METHODS FOR POPULATION PHARMACOKINETICS AND PHARMACODYNAMICS

Emily Anne Colby

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics.

Chapel Hill
2012

Approved by:

Dr. Eric Bair
Gary Koch
Anastasia Ivanova
Michael Kosorok
Dan Weiner

Abstract

EMILY ANNE COLBY: Methods for Population Pharmacokinetics and Pharmacodynamics
(Under the direction of Dr. Eric Bair)

Current applications of cross validation have been unsuccessful at identifying covariate effects in the population Pharmacokinetic/Pharmacodynamic (PK/PD) setting when other methods find a covariate effect may exist. Software that does population PK/PD modeling has a nice feature of being able to do a post hoc step without any major iterations to obtain Bayesian parameter estimates and hence predictions for subjects that were not in the dataset that was used to fit the model. This work proposes cross validation methods for longitudinal mixed effects models that are effective at identifying covariate effects when they exist.

Acknowledgments

I wish to thank Dr. Eric Bair for his encouragement and assistance throughout this process and for sharing his extensive knowledge of cross validation with me. I would like to thank Dr. Dan Weiner for employing me at Pharsight and sharing his vast knowledge of PK/PD. I thank several of my co-workers including Jason Chittenden, Bob Leary, Mike Dunlavey, Jon Monteleone, and Serge Guzy for sharing their knowledge. I thank Pharsight for providing financial support in the last years of my education. I thank Dr. Gary Koch for providing financial support and employment in the Biometrics Consulting Lab for four years, showing me how to conduct statistical consulting, and connecting me with an internship at GlaxoSmithKline while generously providing full financial support so that the work requirement could be met. I would like to thank Dr. Sharon Murray, Nikki Arya, and others at GlaxoSmithKline for showing me how statistical work is conducted at a pharmaceutical company. I thank Pharsight for sending me to conferences such as the PAGE, where a lot of the inspiration for this dissertation began. I thank Pharsight for the skills I've gained and shared as a PK/PD modeling and simulation instructor, going to conferences, client locations, and the FDA– hence having the ability to interact with and learn from major thought leaders in the field. Thanks to all the traveling, I have had the great honor and privilege of meeting in person and talking with several of the authors listed in the bibliography and several others who made significant contributions to the field of PK/PD not directly related to this work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 Introduction

Cross validation has been used in various forms in the population pharmacokinetic (PK) setting. With all the variations, there are two common uses of cross validation currently being used for population Pharmacokinetic/Pharmacodynamic (PK/PD) modeling. Those are final model validation and model comparison.

For model comparison, cross validation has been unsuccessful at finding covariate effects when other methods seem to imply that covariate effects exist (Zomorodi et al., 1998), (Fiset et al., 1995). However, cross validation has been successful at identifying models with major structural differences (Valodia et al., 2000).

It will be shown in Chapter 2 that when covariate effects are present in an underlying population PK/PD model, a misspecification of failing to include a covariate effect may not hurt the overall predictive performance of the model in the outcome variable or concentration. Random effects in the pharmacokinetic parameters can make up for the lack of the covariate. Therefore, cross validation metrics that involve the predicted concentration errors will fail to identify a covariate effect. We instead propose using the post hoc estimates of the random effects as metrics for identifying covariate effects in population PK/PD models.

First, a review of the literature is presented. Then, the methods are proposed and evaluated using simulated data examples and real data examples.

1.2 Literature Review

This chapter reviews the literature pertaining to population pharmacokinetics and pharmacodynamics.

1.2.1 Background

Population pharmacokinetic and pharmacodynamic (PK/PD) modeling is the characterization of the distribution of probable PK/PD outcomes (parameters, concentrations, responses, etc.) in a population of interest. These models consist of fixed and random effects. The fixed effects describe the relationship between explanatory variables such as age, body weight, gender, and pharmacokinetic outcomes. The random effects quantify unexplained variation in PK/PD outcomes (FDA, 1999).

Population PK/PD modeling is useful for identifying influential covariates that may warrant some action, such as changes in labeling, dose adjustment, contraindication, and modification of design of future clinical trials. Quantification of unexplained variation in PK may be relevant to assessing safety risks and determining whether dose individualization is desirable or necessary. It can answer questions like "Is it all right to give everyone the same dose, regardless of body weight? If not, how should the doses be scaled?" (FDA, 1999).

In some cases, population PK modeling is a de facto requirement. The FDA may require that population PK modeling be performed for a new drug. Population analysis of patient data may be used in lieu of some types of PK trials, e.g., renal impairment or drug-drug interaction studies (FDA, 1999).

There are two approaches to population modeling: the two-stage approach and non-linear mixed effects (NLME) modeling (FDA, 1999). The two-stage approach consists of fitting PK models for each individual separately, then summarizing the PK parameters across individuals. Covariate relationships may be found by regressing the natural log of the PK parameters with covariates of interest. The NLME approach differs in that it fits one model across all individuals. This paper will focus on the NLME approach.

1.2.2 Structure of Population PK/PD Models

Population PK models are hierarchical (Davidian and Giltinan, 1995). There is a model for the individual, a model for the population, and a model for the residual error. The individual model consists of the curve of drug concentrations over time.

To explore the model for the individual, one must have a basic understanding of drug pharmacokinetics. There are four basic phases of drug pharmacokinetics: Absorption, Distribution, Metabolism, and Excretion (ADME). Typically a drug is given as an injection (intravenous), an infusion, or extravascular dose (oral, sublingual, inhalation, patch). Once the drug enters the body, it may undergo an absorptive phase prior to being taken into the plasma. If it is injected as a bolus, this phase does not occur. Once in plasma, the drug is distributed to various organs and tissues. It is often metabolised by an organ such as the liver or kidney, then excreted in urine, feces, or by exhalation.

Drug concentration data can be modeled with compartmental modeling, which involves curve fitting, or non-compartmental analysis (NCA). Non-compartmental analysis consists of calculation of pharmacokinetic parameters based on the data alone, with very few assumptions involved. Parameters such as Tmax, the time at which the maximum concentration, Cmax, occurs, and AUC, the area under the concentration-time curve are calculated. The AUC is calculated using a trapezoidal method, where the

concentration data points are connected with straight lines, and lines are drawn to the x-axis (time) to get trapezoidal areas for each time segment. The sum of the trapezoidal areas approximates the AUC. Simple linear regression is used to estimate the slope of the line in the elimination phase, referred to as lambdaZ, or rate of elimination (Gabrielsson and Weiner, 2000). There are many variations on this method, including ones that assume the decline in concentrations is log-linear (Gabrielsson and Weiner, 2000).

In population pharmacokinetic models, curve fitting is used to derive mean/population pharmacokinetic parameters of interest as well as to predict corresponding concentration values for individual patients. The curve of the plasma concentrations over time is sometimes modeled by a compartmental model, which assumes that the body is made up of "compartments" through which the drug passes prior to being excreted. The pharmacokinetic compartmental model is similar to a "black box" engineering model. Each of the compartments is a "black box," where a system of differential equations is derived based on the law of conservation of mass. By the law of conservation of mass, the change in the amount of drug versus time is equal to the sum of the contributing mass flow rates for each compartment (Sandler, 1999). See Figure 1.1 for an example of such a model. For example, the changes in drug amounts over time for a three compartment pharmacokinetic model with extravascular (non-intravenous) administration can be represented by the following differential equations. Note that there is an equation

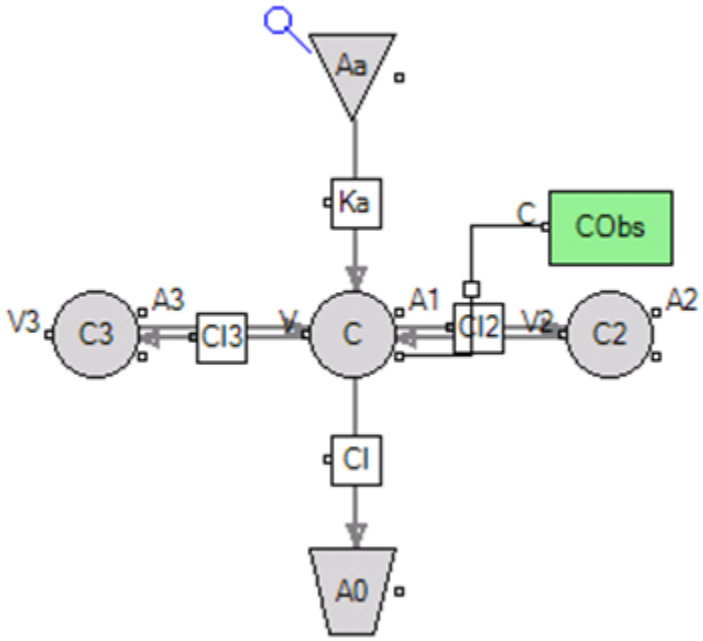

Figure 1.1: A three-compartment pharmacokinetic model

for each compartment.

$$\begin{aligned}
\frac{dA_a}{dt} &= -K_a \cdot A_a \\
\frac{dA_1}{dt} &= K_a \cdot A_a - Cl \cdot C - Cl_2 \cdot (C - C_2) - Cl_3 \cdot (C - C_3) \\
\frac{dA_2}{dt} &= Cl_2 \cdot (C - C_2) \\
\frac{dA_3}{dt} &= Cl_3 \cdot (C - C_3)
\end{aligned}$$

where A_a is the amount in the absorption compartment, and A_1, A_2, and A_3 are the amounts in the central and peripheral compartments, respectively. The rate of absorption of the drug into the central compartment is denoted K_a. The flows of the drug into and out of the peripheral compartments are denoted Cl_2 and Cl_3, and the flow of drug out of the body is denoted Cl. The corresponding volumes associated with

each compartments are V, V_2, and V_3, respectively. Then, we have the concentrations

$$C = \frac{A_1}{V}$$
$$C_2 = \frac{A_2}{V_2}$$
$$C_3 = \frac{A_3}{V_3}$$

The equation for C represents the model for the individual in the hierarchy of models. However, one needs to account for unexplained variability. (Note: The above example assumes the kinetics of drug transfer are first order. However, variations of such a model that employ non-linear kinetics can also be accommodated.)

The model for the residual error accounts for overall uncertainty in the concentrations over time. It captures all variability not captured by the specified fixed and random effects. The errors may weighted so that measurements with higher variability are given less weight compared with measurements with smaller variability. For example, under a constant CV percentage error model,

$$C_{Obs} = C \cdot (1 + C_\epsilon)$$

where C_{Obs} is the observed concentration, C is the predicted concentration, and C_ϵ is the residual error. C_ϵ is almost always assumed to follow a univariate normal distribution with mean 0 and variance σ^2. With a constant CV percent error model, higher concentration measurements (which tend to be more variable) are given less weight (Gabrielsson and Weiner, 2000). Other options for weighting include

1. Additive (Uniform): $C_{Obs} = C + C_\epsilon$

2. Log-Additive (equivalent to fitting a model to the log of the observations): $C_{Obs} = C \cdot exp(C_\epsilon)$, would reduce to an Additive error model if one were to take the log:

$ln(C_{Obs}) = ln(C) + C_\epsilon$

3. Power: $C_{Obs} = C + C^{power} \cdot C_\epsilon$. Special case: Power=0.5 is Poisson weighting: $C_{Obs} = C + C^{0.5} \cdot C_\epsilon$

4. Mixed is a combination of Proportional and Additive: $C_{Obs} = C + C_\epsilon + C \cdot C_\epsilon \cdot C_{MixRatio}$

5. Custom

For a non-population model, the parameters K_a, V, Cl, V_2, Cl_2, V_3, and Cl_3 are modeled with fixed effects only– that is, they are estimated separately for each individual. For a population model, one estimates the population mean values, and the amount each subject's values deviate from the population means in a simultaneous fit of all subject's data. In a population model, the PK parameters can be modeled with regression equations containing fixed effects, covariates, and random effects. The equations for the PK parameters represent the model for the population in the hierarchy of models. For example,

$$\begin{aligned}
K_a &= \theta_{K_a} \cdot \exp(\eta_{K_a} + \eta_{K_a,P_1} P_1 + \eta_{K_a,P_2} P_2) \\
V &= (\theta_V + dVdTrt \cdot Trt) \cdot \exp(\eta_V) \\
V_2 &= (\theta_{V_2} + dV_2dFed \cdot Fed) \cdot \exp(\eta_{V_2}) \\
V_3 &= (\theta_{V_3} + (W/\bar{W}t)^{dV_3dWt}) \cdot \exp(\eta_{V_3}) \\
Cl &= (\theta_{Cl} + dCldGene \cdot Gene) \cdot \exp(\eta_{Cl}) \\
Cl_2 &= \theta_{Cl_2} \cdot \exp(\eta_{Cl_2}) \\
Cl_3 &= \theta_{Cl_3} \cdot \exp(\eta_{Cl_3})
\end{aligned}$$

where θ_x denotes the fixed effect or typical value of a PK parameter x, and η_x denotes a random effect for a PK parameter x. The distribution of PK parameters is generally skewed to the right, and is often model with a log-normal distribution (this is why the

random effects are often exponentiated in the equations for the PK parameters). The vector of random effects is assumed to follow a multivariate normal distribution with mean 0 and variance-covariance matrix Ω. Ω may be diagonal, full block, or block diagonal.

Covariates such as Trt, an indicator that a specific drug was given, can be included. In the example above, Fed and $Gene$ are indicators that a subject was fed and that a certain gene is present, and Wt is a continuous variable for the body weight of a subject. ($\bar{Wt}$ represents the mean of Wt across all subjects.) The effects of the covariates on the PK parameters are given by $dVdTrt$, dV_2dFed, dV_3dWt, and $dCldGene$.

Occasion covariates such as P_1, an indicator for the first set of visits, and P_2, an indicator for the second set of visits, are typically included with random effects such as η_{K_a,P_1} and η_{K_a,P_2}. They are usually assumed to be independent, normally distributed with mean 0 and equal variance.

Hence, population pharmacokinetic models are non-linear mixed effects models. The differential equations may or may not have a closed-form solution, and are solved either analytically or numerically. The parameters are estimated using one of the various algorithms available such as first order conditional estimation with interaction (FOCEI) (Bonate, 2006).

1.2.3 Population PK/PD Modeling Procedure

Based on the FDA guidance for Population Pharmacokinetics (1999), population PK modeling can be carried out in three, interwoven steps: Exploratory Analysis, Model Development, and Model Validation (FDA, 1999).

Exploratory Analysis

The exploratory analysis consists of plotting and summarizing the data in a tabular format. Individual modeling of concentration data, via compartmental modeling or non-compartmental analysis, may be performed to obtain initial estimates for the population model. Linear regression of the natural log of the PK parameters to the covariates of interest may be done as part of the exploratory analysis. Linear regression may also be used to determine the structure of the PK model (for example, if clearance changes with dose, one may consider a Michaelis-Menten model for clearance) (Gabrielsson and Weiner, 2000).

A Michaelis-Menten model for clearance may have the form Cl = (Vmax/(Km + C)) where Vmax (maximum metabolic rate) and Km (Michaelis Menten constant) are parameters, and C is the predicted concentration. Drugs such as Ethanol exhibit Michaelis-Menten pharmacokinetics. From inspection of the equation for clearance, one can see that the Michaelis-Menten clearance decreases as the concentration increases. This can happen when the metabolizing enzymes become saturated, making the process of metabolism slower with an increase in drug concentration (Gabrielsson and Weiner, 2000).

Population Model Development

Model development consists of spelling out objectives, hypotheses, and assumptions, followed by model building (FDA, 1999). The proposed model building procedure will depend on the objectives, hypotheses, and assumptions. For example, if whether or not a subject is fed is expected to have an effect on the PK, one will plan to include a covariate for fed/fasted state prior to doing the model building and plan to test the hypothesis that fed/fasted state has no effect on the PK during the model building

process.

Population Model Building

Model building consists of three steps: Base/Structural Model, Covariate Model, and Covariance Model (FDA, 1999).

Base/Structural Model

The structure of the PK model is determined largely during the exploratory analysis, where the concentrations are plotted on the log scale versus time. The number of compartments may be determined by observing the number of distinct phases visible in the plot (Gabrielsson and Weiner, 2000). For example, Figure 1.2 is a plot of the drug concentration versus time for a drug that exhibits two compartment pharmacokinetics. In Figure 1.2, the concentration increases from zero as the drug is absorbed, until it reaches the maximum concentration, C_{max}. After reaching C_{max}, the drug concentration in plasma decreases sharply at first if the distribution is rapid, then decreases again at a different rate. The two distinct phases after C_{max} are modeled with two compartments: a central compartment and a peripheral compartment. However, the steeper decline may represent elimination if distribution is slower than elimination.

Imagine having several curves like Figure 1.2 in a single plot, varying slightly from one another, representing drug concentrations for several individuals (see Figure 2.5 for an example). In that case, the base/structural population PK model would be a two compartment model with random effects for the PK parameters K_a, V, Cl, V_2, and Cl_2. A non-linear mixed effects model is generally used to predict average/population PK values, and then the etas (random effects) are used to estimate how much each

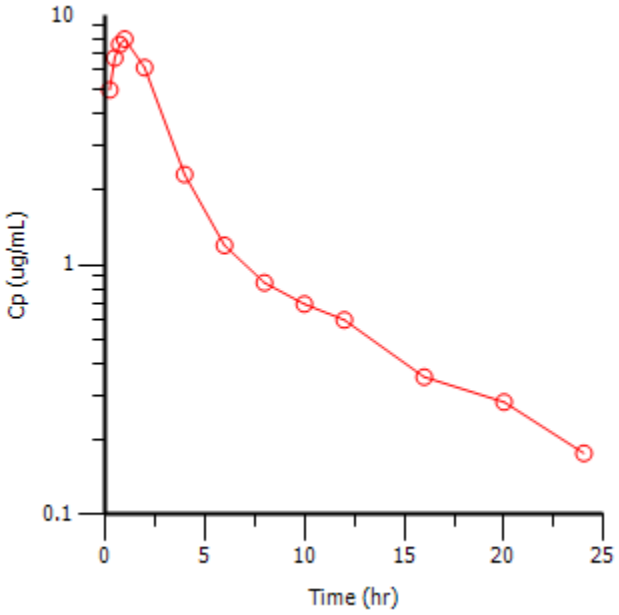

Figure 1.2: Drug concentration versus time for a two-compartment model

individual deviates from the population PK value.

To obtain initial estimates for the fixed effects PK parameters in the base/structural model, one may use traditional methods such as curve stripping (Gibaldi and Perrier, 1975). This method is built in to the WinNonlin Classic models and is performed automatically when the defaults are accepted. One may also use non-compartmental analysis (Gibaldi and Perrier, 1975) for obtaining initial estimates. Also, the Naive Pooled method in Phoenix NLME (Pharsight) can be used to get rough estimates for the fixed effects (essentially FOCEI with random effects parameters frozen to 0), especially when the data are relatively sparse.

Once fixed effects initial estimates are found, one builds up to obtaining rough estimates for the variances and covariances of the random effects. A method called First Order (FO) can be used to accomplish this (Sheiner, Rosenberg and Melmon, 1972), as well as iterative two stage Expectation-Maximization (IT2S-EM) in Phoenix NLME (Ette and Williams, 2004).

