AAPS PharmSci 2002; 4 (4) article 27 (http://www.aapspharmsci.org).

Comparison of Stepwise Covariate Model Building Strategies in Population Pharmacokinetic-Pharmacodynamic Analysis

Submitted: May 8, 2002; Accepted: September 2, 2002; Published: October 7, 2002

Ulrika Wählby [1], E. Niclas Jonsson [1] and Mats O. Karlsson [1]

[1] Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, 751 24 Uppsala, Sweden

ABSTRACT The aim of this study was to compare 2 stepwise covariate model-building strategies, frequently used in the analysis of pharmacokinetic-pharmacodynamic (PK-PD) data using nonlinear mixed-effects models, with respect to included covariates and predictive performance. In addition, the effects of stepwise regression on the estimated covariate coefficients were assessed. Using simulated and real PK data, covariate models were built applying (1) stepwise generalized additive models (GAM) for identifying potential covariates, followed by backward elimination in the computer program NONMEM, and (2) stepwise forward inclusion and backward elimination in NONMEM. Different versions of these procedures were tried (eg, treating different study occasions as separate individuals in the GAM, or fixing a part of the parameters when the NONMEM procedure was used). The final covariate models were compared, including their ability to predict a separate data set or their performance in cross-validation. The bias in the estimated coefficients (selection bias) was assessed. The model-building procedures performed similarly in the data sets explored. No major differences in the resulting covariate models were seen, and the predictive performances overlapped. Therefore, the choice of model-building procedure in these examples could be based on other aspects such as analyst- and computer-time efficiency. There was a tendency to selection bias in the estimates, although this was small relative to the overall variability in the estimates. The predictive performances of the stepwise models were also reasonably good. Thus, selection bias seems to be a minor problem in this typical PK covariate analysis.

KEYWORDS: stepwise model building, covariate analysis, GAM, NONMEM, selection bias.

INTRODUCTION Population pharmacokinetics (PK) and pharmacodynamics (PD) involve analysis of data originating from clinical settings (ie, sparse data from a large number of subjects). Population mean parameters and between- and within-individual variability are simultaneously estimated by the use of nonlinear mixed-effects models. One of the important aims in population PK-PD modeling is the establishment of relationships between parameters and covariates (ie, patient specific variables) to explain parameter variability and facilitate dose adjustment decisions.

Stepwise model building is frequently employed in population PK(-PD) covariate model building with NONMEM (the most widely used program for population PK(-PD) modeling). [1] For example, 58 of 60 papers (1999-2001) related to covariate model building in NONMEM included stepwise procedures (the reference list is available from the authors). No previous comparison between covariate model building procedures has been reported in the literature. In this article we assess the performances of 2 (objective) model-building strategies that make use of stepwise procedures. In addition, we explore how the results of a (typical) covariate analysis can be affected by selection bias, a general problem in stepwise regression.

The search for important covariates is not always straightforward, especially if there are large numbers of covariate-parameter relations to consider. In addition to finding the "right" (ie, true or predictive) covariate, the right functional form of the relation must be identified. This may require considerable analyst- and/or computer run-time. Different approaches to aid model building and reduce the number of models to explore have been considered. Maitre et al [2] suggested plotting individual empirical Bayes (posthoc) estimates of the parameters versus each of the covariates to discover potentially important covariates. These need to be included in the mixed-effects model in some way, often involving stepwise testing. Mandema et al presented a stepwise procedure, using generalized additive models (GAM), [3] which is commonly used (for example, see references [4-13]) and has been implemented in the program Xpose. [14] Built in a stepwise additive way, the GAM makes use of posthoc parameter estimates that are regressed on the individual covariate values. Possible covariate relations are identified, then incorporated and tested in the population model. Another stepwise routine has been suggested in which the covariate model is built within the population model in NONMEM. [15] In this automated procedure the model is first built up by including a new parameter-covariate relation term in each step and, when no more terms can be included (based on a prespecified criterion), all terms are retested in the model by exclusion.

The stepwise model-building procedures suffer the risk of being affected by the general problems related to

Correspondence to:
Ulrika Wahlby
Telephone: +46 18 471 43 05
Facsimile: +46 18 471 40 03
E-mail: Ulrika.Wahlby@farmbio.uu.se

stepwise regression. Some of these include selection bias (ie, overestimation of the effects of the selected covariates), [16-18] possibly overstated importance of retained variables, and invalid distributional assumptions. [19] There are also concerns regarding the number of candidate variables and variables being collinear, as these could affect the number of noise variables and accurate predictor variables in the final model. [20] One suggested way to avoid the problem of selection bias and get unbiased coefficients is to specify a best guess model based on prior knowledge. [21,22] As there is no investigation described in the literature, we also wanted to address the magnitude of the selection bias problem in the area of nonlinear mixed-effects modeling.

MATERIALS AND METHODS

Covariate models were built using 2 model-building procedures: (1) GAM for identifying candidate covariate effects and NONMEM to test the importance of found covariate terms (GAM-NM) and (2) NONMEM both in a stepwise search for possible covariate effects and in a following backward deletion step(s) (NM-NM). The procedures were applied to both simulated and real data. Different versions of the procedures were also tried; for example, to consider run-time saving possibilities or to account for time-varying covariates. A GAM constrained to mimic stepwise multiple linear regression (ie, no nonlinear relationships were allowed) was also tried. The predictive performances of the found models were assessed using either a separate data set (in the case of simulated data) or cross-validation (in the case of real data). The GAM, implemented in Xpose 3.0, [14] and NONMEM (version VI beta) were used. Estimation in NONMEM was performed using the first-order conditional estimation (FOCE) method with the interaction option.

In the following, descriptions of the model-building strategies are followed by the details of the simulated and real PK data, and the different versions of the procedures. The simulations made to assess the magnitude of the selection bias incurred by the use of stepwise model-building are then explained.

Model building procedures

GAM-NM

The individual empirical Bayes estimates of the parameters were first obtained from a fit of a model without covariate effects (basic model). The GAM [3] was used, separately for each fixed effects parameter, regressing the parameter on the covariates trying both linear and nonlinear (natural cubic spline with 1 internal breakpoint) models. Model discrimination was based on the Akaike information criteria (AIC); in each step the model that decreased the AIC the most was retained. The search ended when no model decreased the AIC further. The covariate relations identified in the final GAM models, for all parameters, were incorporated in the mixed-effects model (full model), and covariates were deleted from the model in a stepwise manner using NONMEM. In each step all covariates were left out of the model, one at a time. The least important covariate, according to the likelihood ratio test based on the objective function value (OFV, approximately -2 times the log likelihood of the data), was dropped from the model unless the difference in OFV (DOFV, likelihood ratio) was larger than 6.63 (corresponding to $P < .01$). The final model was established when no more covariates could be excluded from the model.

