. Also, some criteria such as different underlying structural model, presence of time-varying covariates or interoccasion variability were never investigated by the authors. Similarly, methods which were evaluated using real clinical data often only made use of a single dataset: out of 18 articles using real clinical data, 13 articles used a single one.

Note that symbolic regression network, clustering analysis technique, the genetic algorithm developed by Ismail et al. (GA1), gene expression programming (GEP), and Wald's approximation method (WAM) were evaluated using a single real and/or simulated dataset, limiting the generalizability of the communicated results for these methods.

Table 3 shows the heterogeneity of performance metrics used in the articles by displaying them all exhaustively. Metrics designating power and sensitivity were systematically reported under true-positive rate (TPR). Similarly, metrics designating Type I error rate were reported under false-positive rate (FPR), and those designating specificity under true-negative rate (TNR). Note, a single article might have used different performance metrics if it handled methods with different objectives, hence methods were grouped by their objective in the table. Note that the paper from Philipp et al. [3] appeared twice in the same "method objective" group in the table, because different methods required different metrics depending on the nature of the method being investigated.

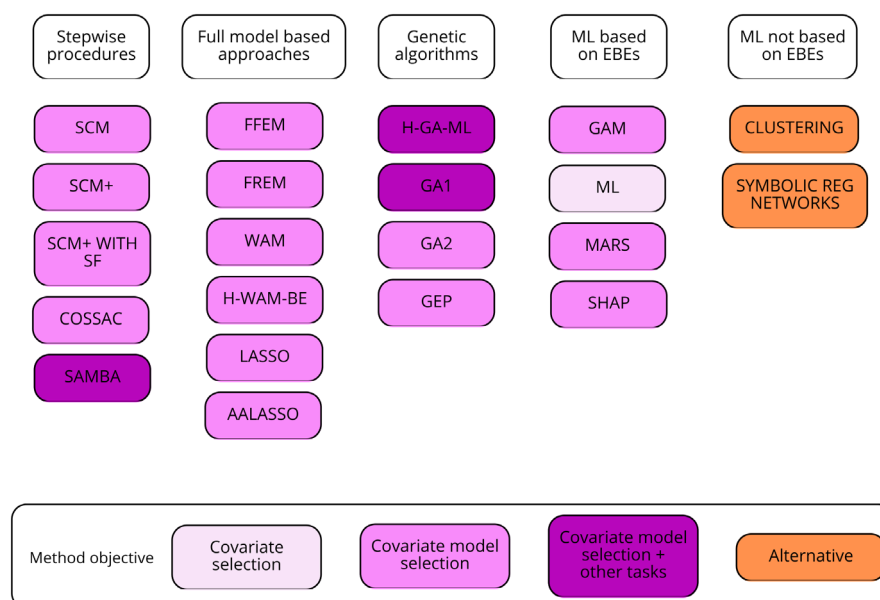


FIGURE 2 | Methods handled in the database of articles according to their category and objective. Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; COSSAC, COnditional Sampling use for Stepwise Approach based on Correlation tests; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GA1, Genetic Algorithm developed by Ismail et al.; GA2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid-Genetic Algorithm-Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; MARS, Multivariate Adaptive Regression Splines; ML, Machine Learning includes: (regularized) (stepwise) linear regression, random forests, neural networks, extreme gradient boosting, support vector machines. REG, Regression; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, Stepwise covariate model; SCM+ with SF, Stepwise covariate model+ with Stage-wise Flitering; SHAP, SHapley Additive exPlanations; WAM, Wald's Approximation Method.

3.3 | Syntheses of Results per Category of Methods

3.3.1 | Stepwise Procedures (SP)

The database contained 14 articles handling stepwise procedures, four of which were introductory articles of novel methods [4–7], and one of which evaluated an existing method [8]. All other articles were cases where SCM was only used for performance comparison to a method belonging to another category, therefore they will not be mentioned in this section.

SCM aims at selecting the covariate model and involves two sequential phases: a forward inclusion phase followed by a backward elimination phase. It was formally introduced jointly with a FO-linearization based technique in 1998 by Jonsson and Karlsson [4] in an article that evaluated its performance on 10 datasets simulated using a single scenario, meaning that no simulation criteria were used. Jonsson and Karlsson [4] compared the standard automated SCM to two different ways of applying the linearization technique in terms of concentrations prediction accuracy using root mean squared error (RMSE), and in terms of true positive (TP), that is the number of times the model correctly identified the true covariate with or without its true functional form and False Positive (FP), the number of times spurious covariates were selected. Results suggested that all methods performed equally well, although SCM without linearization had more TP (between 6 and 10 for SCM and between 2 and 10 for the other techniques) but at the cost of more FP (between 7 and 9 for SCM and between 4 and 5 for the other techniques) and

four times as many runs. In 2011, Khandelwal et al. [9] proposed to upgrade SCM with linearization by means of linearizing FOCE. Later occurrences of linearized-SCM in the review refer to this FOCE-linearization.

Since 1998, SCM has become the most well-known and broadly used procedure for popPK covariate models. Therefore, it is also the method that has been evaluated in some of the most robust and generalizable fashion, both in terms of numbers of real clinical datasets (17 real datasets) on which it has been assessed for comparison purposes [5], and numbers of criteria (four criteria) used for building the different scenarios on which it has been evaluated [8]. This evaluation performed in 2019 by Ahamadi et al. [8] where 16 scenarios were simulated according to 4 criteria aimed at evaluating SCM in terms of precision of the true covariate effect estimation (RMSE and conditional relative RMSE (crRMSE)) and TPR for identifying the correct model. SCM seemed to be sensitive to model complexity and dataset sample size, as these had an impact on TPR to identify the true covariate model and the precision of the parameter estimates. Moreover, highly correlated covariates (correlation = 0.89) had a high likelihood of being incorrectly selected by SCM.

In 2021, COnditional Sampling use for Stepwise Approach based on Correlation tests (COSSAC) was developed by Ayril et al. [5]. It is an iterative stepwise procedure aiming at selecting the covariate model, alternating between deletion and addition of covariates based on p values of correlation tests. The authors used 17 different real datasets for the purpose of

TABLE 1 | Criteria used for setting up the simulation scenarios, for each method and article concerned.

Method objective	Method	Criteria used for scenario construction								References
		N-Cov	E-Cov	C-Cov	N-Ind	Design	PK-Cor	IIV	σ	
Covariate selection	LR methods, sLR (AIC/RMSE), RF, ANN, XGB		✓							[28]
	RF, ANNs, SVR	✓	✓	✓						[27]
Covariate model selection	SCM, LASSO									[23]
	SCM, COSSAC, LASSO, GAM	✓	✓	✓						[27]
	SCM, GAM									[26]
	SCM, COSSAC									[7]
	SCM, GAM		✓							[28]
	SCM, GA2					✓				[20]
	FREM, FFEM	✓		✓						[10]
	FFEM	✓	✓	✓	✓	✓				[14]
	SCM, FREM		✓	✓	✓			✓		[15]
	SCM									[4]
Covariate model selection + other tasks	SCM, SCM+, SCM+ with SF									[6]
	SCM	✓	✓		✓	✓				[8]
	SCM, LASSO				✓					[2]
	SCM, H-WAM-BE				✓	✓	✓			[12]
	SCM, SCM+, FFEM	✓								[3]
	LASSO, ALASSO, AALASSO	✓	✓	✓	✓	✓	✓		✓	[13]
	H-GA-ML									[23]
	SAMBA									[7]
	GA1									[21]

Note: Blue: method evaluated on simulated data only. Mauve: method evaluated on real data only. Beige: method evaluated on both real and simulated data. Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; ALASSO, Adaptive Least Absolute Shrinkage and Selection Operator; ANN, Artificial Neural Networks; COSSAC, COnditional Sampling use for Stepwise Approach based on Correlation tests; ETR, Extra Tree Regressor; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GA1, Genetic Algorithm developed by Ismail et al.; GA2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid—Genetic Algorithm—Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; (s)LR, (stepwise) Linear Regression; MARS, Multivariate Adaptive Regression Splines; RF, Random Forests; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, stepwise covariate model; SCM+ with SF, stepwise covariate model+ with Stage-wise Flitering; SHAP, SHapley Additive exPlanations; SVR, Support Vector Regressor; WAM, Wald's Approximation Method; XGB, eXtreme Gradient Boosting. Other abbreviations, N-Cov, number of covariates; E-Cov, extent of the covariate effect; C-Cov, correlations between the covariates; N-Ind, number of individuals; Design, sampling design; PK-Cor, correlations between the PK parameters; IIV, presence of interindividual variability; σ , residual variabilities.

evaluating this novel method, by comparing it to SCM based on the number of runs, final log likelihood (LL) and corrected Bayesian Information Criterion (BICc) obtained. COSSAC procedure completed on average with 2 to 17 times fewer runs than SCM. Out of the 17 datasets, 11 of them resulted in the exact same final covariate model between COSSAC and SCM. They corresponded to datasets with few relationships in the final model. Four other datasets led to slightly different models but for which the difference in LL was less than 4, which was considered nonsignificant. The two last datasets led to different models, one for which the SCM fit was better by 8 LL

points than that of COSSAC, and the other one for which the COSSAC fit was better than that of SCM by 40 LL points. The difference in number of runs between the two methods grew linearly with the number of parameters and number of covariates to be tested but decreased with the square of the number of relationships to be tested.

