The AAPS Journal 2007; 9 (1) Article 7 (http://www.aapsj.org).

Themed Issue: Bioinformatics and Computational Advances in the Pharmaceutical Sciences
Guest Editor - Murali Ramanathan

A Survey of Population Analysis Methods and Software for Complex Pharmacokinetic and Pharmacodynamic Models with Examples

Submitted: November 9, 2006; Accepted: January 13, 2007; Published: March 2, 2007

Robert J. Bauer,[1] Serge Guzy,[1] and Chee Ng[2]

[1]Pharmacokinetics, Pharmacodynamics, and Bioinformatics, XOMA (US) LLC, Berkeley, CA
[2]Institute for Drug Development/Cancer Research and Therapy Center, San Antonio, TX

ABSTRACT

An overview is provided of the present population analysis methods and an assessment of which software packages are most appropriate for various PK/PD modeling problems. Four PK/PD example problems were solved using the programs NONMEM VI beta version, PDx-MCPEM, S-ADAPT, MONOLIX, and WinBUGS, informally assessed for reasonable accuracy and stability in analyzing these problems. Also, for each program we describe their general interface, ease of use, and abilities. We conclude with discussing which algorithms and software are most suitable for which types of PK/PD problems. NONMEM FO method is accurate and fast with 2-compartment models, if intra-individual and interindividual variances are small. The NONMEM FOCE method is slower than FO, but gives accurate population values regardless of size of intra- and interindividual errors. However, if data are very sparse, the NONMEM FOCE method can lead to inaccurate values, while the Laplace method can provide more accurate results. The exact EM methods (performed using S-ADAPT, PDx-MCPEM, and MONOLIX) have greater stability in analyzing complex PK/PD models, and can provide accurate results with sparse or rich data. MCPEM methods perform more slowly than NONMEM FOCE for simple models, but perform more quickly and stably than NONMEM FOCE for complex models. WinBUGS provides accurate assessments of the population parameters, standard errors and 95% confidence intervals for all examples. Like the MCPEM methods, WinBUGS's efficiency increases relative to NONMEM when solving the complex PK/PD models.

KEYWORDS: population, pharmacokinetics, pharmacodynamics, clinical, software, computation methods

INTRODUCTION

The area of pharmacokinetic (PK) and pharmacodynamic (PD) modeling has advanced considerably over the past several decades, as the desire to understand drug pharmacokinetics, distribution, and their biological actions at the dynamic and quantitative level has increased. The advantage of performing advanced PK/PD modeling is that it allows drug developers and clinicians to more accurately determine the dosing regimen that is most efficacious and cost effective, and also to aid in the design of pharmaceuticals with the desired properties. Providing an analysis environment that can be robust and versatile enough to deal with the complex numerical methods required to implement the PK/PD model and account for irregular dosing regimens that subjects are actually given is demanding. In addition to characterizing the interaction of drug and body at the individual level, one must also understand the range of PK and PD profiles one can expect from a population of subjects. Therefore, the program must also have suitable statistical capabilities of accounting for measurement error, and to account for the distribution of PK and PD parameters that are observed among patients. Solving the combination of the PK/PD structural, as well as the statistical, modeling demands is not straightforward, and has been among the greatest software challenges in the biological/medical field.

A comprehensive review has been recently published on the history of PK/PD modeling development.[1] These authors cover the general principles of modeling dynamical systems on which much of PK/PD modeling methods are based and describe some of the most widely used PK/PD modeling paradigms, such as linear and nonlinear dose/concentration-dependent effects, delayed effect, indirect actions, cell-trafficking models, feedback paradigms, and oscillatory phenomena.

Perhaps the most widely used and oldest of the population analysis programs available is NONMEM, developed in the early 1980s.[2] The historical development of statistical methods on population PK/PD analysis by Pillai et al[3] describes Lewis Sheiner's contribution to introducing nonlinear mixed-effects modeling (NLME) to the pharmaceutical sciences community. Together with Stuart Beal, he developed the NONMEM program, which used least squares methods originally developed for individual subject curve fitting, and extended it to population data analysis (first-order

Corresponding Author: Robert J. Bauer, XOMA (US) LLC, 2910 Seventh Street, Berkeley, CA, 94710. Tel: 510-204-7562; Fax: 510-841-7805; E-mail: bauer@xoma.com

method) by linearizing the otherwise daunting statistical problem.

