Stepwise Covariate Modeling Acceleration Using Individual Uncertainty Adjusted Linear Covariate Screening Method

Peng Man[1], Yevgen Ryeznik[1], Xiaomei Chen[1], Rikard Nordgren[1], Stella Belin[1], Chee M. Ng[2], Robert J. Bauer[3], Mats O. Karlsson[1], and Andrew C. Hooker[1]

[1]Department of Pharmacy, Uppsala University, Sweden [2]NewGround Pharm Consulting LLC, Foster City, CA, USA 3Pharmacometrics, R&D, ICON Clinical Research, LLC, Gaithersburg, Maryland, USA

Introduction

Stepwise covariate modeling (SCM) offers an automated, data-driven approach for covariate model selection via forward inclusion and backward elimination (1). However, its computational efficiency is limited by the need to evaluate numerous covariate effects through nonlinear mixed effects models. To address this, Stochastic Approximation for Model Building Algorithm (SAMBA) was proposed (2). SAMBA fits linear covariate models using Monte Carlo samples of individual random effects to select covariate effects, thereby reducing nonlinear model evaluations. By leveraging samples from SAEM-derived ETA distributions, SAMBA is believed to be less prone to the impact of high shrinkage than methods based on empirical Bayes estimates (EBE) (3). In this study, we applied EBE-based linear covariate screening (LCS) to SCM and developed two methods: SCM using non-stepwise LCS (SCM-LCS) and SCM using stepwise LCS (SCM-SLCS). We evaluated their performance under scenarios varying in shrinkage levels, covariate effect magnitudes, sample sizes, and linear model fitting methods.

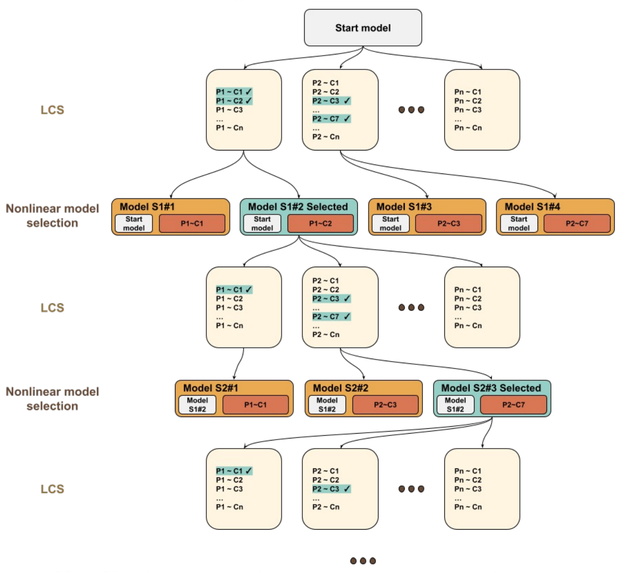


Figure 1. SCM-LCS: Using linear covariate screening as a scope reduction strategy.

Methods

Data were simulated using a one-compartment pharmacokinetic (PK) model with first-order absorption and linear elimination. Six base scenarios were created by varying sampling designs and adjusting the ratio of interindividual variability (IIV) to residual unexplained variability (RUV) to achieve different levels of ETA shrinkage (4). Covariates were sampled from the 2017-2020 NHANES database (5). A true covariate effect of body weight on volume of distribution was implemented with a power model. For each scenario, 500 replicate datasets with 500 individuals each were generated. Unless otherwise stated, SCM methods used likelihood ratio test (LRT), SAMBA used corrected BIC (dividing the log-likelihood by number of samples), and other methods used BIC for model selection.

Performance was assessed with single-covariate search space for power and type I error rate, and with multiple-covariate search space for true positive rate (TPR), true negative rate (TNR), and inclusion frequency.

- Model-level TPR: % of times the exact true model is recovered when a true covariate effect exists
- Model-level TNR: % of times the covariate-free model is retained when no covariate effect exists
- Covariate-level TPR: % of times the true covariate effect is included, regardless of spurious inclusion
- Covariate-level TNR: % of times no spurious covariates are included, regardless of true effect inclusion

Computational efficiency was evaluated using CPU runtime. All methods were implemented in Pharmpy (6) and NONMEM.

S1: parameter set with large IIV-to-RUV ratio (ω:σ=0.5:0.3); S2: parameter set with small IIV-to-RUV ratio (ω:σ=0.2:0.4); N5: rich design with 5 sampling points; N3: sparse design with 3 sampling points; N2: very sparse design with 2 sampling points

Results

An increase in shrinkage of ETA_{VC} and ETA_{CL} from scenario S1N5 to S2N2 was achieved, ranging from 8.97% to 55.7% and 8.88% to 51.6% respectively.