Estimation Methods

The covariate selection, covariance structure selection, and final model are fitted with an FOCEI method, QRPEM, or a Laplacian method. Two kinds of FOCEI, First Order Conditional Estimation- Extended Least Squares (FOCE-ELS) and First Order Conditional Estimation Lindstrom-Bates (FOCE L-B), are implemented in Phoenix NLME. Of the FOCEI methods, the Lindstrom Bates method tends to run faster than the ELS method. The Laplacian method is considered to be the most numerically correct of all the methods, and can be used for things like Poisson regression or other regression models where the likelihood function is specified by the user, though it generally takes longer to run. See (Ette and Williams, 2004) for a description of these methods. QRPEM was recently implemented in Phoenix NLME (in March of 2012) and has been excellent at fitting the more complex models for this work (Leary and Dunlavey, 2012).

Covariate Model

With the base/structural model determined, one can plot random effects versus covariates to determine whether covariates should be added. Then, one proceeds to the covariate model building stage. Covariate model building may be carried out using likelihood ratio tests (LRTs) when candidate models are nested. There may be an inflated type I error rate associated with the LRTs (Bertrand et al., 2009). Sometimes, the stepwise procedure is used. One can also look at the relative standard error percentages of the covariate effect estimates. If the relative standard error percentage is large, there may not be enough data to support the covariate in the model. If the relative standard error percentage is small (below 30), one may consider keeping the covariate in the model. Once covariate modeling is complete, the inter-subject covariance model

is determined.

Covariance Model

The inter-subject covariance model is defined by the structure of the Ω matrix. Random effects with low shrinkages (less than around 0.3) are kept in the model (Karlsson and Savic, 2007) (Savic and Karlsson, 2007), otherwise, random effects may be removed. For the random effects that are kept, the Ω matrix may be full block (all random effects correlated), block diagonal (some random effects correlated, some independent of the others), or diagonal (no random effects correlated). A scatter plot of the etas versus the etas for the final covariate model may be used to help determine the structure of Ω.

To obtain a robust estimate of the Ω matrix, one may use a non-parametric method. One is built in to Phoenix NLME. It starts with a parametric solution, then performs non-parametric iteration(s). If the estimate of the Ω matrix found using the non-parametric method is vastly different from that of the parametric method, it may indicate that one or more of the random effects has a bimodal distribution or some other deviation from a normal distribution. This may mean that a covariate was left out that should have been included (Davidian and Giltinan, 1995).

Model Validation

Once a final model is determined, model validation is done. Model validation may be performed using bootstrapping or predictive check. Consider a dataset containing concentration-time data for n subjects. Suppose one were to take a simple random sample with replacement of size n, fit a model, and obtain estimates. Then, perform

this x times, and summarize the model estimates across the x samples. This procedure is known as bootstrapping (Efron and Tibshirani, 1986) for population PK. Histograms of the bootstrapped model estimates may also be generated. It is known that the estimates of the standard errors of the model estimates can be biased. To account for this, boot-strapping is often employed to get better estimates of the variability. In many cases, it is the only way to obtain estimates of the variability of the model estimates, because the standard errors cannot always be calculated via matrix decomposition (Yafune and Ishiguro, 1999).

Predictive check is used to generate a population of subjects based on the fitted model, and then visually determine if this distribution provides good coverage of the underlying data on which the model was based (Karlsson and Holford, 2008). Beginning with a final model, the final model estimates are assumed to be correct. Then, based on the model assumptions that C_ϵ is normally distributed with mean 0 and variance σ^2, and η follows a multivariate normal distribution with mean 0 and variance Ω, new concentration-time observations are simulated for x replicates of n subjects. If the measurements are not taken at the same times for all subjects, similar times are binned (using an algorithm such as the k-means clustering algorithm). For each time bin, quantiles of the observed and simulated concentrations are calculated. A visual plot of the observed data with bands for the observed and simulated quantiles is used to determine whether the model fits the data.

1.2.4 Challenges in Population PK/PD Modeling

There are various challenges associated with population PK modeling. The structural model generally cannot be identified based on sparse data alone. In this case, one may need meta-data which includes some rich data. Convergence is often difficult,

resulting in having to set some parameters to a constant (or zero) or fitting a less complex model. Covariate model building can be time consuming and lead to inflated type I error rates (Wahlby, Jonsson and Karlsson, 2001). Estimates of precision of parameters are often biased and require bootstrapping or other techniques. Depending on model complexity, richness of data and other issues, it may take hours (or days) to achieve convergence. Message Passing Interface (MPI) may be used to take advantage of multiple processors on a single machine in Phoenix NLME. Also, a grid or cluster of multiple computers may be used for parallel processing.

1.2.5 Smoothing Splines

A discussion of smoothing splines is given here because in Chapter 4 we will compare population PK models to smoothing splines for prediction of concentrations. Suppose we are given a set of response variables $\{y_i\}_{i=1}^n$ and predictor variables $\{x_i\}_{i=1}^n$ and we wish to estimate each y_i based on $f(x_i)$, where f is a function that minimizes

$$\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int \{f''(t)\}^2 \, dt \tag{1.1}$$

Any such f must be an element of the Sobolev space of functions with second derivatives that are square integrable. The tuning parameter λ controls the tradeoff between goodness of fit and smoothness. When $\lambda = \infty$, no second derivative is allowed for f, meaning that f must be linear and (4.2) reduces to the ordinary least squares criteria. When $\lambda = 0$, then any f that interpolates the data will minimize (4.2).

It can be shown that (4.2) is minimized when f is a natural cubic spline with knots at each x_i (Hastie, Tibshirani and Friedman, 2008). Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the order statistics of the x_i's. Then a natural cubic spline $f(x)$ with knots $x_1, x_2, \ldots, x_n$ satisfies the following properties:

1. $f(x)$ is a a piecewise cubic polynomial. In particular, $f(x)$ is a cubic polynomial on $[x_{(1)}, x_{(2)}], [x_{(2)}, x_{(3)}], \ldots, [x_{(n-1)}, x_{(n)}]$.

2. $f(x)$ and its first two derivatives are continuous on $[x_{(1)}, x_{(n)}]$.

3. $f^{(j)}(x_{(1)}) = f^{(j)}(x_{(n)}) = 0$ for $j = 2, 3$. In other words, the second and third derivatives of f are zero at the boundary knots, which implies that f is linear outside the boundary knots.

See (Welham, 2009) or (Dierckx, 1995). For a complete description of smoothing splines and methods for fitting spline models (including the choice of the tuning parameter λ), see (Hastie, Tibshirani and Friedman, 2008).

1.2.6 Cross-Validation

A discussion of cross-validation is given here because in Chapters 2 and 3 we propose new cross-validation methods for population PK/PD covariate model building. Cross-validation is a method for evaluating the expected accuracy of a predictive model. Suppose we have a response variable Y and a predictor variable X and we seek to estimate Y based on X. Using the observed X's and Y's we may estimate a function $\hat{f}$ such that our estimated value of Y (which we call $\hat{Y}$) is equal to $\hat{f}(X)$. Cross-validation is an estimate of the expected loss function for estimating Y based on $\hat{f}(X)$. If we use squared error loss (as is conventional in population PK modeling), then cross-validation is an estimate of $E\left[\left(Y - \hat{f}(X)\right)^2\right]$.

A brief explanation of cross-validation is as follows: First, the data is divided into K partitions of roughly equal size. For the kth partition, a model is fit to predict Y based on X using the $K-1$ other partitions of the data. (Note that the kth partition is not used to fit the model.) Then the model is used to predict Y based on X for the data in the kth partition. This process is repeated for $k = 1, 2 \ldots, K$, and the K

estimates of prediction error are combined. Formally, let $\hat{f}^{-k}$ be the estimated value of f when the kth partition is removed, and suppose the indices of the observations in the kth partition are contained in K_k. Then the cross-validation estimate of the expected prediction error is equal to

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j\in K_i}\left(y_j - \hat{f}^{-i}(x_j)\right)^2$$

Here n denotes the number of observations in the data set. For a more detailed discussion of cross-validation, see (Hastie, Tibshirani and Friedman, 2008).

1.2.7 Current Uses of Cross-Validation in Population PK/PD

As mentioned earlier, covariate model building may be carried out using likelihood ratio tests (LRTs) when candidate models are nested. However, there may be an inflated type I error rate associated with the LRTs (Bertrand et al., 2009). For model comparison, cross validation has been unsuccessful at finding covariate effects when other methods seem to imply that covariate effects exist (Zomorodi et al., 1998), (Fiset et al., 1995). However, cross validation has been successful at identifying models with major structural differences (Valodia et al., 2000).

Cross validation is not often done with population PK modeling (Brendel et al., 2007). In one case (Bailey, Mora and Shafer, 1996), data was pooled across subjects to fit a model as though the data were obtained from a single subject. Subjects were removed, one at a time, and the accuracy of the predicted observations with subsets of the data was assessed. The method we propose is different because it does not pool the data across subjects prior to modeling, and we use it to compare candidate models rather than to assess accuracy of prediction. Another paper (Hooker et al., 2008) describes removing a subject at a time to estimate model parameters, then predicting

PK parameters using the covariate values for the subject that was removed and comparing those with the PK parameters obtained using the full data set, to evaluate the final model and identify influential individuals. The method we propose is different in that it uses a post-hoc step to calculate random effect values for the subject that is removed, and instead of evaluating a final model or identifying influential individuals we use cross validation to compare candidate models.

One article, (Ralph et al., 2006) calculates a prediction error for each subject in the model parameters, and a paired t-test is done on the prediction error between a base and full model to assess whether difference in imprecision of clearance between models is significant. The prediction error is calculated as the difference in the individual and population estimate divided by the individual estimate, times 100 percent, where the individual estimate is obtained using cross validation. The full model is only found to be correct with high levels of the covariate. This is fairly similar to the method we propose, except the statistic is different and a t-test is not employed, thus making it easier to find a covariate effect if there is one.

In (Zomorodi et al., 1998), cross validation is performed and weighted residuals for subjects left out are used to compare a base and full model. Predictions obtained for subjects left out may or may not have been based on the post hoc parameter estimates (article not clear). The base model is found to be better with the cross validation approach, but in other parts of paper the covariate is found to be significant. Actual model development was performed using a likelihood based approach. Later in this work, we explain why covariate effects go unidentified when the cross validation prediction error in the y's is used for comparing models in the population PK/PD setting.

In (Mulla et al., 2003), (Kerbusch et al., 2001), and (Rajagopalan and Gastonguay, 2003), actual model development was performed using a likelihood based approach.

Predictive performance of the final model is assessed using cross validation. In (Kerbusch et al., 2001), "if model predictions based on partial dataset were in accordance with predictions of full dataset, predictive ability of model was confirmed" (here authors cited (Efron and Tibshirani, 1993)).

Covariate models are compared using cross validation in (Fiset et al., 1995). Cross validation error in concentrations ((Obs - Pred)/Pred)*100 was used for comparing covariate models. All models in the comparison had similar cross validation results. Actual model development was performed using likelihood based approaches.

A poster presented in 2001 at PAGE by Ribbing (Ribbing and Jonsson, 2001) proposes a method for cross validation, referred to as cross model validation (CMV). With this method, cross validation is used with the objective function value (similar to log likelihood function) to select a covariate model. A similar method is proposed in (Katsube et al., 2011).

It may be that researchers were finding that cross validation as it is typically done for population PK/PD modeling is not helpful for detecting covariates. In (Wahlby, Jonsson and Karlsson, 2001), for the cross validation, one concentration data point for each parameter, the point at which the parameter is most sensitive, was chosen based on partial derivatives. In the cross-validation, the models showed similar predictive ability with respect to both measures of the concentration prediction errors defined in the article. It seems that even using the cross validation prediction error in the y's at the points that are most sensitive to the PK parameter with the covariate of interest does not help to elucidate a covariate relationship when one appears to exist.

It will be shown in Chapter 2 that when covariate effects are present in an underlying population PK/PD model, a misspecification of failing to include a covariate effect may not hurt the overall predictive performance of the model in the outcome variable or concentration. Random effects in the pharmacokinetic parameters can make up for

the lack of the covariate. Therefore, cross validation metrics that involve the predicted concentration errors will fail to identify a covariate effect. We instead propose using the post hoc estimates of the random effects as metrics for identifying covariate effects in population PK/PD models.

1.2.8 Automated Covariate Selection in Population PK/PD

A commonly used method for automated covariate selection in population PK/PD modeling is forward addition then backward elimination. It is often referred to as "stepwise", though it's different from the stepwise procedure used in traditional linear regression in that it does forward once, then backward once (Jonsson and Karlsson, 1998). Another method is GAM (Mandema, Verotta and Sheiner, 1992). A comparison of these methods can be found in (Wahlby, Jonsson and Karlsson, 2002). Maitre first proposed looking at the plots of the random effects versus the covariates to aid covariate model selection (Maitre et al., 1991). It was found that tree based modeling with cross validation to determine the tree size can help identify possible covariate models (Jonsson and Karlsson, 1999), but it does not seem that the cross validation method described involved re-fitting of the population model. This paper further explores the use of cross validation for automated covariate selection, with cross validation in the post hoc etas obtained from re-fitting population models.

1.3 Proposed Research

The following chapters will show why current uses of cross validation have failed to detect covariate relationships when they seem to exist and propose new methods for covariate model selection using cross validation. Finally, a comparison of mixed effect spline models to population pharmacokinetic models will be made.

Chapter 2

Cross validation for Longitudinal Mixed Effects Models

2.1 Overview

Current applications of cross validation have been unsuccessful at identifying covariate effects in the population PK/PD setting when other methods find a covariate effect may exist, due to the fact that the cross validation error used for covariate model comparison was that of the y's instead of the etas. Cross validation error in the y's is useful for identifying structural models but not for identifying covariate models in the population PK/PD setting. Software that does population PK/PD modeling has a nice feature of being able to do a post hoc step without any major iterations to obtain Bayesian parameter estimates and hence predictions for subjects that were not in the dataset that was used to fit the model. This article propose a cross validation method for longitudinal mixed effects models that is effective at identifying covariate effects when they exist, and two other methods for identifying a structural model.

2.2 Introduction

Cross validation has been used in various forms in the population pharmacokinetic setting. With all the variations, there are two common uses of cross validation currently being used for population PK/PD modeling. Those are final model validation and model comparison.

For model comparison, cross validation has been unsuccessful at finding covariate effects when other methods seem to imply that covariate effects exist (Zomorodi et al., 1998), (Fiset et al., 1995). However, cross validation has been successful at identifying models with major structural differences (Valodia et al., 2000). In these instances, cross validation error in the y's was used for model comparison. There are other methods available to compare population pharmacokinetic/pharmacodynamic (PK/PD) models, such as the likelihood ratio test (LRT), however there may be an inflated Type I error rate associated with these methods in the population PK/PD setting (Bertrand et al., 2009).

When covariate effects are present in an underlying population PK/PD model, a misspecification of failing to include a covariate effect may not hurt the overall predictive performance of the model in the outcome variable y or concentration. Random effects in the pharmacokinetic parameters can make up for the lack of the covariate. Therefore, cross validation metrics that involve the predicted outcome or concentration errors (y's) will often fail to identify a covariate effect.

This work proposes using the cross validation post hoc estimates of the random effects (etas) as metrics for identifying covariate models in the population PK/PD setting, and the Bayesian prediction errors in the y's for identifying structural models. First, some background information on population PK/PD will be provided.

Population pharmacokinetic and pharmacodynamic (PK/PD) modeling is the characterization of the distribution of probable PK/PD outcomes (parameters, concentrations, responses, etc.) in a population of interest. These models consist of fixed and random effects. The fixed effects describe the relationship between explanatory variables such as age, body weight, gender, and pharmacokinetic outcomes. The random effects quantify unexplained variation in PK/PD outcomes.

Population PK models are hierarchical. There is a model for the individual, a model for the population, and a model for the residual error. The individual model consists of the curve of drug concentrations over time, a compartmental model. The pharmacokinetic compartmental model is similar to a black box engineering model. Each of the compartments is like a black box, where a system of differential equations is derived based on the law of conservation of mass (Sandler, 1999).

The equations for the PK parameters represent the model for the population in the hierarchy of models. The PK parameters are modeled with regression equations containing fixed effects, covariates, and random effects (etas). The vector of random effects (eta) is assumed to follow a multivariate normal distribution with mean 0 and variance-covariance matrix Ω. Ω may be diagonal, full block, or block diagonal.

The model for the residual error accounts for overall uncertainty in the concentrations over time. The errors may weighted so that measurements with higher variability are given less weight compared with measurements with smaller variability.

Hence, population pharmacokinetic models are non-linear mixed effects models. The differential equations may or may not have a closed-form solution, and are solved either analytically or numerically. The parameters are estimated using one of the various algorithms available such as first order conditional estimation with interaction (FOCEI). See (Wang, 2007) for a mathematical description of these algorithms.

Once model parameters are estimated using an algorithm such as FOCEI, one may

fix the values of the model estimates and perform a post-hoc calculation to obtain random effect values (etas) for each subject. Thus, one may fit a model to a subset of the data and obtain random effect values for the full data set.

2.2.1 Cross-Validation

Cross-validation is a method for evaluating the expected accuracy of a predictive model. Suppose we have a response variable Y and a predictor variable X and we seek to estimate Y based on X. Using the observed X's and Y's we may estimate a function $\hat{f}$ such that our estimated value of Y (which we call $\hat{Y}$) is equal to $\hat{f}(X)$. Cross-validation is an estimate of the expected loss function for estimating Y based on $\hat{f}(X)$. If we use squared error loss (as is conventional in population PK modeling), then cross-validation is an estimate of $E\left[\left(Y - \hat{f}(X)\right)^2\right]$.

A brief explanation of cross-validation is as follows: First, the data is divided into K partitions of roughly equal size. For the kth partition, a model is fit to predict Y based on X using the $K-1$ other partitions of the data. (Note that the kth partition is not used to fit the model.) Then the model is used to predict Y based on X for the data in the kth partition. This process is repeated for $k = 1, 2 \ldots, K$, and the K estimates of prediction error are combined. Formally, let $\hat{f}^{-k}$ be the estimated value of f when the kth partition is removed, and suppose the indices of the observations in the kth partition are contained in K_k. Then the cross-validation estimate of the expected prediction error is equal to

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j\in K_i}\left(y_j - \hat{f}^{-i}(x_j)\right)^2$$

Here n denotes the number of observations in the data set. For a more detailed discussion of cross-validation, see (Hastie, Tibshirani and Friedman, 2008).

Cross validation is not often done with population PK modeling (Brendel et al., 2007). In one case (Bailey, Mora and Shafer, 1996), data was pooled across subjects to fit a model as though the data were obtained from a single subject. Subjects were removed, one at a time, and the accuracy of the predicted observations with subsets of the data was assessed. The method we propose is different because it does not pool the data across subjects prior to modeling, and we use it to compare candidate models rather than to assess accuracy of prediction. Another paper (Hooker et al., 2008) describes removing a subject at a time to estimate model parameters, then predicting PK parameters using the covariate values for the subject that was removed and comparing those with the PK parameters obtained using the full data set, to evaluate the final model and identify influential individuals. The method we propose is different in that it uses a post-hoc step to calculate random effect values for the subject that is removed, and instead of evaluating a final model or identifying influential individuals we use cross validation to compare candidate models.