Covariate effects were modeled, in the mixed-effects model, as being proportional to the typical value of the parameters:

$$\tilde{P} = \theta_1 * [1 + \theta_2 * COV] \quad (1)$$

q1 is the typical value of the parameter, and θ_2 is the fractional change in the typical parameter with the covariate (COV). Continuous covariates were centered on their median values (medianCOV); θ_1 then represents the typical value of the parameter in an individual with medianCOV, and θ_2 the fractional change in the typical parameter with each unit's change from the median covariate value. Nonlinear models for continuous covariates were parameterized in terms of 2 slopes, joined at the median covariate value:

$$\tilde{P} = \begin{cases} \theta_1 * [1 + \theta_2 * (COV - medianCOV)] & \text{if } COV \le medianCOV \\ \theta_1 * [1 + \theta_3 * (COV - medianCOV)] & \text{if } COV > medianCOV \end{cases} \quad (2)$$

NM-NM

With the NM-NM routine the covariate model is built altogether within NONMEM in a stepwise manner, starting from the basic model. In each step, all possible parameter-covariate combinations are tried, and inclusion of covariate effects is based on the likelihood ratio test. Continuous covariates are first tried in linear relations to the parameters. Once a covariate has been included in the model it is also tried, in the next step, in a nonlinear relation to the parameter. Covariate effects were parameterized as described above. The covariate relationship that gives rise to the largest ΔOFV is retained in the model, given that inclusion results in ΔOFV>3.84 (corresponding to $P < .05$). The full model is established when no more covariates can be included according to this criterion. The covariate relations are then left out of the full model 1 at a time and tested using a stricter criterion (ΔOFV>6.63), in the same way as for GAM-NM, described above.

Simulated data

Data sets

Ten simulated data sets were used in the comparison (previously used in Jonsson and Karlsson [15]). These were generated from a 1-compartment model with first-order absorption under steady-state conditions. The data sets contained on average 3 (range, 1-4) observations from each of 64 individuals, 230 observations in total. (Samples were drawn at 0.5, 2, 4, and 6 or 0.5, 2, 8, and 12 hours post-dose in half of the individuals.) The inter-

individual variability (IIV) was described by an exponential model (Equation 3), where η_i is a normally distributed random variable, with zero mean and variance ω^2, accounting for the difference between the individual (P_i) and typical ($\tilde{P}$) parameter estimate in the population. A proportional residual error was used in the simulations (Equation 4), where ε_{ij} is a zero mean, normally distributed random variable with variance σ^2.

$$P_i = \tilde{P} * e^{\eta_i} \quad (3)$$

$$C_{ij,obs} = C_{ij,pred} * (1 + \varepsilon_{ij}) \quad (4)$$

In the simulation model, apparent clearance (CL/F, $\theta_{CL/F}$ = 20 L/h, $\omega_{CL/F}$ 25%) was related to gender (θ_{SEX} = 0.195) and nonlinearly to age (constant below median age, linearly decreasing above median age (years), θ_{AGE} = -0.0385), and apparent volume of distribution (V/F, $\theta_{V/F}$ = 100 L, $\omega_{V/F}$ 25%) was related to concomitant medication with hydrochlorothiazide (θ_{HCTZ} = -0.560) and bodyweight (kg) (θ_{WT} = 0.006). The absorption rate constant (θ_{Ka} = 2 h^{-1}, ω_{Ka} 20%) was not influenced by covariate effects. The residual error (σ) was 15%. The covariate data, summarized in **Table 1** Table 1, were taken from a clinical trial of the antihypertensive drug prazosin. [3]

GAM-NM simulated data

To explore possible differences in the resulting covariate models, the GAM-NM procedure was repeated for all simulated data sets using both the individual empirical Bayes estimates of the parameters (GAM(parameter)-NM) and the hi (GAM(eta)-NM). Covariate models were built for CL/F and V/F. A GAM, using posthoc parameter estimates, in which the model scope was limited to linear relationships (similar to multiple linear regressions) was also tried (Linear GAM-NM).

NM-NM simulated data

Covariate models for all data sets were established for CL/F and V/F. Fixing some parameter values during model building may shorten the run-times of the NM-NM procedure. To explore the effect of fixing parameter values, the NM-NM covariate models were built using the following variations: (1) estimating all parameters, (2) fixing structural parameter values (eg, CL/F and V/F) to the estimates from the fit of the basic model, and (3) fixing all structural and variance parameters (apart from the residual error and the explored covariate effects). Last, all parameters of the final models were estimated.

Assessment of predictive performance

The covariate models found in the inclusion step (full models) and after the backward step (final models), by the different procedures, were compared with respect to covariate terms. Two covariate models may be equally valuable (ie, have equal predictive capability) even though they differ with respect to included covariate terms—for instance, if covariates are correlated. To consider this aspect, the predictive performance of the found models was evaluated using a separate test data set. This test data set was generated by replication of the covariate data 200 times (ie, generating 12 800 subjects), and simulating new concentrations from the simulation model (with new hs and es). The parameters of each final model were fixed and used to predict the observations in the test data set. The root mean square error (RMSE) and the mean prediction error (MPE) were calculated for measuring predictive performance. This calculation was made using the log of the observed (simulated) concentrations and the log of the predictions for the typical individual to account for heteroscedasticity in the data. The MPE and RMSE of the models derived by the GAM-NM and NM-NM procedures were compared with the same quantities obtained from both the fits of the basic and simulation models and the part of the simulation models supported by the data. The latter was obtained by applying stepwise backward elimination in NONMEM, starting from the simulation covariate model.

