Operating Characteristics of Stepwise Covariate Selection in Pharmacometric Modeling

Ahamadi M[1*]; Anna Largajolli[2*]; Paul M Diderichsen[2]; Rik de Greef[2]; Thomas Kerbusch[2]; Han Witjes[2]; Akshita Chawla[1]; Casey B Davis[1]; Ferdous Gheyas [1]

[1]Pharmacokinetics, Pharmacodynamics & Drug Metabolism, MRL, Merck & Co., Inc., Kenilworth, NJ, USA; [2]Certara Strategic Consulting
*Equal contributions

Introduction

Stepwise covariate modeling (SCM) is a widely used tool in pharmacometric analyses to identify covariates that explain between subject variability (BSV) in exposure and exposure-response relationships. However, potential weaknesses of this approach include over-estimated covariate effects [1] and incorrect selection of covariates due to collinearity [2].

Objectives

In this work we have investigated the operating characteristics of SCM in a controlled simulated setting in order to assess the impact of the effect over-estimation and collinearity on covariate inclusion.

Methods

Model

A two-compartment model with first-order absorption was coupled with sixteen different covariates relations (*scenarios*) obtained by permuting four covariates (body weight (BW) and creatinine clearance (CrCL) on apparent clearance, BW and SEX on volume of distribution - *Table I*).

Scenario	θ_1: CL(WGT)	θ_2: CL(CRCL)	θ_3: V(WGT)	θ_4: V(SEX)
1	0	0	0	0
2	0	0	0	0.5
3	0	0	1	0
4	0	0	1	0.5
5	0	0.5	0	0
6	0	0.5	0	0.5
7	0	0.5	1	0
8	0	0.5	1	0.5
9	0.75	0	0	0
10	0.75	0	0	0.5
11	0.75	0	1	0
12	0.75	0	1	0.5
13	0.75	0.5	0	0
14	0.75	0.5	0	0.5
15	0.75	0.5	1	0
16	0.75	0.5	1	0.5

Table I: Map of covariate coefficients in 16 simulation scenarios

$$CL_i = TVCL\left(\left(\frac{Weight}{70\,kg}\right)^{\theta_1}\cdot\left(\frac{CrCL}{95}\right)^{\theta_2}\right)\cdot\exp(\eta_{CL,i})$$

$$Vc_i = TVVc\left(\left(\frac{Weight}{70\,kg}\right)^{\theta_3}\cdot(1+sex\cdot\theta_4)\right)\cdot\exp(\eta_{Vc,i})$$

$$\begin{bmatrix}\eta_{CL,i}\\ \eta_{Vc,i}\end{bmatrix}\sim N(0,\Omega),\ \Omega=\begin{bmatrix}\omega_{CL}^2 & cov(CL,Vc)\\ cov(CL,Vc) & \omega_{Vc}^2\end{bmatrix}$$

Simulated data

250 datasets were simulated for each scenario with a sample size of 300 subjects and 6 observations per subject (t=0,0.05,0.1,0.5,1,3). Virtual patients defined by 5 covariates (BW, BMI, CrCL, SEX, RACE) were bootstrapped from the NHANES dataset [3].

Analysis

The identifiability of scenarios was assessed by stochastic simulation and estimation (SSE). For each scenario, the relative mean root squared error (RMRSE) of parameter estimates and model stability information (convergence, covariance step, condition number) was derived. Subsequently, each scenario was analyzed by a full SCM procedure and the power to select the true covariate model and RMRSE were derived. Note that initially were used the default SCM boundary conditions on the covariate parameters.

Software

Each scenario was analyzed by a full SCM procedure, as implemented in PsN 4.6.0 [4] coupled with NONMEM 7.2 [5].

