Review Article **Open Access**

Covariate Modeling in Population PK/PD Models: an Open Problem

Davide Verotta[1,2]*

[1]*Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California, USA*
[2]*Department of Biostatistics, University of California, San Francisco, California, USA*

Abstract

Covariate Model selection for population PK/PD models represents a daunting task because of the large variety of possible alternative covariates that can enter a structural model, the different models that can express the relationship parameter/covariates, and the number of alternative models that can be considered. After describing the problem and briefly reviewing the past literature dedicated to the solution of the problem we use simulations to show the limitations of current approaches and propose an alternative based on the sequential use of Bayesian Trans Dimensional Models. Although the alternative mollifies the dimensionality problem associated with covariate selection, we argue that the overall approach to covariate modeling within PKPD models might need to be reconsidered.

Keywords: Population PKPD; Bayes; Trans dimensional models

Introduction

In the last decade PK/PD modeling has reached a new level of maturity, acceptance, and influence and is increasingly considered a central technology for improving both drug development and clinical therapy. Pharmacological effects can be measured at the biochemical, cellular, organ, individual, or clinical level. The integration of diverse pharmacological effects into a coherent picture of drug action is one of the goals of current research in modeling of PK/PD. Two communicating dynamic systems, viz., pharmacokinetics and pharmacodynamics, determine the response of a given individual to a particular drug. Pharmacokinetics (PK) includes drug absorption, distribution, biotransformation, and elimination (processes that involve receptors, transporters and enzymes). Pharmacodynamics (PD) includes the interaction of drug species with enzymes or receptors of a target cell (or bacteria or v.irus) and the often-complex chain of molecular and physiological events, which ultimately determine the physiological responses to the drug. The modeling of PK [1] has a long tradition, and the history of dynamic PK/PD models goes back (as far as we know) to the seminal paper of Segre [2], while the theoretical framework for PK/PD models is described in a number of publications [3-5]. General classes of PK/PD models, which allow a unified approach to modeling, and investigation of PK/PD systems, have been described in [6,7]. What complicates PKPD modeling is the fact that individuals differ in their responses to a given drug. These differences can be expressed as individual differences in PK and PD, which, in turn, can partially be explained by variation in such factors as age, gender, weight, or genetic components. The elucidation and estimation of such relationships is carried out using PK/PD data from samples of individuals from populations of interest. Population experiments are often conducted in a very different manner than those carried out on a small number of individuals where rich data sets can be collected, since the resources required to execute an experiment that fully quantifies each individual are prohibitive. As a result, non-linear hierarchical mixed-effects models are routinely used in attempts to identify PK/PD models from a population study [8-10]. Determining the relations between the parameters of a population PKPD model and covariates is a major aim of population PK/PD modeling. The covariates can improve the estimates precision, explain part of the inter-individual variability, and may also increase the mechanistic interpretability of the model or generate hypothesis. An appropriate covariate might be selected according to its clinical importance, although this is very rare given the semi-empirical nature of PKPD models and limited knowledge about drug/covariate relationships, or, as it is the norm, identified from the statistical significance of its relation(s) with parameter(s).

The main purpose of this paper is to show out how the currently accepted paradigm for the inclusion of covariates in PKPD models constitutes an open problem, since the accepted model for the relationship parameters/covariate can easily generate intractably complex models selection problems. The paper is organized as follows: (i) briefly describe Bayesian methodology, that is at the core of the methodologies we investigate, (ii) describe the covariate selection problem associated with PKPD models, showing how the accepted paradigm that associates covariates to each of the parameters in the model, generates an extremely complex model selection problem, (iii) show, by means of simulations, how current approaches have difficulties in dealing with the dimensionality of the covariate model families, (iv) propose an alternative that reduces the dimensionality problem, (v) end with a short discussion.