In 2022, Svensson et al. [6] developed two upgraded SCMs, namely SCM+ and SCM+ with stage-wise filtering (SF). SCM+ is essentially the same as SCM, with extra steps including covariate scope reduction and retesting of previously removed

TABLE 2 | Number of real datasets, number of criteria used for scenario setup and number of simulated datasets per scenario for each method and article.

Method objective	Method	Setting used for method evaluation				References
		Criteria used (<i>n</i>)	Scenarios constructed (<i>n</i>)	Datasets simulated per scenario (<i>n</i>)	Real datasets used (<i>n</i>)	
Covariate selection	LR methods, sLR (AIC/RMSE), RF, ANN, XGB	1	2	100	1	[28]
	RF, ANNs, SVR	3	36	100	1	[27]
	ETR	NA	NA	NA	6	[31]
	RF	NA	NA	NA	1	[30]
Covariate model selection	SCM, LASSO	0	1	1	7	[23]
	SCM, COSSAC, LASSO, GAM	3	36	100	1	[27]
	SCM, GAM	0	1	10	1	[26]
	SCM, COSSAC	0	1	100	10	[7]
	SCM, GAM	1	2	100	1	[28]
	SCM, GA2	1	3	20	1	[20]
	FREM, FFEM	2	2	189	1	[10]
	FFEM	5	192	10,000	0	[14]
	SCM, FREM	4	Unclear, > 15	1	0	[15]
	SCM	0	1	10	0	[4]
	SCM, SCM+, SCM+ with SF	0	1	100	0	[6]
	SCM	4	16	250	0	[8]
	SCM, LASSO	1	4	100	0	[2]
	SCM, H-WAM-BE	3	3	300	0	[12]
	SCM, SCM+, FFEM	1	2	100	0	[3]
	LASSO, ALASSO, AALASSO	7	64	100	0	[13]
	SCM, COSSAC	NA	NA	NA	17	[5]
	SCM	NA	NA	NA	1	[19]
	GEP	NA	NA	NA	1	[22]
	GAM	NA	NA	NA	4	[25]
MARS	NA	NA	NA	1	[29]	
SHAP	NA	NA	NA	1	[32]	
WAM	NA	NA	NA	1	[11]	
Covariate model selection + other tasks	H-GA-ML	0	1	1	7	[23]
	SAMBA	0	1	100	10	[7]
	GA1	0	1	1	0	[21]
	H-GA-ML	NA	NA	NA	1	[19]

(Continues)

TABLE 2 | (Continued)

Method objective	Method	Setting used for method evaluation				References
		Criteria used (<i>n</i>)	Scenarios constructed (<i>n</i>)	Datasets simulated per scenario (<i>n</i>)	Real datasets used (<i>n</i>)	
Alternative	Clustering	NA	NA	NA	1	[33]
	Symbolic regression networks	NA	NA	NA	1	[34]

Note: Blue: method evaluated on simulated data only. Mauve: method evaluated on real data only. Beige: method evaluated on both real and simulated data.

Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; ANN, Artificial Neural Networks; COSSAC, Conditional Sampling use for Stepwise Approach based on Correlation tests; ETR, Extra Tree Regressor; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GA1, Genetic Algorithm developed by Ismail et al.; GA2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid-Genetic Algorithm-Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; (s)LR, (stepwise) Linear Regression; MARS, Multivariate Adaptive Regression Splines; RF, Random Forests; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, stepwise covariate model; SCM+ with SF, stepwise covariate model+ with Stage-wise Flitering; SHAP, SHapley Additive exPlanations; SVR, Support Vector Regressor; WAM, Wald's Approximation Method; XGB, eXtreme Gradient Boosting.

covariate–parameter relationships. In SCM+ with SF, covariates are divided into three categories according to expert knowledge: mechanistic, structural, and exploratory. Mechanistic covariates are directly included in the base model without being tested, whereas structural and exploratory covariates undergo stepwise inclusion, with structural covariates being tested before exploratory covariates. The authors evaluated these methods using a benchmark of 100 simulated datasets from a single scenario, meaning that no criteria were used for simulating different scenarios. The new methods were compared with the original SCM in terms of efficiency and ability to identify relevant relationships. Efficiency was measured by the total number of function evaluations and the number of executed NON-linear Mixed Effects Modeling (NONMEM) models, while ability to identify relationships was determined by the number of correctly identified covariate–parameter relationships (TP). The two SCM+ methods were significantly more efficient than SCM, with a reduction in the number of executed models ranging of from 44% to 70%, while SCM and SCM+ selected comparable covariate–parameter relationships across parameters. However, SCM+ with SF surpassed both SCM and SCM+ significantly in terms of ability to identify true covariate–parameter relationships.

The last stepwise method that was developed to this date is Stochastic Approximation for Model Building Algorithm (SAMBA), a simulation-based model building procedure for covariate, correlation, and error model selection introduced in 2022 by Prague and Lavielle [7]. The authors insisted that this automatic method should not be blindly used, as mechanistic biological knowledge is irreplaceable. For performance evaluation of the novel method, they gathered 10 real datasets which were evaluated in terms of BICc, and a benchmark of 100 simulated datasets using a single scenario (no criteria used) which were evaluated in terms of BICc, TPR, and FPR. Method performances were compared to those of SCM and COSSAC. For 6 out of the 10 real datasets, covariate models selected by SAMBA were either better or the same as the ones selected by SCM or COSSAC. For three other real datasets, covariate models selected by SAMBA were slightly worse (BICc difference smaller than 6) than the ones selected by SCM or COSSAC. Lastly, there was one real dataset for which BICc for SAMBA was worse by at least 10 points compared with the other methods. In the

simulated datasets, SAMBA correctly identified all existing covariate–parameter relationships in all 100 datasets, granting it with a TPR of 100%, and FPR (i.e., detection of spurious relationship) was 2%. SAMBA always found a BICc that was equal or better than that of the true model, which is consistent because SAMBA was specifically designed to choose the model with best BICc in this use case. Moreover, SAMBA computational times were 6 to 149 times smaller than those for SCM and 2 to 11 times smaller than for COSSAC.

In summary, all methods in this category, namely COSSAC, SCM+ with or without SF, as well as SAMBA claimed equivalent or better performance than SCM, while requiring fewer runs.

3.3.2 | Full-Model Based Approaches (FMA)

The database contained eight articles handling FMA, five of which were introductory articles [2, 10–13], two of which were evaluations of existing methods [14, 15] and one of which was a comparison of an existing method with SCM and SCM+ [3]. The readers should note that FMA require initial user inference of a full model. Some methods described below require certain post-hoc tools to allow to consider them as covariate model selection techniques.

The full fixed effects model (FFEM) was introduced at a conference in 2004 by Gastongay [16] and presented as an alternative to stepwise procedures. This method includes predefined candidate parameter–covariate relations into the model as fixed effects. It is systematically used in combination with a clinical relevance criterion to actually perform parameter–covariate selection. It was first evaluated in a simulated setting in 2018 by Xu et al. [14] (number of criteria used for simulation = 5) on its ability to identify true covariates by means of FPR and the TPR, where a true positive was considered as such if no nonsignificant variables were identified at the same time. In their study, the authors concluded that the FFEM was sensitive to the number of covariates in the underlying model, an increase of 10–20 covariates raising the chances of incorrectly selecting at least one covariate from 5% to 40%–70%. The authors noted that their estimation of false-positive effects might even be optimistic due

TABLE 3 | Performance metrics used by the authors for each method and article.

