Evaluation of Stepwise Covariate Model Building Combined with Cross-Validation

Takayuki Katsube[1, 2], Akash Khandelwal[1], Kajsa Harling[1], Andrew C Hooker[1], Mats O Karlsson[1]
(1) Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden
(2) Clinical Research Department, Shionogi & Co., Ltd., Japan

Background

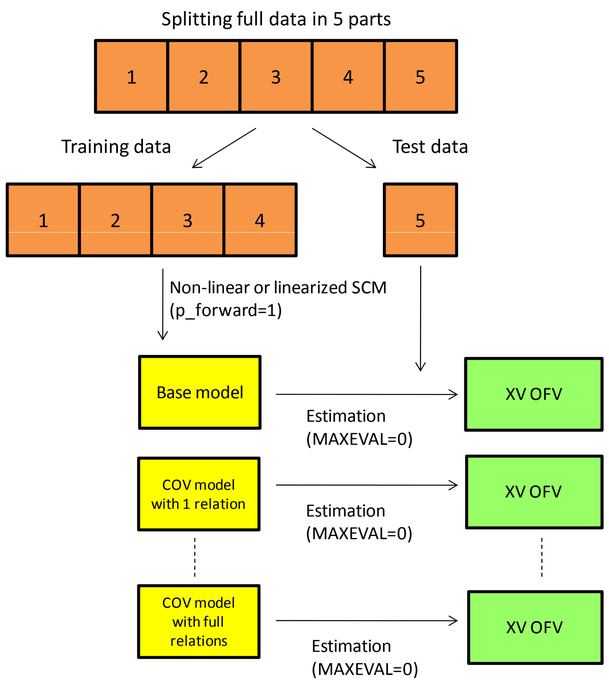
Covariate models are often built using a stepwise covariate model building (SCM), a procedure which is not intrinsically designed for providing good predictive performance. Cross-validation (XV) is a procedure for estimating the prediction error using multiple subsets of a dataset and may be used to select an appropriate model size [1]. If the main goal is predictive modeling, SCM combined with XV (XV SCM) for determining a model size may be useful.

Objective

The objective of this study was to evaluate XV SCM for determining a model size using both linearized [2] and non-linear models.

Methods

Ten times 5-fold XV SCM was used for simulated data and real data.



- Each part was used as test data.
- The sum of XV OFV for 5 test data was calculated by covariate model.
- 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.

Figure 1. Scheme of 5-fold XV SCM.

Simulated data

Simulated data were generated using SCM models with certain relations (0, 2, 4, 6, 10 and 14) based on real data of pefloxacin (true models).

Real data

Phenobarbital

- 155 observations of plasma concentrations from 59 subjects
- 1-compartment model with inter-individual variability (IIV) on CL and V.
- 4 test relations – 2 continuous covariates (weight and APGAR score) on CL and V.

Moxonidine

- 1022 observations of plasma concentrations from 74 patients.
- 1-compartment model with IIV and inter-occasion variability (IOV) on CL, V and Ka.
- 13 test relations – 4 dichotomous covariates (sex and co-medications with digoxin, diuretic or ace inhibitors) and 3 continuous covariates (age, weight, creatinine clearance (CLCR)) on CL and V, except for CLCR on only CL.

Pefloxacin

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

Software

XV SCM was implemented as a Perl-speaks-NONMEM tool (PsN ver. 3.4.2).

Conclusions

- XV SCM is useful to determine suitable model size expected to provide good predictive performance.
- Linearized XV SCM can be used instead of non-linear (regular) XV SCM to speed the process up.

References

[1] Breiman L, Spector P. Int Stat Rev. 1992. 60: 291-319.
[2] Khandelwal A, Harling K, Jonsson EN, Hooker AC, Karlsson MO. PAGE 2010.