2.3 Methods

2.3.1 Comparing models with major structural differences

In this case, a researcher may want to compare models with different numbers of compartments, such as a one-compartment model with a two-compartment model. This method is designed to detect differences in models that affect the overall shape of the curve.

Consider a dataset with subjects i, $i = 1, ..., n$. Each subject has observations y_{ij} for $j = 1, ..., t_i$ (t_i being the number of time points or discrete values of the independent variable for which there are observations for subject i). The statistic can be calculated as follows.

For $i = 1$ to n:

1. Remove subject i from the dataset
2. Fit a mixed effects model to the subset of the data
3. Accept all parameter estimates from the last run, and freeze the parameters to those values
4. Fit the same model to the whole dataset, without any major iterations, estimating only the post hoc values of the random effects (Phoenix NLME: NITER=0. NONMEM: MAXITER=0, POSTHOC=Y)
5. Calculate predicted values for subject i (the subject that was left out)
6. Take the average of the squared individual residuals for the subject that was left out (over all time points or over all values of the independent variable t_i)

Take the average of the quantity in step 6 over all subjects.

This sequence of steps can also be represented by the equation

$$mPRESS = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{ti}(y_{ij} - \hat{y}_{ij,-i})^2}{ti} \tag{2.1}$$

where y_{ij} is the observed value for the ith subject at the jth time point or independent variable value. $\hat{y}_{ij,-i}$ is the predicted value for the ith subject at the jth time point or independent variable value in a model where subject i is left out and post hocs are obtained. The number of time points or independent variable values for which there are observations for subject i is represented by t_i, n is the number of subjects.

For purposes of exploration, another statistic that takes into account the weighting can be calculated

$$wtmPRESS = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{ti} WTIRES_{ij,-i}^2}{ti}, WTIRES_{ij,-i} = \frac{\sqrt{wt_{ij,-i}}(y_{ij} - \hat{y}_{ij,-i})}{\hat{\sigma}_{-i}} \quad (2.2)$$

where $WTIRES_{ij,-i}$ is the individual weighted residual for subject i at time or independent variable value j in a model where subject i is left out and post hocs are obtained, and $wt_{ij,-i}$ is the weight defined by the residual error model (equal to the squared reciprocal of $\hat{y}_{ij,-i}$ for constant CV error models or 1 for additive error models), and $\hat{\sigma}^2_{-i}$ is the estimated residual variance.

When comparing models, the following steps should be applied. If the model with less parameters has a value of the statistic less than or equal to that of the model with more parameters, the model with less parameters should be chosen. For cases where the statistic for the model with more parameters is smaller than that of the model with less parameters, and furthermore, if the statistic for the model with less parameters is within one standard error of the statistic of the model with more parameters, the model with the smaller number of parameters should be chosen. Otherwise, if the model with more parameters has a value of the statistic that is more than one standard error below that of the model with less parameters, the model with more parameters should be chosen. The standard error employed should be that of the model with the smallest value of the statistic.

Alternatively, one may follow the same procedure, removing more than one subject at a time. For example, remove 10 percent of subjects at a time, fit a model, obtain predictions for the subjects left out including the post hoc values of the parameters. Square the individual residuals, average those over the independent variable for each

subject, average over subjects.

This method is similar to, or possibly the same as, cross validation methods already established, though it's not clear whether current methods include calculating the post hoc parameter values to obtain predictions for the subjects that are left out.

2.3.2 Comparing covariate models

In this case, a researcher may want to compare models with and without covariate effects, such as a model with an age effect on clearance versus a model without an age effect on clearance. This method is designed to detect differences in models that affect the equations for the parameters.

Consider a dataset with subjects i, $i = 1, ..., n$. Each subject has observations y_{ij} for $j = 1, ..., t_i$ (t_i being the number of time points or discrete values of the independent variable for which there are observations for subject i). The question of interest is whether or not a fixed effect dPdV for a covariate V should be included in an equation for a parameter P, having fixed effect tvP and random effect η_P. The equation for P could have any of the typical forms used in population PK/PD modeling, for example,

$$P = tvP \cdot (V/mean(V))^{dPdV} \cdot exp(\eta_P) \tag{2.3}$$

and one wishes to compare it with a model having no covariate effect

$$P = tvP \cdot exp(\eta_P) \tag{2.4}$$

If a covariate, V, has an effect on a parameter, P, the unexplained error in P, modeled by η_P, when V is left out of the model tends to have higher variance. By including covariate V in the model, we wish to reduce the unexplained error in P, which is represented by η_P. Therefore, metrics involving η_P are useful for determining whether a covariate V

is needed. While the distribution of η_P under the null and alternative hypotheses is unknown, cross validation can be performed. We propose a statistic for determining whether a covariate, V, is needed for explaining variability in a parameter, P, when P is modeled with a random effect "eta", η_P.

The statistic can be calculated as follows.

For $i = 1$ to n:

1. Remove subject i from the dataset
2. Fit a mixed effects model to the subset of the data
3. Accept all parameter estimates from the last run, and freeze the parameters to those values
4. Fit the same model to the whole dataset, without any major iterations, estimating only the post hoc values of the random effects (Phoenix NLME: NITER=0. NONMEM: MAXITER=0, POSTHOC=Y)
5. Square the post hoc eta estimate for the subject that was left out for the parameter of interest

Take the average of the quantity in step 5 over all subjects.

This sequence of steps can also be represented by the equation

$$nPRESS = \frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_{P_{i,-i}})^2 \tag{2.5}$$

Where $\hat{\eta}_{P_{i,-i}}$ is the post hoc "eta" estimate for the ith subject for parameter P in a model where the ith subject was removed, and n is the number of subjects.

When comparing models, the following steps should be applied. If the model with less parameters has a value of the statistic less than or equal to that of the model with more parameters, the model with less parameters should be chosen. For cases where the statistic for the model with more parameters is smaller than that of the model with less parameters, and furthermore, if the statistic for the model with less parameters is within one standard error of the statistic of the model with more parameters, the model with the smaller number of parameters should be chosen. Otherwise, if the model with more parameters has a value of the statistic that is more than one standard error below that of the model with less parameters, the model with more parameters should be chosen. The standard error employed should be that of the model with the smallest value of the statistic.

Alternatively, one may follow the same procedure, removing more than one subject at a time. For example, remove 10 percent of subjects at a time, fit a model, calculate the post hoc values for the subjects left out, square the post hoc etas, average them over subjects.

2.4 Simulation Example 1

A one-compartment, extravascular model was simulated with eight subjects using Pharsight's Trial Simulator. The equations for the model are as follows.

$$\begin{aligned} \frac{dAa}{dt} &= -Ka \cdot Aa \\ \frac{dA1}{dt} &= Ka \cdot Aa - Cl \cdot C \\ C &= \frac{A1}{V} \end{aligned}$$

A 10 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.01

A fixed effect was added to the absorption rate parameter, Ka. All other parameters were simulated with fixed and random effects.

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot exp(nCl) \end{aligned}$$

The fixed effects for the PK parameters were assumed to be normally distributed at the study level (varying across replicates) with means listed below and standard deviations of 0.1.

$$\begin{aligned} mean(tvKa) &= 0.35 \\ mean(tvV) &= 13.5 \\ mean(tvCl) &= 7.4 \end{aligned}$$

The random effects (nV and nCl) were simulated to be independent and normally distributed at the subject level (varying across subjects) with means of 0 and variances of 0.01. A covariate, GENDER, was simulated, so that there were 50 percent males and 50 percent females. A covariate, BODYWEIGHT, was simulated with a mean of 70 kg for males, 65 kg for females and a standard deviation of 15 for both groups. A covariate, Age, was simulated, with a mean of 40 years and a standard deviation of 10. None of the covariates were simulated to have any effect on the parameters. The true underlying model had no covariate effects. A dose of 5617 was administered at time 0, as an extravascular dose. Two hundred replicates were simulated. See Figure 2.1 for a plot of the simulated data.

The base model was a one compartment extravascular model with random effects for V and Cl and no age effect on clearance. The full model was a one compartment

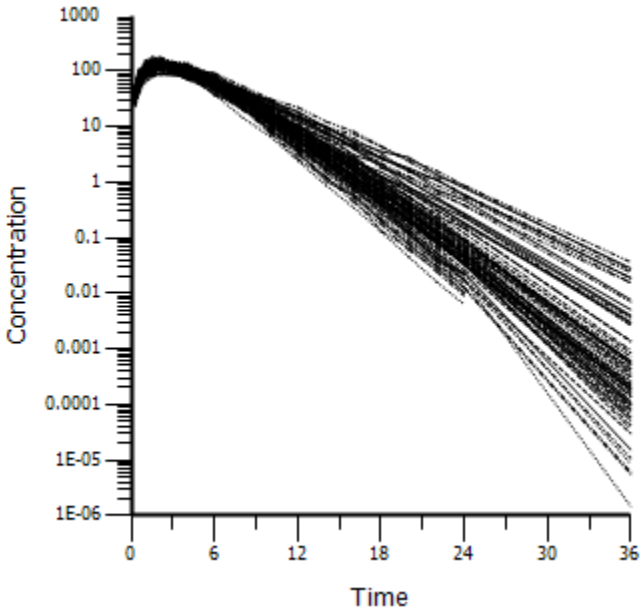

Figure 2.1: Simulation Example 1 data

extravascular model with random effects for V and Cl and an age effect on clearance.

Base (correct) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot exp(nCl) \end{aligned}$$

Full (incorrect) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot (Age/40)^{dCldAge} \cdot exp(nCl) \end{aligned}$$

Where tvKa, tvV, tvCl, and dCldAge are fixed effects parameters to be estimated. Initial estimates for the fixed effects PK parameters (tvKa, tvV, and tvCl) were set to the true (simulated) parameter values. The initial estimate for the covariate effect (dCldAge) was set to -3. The initial estimates of the variances of the random effects

were all 0.1, close to the true value of 0.01.

2.5 Simulation Example 2

A one-compartment, extravascular model was simulated with eight subjects using Pharsight's Trial Simulator. The equations for the model are as follows.

$$\begin{aligned}\frac{dAa}{dt} &= -Ka \cdot Aa \\ \frac{dA1}{dt} &= Ka \cdot Aa - Cl \cdot C \\ C &= \frac{A1}{V}\end{aligned}$$

A 10 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.01

A fixed effect was added to the absorption rate parameter, Ka. All other parameters were simulated with fixed and random effects. The systemic clearance was simulated with an age effect.

$$\begin{aligned}Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot (Age/40)^{dCldAge} \cdot exp(nCl)\end{aligned}$$

The fixed effects (tvKa, tvV, tvCl, and dCldAge) were assumed to be normally distributed at the study level (varying across replicates) with means listed below and

standard deviations of 0.05, 0.1, 0.05, and 0.04, respectively.

$$\begin{aligned} mean(tvKa) &= 0.35 \\ mean(tvV) &= 13.5 \\ mean(tvCl) &= 1.2 \\ mean(dCldAge) &= -0.9 \end{aligned}$$

The random effects (nV and nCl) were simulated to be independent and normally distributed at the subject level (varying across subjects) with means of 0 and variances of 0.01. A covariate, GENDER, was simulated, so that there were 50 percent males and 50 percent females. A covariate, BODYWEIGHT, was simulated with a mean of 70 kg for males, 65 kg for females and a standard deviation of 15 for both groups. A covariate, Age, was simulated, with a mean of 40 years and a standard deviation of 10. The true underlying model had a covariate effect– an age effect on clearance. A dose of 5617 was administered at time 0, as an extravascular dose. Two hundred replicates were simulated. See Figure 2.2 for a plot of the simulated data, with clearance decreasing with age.

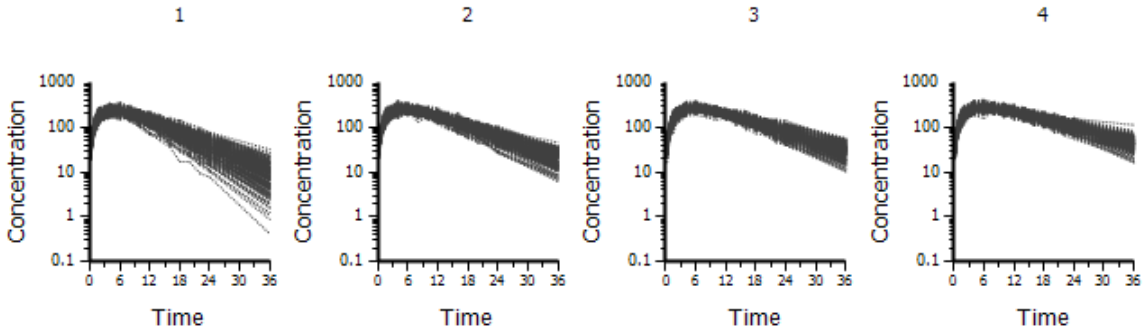

Figure 2.2: Simulation Example 2 data, by age quartiles

The base model was a one compartment extravascular model with random effects on V and Cl and no age effect on clearance. The full model was similar to the base model, but with an age effect included for Cl.

Base (incorrect) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot exp(nCl) \end{aligned}$$

Full (correct) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot (Age/40)^{dCldAge} \cdot exp(nCl) \end{aligned}$$

Where tvKa, tvV, tvCl, and dCldAge are fixed effects parameters to be estimated. Initial estimates for the fixed effects parameters (tvKa, tvV, tvCl, and dCldAge) were set to the true (simulated) parameter values. The initial estimates of the variances of the random effects were all 0.1, close to the true values of 0.01.

2.6 Simulation Example 3

A two-compartment, extravascular model was simulated with eight subjects using Pharsight's Trial Simulator. The equations for the model are as follows.

$$\frac{dAa}{dt} = -Ka \cdot Aa$$
$$\frac{dA1}{dt} = Ka \cdot Aa - Cl \cdot C - Cl2 \cdot (C - C2)$$
$$\frac{dA2}{dt} = Cl2 \cdot (C - C2)$$
$$C = \frac{A1}{V}$$
$$C2 = \frac{A2}{V2}$$

A 10 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.01

A fixed effect was added to the absorption rate parameter, Ka. All other parameters were simulated with fixed and random effects. The systemic clearance was simulated with an age effect.

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ V2 &= tvV2 \cdot exp(nV2) \\ Cl &= tvCl \cdot (Age/40)^{dCldAge} \cdot exp(nCl) \\ Cl2 &= tvCl2 \cdot exp(nCl2) \end{aligned}$$

The fixed effects (tvKa, tvV, tvV2, tvCl, tvCl2, and dCldAge) were assumed to be normally distributed at the study level (varying across replicates) with means listed

below and standard deviations of 0.05, 0.1, 0.1, 0.05, 0.05, and 0.04, respectively.

$$\begin{aligned} mean(tvKa) &= 0.35 \\ mean(tvV) &= 13.5 \\ mean(tvV2) &= 36 \\ mean(tvCl) &= 1.2 \\ mean(tvCl2) &= 0.62 \\ mean(dCldAge) &= -0.9 \end{aligned}$$

The random effects (nV, nV2, nCl, and nCl2) were simulated to be independent and normally distributed at the subject level (varying across subjects) with means of 0 and variances of 0.01. A covariate, GENDER, was simulated, so that there were 50 percent males and 50 percent females. A covariate, BODYWEIGHT, was simulated with a mean of 70 kg for males, 65 kg for females and a standard deviation of 15 for both groups. A covariate, Age, was simulated, with a mean of 40 years and a standard deviation of 10. The true underlying model had a covariate effect– an age effect on clearance. A dose of 5617 was administered at time 0, as an extravascular dose. Two hundred replicates were simulated. See Figure 2.3 for a plot of the simulated data, with clearance decreasing with age.

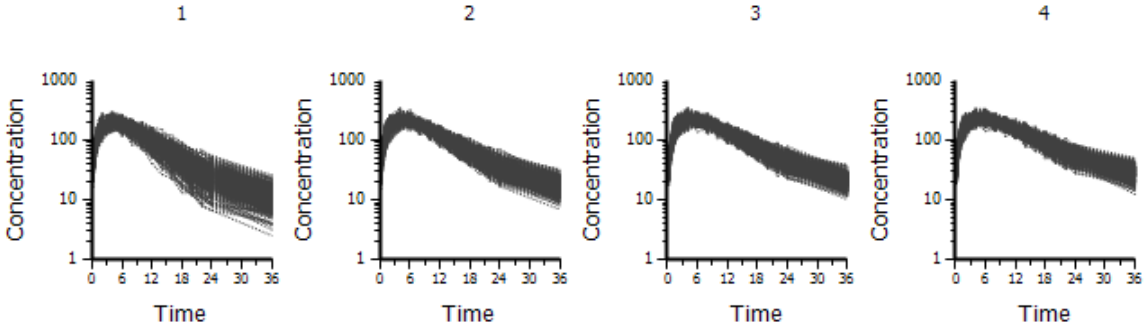

Figure 2.3: Simulation Example 3 data, by age quartiles

The base model was a two compartment extravascular model with random effects

on V, V2, Cl, and Cl2 and no age effect on clearance. The full model was similar to the base model, but with an age effect included for Cl.

Base (incorrect) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ V2 &= tvV2 \cdot exp(nV2) \\ Cl &= tvCl \cdot exp(nCl) \\ Cl2 &= tvCl2 \cdot exp(nCl2) \end{aligned}$$

Full (correct) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ V2 &= tvV2 \cdot exp(nV2) \\ Cl &= tvCl \cdot (Age/40)^{dCldAge} \cdot exp(nCl) \\ Cl2 &= tvCl2 \cdot exp(nCl2) \end{aligned}$$

Where tvKa, tvV, tvV2, tvCl, tvCl2, and dCldAge are fixed effects parameters to be estimated. Initial estimates for the fixed effects parameters (tvKa, tvV, tvV2, tvCl, tvCl2, and dCldAge) were set to the true (simulated) parameter values. The initial estimates of the variances of the random effects were all 0.1, close to the true values of 0.01.

2.7 Simulation Example 4

A one-compartment, extravascular model was simulated with eight subjects using Pharsight's Trial Simulator. The equations for the model are as follows.

$$\frac{dAa}{dt} = -Ka \cdot Aa$$
$$\frac{dA1}{dt} = Ka \cdot Aa - Cl \cdot C$$
$$C = \frac{A1}{V}$$

A 10 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.01

A fixed effect was added to the absorption rate parameter, Ka. All other parameters were simulated with fixed and random effects. The systemic volume was simulated with a body weight effect. The systemic clearance was simulated with body weight (BW), age (Age), gender (Gender), and hepatic impairment (HI) effects.

