*Themed Issue: Population Pharmacokinetics - A Memorial Tribute To Lewis Sheiner, M.D.*
*Guest Editors - Peter Bonate and Diane Mould*

# Recommended Reading in Population Pharmacokinetic Pharmacodynamics

*Submitted: March 30, 2005; Accepted: April 12, 2005; Published: October 5, 2005*

Peter L. Bonate[1]

[1]Genzyme Oncology, San Antonio, TX

## ABSTRACT

Developing the skills or expertise to create useful population pharmacokinetic-pharmacodynamic models can be a daunting task-the level of mathematical and statistical complexity is such that newcomers to the field are frequently overwhelmed. A good place to start in learning the field is to read articles in the literature. However, the number of articles dealing with population pharmacokinetic pharmacodynamics is exponentially increasing on a yearly basis, so choosing which articles to read can be difficult. The purpose of this review is to provide a recommended reading list for newcomers to the field. The list was chosen based on perceived impact of the article in the field, the quality of the article, or to highlight some important detail contained within the article. After reading the articles in the list, it is believed that the reader will have a broad overview of the field and have a sound foundation for more-detailed reading of the literature.

**KEYWORDS:** NONMEM, influential articles, review, first-order approximation

## INTRODUCTION

In the 1970s and early 1980s, pharmacokinetics was moving away from compartmental models, which were the foundation of the field, toward noncompartmental or so-called model-independent analyses because of the perceived problems with compartmental models. For example, one modeler might choose a 1-compartment model with first-order (FO) absorption, whereas another might choose the same model but include a lag time. Alternatively, within the same cohort of subjects, data for most individuals may be consistent with one type of model, but with a handful of individuals, their data may be more consistent with another type of model. For example, most patients may be more consistent with a 2-compartment model, but others may be more consistent with a 3-compartment model. These inconsistencies, in addition to other problems, led to criticism of the compartmental approach. Hence, data analysis began moving from a kinetic foundation to a summary statistic approach.

**Corresponding Author:** Peter L. Bonate, Genzyme Corporation, 4545 Horizon Hill Blvd., San Antonio, TX 78229; Tel: (210) 949-8662; Fax: (210) 949-8219; E-mail: peter.bonate@genzyme.com

In the late 1970s and early 1980s, Sheiner, Beal, and colleagues[1-4] published a series of articles describing a new approach to pharmacokinetic data analysis, which was later coined population pharmacokinetics (PopPK), and introduced a new software package called NONMEM (currently distributed by GloboMax LLC, Hanover, MD) that was capable of performing the necessary operations to do this new type of analysis. Using phenytoin data collected as part of a therapeutic drug monitoring program in epileptic patients, Sheiner and Beal[2,3] showed how estimates of the population-pharmacokinetic parameters could be obtained, although most patients had only 2 or 3 samples collected. They also showed how naïve pooling of the data and the 2-stage approach (which will be discussed later) both result in biased parameter estimates.

Although it is clear that population-based methods were initially developed with an eye toward the analysis of data obtained from routine clinical monitoring and developing individualizing dose regimens based on patient-specific covariates, population-based methods now play a large role in drug development. It is fair to say that PopPK has revolutionized how data from clinical studies is analyzed. Granted, a noncompartmental approach is still used for many phase I studies, but population-based methods are used almost exclusively for phase II and III studies and to summarize data across a drug development program. Hence, noncompartmental and compartmental models, including PopPK, are seen as complementary approaches with each having their own role.

PopPK was slow to take off, like most new technologies, but is now growing rapidly. A MEDLINE search using the key terms “population pharmacokinetics” or “NONMEM” shows that the number of publications has been exponentially increasing for a number of years (Figure 1). One of the problems with PopPK for a long period of time was the difficulty in understanding the nomenclature and mathematics, the heavy reliance on statistics, and the difficulty in using NONMEM. These problems still exist, but now, after many years and thanks to such resources as the NONMEM Users Group (http://www.cognigencorp.com/nonmem/nm/) and more introductory courses, users are more comfortable with the problems. For a newcomer, however, PopPK is as daunting as ever–more so since the mathematics and statistical foundation have become more complex with the introduction of conditional estimation algorithms. A newcomer

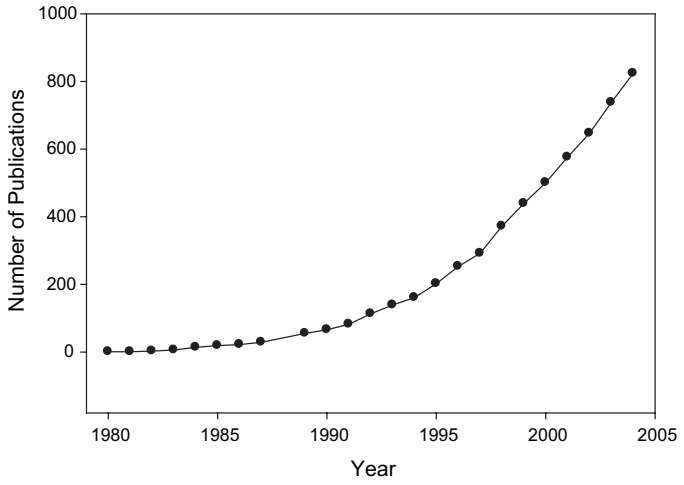


**Figure 1.** Number of articles published each year as reported by MEDLINE using the keywords "population pharmacokinetics" or "NONMEM."

can read the literature as a source to learn about PopPK. But even this task can be difficult, because there are almost 1,000 publications related to PopPK as of the end of 2004.

The purpose of this article is to provide a recommended reading list for newcomers to the field. The list was chosen based on the perceived impact of the manuscript in the field, on the quality of the article, or to highlight some important detail contained within the manuscript. After reading the articles in the list, it is believed that the reader will have a broad overview of the field and will have a more sound foundation for more detailed reading of the literature.