The real data set

Data set

The real PK data set has been analyzed previously and is described in the literature. [3, 23,24] The basic structural pharmacokinetic model was adopted from these earlier analyses, and only the influence of covariate effects was explored. The data originates from multiple IV infusions of a broad-spectrum antibiotic (pefloxacin) to 74 critically ill patients (200 or 400 mg over 1 hr). Approximately 3 samples were drawn from each individual at each of 1 to 3 occasions (visits) (time-course 2.5-14 days); in total 337 plasma concentrations were obtained. The basic model describing the data was a 1-compartment model parameterized in terms of clearance (CL) and volume of distribution (V), with interoccasion variability (IOV) in CL and V, and IIV in CL.

Equation 5 shows the model for IIV and IOV. This equation differs from Equation 1 by the term kik that accounts for the random variation of a parameter within an individual i between occasions k, where p is the approximate coefficient of variation of kik. The residual error was modeled to be proportional to the predicted concentrations (Equation 4)

$$P_{ik} = \tilde{P} * e^{\eta_i + \kappa_{ik}} \quad (5)$$

The covariate data are summarized in **Table 2** .

Table 1. Covariates Simulated Prazosin Data

Continous	Median	Range	Symbol
Age (years)	56	(24-69)	AGE
Height (cm)	173	(140-188)	HT
Weight (kg)	85	(51-137)	WT
Serum creatinine (mg/dL)	1.1	(0.7-1.8)	SECR
Categorical			
Sex (M / F)	42	22	SEX
Race (caucasian/black)	44	20	RACE
Smoking (Y/N)	16	48	SMOK
Propranolol medication (Y/N)	10	54	PROP
Hydrochlorothiazide medication (Y/N)	35	29	HCTZ
Other concomitant medication (Y/N)	51	13	CON

Table 2. Covariates Pefloxacin Data

Continuous	Median	Range	Symbol
Weight (kg)	67	(43-125)	WT
Age (years)	46	(18-84)	AGE
Creatinine clearance (mL/min)	104	(0.4-312)	CLCR
Glasgow score	10	(3-15)	GLAS
Simplified acute physiology score	11	(1-26)	SAPS
Albumin (g/L)	28	(17-40)	ALB
Bilirubin (mmol/L)	21	(4-150)	BIL
Alanine aminotransferase (IU/L)	43	(3-200)	ALAT
Alkaline phosphatase (IU/L)	164	(32 - 615)	AP
Prothrombin level (% normal)	67	(36-100)	PT
Systolic blood pressure (mmHg)	116	(60-175)	SBP
Heart rate (beats/min)	109	(60-170)	HR
Categorical			
Sex (M/F)	57	17	SEX
Center (1/2)	30	44	CEN
Catecholamine (Y/N)	22	91	CAT
Artificial ventilation (Y/N)	71	42	AV

GAM-NM real data

Potential covariate relations were identified by the GAM procedure for the posthoc individual estimates of CL and V from 3 different models: (1) a model describing only IIV in the structural parameters (ie, without IOV), (2) a model with IIV and IOV (the model described above), and (3) a model without IOV but in which the observations originating from different study occasions were treated as being from separate individuals—to account for covariates varying over time. For model (3), a GAM with the model scope constrained to only linear relationships was also tried. The relations found were incorporated into the model with IOV, using the original individual data structure for all models, and then deleted in a stepwise manner in NONMEM.

NM-NM real data

Stepwise model building was performed using 2 approaches. In the first the covariate model was built starting from the model including IIV and IOV, and in the second both the IIV and the IOV was set to zero (ie, all observations were treated as originating from a single individual [naïve/data pooling]).

Assessment of performance—cross-validation

The final models were compared with respect to the found covariate relations. To get a nearly unbiased measure of the ability of the models to predict the data, a cross-validation approach was used. This was accomplished by splitting the data set in 10 parts (random allocation of individuals). The parameters of the final model were estimated using 9/10 of the data and were used to predict the remaining 10th. Estimation and prediction were repeated so that predictions for all individuals were generated. Predictive performance was assessed by calculation of standardized mean prediction error (SMPE) as a measure of bias, [25] and root mean square (standardized) error (RMSSE) as a measure of precision. SMPE is the mean of the difference between the observed and predicted concentration divided by the estimate of the standard deviation of the predicted concentration (SDCp).

$$SMPE = \frac{\sum (C_{Obs} - C_{Pred}) / SD_{Cp}}{n} \quad (6)$$

The standardization was done for the same reason that the MPE and RMSE were computed on the log-scale for the simulated data. Because there is more than 1 observation per individual, and the prediction errors within an individual are correlated, only 1 observation per individual was used when calculating SMPE and RMSSE. These observations were chosen to be parameter-information–rich (sensitive to changes in the parameter, P) based on the partial derivatives with respect to P from the basic model. The point with the largest partial derivative with respect to P was chosen to assess the ability to predict P. Ninety-five percent confidence intervals (CI) were constructed for SMPE and RMSSE.

Estimation of selection bias in NM-NM

To assess the magnitude of selection bias (ie, overestimation of the regression coefficients due to stepwise selection), when stepwise covariate model building is used in a relatively typical population PK model, 100 new data sets were simulated from the final (NM-NM with IOV) model—referred to as the simulation model—for the real data set. The same model was fitted to each of the new data sets to get unbiased estimates of the covariate coefficients. The NM-NM model-building routine was also applied to each of the simulated data sets, starting from a (basic) model without covariate relations. This resulted in 2 sets of estimates of the covariate coefficients: simulation model estimates (estimated with simulation model) and stepwise model estimates (estimated with final NM-NM model for each of the 100 data sets). All estimates were normalized by division by their respective simulation coefficients. Box-and-whiskers plots of normalized coefficients were constructed to visualize potential differences.

The biases of the stepwise model coefficients—the MPE between true coefficients and stepwise model estimates (expressed as a percentage of the true coefficient)—were determined. The true coefficient refers to the estimate from the selected covariate model in a large test

data set (see below); ie, the true coefficient was recalculated for each set of selected covariates. [17] The bias of the stepwise model coefficients was compared with the precision of the simulation model estimates (ie, the precision resulting if the simulation model was treated as the best guess model). Precision was expressed as the RMSE between simulation coefficients and simulation model estimates (in all 100 data sets) divided by the simulation coefficients for this model.

Predictive performance—simulated pefloxacin models

The ability of the above final NM-NM covariate models to predict the concentrations of a separate test data set was also assessed. This test data set was obtained by replication of the pefloxacin covariate data 200 times (ie, generating 14 800 subjects), and simulation of new concentrations from the simulation model (with new ηs and εs). Predictive performance of the 100 models found was calculated as RMSE and MPE (as described previously for the simulated prazosin data sets). For comparison, predictive performance was also computed for the fits of the simulation model, the fits of the basic model, and the simulation coefficients.