Method objective	Metrics used on real data only										Metrics used on simulated data only										Metrics that can be used both on real and simulated data									
	Error on reference PK parameter					Error on true covariate effect					Ability to identify covariates and relationships					Error on predicted concentrations					Model fitness									
	Method	R ²	MSE	Bias	Dispersion	PK parameter	Precision	RMSE	Other	TP(R)	TN(R)	FP(R)	FI	AUROC	Other	MPE	Accuracy	Other	RMSE	MAE	OFV	BIC	AIC	References						
Covariate selection	LR methods, sLR (AIC/RMSE), RF, ANN, XGB											✓*												[28]						
	RF, ANNs, SVR											✓*	✓											[27]						
Covariate model selection	ETR	✓	✓		✓																			[31]						
	RF																							[30]						
	SCM, LASSO								✓*							✓real			✓real		✓both	✓real	✓real	[23]						
	SCM, COSSAC, LASSO, GAM											✓***												[27]						
	SCM, GAM															✓both			✓both					[26]						
	SCM, COSSAC																					✓both		[7]						
	SCM, GAM												✓*											[28]						
	SCM, GA2																						✓sim	✓both	[20]					
	FREM, FFEM																								[10]					
	FFEM									✓* _{R,c}															[14]					
SCM, FREM															✓ _a			✓ _r						[15]						
SCM									✓***														✓	[4]						

(Continues)

TABLE 3 | (Continued)

Method objective	Metrics used on real data only										Metrics used on simulated data only										Metrics that can be used both on real and simulated data									
	Error on reference PK parameter					Error on true covariate effect					Ability to identify covariates and relationships					Error on predicted concentrations					Model fitness									
	Method	R ²	MSE	IIV	PK parameter dispersion	Bias	Precision	RMSE	Uncertainty	Other	TP(R)	TN(R)	FP(R)	FI	AUROC	Other	MPE	Other	Accuracy	RMSE	MAE	OFV	BIC	AIC	References					
SCM, SCM+, SCM+ with SF										✓***															[6]					
SCM							✓ _{r,cr}								✓ _d										[8]					
SCM, LASSO																					✓				[2]					
SCM, H-WA, M-BE										✓**R											✓				[12]					
SCM, SCM+					✓ _r			✓ _b		✓*R					✓ _{d,e}										[3]					
FFEM					✓ _r			✓ _b							✓ _e										[3]					
LASSO, ALASSO, AALASSO					✓																✓				[13]					
SCM, COSSAC																					✓	✓			[5]					
SCM																					✓				[19]					
GEP																					✓				[22]					
GAM																					✓				[25]					
MARS	✓																				✓		✓		[29]					
SHAP																									[32]					
WAM																									[11]					

(Continues)

TABLE 3 | (Continued)

Method objective	Metrics used on real data only										Metrics used on simulated data only										Metrics that can be used both on real and simulated data									
	Error on reference PK parameter					Error on true covariate effect					Ability to identify covariates and relationships					Error on predicted concentrations					Model fitness									
	Method	R ²	MSE	IIV	Bias	EE	RMSE	Precision	Uncertainty	Other	TP(R)	TN(R)	FP(R)	F1	AUROC	Other	MPE	Accuracy	Bias/Accuracy	RMSE	MAE	OFV	BIC	AIC	References					
Covariate model selection + other tasks	H-GA-ML									✓*	✓*	✓*				✓real			✓real	✓real	✓real	✓real	✓real	✓real	[23]					
	SAMBA									✓**		✓**R										✓both			[7]					
	GAI																					✓			[21]					
Alternative	H-GA-ML			✓																		✓			[19]					
	Clustering																					✓			[33]					
	Symbolic regression networks																	✓f							[34]					

Note: Blue: method evaluated on simulated data only. Mauve: method evaluated on real data only. Beige: method evaluated on both real and simulated data. Metrics abbreviations: AIC, Akaike Information Criterion; AUROC, Area Under the Receiver Operating Characteristics curve; BIC(C), (corrected) Bayesian Information Criterion; (r)EE, (relative) estimation error; FP(R), False Positive (Rate); LL, Log Likelihood; LRT, likelihood ratio test; MAE, Mean Absolute Error; mMdALE, mean Median Absolute Logarithmic Error; mMdAPE, mean Median Absolute Prediction Error; MPE, Mean Prediction Error; MSE, Mean Squared Error; OFV, Objective Function Value; (c)(r)RMSE, (conditional) (relative) root mean squared error; TP(R), True Positive (Rate); TN(R), True Negative (Rate); Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; ANN, Artificial Neural Networks; COSSAC, Conditional Sampling use for Stepwise Approach based on Correlation tests; ETR, Extra Tree Regressor; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GAI, Genetic Algorithm developed by Ismail et al.; GA 2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid—Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; (s) LR, (stepwise) Linear Regression; MARS, Multivariate Adaptive Regression Splines; RF, Random Forests; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, stepwise covariate model; SCM+ with SF, stepwise covariate model+ with Stage-wise Filtering; SHAP, SHapley Additive exPlanations; SVR, Support Vector Regressor; WAM, Wald's Approximation Method; XGB, eXtreme Gradient Boosting; ✓f, relative metric evaluation (e.g., if ✓f lies in the RMSE column, it means that the relative RMSE was computed); ✓+r, metric evaluation and relative metric evaluation (e.g., if ✓+r lies in the RMSE column, it means that both RMSE and the relative RMSE were computed); ✓cr, conditional relative metric evaluation (e.g., if ✓cr lies in the RMSE column, it means that the conditional relative RMSE was computed); ✓both/sim/real, concerns the methods that were assessed on both real and simulated data, and for which performance metrics could have been used on either (sim/real) or both types of data; ✓R, Rate of the metric (e.g., if ✓R lies in the FP(R) column, it means that the FP Rate was computed); ✓a, mean relative EE; ✓b, SE, rSE, empirical SE and empirical rSE; ✓c, sensitivity to detect the true significant covariate without falsely identifying any other non-significant covariate; ✓d, sensitivity for identifying the correct model; ✓e, bias (EE and rEE), precision (rRMSE) and uncertainty (coverage rates) of covariate ratios estimation; ✓f, predictions bias: mMdLE, mMdPE. Prediction accuracy: mMdALE, mMdAPE, ✓**/✓***/✓****, respectively, covariate selection (✓*), covariate model selection (✓*), covariate selection and covariate model selection (✓**); BIC and BICc were indistinguishably reported under BIC.

to using an intensive PK sampling scheme, whereas real-world sparse sampling in late-stage clinical trials could result in larger estimation errors and thus higher false positive magnitudes.

FFEM was compared to SCM and SCM+ by Philipp et al. [3] in 2024. In their article, SCM and SCM+ selected the same covariate model in at least 98.5% of cases, therefore all comparison results between FFEM and SCM also applied to FFEM and SCM+. They used 200 simulated datasets from two simulated scenarios: a base model including two covariate–parameter relationships and a covariate model including six. They evaluated covariate parameter estimation by means of accuracy (estimation error [EE], relative estimation error [rEE]) and uncertainty (standard error [SE], relative standard error [rSE], empirical SE, and empirical rSE). They also evaluated covariate ratios estimation by means of accuracy (EE and rEE), precision (relative root mean squared error (rRMSE)) and uncertainty (coverage rates), from which they constructed a clinical relevance decision-making process as a mean to select covariate–parameter relationships. They also reported percentages of correctly selected covariate–parameter relationships (TPR) and correctly unselected covariate–parameter relationships (TNR) in the case of SCM/SCM+. In summary, the three methods were all equally well able to identify clinical relevance of true covariate–parameter relationships. They were also all equally well able to either consider as clinically irrelevant (FFEM) or simply not select (SCM/SCM+) false ones. Regarding computational time, FFEM was 4 to 20 times faster than SCM/SCM+ under the covariate model scenario.

The full random effects model (FREM), a technique where covariates are treated as random variables to capture the influence of covariates through the estimated covariances between individual parameters and covariates, was proposed in 2012 by Karlsson [17]. Unlike FFEM, FREM is designed to handle correlated covariates. This method also allows to retrieve the coefficients of any combination of covariate inclusion by computing the ratio of the subcovariance matrix between parameters and covariates of interest, to that of the covariates of interest. Results of this poster presentation implied that FREM had higher precision of estimates than FFEM and that pre-defining subsets of covariate–parameter relationships led to biased parameters with FFEM. The first introductory article was written by Yngman et al. in 2022 [10], in which the authors performed both a real data-based and a simulation-based comparison of FREM with FFEM using two scenarios (number of criteria used for scenario setup = 2). The models built using the real dataset were compared subjectively based on qualitative similarity of the estimated coefficients, whereas the models built using the simulated datasets were compared in terms of accuracy of coefficient estimation. Results on the real dataset suggested that both FREM and FFEM estimated similar covariate–parameter relationship coefficients and IIV covariance matrices. For simulated data, FREM showed slightly higher precision in estimating the IIV covariance matrix, while both methods had similar accuracy in common model parameters and coefficient estimations. For scenarios involving high correlation and different parameterizations, FREM and FFEM produced comparable results. In 2023, Amann et al. [15] performed a more generalizable simulation-based evaluation of FREM (number of criteria used for simulation = 4) by comparing it to SCM using TPR, rRMSE

and rEE as performance metrics, reflecting power of identifying true covariates, precision, and bias, respectively. In the case of FREM, covariate identification was defined as the covariate having the highest effect size different from zero and non-overlapping 90% confidence interval. The conclusion was that in small datasets, the TPR of FREM was substantially higher, leading to a lower rEE, compared to SCM. In larger datasets ($n > 100$), TPR, rRMSE, and rEE of FREM were comparable to SCM. No comparative computational times were given in neither of the articles handling FFEM, however by design of the methods, FFEM is expected to be more time-efficient than SCM, similarly as for FREM.