### *Articles of Historical Importance*

The articles by Sheiner et al.[1] and Sheiner and Beal[2-4] formed the basis for the field today as we know it by introducing the concept of first-order (FO) approximation. The discussion that follows is the development of FO approximation based on the original set of papers and may not reflect the current understanding. Given a set of {x, Y} data, where x is the design matrix of predictor variables measured without error and Y is the dependent variable with a single observation per individual, and a model $f(\theta, x)$, ordinary least-squares finds the vector of $\theta$ called $\hat{\theta}$ that minimizes the function

$$O(\hat{\theta}, Y) = \sum_{i=1}^{n} \left[Y_i - f(\hat{\theta}, x_i)\right]^2 \qquad (1)$$

Equation 1 is a type of objective function; there are many others as will be shown. This approach assumes that the residuals are identically distributed and independent with mean zero and constant variance $\sigma^2$. If the assumption that the residuals are normally distributed is added, the model parameter estimates $\hat{\theta}$ are also the maximum likelihood estimates.

When applied to pharmacokinetic data where there is at least 1 observation per subject, this method is called the naïve pooled approach with the word "naïve" added, because the method violates the assumption of independence in the residuals, as it is likely that observations within an individual are correlated. With this method, in terms of the variability in the data, only an estimate of $\sigma^2$ can be obtained. Estimates of the between-subject variability in $\theta_i$ cannot be obtained, and so $\sigma^2$ represents all of the sources of variability. If enough data are available on each subject, then the model can be fit on the data of each subject to obtain a set of estimates $\{\hat{\theta}_1, \hat{\theta}_2, ...\hat{\theta}_n\}$, each of which are independent realizations for $\theta$, with the subscript in this case denoting that $\theta_i$ represents the vector of estimates for the ith subject. It should be noted that $\theta$ is a vector of population parameters, such as clearance and volume of distribution, and that $\hat{\theta}_i$ denotes the estimate of clearance and volume of distribution for each subject. The estimates of the population mean and variance of $\theta$ are then computed using the set of $\hat{\theta}_i$s. $\sigma^2$ can be estimated from the sum of the residual errors from each individual fit divided by the total degrees of freedom. This method is called the 2-stage approach.

If the variance of the observations change from observation to observation, such as is observed when examining the variability around maximal concentrations to the variability observed some time later in a concentration-time curve, then weighed least-squares (WLS) can be used to obtain an estimate of $\theta$. WLS minimizes the function

$$O(\hat{\theta}, Y) = \sum_{i=1}^{n} \frac{1}{w_i} \left[Y_i - f(\hat{\theta}, x_i)\right]^2 \qquad (2)$$

where $w_i$ is proportional to the variance of the observation. If the variances of the observations are not known, then they too could be modeled such that the objective function becomes

$$O\left(\hat{\theta}, Y, \hat{\xi}\right) = \sum_{i=1}^{n} \frac{1}{g\left(\hat{\theta}, x_i, \hat{\xi}\right)} \left[Y_i - f(\hat{\theta}, x_i)\right]^2 \qquad (3)$$

where $g(\theta, x, \xi)$ models the variance of the observation, and $\xi$ is a parameter that is unique to the variance model. It should be noted that when $g(\hat{\theta}, x, \hat{\xi}) = f(\hat{\theta}, x)^2$ this corresponds with a constant coefficient of variation residual error model.

However, it is not possible to model both the expected value function $f(\theta, x)$ and the variance function $g(x, \theta, \xi)$ simultaneously using WLS, because the variance function could be made arbitrarily large, thereby causing the objective function to continue to decrease until such a point as the modeled variance is infinite and the objective function is zero, regardless of the value of $\theta$. Hence, a penalty term is added to the objective function such that the objective

function increases in response to an increase in the variance model

$$O\left(\hat{\theta}, Y, \hat{\xi}\right) = \sum_{i=1}^{n} \left\{ \frac{1}{g\left(\hat{\theta}, x_i, \hat{\xi}\right)} \left[Y_i - f\left(\hat{\theta}, x_i\right)\right]^2 + Ln\left[g\left(\hat{\theta}, x_i, \hat{\xi}\right)\right] \right\} \quad (4)$$

This objective function forms the basis of extended least-squares.[5,6] In matrix notation, which allows θ to be multivariate, the objective function in equation 4 can be written as follows:

$$O\left(\hat{\theta}, Y, \hat{\xi}\right) = \sum_{i=1}^{n} \left[Y_i - f\left(\hat{\theta}, x_i\right)\right]^T g\left(\hat{\theta}, x_i, \hat{\xi}\right)^{-1} \times \left[Y_i - f\left(\hat{\theta}, x_i\right)\right] + Ln\left[\left|g\left(\hat{\theta}, x_i, \hat{\xi}\right)\right|\right] \quad (5)$$

where T is the transpose, $^{-1}$ is the inverse, and $|\cdot|$ is the determinant function for some matrix.

With nonlinear mixed effects, models to f(θ, x) and g(θ, x, ξ) must be expanded to include the random effects in the model to f(θ, x, η) and g(θ, x, ξ, η), respectively. However, the random effects in the model are nuisance variables, and an approximation to f(θ, x) and g(θ, x, ξ) must be found before equation 5 can be used. This is accomplished by first expanding the nonlinear mixed effects model using a Taylor series approximation to find f(θ, x) and then applying statistical theory based on a linear combination of independent random variables to find g(θ, x, ξ). To best understand these concepts is by example. Suppose the nonlinear mixed effects model was a 1-compartment model parameterized in terms of clearance (CL) and central volume (V) where both CL and V can vary across subjects:

$$C(t) = \frac{D}{V + \eta_V} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right) + \varepsilon \quad (6)$$

where t is time and D is dose, both of which comprise the design matrix, ε is the vector of independent, normally distributed residuals having zero mean and variance $\sigma^2$, and $\eta_{CL}$ and $\eta_V$ are the between-subject random effects for CL and V having zero mean and variance $\omega^2_{CL}$ and $\omega^2_V$, respectively. $\omega^2_{CL}$ and $\omega^2_V$ collectively comprise Ω, a diagonal matrix with zero off-diagonal elements.