Results Simulated data

The results from fitting a GAM to the ηs or to the individual parameter estimates were similar (**Tables 3** and **4**). When the hs were used, fewer false covariates were identified in the full (GAM) model than when the parameter estimates were used. However, most of the false covariates were excluded from the model in the backward NONMEM step, and the final models were comparable. Note that for data set 8, no covariates were found in the final GAM(parameter) model for CL/F, even though both age and sex were supported in this data set. No difference in bias could be detected (data not shown), but the RMSE was somewhat smaller for the models developed based on the ηs (**Figure 1**).

When the covariate models were built stepwise within NONMEM, the results were similar to when the GAM(eta)-NM was used. The final models were identical, although the full NM-NM model included both some additional true and some additional false covariates. Both these procedures detected almost all of the covariates supported by the data. The predictive performances of the final NM-NM and GAM(eta)-NM models (both with respect to RMSE and MPE) were the same, and not very different from that of the supported models.

When the structural parameters were fixed to estimates from the basic model during the NM-NM procedure, fewer true covariate relations were found. The ability to find true covariate relations diminished further if the variance parameters also were fixed during model building; however, this also resulted in fewer false covariates (**Table 4**). The precision in predicting the large simulated data set was reduced when the parameters were fixed while the model was built compared with when parameters were estimated.

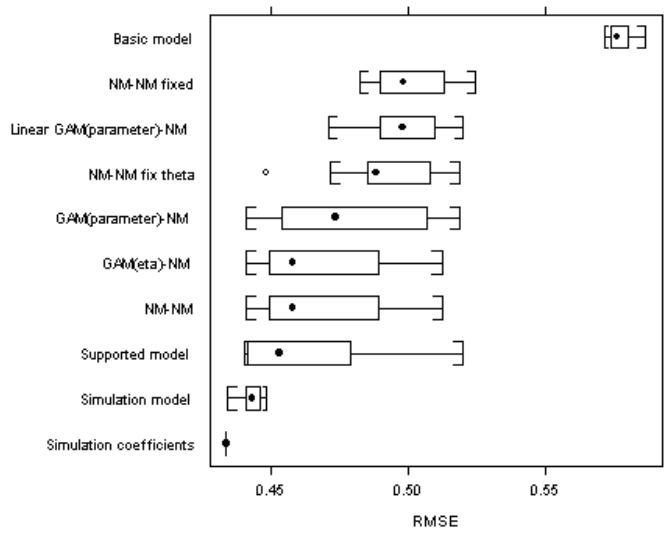

Figure 1. Box-and-whiskers plots of root mean square error (RMSE) between the log of the observed (simulated) concentrations and the log of the predicted concentrations obtained with the basic, NM-NM fixed parameters, NM-NM fixed thetas, GAM(parameter)-NM, GAM(eta)-NM, linear GAM(parameter)-NM, NM-NM, supported, simulation (estimated) models and the simulation coefficients for the 10 simulated prazosin data sets. The filled dot represents the median, the box range is the interquartile range, and the whiskers extend to values less than/greater than/equal to 1.5 times the interquartile range. Values beyond these limits are shown as empty dots.

Table 4. The Total Number of True and False Covariates in the Final Models Found for CL/F and V/F, by the different Stepwise Procedures, in the 10 Simulated Prazosin Data Sets*

	CL/F			V/F		
	AGE†	SEX	False	HCTZ	WT	False
GAM(eta) -NM	9 (7)	5	2	10	3	2
GAM(parameter) -NM	8 (6)	4	2	10	2	2
Linear GAM(parameter)-NM	7 (0)	3	1	10	2	2
NM-NM	9 (7)	5	2	10	3	2
NM-NM fixed thetas	8 (2)	3	2	10	3	2
NM-NM fixed parameters	6 (0)	2	2	10	1	-
Supported	9 (7)	6		10	4	

*CL/F indicates apparent clearance; V, apparent volume of distribution; HCTZ, hydrochlorothiazide medication ; GAM, indicates generalized additive models; NM, NONMEM computer program.
†Right functional form in parentheses

Table 3. Full and Final (in bold) Covariate Models Found for CL/F, Using the Different Stepwise Procedures, in the Simulated Prazosin Data Sets*

Data set		Full/**Final** Model				
1	NM-NM	**AGE2**†	**SEX**	WT		
1	NM-NM fixed thetas	**AGE2**†	**SEX**			
1	NM-NM fixed parameters		**SEX**			
1	GAM(eta)-NM	**AGE2**†	**SEX**	WT		
1	GAM(parameter)-NM	**AGE2**†	**SEX**	WT		
1	Linear GAM(parameter)-NM	AGE	**SEX**			
1	Supported	**AGE2**†	**SEX**			
2	NM-NM	**AGE2**†	**SEX**			
2	NM-NM fixed thetas	**AGE**2†	SEX			
2	NM-NM fixed parameters	**AGE**2†	SEX			
2	GAM(eta)-NM	**AGE2**†	**SEX**			
2	GAM(parameter)-NM	**AGE2**†	**SEX**			
2	Linear GAM(parameter)-NM	**AGE**		HT	HCTZ	
2	Supported	**AGE2**†	**SEX**			
3	NM-NM	**AGE**2†		RACE		
3	NM-NM fixed thetas	**AGE**		RACE		
3	NM-NM fixed parameters	**AGE**				
3	GAM(eta)-NM	**AGE**2†		RACE		
3	GAM(parameter)-NM	**AGE**2†		RACE		
3	Linear GAM(parameter)-NM	**AGE**	SEX	RACE	CON	HCTZ
3	Supported	**AGE**				
4	NM-NM	**AGE2**†				
4	NM-NM fixed thetas	**AGE**2†				
4	NM-NM fixed parameters	**AGE**				
4	GAM(eta)-NM	**AGE2**†		HT		
4	GAM(parameter)-NM	**AGE2**†		HT		
4	Linear GAM(parameter)-NM	**AGE**		HT	HCTZ	
4	Supported	**AGE2**†				
5	NM-NM	**AGE2**†				
5	NM-NM fixed thetas	**AGE2**†				
5	NM-NM fixed parameters	AGE2†				
5	GAM(eta)-NM	**AGE2**†				
5	GAM(parameter)-NM	**AGE2**†				
5	Linear GAM(parameter)-NM	**AGE**	SEX	RACE		
5	Supported	**AGE2**†				
6	NM-NM	**AGE2**†	**SEX**			
6	NM-NM fixed thetas	**AGE**2†	**SEX**			
6	NM-NM fixed parameters	**AGE**	SEX			
6	GAM(eta)-NM	**AGE2**†	**SEX**			
6	GAM(parameter)-NM	**AGE2**†	**SEX**			
6	Linear GAM(parameter)-NM	**AGE**	**SEX**			
6	Supported	**AGE2**†	**SEX**			
7	NM-NM	**AGE**2†	**SEX**			
7	NM-NM fixed thetas	**AGE**	**SEX**			
7	NM-NM fixed parameters	**AGE**	**SEX**			
7	GAM(eta)-NM	**AGE**2†	**SEX**			
7	GAM(parameter)-NM	**AGE**	**SEX**			
7	Linear GAM(parameter)-NM	**AGE**	**SEX**			