LASSO was introduced as a covariate model selection tool in 2007 by Ribbing et al. [2] in a study leveraging a benchmark of bootstrapped datasets generated from a real dataset, where different scenarios corresponding to different numbers of individuals (number of criteria used for simulation = 1) were tested. In this LASSO implementation, L1 regularization is performed on all covariate coefficients simultaneously during OFV minimization of the full model. Performance was evaluated through the ability to correctly predict external data by means of Mean Absolute Error (MAE) of concentration predictions and results showed that LASSO surpassed SCM in that matter, regardless of what p value (0.05, 0.01 or 0.001) was used in the SCM procedure. However, the benefit of using LASSO over SCM was negligible in terms of MAE for larger datasets (120 or 180 individuals). In 2017, a method called Adjusted Adaptive LASSO (AALASSO) was developed for covariate modeling by Haem et al. [13], also performing regularization on all covariate coefficients simultaneously. It is an upgrade of Adaptive LASSO (ALASSO) developed by Zou [18] where weights are attributed to each coefficient to be estimated by LASSO, but this method was never applied to popPK covariate modeling to our knowledge. In AALASSO, these initial weights were set to the ratio of the SE of the maximum likelihood estimator to the maximum likelihood estimator itself of the coefficients to be estimated. For method evaluation, Haem et al. used a benchmark of datasets simulated using 64 scenarios according to 7 different simulation criteria, the highest number of criteria used in our review. For performance assessment, MAE of the concentrations and Error of the estimated coefficient of the first covariate were computed for LASSO, ALASSO and AALASSO. The results showed that AALASSO was the best of the three methods. Indeed, it demonstrated better predictive performance (lower MAE) than ALASSO and LASSO, particularly with small sample sizes and high covariates correlation (correlation = 0.7). Moreover, AALASSO consistently outperformed both ALASSO and LASSO when individual data was less informative (smaller number of observations per subject and increased residual variability), even when dealing with a high number of subjects. AALASSO also outperformed both LASSO and ALASSO as the number of covariates increased, regardless of correlation levels between covariates. No comparative computational times were provided but they are expected to be comparable.

The Wald's approximation method (WAM) to the likelihood ratio test (LRT) was introduced in 2001 by Kowalski et al. [11]. This method aims at finding the submodels, that is, the subcovariate–parameters relationships from a full model that achieves the best LRT, without having to explicitly compute it. The authors used a

single real dataset for the purpose of evaluating the method and it was the only article handling WAM in the review database. There was no actual assessment of model performances; however, the authors used several bootstrapped datasets from the real dataset to assess stability of the model found. The author claimed stability of the method on their bootstrapped data.

In 2021, an upgrade of WAM was introduced by Zou et al. under the name of Hybrid WAM with Backward Elimination (H-WAM-BE) [12]. It essentially performs backward elimination after the WAM step and is therefore a method that aims at finding covariate–parameter relationships too. The authors used a benchmark of simulated scenarios based on 2 criteria for method evaluation. TPR and TNR were used for method evaluation and comparison with SCM. It was concluded that best-performing H-WAM-BE methods showed TPR and TNR comparable to those of SCM. H-WAM-BE required shorter or comparable computation times than SCM regardless of the model structure, sample size, or sampling design used in this study.

To summarize, FFEM had performances comparable to FREM, while a study found that FREM has similar or better performances than SCM. However, another study found that FFEM has similar performances to those of SCM and SCM+. WAM itself was not compared to any method, but H-WAM-BE was compared to SCM and showed equivalent performance yet with shorter required computational time. AALASSO and ALASSO showed better performances than LASSO, which itself showed better ability than SCM at selecting the covariate model.

3.3.3 | Genetic Algorithm Based Methods (GA)

There were five articles in the database that handled genetic algorithms (GA) for covariate modeling. Four of them were introductions to novel methods [19–22] and one of them was a comparison of an existing GA with SCM [23].

The first article related to GA was published in 2006 by Bies et al. who introduced the Hybrid—Genetic Algorithm—Machine Learning (H-GA-ML) [19] in which they used a GA in combination with Downhill search after every 10 generations of GA to construct all components of the model, namely the structural, covariate, and pharmacostatistical model. They used a single real dataset to apply this novel method and evaluated its performance by comparing it with SCM in terms of objective function value (OFV), IIV and robustness (successful covariance step) of parameter estimates. The results showed that OFV, IIV and parameter estimates robustness were better for Hybrid-GA than SCM. The authors did not provide computational time but stated that the genetic algorithms require dramatically more computational time, although less personnel time, than SCM.

In 2012, a comparison of a single objective hybrid genetic algorithm with SCM and LASSO was performed on seven real clinical datasets and a single simulated dataset by Sherer et al. [23]. Their method was essentially the same as Bies et al.'s H-GA-ML, therefore we refer to it as H-GA-ML in the tables, figures and text. For model evaluation based on real data, they

used delta OFV as well as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC), RMSE and Mean Prediction Error (MPE) of predicted concentrations, reflecting model fitness, precision and bias respectively. For the simulated dataset, performance was evaluated based on TP and FP for covariate identification. For clinical datasets, H-GA-ML provided improved model fits over the stepwise approach for 4 out of 7 compounds based on AIC, which is consistent with Bies' results, but had slightly higher RMSE yet lower MPE compared to the stepwise approach. For the simulated dataset, H-GA-ML and SCM performed equally well at identifying true covariates (TP=3) but H-GA-ML was better at avoiding spurious covariates than SCM (FP=1 for H-GA-ML and FP=2 for SCM). Both methods were better than LASSO in both matters.

In 2017, the GEP was introduced by Yamashita et al. [22] for the first time as a method to determine the covariate model. The authors used a single real dataset as an example, and the evaluation was done by comparing the OFV of the GEP model with that of a literature model from Aarons et al. [24]. The GEP model was slightly different than that of the reported literature model with an extra covariate being selected for one of the two PK parameters that underwent covariate modeling, granting GEP with a better OFV.

In 2022, Ismail et al. [21] proposed an upgrade of Bies and Sherer's GA (referred to as GA1 in the tables and figures) in which the encoding of the model (the chromosome) was done as character string, as opposed to binary strings. As an example, they used a single simulated dataset on which the GA was compared in terms of OFV with both the true model and the SCM built by a pharmacometrician external to the study. The model identified by the GA had the right number of compartments but not the right covariate model structure. The SCM was different from the true model but had lower OFV than the one discovered using the GA (OFV = -2187 for SCM, OFV = -2153 for GA, while OFV = -2174 for the true model). True-positive and false-positive rates were not reported as such in the article; however, they could be deduced from a reported table displaying the selected covariate–parameter relationships. It appeared that GA and SCM performed equally well in selecting true covariates (TP=3 out of a total of 5 true covariate–parameter relationships), however SCM surpassed GA in avoiding spurious covariate (FP=0 for SCM, and FP=4 for GA, out of a total of five possible spurious relationships).

In 2023, Ronchi et al. introduced another GA [20] derived from Bies et al. (referred to as GA2 in the tables and figures), but this time solely dedicated to covariate model selection. They based their evaluation of performance on a single simulated dataset and on one real clinical dataset. Evaluation using the simulated data was performed by assessing TPR of multiple GA runs and the number of times where AIC was lower than that of the true model. Evaluation using the real data was done by comparing AIC between GA and SCM. Computational times were also provided in both cases. Results using simulated data suggested that, for GA, AIC was better than that of the true model in 18 out of 20 cases, and for 95% of runs, GA selected the best solution. Results using real data suggested that the models resulting from GA and SCM were very similar,

with a slightly better AIC in the case of GA. In terms of computational time, GA performed worse than SCM both for the simulated case (GA = 1 h, SCM = 2.5 min) and the real case (GA = 100 h, SCM = 2.5 h).

Overall, evaluation settings of all GA-based methods had poor generalizability. Moreover, it is inherent to the method itself that genetic algorithms require long computational time. Although they were rarely provided, Ronchi et al.'s article showed that a GA dedicated to covariate selection required at least 24 times longer to run than SCM. Indeed, this family of methods has the particularity of being computationally expensive regarding the number of function evaluations required at each iteration. The communicated results only raise little hope on their ability to be an alternative approach to SCM, provided the low performance gain at the cost of the high computational time required.

3.3.4 | EBEs Based ML Methods (ML-EBE)

There were eight articles in this category, with only one article introducing a new method [25], three articles comparing ML to a method from another category [26–28], and four remaining articles being simple applications of ML to a PK dataset [29–32].

GAM are flexible statistical models that predict outcomes by summing smooth functions of predictor variables, allowing for non-linear relationships and easy interpretation of individual predictor effects. It was introduced in 1992 by Mandema et al. [25] as a covariate model selection method in which the EBEs are regressed on the covariates. It is the first ever ML-like method to be applied to covariate modeling, and it is often used in a stepwise manner. The authors evaluated their newly introduced method using four different real clinical datasets without comparing the performances with those of another method. However, in 2002 Wählby et al. [26] performed a comparison of GAM with SCM in an article that used a benchmark of simulated datasets from a single scenario by means of bias and precision (MPE and RMSE, respectively) and run time for performance assessment. Their results showed very similar predictive performance between GAM and SCM both in terms of accuracy and precision, with GAM being quicker to run than SCM.