To obtain approximate expressions for f(θ, x) and g(θ, x, ξ), Sheiner and Beal[1] proposed taking a FO approximation to equation 6 around the random effects:

$$C(t) = \frac{D}{V + \eta_{0,V}} \exp\left(-\frac{CL + \eta_{0,CL}}{V + \eta_{0,V}} t\right) - \frac{Dt}{(V + \eta_V)^2} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right)(\eta_{CL} - \eta_{0,CL}) + \left[\frac{Dt(CL + \eta_{CL})}{(V + \eta_V)^3} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right) - \frac{D}{(V + \eta_V)^2} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right)\right](\eta_V - \eta_{0,V}) + \varepsilon^* \quad (7)$$

and then evaluating the expression around $\eta_0 = 0$ for all of the parameters associated with a random effect η. Hence, equation 7 simplifies to the following:

$$C(t) = \frac{D}{V} \exp\left(-\frac{CL}{V} t\right) - \frac{Dt}{(V + \eta_V)^2} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right)(\eta_{CL}) - \left[\frac{D}{(V + \eta_V)^2} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right) + \frac{Dt(CL + \eta_{CL})}{(V + \eta_V)^3} \exp\left(-\frac{CL + \eta_{CL}}{V + \eta_V} t\right)\right](\eta_V) + \varepsilon^* \quad (8)$$

which is of the form

$$C(t) = f(x, \theta) + G_{CL}(x, \theta)\eta_{CL} + G_V(x, \theta)\eta_V + \varepsilon^*. \quad (9)$$

It should be noted that $\varepsilon \neq \varepsilon^*$ as the latter term also includes truncation error by the approximation, although it is still assumed that ε* is independent with zero mean and variance $\sigma^2$. The expected value of equation 9 is f(x, θ). Hence, an approximation to f(θ, x, η) is f(θ, x). Given a linear function of independent random variables $a_1, a_2, \ldots a_n$ and constants $b_0, b_1, b_2, \ldots b_n$

$$Z = b_0 + b_1 a_1 + b_2 a_2 + \ldots . b_n a_n \quad (10)$$

the variance of the function is given by

$$Var[Z] = \sum_{i=1}^{n} b_i^2 E(a_i), \quad (11)$$

where $E(a_i)$ is the expected value of $a_i$. Applying this expression to equation 9 in multivariate form gives the approximation to the variance model as:

$$g\left(x_i, \hat{\theta}, \hat{\xi}, \hat{\eta}\right) \cong g\left(x_i, \hat{\theta}, \hat{\xi}\right) = G(\theta, x)\Omega G(\theta, x)^T + \sigma^2 I \quad (12)$$

where $G = [G_{CL}\ G_V]$ is the Jacobian matrix of first derivatives of the function f(θ, x) evaluated at $\eta_0 = 0$ and I is the identity matrix. Thus, by knowing the expected value and

variance of the observations, extended least-squares can be used to find the population estimate of θ while simultaneously estimating the between-subject variability Ω and residual variability $\sigma^2$. This is the FO-approximation algorithm applied to the model as presented in equation 6; the expected value and variance will change for different models.

In the first article in the series, Sheiner and Beal[1] used this methodology to characterize the pharmacokinetics of digoxin in plasma and urine. The article does an excellent job of laying the groundwork and providing a rationale for the method, but the technical details are not very illuminating. Sheiner and Beal[2] followed up their work and renamed their estimation method "FO-approximation." It is in this article that the methodology is more fully clarified and illustrated by predicting the phenytoin dose needed to obtain a particular average steady-state concentration. Using simulation, they showed that the population approach produced less-biased estimates of the population means and variances than either the naïve-pooled or 2-stage approach. In their third article,[3] they studied the method using data simulated under a 2-compartment model after intravenous administration for 3 different drugs. Their last paper[4] studied the behavior of the method using data simulated from a 1-compartment model after repeated intravenous infusions to steady state. They examined how the population estimates were influenced by the number of subjects, the number of observations per subject, and the choice of sampling design.

### *Early Covariate Models*

Vozeh et al and others[7-10] represent the first set of reports detailing how covariates can be used to predict a patient's pharmacokinetics through the development of covariate submodels in a population analysis. For example, Grasela and Donn[9] showed that in pediatric patients with seizures or intraventricular hemorrhage treated with phenytoin, total systemic clearance was a function of weight (WT), sex (SEX), and a series of binary categorical variables, as follows:

$$\begin{aligned} CL = {} & CL_0(WT)(1 + \theta_{APGR}APGR)(1 + \theta_{SEX}SEX) \\ & (1 + \theta_{GA}GA)(1 + \theta_{AGE}AGE) \end{aligned} \quad (13)$$

where $CL_0$ was baseline clearance; APGR was coded as 1 if the 5-minute Apgar score was $< 5$ and 0 otherwise; GA was coded as 1 if the patient was born at $< 34$ weeks gestation and 0 otherwise; AGE was coded as 1 if the sample was collected after the 6 $^{th}$ postnatal day and 0 otherwise; and $\theta_{APGR}$, $\theta_{SEX}$, $\theta_{GA}$, and $\theta_{AGE}$ represent the fractional increase or decrease associated with the indicator variable. Grasela and Donn[9] also made the first attempt at model validation.

The article by Grasela et al.[11] is important, because it showed how population models can be used to detect a potential drug-drug interaction from data collected during a phase I clinical trial. Noncompartmental analysis of 10 patients with intensive pharmacokinetic sampling and population analysis of 28 subjects with sparse sampling both detected a drug interaction between imipramine and alprazolam clearance. A better fit was obtained under the population model when imipramine clearance was modeled as a function of alprazolam concentration at the time of sampling than when imipramine clearance was modeled using a dichotomous covariate equal to 1 if the patient was taking both alprazolam and imipramine and 0 if the patient was taking only imipramine. Thus, the population approach provided a more mechanistic interpretation as to the nature of the interaction than the noncompartmental approach. If these results could be extrapolated to other settings, then this report sets the stage for drug companies using population methods during drug development.

### *Model Building Strategies*

Early reports of population models did not elaborate on how particular covariates were chosen and included with particular pharmacokinetic parameters (eg, see Refs.[7,9]). These reports simply presented the model including covariates. Testing for whether inclusion of a covariate resulted in an improvement in the goodness of fit requires at least 2 models: one with and one without the covariate of interest. Hence, for a model with a modest number of parameters, testing of covariates can become prohibitive, especially in the early 1980s, when personal computers were a luxury, and the speed of mainframe computers was slower than most personal computers today. One of the common practices today is to fit a model without any covariates, called the base or structural model, and then to determine each subject's pharmacokinetic parameters using Bayesian estimation. From these empirical Bayes estimates (EBEs) some type of screening is done, usually regression-based, to filter the number of covariates that need to be tested using NONMEM, thereby reducing the number of models that are evaluated, as well as the total run times. A modification of this is to regress the residuals or partial residuals, instead of the EBEs, against the covariate of interest (see[11] for a comparison of some of these methods).