Performance of Benchmark Methods

SCM-FORWARD maintained high power across scenarios but showed lower specificity. SAMBA showed high specificity and more cautious covariate inclusion, with reduced sensitivity in high shrinkage scenarios (S2N2). Changing model selection criterion to LRT improved SAMBA's power, but at the expense of specificity.

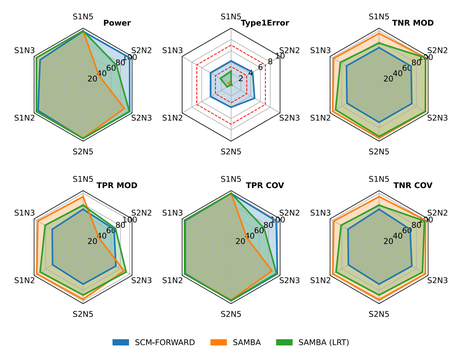


Figure 2. Benchmark method performance across all scenarios with strong covariate effect magnitude. Red dashed lines in the type I error radar chart indicate the prediction interval around 5% type I error rate.

Table 1. Average performance over scenarios.

	Covaraite Effect Size	RA_{Power} (%)	$RA_{Type1Error}$ (%)	$RA_{TPR,MOD}$ (%)	$RA_{TNR,MOD}$ (%)	$RA_{TPR,COV}$ (%)	$RA_{TNR,COV}$ (%)
SCM-FORWARD	Strong	95.79	34.68	46.09	48.12	97.35	46.27
SAMBA	Strong	75.22	0.15	71.72	97.09	74.51	95.58
SAMBA (LRT)	Strong	90.30	2.20	70.12	83.70	88.93	78.19
SCM-FORWARD	Weak	61.04	-	25.74	-	51.21	41.80
SAMBA	Weak	18.38	-	15.89	-	16.78	93.34
SAMBA (LRT)	Weak	31.81	-	20.73	-	28.24	77.70

RA, relative area. For type I error, the reference total area was the area of a regular hexagon with a side equal to the upper bound of the prediction interval; thus, an area exceeding 100% suggests an issue of type I error inflation.

Performance of SCM-(S)LCS Methods

SCM-LCS and SCM-SLCS achieved high power and well-controlled type I error, comparable to SCM-FORWARD. SCM-LCS improved specificity.

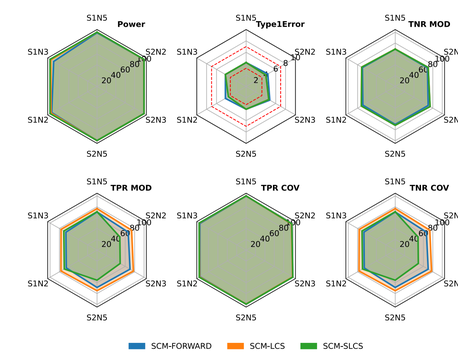


Figure 3. SCM-(S)LCS performance across all scenarios with strong covariate effect magnitude. Each radar chart represents one performance metric.

Table 2. Average performance over scenarios.

	Covaraite Effect Size	RA_{Power} (%)	$RA_{Type1Error}$ (%)	$RA_{TPR,MOD}$ (%)	$RA_{TNR,MOD}$ (%)	$RA_{TPR,COV}$ (%)	$RA_{TNR,COV}$ (%)
SCM-FORWARD	Strong	95.79	34.68	46.09	48.12	97.35	46.27
SCM-LCS	Strong	98.67	31.51	57.70	51.03	97.81	57.75
SCM-SLCS	Strong	99.93	31.51	37.21	50.89	98.67	37.21
SCM-FORWARD	Weak	61.04	-	25.74	-	51.21	41.80
SCM-LCS	Weak	61.11	-	30.37	-	50.52	48.55
SCM-SLCS	Weak	61.94	-	24.61	-	52.24	40.78

Computational Efficiency

SCM-LCS and SCM-SLCS substantially reduced runtime, highlighting the efficiency of linear covariate screening.

Table 3. Computational efficiency measured as average CPU runtime in minutes.

	Mean	SD	Min	Median	Max
SCM-FORWARD	12.94	5.02	3.77	11.88	43.87
SAMBA	17.48	5.45	5.04	17.26	40.15
SAMBA (LRT)	18.74	5.09	5.63	18.58	42.56
SCM-LCS	5.37	1.75	2.83	5.09	16.37
SCM-SLCS	4.12	0.79	2.67	4.15	7.94
SCORE	40.47	10.14	18.76	39.48	107.50
WAM	23.68	5.55	10.17	22.79	54.27

Individual Uncertainty in LCS

Incorporating individual uncertainty—via weighted least squares (WLS) and linear mixed effects model (LME)—improved SAMBA's power but had minimal effect on SCM-LCS performance. All methods maintained controlled type I error.