Statistical comparisons of population methods and software have been explored in earlier reviews,[4-6] which covered the following methods: naive-pooled data analysis, 2-stage, global 2-stage, linearized expectation-maximization (EM) (iterative 2-stage), nonlinear mixed effects methods using first-order and Gaussian-Quadrature methods, nonparametric, and semi-parametric methods.

In the past 10 years, a series of new tools for PK/PD modeling and population analysis have become available to the pharmaceutical scientist. Thanks to ever-increasing computer power, additional population analysis methods have been developed for PK/PD modeling, such as the expectation-maximization methods that evaluate the problem without imposing linear approximation, addressing the bias sometimes introduced from the linearization methods in some of the traditional methods.[3,7,8] Often the results of these methods can be compared with those of the least-squares methods, as the objective function whose optimum is sought may not be identical, but may have a similar statistical basis. Full Bayesian techniques employing 3-stage hierarchical analysis are also becoming easier to implement with the greater computer power available. The program WinBUGS[9,10] uses not only a different approach in analyzing the data, but also a different statistical goal. In the WinBUGS program, not just a single best-fit population parameter set is provided, but rather a collection of several thousand population parameter sets is reported, clustered in proportion to their likelihood of representing the data. Descriptive statistics of the collection of representative parameters may then easily be obtained, such as mean, variance, covariance between any 2 parameters, and quantile ranges (such as 2.5%, median, 97.5% levels), and the mode could be extracted from a narrow range of values that occur most frequently, which is the parameter set with the greatest likelihood of representing the data. If the analysis is performed with very weak initial (called prior) information, the parameter set representing the mode of the analysis (equivalent to the mean, if parameters are symmetrically distributed) may be compared with the optimal parameter set from the 2- stage hierarchical maximum likelihood methods.

NONMEM continues to be the most widely used population analysis software, and particularly its first-order conditional estimation (FOCE) method is still useful as a method for many population analysis problems that have a high degree of intersubject variability or model nonlinearity.[11]

The comparisons between methods and software that have been described in the literature to date involved 1- or 2-compartment PK models or fairly simple analytical function descriptions of the data. The present paper is designed to provide an overview of the statistical basis of some of the present population analysis methods, their implementation in some of the software packages available, and an assessment of which software packages are most appropriate for various PK/PD modeling problems, including fairly complex ones requiring the solution of ordinary differential equations. We begin with the theoretical bases of the algorithms typically used in these software packages. Then, 4 PK/PD example problems are solved using the programs NONMEM,[12] PDx-MCPEM,[13] S-ADAPT,[14] MONOLIX,[15] and WinBUGS,[16] informally assessed for accuracy, ease of use, versatility, and stability in analyzing these problems. We conclude with discussing which algorithms and software are most suitable for which types of PK/PD problems.

Theoretical Bases of Methods

The PK/PD Model

All of the methods presented here rely on the PK/PD model that is designed or selected by the user to describe the data as a function of time. The model function we shall designate as *f*, and may be as simple as a single decaying exponential as a function of time *t*, for a 1-compartment model for single IV bolus dose D[4,17]:

$$f(t,k,V) = \frac{De^{-kt}}{V} \tag{1}$$

the result of which is to be compared against measured serum levels of a drug. The model parameters in this case are V (volume of distribution), and k, the rate constant of elimination. The purpose of a statistical analysis is to find the values of V and k that best represent the data. Or, the model may be as complex as a series of functions that are in turn related to a set of differential equations that represent biologically mechanistic-based mass transfer and elimination of compounds, and interaction of drug with a biological target, each function describing one of several possible data types and/or physiological compartments (such as serum level of parent compound, metabolite, peripheral compartments, levels of biological markers, and so forth). The differential equations must be numerically integrated during the statistical fitting process, and can lead to long computation times.