Results

- Distributions of XV OFV profiles were characterized by using 10 times 5-fold XV SCM (non-linear/linearized). (Fig. 2)
- The most frequent model size predicted using XV SCM for simulated data was equal to the true model size. The frequency of underestimated model size using XV SCM was lower than when using standard SCM. (Fig. 3)
- Mean XV OFV profiles using XV SCM were minimal at 2, 2 and 13 relations for phenobarbital, moxonidine and pefloxacin, respectively (Fig. 4) and the sizes were the same (phenobarbital and moxonidine) or larger (pefloxacin) than when using standard SCM.
- XV OFV profiles and predictive model sizes were similar between linearized and non-linear XV SCM. (Figs. 2 – 4)

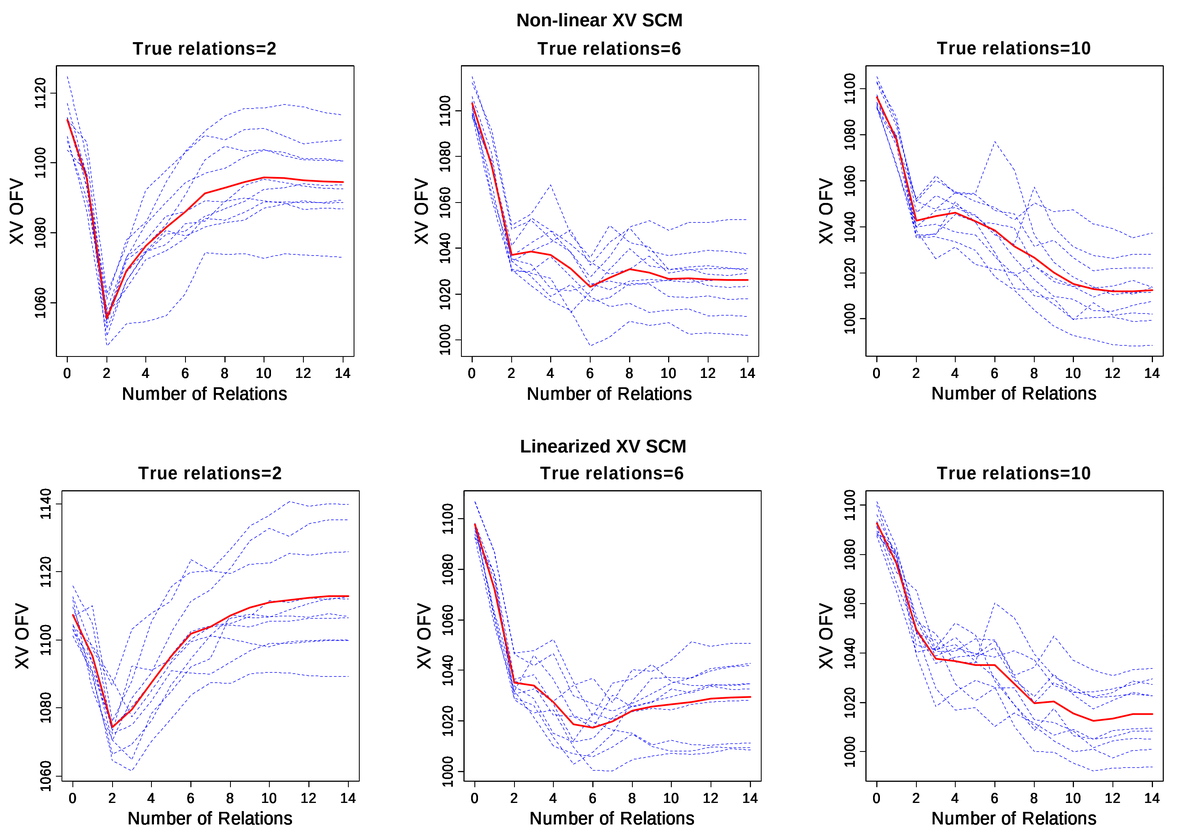


Figure 2. XV OFV profiles for a single simulated data using covariate models with 2, 6 or 10 true relations. Red solid line: mean XV OFV over splits. Blue dotted line: sum of XV OFV within the same split.

