A Novel Screening Method Using Score Test for Efficient Covariate Selection in Population Pharmacokinetic Analysis

Yixuan Zou[1], Chee M. Ng[1]

[1]College of Pharmacy, University of Kentucky, Lexington, KY

May 30, 2018

College of Pharmacy

Outline

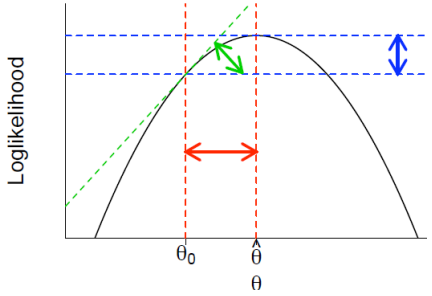
Covariates Selection in Population Pharmacokintics

- One of the important goals in population pharmacokinetics (PPK) is to establish the correct relationships between parameters and covariates
- PPK mainly uses nonlinear mixed effects modeling, which makes covariate-parameter relationship selection more difficult

Score Test Introduction

Problem

$H_0 : \theta = \theta_0$ (a p-vector), $\hat{\theta}$ is the maximum likelihood estimate (MLE) and $\ell(\theta)$ is the log likelihood at θ.



Likelihood Ratio Test (LRT)

$S_{LRT} = 2(\ell(\hat{\theta}) - \ell(\theta_0))$.

Wald Test

$S_{wald} = (\hat{\theta} - \theta_0)' I(\hat{\theta})(\hat{\theta} - \theta_0)$.

Score Test

$S_{score} = [s(\theta_0)]'[I(\theta_0)]^{-1}[s(\theta_0)]$.

These three statistics have asymptotic chi square distribution on p degrees of freedom.

Covariates Selection Method: LRT-based Stepwise Covariate Modeling (SCM)[1]

SCM: forward selection is followed with backward elimination based on LRT

- Pros: It is still among the most popular methods to do covariate selection in PPK due to good interoperability and relative economic computation
- Cons: The model runs in NONMEM could be prohibitively large with large number of covariates to be tested. The selection bias and inflated P-value are also issues

[1] Jonsson and Karlsson 1998.

Covariates Selection Methods: Wald Approximation Method (WAM)[2]

WAM: full covariate model fit to calculate the approximate LRT statistic for all possible restricted models using Wald's approximation. Final model is selected based on the maximum (actual) Bayesian information criterion (BIC) derived from NONMEM model fits for the 10–15 most probable models. It screens all possible submodels only by fitting the full model with all covariates.

- Pros: Fewer NONMEM runs (10-15) compared to SCM
- Cons: A full covariate model with a covariance matrix is required, which is hard to obtain and MLE could be easily trapped into local minimum, so covariance matrix may not be reliable even if it can be obtained

[2] Kowalski and Hutmacher 2001.

Other Covariates Selection Methods

There are other covariate selection methods, such as genetic algorithem (GA)[3], LASSO[4] etc. These methods are beyond the scope of this presentation. For a full dicussion about the pros and cons of different methods, please refer to the referenced review paper[5]

[3] Bies et al. 2006.
[4] Ribbing et al. 2007.
[5] Hutmacher and Kowalski 2015.

Covariates Selection Methods: Score Test-Based Covariate Method

- Score test-based covariate method has never been used in PPK with nonlinear mixed effect modeling
- **No model fit** is needed for score test after the base model has been successfully identified. It is potentially very useful to efficiently screen the potential covariate-parameter relationships in the presence of long model run due to complex population structural model and/or large number of tested covariates

Objective

The objectives were

- To conduct the type I error analysis and power analysis of score test in nonlinear mixed effect modeling
- To develop the first score test-based covariate selection method in the PPK using nonlinear mixed effects modeling approach

Score Test in Nonlinear Mixed Effects Model

Notations

- Assume $\ell(\gamma, \boldsymbol{\theta_c})$ is the log likelihood function for a certain PPK covariate model, where $\gamma = (\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$, $\boldsymbol{\theta}$ denotes fixed effects parameters for the base model, $\boldsymbol{\Omega}$ denotes the covariance matrix of inter-individual random effects, $\boldsymbol{\Sigma}$ denotes the covariance matrix of intra-individual random effects, and $\boldsymbol{\theta_c}$ denotes covariate parameters
- γ_0 is the MLE of the model when $\boldsymbol{\theta_c} = 0$
- Score function is defined as $S(\gamma, \boldsymbol{\theta_c}) = (\frac{\partial \ell}{\partial \gamma}, \frac{\partial \ell}{\partial \boldsymbol{\theta_c}})$
- The observed fisher information matrix (negative hessian matrix) is

$$I_{obs} = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial^2 \gamma} & \frac{\partial^2 \ell}{\partial \gamma \partial \boldsymbol{\theta_c}} \\ \frac{\partial^2 \ell}{\partial \gamma \partial \boldsymbol{\theta_c}} & \frac{\partial^2 \ell}{\partial^2 \boldsymbol{\theta}} \end{bmatrix}$$

