Characterization of Stepwise Covariate Model Building Combined with Cross-Validation

Takayuki Katsube[1,2], Akash Khandelwal[1], Andrew C Hooker[1], E. Niclas Jonsson[1], Mats O Karlsson[1]

(1) Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

(2) Clinical Research Department, Shionogi & Co., Ltd., Japan

Background

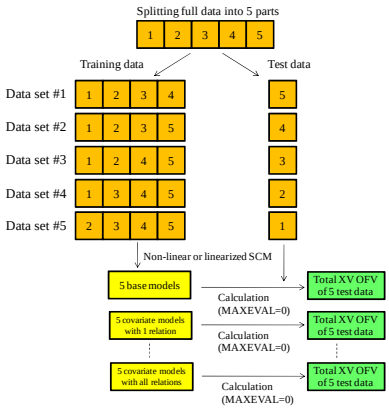
A stepwise covariate model building combined with cross validation (XV SCM) was reported [1]. This method is useful to determine suitable model size based on prediction error using multiple data

Objective

The objective of this study is to further characterize XV SCM and make comparisons of estimated model sizes and predictive performance of developed models with other covariate modeling methods, standard SCM [2] and lasso [3].

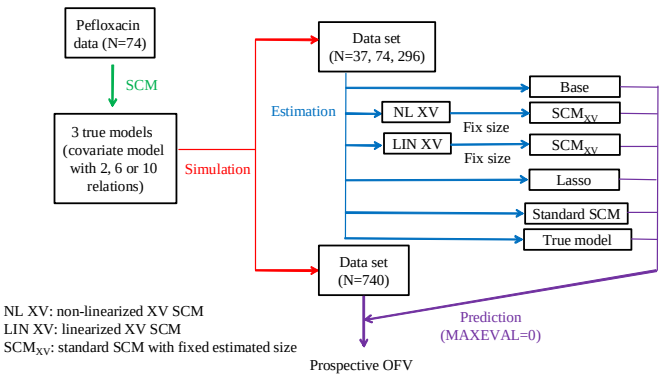
Methods

Ten times 5-fold XV SCM was used.



- In 5-fold XV, repeatedly 80% of data (training data) is used to develop a model which subsequently is used to predict XV OFV in left out 20% of data (test data).
- Total XV OFV of 5 test data was calculated for each model size.
- The random splits of full data was performed 10 times.
- The number of relations where the mean XV OFV over splits was minimal was taken to be an appropriate model size in this procedure.

Figure 1. Scheme of 5-fold XV SCM.



- Simulated data were generated using SCM models (true models) with certain relations (2, 6 and 10) based on real PK data of pefloxacin.
- 9 simulation scenarios (3 model sizes x 3 covariate models).
- Standard SCM used forward addition ($p<0.05$) and backward deletion ($p<0.01$).
- The effect of the number of splits on the predictive performance using 5-fold XV SCM was assessed by changing the number of splits.

Figure 2. Strategy for evaluation of predictive performance.

Real PK data of pefloxacin

- 337 observations of plasma concentrations from 74 patients
- 1-compartment model with IIV and IOV on CL and V
- 14 test relations on CL and V – 5 continuous (weight, age, CLCR, bilirubin and systolic blood pressure) and 2 dichotomous (sex and center) covariates.

Software

XV SCM, standard SCM and lasso were performed using PsN ver. 3.4.2 or higher.

References

[1] Katsube T, Khandelwal A, Harling K, Hooker AC, Karlsson MO. PAGE2011.
[2] Jonsson EN, Karlsson MO. Pharm Res. 1998. 15:1463.
[3] Ribbing J, Nyberg J, Caster O, Jonsson EN. J PK PD. 2007. 34:485.

Conclusions

- XV SCM provides a suitable covariate model size, typically the same as the true size, with good predictive performance except for extremely small data.
- XV SCM performs well even with few splits.

Results

- Distributions of XV OFV profiles were characterized by using 10 times 5-fold XV SCM (non-linear/linearized) and 3 times 5-fold XV lasso (Fig. 3).
- All methods improved the prospective OFV compared to the base model (Fig. 4). Predictive performance was comparable for different methods except for the small data.
- The typical model sizes of final models when using XV SCM were the same as the true sizes (Fig. 5). The sizes when using lasso and standard SCM were 2 relations more and 1 relation less relative to the true sizes, respectively.
- The predictive performance of the XV methods was similar regardless of the number of splits (Fig. 6).

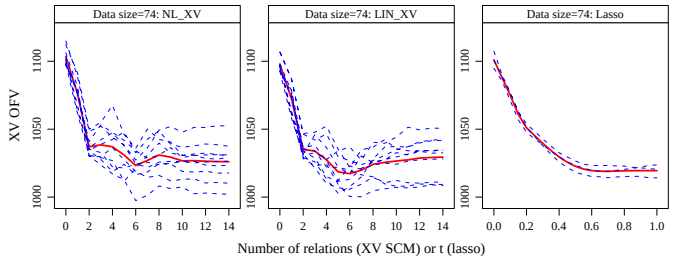


Figure 3. XV OFV profiles for a single simulated data based on covariate models with 6 true relations by method. Red line: mean XV OFV over splits. Blue line: individual XV OFV.

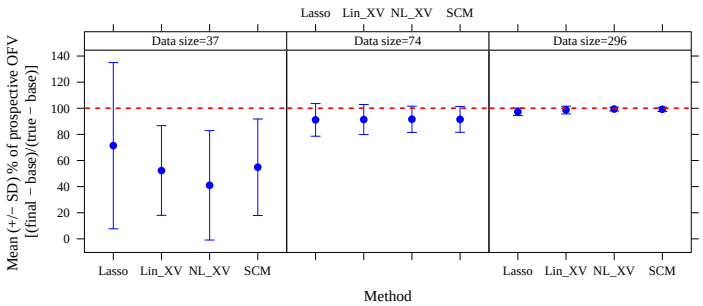


Figure 4. Predictive performance in percent [(final – base)/(true – base)] by simulated data size using each method.

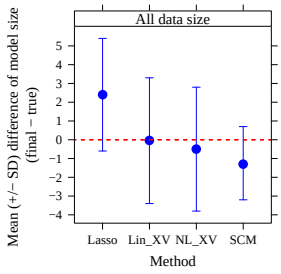


Figure 5. Differences of model size (final – true) using each method.

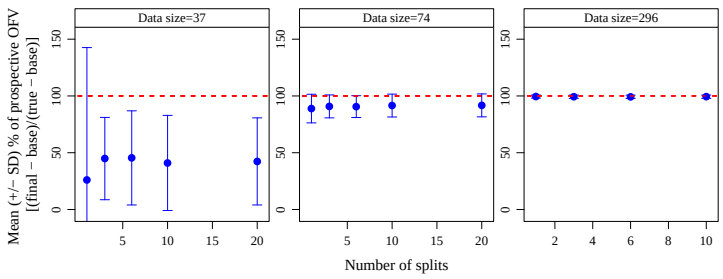


Figure 6. Effect of number of splits on predictive performance using non-linear XV SCM by simulated data size.