Methods

Bayesian estimation

We briefly summarize the main concept and problem involved in the Bayesian approach to inference, see, e.g., [11] for details. The Bayesian approach is developed in the presence of observations Y whose value is initially uncertain and described though a probability distribution that depends on some parameters θ. The Bayesian approach incorporates this information into the analysis through a density $p(\theta)$ that represents the prior knowledge before the observations Y are collected. Inference

---

*Corresponding author: Davide Verotta, PhD, Department of Bioengineering and Therapeutic Sciences University of California, Box 0912, San Francisco, CA 94143, USA, E-mail: davide.verotta@ucsf.edu

is based on the probability distribution of θ after observing Y and is obtained through its posterior distribution using Bayes' theorem. In trans-dimensional models (TDM), the prior uncertainty regarding the model structure is acknowledged, and the model itself becomes a parameter. For example, in variable selection we wish to select the "best" subset of all predictor variables on which to regress the response variable; the number of selected variables and the variables themselves, which together define the model structure, are unknown parameters. These models are referred to as trans-dimensional because each possible model structure has a potentially distinct set of coefficients/parameters associated with it. In general, these sets will be of different sizes and so as we move from one model to another, the parameter space of interest changes dimension. Until relatively recently the main restriction of Bayesian inference has been that the computation of the posterior is analytically intractable other than for relatively simple cases. Besides the controversy about the appropriateness of incorporating historical information, $p(\theta)$ in the analysis and specifying a complex model for the data, the main restriction (a fatal flaw, for practical purposes) of this approach has been the computation of this posterior, which, other than for relatively simple cases, is analytically intractable. Computation of the posterior involves solving multi-dimensional integrals, and many computational methods have been devised to do so. In the last ten years a large body of work on stochastic methods has made possible the applications of the Bayesian approach to complex problems. Markov Chain Monte Carlo (MCMC) methods [12-16] opened the way to analysis of complex models. Coupled with the advent of MCMC methods for posterior computation, the development and application of Bayesian methods for model uncertainty [13,17-21] have seen remarkable evolution over the past decade. Some of these methods have been applied to PK/PD population models in [22,23] as well as to models for individual responses [24,25]. In general the field has achieved a high level of maturity, and computing systems and methods are available for a great number of Bayesian problems, we mention in particular BUGS [26], PKBUGS and JUMP for reversible jump MCMC [27].

Model selection

Variable selection in, for example, linear regression [28] consists of a set of potential predictors $\{X_i\}$ for a response Y of interest: each model under consideration corresponds to a distinct subset of $\{X_i\}$, and is of the form

$$Y = \theta_1 + X_\gamma \theta_\gamma + \varepsilon \tag{1}$$

where, X_γ is the design matrix whose columns correspond to the γ-th subset of covariates, $\{X_i\}_\gamma$, and θ_γ is the vector of regression coefficients for the γ-th subset. In PK/PD modeling the predictors (covariates) are not directly related to the response, but instead to the parameters of the model [29]. This fact immediately complicates the modeling selection problem, because each parameter can be related to different covariates. Moreover, different models can be used to express the relationship between a parameter and covariates (for example, clearance might be related to creatinine clearance linearly, but volume might be related to weight non-linearly). An instance of a general PK/PD covariates model is of the form

$$Y = f\left(t, \vartheta_\gamma, g_\eta\right) + \varepsilon$$
$$\vartheta_{1,\gamma_1} = g_{\eta_1}\left(\{X_i\}_{\gamma_1}, \theta_{\gamma_1}\right)$$
$$\cdots$$
$$\vartheta_{m,\gamma_2} = g_{\eta_m}\left(\{X_i\}_{\gamma_m}, \theta_{\gamma_m}\right) \tag{2}$$

Where, m is the number of structural parameters in the model, $\gamma = \left(\gamma_1, \cdots, \gamma_m\right)$, the collection of sub-sets of covariates for each parameter, $g_\eta = \left(g_{\eta_1}, \cdots, g_{\eta_m}\right)$ and is the collection of functions characterizing the relationship of the parameters to covariates. For simplicity we omit details on the specification of individual specific random effects in equation (2), we refer to, e.g., [25] for the, Bayesian, formulation of the mixed effect model.