Maitre et al[12] were the first to use correlation analysis of the EBE for a pharmacokinetic parameter and a particular covariate whereby only covariates showing a significant correlation were passed through the filter. Using this approach with midazolam, they found that among 11 covariates, only body weight and clearance, body weight and central volume, liver disease and clearance, and age and clearance showed a significant correlation. These covariates were then subsequently used in the NONMEM model.

Mandema et al[13] expanded this approach and studied the use of generalized additive models[14] as a stepwise regression-based covariate screen, instead of simple correlations. They also discussed options for when the covariates are time-varying and when it is not possible to obtain a single time-independent estimate of a pharmacokinetic parameter, that is, when clearance changes over time. This article also discusses the importance of the quality of the EBEs and how they "shrink" toward the population mean when there are few observations per subject or when the residual error is large compared with between-subject variability.

Bruno et al[15] combined the various techniques that had been reported to that time and developed a state-of-the-art population analysis of docetaxel in patients with solid tumors. In their analysis, data from 2 phase I studies and 22 phase II studies involving 547 patients were pooled and then split to a model development data set with 280 patients and a validation data set with 267 patients. The base model was a 3-compartment model. The EBEs were determined for the random effects, and a total of 18 covariates were screened using graphical inspection, multiple linear regression, and generalized additive models. Those covariates that were deemed important during the screening process were then tested in NONMEM for statistical significance using the likelihood ratio test (LRT). The model was then "validated" using the validation data set. The 2 data sets were then combined to refine the population model estimates. In the end, docetaxel clearance was dependent on body surface area, $\alpha_1$-acid glycoprotein concentration, and hepatic function. This report still remains as one of the most frequently cited references for model development in population analyses.

Before the report by Karlsson and Sheiner,[16] 2 sources of variability were modeled: between-subject variability and residual variability. For example, the ith patient's clearance ($CL_i$) may be modeled as follows:

$$CL_i = \theta_0 \exp(\eta_i) \qquad (14)$$

where $\theta_0$ is the population mean and $\eta_i$ represents the deviation for the ith subject from the population mean on a log-scale. However, for many reasons, a pharmacokinetic parameter may change from occasion to occasion. For instance, a change in apparent oral clearance may be attributable to changes in bioavailability from occasion to occasion resulting in differential drug exposure across occasions. Karlson and Sheiner[16] showed how between-subject variability may be additionally refined into interindividual variability and interoccasion variability, so, for example, equation 14 may be written as follows:

$$CL_i = \theta_0 \exp(\eta_i + \kappa_i) \qquad (15)$$

where $\kappa_i$ is the ith occasion under the assumption that $\kappa \sim N(0, \omega_\kappa^2)$. Hence, between-subject variability may be decomposed into interindividual variability and interoccasion variability (IOV). Using simulated data, they showed that not including IOV in the model when it is truly present can lead to appreciable bias in the structural model parameter estimates, but which parameter is affected and to what extent depends on many different variables with no clear pattern emerging. Ignoring IOV always inflated the residual variance estimate, whereas the bias in the estimate of BSV was negligible when BSV was large relative to IOV. On the other hand, when IOV was large relative to BSV, then biases in the estimates of BSV were large. Others have used this approach since then to model different sources of variability, such as between-study variability.[17]

The article by Wade et al[18] caused quite a stir and is disconcerting for a number of reasons. They simulated data from a 2-compartment model having an approximate $\alpha$-half-life and $\beta$-half-life of 35 minutes and 36 hours, respectively, with samples collected under 4 different time ranges: 3 to 120 hours, 6 to 120 hours, 12 to 120 hours, and 48 to 120 hours. They then fit a 1-compartment and 2-compartment model to the data where clearance and volume of distribution were treated as uncorrelated random effects. The frequency of selecting the 2-compartment model was much larger than expected, even with samples in the 48-to-120-hour group, which is >80 half-lives after the end of the distribution phase. Furthermore, the frequency of choosing the 2-compartment model changed depending on whether or not a covariate was included in the base model. When clearance was a function of a dichotomous covariate representing the effect of a patient subgroup, like a drug-drug interaction, the frequency of detecting the covariate was always higher under the 2-compartment model than the 1-compartment model. Despite the fact that the model used to simulate the data did not contain a covariance term, only when a covariance term was added between the clearance and volume of distribution did the frequency of choosing the 1-compartment model fall in line with expectations. What this article illustrated was that the choice of the covariate and structural model was dependent on the choice of the covariance model and that the most complex covariate or variance models could occur with the simplest structural model. Simply put, as the title states, an interaction exists among the structural, covariate, and covariance models in choosing the best model, which may change depending on how the submodels are formulated.

Wade et al[18] then recommended the following steps for model development, a process that has held to this day. First, the structural model without covariates should be chosen. Second, the influence of the covariates on all of the model parameters in the structural model should be explored. Third, if before covariate model development a more complex structural model seems just as likely, then the covariate model built with the simpler structural model

should be tested on the more-complex structural model. Rejected covariates under the more-simple model should then be reevaluated. Lastly, superfluous covariates can then be removed by stepwise deletion.

Missing from the article by Wade et al[18] is how to handle the covariance matrix. Should an unstructured or block covariance always be used and then reduced, or should a diagonal covariance be used and then tweaked at some later time? Most texts dealing with linear mixed effects models[19] recommend that an overparameterized covariance matrix be used, because whereas an overparameterized covariance may lead to poor estimation of the SEs, an underparameterized covariance results in biased inference about the fixed effects. This reasoning may apply to nonlinear mixed effects models as well.

A type I error is committed when the null hypothesis of no difference between groups is rejected when in fact it should not be. In other words, the null hypothesis assumes that there is no difference between groups; a type I error is made when it is declared that a difference between groups exists when in fact the difference does not exist. In the case of nonlinear mixed-effects modeling, a type I error is committed when a covariate is added to a model when it should not be. Wahlby et al[20,21] published as series of papers studying the type I error rate for FO approximation, conditional estimation methods for covariate models, and the addition of variance terms in a model. The articles are quite comprehensive, and it will be difficult to summarize their results in a few sentences. In general, FO approximation resulted in a higher type I error rate than conditional estimation methods, and the error rate was influenced by many factors, including sample size, the number of observations per subject, sampling scheme, and magnitude of residual error. FO conditional estimation with interaction (FOCE-I) resulted in type I error rates near their nominal values under most conditions.