Score Test in Nonlinear Mixed Effects Model

Hypothesis Testing

To test $H_0 : \boldsymbol{\theta_c} = \mathbf{0}$, the score statistic is

$$S_{score} = S(\gamma, \boldsymbol{\theta_c})' I_{obs}(\gamma, \boldsymbol{\theta_c})^{-1} S(\gamma, \boldsymbol{\theta_c})|_{\gamma=\boldsymbol{\gamma_0}, \boldsymbol{\theta_c}=\mathbf{0}},$$

since $\boldsymbol{\gamma_0}$ is the MLE under the H_0, thus

$$S_{score} = (\mathbf{0}, \frac{\partial \ell}{\partial \boldsymbol{\theta_c}}|_{\boldsymbol{\theta_c}=\mathbf{0}})' I_{obs}(\boldsymbol{\gamma_0}, \mathbf{0})^{-1} (\mathbf{0}, \frac{\partial \ell}{\partial \boldsymbol{\theta_c}}|_{\boldsymbol{\theta_c}=\mathbf{0}}),$$

Under H_0 hypothesis and some regularity conditions, S_{score} and LRT statistic have the same asymptotic chi-square distribution.

Model Selection Criterion

Similar to BIC is penalized score chi square (ScoreP) used in the model selection process.

Connection between BIC and ScoreP

By omitting the constant term for each model,

$$BIC = -S_{LRT} + klog(n) \text{ and } ScoreP = -S_{score} + klog(n),$$

where $S_{LRT} = -2(LL_b - LL_c)$, S_{score} is the score statistic, LL_b is the log likelihood for the base model and LL_c is the log likelihood for the covariate model, k is the number covariate parameters added in the model and n is the number of observations in the dataset.

$\uparrow$ BIC and ScoreP $\rightarrow$ Model Performance $\downarrow$. We are going to use ScoreP to identify the non-informative covariates.

Score Function and Observed Fisher Information Matrix Calculation

Finite Difference

$$\frac{\partial \ell}{\partial \theta_i}|_{\theta_i=a} = \frac{\ell(a+h) - \ell(a-h)}{2h}, \frac{\partial^2 \ell}{\partial^2 \theta_i}|_{\theta_i=a} = \frac{\ell(a+h) + \ell(a-h) - 2\ell(a)}{-2h^2}$$

$$\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}|_{\theta_i=a,\theta_j=b} = \frac{\ell(a+h,b+h) - \ell(a+h,b-h) - \ell(a-h,b+h) + \ell(a-h,b-h)}{4h^2}.$$

Implementation in NONMEM

Step size is set to be $h = 0.0001$, METHOD=FOCE, SIGL=10, NSIG=3 is used for the fitting of base model without any covariates, each log likelihood function is evaluated in NONMEM using the option MAXEVAL=0 without any model fit.

Type I Error and Power Analysis of Score Test

One compartment PK model with IV bolus

- Typical values $TV_{CL} = 0.1L/h$, $TV_V = 1.0L$
- Inter-individual variability $\omega^2_{CL} = \omega^2_V = 0.1$
- Intra-individual variability only has the proportional error $\sigma^2 = 0.1$
- Sample size 50 and 200 with intensive sampling design (six sampling points per subject)
- Two types of analysis
  - Type I Error: The dataset was simulated without any covariates. The actual significance level[6] for score test and LRT was compared using nominal significance levels 0.1, 0.05 and 0.01
  - Power Analysis: The dataset was simulated with covariate weight on clearance, the covariate parameter is $CL_{WT} = 0.25, 0.75$ respectively, the power of score test and LRT was compared using significance levels 0.1, 0.05 and 0.01

[6] Wählby, Jonsson, and Karlsson 2001.