Consider the simplest possible PKPD model a single compartment model, with two parameters:

$$Y = D / \vartheta_1 \; e^{-\vartheta_1/\vartheta_2 t} \tag{3}$$

where ϑ_2 represents volume of distribution, clearance, D the drug dose and CrCl (Creatinine clearance). Now assume that in the clinical trial one has collected just four covariates, say Age, Weight, Height and Creatinine clearance (CrCl); assuming that only one covariate can influence a parameter the possible alternative models to consider are 4 for each parameter generating $4 \times 4 = 16$ possible combinations. If we relax the assumption that only one covariate can influence a parameter, then for each parameter we have 4 models with 1 covariate, 6 with 2, 4 with 3, and 1 with 4 covariates, that is 15 models, and correspondingly $15^2 = 225$ models for two parameters. In absence of prior information, for c covariates the number of possible covariate models that can be considered for a parameter in a PKPD model can be indexed as a family $\{J_k\}$. The dimension of the family equals the sum of the possible combination of c elements taken 1, 2, …,c at a time, that is

$$\#\{J_k\} = \sum_{i=1}^{c}\binom{c}{i} = \sum_{k=1}^{c}\frac{c!}{i!(c-i)!} \tag{4}$$

For m parameters the total number of models to consider, indexed as a family $\{M_k\}$ equals the number of permutations with repetition taken p at a time from $\{J_k\}$, the dimension of is $\{M_k\}$

$$\#\{M_k\} = \left(\#\{J_k\}\right)^p = \left(\sum_{i=1}^{c}\binom{c}{i}\right)^p \tag{5}$$

For a relatively small PK model with 4 parameters and 10 covariates the number of possible models is rather large: $\#\{M_k\} = \left(\#\{J_k\}\right)^4 = 1024^4 = 109{,}951{,}162{,}776$. A rather large number, to say the least, that does not even consider possible interactions between covariates or alternative models for the relationship covariate/parameters.

Model selection

Obviously the number of models indicated by (5) is too large to attack the problem by brute force. Putting aside the problem of selection bias that such a rather absurd number of alternative models can generate (exaggerating the importance of covariates is all but guaranteed with such a number of alternatives [30]), the main concern in the PKPD literature has been in trying to identify a practical and reliable selection method. There are various statistical approaches to model selection, of which one of the most common is stepwise addition (covariates are added one at a time until a maximum, large, number allowed in the model is reached) followed by backward elimination (the model is "pruned" by removing one of the selected covariates at a time until the removal corresponds to a significant increase in the

objective function value) [31]. The adaptation of the methodology to PKPD is generally carried out by introducing three simplifications.

Algorithm 1: (i) First, one considers one parameter at a time, adding covariates iteratively using stepwise addition.

(ii) Second, covariates are only added to the model if the addition corresponds to a significant drop in the objective function accordingly to a statistical selection criterion, e.g. [32].

(iii) Third, the backward elimination process is not carried out.

Adopting (i) reduces the total number of models to be considered for each parameter to $\sum_{i=1}^{c}\binom{c-i+1}{i}$ and the total number now growths linearly with *p*, not geometrically:

$$\#\{M_k\} = \sum_{i=1}^{c}\binom{c-i+1}{i}\times p \qquad (6)$$

The approach just described can easily be criticized. Simple stepwise addition/elimination is often ineffective even when used with much simpler statistical models. Using the PKPD variant just described it is rather obvious that whatever model is selected it would not represent the best possible one. However the approach has the advantage of making the intractable problem corresponding to equation (5) into something that can somewhat be managed, and it has became a sort of a standard in the literature. For the previous example of 4 parameters and 10 covariates the number of possible models drops from $\#\{M_k\}$ =1024^4, to $\#\{M_k\} = \sum_{i=1}^{10}\binom{10-i+1}{i}\times 4 = 143\times4=572$, a number that is further reduced under (ii) above.