Correlation between covariates within simulated dataset

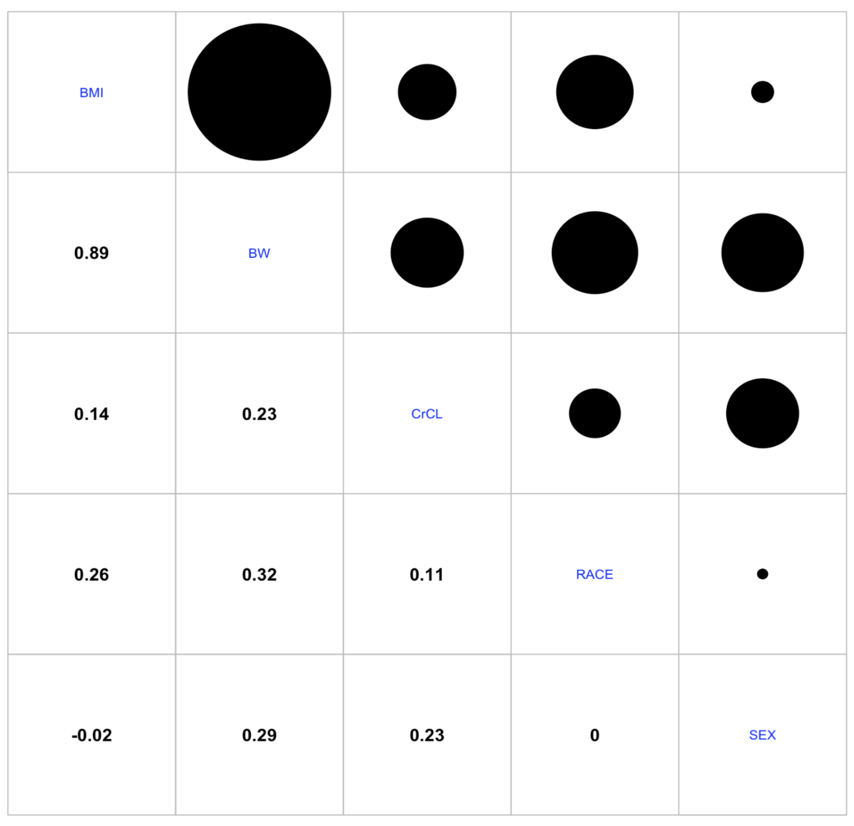

Figure 1: Pearson correlation matrix from bootstrapped covariates of 250x300 simulated subjects

- Covariates were generated based on bootstrap from NHANES dataset (~3000 subjects)
  - Race was dichotomized in Asian/ non-Asian (white)
  - Age >18 yrs
  - Missing values/missing IDs among different dataset --> ignored subjects
- In the figure: size of the circles are proportional to the correlation coefficients.

Strong correlation between BMI and BW [89%] (Figure 1)