Type I Error and Power Analysis of Score Test

- The actual significance level is calculated using the estimated upper tail probabilities of the statistic S under null hypothesis $H_0 : \theta = 0$

$$\Sigma_{i=1}^{500} I[S > \chi_1^2(1-\alpha)|H_0]/500$$

- The empirical power is calculated using the estimated upper tail probabilities of the statistic S under alternative hypothesis $H_1 : \theta = \theta_0$

$$\Sigma_{i=1}^{500} I[S > \chi_1^2(1-\alpha)|H_1]/500$$

All Possible Subset Screening Method Based on Score Test

Algorithm

1. Identify the best base model without any covariates and fit the model in NONMEM to get the MLE γ_0

All Possible Subset Screening Method Based on Score Test

Algorithm

1. Identify the best base model without any covariates and fit the model in NONMEM to get the MLE γ_0
2. Add all the pre-specified covariate parameters into the base model and use finite difference method to get the score function and observed Fisher information matrix for each parameter

All Possible Subset Screening Method Based on Score Test

Algorithm

1. Identify the best base model without any covariates and fit the model in NONMEM to get the MLE γ_0
2. Add all the pre-specified covariate parameters into the base model and use finite difference method to get the score function and observed Fisher information matrix for each parameter
3. Calculate the score statistic for each combination of covariates and the corresponding ScoreP

All Possible Subset Screening Method Based on Score Test

Algorithm

1. Identify the best base model without any covariates and fit the model in NONMEM to get the MLE γ_0
2. Add all the pre-specified covariate parameters into the base model and use finite difference method to get the score function and observed Fisher information matrix for each parameter
3. Calculate the score statistic for each combination of covariates and the corresponding ScoreP
4. Selection and elimination of the non-informative covariate models: Select the model with the biggest ScoreP, which has the least information, the covariates which are not identified in the model are kept

Case Study Dataset Summary

Rituximab dataset was simulated with the original PK sampling design using the model[7]:

- $N = 107$ with IV rituximab 1000 mg on days 1 and 15
  - On day 1, rituximab was given IV over about 255 minutes, and PK samples were collected at predose, 3 hours, the end of infusion, and 6 and 48 hours after the start of the infusion
  - On day 15, the second dose was given IV over about 195 minutes, and PK samples were collected at predose, 3 hours, the end of infusion, and 6, 48, 336, 1088, 2352, and 3696 hours after the start of the infusion

[7] Ng et al. 2005.

Case Study Dataset Summary

Model used for the simulation:

- Two compartment linear PK model with the following paremeter-covariate relationships: body surface area (BSA) and Gender (GEN) on both CL and V_c

$$CL = TV_{CL}(\frac{BSA}{1.72})^{CL_{BSA}} \times exp(CL_{GEN}) \times exp(\eta_{CL}),$$

$$V_c = TV_{V_c}(\frac{BSA}{1.72})^{(V_{c\,BSA})} \times exp(V_{c\,GEN}) \times exp(\eta_{V_c})$$

- Mixed additive and proportional error model was used
- Four covariates, AGE, GEN, BSA and baseline B cell levels (BCF) in the simulated dataset were tested

Type I Error Analysis

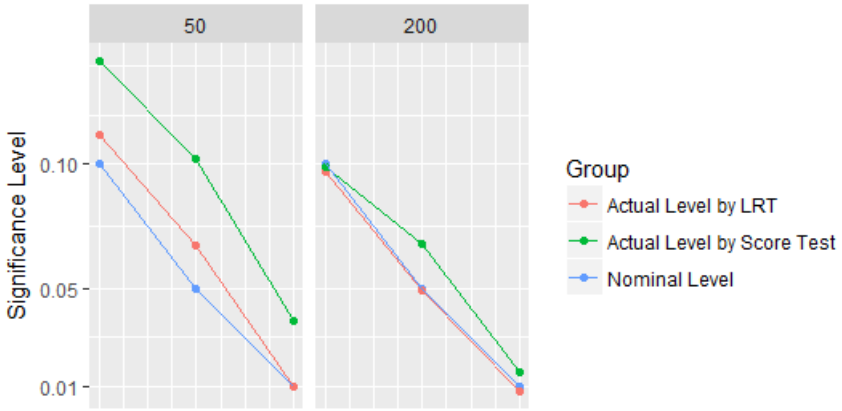


Figure 1: Actual significance level of the score statistic and LRT statistic based on 500 replications generated from one compartment model

Power Analysis

		$\theta_{CL_{WT}} = 0.75$			$\theta_{CL_{WT}} = 0.25$		
n	Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
50	LRT	0.968	0.924	0.816	0.366	0.238	0.084
	Score Test	0.962	0.926	0.828	0.378	0.272	0.15
200	LRT	1	1	1	0.744	0.644	0.398
	Score Test	1	1	1	0.764	0.644	0.42

Table 1: Empirical power of the score statistic and LRT statistic based on 500 replications generated from one compartment model