7	Supported	**AGE**	**SEX**			
8	NM-NM	**AGE2**†	**SEX**			
8	NM-NM fixed thetas		SEX			
8	NM-NM fixed parameters		SEX			
8	GAM(eta)-NM	**AGE2**†	**SEX**			
8	GAM(parameter)-NM	RACE	SEX	SMOK	HCTZ	HT
8	Linear GAM(parameter)-NM	RACE	SEX	SMOK	HCTZ	HT
8	Supported	**AGE2**†	**SEX**			
9	NM-NM	AGE2†		**SECR**		
9	NM-NM fixed thetas	AGE		**SECR**		
9	NM-NM fixed parameters	AGE		**SECR**		
9	GAM(eta)-NM	AGE2†		**SECR**2†		
9	GAM(parameter)-NM	AGE2†		**SECR**2†	CON	
9	Linear GAM(parameter)-NM	AGE		SECR	HT	
9	Supported					
10	NM-NM	**AGE2**†	SEX	**HT**2†		
10	NM-NM fixed thetas	**AGE**		**HT**		
10	NM-NM fixed parameters	**AGE**		**HT**		
10	GAM(eta)-NM	**AGE2**†	SEX	**HT**2†		
10	GAM(parameter)-NM	**AGE2**†	SEX	**HT**2†		
10	Linear GAM(parameter)-NM	**AGE**		**HT**		
10	Supported	**AGE2**†	**SEX**			

*GAM indicates generalized additive models; NM, NONMEM computer program.
†Nonlinear relationship

There was a tendency of the constrained linear GAM to find more false covariate relationships. However, after the backward step most of these were eliminated, and the predictive performance was of the same magnitude as when the NM-NM procedure with fixed parameters was used (**Figure 1**).

Table 5. Full and Final (in bold) Covariate Models Found by the Different Stepwise Procedures in the Real Pefloxacin Data*

	CL	V
	Full/**final** models	Full/**final** models
NM-NM$_{IOV}$	**CLCR BIL AGE SEX** AP **CEN** SBP	**WT CLCR BIL AP** AV
GAM$_{Occasion}$†-NM	**CLCR BIL AGE SEX** AP **CEN** SBP	**WT CLCR BIL** SEX **AP** CRCL2‡
GAM$_{IOV}$-NM	**CLCR BIL AGE SEX** AP	**WT CLCR** SEX **AP AV**
GAM$_{without\ IOV}$-NM	**CLCR BIL AGE SEX** AP	**WT CRCL BIL** SEX SAPS
NM-NM$_{POOLED}$	**CLCR BIL AGE SEX AP CEN SBP WT2**‡ CAT	**WT CLCR BIL** AP SAPS

*CL indicates clearance; V, volume of distribution; NM, NONMEM computer program; GAM generalized additive models; CLCR, creatinine clearance; BIL, bilirubin; AP, alkaline phosphatase; CEN, center; SBP, systolic blood pressure; SAPS, simplified acute physiology score; AV, artificial ventilation.
†Each occasion treated as a separate individual
‡Nonlinear models.

Table 6. Final Parameter Estimates from the Models Found by the Different Stepwise Procedures in the Real Pefloxacin Data*

	BASIC$_{IOV}$	NM-NM$_{IOV}$	GAM$_{Occasion}$†-NM	GAM$_{IOV}$-NM	GAM$_{without\ IOV}$-NM	NM-NM$_{POOLED}$
OFV	1054.5	929.9	929.9	940.3	947.5	1100.07
θ_{CL}	3.48	3.51	3.51	4.13	4.17	3.37
θ_{CLAGE}	-	-0.0064	-0.0064	-0.0062	-0.0064	-0.0099
θ_{CLAP}	-	-	-	-	-	0.00077
θ_{CLBIL}	-	-0.0048	-0.0048	-0.0050	-0.0047	-0.0047
θ_{CLCEN}	-	0.30	0.30	-	-	0.29
θ_{CLCLCR}	-	0.0027	0.0027	0.0033	0.0031	0.0018
θ_{CLSBP}	-	-	-	-	-	0.0020
θ_{CLSEX}	-	-0.24	-0.24	-0.27	-0.28	-0.13
θ_{CLWT}‡	-	-	-	-	-	0.018
θ_{CLWT2}‡	-	-	-	-	-	-0.0025
θ_{V}	67.9	69.3	69.3	75.3	66.4	76.9
θ_{VAP}	-	-0.00058	-0.00058	-0.00063	-	-
θ_{VAV}	-	-	-	-0.17	-	-
θ_{VBIL}	-	-0.0031	-0.0031	-	-0.0028	-0.0038
θ_{VCLCR}	-	0.0017	0.0017	0.0015	0.0016	0.0023
θ_{VWT}	-	0.016	0.016	0.017	0.017	0.023
ω_{CL}	0.58	0.25	0.25	0.26	0.28	0.30
π_{CL}	0.21	0.22	0.22	0.25	0.22	-
π_{V}	0.28	0.17	0.17	0.15	0.22	-
σ	0.15	0.15	0.15	0.15	0.15	-

*NM indicates NONMEM computer program; GAM, generalized additive models; CLWT and CLWT2 are the slopes below and above median WT, respectively.
†Each occasion treated as a separate individual.
‡Nonlinear model.