Thus, we may generally describe the model as

$$f(t_{ij}, \boldsymbol{\theta}) \tag{2}$$

where vector $\boldsymbol{\theta}$ is the set of model parameters (in the simple example, these would be V and k) whose values are sought to best represent the data of a subject i, and a specific function and time t is associated with a particular data point j and subject i.

Least Squares Methods of Estimation

The next step is to develop a statistical test that is small in value when the function $f(t_{ij}, \theta)$ represents the data closely for a given θ, and large when the function $f(t_{ij}, \theta)$ represents the data poorly. The ordinary least squares method uses the following test value to compare the function $f(t_{ij}, \theta)$ with the measured data y_{ij}[17]:

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - f(t_{ij}, \theta))^2 \quad (3)$$

called the objective function. When the set of $\theta = \hat{\theta}$ is found so that L is at its smallest value, the minimum of the objective function has been achieved, and the average distance of the data from the model function values has been reduced to the minimum, that is, the function $f(t_{ij}, \theta)$ at $\hat{\theta}$ fits the data most closely among all possible values of θ.

There is measurement or assay error in data collected, however, and some data are more precisely known than others. So, the distance of the function from data that are more precisely known should be given greater weight in the least squares function. For example, assay error often imparts a measurement error that is in direct proportion to the concentration, so the error is greater for higher concentrations than lower concentrations. As an example, the residual variance (the square of the error) could be described therefore by

$$g_{ij} = \sigma^2 f^2(t_{ij}, \theta) \quad (4)$$

where σ is the residual error coefficient. A further refinement in the least squares objective function may thus be obtained by incorporating this variable uncertainty in the measurement error as a weighting factor:

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - f(t_{ij}, \theta))^2 / g_{ij} \quad (5)$$

Thus, distances of function values from data with lower variance will be given greater weight to the overall sum of squares. Additional sources of this residual error include variations in a subject's disposition of drug throughout the sampling period, as well as inaccuracy of the model in describing the drug disposition (model mis-specification).

Maximum Likelihood Methods of Estimation

With the least squares method, an intuitive desire to obtain parameter values that most closely fit the data leads to the least squares objective function. No attention is paid to how the data are statistically distributed. With the maximum likelihood method, consideration is made about the statistical distribution of data y_{ij} about the model function $f(t_{ij}, \theta)$. For example, an assumption is often made that the measured data are normally distributed about its mean, so the maximum likelihood method seeks to find the set of $\hat{\theta}$ for which $f(t_{ij}, \hat{\theta})$ is the mean for each y_{ij}. The normal distribution of y_{ij} is thus expressed as

$$l(y_{ij} \mid \theta, \sigma) = \frac{1}{\sqrt{2\pi g_{ij}}} \exp\left(-\frac{(y_{ij} - f(t_{ij}, \theta))^2}{2 g_{ij}} \right) \quad (6)$$

It is convenient to express this as its –2 times the logarithm form (up to a constant term):

$$\frac{(y_{ij} - f(t_{ij}, \theta))^2}{g_{ij}} + \ln(g_{ij}) \quad (7)$$

The total twice negative log-likelihood for a set of independently distributed data among m subjects is expressed as the sum

$$L = \sum_{i=1}^{m} \sum_{j=1}^{m_i} \left[\frac{(y_{ij} - f(t_{ij}, \theta)^2}{g_{ij}} + \ln(g_{ij}) \right] \quad (8)$$

The parameter set $\boldsymbol{\theta}$ may then be varied until the objective function L is at its minimum (and thus maximum probability). This parameter estimate $\boldsymbol{\theta}$ at which L is minimum thus represents the "most likely" fit, that is, it fits the data more closely than all other possible $\boldsymbol{\theta}$, so that

$$L_{\hat{\boldsymbol{\theta}}} \le L_{\boldsymbol{\theta}} \textit{ for all } \boldsymbol{\theta} \quad (9)$$

Comparison with the least squares objective function shows that for constant g_{ij}, their minimum would be at identical values of $\boldsymbol{\theta}$[17].