Alternatives methods to stepwise addition have been proposed, see [33] for comparisons, and in addition a number of methods decoupling the search for covariates from the overall likelihood function attempt to mollify the intractability of the task, see [34] using generalized additive models, or [35] using cluster analysis. Here we consider a methodology that takes advantage of the algorithms and formalism of trans-dimensional models (TDM). Briefly, to define a TDM one starts from a family of possible models, $\{M_k\}$, where each model consists of a distribution $p(Y|\theta_k, M_k)$ where, θ_k is a vector of parameters with dimension d_k. The Bayesian approach proceeds by assigning a prior probability distribution $p(\theta_k|M_k)$ to the parameters of each model, and a prior probability $p(M_k)$ to each model. This prior formulation induces a joint distribution over the data, parameters, and models. In effect, these priors serve to embed the various separate models within one large hierarchical mixture model. Conditioning on the data, Y yields the posterior model probabilities $p(M_k|Y)$ through the marginal likelihood of M_k, obtained using MCMC [36]. The priors $p(\theta_k|M_k)$ and $p(M_k)$ provide an initial representation of model uncertainty, the model posteriors, provides a complete representation of post-data model uncertainty that can be used for a variety of inferences and decisions. The TDM framework, proposed in PKPD by [27,37], can help with the problems associated with addition/deletion, however as we will see, it does not solve the problems related to the high-dimensionality of the model space.

I. Simulation 1

We consider a simple model of the form of equation (3), where:

$$\vartheta_1 = \theta_1 + \theta_2 CrCl$$
$$\vartheta_2 = \theta_3 + \theta_4 Weight \qquad (7)$$

Weight and CrCl are the individual weight and creatinine clearance, respectively. θ_1=1, θ_2=0.2, θ_3=1, θ_4=.2, that describes a linear relationship of volume (ϑ_2) and clearance (ϑ_1) with Weight and CrCl. We simulate a study with 100 individuals, and assume lognormal distribution for the individuals' parameters with 50% inter-subject variability; 10% proportional error model for intra-subject variability. We assume that c=8 covariates are collected (corresponding to *Sex, Ethnicity, Genotype 1, Genotype 2, Weight, Height, Age, CrCl*). To generate a simulated covariates data set we use a multivariate normal model with a single correlation parameter ρ. The correlation between the i-th and j-th covariate equals $\rho^{|i-j|}$, and the variance of each covariate is 1. For implementation of the model we rely on the interface and set up developed by [27]. We consider two chains run a 25,000 iterations burn-in, and then a further 175,000 iterations from which we shall draw posterior inferences. The analysis involves one Markov chain, initialized corresponding to the model with no covariates included.
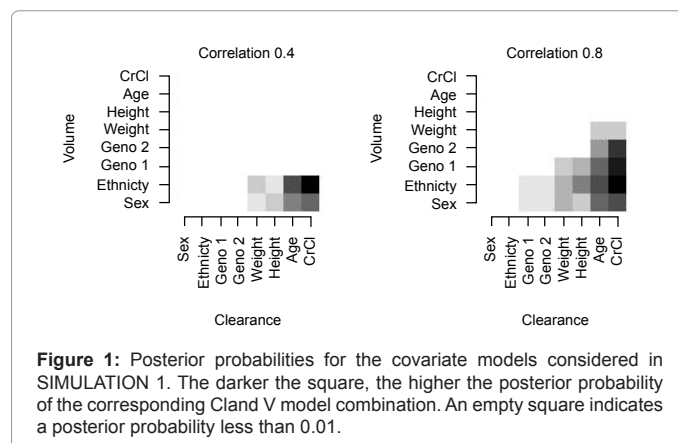
Figure 1 shows the posterior probabilities corresponding to the 64 instances of the TDM. The darker the square, the higher the posterior probability of the corresponding Cl and V model; an empty square indicates a posterior probability less than 0.01. Note the effectiveness of the methodology in detecting the combination that shows the correct relationship of Cl to CrCl and V and to Weight. The plots also reveal a second main motivation for the use of TDM: when the correlation between covariates increases posterior probabilities for models corresponding to such covariates increases. In these cases model averaging, might be a preferable alternative to selecting a unique "best" model.