The real data set

The covariate models identified in the GAM for the PK data set, using models without IOV, with IOV, or with different occasions treated as separate individuals, were comparable (**Table 5**). The full model with each occasion treated as a separate individual included the most covariates, and this model also had the lowest OFV. When the models with and without IOV were used, the same covariates were identified to influence CL, while the models for V differed some. The OFV for the model with IOV was lower than that for the model without IOV. After the backward deletion step in NONMEM, the covariate models for CL were similar. The models with and without IOV differed by only 1 parameter (CEN) from the model in which each occasion was treated as a separate individual. The covariate models for V also differed by only 1 parameter. Parameter estimates from the covariate models are shown in **Table 6** . When the GAM was constrained to use only linear relationships, the final model was exactly the same as when the GAM was unconstrained; the results are therefore only shown for the latter.

The full NM-NM model with IOV had a lower OFV than the full GAM-model in which each occasion was treated as a separate individual. However, the final models for these 2 model-building strategies were exactly the same. When the covariate model was built with NM-NM and the pooled data approach, the largest number of covariates were found, in both the full and final models.

In the cross-validation, the models showed similar predictive ability with respect to both SMPE and RMSSE (**Table 7**). There was a small negative bias in the points influencing CL, while a small positive bias and a tendency to better predictions of the points influenced by V could be observed. Worth noting is that the model from the pooled data approach also predicts well.

Table 7. Predictive Performance, of the Final Covariate Models Found by the Different Stepwise Procedures in the Real Pefloxacin Data, Based on Cross-Validation, Using One Parameter-Information-Rich Concentration per Individual*

	Bias - SMPE (95% CI)				Precision - RMSSE (95% CI)			
	$Cl^{\dagger}$		$V^{\dagger}$		$Cl^{\dagger}$		$V^{\dagger}$	
NM-NM_{IOV}	-0.18	(-0.36 - -0.01)	0.15	(-0.02 - 0.31)	0.79	(0.60 - 0.95)	0.72	(0.57 - 0.85)
$GAM_{Occasion}$‡-NM	-0.18	(-0.36 - -0.01)	0.15	(-0.02 - 0.31)	0.79	(0.60 - 0.95)	0.72	(0.57 - 0.85)
GAM_{IOV}-NM	-0.20	(-0.37 - -0.03)	0.15	(-0.01 - 0.32)	0.78	(0.62 - 0.91)	0.73	(0.61 - 0.83)
$GAM_{without IOV}$-NM	-0.16	(-0.33 - 0.01)	0.16	(0.01 - 0.31)	0.76	(0.61 - 0.89)	0.69	(0.69 - 0.79)
NM-NM_{POOLED}	-0.23	(-0.38 - -0.07)	0.14	(-0.02 - 0.30)	0.72	(0.54 - 0.87)	0.69	(0.56 - 0.81)

*SMPE indicates standardized mean prediction error; CI, confidence interval; RMSSE, root mean square (standardized) error; CL, clearance; V, volume of distribution; NM, NONMEM computer program; GAM, generalized additive models; IOV, interoccasion variability.
†The subscript denotes the parameter to which the concentration was information-rich
‡Each occasion treated as a separate individual

Table 8. Bias of Stepwise Model Estimates, from the Subset of Data Sets Where the Covariate Effect was Identified, and Precision of Simulation Model Estimates, from all (100) Data Sets, for Covariate Effects on CL in the Simulated Pefloxacin Data*

Covariate	Bias Stepwise Estimates, Subset %MPE†	Precision Simulation Model Estimates, 100 %RMSE‡	Ratio Bias/Precision
AGE	-12	35	-0.34
BIL	2	11	0.19
CEN	-14	39	-0.35
CRCL	8	22	0.38
SEX	-8	27	-0.30

*BIL indicates bilirubin; CEN, center; CRCL, creatinine clearance.
†Expressed as the percentage of the true coefficient.
‡Expressed as the percentage of the simulation coefficient.

Estimation of selection bias in NM-NM

The stepwise NM-NM procedure selected the relationships CL-BIL and V-WT in 99 of the 100 simulated data sets, CL-CLCR in 96, V-CLCR in 80, CL-CEN in 75, CL-SEX and CL-AGE in 66, V-AP in 63, and V-BIL in 60. The true (ie, simulation) model was found in 7 data sets. Altogether, 54 false covariates were found in 40 of the data sets (24 on CL and 30 on V). Box-and-whiskers plots of the simulation model and stepwise (normalized) estimates for each covariate effect for CL are shown in **Figure 2** . The simulation model estimates are shown for both all data sets and for the subset of data sets in which the covariate effect was selected by the stepwise procedure. For the covariate effects that were found in 99 of the 100 data sets (CL-BIL and V-WT [data not shown]), only minor differences could be detected between the simulation model and stepwise estimates. However, in the cases where the effects were found less frequently (eg, CL-SEX and CL-AGE) upward biases in the stepwise estimates were evident when compared with the simulation model estimates [from all data sets]. The simulation model estimates from the subset of data sets also showed a tendency of being upwards biased, which indicates that there is a selection bias. It could also be seen that the RMSE of the simulation model estimates always exceeded the MPE of the stepwise estimates (**Table 8**), indicating that the bias is of less importance.

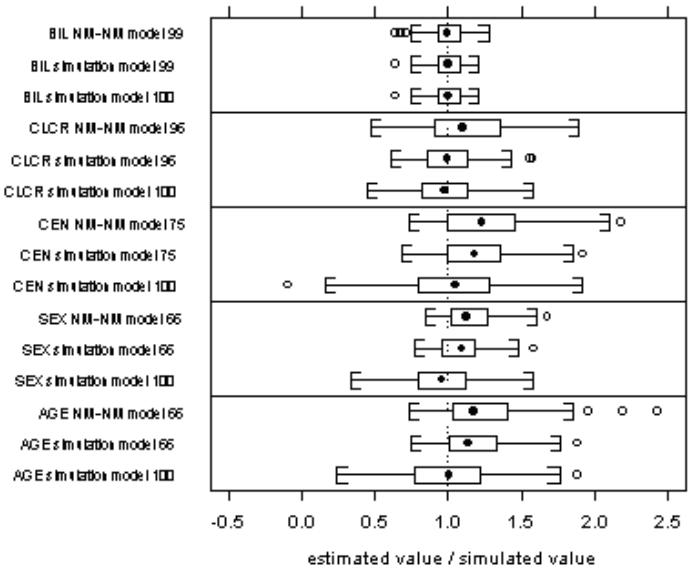

Figure 2. Box-and-whiskers plots of estimated covariate coefficients for CL in the simulated pefloxacin data sets; all values are normalized by division with the respective values used for simulation. For each covariate, from top to bottom, the stepwise NM-NM model estimates, the simulation model estimates from the subset of data sets where the covariate effects were found (indicated by the number), and the simulation model estimates from all 100 data sets, are shown. The filled dot represents the median, the box range is the interquartile range, and the whiskers extend to values less than/greater than/equal to 1.5 times the interquartile range. Values beyond these limits are shown as empty dots. The dotted line indicates the reference value 1.