The articles also introduce the use of covariate randomization to determine the change in objective function value needed for significance under the null hypothesis of no covariate effect. With this method, the covariate of interest is randomized (permutated) across subjects, thereby breaking any covariate-pharmacokinetic parameter relationship. The objective function is determined using the model that contains the covariate of interest, and the change in objective function value ($\Delta$OFV) relative to the same model without the covariate of interest is computed. This process is repeated many times such that a distribution of $\Delta$OFVs is obtained. The $\Delta$OFVs are ranked, and the critical value for significance is determined from the percentiles, that is, for 5% statistical significance the fifth percentile is used. They showed that compared with a theoretical value of 3.84 for 0.05 level significance using the LRT with 1 degree of freedom, the actual $\Delta$OFV under FO approximation ranged from ~13 to 17, depending on the covariate studied, but was approximately 4 using FOCE-I. Also, buried in the article is a method for developing a generalized least-squares approach to parameter estimation (Table 2 in Ref. 20).

In their second article, Wahlby et al[21] examined the influence of a covariate on a variance component, which does not occur that often but is still of interest. More importantly, however, are their other simulations. They also compared the type I error rate using the LRT estimated using FO approximation and FOCE-I to determine whether inclusion of a covariance term between pharmacokinetic parameters resulted in a significant improvement in the goodness of fit. They found that the type I error rates were near nominal values when the residual error was low but that higher residual variability resulted in higher type I error rates. Lastly, they studied the impact of including a false variance term on a pharmacokinetic parameter when one was not needed. The type I error rates were below expected using FOCE-I and were not influenced by number of observations per subject, number of individuals, or residual error magnitude.

All of the tests that Wahlby et al[20,21] examined were based on the LRT. Stram and Lee[22] have shown that the LRT is not valid when the alternative hypothesis falls on the boundary of the null hypothesis, that is, the distribution of the LRT is no longer $\chi^2$. Instead, the distribution becomes a mixture of $\chi^2$ distributions, and in the special case where a single variance component is being tested, the mixture is 50:50 of $\chi^2$ random variables with 1 and 2 degrees of freedom. Hence, instead of a value of 3.84 or 6.64 for 0.05 or 0.01 levels of significance, respectively, the critical values become 5.14 and 8.27, respectively. Pinheiro and Bates[23] used Monte Carlo simulation to examine this issue and came up with a ratio of 0.65:0.35 instead of 50:50. Interestingly, this boundary problem does not become an issue with covariance terms, because a covariance does not have a zero boundary; a covariance can fall anywhere on the real line. So when Wahlby et al[21] found the type I error rate for covariance terms to be near nominal values but type I error rates for variance terms to be too low, these results were entirely consistent with the results of Stram and Lee[22] but allowed the results of Stram and Lee to extend to the nonlinear mixed-effects model case.

### *Experimental Design*

Early PopPK models used data that were usually collected as part of a therapeutic drug monitoring program. Later, the pharmaceutical industry recognized the utility of PopPK in their analysis of data collected during phase III, which was typically considered too sparse to model at an individual level. It was recognized, however, that not all

experimental designs are equal. A better experimental design can result in more efficient parameter estimates, that is, less bias and smaller SEs. For example, at the simplest level, the optimum design for modeling a drug that shows monophasic elimination kinetics after intravenous administration is to collect samples as early as possible and as late as possible so that the length of time between samples is as long as possible. It is easy to envision that a design where the second sample is collected shortly after the first sample will not lead to a very good estimate of the rate of elimination. Hence, naïve collection of samples may result in biased and imprecise estimates.

In light of these observations, Aarons et al[24] present a short overview of a 1995 meeting held in Brussels to discuss issues related to experimental design in PopPK studies in drug development. It was recognized that most PopPK studies are done using sparse data collected during phase II and III where the purpose of such studies is not to characterize the pharmacokinetics of the drug but to demonstrate evidence of or confirm efficacy. PopPK is a secondary objective. Hence, whatever experimental design recommendations could be made would have to be done within the confines of the study, for example, samples cannot be collected at night during an outpatient phase III study, because no physician offices are open during those hours to collect the samples. A number of issues were addressed, including current practices, logistic issues, covariate assessment, protocol design, and the importance of communication. No specific conclusions were drawn, because too often these issues have to be addressed on a case-by-case basis. Nevertheless, the article presents an excellent overview of the issues involved in collecting and analyzing data from late-phase clinical trials.

Kowalski and Hutmacher[25] report on the development of a new drug wherein pharmacokinetic analysis from data-rich, intensive within-subject sampling collected during phase I showed that the drug had biexponential kinetics. However, it was believed that the sampling times for the proposed phase III study would not support such a model and that only a 1-compartment model could be supported. The authors then used clinical trial simulation to fit a 1-compartment model to data simulated from a 2-compartment model and to estimate the degree of bias in the resulting model parameters, assuming that the pharmacokinetics in phase III behaved the same as in phase I. A second simulation was used to assess the sample size needed to detect a 40% reduction in clearance in patients in a subpopulation compared with the main population with at least 90% power. The results showed that fitting a less-complex model resulted in unbiased estimation of the population means, although volume of distribution under the 1-compartment model really represented volume of distribution at steady-state under the 2-compartment model. The estimates of between-subject and residual variability were substantially biased, however. They also showed that under the specified conditions, if 5% of patients were in the subpopulation, there was an 86% chance of detecting the subpopulation given 150 subjects total and 94% given 225 subjects total. These results need to be tempered by the observation that the type I error rate was considerably inflated and higher than the nominal level of 0.05. To adjust for the inflation in the type I error rate, the critical value (fifth upper percentile) from the simulated null distribution of the test statistic (ΔOFV between the models with and without the covariate for subpopulation differences in apparent oral clearance simulated under the null model of no covariate effect) was used to downward adjust the estimates of power to 73% for 150 subjects and 84% for 225 subjects. Based on these adjusted estimates of power, the population pharmacokinetic substudy was conducted with a sample size of 225 subjects. In conversations with K. Kowalski (April, 2005) after the conclusion of the study, he indicated that the population estimate of apparent oral clearance for this patient population was within 5% of the healthy volunteer estimate and that the covariate effects included in the final model for this patient study were very much consistent with findings from healthy volunteer studies. This article nicely illustrates how a PopPK model can be applied to answer questions about future studies through the use of computer simulation.