In 2017, Multivariate Adaptive Regression Splines (MARS) was first applied as a covariate modeling method by Hall et al. [29]. This method has the advantage of evaluating both linear and nonlinear relationships simultaneously. It was applied using a real clinical dataset, where the authors used R-squared values and mean squared error (MSE) of concentrations to select covariates. They also sporadically assessed explainability by means of explained variance on PK parameters. This article is the only instance of application of MARS in our database; therefore, no comparison of MARS with other covariate modeling methods has been done to this day to our knowledge.

In 2021, Sibieude et al. [27] published a comparison of ML methods (random forests [RF], artificial neural networks (ANN), support vector regressor [SVR]) regressed on EBEs, with more

classical methods (SCM, COSSAC, LASSO) regarding their ability to identify clinically relevant covariates. The ML methods providing a ranking of covariates, they used several automated approaches such as top-M, order of importance and minimum degree of importance in order to actually perform covariate selection. One real clinical dataset and one benchmark of datasets simulated according to 36 scenarios (number of criteria used for scenario construction = 3) were used for evaluation. ML methods were compared with each other using AUROC, whereas F1 score (i.e., the harmonic mean of precision) and TPR were used for comparison across both ML and classical pharmacometrics methods. Authors reported that correlations between covariates didn't change the performance of any method, but effect size did. Indeed, when the covariate effect size was high, the importance scores for true covariates were consistently greater than those for false covariates. In contrast, low effect sizes resulted in less precise selections, with AUROCs for methods such as ANN and SVR being close to random guesses. The authors performed an additional comparison of the ML methods to linearized-SCM and GAM based on a single scenario. This comparison showed that the use of linearization divided SCM runtimes roughly by 4, in agreement with Jonsson and Karlsson's [4] FO-linearization, while preserving its accuracy. GAM computational time was similar to that of neural networks, but with the lowest F1 score across all considered methods. Overall, this article suggested that ML methods surpassed classical pharmacometrics methods, both in terms of performance for performing covariate selection, and computational time.

In that same year, several cases of applications of tree-based methods using real clinical PK data have emerged, namely RF by Ibrahim et al. [30] and Extra Tree Regressors (ETR) by Tang et al. [31].

Later, in 2024, an article by Asiimwe et al. [28] aiming again to compare ML methods (linear regression-based methods including Ridge, LASSO, GAM, stepwise Linear Regression (sLR) (both AIC and RMSE based), RF, ANN, eXtreme Gradient Boosting [XGB]) with SCM and linearized-SCM was published. Note that they implemented LASSO as a regressor on EBEs—it is therefore different from the common LASSO implementation presented in the next section. This implementation of LASSO will be referred to as LASSO-EBEs in tables and figures. The authors first used simulated data from a real clinical dataset to compare all methods based on their ability to select the true covariates by means of the F1 score proposed earlier by Sibieude et al. [27]. Then, the best covariate selection method found with the simulated data was applied on the real clinical dataset from which the data was simulated, after which they applied SCM. When using simulated data, ridge regression was the best ML technique with a F1 score of 0.475 ± 0.231 for both scenarios. Classical pharmacometrics techniques performed worse than ML methods combined with SCM, having an F1 score of 0.251 ± 0.274 (0.198 ± 0.240 with linearization) with only forward selection, and 0.206 ± 0.228 (0.172 ± 0.213 with linearization) with backward selection too. On the real clinical dataset, applying a preselection of covariates with ridge regression followed by SCM resulted in the same final model as with SCM alone, yet with a runtime decrease from 1.75 to 1 day with ML-based preselection. Using FOCE-linearization with SCM divided its runtime by 15.

SHapley Additive exPlanations (SHAP) was first introduced as a covariate modeling method in 2022 by Janssen et al. in an application using a real clinical dataset [32]. No performance was reported for this application.

To summarize this section, all ML methods regressed on EBEs (GAM, linear regression-based methods including LASSO, Ridge and elastic-net, RF, ANN, SVR, and XGB) proved to have equal or better performances than SCM regarding covariate selection, while being more time efficient. However, it is important to acknowledge that these covariate selection methods require model parameter estimation prior and after their use in order to perform covariate model building.

3.3.5 | Non-EBEs Based ML Methods (ML-non-EBE)

There were only two articles in this category, and both introduced a novel method [33, 34].

Semmar et al. presented in 2005 a clustering method used for covariate modeling [33]. The clustering method was built with euclidean distance and complete-link clustering algorithm, allowing to create a categorical covariate describing the patients groups (clusters). They evaluated this method, using a single real clinical dataset, through OFV, model performances in terms of precision of estimates, IIV and residual variabilities with or without clustering. Similar performances were recorded when constructing a covariate model with all original covariates versus the single categorical covariate built with clustering. This method is not considered as a covariate model selection technique, but rather as an alternative technique. In 2023, another alternative technique was introduced by Wahlquist et al. in which they used symbolic regression networks [34], a particular type of neural network. The specificity of this method lies in the fact that each PK parameter has its own symbolic regression network, and each of these networks are trained in parallel on their ability to correctly predict all together the observed concentrations. The advantage of symbolic regression networks over classical neural networks is that they give equations of covariate relationship to PK parameter without requiring user pre-specification, besides being constrained to the relationships dictated by the neural network architecture. The authors used a single real dataset aggregated from 30 different studies of the same drug for model evaluation. The model resulting from the networks was compared to a published state-of-the-art model on the same dataset through assessment of mean Median Absolute Logarithmic Error (mMdALE) and mean Median Absolute Prediction Error (mMdAPE). The article's method found a covariate model structure and corresponding parameter values with a slightly better fit, while relying on notably fewer covariates than the state-of-the-art model. Although no comparison of computational time was reported, it is known that neural networks are usually faster to train than classical pharmacometrics methods but they require more data.

To summarize, clustering showed that it brought interpretability rather than performance, and symbolic regression networks showed better performances than the published state-of-the-art model to which it was compared to.

3.4 | Summary of Reported Comparative Performances With Associated Evaluation Settings

Actual meta-analyses were not possible due to the heterogeneity of the settings used for evaluation and comparison, and the heterogeneity of the methods objectives. Indeed, two methods with different objectives are not expected to be evaluated similarly. Furthermore, published comparisons may be conducted for different purposes, whether exploratory or confirmatory. Advantages of each method can vary depending on these different intentions. Yet, we combined in a single figure the comparative performances (Figure 3) and comparative computational times (Figure 4) between methods, allowing the reader to have an overview of all the articles general comparative conclusions. The existence of these overviews is only fair in presence of reminders of the methods' objectives and comparison settings. Therefore, lines with varying colors and widths were used to convey information about the setting of comparison. We again invite the reader to consider this information as a mean for assessing the uncertainty to be put on the communicated results.

Figure 3 displays comparative performances reported within each paper, when available. ML methods, LASSO, AALASSO, H-GA-ML, FREM, and SCM+ with SF have better reported performances than SCM, although one instance of LASSO comparison with SCM reported better performances of the latter. ML methods are the only methods solely dedicated to covariate selection. They are the best reported methods for this task, and their evaluation was performed on a great number of scenarios, as showed by the width of the line. AALASSO was the best reported method for covariate model selection, while SCM+ with SF, H-GA-ML and FREM were the second best reported methods for this task. There is no information regarding the comparative performances of MARS, WAM, Cluster analysis, SHAP, Symbolic regression networks, and GEP.

Figure 4 displays the comparative computational times reported within each paper, when available. It shows that all methods for which this information is provided claim to have lower computational times than the classical SCM except for GA2, with the ML-based methods being among the fastest methods, as expected. However, it is important to acknowledge that EBEs-based ML methods involve necessary preliminary and follow-up runs, leading to extra computational time not account for in the figure.

4 | Discussion

The objective was to list all covariate modeling techniques and gather all available information regarding their comparative performances and computational times.