Predictive performance—simulated pefloxacin

The ability of the stepwise model estimates to predict the concentrations of the simulated pefloxacin test data set was considerably better than that of the basic model without covariates, and slightly worse than that of the simulation model estimates regarding precision (**Figure 3**). The overall biases (MPE) of the predicted concentrations were comparable between the stepwise and basic models (data not shown).

DISCUSSION

The main aim of the study was to explore and compare the properties of the GAM-NM and the NM-NM with respect to inclusion of true and false covariates and predictive performance. As the model-building procedures performed similarly in the data sets explored, we will not make a systematic comparison but focus on some main conclusions. The secondary aim was to address how a “typical” population PK analysis is affected by the known problems associated with stepwise regression, as the implications of its use in the field of PK-PD modeling have not been investigated previously. A tendency toward selection bias was seen, but it was relatively small compared with the variability in the estimates between data sets.

While some minor differences could be seen between the GAM-NM and NM-NM covariate models, the variability between data sets was as large as the

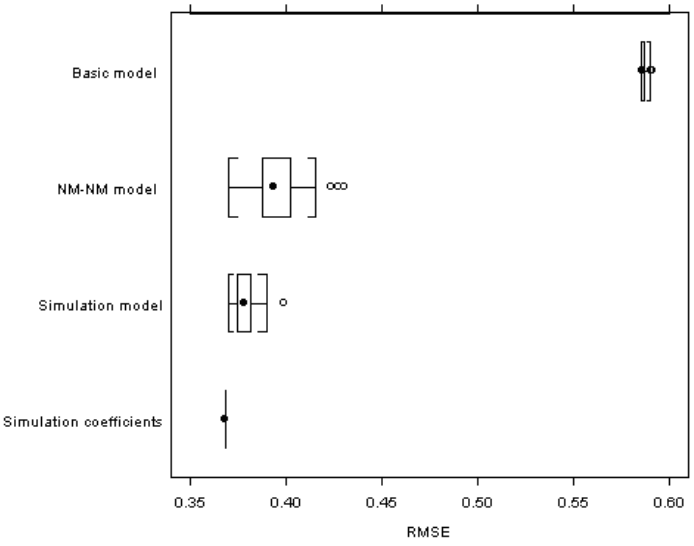

Figure 3. Box-and-whiskers plots of root mean square error (RMSE) between the log of the observed (simulated) concentrations and the log of the predicted concentrations obtained with the basic, stepwise NM-NM, and simulation models, and the simulation coefficients for the 100 simulated pefloxacin data sets. The filled dot represents the median, the box range is the interquartile range, and the whiskers extend to values less than/greater than/equal to 1.5 times the interquartile range. Values beyond these limits are shown as empty dots.

variability between procedures, and the predictive performances overlapped. Also, minor differences within procedures were noted; but the final covariate models were essentially the same, regardless of whether the hs or parameters were used in the GAM. When there were true nonlinear relationships in the data, and the GAM was constrained to use only linear relationships (similar to stepwise linear regression), these relationships could obviously not be found; in addition, other true relationships were overlooked, which lead to a lower predictive ability. For the NM-NM search, there was a tendency toward a lower inclusion frequency of both true and false covariates when parameters were fixed.

In these data sets, with sampling frequent enough to yield reasonably good individual parameter estimates, it is of little consequence, with respect to performance, which method is chosen. Hence, the choice must be based on other aspects of the procedures. The GAM is fast and, as it reduces the number of covariate relations to incorporate in the mixed-effects model, it is run-time efficient compared with the NM-NM procedure. Computer run-times can be a problem for NM-NM, especially since the procedure relies on significance levels derived from the likelihood ratio test. For many models (nonlinear models with heteroscedastic errors), the computer-intensive FOCE method with η-ε interaction must be used for these levels to equal the theoretical levels. [26] The relations found in the GAM must be included and tested in the NONMEM model; accordingly, the GAM-NM is not as analyst-time efficient as the NM-NM, which is fully automated. In many practical situations it is the data analyst time that is rate-limiting. The GAM does not account for covariates/parameters varying over time unless different occasions (visits) are treated as separate individuals. In circumstances in which this is not possible, the NM-NM would be preferred.

Both procedures are prone to be affected by selection bias; however, the 2 methods were equal with respect to performance (and the GAM is often combined with a stepwise inclusion of the identified covariates in the mixed-effects model), thus the issue was only addressed with NM-NM. Stepwise model building did result in upward-biased estimates of the covariate coefficients, but the bias seemed minor compared with the imprecision of the estimates yielded when the true (simulation) model was applied. The overestimated effects may lead to poorer predictive performance of the model. However, the difference in predictive performance between the true (simulation) and stepwise models, in our example, was small. The frequency of false inclusion in the simulated pefloxacin and prazosin data sets indicated that when more variables were tested more false covariates were found in the final models, probably an effect of multiple testing. However, the term “false” may not be appropriate as one covariate could be substituting for another due to correlations and, therefore, might not be false at random. These results are limited to 1 data set, and the properties may be different in other (especially smaller) data sets. However, data sets for which exploratory covariate analysis is performed are usually larger (with respect to number of individuals) than our example (n = 74).