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot (BW/70)^{dVdBW} \cdot exp(nV) \\ Cl &= tvCl \cdot (BW/70)^{dCldBW} \cdot (Age/40)^{dCldAge} \cdot (1 + dCldG \cdot Gender) \\ &\quad \cdot (1 + dCldHI \cdot HI) \cdot exp(nCl) \end{aligned}$$

The fixed effects (tvKa, tvV, tvCl, dVdBW, dCldBW, dCldAge, dCldG, and dCldHI) were assumed to be normally distributed at the study level (varying across replicates) with means listed below and standard deviations of 0.05, 0.1, 0.05, 0.1, 0.1, 0.04, 0.05,

and 0.05 respectively.

$$\begin{aligned} mean(tvKa) &= 0.35 \\ mean(tvV) &= 13.5 \\ mean(tvCl) &= 1.2 \\ mean(dVdBW) &= 1 \\ mean(dCldBW) &= 0.75 \\ mean(dCldAge) &= -0.9 \\ mean(dCldG) &= 0.1 \\ mean(dCldHI) &= -0.2 \end{aligned}$$

The random effects (nV and nCl) were simulated to be independent and normally distributed at the subject level (varying across subjects) with means of 0 and variances of 0.01. A covariate, Gender, was simulated, so that there were 50 percent males (Gender=1) and 50 percent females (Gender=0). A covariate for body weight, BW, was simulated with a mean of 70 kg for males, 65 kg for females and a standard deviation of 15 for both groups. A covariate, Age, was simulated, with a mean of 40 years and a standard deviation of 10. A covariate for hepatic impairment, "HI", was simulated, with 70 percent not hepatically impaired (HI=0) and 30 percent hepatically impaired (HI=1). The true underlying model had five covariate effects– a body weight effect on volume, and age, body weight, gender, and hepatic impairment effects on clearance. A dose of 5617 was administered at time 0, as an extravascular dose. Two hundred replicates were simulated. See Figure 2.4 for a plot of the simulated data.

The base model was a one compartment extravascular model with random effects on V and Cl, a body weight effect on V, and age, gender, and body weight effects on clearance. The full model was similar to the base model, but with a hepatic impairment effect also included for Cl.

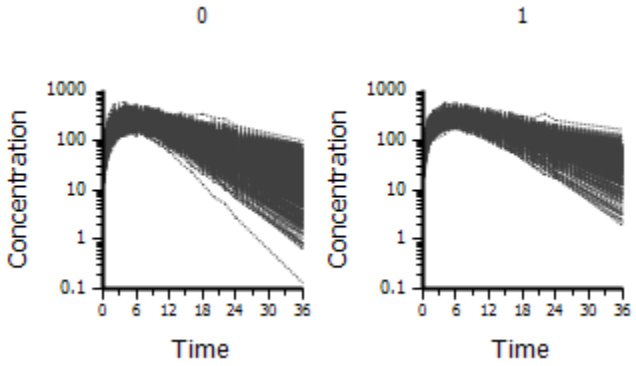

Figure 2.4: Simulation Example 4 data, by hepatic impairment

Base (incorrect) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot (BW/70)^{dVdBW} \cdot exp(nV) \\ Cl &= tvCl \cdot (BW/70)^{dCldBW} \cdot (Age/40)^{dCldAge} \cdot (1 + dCldG \cdot Gender) \cdot exp(nCl) \end{aligned}$$

Full (correct) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot (BW/70)^{dVdBW} \cdot exp(nV) \\ Cl &= tvCl \cdot (BW/70)^{dCldBW} \cdot (Age/40)^{dCldAge} \cdot (1 + dCldG \cdot Gender) \\ &\quad \cdot (1 + dCldHI \cdot HI) \cdot exp(nCl) \end{aligned}$$

Where tvKa, tvV, tvCl, dVdBW, dCldBW, dCldAge, dCldG, and dCldHI are fixed effects parameters to be estimated. Initial estimates for the fixed effects parameters were set to the true (simulated) parameter values. The initial estimates of the variances of the random effects were all 0.1, close to the true values of 0.01.

2.8 Simulation Example 5

A two-compartment, extravascular model was simulated with six subjects using Pharsight's Trial Simulator. The equations for the model are as follows.

$$\begin{aligned}
\frac{dAa}{dt} &= -Ka \cdot Aa \\
\frac{dA1}{dt} &= Ka \cdot Aa - Cl \cdot C - Cl2 \cdot (C - C2) \\
\frac{dA2}{dt} &= Cl2 \cdot (C - C2) \\
C &= \frac{A1}{V} \\
C2 &= \frac{A2}{V2}
\end{aligned}$$

A 10 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.01

A fixed effect was added to the absorption rate parameter, Ka. All other parameters were simulated with fixed and random effects.

$$\begin{aligned}
Ka &= tvKa \\
V &= tvV \cdot exp(nV) \\
V2 &= tvV2 \cdot exp(nV2) \\
Cl &= tvCl \cdot exp(nCl) \\
Cl2 &= tvCl2 \cdot exp(nCl2)
\end{aligned}$$

The fixed effects for the PK parameters were assumed to be normally distributed at the study level (varying across replicates) with means listed below and standard deviations of 0.1, except in this example the fixed effect for Ka was simulated with a

standard deviation of 0.05 because when the absorption rate was smaller the portion of the curve for the first compartment became less pronounced in relation to the portion for the second compartment. Having a smaller standard deviation for Ka increased the chance that all the simulated profiles would have a characteristic two compartment shape.

$$\begin{aligned} mean(tvKa) &= 0.35 \\ mean(tvV) &= 13.5 \\ mean(tvV2) &= 34 \\ mean(tvCl) &= 7.4 \\ mean(tvCl2) &= 1.2 \end{aligned}$$

The random effects (nV, nV2, nCl, and nCl2) were simulated to be normally distributed at the subject level (varying across subjects) with means of 0 and variances of 0.01. A covariate, GENDER, was simulated, so that there were 50 percent males and 50 percent females. A covariate, BODYWEIGHT, was simulated with a mean of 70 kg for males, 65 kg for females and a standard deviation of 15 for both groups. A covariate, Age, was simulated, with a mean of 40 years and a standard deviation of 10. The true underlying model had no covariate effects. A dose of 5617 was administered at time 0, as an extravascular dose. One hundred replicates were simulated. See Figure 2.5 for a plot of the simulated data.

Pharsights Phoenix NLME was used to fit models to the simulated data. A base model with one compartment was fit to the simulated data. A full model with two compartments was fit.

Base (incorrect) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ Cl &= tvCl \cdot exp(nCl) \end{aligned}$$

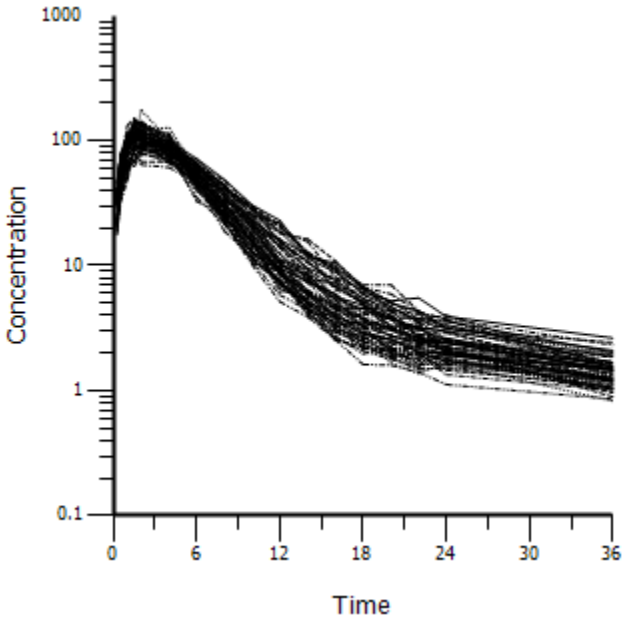

Figure 2.5: Simulation Example 5 data

Full (correct) model

$$\begin{aligned} Ka &= tvKa \\ V &= tvV \cdot exp(nV) \\ V2 &= tvV2 \cdot exp(nV2) \\ Cl &= tvCl \cdot exp(nCl) \\ Cl2 &= tvCl2 \cdot exp(nCl2) \end{aligned}$$

Where tvKa, tvV, tvV2, tvCl, and tvCl2 are fixed effects parameters to be estimated. Initial estimates for the fixed effects parameters (tvKa, tvV, tvV2, tvCl, and tvCl2) were set to the true (simulated) parameter values. The initial estimates of the variances of the random effects were all 0.1, close to the true values of 0.01.

2.9 Computational details

Pharsights Phoenix NLME was used to fit models to the simulated data from the command line. First, R was used to split replicates of the simulated datasets into

different files, then to split replicate datasets into training datasets in separate folders. Each training dataset consisted of the full dataset for the given replicate except with concentrations and amounts set to missing for one subject. In each folder with each training dataset resided the full dataset for the corresponding replicate. One batch file was used to call another batch file to execute NLME in all the folders until all training datasets and all replicates were processed. First the batch files were called to run NLME with the training datasets to obtain model parameter estimates using the Lindstrom-Bates method (Lindstrom and Bates, 1990), then called to run NLME again in the same folder with the full datasets without any major iterations, starting from the last solution (!iflagrestart=1 in nlmeflags.asc), to obtain post hoc estimates for the subjects that had been removed.

This entire process was completed for a base model and a full model for each simulation scenario.

2.10 Simulation Results

The results of the simulations are summarized in Tables 2.1, 2.2, 2.3, 2.4, and 2.5. If the true underlying model was the base model (as in the first scenario), a model comparison method was considered correct if it selected the base model. For AIC, BIC, and nPRESS, if the value for the base model was smaller than that of the full model, it was considered to be correct. For the mPRESS and wtmPRESS, the method was considered to be correct if the value of the statistic for the base model was less than the value of statistic for the full model. Otherwise, if the value of the statistic for the full model was smaller, it was still correct if the statistic for the base model was within one standard error of the statistic for the full model (employing the standard error of the statistic for the full model).

If the true underlying model was the full model, a model comparison method was

Table 2.1: Proportion correct out of 200 replicates

True Model	Comparison	AIC	BIC	wtmPRESS	mPRESS	nPRESS
1 Cpt	1 Cpt, Age-Cl	0.885	0.945	0.940	0.965	0.970
1 Cpt, Age-Cl	1 Cpt	0.985	0.930	0.000	0.000	0.925
2 Cpt, Age-Cl	2 Cpt	0.975	0.940	0.005	0.010	0.930
1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl, HI-Cl	1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl	0.715	0.640	0.015	0.005	0.970
2 Cpt	1 Cpt	1.00*	1.00*	1.00*	1.00*	N/A

Cpt=Compartment, Age-Cl indicates age effect on clearance, BW=Body Weight, V=Volume, G=Gender, HI=Hepatic Impairment
*Based on 100 replicates

considered correct if it selected the full model. For AIC, BIC, and nPRESS, if the value was greater for the base model than for the full model, it was considered correct. For the mPRESS and wtmPRESS, the method was considered incorrect if the value of the statistic for the base model was smaller than the value of the statistic for the full model. Otherwise, if the value of the statistic for the base model was greater than the value of the statistic for the full model plus one standard error, it was considered to be correct (employing the standard error of the statistic for the full model).

The modified nPRESS statistic was correct in 97.0 percent of the 200 cases where the true one compartment underlying model had no covariate effect, whereas AIC was correct in 88.5 percent of cases and BIC was correct in 94.5 percent of cases. It correctly identified the full model when the true underlying model was a one compartment model with an age effect on clearance in 92.5 percent of the 200 cases, where AIC found the correct model in 98.5 percent of cases and BIC found the correct model in 93 percent of cases. When the true underlying model was two compartment with an age effect on clearance, nPRESS was correct in 93.0 percent of cases, whereas AIC and BIC were correct in 97.5 and 94.0 percent of cases, respectively. When the true underlying

model was one compartment with a body weight effect on volume, and body weight, gender, age, and hepatic impairment effects on clearance, nPRESS was correct in 97.0 percent of cases, whereas AIC and BIC were correct in 71.5 and 64.0 percent of cases, respectively.

To emphasize the fact that the mPRESS and wtmPRESS statistics, which use the predicted values rather than the random effects for the parameters, should not be used for comparing different covariate models in the population PK/PD setting, mPRESS and wtmPRESS were calculated for all scenarios. They were wrong almost every time when the true model had a covariate effect. The predicted values are just as accurate with and without the covariate effect when the true model has a covariate effect, because the etas (e.g., nCl) will always compensate for a missing covariate in a parameter (e.g., Cl). This is why the nPRESS and not the mPRESS, nor the wtmPRESS, should be employed for situations when one wishes to compare different covariate models.

All four applicable methods (AIC, BIC, mPRESS, and wtmPRESS) correctly identified the two compartment model with random effects on V, V2, Cl, and Cl2 as the correct model when the base model was a one compartment model with random effects on V and Cl in 100 out of 100 cases. This finding is of interest because the standard likelihood ratio test cannot be applied when there are random effects in the full model that aren't present in the base model.

While AIC and BIC were not correct in all cases, the average AIC and BIC were smaller for the true underlying models when averaging across replicates.

For each replicate, the value of mPRESS, wtmPRESS, or nPRESS has a standard error associated with it. The standard error for each replicate is calculated as the sample standard error of mPRESS, wtmPRESS, or nPRESS divided by the square root of the number of subjects. The standard errors are summarized in Tables 2.3 and 2.4, along with the mean values of nPRESS, mPRESS, and wtmPRESS. The values

Table 2.2: Summary of AIC and BIC in simulation scenarios

True Model	Comparison		AIC		BIC	
			Base	Full	Base	Full
1 Cpt	1 Cpt,	N	200	200	200	200
	Age-Cl	Mean	384	404	401	424
		SD	114	115	114	115
1 Cpt,	1 Cpt	N	200	200	200	200
Age-Cl		Mean	1106	1093	1124	1113
		SD	25	25	25	25
2 Cpt,	2 Cpt	N	200	200	200	200
Age-Cl		Mean	1039	1026	1068	1058
		SD	27	27	29	29
1 Cpt, BW-V;	1 Cpt, BW-V;	N	200	200	200	200
BW-Cl, G-Cl,	BW-Cl, G-Cl,	Mean	1106	1100	1135	1132
Age-Cl, HI-Cl	Age-Cl	SD	33	32	33	32
2 Cpt	1 Cpt	N	100	100	100	100
		Mean	648	417	664	443
		SD	218	26	218	26

Cpt=Compartment, Age-Cl indicates age effect on clearance, BW=Body Weight, V=Volume, G=Gender, HI=Hepatic Impairment

Table 2.3: Summary of nPRESS in simulations

True Model	Comparison		nPRESS Base	nPRESS Full	SE Base	SE Full
1 Cpt	1 Cpt, Age-Cl	N	200	200	200	200
		Mean	0.136	0.919	0.022	0.403
		SD	0.236	0.920	0.032	0.512
1 Cpt, Age-Cl	1 Cpt	N	200	200	200	200
		Mean	0.078	0.012	0.031	0.005
		SD	0.056	0.010	0.022	0.005
2 Cpt, Age-Cl	2 Cpt	N	200	200	200	200
		Mean	0.144	0.024	0.077	0.020
		SD	0.728	0.155	0.453	0.155
1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl, HI-Cl	1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl	N	200	200	200	200
		Mean	1.700	0.124	0.626	0.081
		SD	2.577	0.281	2.515	0.214

Cpt=Compartment, Age-Cl indicates age effect on clearance, BW=Body Weight, V=Volume, G=Gender, HI=Hepatic Impairment

of nPRESS tended to be lower for the true underlying model in all scenarios. Because the eta will always compensate for a missing covariate, the mPRESS and wtmPRESS statistics should not be used for comparing covariate models, although they did perform well when the true underlying model had no covariate effect as well as for identifying the correct structural model.

2.11 Indomethacin Example

Pharsights Phoenix NLME was used to fit models to the published indomethacin dataset (Kwan et al., 1976). The indomethacin dataset, containing six subjects with eleven observations each, was fit using a two-compartment IV bolus model with Clearance parameterization and a proportional residual error model. Concentration units of ug/mL were assumed, and a dose of 25000 ug at 0 hours was assumed. Random effects were added to the PK parameters V, Cl, V2, and Cl2, in the form ThetaX*exp(nX),

Table 2.4: Summary of mPRESS in simulations

True Model	Comparison		mPRESS Base	mPRESS Full	SE Base	SE Full
1 Cpt	1 Cpt, Age-Cl	N	200	200	200	200
		Mean	41.1	46.4	13.3	18.5
		SD	58.4	82.0	51.6	73.6
1 Cpt, Age-Cl	1 Cpt	N	200	200	200	200
		Mean	247	254	45.6	47.2
		SD	54.0	59.0	24.4	25.5
2 Cpt, Age-Cl	2 Cpt	N	200	200	200	200
		Mean	159	199	28.0	59.9
		SD	34.9	171	12.5	160
1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl, HI-Cl	1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl	N	200	200	200	200
		Mean	323	472	85.1	173
		SD	116	401	72.2	284
2 Cpt	1 Cpt	N	100	100	100	100
		Mean	206.11	30.28	32.32	8.32
		SD	90.3	9.60	20.7	5.74

Cpt=Compartment, Age-Cl indicates age effect on clearance, BW=Body Weight, V=Volume, G=Gender, HI=Hepatic Impairment

Table 2.5: Summary of wtmPRESS in simulations

True Model	Comparison		wtmPRESS Base	wtmPRESS Full	SE Base	SE Full
1 Cpt	1 Cpt, Age-Cl	N	200	200	200	200
		Mean	1.14	1.20	0.157	0.176
		SD	0.303	0.388	0.064	0.154
1 Cpt, Age-Cl	1 Cpt	N	200	200	200	200
		Mean	0.967	1.01	0.144	0.154
		SD	0.133	0.130	0.120	0.111
2 Cpt, Age-Cl	2 Cpt	N	200	200	200	200
		Mean	1.0096	1550*	0.154	1550*
		SD	0.0995	21700*	0.071	21700*
1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl, HI-Cl	1 Cpt, BW-V; BW-Cl, G-Cl, Age-Cl	N	200	200	200	200
		Mean	1.25	2.78	0.305	1.56
		SD	0.485	11.6	0.432	11.3
2 Cpt	1 Cpt	N	100	100	100	100
		Mean	14.7	1.14	3.09	0.233
		SD	10.1	0.143	2.86	0.130

Cpt=Compartment, Age-Cl indicates age effect on clearance, BW=Body Weight, V=Volume, G=Gender, HI=Hepatic Impairment

*One of the replicates (161) had an inflated value for wtmPRESS. One subject was significantly younger than the others, and the effect of age on clearance was estimated to be around -22 instead of the true value of -0.9 in the model where this subject was left out. Hence the younger subject had inflated residuals.

Table 2.6: Theta from final model of Indomethacin dataset

Parameter	Estimate	Units	Stderr	Stderr%
tvV	8898	mL	574.84	6.46
tvV2	19527.3	mL	3169.70	16.23
tvCl	7905.99	mL/h	608.53	7.70
tvCl2	5252.15	mL/h	768.93	14.64
stdev0	0.1440		0.02	13.25

stdev0 = estimated residual standard deviation
prefix of 'tv' denotes fixed effect or typical value

where X is the parameter of interest.

Exploratory analysis (plot, Figure 4.2) showed two compartment PK. Individual initial estimates were obtained using the curve stripping method (Gibaldi and Perrier, 1975) with a WinNonlin Classic model. The averages of the individual PK parameters were used as initial estimates for the pop PK model.