Stability of simulated dataset

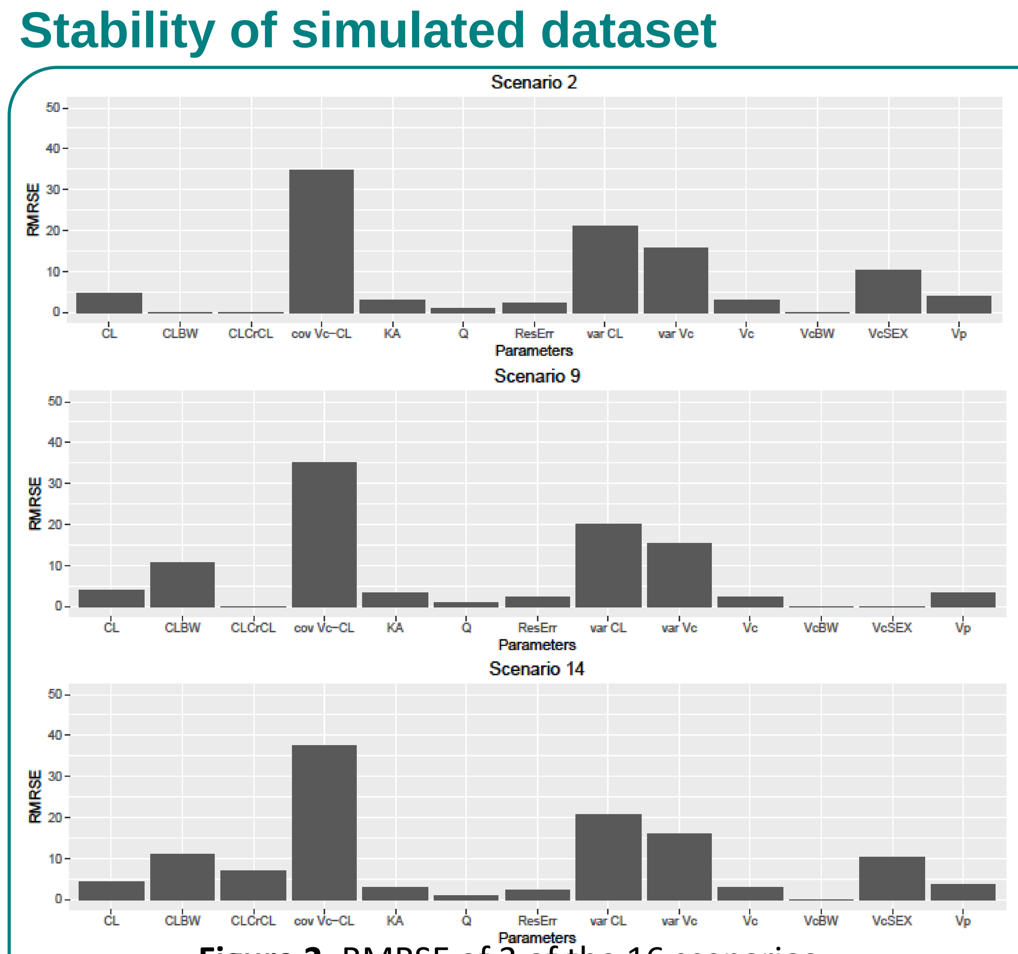

Figure 2: RMRSE of 3 of the 16 scenarios

$$*\,RMRSE_{par_k} = \sqrt{\frac{1}{250}\sum_{i=1}^{250}\left(\frac{par_k^{obs}-par_{k_i}^{est}}{par_k^{obs}}\right)^2}$$

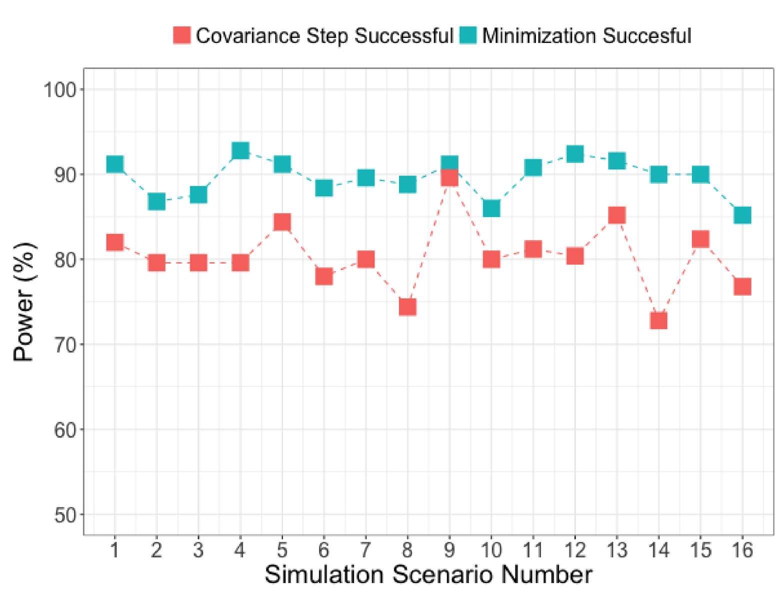

Figure 3: Model stability information: percentage of models with minimization successful and covariance step successful

The RMRSE(*) (Figure 2) in selected scenarios and model stability parameters (Figure 3) confirmed that all scenarios could be estimated and were numerically stable.