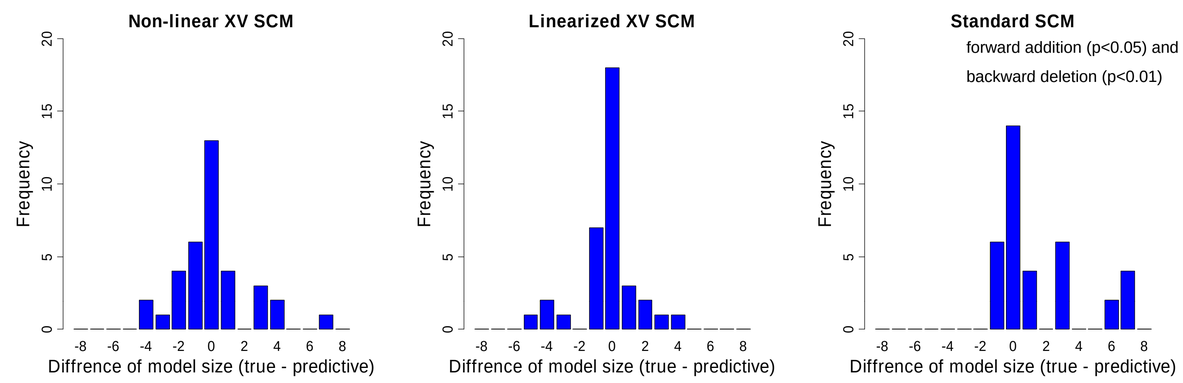


Figure 3. Distributions of differences of model size (i.e. true size – predictive size) for 6 simulated data using covariate models with 0, 2, 4, 6, 10 or 14 true relations.

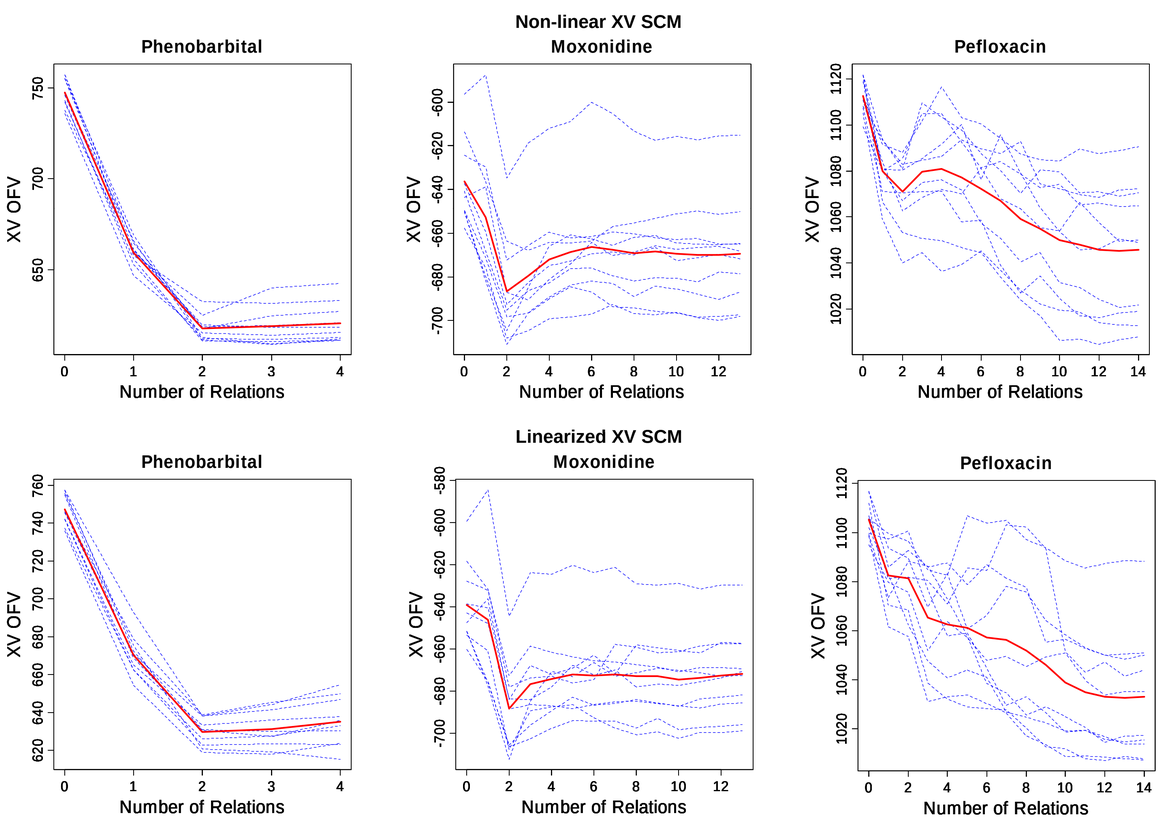


Figure 4. XV OFV profiles for real data. Red solid line: mean XV OFV over splits. Blue dotted line: sum of XV OFV within the same split.