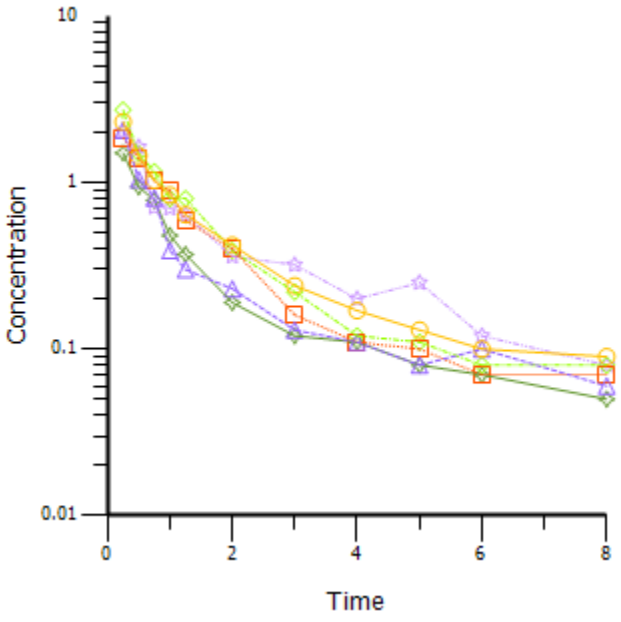

Figure 2.6: Concentration versus time from Indomethacin dataset

The model appeared to fit well based on diagnostic plots (Figures 4.3, 4.4, 4.5).

Table 2.7: Omega from final model of Indomethacin dataset

	nV	nCl	nV2	nCl2
nV	0.0017			
nCl	0	0.0338		
nV2	0	0	0.0666	
nCl2	0	0	0	0.1202
Shrinkage	0.7064	0.0329	0.3321	0.0727

The shrinkage was high for nV, but it was kept in. The relative standard errors for the fixed effect parameter estimates were all well below 30. The value of the original (Allen, 1974) PRESS statistic for the final model was 1.54, based on removal of 66 data points, one at a time. The average of Allen's PRESS over all data points was 0.02337.

The likelihood ratio test and the mPRESS statistic were used to compare a two compartment model with a one compartment model, without any random effects on the PK parameters (nV, nCl, nV2, nCl2 were removed so that comparison could be made with likelihood ratio test). The likelihood ratio test favored the two compartment model ($p < 0.0001$). The mPRESS (without post hocs) was in agreement with the likelihood ratio test, having a value of 0.1419 (SE 0.03393) for the one compartment model, and 0.0428 (SE 0.01355) for the two compartment model. Because $0.0428 + 0.01355 < 0.1419$, the mPRESS (without post hocs) favored the two compartment model.

The mPRESS with post hocs also favored the full (two compartment) model over the base (one compartment) model, when the random effects were added back in. The mPRESS (with post hocs) for the full model was 0.01679 (SE 0.004194) and 0.1406 (SE 0.03358) for the base model.

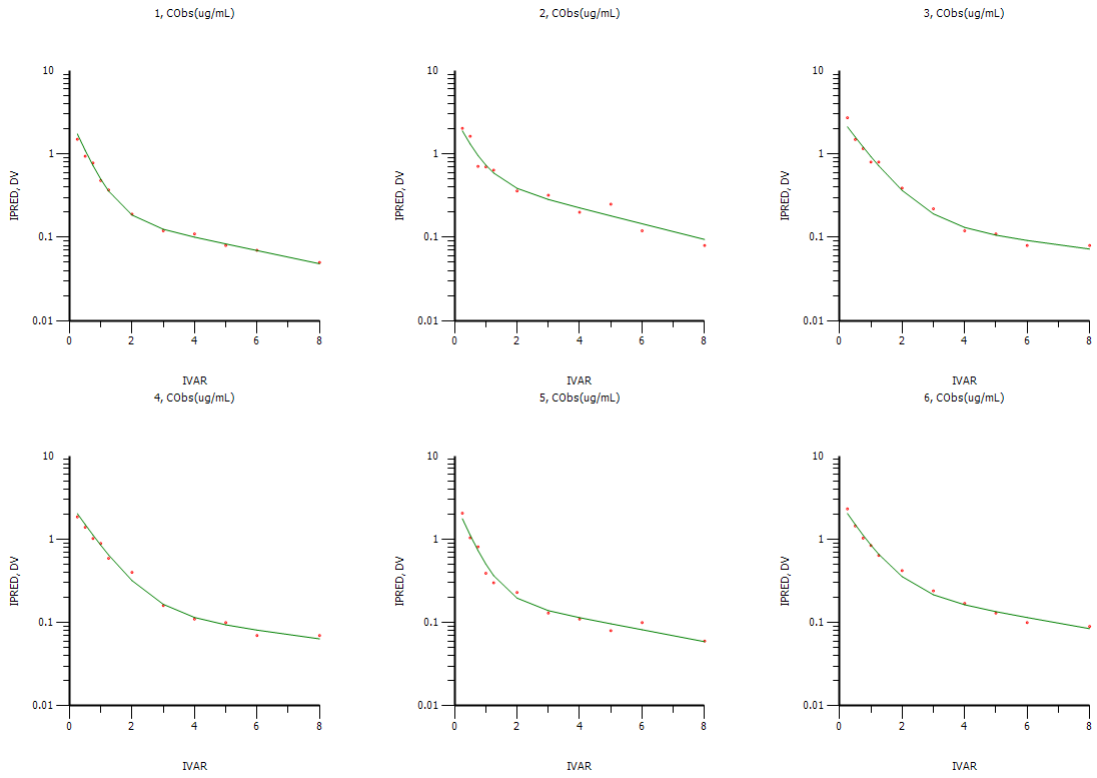

Figure 2.7: Final model of Indomethacin dataset

2.12 Theophylline Example

Pharsights Phoenix NLME was used to fit models to the published theophylline dataset (Boeckmann, Sheiner and Beal, 1992). The theophylline dataset, containing twelve subjects with eleven observations each, was fit using a one-compartment extravascular model with Clearance parameterization and an additive residual error model. Concentration units of mg/L were assigned, and doses of 267-320 mg/kg were given at 0 hours. A covariate for weight in kg was present in the dataset. Random effects were added to the PK parameters Ka, V, and Cl and in the form ThetaX*exp(nX), where X is the parameter of interest.

Exploratory analysis (plot, Figure 2.10) showed one compartment PK. Initial estimates for Ka, V, and Cl were 2, 1, and 1, respectively.

The likelihood ratio test and the mPRESS statistic were used to compare a model

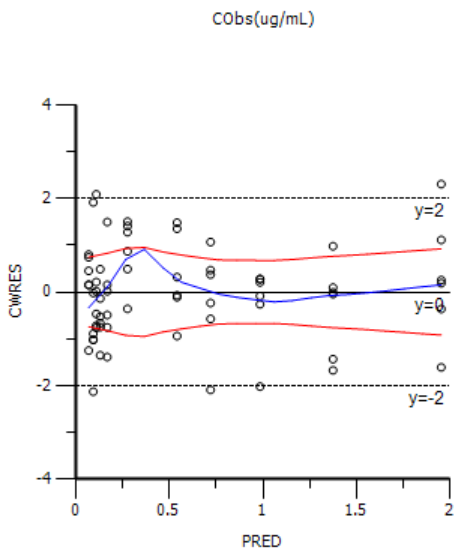

Figure 2.8: Residuals from final model of Indomethacin dataset

with a time lag parameter (tlag) to a model without a tlag parameter, without a random effect on the tlag parameter. The likelihood ratio test favored the tlag model ($p < 0.0001$). The mPRESS was in agreement with the likelihood ratio test, having a value of 0.2546 (SE 0.05727) for the model with tlag, and 0.3927 (SE 0.10001) for the model without tlag. Because $0.2567 + 0.05727 < 0.3927$, the mPRESS favored the tlag model.

The likelihood ratio test and the nPRESS statistic were used to compare a model with the tlag parameter and body weight effect on Ka to the model with the tlag parameter. The covariate plots for the model with tlag seemed to indicate a weight effect on Ka might be needed (Figure 2.11).

$$Ka = tvKa \cdot (wt/mean(wt))^{dKadwt} \cdot exp(nKa)$$

The likelihood ratio test had a borderline result ($p = 0.0667$). The nPRESS favored the full model with a tlag and a weight effect on Ka, having a value of 0.06220 (SE 0.02942) for the full (tlag and wt) model, and 0.7819 (SE 0.2846) for the base (tlag)

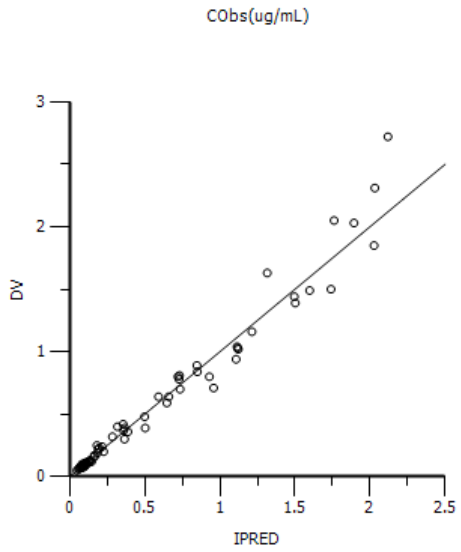

Figure 2.9: Observed versus predicted values for Indomethacin model

model. Because 0.06220 + 0.02942 <0.7819, the nPRESS favored the full model with tlag and wt.

2.13 Discussion

Cross validation can be used to identify structural models and covariate models for mixed effects longitudinal modeling, such as population PK/PD modeling. The nPRESS method outperformed AIC and BIC for finding the true underlying model when the true underling model had five covariates in a simulation study. It was specific enough not to find a covariate effect when one did not exist in the true underlying model. The mPRESS was in agreement with the likelihood ratio test in a real data example comparing models with different numbers of compartments. It correctly identified the two compartment model in a simulation study with a one compartment model for comparison.

In the Theophylline example, the population covariate plots (etas versus covariates) seemed to suggest a weight effect on Ka, while the likelihood ratio test gave a result of

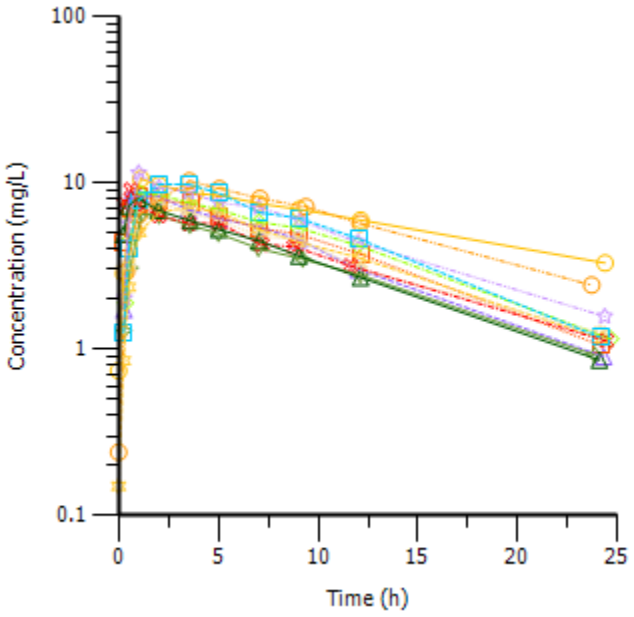

Figure 2.10: Concentration versus time from Theophylline dataset

borderline significance. The nPRESS statistic clearly favored the model with the weight effect on Ka. It seems that the likelihood ratio test cannot always identify covariate effects when they exist, therefore it is useful to have another method that can elucidate the underlying model.

These modified average PRESS statistics are easier to calculate (less computationally intensive) than the original PRESS statistic (Allen, 1974) when there are multiple observations per subject. The methods proposed in this paper are no more computationally intensive than bootstrapping, which is commonly done in population PK/PD modeling. Ten fold cross validation (removal of 10 percent of cases at a time) can be performed if it becomes cumbersome taking one subject out at a time, as long as the subset of ninety percent of the data can support the model of interest and is fairly well balanced in the observed covariate values.

This method might be applied with modifications in the linear mixed effects and generalized linear mixed effects modeling setting.

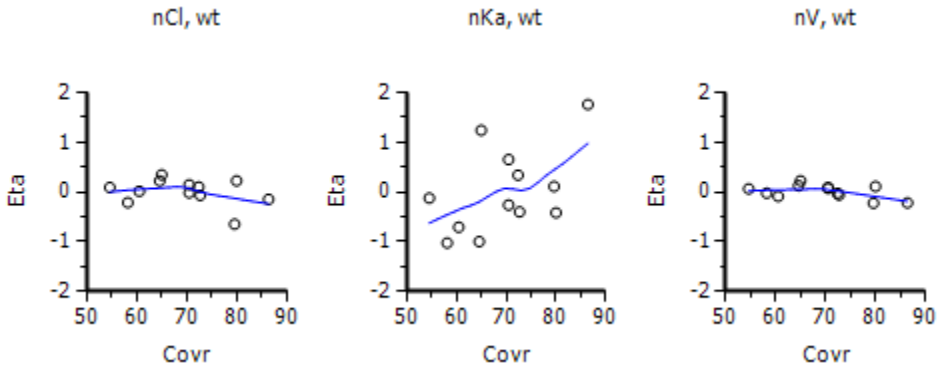

Figure 2.11: Eta versus covariate plots for Theophylline dataset

Note that as a fairly small number of replicates were generated for the simulation studies, this was not an attempt to estimate the true proportion of times the statistics will select the correct model in the scenarios that were considered. It was simply to show that the methods proposed in this paper are worth considering. More extensive simulation studies would need to be completed in order to estimate the true proportions.

Chapter 3

Automated Model Building Procedure

3.1 Overview

One problem when doing population modeling, is that there may be several covariates to consider and it is not feasible to fit all possible covariate models. It has been found that the average squared cross validation post hoc eta is a useful metric for identifying the correct covariate model. This article proposes a combination of stepwise and cross validation procedures with shrinkage and the average squared cross validation post hoc eta as a criteria for selecting covariates for inclusion in population PK/PD models.

3.2 Introduction

A commonly used method for automated covariate selection in population PK/PD modeling is forward addition then backward elimination. It is often referred to as "stepwise", though it's different from the stepwise procedure used in traditional linear regression in that it does forward once, then backward once (Jonsson and Karlsson, 1998). Another method is GAM (Mandema, Verotta and Sheiner, 1992). A comparison

of these methods can be found in (Wahlby, Jonsson and Karlsson, 2002). Maitre first proposed looking at the plots of the random effects versus the covariates to aid covariate model selection (Maitre et al., 1991). It was found that tree based modeling with cross validation to determine the tree size can help identify possible covariate models (Jonsson and Karlsson, 1999), but it does not seem that the cross validation method described involved re-fitting of the population model. This paper further explores the use of cross validation for automated covariate selection, with cross validation in the post hoc etas obtained from re-fitting population models.

When covariate effects are present in an underlying population PK/PD model, a misspecification of failing to include a covariate effect may not hurt the overall predictive performance of the model in the outcome variable or concentration. Random effects in the pharmacokinetic parameters can make up for the lack of the covariate. Therefore, cross validation metrics that involve the predicted outcome or concentration errors will often fail to identify a covariate effect (as shown in paper 2). AIC and BIC are also error prone when there are several covariates (as shown in paper 2). Likelihood based approaches can result in inflated Type I error rates (Bertrand et al., 2009).

This work proposes a variation of the traditional stepwise model building procedure, with shrinkage (Karlsson and Savic, 2007) (Savic and Karlsson, 2007) as a primary consideration, and the average of the squared post hoc estimates of the random effects obtained from cross validation as a secondary consideration for identifying complex covariate models in the population PK/PD setting. First, some background information on population PK/PD will be provided.

3.3 Methods

3.3.1 Comparing covariate models

This method was already introduced and evaluated in a separate paper by the same authors. It is reiterated here to show the calculation formula and decision process. In this case, a researcher may want to compare models with and without covariate effects, such as a model with an age effect on clearance versus a model without an age effect on clearance. This method is designed to detect differences in models that affect the equations for the parameters.

Consider a dataset with subjects i, $i = 1, ..., n$. Each subject has observations y_{ij} for $j = 1, ..., t_i$ (t_i being the number of time points or discrete values of the independent variable for which there are observations for subject i). The question of interest is whether or not a fixed effect dPdV for a covariate V should be included in an equation for a parameter P, having fixed effect tvP and random effect η_P. The equation for P could have any of the typical forms used in population PK/PD modeling, for example,

$$P = tvP \cdot (V/mean(V))^{dPdV} \cdot exp(\eta_P) \tag{3.1}$$

and one wishes to compare it with a model having no covariate effect

$$P = tvP \cdot exp(\eta_P) \tag{3.2}$$

If a covariate, V, has an effect on a parameter, P, the unexplained error in P, modeled by η_P, when V is left out of the model tends to have higher variance. By including covariate V in the model, we wish to reduce the unexplained error in P, which is represented by η_P. Therefore, metrics involving η_P are useful for determining whether a covariate V is needed. While the distribution of η_P under the null and alternative hypotheses is

unknown, cross validation can be performed. We propose a statistic for determining whether a covariate, V, is needed for explaining variability in a parameter, P, when P is modeled with a random effect "eta", η_P.

The statistic can be calculated as follows.

For $i = 1$ to n:

1. Remove subject i from the dataset
2. Fit a mixed effects model to the subset of the data
3. Accept all parameter estimates from the last run, and freeze the parameters to those values
4. Fit the same model to the whole dataset, without any major iterations, estimating only the post hoc values of the random effects (Phoenix NLME: NITER=0. NONMEM: MAXITER=0, POSTHOC=Y)
5. Square the post hoc eta estimate for the subject that was left out for the parameter of interest

Take the average of the quantity in step 5 over all subjects.

This sequence of steps can also be represented by the equation

$$nPRESS = \frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_{P_{i,-i}})^2 \tag{3.3}$$

Where $\hat{\eta}_{P_{i,-i}}$ is the post hoc "eta" estimate for the ith subject for parameter P in a model where the ith subject was removed, and n is the number of subjects.

When comparing models, the following steps should be applied. If the model with less parameters has a value of the statistic less than or equal to that of the model with more parameters, the model with less parameters should be chosen. For cases where the statistic for the model with more parameters is smaller than that of the model with less parameters, and furthermore, if the statistic for the model with less parameters is within one standard error of the statistic of the model with more parameters, the model with the smaller number of parameters should be chosen. Otherwise, if the model with more parameters has a value of the statistic that is more than one standard error below that of the model with less parameters, the model with more parameters should be chosen. The standard error employed should be that of the model with the smallest value of the statistic.

Alternatively, one may follow the same procedure, removing more than one subject at a time. For example, remove 10 percent of subjects at a time, fit a model, calculate the post hoc values for the subjects left out, square the post hoc etas, average them over subjects.

3.3.2 Automated Model Selection Procedure

When there are multiple covariates to consider, use this procedure.