Power to obtain the correct final model after SCM procedure

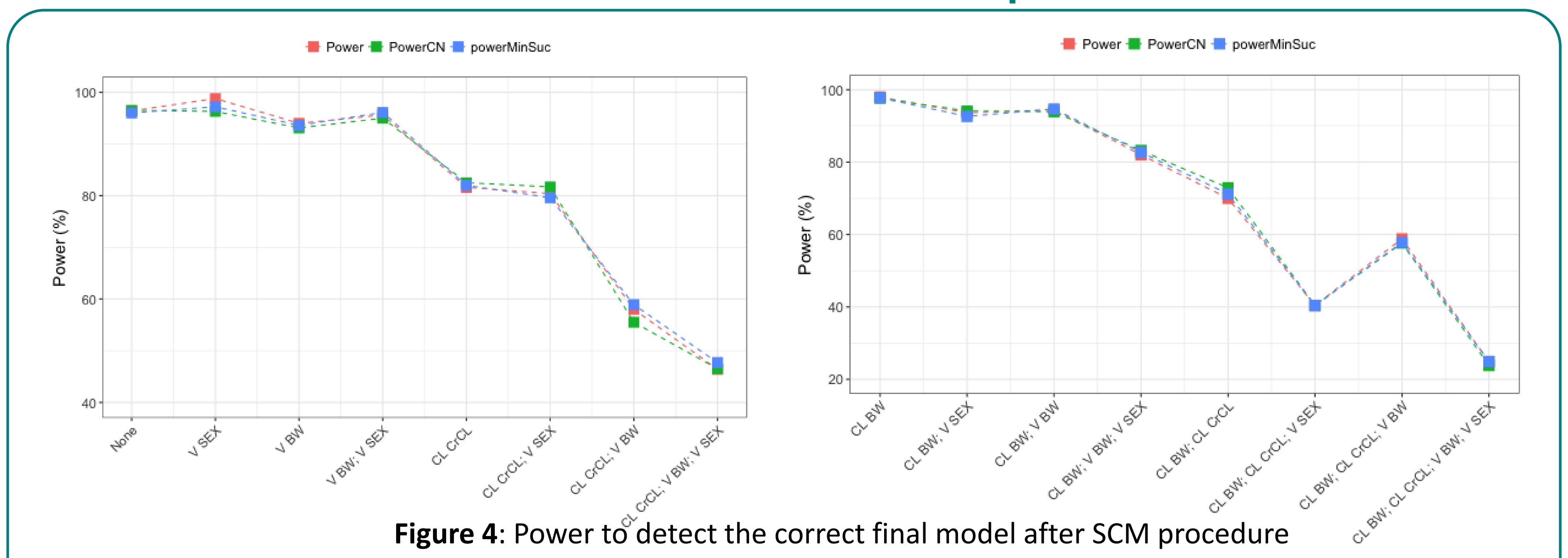

Figure 4: Power to detect the correct final model after SCM procedure

- ***PowerCN***: power conditioned on the condition number (CN); i.e. based only on datasets in which the true model had CN<1000 in the SSE
- ***PowerMinSuc***: power conditioned on the minimization successful; i.e. based only on datasets in which the true model converged successfully in the SSE

Estimated power of SCM decreases dramatically as the complexity of the true model increases

Summary of false relation detected during SCM

	true relations	*most frequent false relation*	*freq of the false relation*
scenario 1	none	"Vc SEX " & "CL RACE "	2/250
scenario 2	Vc SEX	"CL RACE Vc SEX "	3/250
scenario 3	Vc BW	"Vc BMI Vc BW "	4/250
scenario 4	Vc BW; Vc SEX	"CL RACE Vc SEX Vc BW "	3/250
scenario 5	CL CrCL	"CL BMI CL CrCL "	10/250
scenario 6	CL CrCL; Vc SEX	"CL BW Vc SEX CL CrCL "	8/250
scenario 7	CL CrCL; Vc BW	"Vc CrCL Vc BW CL CrCL "	19/250
scenario 8	CL CrCL; Vc BW; Vc SEX	"CL CrCL Vc BW "	20/250
scenario 9	CL BW	"Vc RACE CL RACE CL BW"&"CL BMI" & "Vc SEX CL BW"&"CL RACE CL BW"&"CL SEX CL BW"	1/250
scenario 10	CL BW; Vc SEX	"CL BMI CL BW Vc SEX "	5/250
scenario 11	CL BW; Vc BW	"Vc CrCL CL BW Vc BW "	5/250
scenario 12	CL BW; Vc BW: Vc SEX	"CL BMI CL BW Vc SEX Vc BW "	24/250
scenario 13	CL BW; CL CrCL	"CL BMI CL CrCL "	56/250
scenario 14	CL BW; CL CrCL; Vc SEX	"CL BMI Vc SEX CL CrCL "	52/250
scenario 15	CL BW; CL CrCL; Vc BW	"Vc BW CL BMI CL CrCL "	30/250
scenario 16	CL BW; CL CrCL; Vc BW: Vc SEX	"CL BMI Vc BW Vc SEX CL CrCL "	33/250