Population Analysis

The above statistical methods are suitable for fitting data for individual subjects or for naïve pooled analysis of data from a collection of individuals. An expansion of the likelihood method is needed to take into consideration the intersubject variability of parameters $\boldsymbol{\theta}$ among the subjects, in addition to the within-inidividual error g_{ij} that we have already considered.

Often in PK/PD modeling analysis, the variability of the vector of parameters $\boldsymbol{\theta}$ among the population is also assumed to be normally distributed, or perhaps some simple transformation of the parameters is normally distributed (such as via logarithmic transformation), with population mean vector $\boldsymbol{\mu}$, and population variance matrix $\boldsymbol{\Omega}$. The probability density of a given θ is therefore, not including a constant term,

$$h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) = \frac{1}{\sqrt{|\boldsymbol{\Omega}|}} \exp[-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{\Omega}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu})] \quad (10)$$

where we shall designate $h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})$ as the parameter population density for $\boldsymbol{\theta}$, given $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$. Then, the joint probability density for some vector $\boldsymbol{\theta}$ and a set of data $\boldsymbol{y}_i$ for subject i is

$$p(\mathbf{y}_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma}) = l(\mathrm{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) \quad (11)$$

where

$$l(\mathbf{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) = \prod_{j=1}^{m_i} l(y_{ij} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) \tag{12}$$

is the individual observed data likelihood for subject *i*. For a particular subject, the data y_i is observed and therefore fixed throughout the analysis, whereas the parameter vector $\boldsymbol{\theta}$ describing the pattern in the data are unknown but imputable, based on the model and the observed data. It is therefore best to consider all possible values of $\boldsymbol{\theta}$, taking into consideration the probability of occurrence of each $\boldsymbol{\theta}$ for the particular population in question. To do so, we integrate the density over all possible $\boldsymbol{\theta}$, producing the following contribution of the objective function by subject *i*:

$$L_i = -2\log(p(\mathbf{y}_i \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma})) = -2\log(\int_{-\infty}^{+\infty} l(\mathbf{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) d\,\boldsymbol{\theta}) \tag{13}$$

which is the twice negative logarithm of the marginal density of the data y_i for subject *i*, up to a constant. The twice negative logarithm of the joint marginal density for all *m* subjects is then

$$L = -2\log(p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma})) = \sum_{i=1}^{m} L_i = -2\sum_{i=1}^{m} \log(\int_{-\infty}^{+\infty} l(\mathbf{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\sigma}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) d\,\boldsymbol{\theta}) \tag{14}$$

where *L* is now the total objective function.

Linearized Approximation Method of Analysis

To find the set of mean population parameters $\boldsymbol{\mu}$, population variance $\boldsymbol{\Omega}$, and residual error coefficients $\boldsymbol{\sigma}$ that best fits the data from *m* subjects, one maximizes the above marginal density of $\mathbf{y}$ with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, minimizing the resulting *L*. However, the integration steps needed to evaluate *L* are computationally expensive. The first-order conditional estimation method (FOCE,[11,18]) minimizes an alternative objective function:

$$L_N = \sum_{i=1}^{m} L_{iN} \tag{15}$$

where

$$\mathrm{L}_{iN} = -2\log(p(\mathbf{y}_i, \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma})) - \log|\hat{\mathbf{B}}_i| \tag{16}$$

is the approximate individual likelihood contribution of individual *i*, $\hat{\theta}_i$ is the mode of the joint density for each subject *i*, and $\hat{\mathrm{B}}_i$ is the first order approximation to the conditional variance matrix of the parameters over the joint density. Equation 16 represents the exact evaluation of Equation 13 only if $p(\boldsymbol{y}_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma})$ with respect to $\boldsymbol{\theta}$ is a normal distribution, which occurs if the model function $f(t_{ij}, \boldsymbol{\theta})$ is linear with respect to θ. Thus, Equation 16 is a reasonable approximation to Equation 13 to the extent that the density $p(y_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma})$ is approximately normal in $\boldsymbol{\theta}$. The minimization of Equation 16 is conceptually done by a 2-step process. First, initial values of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ are set, while the data from each individual are fitted by finding the minimum of L_{iN}, obtaining the parameter set $\hat{\theta}_i$ at the mode (maximum) of the distribution, and obtaining $\hat{\mathrm{B}}_i$ by standard partial derivative assessment at the mode. Once this is done for all subjects, the total L_N is calculated as simply the sum of L_{iN}, which serves as the population likelihood at that given $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$. A quasi-Newton search routine assesses *L* at various $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, and performs a gradient directed search for the set of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ that provide the optimal *L*. Thus, individual maximum likelihood analyses are performed within a larger scale maximization to find $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and σ. A somewhat more accurate assessment of the integral can by obtained if $\hat{\mathrm{B}}_i$ is replaced with a variance-covariance matrix that was derived from a second-order assessment of the information matrix under the individual's conditional parameter distribution, as is done with the Laplace method in NONMEM.[11,19]