1. Start with the base population model with random effects on all of the parameters and no covariates (check plots prior to this step to make sure overall model structure as well as the form of the residual error model are good). Use the eta versus covariate plots to determine the appropriate form for the different potential covariate effects.

2. If none of the random effects have shrinkages below 0.3 in the base model, stop. Otherwise, identify the parameter associated with the random effect with the lowest shrinkage. Calculate nPRESS for the base model.

3. For the parameter identified in the previous step, calculate nPRESS and the standard error of nPRESS for all univariate covariate models. Find the covariate that gives the biggest reduction in nPRESS over the base model. If it is smaller than that of the base model, and still smaller when one standard deviation (of the nPRESS of the full model) is added, include that covariate effect in the model for the given parameter, fit the model without cross validation (this is now the current working model), and go on to the next step. Otherwise, if none of the potential covariates meet the criterion, then determine whether any other random effects have shrinkages below 0.3 in the base model. If none of the remaining random effects have shrinkages below 0.3, then stop. Otherwise, identify the parameter with the next lowest shrinkage and repeat this step.

4. In the current working model, fitted without cross validation, if none of the random effects have shrinkages below 0.3, stop. Otherwise, identify the parameter that corresponds to the random effect with the lowest shrinkage in the current working model. If all the covariate models have already been considered for this parameter, identify the parameter that corresponds to the random effect with the next lowest shrinkage.

5. For the parameter identified in the previous step, calculate nPRESS for each of the remaining potential covariates modeled with the covariate(s) already added. Find the one that gives highest reduction in nPRESS over the nPRESS of the current working model. If value of nPRESS is still smaller when one standard deviation is added, add that covariate to the model for the parameter of interest (current working model now has an additional covariate effect). Otherwise, if none of the remaining potential covariates meet the criterion, then stop (current working model is unchanged).

6. Repeat steps 4-5 (the last two steps) until none of the remaining covariates meet criteria for entry (none of the random effects have shrinkages below 0.3 or it increases nPRESS or decreases nPRESS but the decrease is within one standard deviation to add more parameters).

3.4 Remifentanil Example

Pharsights Phoenix NLME was used to fit models to the published remifentanil dataset (Minto et al., 1997). The remifentanil dataset, containing sixty five subjects, was fitted using a three-compartment IV infusion model with Clearance parameterization and a proportional residual error model. Random effects were added to the PK parameters V, Cl, V2, Cl2, V3, and Cl3 in the form ThetaX*exp(nX), where X is the parameter of interest.

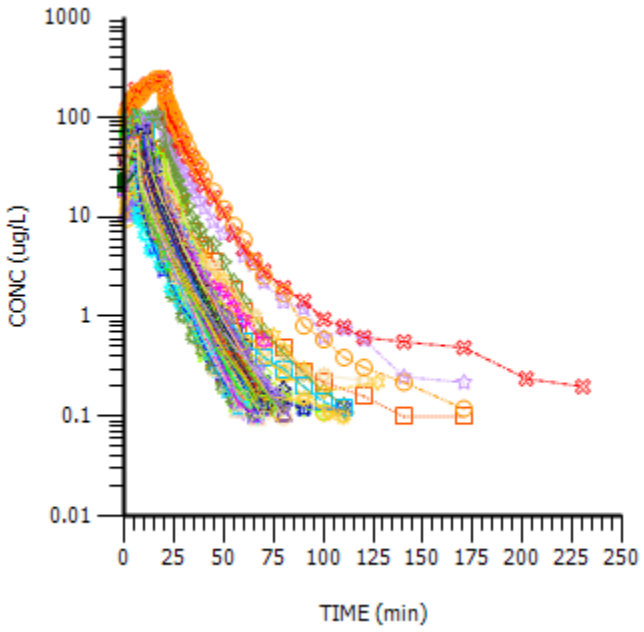

Figure 3.1: Concentration versus time from Remifentanil dataset

Exploratory analysis (plot, Figure 3.1) showed three compartment PK. This particular example was difficult because only a few subjects contributed data for the third compartment (in Figure 3.1, only a few subjects show full three compartment PK). Individual initial estimates were obtained using the curve stripping method (Gibaldi and Perrier, 1975) with a WinNonlin Classic model. The weighted averages of the individual PK parameters were used as initial estimates for the pop PK model, with the reciprocal of the standard deviation of the estimate as the weight. With these values, the Naive Pooled engine (which pools data as though it came from a single subject, but still taking into account different dosing, and all random effects except a residual error term are removed) was used to further refine the initial estimates for the fixed effects in the base population model.

Next, QRPEM (Leary and Dunlavey, 2012) was used to fit the population base model. The additive error model for the residual error appeared to show a trend in CWRES versus PRED (plot, Figure 3.2). When the residual error model was changed to multiplicative (constant CV) weighting, this trend was diminished (plot, Figure 3.4).

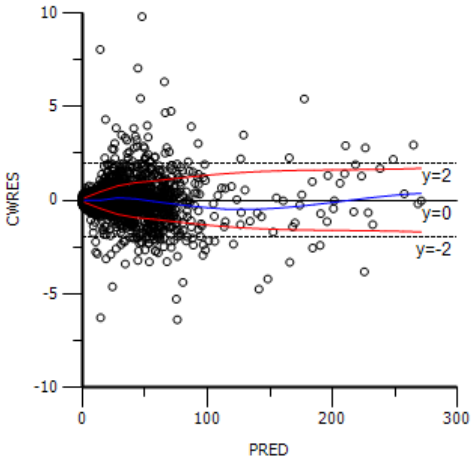

Figure 3.2: Residuals with Additive residual error model

Prior to covariate selection with the automated procedure described in this paper, the subject numbers were randomized because there were clusters of similar subjects

Table 3.1: Theta from final model of Remifentanil dataset

Parameter	Estimate	Units	Stderr	Stderr%
tvV	5.25		0.297	5.66
tvV2	9.68		0.534	5.51
tvV3	10.82	L	0.802	7.41
tvCl	2.39		0.048	2.00
tvCl2	1.55		0.089	5.72
tvCl3	0.08	L/min	0.002	2.35
dCldAGE	-0.29		0.037	-12.79
dCldBSA	0.68		0.130	19.13
dCl2dAGE	-0.47		0.118	-24.93
dV2dMALE0	-0.27		0.062	-22.41
dVdBSA	1.93		0.447	23.15
dV2dAGE	-0.44		0.088	-20.06
stdev0	0.14		0.001	0.87

stdev0 = estimated residual standard deviation
prefix of 'tv' denotes fixed effect or typical value

in the dataset. The method was applied with approximately ten-fold cross validation (five subjects removed at a time). Third compartment covariate effects were not going to be considered to avoid over-fitting and because only a few subjects contributed data for the third compartment. The steps taken are summarized in the appendix.

Once covariate model selection was finished, the covariance model selection was completed by first considering whether any random effects could be removed. The random effect for the volume of the third compartment had a high shrinkage value (above 0.3), and so it was removed. Next, scatter plots of etas versus etas were inspected for trends suggesting correlation between the random effects. A trend was found for nCl, nV2, and nCl2. The final Omega matrix was block-diagonal to reflect this relationship. Later, when predictive check was performed, the variance for nCl3 appeared to be inflated, so that random effect was removed as well.

The model appeared to fit well based on diagnostic plots (Figures 3.3, 3.4, 3.5),

Table 3.2: Omega from final model of Remifentanil dataset

	nV	nCl	nV2	nCl2
nV	0.100			
nCl	0	0.019		
nV2	0	0.029	0.095	
nCl2	0	0.035	0.093	0.161
Correlation				
nV	1			
nCl	0	1		
nV2	0	0.689	1	
nCl2	0	0.634	0.746	1
Shrinkage	0.128	0.028	0.051	0.086

except for in the third compartment (at the tails), where the concentrations began to dip below the limit of quantification. This might have been helped by keeping the random effects for Cl3 and V3, but with high shrinkage for nV3 and a poor estimate for the variance of nCl3 (as determined by a visual predictive check, not shown), the random effects for V3 and Cl3 were removed to keep the model as parsimonious as possible. The relative standard error percentages for the fixed effect parameter estimates were all below 30 percent.

For final model validation, a visual predictive check was performed (Figure 3.6). Assuming the final model parameters were correct, concentration values were simulated based on the model assumptions. Simulated quantiles were plotted against observed data and quantiles of observed data and found to be in fairly good congruence.

3.5 Simulation Example

A three-compartment, IV infusion model was simulated with sixty five subjects using Pharsight's Phoenix NLME. The equations for the model are as follows.

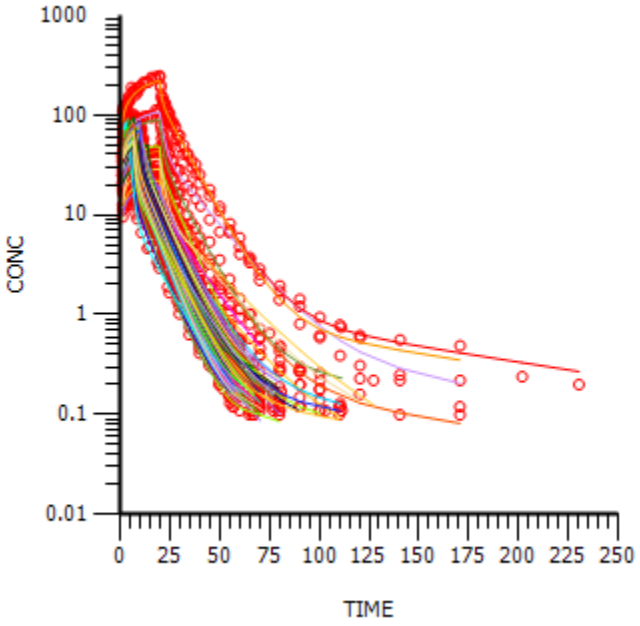

Figure 3.3: Final model of Remifentanil dataset

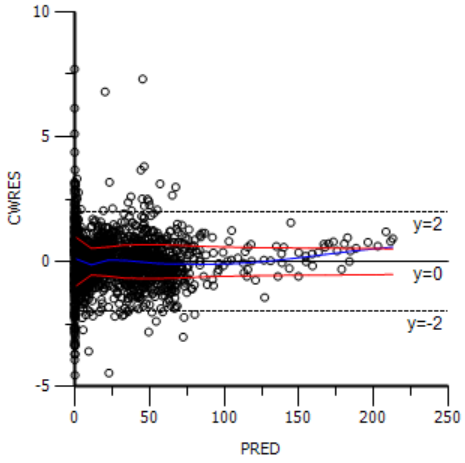

Figure 3.4: Residuals from final model of Remifentanil dataset

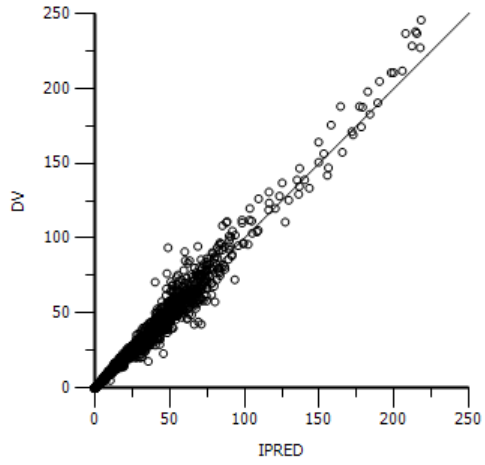

Figure 3.5: Observed versus predicted values from Remifentanil model

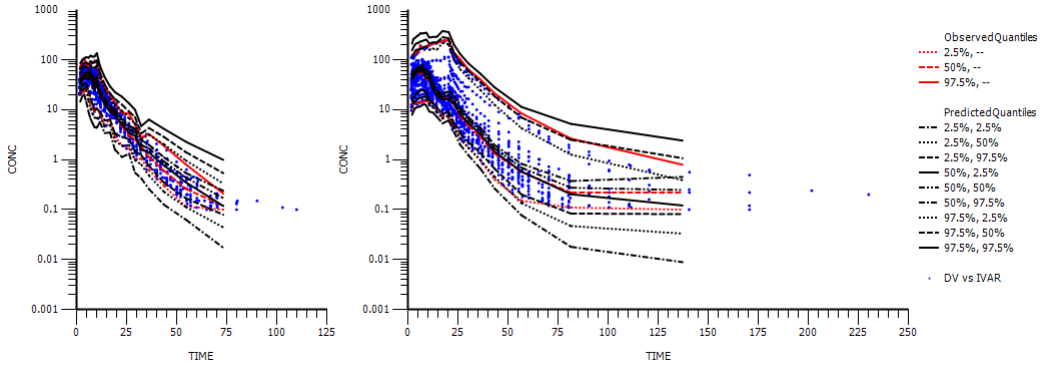

Figure 3.6: Predictive check from final Remifentanil model

$$\frac{dA1}{dt} = In(t) - Cl \cdot C - Cl2 \cdot (C - C2) - Cl3 \cdot (C - C3)$$
$$\frac{dA2}{dt} = Cl2 \cdot (C - C2)$$
$$\frac{dA3}{dt} = Cl3 \cdot (C - C3)$$
$$C = \frac{A1}{V}$$
$$C2 = \frac{A2}{V2}$$
$$C3 = \frac{A3}{V3}$$

where In(t) represents a constant rate infusion of the amounts and rates recorded in the Remifentanil dataset (Minto et al., 1997). A 14 percent constant CV percentage was simulated for the residual error.

CObs = C * (1 + CEps) where Var(CEps) = 0.0196

A covariate, "MALE", was simulated, so that there were approximately 50 percent males and 50 percent females (as in the original dataset). The covariates, body surface area, "BSA", and "AGE", were set to the values in the original Remifentanil dataset. All parameters were simulated with fixed effects, and random effects on V, Cl, V2, and Cl2. Several covariate effects were simulated: Age and BSA effects on Cl and V2, age effects on Cl2 and V, and a gender effect on V. This covariate model was slightly different than the one found in the analysis of the real data, to see whether the method

would select a different model if the true underlying model was different.

$$
\begin{aligned}
V &= tvV \cdot (AGE/mean(AGE))^{dVdAGE} \cdot exp(dVdMALE0 \cdot (MALE0)) \cdot exp(nV) \\
V2 &= tvV2 \cdot (BSA/mean(BSA))^{dV2dBSA} \cdot (AGE/mean(AGE))^{dV2dAGE} \cdot exp(nV2) \\
V3 &= tvV3 \\
Cl &= tvCl \cdot (AGE/mean(AGE))^{dCldAGE} \cdot (BSA/mean(BSA))^{dCldBSA} \cdot exp(nCl) \\
Cl2 &= tvCl2 \cdot (AGE/mean(AGE))^{dCl2dAGE} \cdot exp(nCl2) \\
Cl3 &= tvCl3
\end{aligned}
$$

The fixed effects were set to the following values.

$$
\begin{aligned}
tvV &= 5.83 \\
tvV2 &= 8.65 \\
tvV3 &= 11.84 \\
tvCl &= 2.38 \\
tvCl2 &= 1.54 \\
tvCl3 &= 0.08 \\
dCldAGE &= -0.30 \\
dCldBSA &= 0.61 \\
dV2dBSA &= 0.74 \\
dV2dAGE &= -0.42 \\
dCl2dAGE &= -0.39 \\
dVdAGE &= -0.39 \\
dVdMALE0 &= -0.35
\end{aligned}
$$

The random effects (nV, nV2, nCl, and nCl2) were simulated to be normally distributed at the subject level (varying across subjects) with means of 0 and variance-covariance matrix given below. Ten replicates were simulated. See Figure 3.7 for a plot of the

Table 3.3: Omega for simulation

	nCl	nCl2	nV2	nV
nCl	0.02			
nCl2	0.04	0.17		
nV2	0.03	0.10	0.11	
nV	0.00	0.00	0.00	0.07

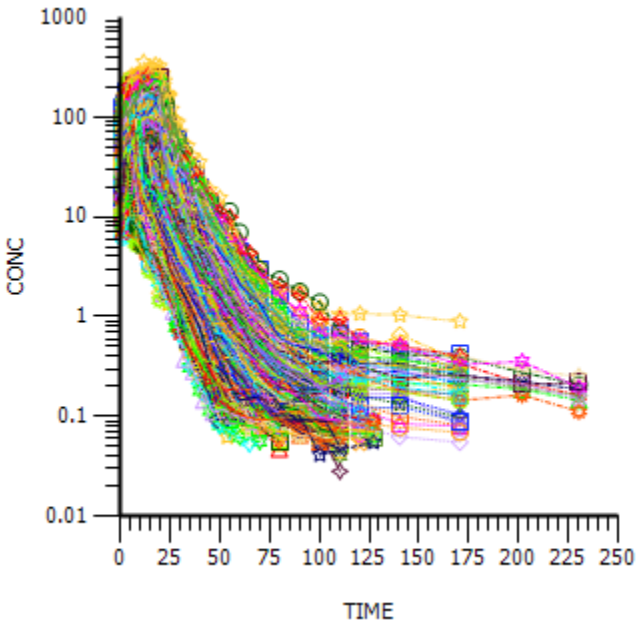

Figure 3.7: Simulated data

simulated data.

Once the data was simulated, the subject numbers were randomized for cross validation. Pharsights Phoenix NLME was used to fit models to the simulated data in batch mode using the automated covariate selection procedure described in this paper, as well as with stepwise with AIC and BIC as criteria. Initial estimates for the fixed effects parameters (tvV, tvV2, tvV3, tvCl, tvCl2, and tvCl3) were set to the true (simulated) parameter values. Initial estimates for the covariate effects were set to 0. The initial estimates of the variances of the random effects were all 0.1. The R program

Table 3.4: Models chosen for each replicate

Replicate	Parameter	Shrinkage-nPRESS	AIC	BIC
1	V	Male, Age	Age, Male	Age, Male
1	V2	Age	Age, BSA	BSA
1	Cl	Age, BSA	Age, BSA	Age, BSA
1	Cl2		Age, Male	
2	V	Age, Male	Age, Male	Age, Male
2	V2	Age	Age, Male	Age
2	Cl	Age, BSA	Age, BSA	Age, BSA
2	Cl2		Age	
3	V	Age, Male	Age, Male	Age, Male
3	V2	Age	Age, Male, BSA	Age
3	Cl	Age	Age, Male, BSA	Age, BSA
3	Cl2	Age	Age, Male	Age
4	V	Male, Age	Age, Male	Age, Male
4	V2	Age, BSA	Age, BSA	Age, BSA
4	Cl	Age, BSA	Age, BSA	Age, BSA
4	Cl2	Age	Age	Age
5	V	Age, Male	Age, Male	Age, Male
5	V2	Age	Age, BSA	Age
5	Cl	Age, BSA	Age, BSA	Age, BSA
5	Cl2	Age	Age	Age

was used to parse the output from the NLME batch models.

3.6 Simulation Results

The results of the simulations are summarized in Tables 3.4 and 3.5.

AIC tended to select models with too many covariates. BIC and Shrinkage-nPRESS tended to select models with too few covariates. In replicate four, all three methods selected the true underlying model.