Model Screening Based on ScoreP

Rank by ScoreP	Theta Selected	ScoreP	Score Chi Square	BIC	LRT Chi Sqaure	Rank by BIC
1	$\mathbf{[CL_{AGE}, Cl_{BCF}, V_{cAGE}, V_{cBCF}]}$	**25.55**	**2.35**	**25.32**	**2.58**	**1**
2	$[CL_{AGE}, V_{cAGE}, V_{cBCF}]$	20.63	0.30	20.56	0.37	2
3	$[Cl_{BCF}, V_{cAGE}, V_{cBCF}]$	18.91	2.01	18.89	2.03	3
4	$[CL_{AGE}, Cl_{BCF}, V_{cAGE}]$	18.72	2.20	18.51	2.41	4
5	$[CL_{AGE}, Cl_{BCF}, V_{cBCF}]$	18.62	2.31	18.39	2.53	5
6	$[V_{cAGE}, V_{cBCF}]$	13.81	0.14	13.80	0.15	6
7	$[CL_{AGE}, V_{cAGE}]$	13.78	0.18	13.71	0.24	7
8	$[CL_{AGE}, V_{cBCF}]$	13.69	0.26	13.62	0.33	8
9	$[CL_{AGE}, Cl_{BCF}, V_{cAGE}, V_{cBCF}, V_{cBSA}]$	12.71	22.17	4.75	30.13	10
10	$[Cl_{BCF}, V_{cAGE}]$	12.10	1.85	12.08	1.87	9

Table 2: Ten worst models selected by ScoreP

The remaining covariate parameters, i.e. those that are not selected by this process CL_{BSA}, Cl_{GEN}, V_{cBSA}, V_{cGEN} are exactly the true covariates in the original model used for simulation; SCM approach with forward inclusion (p=0.05) and backward elimination (p=0.01) also found the true covariates, but it took 38 NONMEM runs

Relationship between Score Chi Square and LRT Chi Square

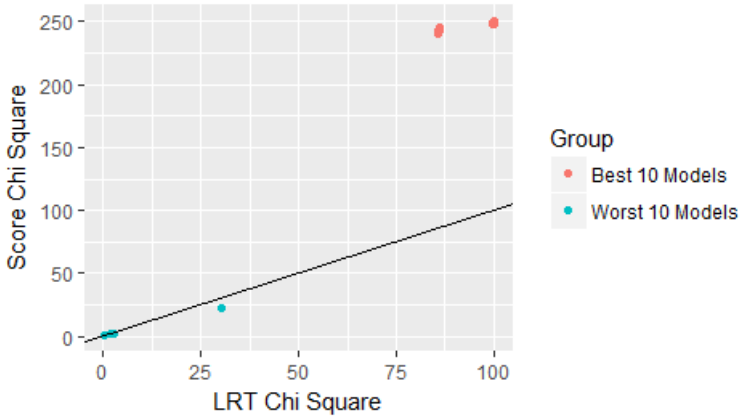


Figure 2: LRT chi square and score chi square for the best 10 models and the worse 10 models

Discussion and Conclusions

- Preliminary type I error and power analyses were conducted for score test under the nonlinear mixed effects setting
  - Score test had comparable performance with LRT in power analysis, but had inflated type I error when sample size was small
  - Regarding to score test's superior computational efficiency (no model fit is required), it may serve as a good screening tool in the forward selection step for SCM when different forms of covariate parameters and large number of covariates need to be tested instead of using LRT
- A fast covariate screening method was proposed based on score test. It could identify those uninformative covariates all at once after base model was obtained without any further NONMEM runs
- Further study is ongoing on the validation process and the performance of this method in different scenarios with real clinical datasets

Thank You!

- Bies, Robert R et al. (2006). "A genetic algorithm-based, hybrid machine learning approach to model selection". In: *Journal of Pharmacokinetics and Pharmacodynamics* 33.2, pp. 195–221.
- Hutmacher, Matthew M and Kenneth G Kowalski (2015). "Covariate selection in pharmacometric analyses: a review of methods". In: *British journal of clinical pharmacology* 79.1, pp. 132–147.
- Jonsson, E Niclas and Mats O Karlsson (1998). "Automated covariate model building within NONMEM". In: *Pharmaceutical research* 15.9, pp. 1463–1468.
- Kowalski, Kenneth G and Matthew M Hutmacher (2001). "Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test". In: *Journal of pharmacokinetics and pharmacodynamics* 28.3, pp. 253–275.
- Ng, Chee M et al. (2005). "Population pharmacokinetics of rituximab (anti-CD20 monoclonal antibody) in rheumatoid arthritis patients during a phase II clinical trial". In: *The Journal of Clinical Pharmacology* 45.7, pp. 792–801.

- Ribbing, Jakob et al. (2007). "The lasso—a novel method for predictive covariate model building in nonlinear mixed effects models". In: *Journal of pharmacokinetics and pharmacodynamics* 34.4, pp. 485–517.
- Wählby, Ulrika, E Niclas Jonsson, and Mats O Karlsson (2001). "Assessment of actual significance levels for covariate effects in NONMEM". In: *Journal of pharmacokinetics and pharmacodynamics* 28.3, pp. 231–252.