An Even More Linearized Approximation Method: FO

Despite the linearization reducing the complexity of the integration step, FOCE still requires considerable computation time, although this has improved recently with ever-faster computers. To simplify the process further, one can combine the deviation of the data from the model function values within the individual (the residual error) with the deviation of the model function values between individuals (from the intersubject model parameter variation). This simplification is an approximation that is reasonable if the intersubject variability plus residual variability are not large, so that one can approximate all intra and interindividual deviations as a first order approximation. What results is an objective function of the form[2,4,12,17]:

$$L_{FO} = \sum_{i=1}^{m} [(\boldsymbol{y}_i - f(\mathrm{t}_i, \boldsymbol{\mu}))' \mathrm{C}_i^{-1} (\boldsymbol{y}_i - f(\mathrm{t}_i, \boldsymbol{\mu})) + \log(\det(\mathrm{C}_i))] \tag{17}$$

where

$$\mathrm{C}_i = \mathbf{G}_i \boldsymbol{\Omega} \mathbf{G}_i' + \mathrm{diag}(\mathrm{g}_i) \tag{18}$$

is the matrix modeling the contribution of the interindividual variance of the parameters and the residual variance of measurement error to the total variations observed in the data,

$$\mathbf{G}_i = \frac{\partial f(\mathrm{t}_i, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \tag{19}$$

and

$$\mathrm{diag}(g_i) \tag{20}$$

is a diagonal matrix with diagonal elements g_{ij}. Equation 17 can be minimized in place of Equation 15 or Equation 14.

The model function *f* and residual error function *g* are always evaluated at the population mean position $\boldsymbol{\mu}$, and an individual's best fit values are never determined during the population analysis. Rather, the effect of that person's deviations of its parameters on his data are always interpolated via the matrix **G**. By using the above approximation, a reasonable sense of what portion of the deviation of the data from the population mean value is due to residual error and what portion is due to interindividual variation can still be obtained. The time savings results from requiring only a minimization for $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, without having to first perform individual optimizations within the larger optimization. Thus while FOCE only linearizes the deviation at the residual error level, FO linearizes the deviation of data at the interindividual and residual error level. This FO method was the original population analysis algorithm available in the first version of NONMEM.[20]

If intra- plus interindividual deviations sum to be large, the FO can be inaccurate.[21,22] A 20% residual variability has been sited as causing inaccurate estimation of the parameters in several examples.[23,24] The degree of inaccuracy that large intra- plus interindividual deviations cause in the FO method also depends on the degree of statistical nonlinearity of the model with respect to the parameters. To avoid all inaccuracies arising from linearizing a problem, exact integrations under each subject's conditional density should be performed.

Return to More Exact Integration Methods

In recent years attempts have been made to increase the accuracy of integration under the individual's conditional distribution. The Gaussian-Quadrature method, and methods similar to it, have been implemented in order to achieve this.[6,19,25] These methods tend to require greater computation times than the linearized approximations.[26] The following section discusses combining integrations/expectations by Monte-Carlo methods with a straightforward method of maximization, and is showing promise for population analysis, especially for PK/PD problems. Monte Carlo methods have been found to be robust and are easy to program.[19]

The Expectation-Maximization (EM) Algorithm with Monte-Carlo Integration Methods for Obtaining Exact Integrals