The definition of a good model varies depending on its purpose. Frequently the model is used for making predictions, and good predictive ability is required. Simplicity is also desired; covariates not improving the predictive ability should not be included. The stepwise search is based purely on statistical criteria, but it may be advantageous to use other approaches, such as including predictive ability as a model-building criterion, or considering clinical significance in the process. Because clinical significance criteria vary from drug to drug incorporating them is not straightforward; however, using Bayesian modeling clinical significance has been included. [27] Another approach to covariate modeling was recently taken by Kowalski and Hutmacher, [28] who suggested evaluating all possible relations from the estimated variance-covariance matrix of the full (including all covariates) model fit. However, the practical usefulness of this procedure is still unclear, and we did not evaluate it.

If the model is built for exploratory purposes, selection bias may be of minor importance; however, bias may represent a larger problem if the model is intended for prognostic purposes. To get unbiased estimates and predictions, one would preferably develop the model on one part of the data, estimate coefficients on a second part, and predict a third part. However, data are usually too precious to waste to be used in this fashion. Another option could be to prospectively consider known or expected characteristics of the drug in a best guess model and estimate these effects [21,22] to get unbiased coefficients. An exploratory search can then be carried out starting from this model. [17] Considering only plausible covariates in the search limits the covariate scope and reduces the risk of erroneous effects (due to multiple testing). In addition, the problems with accuracy and precision that can occur if 2 highly correlated covariates enter a model simultaneously [29] can be avoided by excluding 1 of 2 highly correlated covariates.

In summary, the properties of 2 stepwise covariate model-building procedures have been compared. No difference in performance could be seen in the data sets explored; however, there are differences with respect to practical advantages and disadvantages of the procedures. Upward biases in the stepwise estimated coefficients were evident, but these were small compared with the imprecision in the estimates under the true model.

ACKNOWLEDGMENTS Ulrika Wählby was supported by a grant from AstraZeneca R&D Mölndal, Sweden.

REFERENCES

1. Beal SL, Sheiner LB. NONMEM Users Guides. San Francisco, CA: University of California; 1992.

2. Maitre PO, Buhrer M, Thomson D, Stanski DR. A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam. J Pharmacokinet Biopharm. 1991;19(4):377-384.

3. Mandema JW, Verotta D, Sheiner LB. Building population pharmacokinetic-pharmacodynamic models. I. Models for covariate effects. J Pharmacokin Biopharm 1992;20(5):511-528.

4. Gallo JM, Laub PB, Rowinsky EK, Grochow LB, Baker SD. Population pharmacokinetic model for topotecan derived from phase I clinical trials. J Clin Oncol. 2000;18(12):2459-2467.

5. Troconiz IF, Armenteros S, Planelles MV, Benitez J, Calvo R, Dominguez R. Pharmacokinetic-Pharmacodynamic Modelling of the antipyretic effect of two oral formulations of ibuprofen. Clin Pharmacokinet. 2000;38(6):505-518.

6. Takama H, Tanaka H, Sudo T, Tamura T, Tanigawara Y. Population pharmacokinetic modeling and model validation of a spicamycin derivative, KRN5500, in phase 1 study. Cancer Chemother Pharmacol. 2001;47(5):404-410.

7. du Preez MJ, Botha JH, McFadyen ML, Holford NH. The pharmacokinetics of theophylline in premature neonates during the first few days after birth. Ther Drug Monit. 1999;21(6):598-603.

8. Zhou XJ, Sheiner LB, D'Aquila RT, et al. Population pharmacokinetics of nevirapine, zidovudine, and didanosine in human immunodeficiency virus-infected patients. The National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. Antimicrob Agents Chemother. 1999;43(1):121-128.

9. Zhou H, Chen TL, Marino M, et al. Population PK and PK/PD modelling of microencapsulated octreotide acetate in healthy subjects. Br J Clin Pharmacol. 2000;50(6):543-552.

10. Gwilt P, Tempero M, Kremer A, Connolly M, Ding C. Pharmacokinetics of levamisole in cancer patients treated with 5- fluorouracil. Cancer Chemother Pharmacol. 2000;45(3):247-251.

11. Gidal BE, Anderson GD, Rutecki PR, Shaw R, Lanning A. Lack of an effect of valproate concentration on lamotrigine pharmacokinetics in developmentally disabled patients with epilepsy. Epilepsy Res. 2000;42(1):23-31.

12. Ouellet D, Periclou AP, Johnson RD, Woodworth JR, Lalonde RL. Population pharmacokinetics of pemetrexed disodium (ALIMTA) in patients with cancer. Cancer Chemother Pharmacol. 2000;46(3):227-234.

13. Freyer G, Tranchand B, Ligneau B, et al. Population pharmacokinetics of doxorubicin, etoposide and ifosfamide in small cell lung cancer patients: results of a multicentre study. Br J Clin Pharmacol. October 2000;50(4):315-324.

14. Jonsson EN, Karlsson MO. Xpose--an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. Comput Methods Programs Biomed. 1999;58(1):51-64.

15. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. Pharm. Res. 1998;15(9):1463-1468.

16. Miller AJ. Subset Selection in Regression. Vol 40. 1st ed. London, UK: Chapman & Hall; 1990.

17. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. October 1999;52(10):935-942.

18. Vach K, Sauerbrei W, Schumacher M. Variable selection and shrinkage: comparison of some approaches. Statistica Neerlandica. 2001;55(1):53-75.

19. Harrell FE. Regression Modeling Strategies. New York, NY: Springer; 2001.

20. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. Br J of Math Stat Psychol. 1992;45:265-282.

21. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-387.

22. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059-1079.

23. Bruno R, Iliadis MC, Lacarelle B, et al. Evaluation of Bayesian estimation in comparison to NONMEM for population pharmacokinetic data analysis: application to pefloxacin in intensive care unit patients. J Pharmacokinet Biopharm. 1992;20(6):653-669.

24. Karlsson MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. J Pharmacokinet Biopharm. 1993;21(6):735-750.

25. Vozeh S, Maitre PO, Stanski DR. Evaluation of population (NONMEM) pharmacokinetic parameter estimates. J Pharmacokinet Biopharm. 1990;18(2):161-173.

26. Wahlby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn. 2001;28(3):231-252.

27. Wakefield J, Bennett J. The Bayesian Modeling of Covariates for Population Pharmacokinetic Models. J Am Stat Assoc. 1996;91:917-927.

28. Kowalski KG, Hutmacher MM. Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. J Pharmacokinet Pharmacodyn. 2001;28(3):253-275.

29. Bonate PL. The effect of collinearity on parameter estimates in nonlinear mixed effect models. Pharm Res. 1999;16(5):709-717.