Lee[26] used simulation to examine how degree of compliance, number of samples per subject, choice of sampling time, total number of subjects, and inclusion of individuals with data-rich sample collection interact to detect a 30% increase in clearance in a subpopulation in a sparsely design phase III study. Many conclusions were reported, but some of the more useful ones are as follows: (1) taking 3 samples resulted in greater power than taking only 2 samples, (2) collecting only trough samples resulted in poor power and could not accurately estimate the magnitude of difference between the main population and the subpopulation, (3) adding subjects with complete pharmacokinetic profiles to sparsely sampled subjects did not improve the estimate of the clearance difference between the groups and actually increased the type I error rate for detecting the subpopulation, and (4) the clearance difference was poorly estimated if the timing of a dosing event was missing. This article additionally illustrates the use of simulation as a tool to help design PopPK studies.

Recently, studies using optimal design in PopPK have received increased attention. The mathematics behind this algorithm is quite complicated (involving minimizing the negative determinant of the Fisher Information Matrix, of which the inversion results in the variance-covariance matrix of the model parameter estimates and the square root of the diagonal elements of the variance-covariance

matrix leads to the SE of the model parameter estimates) but relates the finding to a set of sampling times meeting user-defined constraints (eg, no samples collected between the periods of 6:00 PM and 6:00 AM in the case of a study done in an outpatient setting) that will optimally minimize the SEs of the estimable parameters in the model. Although of great interest, the use of optimal designs in practice has had limited experience. The reader is referred elsewhere[27,28] for more details.

### *Model Validation*

All models are based on assumptions. In ordinary least-squares, the assumptions are that the residuals are independent and identically distributed with zero mean and constant variance. Models that deviate from the underlying assumptions may have unstable or biased parameter estimates. PopPK also has many assumptions, such as adequacy of the FO approximation, error-free sampling times, and so forth. Karlsson et al[29] present a comprehensive list of assumptions, falling into the following categories, that should be tested: the estimation algorithm, data quality, structural model, covariate model, statistical models, and general aspects related to modeling. Not all of these assumptions are usually tested and confirmed, but many should be, and not just on the final model. For example, if the FO approximation is inadequate, it would be tragic to find this out at the end of the model development process, because the entire process then becomes questionable.

Bonate[30] showed that model parameter estimates may be unstable when correlated covariates ($\rho > 0.5$) are entered into a model simultaneously, similar to the collinearity problem that arises in linear models, and that instability increases as the degree of correlation increases. So, for example, a modeler would not want to use weight and body surface area as covariates to predict clearance, because weight and body surface area tend to be highly correlated. Collinearity can also arise when the correlated covariates are on different parameters, so if age is on clearance, weight is on central volume, and age and weight are highly correlated, then all of the parameter estimates in the model have the potential to be unstable. Various indices are also presented that quantify the degree of collinearity.

Hartford and Davidian[31] showed what happens to model parameter estimates when the assumptions of the model are violated. A large portion of the article is a technical discussion of the differences in estimation algorithms and the assumptions each algorithm makes. A key assumption is that the random effects ($\eta$) are normally distributed if clearance is modeled as $\theta\exp(\eta)$. Using simulation and SAS NLINMIX macro (SAS Institute, Cary, NC), which is no longer supported by SAS and has been supplanted by the NLMIXED procedure with version 8.0 and higher, they studied the impact on the parameter estimates when the random effects were not normally distributed, but were heavy tailed normal, a mixture of normal distributions, asymmetric, or bimodal symmetric in distribution. When the clearance model is correctly specified and the only violation is the assumption of the random-effects distribution, FO approximation has a very high rate of convergence, whereas conditional methods are less so. Unbiased estimation of the population means, regardless of the estimation method, was not affected to any significant extent when the random effects were not normally distributed. The estimation of the variances, however, was sensitive to the distribution of the random effects, but the estimation bias tended to be acceptable as long as the distribution was unimodal.

Ette et al[32] present an excellent overview of the current state of model validation in PopPK. Model validation is, unfortunately, one of the great misnomers in our field. No model can ever truly be validated under all of the conditions. Instead, the authors use the term "model appropriateness" to indicate that the model is validated for the specific purpose in which it was intended and that purpose is clearly stated a priori. They then illustrate the concepts in the article using a PopPK analysis of 5-fluorocytosine in patients with infection because of *Cryptococcus neoformans* and Candida species.

A central point is that model validation needs to be designed to support the intended use of the model. There are many approaches available to test a model, but the selection of test(s) is dependent on what aspects of the model require evaluation. Model validation should be designed prospectively and clearly stated in a data analysis plan. Prospective validation increases the credibility of the results with reviewers and the intended audience. The reviewer is referred to the Food and Drug Administration Guidance to Industry on Population Pharmacokinetics (which is discussed later in the article) for details on these data analysis plans.

### *Case Studies*

Enoxaparin (Lovenox) is a low molecular weight heparin that cannot easily be measured by analytical means. Instead, enoxaparin is indirectly estimated by factor Xa concentrations, which are thought to reflect unbound enoxaparin concentrations. Bruno et al[33] presented a PopPK analysis of factor Xa (or equivalently enoxaparin) in patients with unstable angina and non-ST-segment elevation myocardial infarction. Clearance was found to be a function of age, weight, creatinine clearance, and patient sex. The pharmacokinetic model was then used to predict area under the curve at steady state using the EBEs for

clearance under the final population model. Univariate and multivariate logistic regression was then used to relate area under the curve at steady state and dose, as measures of drug exposure, to the probability of experiencing major hemorrhage or any hemorrhage (which includes major hemorrhage) during the study. Other patient covariates, such as age, sex, weight, and platelet count, were also examined in the model. The pharmacokinetic model was then combined with the pharmacodynamic model to predict the risk of hemorrhagic events. This article is an excellent example of PopPK-PD for a number of reasons. Model development and validation is of sufficient detail and clearly described. The application of the model is also evident and useful, rather than the standard fare of just presenting the pharmacokinetics. Unfortunately, this information has failed to be included on the most recent package insert for the product.