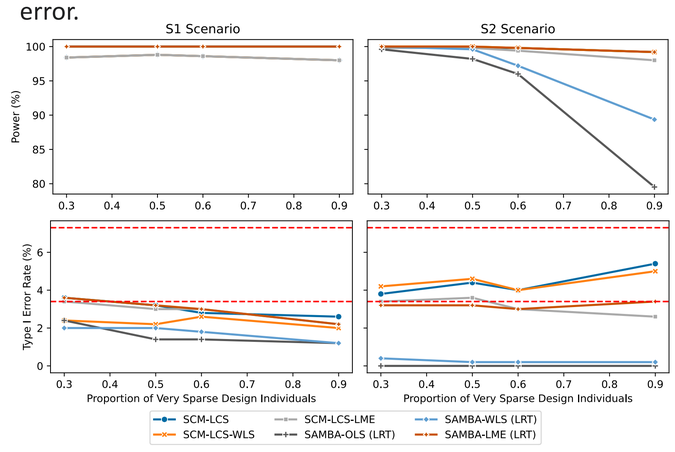


Figure 4. Change of power and type I error rate as the proportion of very sparse design individuals increased. All methods used likelihood ratio test (LRT) for model selection to ensure comparability.

Overall Performance Comparison

All methods showed reduced performance as sample size decreased, or shrinkage increased. SAMBA achieved the highest performance under low shrinkage and large sample size scenarios but declined markedly with smaller sample sizes or sparser designs. SCM-LCS demonstrated relatively stable performance across varying shrinkage and sample size, and its BIC variant further improved specificity and overall performance.

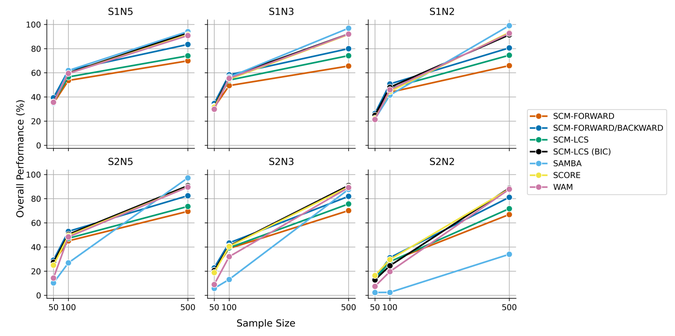


Figure 5. Overall performance under various sample size and shrinkage conditions. Overall performance was measured as the relative area formed by power, 1 - type I error, model-level TPR, model-level TNR, covariate-level TPR, and covariate-level TNR within each scenario.

Conclusions

- SCM-LCS and SCM-SLCS demonstrated improved computational efficiency and maintained comparable power to SCM-FORWARD, even under challenging conditions
- SCM-LCS and SCM-FORWARD/BACKWARD consistently outperformed SCM-FORWARD, primarily due to improved specificity achieved through preliminary LCS and backward elimination
- BIC further enhanced specificity of SCM-LCS, yielding performance comparable to other BIC-driven methods
- Incorporating individual uncertainty in LCS informs the extent of ETA shrinkage, but added minimal benefit
- SAMBA performed well in easy scenarios but behaved conservatively in covariate effect selection under tough scenarios—being best at avoiding false positives.
- Across tested scenarios, conditional samples offered no clear advantage over EBEs for covariate effect screening
- To address the informativeness issue, future methods should account for both extent and direction of ETA shrinkage

Acknowledgments

I would like to thank my supervisors, Prof. Andrew Hooker and Dr. Yevgen Ryeznik, for their guidance and support throughout this project. I also thank the Pharmpy team and Dr. Robert Bauer for their valuable discussions and feedback. The computations were enabled by resources at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

UPPSALA UNIVERSITET

References

1. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. Pharm Res. 1998 Sep;15(9):1463–8.
2. Prague M, Lavielle M. SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. CPT: Pharmacometrics & Systems Pharmacology. 2022;11(2):161–72.
3. Lavielle M, Ribba B. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. Pharm Res. 2016 Dec;33(12):2979–88.
4. Combes FP, Retout S, Frey N, Mentré F. Powers of the Likelihood Ratio Test and the Correlation Test Using Empirical Bayes Estimates for Various Shrinkages in Population Pharmacokinetics. CPT Pharmacometrics Syst Pharmacol. 2014 Apr;3(4):e109.
5. National Health and Nutrition Examination Survey, Questionnaires, Datasets, and Related Documentation, 2017-March 2020 Pre-Pandemic Data. Centers for Disease Control and Prevention. Accessed 20 Nov 2025
6. Chen X, Nordgren R, Belin S, Hamdan A, Wang S, Yang T, et al. A fully automatic tool for development of population pharmacokinetic models. CPT: Pharmacometrics & Systems Pharmacology. 2024;13(10):1784–97.