II. Simulation 2

We use the same set up, but now the data are simulated using:

$$\vartheta_1 = \theta_1 + \theta_2 CrCl + \theta_5 Heigth$$
$$\vartheta_2 = \theta_3 + \theta_4 Weight + \theta_6 Sex \qquad (8)$$

Considering just two covariates increases the number of possible models quite dramatically, for each parameter we have 52 (8 alternative covariate models with 1 covariate, plus 45 with 2). The total number of

Figure 1: Posterior probabilities for the covariate models considered in SIMULATION 1. The darker the square, the higher the posterior probability of the corresponding Cl and V model combination. An empty square indicates a posterior probability less than 0.01.

possible models is 52^2=2704. The analysis now involves two Markov chains, initialized corresponding to the model with no covariates included, and to the model with highest posterior probability when considering one covariate. As suggested in [38], we check convergence qualitatively by plotting 'model state' against iteration number for each Markov chain and then quantitatively by comparing tables of posterior model probabilities from each chain. In addition at the end of the runs we simply counted the number of visits for each covariate model. Generating a data set with ρ =0.8 and 100 subject, and using, as above, a total number of 200,000 iterations a relatively large number of models, approximately 12%, receives less than 5 visits, a number that we consider too low to guarantee a proper evaluation of those model (note that the Markov chain should not only visit the models, but also spend "enough time" in each one to estimate the parameter associating covariates to parameters correctly). Doubling the length of the chains did not improve the situation significantly, making it questionable if the TDM search is capable to fully explore model spaces of this dimension.

The simulation points out how relatively quickly the methodology can get in trouble. The approach seems to be rather unfeasible using a model as simple as (3), and it would seems all but quite intractable with just slightly more complicated PK models and a larger number of covariates.

An alternative to the use of a "brute force" TDM, where all possible models are candidates, is to adapt a sequential approach, where model with 1, 2, … covariates are considered sequentially, as spelled out by the following simple algorithm.

Algorithm 2: *Fit the base model with no covariates, B_0*

For j =1,…

Define the family of covariate models with j covariates obtained conditional on B_{j-1} $\left\{ M_k^j \right\}$.

Fit the Trans-Dimensional Model $\left\{ M_k^j \right\}$, and obtain the model with highest posterior probability, *Bj*

Stop when the increase of posterior probability of B_j does not reach a desired threshold (or a model selection criterion indicates that B_{j-1} is the preferred model).

The family $\left\{ M_k^j \right\}$ corresponds to all the possible permutations with repetition of the models that can be obtained by adding 0 or 1 additional covariate to each parameter, starting from B_{j-1}. The approach has the advantage of bringing the dimension of the model space under control: at each step the dimension of the TDM is constant: only c models need to be considered for each parameter, and the total number of models for each step is c^p.

When we apply the sequential approach to the data generated for simulation 2, we need to perform three steps (corresponding to 1, 2, 3, covariates, and 64 alternative covariate models for each step) before the Akaike criterion [32] indicates that the model with two covariates is selected to represent the data. The result indicates a cluster of models grouped around the correct quadruplet *Sex, Weight, Height, CrCl* (Equation (8)) thoroughly follows the correlation structure adopted in the simulation. For the simulated data set the model with highest posterior has *Sex and Weight for Cl, and Weight and Sex for V*.

Discussion

The main purpose of this paper is to point out how the currently accepted modeling paradigm for the inclusion of covariates in PKPD models, that is model (2), can easily generate intractably complex model selection problems. The examples shown in the paper illustrate the fundamental problems with the approach. Problems that we summarize below.

Equations (5) indicates how the total number of models can grow rather astronomically with the number of parameters in the model and the number of covariates. However that number is still an underestimate of the models that one might, in principle, want to consider. In the computation of the dimension of the space of covariate models we have not included (i) interactions between covariates, that is models of the form:

$$\vartheta_i = \theta_k + \theta_{k+1} Cov_p + \theta_{k+2} Cov_q + \theta_{k+2} Cov_r Cov_s$$

Where Cov_j indicates the *j*-th covariate. We have also omitted (ii) alternative models for the relationship between parameters and covariates. For example polynomial model of the form:

$$\vartheta_i = \theta_k + \theta_{k+1} Cov_l^{\theta_{k+2}}$$

instead of the linear models of the form $\vartheta_i = \theta_k + \theta_{k+1} Cov_l$. In addition, to give a complete picture of the potential complexity involved, we (iii) strictly speaking did not consider PKPD models, but only the simplest possible PK model. Introducing a PD component adds parameters to the model, and therefore further adds to the number of covariate relationships that need to be elucidated. And finally, (iv), if anything PKPD models are getting more complex in the literature [5] and the complexity again translates into more parameters and even more possible relationships with covariates.