Rajagopalan and Gastonguay[34] present a nice example of validation methods in their analysis of ciprofloxacin pharmacokinetics in pediatric patients having various infections. Using the parametric bootstrap, nonparametric bootstrap, leverage analysis, and cross-validation, the authors evaluated the performance and stability of the model. Using the results of the model, they then develop a dosing scheme to help clinicians dose pediatric patients with complicated urinary tract infections. This article also presents the issue of weight in modeling pediatric data. Weight is frequently a covariate in pediatric studies that is modeled using a power function, for example:

$$\mathrm{CL} = \theta_1(\mathrm{Wgt})^{\theta_2}\exp(\eta). \qquad (16)$$

However, one camp of PopPK modelers believes that the exponent in the model, $\theta_2$, should be empirically estimated based on the data so that the best possible fit is obtained. Another camp believes that based on allometric theory, $\theta_2$ should be a priori fixed to 0.75 for clearance terms and 1.0 for volume terms.[35] There is no consensus for which method should be preferred.

Grasmader et al[36] studied the PopPK of mirtazapine, a new antidepressant, in 49 patients with a diagnosis of clinical depression. What makes this paper notable are 2 things. One, patients were genotyped for isozymes of cytochrome (CYP) 2D6, 2C9, and 2C19. In the base model without covariates, the random effects for clearance ($\eta_{CL}$) showed a distinct bimodal distribution. When a categorical variable was added to clearance assigning patients a value of 2 if they were a CYP 2D6-extensive metabolizer or 1 if they were a poor or intermediate metabolizer, the bimodal distribution became unimodal and approximately normally distributed. This illustrates nicely the impact of a covariate on model assumptions and how proper use of a covariate can result in a model having violated assumptions correcting itself. This article also shows their covariate model development process for clearance allowing a newcomer to get a better feel for how models are tested and either accepted or rejected. One problem with this article is that how missing genotype data were handled is not adequately discussed.

### *Review Articles*

A number of useful review articles[37-39] have been written on population pharmacokinetics. Beal and Sheiner[37] present an overview of the statistical foundation and development of the FO approximation in a readable format and compare this with the 2-stage method for obtaining population estimates. Sheiner and Ludden[38] present a nontechnical overview of the PopPK, including a review of the methods, an example, and a summary of the results obtained with population methods. Although the article is a little dated, being published more than a decade ago, it still is useful in that it provides the reader with a feel for what PopPK is capable of. Tett et al[39] provide an excellent nontechnical overview of PopPK that would be useful for someone who needs to know what PopPK is, its advantages and disadvantages, and some of its uses. This review would be helpful for clinical research associates, project team leaders, physicians, or nontechnical members of a project team in a pharmaceutical company that are involved with a clinical trial in which PopPK is a component.

Two books have been written on the topic of nonlinear mixed effects models, but neither of these books are geared toward the pharmacokineticist; rather, they are written by statisticians for a mostly statistical audience and, as such, are difficult to read for a newcomer to the field. Davidian and Giltinan[40] is the more user-friendly of the two and is a must-read once the basics are understood. Pinheiro and Bates[23] is geared exclusively toward the computer language S and the software system *S*-Plus (Insightful Corp, Seattle, WA). Neither of these books should be read at the outset of learning the material. The book by Verbeke and Molenberghs,[41] although not dealing with PopPK, deals with linear mixed effects models in general. A good understanding of linear mixed-effects models can only lead to a better understanding of nonlinear mixed-effects models, just like a good understanding of linear regression leads to a better understanding of nonlinear regression. They published a similar book in 1997 that includes all of the material in the later book, less some more complex material, but it is more SAS-oriented than the later book.[19]

### *Regulatory Documents*

Beginning with Carl Peck and later Janet Woodcock as Directors for the Center for Drug Evaluation and Research

at the Food and Drug Administration (FDA), a number of prominent experts in the field of PopPK were brought into the agency through the formation of a dedicated pharmacometrics group that would be used to analyze PopPK data submitted by pharmaceutical companies in support of a new drug application. In 1999, the FDA issued a guidance to industry on PopPK.[42] The purpose of this was to make recommendations to industry on the design, analysis, and presentation of PopPK analyses made by industry. The guidance is actually a very good synthesis of simulation studies using the various estimation algorithms regarding experimental design issues, as well as some practical and logistic recommendations regarding design and execution of PopPK analyses. A key component of the FDA guidance is the concept of a "study protocol," which is more commonly referred to by the International Conference on Harmonisation as a "statistical analysis plan" or "data analysis plan." The study protocol should detail before the start of the analysis the objectives of the analysis and how the data will be analyzed, the idea being that the results will be more credible than if the analysis were after having an initial "look" at the data. Although laudable, study protocols in practice only provide a rough measure of credibility, because, in contrast to phase III studies where the analysis is locked into place with the null hypothesis carefully defined a priori, PopPK is still an exercise in exploratory analysis and, as such, it is impossible to plan every contingency. Nevertheless, detailing the objectives of the analysis, how missing data and outliers will be handled, and other general details is useful. The FDA guidance also provides guidelines on how the PopPK report should be written, which I could imagine was a horrible hodge-podge from different companies before publication of the FDA report.

## SUMMARY

Granted, the author has recommended a lot of papers to be read. The field has shown great progress since its inception, and continues to show greater and greater acceptance by other scientists and regulatory reviewers, as its concepts and terms become more understood, as its uses and limitations become more realized, and as its successes continue to be reported. However, this field is not easily understood, even for those with a firm statistical and mathematical foundation. Lynne Cox, the long-distance swimmer who broke the English Channel record at age 15, who swam the Bering Strait, and who swam the Arctic Ocean once said that "as long as you hang in there and keep going, you have a chance at succeeding. Once you give up, you're done." Learning PopPK is not easy, and understanding it can be a long, slow process, but it can be very rewarding once understood and applied.

## ACKNOWLEDGMENTS

The author thanks Ken Kowalsky for his helpful comments on this article.

## REFERENCES

1. Sheiner LB, Rosenberg B, Marathe V. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm.* 1977;5:445-479.

2. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm.* 1981;9:635-651.

3. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine pharmacokinetic data. *J Pharmacokinet Biopharm.* 1982;11:303-319.

4. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm.* 1980;8:553-571.