We decided to focus the review on general covariate modeling methods and purposely did not include specific methodological comparisons of single methods. This review did not cover the investigations of the impact of missing values either, although some methods are implicitly designed to handle them, such as FREM or tree-based ML methods. It is well established that missing covariates can have an impact on covariate

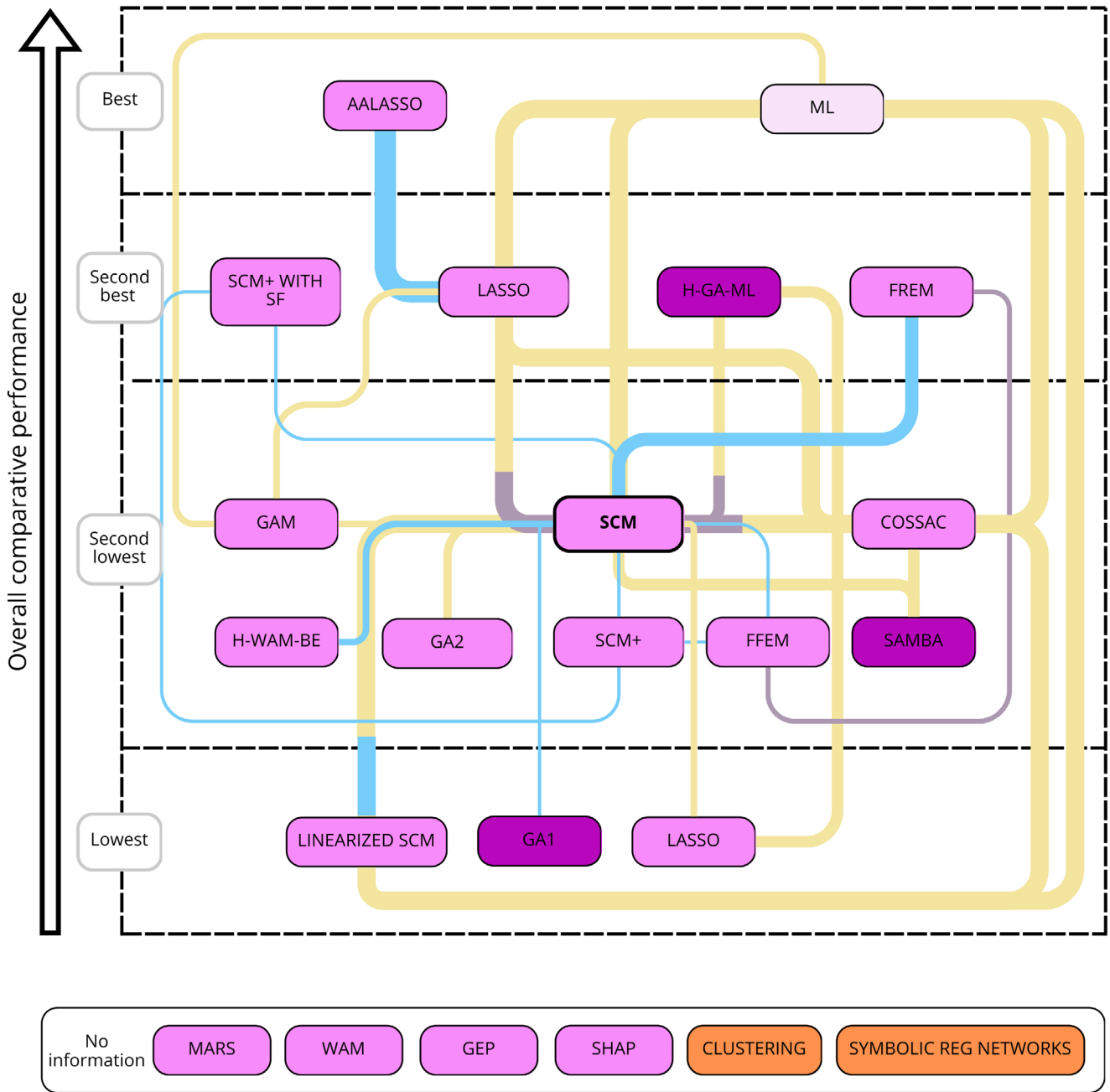


FIGURE 3 | Overall reported comparative performances across methods. The dashed lines separate groups of methods having reportedly equivalent performance regarding their objective. Blue line: Comparison performed only simulated data. Mauve line: Comparison performed only on real data. Beige line: Comparison performed on both real and simulated data. Multiple colors: The comparison has been made across articles using different evaluation settings. Width of the line = the number of scenarios (whether real or simulated) investigated for the comparison (the larger the width, the more scenarios were investigated for the corresponding comparison). In case multiple articles made the same comparison, the maximum number of scenarios investigated across those articles was retrieved. There are two occurrences of LASSO in the figure, because one comparison of LASSO to SCM [23] contradicts all the others. Boxes are colored according to the methods objective. Light pink: Method performing covariate selection. Pink: Method performing covariate model selection. Purple: Method performing covariate model selection and other tasks. Orange: Alternative methods. Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; COSSAC, COnditional Sampling use for Stepwise Approach based on Correlation tests; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GA1, Genetic Algorithm developed by Ismail et al.; GA2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid-Genetic Algorithm-Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; MARS, Multivariate Adaptive Regression Splines; ML, Machine Learning. Includes: (regularized) (stepwise) linear regression, random forests, neural networks, extreme gradient boosting, support vector machines; REG, Regression; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, Stepwise covariate model; SCM+ with SF, Stepwise covariate model+ with Stage-wise Flitering; SHAP, SHapley Additive exPlanations; WAM, Wald's Approximation Method.

model building performance, and careful handling of missing covariates before use of any covariate model building method should be performed. We refer the reader to the paper of Bräm

et al. [35], which shows that classical methods such as mean imputation or simple deletion can be biased under certain conditions, suggesting the need for careful consideration. They

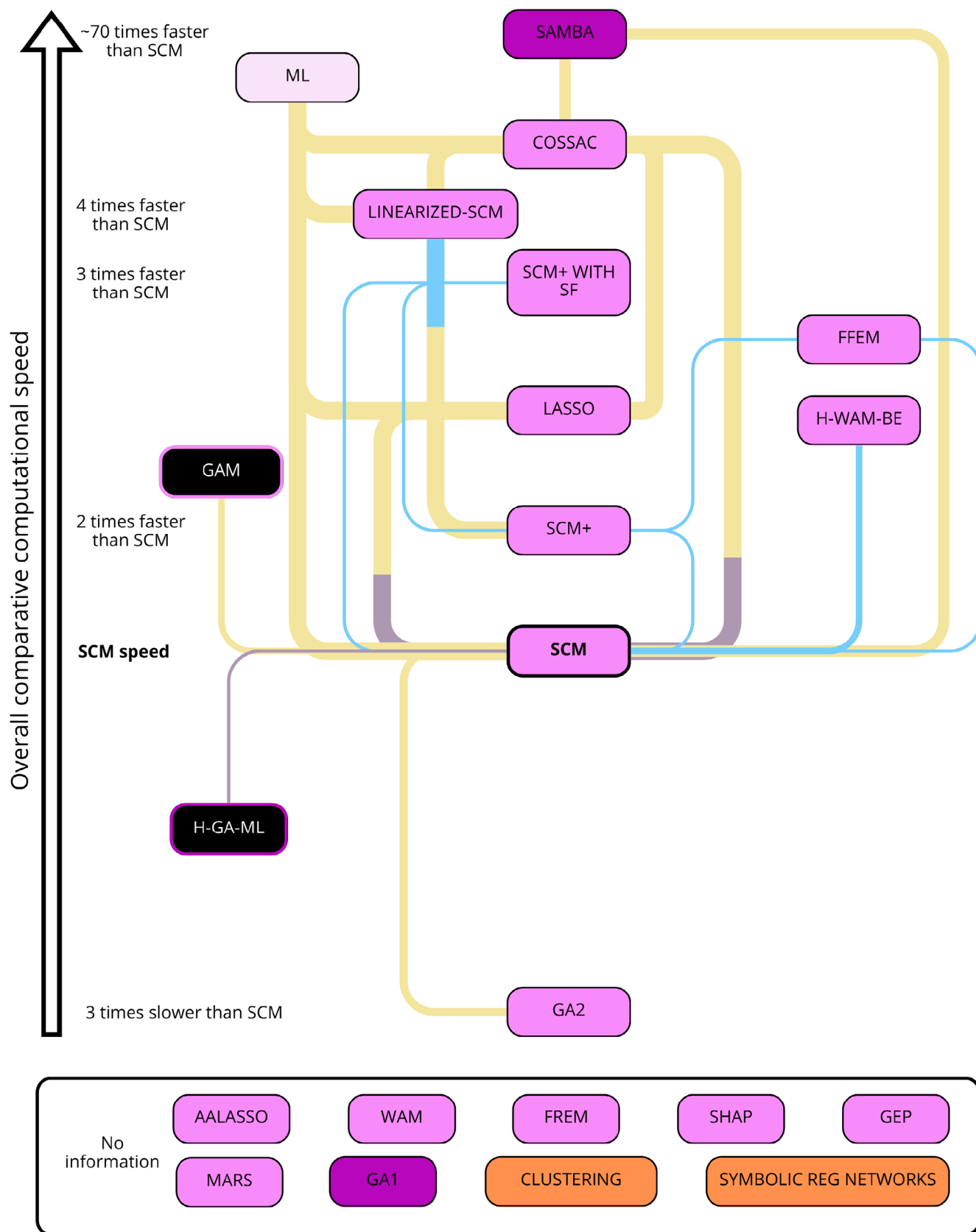


FIGURE 4 | Legend on next page.