The current approach to "solve" the problem basically ignores it. A univariate stepwise addition is adopted, where covariates are attached to one parameter at a time (ALGORYTHM 1 above, and currently distributed software implementing the simple algorithm [39]). However univariate stepwise addition is an approach that is not only not guaranteed to identify the covariate model maximizing the likelihood given the data, but it is not even guaranteed to obtain the same result if a different order of the parameters is considered in the search. (For example, if in a model with, say, four parameters one add covariates to the 1st, 2nd, 3rd and than 4th parameter, the final result will be different if one chooses to add to the 2nd, 3rd, 4th and than 1st).

Given the complexity of the problem it is questionable if slightly more sophisticated versions of stepwise addition could be more successful. Even for much simpler univariate problems stepwise addition, followed by backward elimination, is not guaranteed to obtain a global maximum, and a number of alternatives have been described [40,41] trying to solve its limitations. The methodology described in [27,37] for PKPD models is basically an adaptation of those approaches, and ALGORITHM 1 shows a further elaboration that attempts to mollify the dimensionality of the problem associated with the covariate model (2).

The culprit for this situation is model (2). Its formulation makes a univariate problem into a multivariate one and explodes the dimensionality of the problem geometrically with the number of parameters in the model. The solution to this might simply is to formulate the problem into a relationship of covariates with the response, not with the model parameters. For example, a representation of the form:

$$Y = f(t, \vartheta) g(X_i, \theta) + \varepsilon$$

Might prove to be as effective to incorporate the effect of covariates

on PKPD resposes. The investigation of such models is of course beyond the scope of the present paper. This paper might instead help stimulating discussion on model (2), a model that can be easily formulated in theoretical terms but that has a degree of flexibility, and admits so many variants, that might be unacceptable from both a scientific and statistical point of view.

Acknowledgements

References

1. Gibaldi M, Perrier D (1982) Pharmacokinetics. (2ndedn), New York: Marcel Dekker.

2. Segre G (1968) Kinetics of interaction between drugs and biological systems. Farmaco Sci 23: 907-918.

3. Sheiner LB, Beal SL, Sambol NC (1989) Study design for dose-ranging. Clin Pharmacol Ther 46: 63-77.

4. Veng-Pedersen P, Gillespie WR (1988) A system approach to pharmacodynamics I: theoretical framework. J Pharma Sci 77: 39-47.

5. Csajka C, Verotta D (2006) Pharmacokinetics-Pharmacodynamics Modeling: history and perspectives. J Pharmacokinet Pharmacodyn 33: 227-279.

6. Jusko WJ, Ko HC (1994) Physiologic indirect response models characterize diverse types of pharmacodynamic effects. Clin Pharmacol Ther 56: 406-419.

7. Verotta D, Sheiner LB (1995) A general conceptual model for non-steady state pharmacokinetic/pharmacodynamic data. J Pharmacokinet Biopharm 23: 1-4.

8. Sheiner LB, Ludden TM (1992) Population Pharmacokinetics/Pharmacodynamics. Ann Rev Pharmacol Toxicol 32: 185-209.

9. Whiting B, Kelman AW, Grevel J (1986) Population pharmacokinetics. Theory and clinical applications. Clin Pharmacokinet 11: 387-401.

10. Fan J, Gijbels I (1995) Nonlinear Models for Repeated Measurement Data. Monographs on Statistics and Applied Probability. Chapman & Hall.

11. Bernardo JM, Smith AFM (1994) Bayesian Theory. (2ndedn), New York: Wiley.

12. Metropolis N, Arianna WR, Marshall NR, Augusta HT, Edward T (1953) Equation of state calculations by fast computing machine. J Chem Phys 21: 1087-1092.