Table II: summary of the most frequent false relations detected in each scenario

- The most frequent false relations and its frequency are reported; in bold are underlined the false covariate selected.

Often the wrong/additional covariate selected is a correlated covariate (i.e. BMI instead of BW)

SCM full results – estimates RMRSE

	n relations	Vc	CL	Vp	Q	KA	Err	ω1-Vc	cov (Vc,CL)	ω2-CL	Vc-BW	Vc -SEX	CL - BW	CL - CrCL
scenario 1	0	2.4%	4.1%	3.3%	0.9%	3.2%	2.2%	12.5%	22.2%	15.7%	-	-	-	-
scenario 2	1	2.8%	4.6%	4.0%	0.9%	3.0%	2.1%	12.7%	22.1%	16.1%	-	9.7%	-	-
scenario 3	1	2.5%	4.2%	3.3%	0.9%	2.9%	2.1%	12.7%	23.0%	16.0%	6.5%	-	-	-
scenario 4	2	2.8%	4.6%	3.8%	1.0%	2.7%	2.2%	12.9%	23.4%	16.3%	7.5%	10.4%	-	-
scenario 5	1	2.5%	3.7%	4.1%	1.1%	3.3%	2.2%	12.4%	30.0%	27.9%	-	-	-	23.0%
scenario 6	2	10.0%	4.8%	6.1%	1.1%	4.4%	2.2%	27.5%	39.9%	23.4%	-	22.8%	-	34.6%
scenario 7	2	4.1%	4.1%	5.2%	1.1%	3.5%	2.1%	26.0%	42.0%	23.6%	7.2%	-	-	25.1%
scenario 8	3	8.7%	4.4%	6.1%	1.1%	4.0%	2.2%	38.8%	54.2%	22.2%	13.0%	30.1%	-	39.9%
scenario 9	1	2.4%	4.3%	3.1%	0.9%	3.2%	2.1%	12.4%	22.2%	14.8%	-	-	12.9%	-
scenario 10	2	3.1%	4.8%	3.9%	0.9%	2.9%	2.1%	12.8%	22.3%	13.9%	-	10.5%	16.0%	-
scenario 11	2	2.6%	5.0%	3.3%	0.9%	2.9%	2.1%	11.9%	22.0%	12.1%	8.0%	-	20.6%	-
scenario 12	3	3.1%	5.3%	3.8%	1.0%	2.6%	2.2%	12.3%	23.5%	13.1%	8.8%	11.5%	26.4%	-
scenario 13	2	2.4%	6.8%	4.9%	1.1%	3.1%	2.1%	12.1%	34.1%	37.7%	-	-	7.6%	33.3%
scenario 14	3	10.3%	6.8%	7.8%	1.2%	4.4%	2.2%	26.8%	65.5%	43.1%	-	24.6%	12.0%	43.5%
scenario 15	3	4.9%	5.4%	5.0%	1.2%	3.5%	2.2%	24.4%	129.3%	49.2%	11.0%	-	11.7%	38.1%
scenario 16	4	10.6%	6.5%	7.0%	1.1%	3.8%	2.3%	38.2%	152.3%	58.5%	15.3%	25.2%	20.2%	47.9%