FIGURE 4 | Overall reported comparative computational speed across methods. Blue line: Comparison performed only simulated data. Mauve line: Comparison performed only on real data. Beige line: Comparison performed on both real and simulated data. Multiple colors: The comparison has been made across articles using different evaluation settings. Width of the line = the number of scenarios (whether real or simulated) investigated for the comparison (the larger the width, the more scenarios were investigated for the corresponding comparison). In case several articles made the same comparison, the maximum number of scenarios investigated across those articles was retrieved. Black box = no magnitude on the difference in computational speed was reported. Boxes are colored according to the methods objective. Light pink: Method performing covariate selection. Pink: Method performing covariate model selection. Purple: Method performing covariate model selection and other tasks. Orange: Alternative methods. Methods abbreviations: AALASSO, Adjusted Adaptive Least Absolute Shrinkage and Selection Operator; COSSAC, Conditional Sampling use for Stepwise Approach based on Correlation tests; FFEM, Full Fixed Effects Model; FREM, Full Random Effects Model; GA1, Genetic Algorithm developed by Ismail et al.; GA2, Genetic Algorithm developed by Ronchi et al.; GAM, Generalized Additive Models; GEP, Gene Expression Programming; H-GA-ML, Hybrid-Genetic Algorithm-Machine Learning; H-WAM-BE, Hybrid WAM with Backward Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; MARS, Multivariate Adaptive Regression Splines; ML, Machine Learning. Includes: (regularized) (stepwise) linear regression, random forests, neural networks, extreme gradient boosting, support vector machines; REG, Regression; SAMBA, Stochastic Approximation for Model Building Algorithm; SCM, Stepwise covariate model; SCM+ with SF, Stepwise covariate model+ with Stage-wise Flitering; SHAP, SHapley Additive exPlanations; WAM, Wald's Approximation Method.

also showed the potential of ML-based imputation techniques. We also decided not to distinguish continuous and categorical covariates in the datasets used by the authors for method evaluation, as this level of information was not deemed necessary for the relative performance overview provided by our review and in view of the important differences in method assessment across publications. However, this distinction should be encouraged when comparing covariate model building methods in research papers. We refer to the article of Sanghavi et al. [36] for interesting points of considerations in all these matters.

4.1 | New Standards for Covariate Model Building: Roadblocks and Solutions

This review highlights two main roadblocks that prevent pharmacometricians from moving away from the SCM standard: (i) lack of an open-source benchmark of simulated datasets for investigating the performance of covariate model building methods, and (ii) lack of standardized performance metrics that could be used across different covariate model building methods.

Indeed, despite the growing number of covariate modeling techniques proposed to outperform SCM in terms of speed or predictive performance, to date the level of evidence to support the use of a method over SCM is not sufficient for the latter to be overthrown. There is currently a need for a generalizable evaluation and comparison of all covariate modeling techniques. From the 21 methods proposed for covariate modeling outside of SCM, 6 were not compared to any other covariate modeling method, and from the remaining 15 which underwent comparison, 14 were compared to the standard SCM method. Comparison of covariate model building methods used different evaluation metrics, and different dataset benchmark with real and/or simulated dataset from different scenarios, preventing definite conclusions to be drawn from the collection of papers on the topic.

An open-source benchmark of simulated datasets on a complete setting of different scenarios would allow for direct cross evaluation of new covariate modeling methods. Simulated scenarios found in the database included variability on eight criteria

listed earlier. On top of them, the impact of underlying structural model, interoccasion variability, functional relationships between random effects or covariates and PK parameters could also be investigated in future scenarios. Distinction between continuous and categorical covariates, and time-varying covariates might allow for better discrimination of method performances. Indeed, we did not come across articles handling time-varying covariates, although including such covariates can be essential to account for changes in patient characteristics that can affect drug pharmacokinetics over time.

Similarly, an advised set of performance metrics to universally assess predictive performance and efficiency across covariate modeling techniques is currently lacking. If applicable to the methods objective, we propose to systematically report TPR (sensitivity), FPR (Type I error), FNR (Type II error), and TNR (specificity) on covariate selection and covariate model selection, since all these metrics are complementary and inform on a methods ability to perform selection. We also propose to systematically report covariate parameter error bias (MPE) and precision (RMSE), and clinical relevance as Philipp et al. defined it [3], and model fitness by means of BIC, concentration prediction error bias (MPE), and precision (RMSE). If a method is solely dedicated to performing selection, we propose to combine it with SCM or another covariate model building method in order to compute all performance evaluation metrics. Finally, inclusion of comparison with SCM alone when evaluating another covariate model building method would allow to move away from the current standard, similarly as to what was done when FOCE was introduced, with systematic comparison of FOCE performance to FO, which led to the shift in the standard minimization algorithm.

4.2 | Promising Future Covariate Model Building Methods

ML-based methods regressed on EBEs are promising to speedup covariate model building, with reported greater efficiencies and better performances than SCM for identifying covariates. Other ML-based methods like genetic algorithms seem impractical since they often prove only to be slightly better than SCM at the cost of many more model estimation runs. Similarly as

for SCM, FOCE-linearization could be considered for methods (such as GA2) searching the covariate–parameter relationship space from a base model—and requiring parameter estimation at several points in that space—to speed up their completion. Furthermore, all categories of methods mentioned above can be seen as complementary rather than as competitive. Therefore, combinations of methods could also be investigated in the future. For instance, ML-based methods could be used prior to more classical methods. Just like Sibieude et al. [27] pointed out in their discussion, to enhance efficiency, ML could be integrated into popPK covariate model building for rapid initial covariate screening, and then incorporating the selected covariates into standard pharmacometrics approaches such as SCM, FFEM, and FREM. The challenge now lies in embedding ML into current state-of-the-art methods to boost their performance. The article from Comets et al. [37] on stepwise strategies also suggested a combination of methods for covariate modeling. Including ML-based methods as a screening step before classical pharmacometrics methods was recently done by Asiimwe et al. [28] and showed promising results. Alternatively, ML could be fully integrated in existing methods: a recent talk on the integration of LASSO in the SAMBA algorithm was given by Auriane Gabaut [38]. This new combined method promised better and quicker performances than that of SAMBA used alone. Additionally, the R package “mlcov” [39] has recently been released for covariate screening based on EBEs, and so has a python package “pyDarwin” [40] that uses several ML techniques to perform full population model selection, including the covariate one. It is also to be mentioned that ML methods require a lot of data, therefore the impact of dataset size on these methods shall be thoroughly investigated for these methods.

In the future, ML-based methods might account not only for EBEs, but also for their uncertainty. To date, most ML techniques from the literature are applied to EBEs, and the article introducing GAM stated that there were no differences applying GAM to either EBEs or individual PK parameters. These methods from the literature however never consider the uncertainty of EBEs which are often the mode of some distribution, which could lead to biased decisions. Investigations comparing model selection in stepwise procedures based on either EBEs or LRTs concluded nevertheless that both methods provided almost identical power (sensitivity) for detecting a covariate effect, even in the case of high shrinkage [41].

Author Contributions

All authors wrote the manuscript. M.K. and E.C. designed the research. M.K. performed the research. M.K. analyzed the data.

Conflicts of Interest

M. Karlsen, D. Fabre, D. Marchionni, and E. Calvier are Sanofi employees and may hold shares and/or stock options in the company. All other authors declared no competing interests for this work.

References

1. M. J. Page, J. E. McKenzie, P. M. Bossuyt, et al., “The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews,” *BMJ* 372 (2021): n71.