Table 3.5: Models chosen for each replicate (cont'd)

Replicate	Parameter	Shrinkage-nPRESS	AIC	BIC
6	V	Age, Male	Age, Male	Age, Male
6	V2	Age	Age, BSA	Age
6	Cl	Age	Age, Male, BSA	Age, BSA
6	Cl2	Age	Age	Age
7	V	Male, Age	Age, Male	Age, Male
7	V2	Age	Age	Age
7	Cl	Male, Age	Age, Male, BSA	Age, Male
7	Cl2	Age	Age	Age
8	V	Male, Age	Age, Male	Age, Male
8	V2	Age	Age, Male	Age
8	Cl	Age, BSA	Age, BSA	Age, BSA
8	Cl2	Age	Age	Age
9	V	Male, Age	Age, Male	Male
9	V2	Age	Age, Male, BSA	Age
9	Cl	Age	Age, BSA	Age, BSA
9	Cl2		Age, Male	Age
10	V	Age, Male	Age, Male	Age, Male
10	V2	BSA	BSA	BSA
10	Cl	Age, BSA	Age, BSA	Age, BSA
10	Cl2		Male, BSA	

3.7 Discussion

The Shrinkage-nPRESS automated covariate selection method appeared to select a reasonable model in the real data analysis. In the simulation, BIC and Shrinkage-nPRESS performed similarly, while AIC selected models with too many covariates.

One might consider decreasing the shrinkage cutoff from 0.3 to 0.2 to avoid overfitting. In order to capture more of the variables that should be added, one might modify the method to reconsider adding more covariates to a parameter once covariates are added to other parameters.

This method might be applied with modifications in the linear mixed effects and generalized linear mixed effects modeling setting.

Note that as a fairly small number of replicates were generated for the simulation studies, this was not an attempt to estimate the true proportion of times the statistics will select the correct model in the scenario that was considered. It was simply to show that the methods proposed in this paper are worth considering. More extensive simulation studies would need to be completed in order to estimate the true proportions.

Chapter 4

Comparison of Smoothing Splines to Pop PK

4.1 Introduction

When predicted concentrations are the primary outcome of interest, rather than pharmacokinetic parameters, one may consider using a mixed effects smoothing spline model instead of a population pharmacokinetic model. This work is a comparison of mixed effect smoothing splines to population pharmacokinetic models.

4.2 Methods

A comparison of mixed effect spline models to population PK models will be performed. Simulations, with 400 replicates, each with 65 subjects, will be created for the following scenarios.

1. Rich data, correctly specified model
2. Rich data, model misspecification of population PK model
3. Sparser data, correctly specified model

For each replicate, the following will be reported:

Mean squared error for all predictions: mean of

$$(y_i - \hat{y}_i)^2 \tag{4.1}$$

We will look at standard errors across replicates to compare mixed effect spline models to population PK models.

4.2.1 Traditional Population Pharmacokinetic Modeling

Population pharmacokinetic and pharmacodynamic (PK/PD) modeling is the characterization of the distribution of probable PK/PD outcomes (parameters, concentrations, responses, etc.) in a population of interest. These models consist of fixed and random effects. The fixed effects describe the relationship between explanatory variables such as age, body weight, gender, and pharmacokinetic outcomes. The random effects quantify unexplained variation in PK/PD outcomes.

Population PK models are hierarchical. There is a model for the individual, a model for the population, and a model for the residual error. The individual model consists of the curve of drug concentrations over time, a compartmental model. The pharmacokinetic compartmental model is similar to a black box engineering model. Each of the compartments is like a black box, where a system of differential equations is derived based on the law of conservation of mass (Sandler, 1999).

The equations for the PK parameters represent the model for the population in the hierarchy of models. The PK parameters are modeled with regression equations containing fixed effects, covariates, and random effects (etas). The vector of random effects (eta) is assumed to follow a multivariate normal distribution with mean 0 and variance-covariance matrix Ω. Ω may be diagonal, full block, or block diagonal.

The model for the residual error accounts for overall uncertainty in the concentrations over time. The errors may weighted so that measurements with higher variability are given less weight compared with measurements with smaller variability.

Hence, population pharmacokinetic models are non-linear mixed effects models. The differential equations may or may not have a closed-form solution, and are solved either analytically or numerically. The parameters are estimated using one of the various algorithms available such as first order conditional estimation with interaction (FOCEI). See (Wang, 2007) for a mathematical description of these algorithms.

4.2.2 Smoothing Splines

Suppose we are given a set of response variables $\{y_i\}_{i=1}^n$ and predictor variables $\{x_i\}_{i=1}^n$ and we wish to estimate each y_i based on $f(x_i)$, where f is a function that minimizes

$$\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int \{f''(t)\}^2 \, dt \tag{4.2}$$

Any such f must be an element of the Sobolev space of functions with second derivatives that are square integrable. The tuning parameter λ controls the tradeoff between goodness of fit and smoothness. When $\lambda = \infty$, no second derivative is allowed for f, meaning that f must be linear and (4.2) reduces to the ordinary least squares criteria. When $\lambda = 0$, then any f that interpolates the data will minimize (4.2).

It can be shown that (4.2) is minimized when f is a natural cubic spline with knots at each x_i (Hastie, Tibshirani and Friedman, 2008). Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the order statistics of the x_i's. Then a natural cubic spline $f(x)$ with knots $x_1, x_2, \ldots, x_n$ satisfies the following properties:

1. $f(x)$ is a a piecewise cubic polynomial. In particular, $f(x)$ is a cubic polynomial on $[x_{(1)}, x_{(2)}], [x_{(2)}, x_{(3)}], \ldots, [x_{(n-1)}, x_{(n)}]$.

2. $f(x)$ and its first two derivatives are continuous on $[x_{(1)}, x_{(n)}]$.

3. $f^{(j)}(x_{(1)}) = f^{(j)}(x_{(n)}) = 0$ for $j = 2, 3$. In other words, the second and third derivatives of f are zero at the boundary knots, which implies that f is linear outside the boundary knots.

See (Welham, 2009) or (Dierckx, 1995). For a complete description of smoothing splines and methods for fitting spline models (including the choice of the tuning parameter λ), see (Hastie, Tibshirani and Friedman, 2008).

4.3 Simulated Data Examples

For the simulations, two studies were created in Pharsight's Trial Simulator. Both studies had the same population PK model: a three compartment model with an IV infusion of 20 mg/kg over 20 minutes (see Figure 4.1 for a graphical representation).

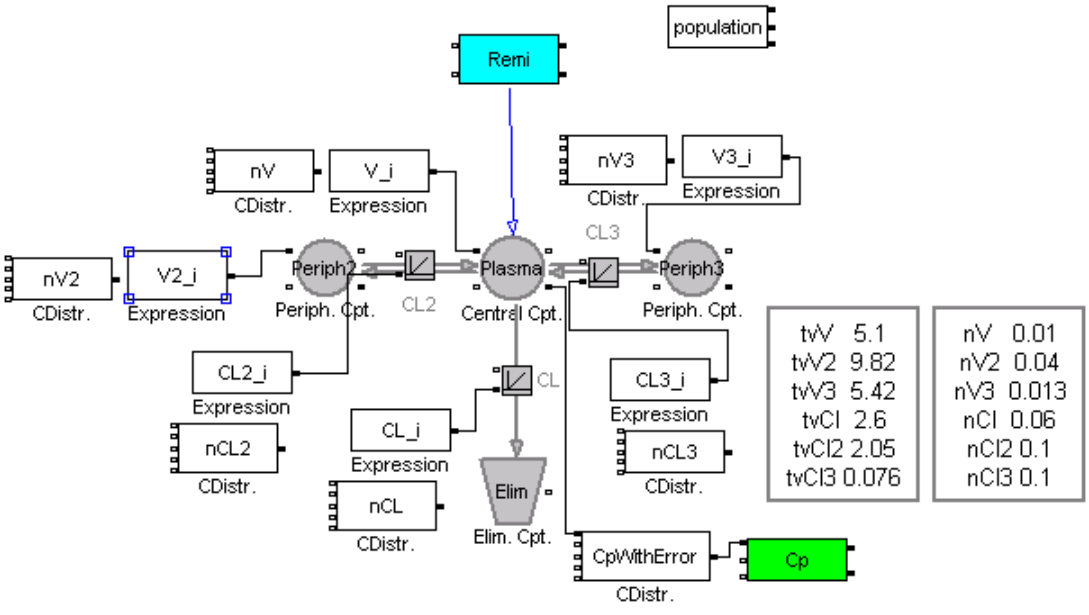

Figure 4.1: Simulated data

Four hundred replicates of concentration-time data for sixty-five subjects was simulated based on that model. The body weights were generated from a normal distribution with a mean of 60 kg and a standard deviation of 15 kg.

The equations for the model are as follows.

$$\frac{dA1}{dt} = In(t) - Cl \cdot C - Cl2 \cdot (C - C2) - Cl3 \cdot (C - C3)$$
$$\frac{dA2}{dt} = Cl2 \cdot (C - C2)$$
$$\frac{dA3}{dt} = Cl3 \cdot (C - C3)$$
$$C = A1/V$$
$$C2 = A2/V2$$
$$C3 = A3/V3$$

where In(t) = Dose/T during infusion time T, and 0 otherwise

Dose = 20 mg/kg

T = 20 min

A constant CV percentage residual error model was generated with a CV percentage of six percent.

CObs = C * (1 + CEps) where Var(CEps) = 0.0036

The PK parameters were simulated with random effects on all the parameters.

$$\begin{aligned} V &= 5.1 \cdot exp(nV) \\ V2 &= 9.82 \cdot exp(nV2) \\ V3 &= 5.42 \cdot exp(nV3) \\ Cl &= 2.6 \cdot exp(nCl) \\ Cl2 &= 2.05 \cdot exp(nCl2) \\ Cl3 &= 0.076 \cdot exp(nCl3) \end{aligned}$$

The random effects were assumed to be independent with means of zero and variances listed below.

$$\begin{aligned} Var(nV) &= 0.01 \\ Var(nV2) &= 0.04 \\ Var(nV3) &= 0.013 \\ Var(nCl) &= 0.06 \\ Var(nCl2) &= 0.1 \\ Var(nCl3) &= 0.1 \end{aligned}$$

The two studies differed only in the observation times (when samples were to be "collected"). Observations were planned for a "rich data" case and a "sparse data" case. For the "rich data" case, samples were to be collected at times (0, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 20.5, 21, 21.5, 22, 22.5, 23, 23.5, 24, 24.5, 25, 26, 27, 28, 29, 30, 32, 34, 36, 38, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110). For the 'sparse data' case, samples were to be collected at times (0, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10, 12, 14, 16, 24, 48).

After the data was simulated, models were fit using Phoenix NLME in command line batch mode, with the FOCE L-B method. There were three cases for the model fitting:

1. Rich data, model correctly specified

2. Rich data, model misspecified (2 compartment instead of 3 compartment)

3. Sparse data, model correctly specified

4.4 Results

The 400 simulated 'Rich data, model correctly specified' datasets each contained 3055 observations. The average MSE across the 400 replicates was 0.93 with a standard deviation of 0.13 for the population PK models, while the average MSE (standard deviation) was 2.26 (0.35) for the spline models.

The 400 simulated 'Rich data, model misspecified' datasets each contained 3055 observations. The average MSE across the 400 replicates was 29.44 with a standard deviation of 2.91.

The 400 simulated 'Sparse data, model correctly specified' datasets each contained 975 observations. The average MSE across the 400 replicates was 2.71 with a standard deviation of 0.40, while the average MSE (standard deviation) was 4.01 (0.73) for the spline models.

4.5 Real Data Examples

Pharsights Phoenix NLME was used to fit models to the published indomethacin dataset (Kwan et al., 1976). The indomethacin dataset, containing six subjects with eleven observations each, was fit using a two-compartment IV bolus model with Clearance parameterization and a proportional residual error model. Concentration units of ug/mL were assumed, and a dose of 25000 ug at 0 hours was assumed. Random effects were added to the PK parameters V, Cl, V2, and Cl2, in the form ThetaX*exp(nX), where X is the parameter of interest.

Table 4.1: Theta from final model of Indomethacin dataset

Parameter	Estimate	Units	Stderr	Stderr%
tvV	8898	mL	574.84	6.46
tvV2	19527.3	mL	3169.70	16.23
tvCl	7905.99	mL/h	608.53	7.70
tvCl2	5252.15	mL/h	768.93	14.64
stdev0	0.1440		0.02	13.25

stdev0 = estimated residual standard deviation
prefix of 'tv' denotes fixed effect or typical value

Exploratory analysis (plot, Figure 4.2) showed two compartment PK. Individual initial estimates were obtained using the curve stripping method (Gibaldi and Perrier, 1975) with a WinNonlin Classic model. The averages of the individual PK parameters were used as initial estimates for the pop PK model.

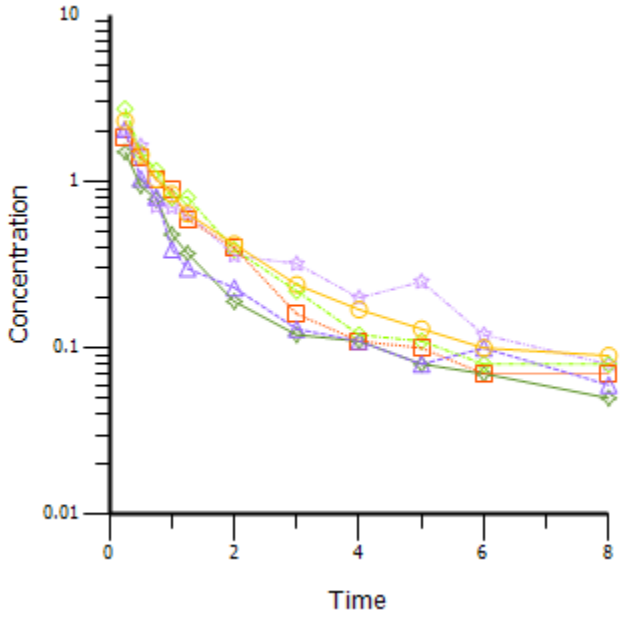

Figure 4.2: Concentration versus time from Indomethacin dataset

The model appeared to fit well based on diagnostic plots (Figures 4.3, 4.4, 4.5). The shrinkage was high for nV, but it was kept in. The relative standard errors for

Table 4.2: Omega from final model of Indomethacin dataset

	nV	nCl	nV2	nCl2
nV	0.0017			
nCl	0	0.0338		
nV2	0	0	0.0666	
nCl2	0	0	0	0.1202
Shrinkage	0.7064	0.0329	0.3321	0.0727

the fixed effect parameter estimates were all well below 30. The value of the original (Allen, 1974) PRESS statistic for the final model was 1.54, based on removal of 66 data points, one at a time. The average of Allen's PRESS over all data points was 0.02337.

4.6 Discussion

As expected, the pop PK models with the model misspecified had a higher MSE on average than the pop PK models with the model correctly specified. The sparser datasets with correctly specified models had significantly higher MSE on average than the rich datasets with the same fitted models. Compared with the spline models, the pop PK models tended to have lower MSE on average, however in the case where the pop PK model was misspecified, a spline model would be far superior in terms of prediction of concentrations.

In some cases, the data is too sparse to fit the correct structural pop PK model. For example, plots of concentration versus time may appear to indicate a two compartment model, but convergence fails when fitting a two compartment model while convergence can be achieved with a one compartment model. In these cases one might consider a spline model instead of a pop PK model if the primary outcome of interest is predicted concentrations.

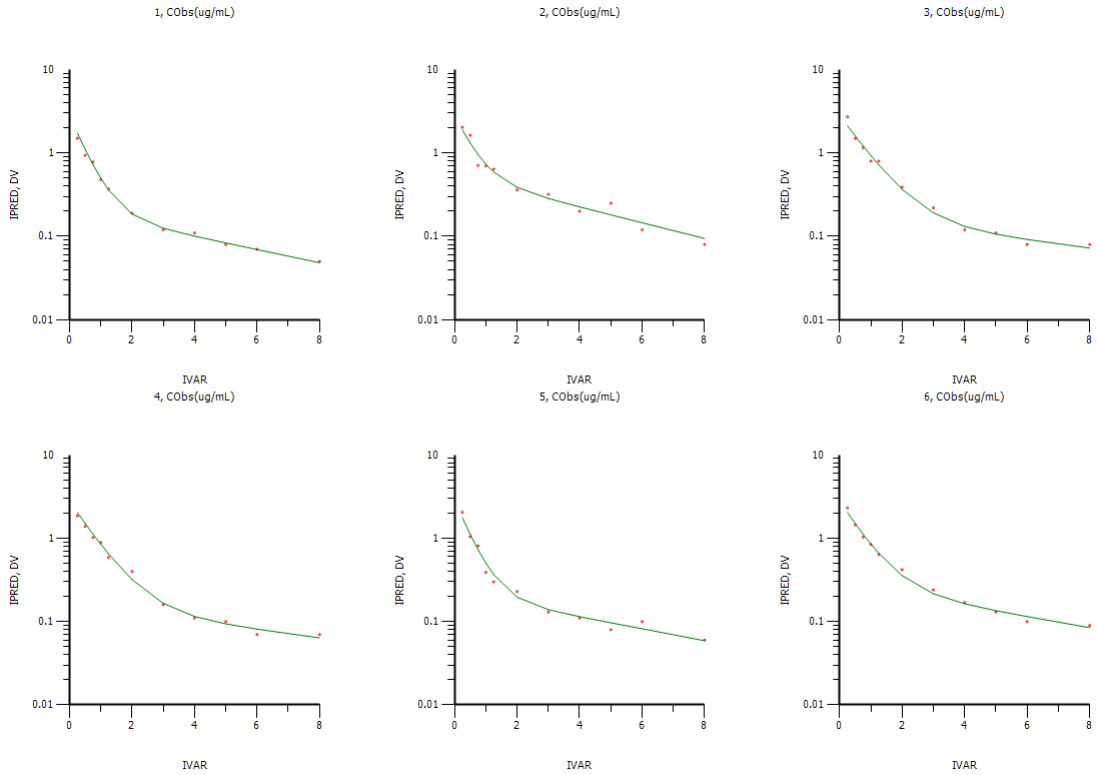

Figure 4.3: Final model of Indomethacin dataset

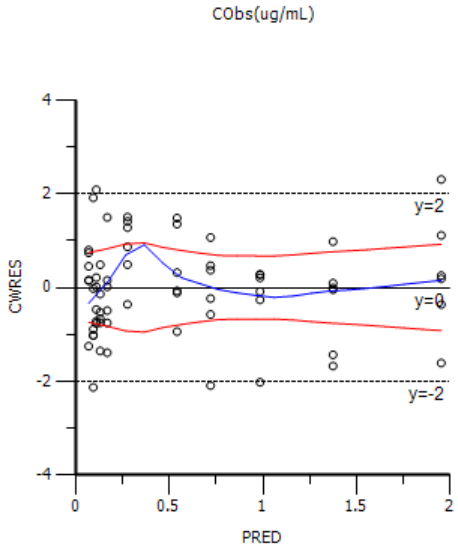

Figure 4.4: Residuals from final model of Indomethacin dataset

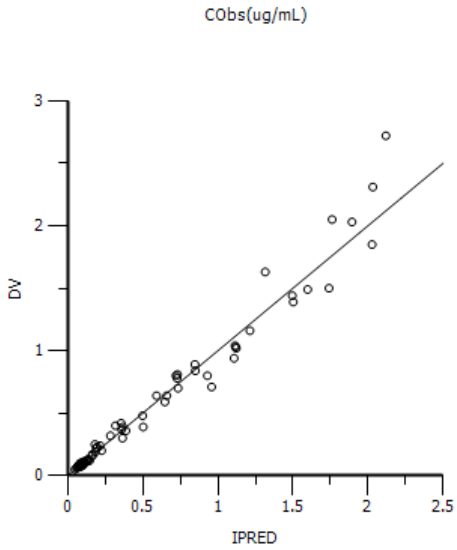

Figure 4.5: Observed versus predicted values from Indomethacin model

Appendix

Results log from automated covariate selection of Remifentanil data

Base

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1071 0.0072 0.0291 0.2678 0.0490 0.1995