Table III: RMRSE of the model parameter common to the 16 scenarios

RMRSE is low for fixed effects not relative to covariate effects; The BSV variances and fixed effect relative to covariates increase dramatically with complexity of the true model;

Impact of default boundary condition provided by SCM in power relations

Results relative to dataset 1 of scenario 16

During third forward step, **both loose and strict bound condition select V2BW-6**

MODEL	OFV	loose bound cond NEW OFV	loose bound cond (DROP)	strict bound cond NEW OFV	strict bound cond (DROP)
CLBMI-6		-3287.86	0.06	-3389.03	22.94
CLBW-6		-3296.69	8.89	-3422.25	56.16
CLRACE-6		-3288.83	1.02	-3366.31	0.22
CLSEX-6		-3288.31	0.50	-3378.1	12.02
V2BMI-6		-3313.46	25.66	-3460.96	94.87
V2BW-6		***-3372.15***	***84.34***	***-3464.48***	***98.39***
V2CrCL-6		-3288.18	0.37	-3374.51	8.42
V2RACE-6		-3287.23	-0.58	-3375.48	9.39

During 4th forward step, **V2CrCL-6 was chosen by loss bound cond** whilst **CLBW-6 was chosen by strict boundary condition**

MODEL	OFV	loose bound cond NEW OFV	loose bound cond (DROP)	strict bound cond NEW OFV	strict bound cond (DROP)
CLBMI-6		-3372.26	0.11	-3548.26	83.78
CLBW-6		-3381.65	9.51	***-3597.68***	***133.20***
CLRACE-6		-3372.45	0.31	-3469.05	4.57
CLSEX-6		-3372.37	0.23	-3473.89	9.41
V2BMI-6		-3372.18	0.03	-3468.04	3.56
V2CrCL-6		***-3411.15***	***39.00***	-3465.37	0.89
V2RACE-6		-3373.54	1.39	-3464.53	0.05

Example of LOOSE BOUNDARY for power model (default PsN)

```
$THETA  (-100000,0.5,100000) ; CLCrCL1
```

Example of adjusted STRICT BOUNDARY

```
$THETA  (-10.00,0.5, 10.00) ; CLCrCL1
```

Simulation showed that default boundary condition provided by SCM lead to higher initial gradient (greater model instability) which influence the choice of covariates.

New boundary condition and sample size investigation

power	*stricter boundary (new Proposal)* 600 subj*	300 subj**	150 subj*	n relations
scenario 5	0.92	0.932	0.888	1
scenario 7	0.936	0.916	0.896	2
scenario 12	0.944	0.944	0.816	3
scenario 16	0.928	0.91	0.84	4

** based on 125 dataset*
*** based on 250 dataset*

Table IV: power to detect the correct final model after model SCM procedure with respect to different sample size

New boundaries condition helped the power to improve. First sample size investigation confirm that the more subjects we have the higher the power.

Conclusions

Model complexity has a great impact on the power to identify the true covariate model and on the accuracy and precision of the parameter estimates

- Default boundary condition handling provided by SCM for power model in PsN have impact on the selection of covariates during the screening.
- Highly correlated covariates have high likelihood to be wrongly selected by SCM.
- In general, all RMRSE tend to increase with model complexity and the power to decrease.

References

1. Ribbing J, Jonsson EN. Power, selection bias and predictive performance of the population pharmacokinetic covariate model. J Pharmacokinet Pharmacodyn. 2004;31(2):109–134.
2. Bonate PL. The effect of collinearity on parameter estimates in nonlinear mixed effect models. Pharm Res. 1999;16(5):709–717.
3. CDC - National Center for Health Statistics - Homepage. http://www.cdc.gov/nchs/.
4. L. Lindbom, P. Pihlgren, and E. N. Jonsson. Psntoolkit: a collection of computer intensive statistical methods for non-linear mixed effect modeling using nonmem. Computer Methods and Programs in Biomedicine, 79(3):241, Sept. 2005
5. Beal SL, Sheiner LB, Boeckmann AJ & Bauer RJ (Eds.) NONMEM Users Guides. 1989-2017. Icon Development Solutions, Ellicott City, Maryland, USA