2. J. Ribbing, J. Nyberg, O. Caster, and E. N. Jonsson, “The Lasso—A Novel Method for Predictive Covariate Model Building in Nonlinear Mixed Effects Models,” *Journal of Pharmacokinetics and Pharmacodynamics* 34 (2007): 485–517.
3. M. Philipp, S. Buatois, S. Retout, and F. Mentré, “Impact of Covariate Model Building Methods on Their Clinical Relevance Evaluation in Population Pharmacokinetic Analyses: Comparison of the Full Model, Stepwise Covariate Model (SCM) and SCM+ Approaches,” *Journal of Pharmacokinetics and Pharmacodynamics* 51 (2024): 653–670.
4. E. N. Jonsson and M. O. Karlsson, “Automated Covariate Model Building Within NONMEM,” *Pharmaceutical Research* 15 (1998): 1463–1468.
5. G. Ayral, J.-F. S. Abdallah, C. Magnard, and J. Chauvin, “A Novel Method Based on Unbiased Correlations Tests for Covariate Selection in Nonlinear Mixed Effects Models: The COSSAC Approach,” *CPT: Pharmacometrics & Systems Pharmacology* 10 (2021): 318–329.
6. R. J. Svensson and E. N. Jonsson, “Efficient and Relevant Stepwise Covariate Model Building for Pharmacometrics,” *CPT: Pharmacometrics & Systems Pharmacology* 11 (2022): 1210–1222.
7. M. Prague and M. Lavielle, “SAMBA: A Novel Method for Fast Automatic Model Building in Nonlinear Mixed-Effects Models,” *CPT: Pharmacometrics & Systems Pharmacology* 11 (2022): 161–172.
8. M. Ahamadi, A. Largajolli, P. M. Diderichsen, et al., “Operating Characteristics of Stepwise Covariate Selection in Pharmacometric Modeling,” *Journal of Pharmacokinetics and Pharmacodynamics* 46 (2019): 273–285.
9. A. Khandelwal, K. Harling, E. N. Jonsson, A. C. Hooker, and M. O. Karlsson, “A Fast Method for Testing Covariates in Population PK/PD Models,” *AAPS Journal* 13 (2011): 464–472.
10. G. Yngman, H. B. Nyberg, J. Nyberg, E. N. Jonsson, and M. O. Karlsson, “An Introduction to the Full Random Effects Model,” *CPT: Pharmacometrics & Systems Pharmacology* 11 (2022): 149–160.
11. K. G. Kowalski and M. M. Hutmacher, “Efficient Screening of Covariates in Population Models Using Wald’s Approximation to the Likelihood Ratio Test,” *Journal of Pharmacokinetics and Pharmacodynamics* 28 (2001): 253–275.
12. Y. Zou, F. Tang, and C. M. Ng, “A Modified Hybrid Wald’s Approximation Method for Efficient Covariate Selection in Population Pharmacokinetic Analysis,” *AAPS Journal* 23 (2021): 37.
13. E. Haem, K. Harling, S. M. Taghi Ayatollahi, N. Zare, and M. O. Karlsson, “Adjusted Adaptive Lasso for Covariate Model-Building in Nonlinear Mixed-Effect Pharmacokinetic Models,” *Journal of Pharmacokinetics and Pharmacodynamics* 44 (2017): 55–66.
14. X. S. Xu, M. Yuan, H. Zhu, et al., “Full Covariate Modelling Approach in Population Pharmacokinetics: Understanding the Underlying Hypothesis Tests and Implications of Multiplicity,” *British Journal of Clinical Pharmacology* 84 (2018): 1525–1534.
15. L. F. Amann and S. G. Wicha, “Operational Characteristics of Full Random Effects Modelling (‘frem’) Compared to Stepwise Covariate Modelling (‘scm’),” *Journal of Pharmacokinetics and Pharmacodynamics* 50 (2023): 315–326.
16. M. R. Gastonguay, “A Full Model Estimation Approach for Covariate Effects: Inference Based on Clinical Importance and Estimation Precision,” *AAPS Journal* 6, no. S1 (2004): W4354.
17. M. O. Karlsson, *A Full Model Approach Based on the Covariance Matrix of Parameters and Covariates* (Athens: PAGE (Population Approach Group Europe), 2011).
18. H. Zou, “The Adaptive Lasso and Its Oracle Properties,” *Journal of the American Statistical Association* 101 (2006): 1418–1429.
19. R. R. Bies, M. F. Muldoon, B. G. Pollock, S. Manuck, G. Smith, and M. E. Sale, “A Genetic Algorithm-Based, Hybrid Machine Learning Approach to Model Selection,” *Journal of Pharmacokinetics and Pharmacodynamics* 33 (2006): 195–221.

20. D. Ronchi, E. M. Tosca, R. Bartolucci, and P. Magni, "Go Beyond the Limits of Genetic Algorithm in Daily Covariate Selection Practice," *Journal of Pharmacokinetics and Pharmacodynamics* 51 (2024): 109–121.
21. M. Ismail, M. Sale, Y. Yu, et al., "Development of a Genetic Algorithm and NONMEM Workbench for Automating and Improving Population Pharmacokinetic/Pharmacodynamic Model Selection," *Journal of Pharmacokinetics and Pharmacodynamics* 49 (2022): 243–256.
22. F. Yamashita, A. Fujita, Y. Sasa, Y. Higushi, M. Tsuda, and M. Hashida, "An Evolutionary Search Algorithm for Covariate Models in Population Pharmacokinetic Analysis," *Journal of Pharmaceutical Sciences* 106 (2017): 2407–2411.
23. E. A. Sherer, M. E. Sale, B. G. Pollock, et al., "Application of a Single-Objective, Hybrid Genetic Algorithm Approach to Pharmacokinetic Model Building," *Journal of Pharmacokinetics and Pharmacodynamics* 39 (2012): 393–414.
24. L. Aarons, S. Vozeh, M. Wenk, P. Weiss, and F. Follath, "Population Pharmacokinetics of Tobramycin," *British Journal of Clinical Pharmacology* 28 (1989): 305–314.
25. J. W. Mandema, D. Verotta, and L. B. Sheiner, "Building Population Pharmacokinetic/Pharmacodynamic Models. I. Models for Covariate Effects," *Journal of Pharmacokinetics and Biopharmaceutics* 20 (1992): 511–528.
26. U. Wählby, E. N. Jonsson, and M. O. Karlsson, "Comparison of Stepwise Covariate Model Building Strategies in Population Pharmacokinetic/Pharmacodynamic Analysis," *AAPS Journal* 4 (2002): 68–79.
27. E. Sibieude, A. Khandelwal, J. S. Hesthaven, P. Girard, and N. Teranova, "Fast Screening of Covariates in Population Models Empowered by Machine Learning," *Journal of Pharmacokinetics and Pharmacodynamics* 48 (2021): 597–609.
28. I. G. Asiimwe, B. S. Ndzamba, S. Mouksassi, G. Pillai, A. Lombard, and J. Lang, "Machine-Learning Assisted Screening of Correlated Covariates: Application to Clinical Data of Desipramine," *AAPS Journal* 26, no. 63 (2024): 63.
29. R. Hall, J. G. Pasipanodya, M. A. Swancutt, C. Meek, R. D. Leff, and T. Gumbo, "Supervised Machine-Learning Reveals That Old and Obese People Achieve Low Dapsone Concentrations," *CPT: Pharmacometrics & Systems Pharmacology* 6 (2017): 552–559.
30. M. E. Ibrahim, J. R. Castillo-Mancilla, and J. Yager, "Individualized Adherence Benchmarks for HIV Pre-Exposure Prophylaxis," *AIDS Research and Human Retroviruses* 37, no. 6 (2021): 421–428.
31. B.-H. Tang, Z. Guan, K. Allegaert, et al., "Drug Clearance in Neonates: A Combination of Population Pharmacokinetic Modelling and Machine Learning Approaches to Improve Individual Prediction," *Clinical Pharmacokinetics* 60 (2021): 1435–1448.
32. A. Janssen, M. Hoogendoorn, M. H. Cnossen, and R. A. A. Mathôt, "Application of SHAP Values for Inferring the Optimal Functional Form of Covariates in Pharmacokinetic Modeling," *CPT: Pharmacometrics & Systems Pharmacology* 11 (2022): 1100–1110.
33. N. Semmar, B. Bruguerolle, S. Boullu-Ciocca, and N. Simon, "Cluster Analysis: An Alternative Method for Covariate Selection in Population Pharmacokinetic Modeling," *Journal of Pharmacokinetics and Pharmacodynamics* 32 (2005): 333–358.
34. Y. Wahlquist, J. Sundell, and K. Soltesz, "Learning Pharmacometric Covariate Model Structures With Symbolic Regression Networks," *Journal of Pharmacokinetics and Pharmacodynamics* 51 (2024): 155–167.
35. D. S. Bräm, U. Nahum, A. Atkinson, G. Koch, and M. Pfister, "Evaluation of Machine Learning Methods for Covariate Data Imputation in Pharmacometrics," *CPT: Pharmacometrics & Systems Pharmacology* 11 (2022): 1638–1648.
36. K. Sanghavi, J. Ribbing, J. A. Rogers, et al., "Covariate Modeling in Pharmacometrics: General Points for Consideration," *CPT: Pharmacometrics & Systems Pharmacology* 13, no. 15 (2024): 710–728.
37. E. Comets, C. Verstyft, M. Lavielle, P. Jaillon, L. Becquemont, and F. Mentré, "Modelling the Influence of MDR1 Polymorphism on Digoxin Pharmacokinetic Parameters," *European Journal of Clinical Pharmacology* 63 (2007): 437–449.
38. A. Gabaut and M. Prague, "High-Dimension Mechanistic Model Building Using LASSO Approaches: Application to Ebola Vaccination," in *Population Approach Group Europe (PAGE)* (Rome: PAGE (Population Approach Group Europe), 2024).
39. I. Rebai, V. Duval, A. Akil, et al., *mlcov: New Machine Learning Based R Package for Covariate Selection* (Rome: PAGE (Population Approach Group Europe), 2024).
40. L. Xinnong, M. Sale, K. Nieforth, et al., "pyDarwin: A Machine Learning Enhanced Automated Nonlinear Mixed-Effect Model Selection Toolbox," *CPT* 115, no. 14 (2024): 758–773.
41. X. S. Xu, M. Yuan, H. Yang, Y. Feng, J. Xu, and J. Pinheiro, "Further Evaluation of Covariate Analysis Using Empirical Bayes Estimates in Population Pharmacokinetics: The Perception of Shrinkage and Likelihood Ratio Test," *AAPS Journal* 19 (2017): 264–273.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.