5. Peck CC, Beal SL, Sheiner LB, Nichols AI. Extended least squares nonlinear regression: a possible solution to the "choice of weights" problem in analysis of individual pharmacokinetic data. *J Pharmacokinet Biopharm.* 1984;12:545-558.

6. Sheiner LB, Beal SL. Pharmacokinetic parameter estimates from several least squares procedures: superiority of extended least squares. *J Pharmacokinet Biopharm.* 1985;13:185-201.

7. Vozeh S, Katz G, Steiner V, Follath F. Population pharmacokinetic parameters in patients treated with oral mexiletine. *Eur J Clin Pharmacol.* 1982;23:445-451.

8. Grasela TH, Sheiner LB. Population pharmacokinetics of procainamide from routine clinical data. *Clin Pharmacokinet.* 1984;9:545-554.

9. Grasela T, Donn SM. Neonatal population pharmacokinetics for phenobarbital derived from routine clinical data. *Dev Pharmacol Ther.* 1985;8:374-383.

10. Grasela TH, Antal EJ, Ereshefsky L, Wells BG, Evans RL Smith RB. An evaluation of population pharmacokinetics in therapeutic trials. Part II. Detection of a drug-drug interaction. *Clin Pharmacol Ther.* 1987;42:433-441.

11. Wahlby U, Jonsson EN, Karlsson MO. Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. *AAPS PharmSci.* 2002;4:E27.

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

13. Mandema J, Verotta D, Sheiner LB. Building population pharmacokinetic-pharmacodynamic models. I. Models for covariate effects. *J Pharmacokinet Biopharm.* 1992;20:511-528.

14. Hastie TJ, Tibshirani RJ. *Generalized Additive Models.* New York: Chapman and Hall, 1990.

15. Bruno R, Vivier N, Vergniol JC, DePhillips SL, Montay G, Sheiner LB. A population model for docetaxel (Taxotere®): model building and validation. *J Pharmacokinet Biopharm.* 1996;24:153-172.

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

17. Laporte-Simitsidis S, Girard P, Mismetti P, Chabaud S, Decousus H, Boissel JP. Inter-study variability in population pharmacokinetic analysis: when and how to estimate it? *J Pharm Sci.* 2000;89:155-166.

18. Wade JR, Beal SL, Sambol NC. Interaction between structural, statistical, and covariate models in population pharmacokinetic analysis. *J Pharmacokinet Biopharm.* 1994;22:165-176.

19. Verbeke G, Molenberghs G. *Linear Mixed Models in Practice: A SAS-Oriented Approach.* New York: Springer Verlag; 1997.

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

21. Wahlby U, Bouw MR, Jonsson EN, Karlsson MO. Assessment of type I error rates for statistical sub-model in NONMEM. *J Pharmacokinet Pharmacodyn.* 2002;29:251-269.

22. Stram DO, Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics.* 1994;50:1171-1177.

23. Pinheiro JC, Bates DM. *Mixed-Effect Models in S and S-Plus.* New York: Springer Verlag; 2000.

24. Aarons L, Balant LP, Mentre F, Morselli PL, Rowland M Steimer J-L, Vozeh S. Practical experience and issues in designing and performing population pharmacokinetic/pharmacodynamic studies. *Eur J Clin Pharmacol.* 1996;49:251-254.

25. Kowalski K, Hutmacher MM. Design evaluation for a population pharmacokinetic study using clinical trial simulation: a case study. *Stat Med.* 2001;20:75-91.

26. Lee PID. Design and power of a population pharmacokinetic study. *Pharm Res.* 2001;18:75-82.

27. Green D, Duffull SB. Prospective evaluation of a D-optimal designed population pharmacokinetic study. *J Pharmacokinet Pharmacodyn.* 2003;30:145-161.

28. Duffull SB, Retout S, Mentre F. The use of simulated annealing for finding optimal population designs. *Comput Methods Programs Biomed.* 2002;69:25-35.

29. Karlsson MO, Jonsson EN, Wiltse CG, Wade JR. Assumption testing in population pharmacokinetic models: illustrated with an analysis of moxonidine data from congestive heart failure patients. *J Pharmacokinet Biopharm.* 1998;26:207-246.

30. Bonate PL. The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharm Res.* 1999;16:709-717.

31. Hartford A, Davidian M. Consequences of misspecifying assumptions in nonlinear mixed effects models. *Comp Stat Data Anal.* 2000;34:139-164.

32. Ette EI, Williams PJ, Kim YH, Lane JR, Liu M-J, Capparelli EV. Model appropriateness and pharmacokinetic modeling. *J Clin Pharmacol.* 2003;43:610-623.

33. Bruno R, Baille P, Retout S, Vivier N, Veyrat-Follet C Sanderlink G-J, Becker R, Antman EM. Population pharmacokinetics and pharmacodynamics of enoxaparin in unstable angina and non-ST-segment elevation myocardial infarction. *Br J Clin Pharmacol.* 2003;56:407-414.

34. Rajagopalan P, Gastonguay MR. Population pharmacokinetics of ciprofloxacin in pediatric patients. *J Clin Pharmacol.* 2003;43:698-710.

35. Holford NHG. A size standard for pharmacokinetics. *Clin Pharmacokinet.* 1996;30:329-332.

36. Grasmader K, Verwohlt PL, Kuhn K-U, Dragicevic A, von Widdern O, Zobel A, Hiemke C, Rietschel M, Maier W, Jaehde U, Rao ML. Population pharmacokinetic analysis of mirtazapine. *Eur J Clin Pharmacol.* 2004;60:473-480.

37. Beal SL, Sheiner LB. Methodology of population pharmacokinetics. In: Garrett ER, Hirtz JL, eds. *Drug Fate and Metabolism: Methods and Techniques, Vol. 5.* New York: Marcel Dekker, Inc; 1985;135-183.

38. Sheiner LB, Ludden TM. Population pharmacokinetics/ pharmacodynamics. *Annu Rev Pharmacol Toxicol.* 1992;32:185-209.

39. Tett S, Holford NHG, McLachlan AJ. Population pharmacokinetics and pharmacodynamics: an underutilized resource. *Drug Inf J.* 1998;32:693-710.

40. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measures Data.* New York: Chapman and Hall; 1995.

41. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer-Verlag; 2000.

42. US Department of Health and Human Services, Food and Drug Administration. Guidance for Industry. *Population Pharmacokinetics.* Washington, DC: Food and Drug Administration; 1999.