13. Tierney L (1994) Markov Chains for exploring posterior distributions. Ann Stat 22: 1701-1762.

14. Gelman A, Roberts GO, Gilks WR (1999) Efficient metropolis jumping rules. in Bayesian Statistics 5. Oxford: Oxford University Press.

15. Gelfand AE, Smith AFM (1992) Bayesian statistics without tears: a sampling-resampling perspective. The Am Stat 46: 84-88.

16. Gilks WR, Richardson S, Spiegelhalter DJ (1995) Markov chain Monte Carlo in practice, New York: Chapman & Hall.

17. Hoeting JA, Raftery AE, Madigan D (2002) Bayesian variable and transformation selection in linear regression. J Comput Graph Statist 11: 485–507.

18. Berger JO, Pericchi L (2001) Objective Bayesian Methods for Model Selection: Introduction and Comparison. IMS Lecture Notes – Monograph Series 38: 135-193.

19. Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82: 711-732.

20. Chipman H, George E, McCulloch R (2001) The Practical Implementation of Bayesian Model Selection. Institute of Mathematical Statistics 65–134.

21. Sisson SA (2005) Transdimensional Markov chains: A decade of progress and future perspectives. J Am Stat Assoc 100: 1077-1089.

22. Wakefield J, Racine-Poon A (1995) An application of Bayesian population pharmacokinetic/pharmacodynamic models to dose recommendation. Stat Med 14: 971-986.

23. Gelman A, Bois F, Jiang J (1996) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. JASA 91: 1400-1412.

24. D'Argenio DZ, Park K (1997) Uncertain pharmacokinetic/pharmacodynamics systems; design estimation and control. Control Eng Practice 5: 1707-1716.

25. Wakefield J, Bennett J (1996) The Bayesian modeling of covariates for population pharmacokinetic models. JASA 91: 917-927.

26. Spiegelhalter DJ, Andrew T, Nicky B, Wally RG (1997) BUGS: Bayesian inference Using Gibbs Sampling. Cambridge, UK: MRC Biostatistic Unit.

27. Lunn DJ (2008) Automated covariate selection and Bayesian model averaging in population PK/PD models. J Pharmacokinet Pharmacodyn 35: 85-100.

28. Miller AJ (2001) Subset selection in regression. (2ndedn), Chapman & Hall.

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

30. Ribbing J, Jonsson EN (2004) Power, Selection Bias and Predictive Performance of the Population Pharmacokinetic Covariate Model. J Pharmacokinet Pharmacodyn 31: 109-226.

31. Breiman L, Friedman J, Charles JS, Olshen RA (1984) Classification and Regression Trees, Belmont, CA: Wadsworth.

32. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19: 716-723.

33. Wählby U, Jonsson EN, Karlsson MO (2002) Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. AAPS Pharm Sci 4: E27.

34. Mandema JW, Verotta D, Sheiner LB (1995) Building population pharmacokinetics-pharmacodynamic models, ed. D.Z. D'Argenio: Plenum, New York.

35. Semmar N, Bruguerolle B, Boullu-Ciocca S, Simon N (2005) Cluster Analysis: An Alternative Method for Covariate Selection in Population Pharmacokinetic Modeling. J Pharmacokinet Pharmacodyn 32: 333-358.

36. Han C, Carlin BP (2001) Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review. J Am Stat Assoc 96: 1122-1132.

37. Verotta D (2007-2012) Modeling Complex Pharmacokinetics/Pharmacodynamics., NIH, RO1 AI50587.

38. Brooks SP, Giudici P (1999) Convergence assessment for reversible jump MCMC simulations. Bayesian statistics 6, ed. J. Bernardo, et al. Oxford: Oxford University Press.

39. Lindbom L, Pihlgren P, Jonsson EN (2005) PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput Methods Programs Biomed 79: 241-257.

40. Denison DGT, Mallick BK, Smith AFM (1998) Bayesian MARS. Stat Comput 4: 337-346.

41. Denison DGT, Mallick BK, Smith AFM (1998) A Bayesian CART Algorithm. Biometrika 85: 363-377.