Rank 4.0000 1.0000 2.0000 6.0000 3.0000 5.0000

Base

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1144 0.0515 0.1558 0.3333 0.2700 0.3693

SE 0.0168 0.0096 0.0294 0.0716 0.0455 0.0664

Sum 0.1313 0.0611 0.1852 0.4050 0.3156 0.4357

dCldAGE

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1157 0.0271 0.1581 0.2779 0.2780 0.3366

SE 0.0164 0.0053 0.0293 0.0598 0.0463 0.0504

Sum 0.1321 0.0323 0.1875 0.3377 0.3243 0.3869 Add ClAge

dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1126 0.0403 0.1582 0.3220 0.2792 0.3599

SE 0.0163 0.0087 0.0295 0.0683 0.0467 0.0629

Sum 0.1290 0.0491 0.1877 0.3903 0.3259 0.4228

dCldMALE0

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1152 0.0455 0.1582 0.2827 0.2733 0.3307

SE 0.0167 0.0090 0.0298 0.0618 0.0459 0.0513

Sum 0.1320 0.0544 0.1880 0.3445 0.3192 0.3820

dCldAGE

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1096 0.0142 0.0304 0.2743 0.0434 0.1905

Rank 4.0000 1.0000 2.0000 6.0000 3.0000 5.0000

dCldAGE

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1157 0.0271 0.1581 0.2779 0.2780 0.3366

SE 0.0164 0.0053 0.0293 0.0598 0.0463 0.0504

Sum 0.1321 0.0323 0.1875 0.3377 0.3243 0.3869

dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1207 0.0199 0.1563 0.2977 0.2712 0.3434

SE 0.0168 0.0043 0.0293 0.0642 0.0447 0.0511

Sum 0.1375 0.0242 0.1857 0.3619 0.3159 0.3945 Add ClBSA

dCldAGE dCldMALE0

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1167 0.0242 0.1563 0.2731 0.2752 0.3598

SE 0.0165 0.0048 0.0295 0.0618 0.0452 0.0538

Sum 0.1332 0.0290 0.1858 0.3349 0.3205 0.4135

dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1060 0.0174 0.0333 0.2705 0.0499 0.1515

Rank 4.0000 1.0000 2.0000 6.0000 3.0000 5.0000

dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1207 0.0199 0.1563 0.2977 0.2712 0.3434

SE 0.0168 0.0043 0.0293 0.0642 0.0447 0.0511

Sum 0.1375 0.0242 0.1857 0.3619 0.3159 0.3945

dCldAGE dCldBSA dCldMALE0

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1197 0.0203 0.1548 0.2736 0.2739 0.3622

SE 0.0169 0.0044 0.0292 0.0627 0.0452 0.0540

Sum 0.1365 0.0247 0.1840 0.3363 0.3191 0.4162 Don't add ClMale

dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1060 0.0174 0.0333 0.2705 0.0499 0.1515

Rank 4.0000 1.0000 2.0000 6.0000 3.0000 5.0000

dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1207 0.0199 0.1563 0.2977 0.2712 0.3434

SE 0.0168 0.0043 0.0293 0.0642 0.0447 0.0511

Sum 0.1375 0.0242 0.1857 0.3619 0.3159 0.3945

dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1175 0.0198 0.1105 0.3153 0.2866 0.3240

SE 0.0167 0.0043 0.0205 0.0631 0.0485 0.0479

Sum 0.1342 0.0241 0.1309 0.3784 0.3351 0.3719 Add V2Age

dV2dBSA dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1161 0.0198 0.1300 0.2683 0.2664 0.3325

SE 0.0165 0.0043 0.0268 0.0600 0.0440 0.0494

Sum 0.1326 0.0241 0.1569 0.3283 0.3104 0.3819

dV2dMALE0 dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1160 0.0198 0.1250 0.2983 0.2685 0.3080

SE 0.0166 0.0043 0.0236 0.0602 0.0442 0.0494

Sum 0.1326 0.0241 0.1486 0.3584 0.3127 0.3573

dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1072 0.0176 0.0457 0.2470 0.0342 0.1993

Rank 4.0000 1.0000 3.0000 6.0000 2.0000 5.0000

dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1175 0.0198 0.1105 0.3153 0.2866 0.3240

SE 0.0167 0.0043 0.0205 0.0631 0.0485 0.0479

Sum 0.1342 0.0241 0.1309 0.3784 0.3351 0.3719

dCl2dAGE dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1134 0.0193 0.0989 0.3370 0.1334 0.3636

SE 0.0158 0.0042 0.0192 0.0667 0.0225 0.0542

Sum 0.1292 0.0236 0.1182 0.4038 0.1559 0.4178 Add Cl2Age

dCl2dBSA dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1141 0.0197 0.1083 0.3345 0.2728 0.3016

SE 0.0162 0.0043 0.0201 0.0657 0.0454 0.0475

Sum 0.1303 0.0239 0.1284 0.4002 0.3182 0.3490

dCl2dMALE0 dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1164 0.0196 0.1075 0.3056 0.2820 0.3509

SE 0.0167 0.0043 0.0205 0.0630 0.0460 0.0489

Sum 0.1331 0.0239 0.1280 0.3686 0.3280 0.3997

dCl2dAGE dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.0958 0.0104 0.0591 0.2808 0.0783 0.1616

Rank 4.0000 1.0000 2.0000 6.0000 3.0000 5.0000

dCl2dAGE dV2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1134 0.0193 0.0989 0.3370 0.1334 0.3636

SE 0.0158 0.0042 0.0192 0.0667 0.0225 0.0542

Sum 0.1292 0.0236 0.1182 0.4038 0.1559 0.4178

dV2dAGE dV2dBSA dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1081 0.0194 0.0870 0.2926 0.1344 0.3580

SE 0.0153 0.0042 0.0164 0.0625 0.0229 0.0518

Sum 0.1235 0.0236 0.1034 0.3552 0.1573 0.4098

dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1148 0.0194 0.0820 0.3061 0.1300 0.3321

SE 0.0157 0.0042 0.0153 0.0613 0.0220 0.0499

Sum 0.1305 0.0236 0.0973 0.3674 0.1520 0.3819 Add V2Male

dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.0975 0.0207 0.0771 0.2351 0.1026 0.2017

Rank 3.0000 1.0000 2.0000 6.0000 4.0000 5.0000

dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1148 0.0194 0.0820 0.3061 0.1300 0.3321

SE 0.0157 0.0042 0.0153 0.0613 0.0220 0.0499

Sum 0.1305 0.0236 0.0973 0.3674 0.1520 0.3819

dV2dAGE dV2dMALE0 dV2dBSA dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1080 0.0194 0.0827 0.2785 0.1301 0.3654

SE 0.0150 0.0042 0.0155 0.0616 0.0225 0.0512

Sum 0.1230 0.0236 0.0981 0.3401 0.1526 0.4166 Don't add V2BSA

dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.0975 0.0207 0.0771 0.2351 0.1026 0.2017

Rank 3.0000 1.0000 2.0000 6.0000 4.0000 5.0000

dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.1148 0.0194 0.0820 0.3061 0.1300 0.3321

SE 0.0157 0.0042 0.0153 0.0613 0.0220 0.0499

Sum 0.1305 0.0236 0.0973 0.3674 0.1520 0.3819

dVdAGE dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0922 0.0194 0.0842 0.2858 0.1320 0.3087

SE 0.0151 0.0042 0.0153 0.0602 0.0225 0.0484

Sum 0.1073 0.0236 0.0995 0.3460 0.1545 0.3572

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0498 0.0195 0.0854 0.3171 0.1356 0.3875

SE 0.0073 0.0043 0.0151 0.0653 0.0240 0.0724

Sum 0.0571 0.0237 0.1006 0.3824 0.1596 0.4599 Add VBSA

dVdMALE0 dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0631 0.0195 0.0835 0.3037 0.1453 0.3300

SE 0.0102 0.0043 0.0155 0.0633 0.0245 0.0504

Sum 0.0733 0.0237 0.0990 0.3670 0.1698 0.3804

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1690 0.0200 0.0553 0.3084 0.0808 0.1667

Rank 5.0000 1.0000 2.0000 6.0000 3.0000 4.0000

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0498 0.0195 0.0854 0.3171 0.1356 0.3875

SE 0.0073 0.0043 0.0151 0.0653 0.0240 0.0724

Sum 0.0571 0.0237 0.1006 0.3824 0.1596 0.4599

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCl2dBSA dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0511 0.0194 0.0835 0.2962 0.1357 0.3966

SE 0.0074 0.0043 0.0152 0.0622 0.0242 0.0725

Sum 0.0585 0.0237 0.0988 0.3584 0.1599 0.4691 Don't add Cl2BSA

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCl2dMALE0 dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0515 0.0195 0.0846 0.2948 0.1354 0.3961

SE 0.0075 0.0043 0.0152 0.0617 0.0244 0.0733

Sum 0.0590 0.0238 0.0998 0.3565 0.1598 0.4694 Don't add Cl2Male

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

Shrinkage 0.1690 0.0200 0.0553 0.3084 0.0808 0.1667

Rank 5.0000 1.0000 2.0000 6.0000 3.0000 4.0000

dVdBSA dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0498 0.0195 0.0854 0.3171 0.1356 0.3875

SE 0.0073 0.0043 0.0151 0.0653 0.0240 0.0724

Sum 0.0571 0.0237 0.1006 0.3824 0.1596 0.4599

dVdBSA dVdAGE dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0453 0.0193 0.0823 0.2937 0.1340 0.3892

SE 0.0070 0.0042 0.0151 0.0648 0.0236 0.0662

Sum 0.0523 0.0236 0.0974 0.3586 0.1577 0.4555 Don't add VAge

dVdBSA dVdMALE0 dV2dAGE dV2dMALE0 dCl2dAGE dCldAGE dCldBSA

nV nCl nV2 nV3 nCl2 nCl3

nPRESS 0.0456 0.0194 0.0831 0.3019 0.1385 0.3795

SE 0.0067 0.0043 0.0152 0.0627 0.0246 0.0648

Sum 0.0523 0.0237 0.0983 0.3646 0.1632 0.4443 Don't add VBSA

Bibliography

Allen, David M. 1974. "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction." *Technometrics* 16(1):125–127.

Bailey, J. M., C. T. Mora and S. L. Shafer. 1996. "Pharmacokinetics of propofol in adult patients undergoing coronary revascularization." *Anesthesiology* 84(1):288–97.

Bertrand, J., E. Comets, C. M. Laffont, M. Chenel and F. Mentre. 2009. "Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm." *J Pharmacokinet Pharmacodyn* 36(4):317–339.

Boeckmann, A.J., L.B. Sheiner and S.L. Beal. 1992. *NONMEM Users Guide: Part V*. University of California, San Francisco.

Bonate, PL. 2006. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer.

Brendel, K, C Dartois, E Comets, A Lemenuel-Diot, C Laveille, B Tranchand, P Girard, CM Laffont and F Mentre. 2007. "Are population pharmacokinetic and/or pharmacodynamic models adequately evaluated? A survey of the literature from 2002 to 2004." *Clin Pharmacokinet.* 46(3):221–234.

Davidian, M. and D. M. Giltinan. 1995. *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall.

Dierckx, P. 1995. *Curve and surface fitting with splines*. Monographs on numerical analysis New York, NY: Clarendon Press.

Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Efron, B. and R. Tibshirani. 1986. "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy." *Statistical Science* 1(1):54–77.

Ette, E. I. and P. J. Williams. 2004. "Population pharmacokinetics II: estimation methods." *Ann Pharmacother* 38(11):1907–1915.

FDA. 1999. "Guidance for industry on Population Pharmacokinetics. Food and Drug Administration, HHS. Notice." *Fed Regist* .

Fiset, P., L. Mathers, R. Engstrom, D. Fitzgerald, S. C. Brand, F. Hsu and S. L. Shafer. 1995. "Pharmacokinetics of computer-controlled alfentanil administration in children undergoing cardiac surgery." *Anesthesiology* 83(5):944–955.

Gabrielsson, J and D Weiner. 2000. *Pharmacokinetic and Pharmacodynamic Data Analysis*. 4 ed. Swedish Pharmaceutical Press.

Gibaldi, M and D Perrier. 1975. *Pharmacokinetics*. 1 ed. M. Dekker.

Hastie, T., R. Tibshirani and J.H. Friedman. 2008. *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics New York, NY: Springer.

Hooker, AC, AJ Ten Tije, MA Carducci, J Weber, E Garrett-Mayer, H Gelderblom, WP Mcguire, J Verweij, MO Karlsson and SD Baker. 2008. "Population Pharmacokinetic Model for Docetaxel in Patients with Varying Degrees of Liver Function: Incorporating Cytochrome P450 3A Activity Measurements." *Clinical Pharmacology and Therapeutics* 84(1):111–118.

Jonsson, E. N. and M. O. Karlsson. 1998. "Automated covariate model building within NONMEM." *Pharm. Res.* 15(9):1463–1468.

Jonsson, E. N. and M. O. Karlsson. 1999. "Xpose–an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM." *Comput Methods Programs Biomed* 58(1):51–64.

Karlsson, M. O. and N. H. G. Holford. 2008. "A Tutorial on Visual Predictive Checks." PAGE 17 (2008) Abstr 1434.
URL: *www.page-meeting.org/?abstract=1434*

Karlsson, M. O. and R. M. Savic. 2007. "Diagnosing model diagnostics." *Clin. Pharmacol. Ther.* 82(1):17–20.

Katsube, T., A. Khandelwal, K. Harling, A. C. Hooker and M. O. Karlsson. 2011. "Evaluation of Stepwise Covariate Model Building Combined with Cross-Validation." PAGE 20 (2011) Abstr 2111.
URL: *www.page-meeting.org/?abstract=2111*

Kerbusch, T., J. de Kraker, R. A. Mathot and J. H. Beijnen. 2001. "Population pharmacokinetics of ifosfamide and its dechloroethylated and hydroxylated metabolites in children with malignant disease: a sparse sampling approach." *Clin Pharmacokinet* 40(8):615–625.

Kwan, K. C., G. O. Breault, E. R. Umbenhauer, F. G. McMahon and D. E. Duggan. 1976. "Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man." *J Pharmacokinet Biopharm* 4(3):255–280.

Leary, R. H. and M. Dunlavey. 2012. "QRPEM, A Quasi-Random Parametric EM Method." PAGE 21 (2012) Abstr 2338.
URL: *www.page-meeting.org/?abstract=2338*

Lindstrom, M. L. and D. M. Bates. 1990. "Nonlinear mixed effects models for repeated measures data." *Biometrics* 46(3):673–687.

Maitre, P. O., M. Buhrer, D. Thomson and D. R. Stanski. 1991. "A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam." *J Pharmacokinet Biopharm* 19(4):377–384.

Mandema, J. W., D. Verotta and L. B. Sheiner. 1992. "Building population pharmacokinetic–pharmacodynamic models. I. Models for covariate effects." *J Pharmacokinet Biopharm* 20(5):511–528.

Minto, C. F., T. W. Schnider, T. D. Egan, E. Youngs, H. J. Lemmens, P. L. Gambus, V. Billard, J. F. Hoke, K. H. Moore, D. J. Hermann, K. T. Muir, J. W. Mandema and S. L. Shafer. 1997. "Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development." *Anesthesiology* 86(1):10–23.

Mulla, H., F. Nabi, S. Nichani, G. Lawson, R. K. Firmin and D. R. Upton. 2003. "Population pharmacokinetics of theophylline during paediatric extracorporeal membrane oxygenation." *Br J Clin Pharmacol* 55(1):23–31.

Rajagopalan, P. and M. R. Gastonguay. 2003. "Population pharmacokinetics of ciprofloxacin in pediatric patients." *J Clin Pharmacol* 43(7):698–710.

Ralph, L. D., M. Sandstrom, C. Twelves, N. A. Dobbs and A. H. Thomson. 2006. "Assessment of the validity of a population pharmacokinetic model for epirubicin." *Br J Clin Pharmacol* 62(1):47–55.

Ribbing, J. and E. N. Jonsson. 2001. "Cross Model Validation As A Tool For Population Pharmacokinetic/Pharmacodynamic Covariate Model Building." PAGE 10 (2001) Abstr 215.
URL: *www.page-meeting.org/?abstract=215*

Sandler, SI. 1999. *Chemical and Engineering Thermodynamics*. Wiley.

Savic, R. M. and M. O. Karlsson. 2007. "Importance of Shrinkage in Empirical Bayes Estimates for Diagnostics and Estimation: Problems and Solutions." PAGE 16 (2007) Abstr 1087.
URL: *www.page-meeting.org/?abstract=1087*

Sheiner, L. B., B. Rosenberg and K. L. Melmon. 1972. "Modelling of individual pharmacokinetics for computer-aided drug dosage." *Comput. Biomed. Res.* 5(5):411–459.

Valodia, P. N., M. A. Seymour, M. L. McFadyen, R. Miller and P. I. Folb. 2000. "Validation of population pharmacokinetic parameters of phenytoin using the parallel Michaelis-Menten and first-order elimination model." *Ther Drug Monit* 22(3):313–319.

Wahlby, U., E. N. Jonsson and M. O. Karlsson. 2001. "Assessment of actual significance levels for covariate effects in NONMEM." *J Pharmacokinet Pharmacodyn* 28(3):231–252.

Wahlby, U., E. N. Jonsson and M. O. Karlsson. 2002. "Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis." *AAPS PharmSci* 4(4):E27.

Wang, Y. 2007. "Derivation of various NONMEM estimation methods." *J Pharmacokinet Pharmacodyn* 34(5):575–593.

Welham, S.J. 2009. Smoothing spline models for longitudinal data. In *Longitudinal data analysis*, ed. G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs. Chapman & Hall/CRC handbooks of modern statistical methods Boca Raton, FL: CRC Press chapter 12, pp. 253–289.

Yafune, A. and M. Ishiguro. 1999. "Bootstrap approach for constructing confidence intervals for population pharmacokinetic parameters. I: A use of bootstrap standard error." *Stat Med* 18(5):581–599.

Zomorodi, K., A. Donner, J. Somma, J. Barr, R. Sladen, J. Ramsay, E. Geller and S. L. Shafer. 1998. "Population pharmacokinetics of midazolam administered by target controlled infusion for sedation following coronary artery bypass grafting." *Anesthesiology* 89(6):1418–1429.