Learning About Covariate Relations in the Patient Population Based on Data from Multiple Studies – Consequences of Different Approaches

UPPSALA UNIVERSITET

Jakob Ribbing and E. Niclas Jonsson

Div. of Pharmacokinetics and Drug Therapy,
Dep. of Pharmaceutical Biosciences,
Uppsala University,
Sweden
Box 591,
SE-751 24 Uppsala, SWEDEN
tel: 018-471 44 37
fax: 018-471 40 03

e-mail:
Jakob.Ribbing@farmbio.uu.se
Niclas.Jonsson@farmbio.uu.se

internet: www.farmbio.uu.se/PKPD/

Introduction: Identification and quantification of covariate relations is an important part of population pharmacokinetic/pharmacodynamic modelling and is often based on data from a single study. However, the individuals in a clinical study are merely a sample and may not truly reflect the underlying patient population. Thus, basing covariate selection on data from a single study may lead to false covariates being included in the model (type-I error), true covariates being omitted (type-II error) and to estimated coefficients being biased due to data-driven selection of covariates (selection bias).[1] Overall this will have negative consequences on the predictive performance of the model.

Aim: To study, through simulation, to what extent different approaches to covariate identification from multiple studies lead to lower type-I and II errors, less biased estimates of the covariate coefficients and better predictive performance.

Method: Six potential covariates were sampled with replacement from an empirical distribution containing 1492 patients[2]. For each replicate, three PK datasets were simulated using a one-compartment model with first-order absorption and a multivariate-linear-covariate model on the typical value of clearance (TVCL).

A stepwise procedure within NONMEM was used to select covariates[3]. A final covariate model was derived from study 1 and 2 with several different approaches. The first and second dataset, corresponding to two different studies, contained 200 subjects in total. The third dataset consisted of 1000 subjects and was only used to assess the predictive performance on TVCL.

The approaches examined include:

- Naive Independent (NI): analysing the second dataset independently of the first dataset
- Merge (ME): merging the first and second datasets into one
- Best Guess (BG): estimating the model selected from the first dataset on the second

Simulation set-ups:

- Default set-up where each of four covariates had 15% effect (10/90 percentile) on TVCL. The four covariates were Weight (WT), SEX and two concomitant medications: ATAR and C2E1
- Adaptive design, where covariates significant on the 5%-level in the first dataset were stratified for the covariate effect to be estimated better in the second
- Reduced signal-to-noise ratio where only WT has an (15 %) effect on TVCL

Results & Discussion: The inclusion ratios for all six potential covariates are shown for the different approaches and set-ups in Fig. 1. The binomial covariates SEX, ATAR and C2E1 all have the same magnitude of covariate effect, but different inclusion ratios due to different binomial distribution and correlation between the six potential covariates. Merging the two datasets improves the selection ratios.

The BG approach is non-data driven and thus without selection bias (Fig. 2). Using the NI approach creates high selection bias in some covariates. The bias is still present although reduced when analysing the merged dataset.

From Fig. 3 it is clear that in all situations investigated the ME approach performs considerably better than the alternatives in terms of predictive performance. Being unbiased, the BG approach performs slightly better than the NI approach.

In a situation with a large number of potential covariates in relation to the moderate number of subjects[4], the BG could possibly yield better predictive performance than the ME.

Conclusions:

- Merging all available studies into one dataset is often the best approach for learning about covariate relations
- Fitting a Best-Guess model without data-driven selection of covariates provides unbiased estimates of covariate relations
- Data-driven selection of covariates may create high selection bias in the least important covariates, but on average predictive performance this approach performs only slightly worse than the Best-Guess approach

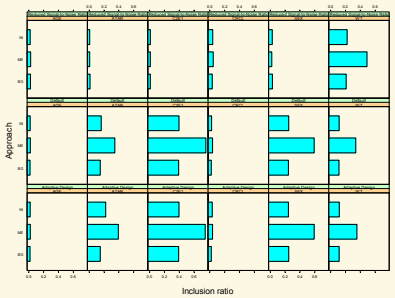

Figure 1. *Inclusion ratios for the six potential covariates tested for inclusion according to three different approaches in three simulation set-ups (top, middle and bottom row). The approaches considered are the Naive Independent (NI), the Merge (ME) and the Best Guess (BG). The difference in inclusion ratio is due to the simulation model used, the distribution of covariate values and the between-covariate correlations, as well as the approach to covariate modelling.*

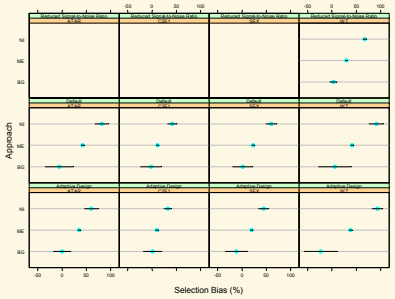

Figure 2. *Selection Bias (point estimate and 95% confidence interval) for the covariates in the simulation model given for different simulation set-ups and approaches to covariate modelling. The approaches considered are the Naive Independent (NI), the Merge (ME) and the Best Guess (BG). The ME has less selection bias than the NI, but the BG creates no selection bias at all.*

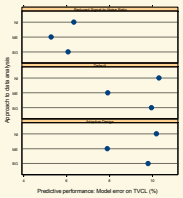

Figure 3. *Predictive performance given for different simulation set-ups and approaches to covariate modelling. The approaches considered are the Naive Independent (NI), the Merge (ME) and the Best Guess (BG). For all set-ups, the ME approach has considerably better predictive performance.*

References:

[1] Ribbing and Jonsson. Power, Selection Bias and Predictive Performance of the Population Pharmacokinetic Covariate Model. In Press (Journal of Pharmacokinetics and Pharmacodynamics, Vol. 31, No. 2, Apr. 2004)

[2] Zingmark, et al. Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. Br J Clin Pharmacol, Vol. 56, No. 2, Aug 2003

[3] Jonsson and Karlsson. Automated covariate model building within NONMEM. Pharm Res. 15:1463-8 (1998).

[4] Harrell. 4.4 Overfitting and Limits on Number of Predictors. In: Regression Modeling Strategies. New York: Springer-Verlag; 2001: 60-61.