The EM algorithm is a means by which the exact objective function *L* of equation 14 may be minimized. It can be shown that if the parameter population density $h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})$ is of the form of a multivariate normal distribution with respect to $\boldsymbol{\theta}$ (or some transformation of $\boldsymbol{\theta}$), then at the minimum of the objective function the following relationships are true[27]:

$$\boldsymbol{\mu} = \frac{1}{m}\sum_{i=1}^{m}\bar{\boldsymbol{\theta}}_i \tag{21}$$

$$\boldsymbol{\Omega} = \frac{1}{m}\sum_{i=1}^{m}\bar{\boldsymbol{\Omega}}_i \tag{22}$$

where

$$\bar{\boldsymbol{\theta}}_i = \int_{-\infty}^{\infty}\boldsymbol{\theta} z(\boldsymbol{\theta}\,|\,y_i,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\,\boldsymbol{\theta} \tag{23}$$

is the conditional mean $\boldsymbol{\theta}$ vector for subject *i*,

$$\bar{\boldsymbol{\Omega}}_i = \int_{-\infty}^{\infty}(\boldsymbol{\theta}-\boldsymbol{\mu})(\boldsymbol{\theta}-\boldsymbol{\mu})' z(\boldsymbol{\theta}\,|\,y_i,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\,\boldsymbol{\theta} \tag{24}$$

is the contribution to the population variance from each subject *i*, and

$$z(\boldsymbol{\theta}\,|\,y_i,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}) = \frac{l(\boldsymbol{y}_i\,|\,\boldsymbol{\theta},\boldsymbol{\sigma})h(\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Omega})}{\int_{-\infty}^{\infty} l(\boldsymbol{y}_i\,|\,\boldsymbol{\theta},\boldsymbol{\sigma})h(\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Omega})d\,\boldsymbol{\theta}} \tag{25}$$

is the conditional density of $\boldsymbol{\theta}$, given data y_i, and population parameters $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\sigma}$. An equivalent form of Equation 24 is:

$$\bar{\boldsymbol{\Omega}}_i = (\bar{\boldsymbol{\theta}}_i-\boldsymbol{\mu})(\bar{\boldsymbol{\theta}}_i-\boldsymbol{\mu})' + \bar{\mathbf{B}}_i \tag{26}$$

where

$$\bar{\mathbf{B}}_i = \int_{-\infty}^{\infty}(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}_i)' z(\boldsymbol{\theta}\,|\,y_i,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\,\boldsymbol{\theta} \tag{27}$$

is the conditional variance matrix of $\boldsymbol{\theta}$ for subject *i*.

The above equations suggest that one may maximize $p(y|\boldsymbol{\mu},\Omega)$ with respect to the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ by first evaluating the conditional mean $\bar{\theta}_i$ by Equation 23 and the conditional variance $\bar{B}_i$ by Equation 27 for each subject *i*, using fixed values of $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ (the expectation step), followed by evaluating updates to $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ using Equations 21, 26, and 22 (the maximization step[27]). The EM update Equations 21 and 22 are specific for a parameter density $h(\boldsymbol{\theta}|\mu,\boldsymbol{\Omega})$ that is multivariate normal, but are generally true for any data density $l(\boldsymbol{y}_i|\boldsymbol{\theta})$. An additional algorithm is required to update $\boldsymbol{\sigma}$.[8,28]

Additionally, we would like to evaluate the integrations given in Equations 23 and 27 more precisely while avoiding complicated integration algorithms. Monte-Carlo integration methods allow one to randomly sample over the entire space of $\boldsymbol{\theta}$, and then calculate a weighted average of the quantity of interest, converting the theoretical integration step into a practically evaluated summation step[8,29]:

$$\bar{\boldsymbol{\theta}}_i = \frac{\int_{-\infty}^{\infty}\boldsymbol{\theta} p(\boldsymbol{y}_i,\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\,\boldsymbol{\theta}}{\int_{-\infty}^{\infty} p(\boldsymbol{y}_i,\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\,\boldsymbol{\theta}} \Rightarrow \frac{\sum_{k=1}^{r}\boldsymbol{\theta}_{(k)} W(l(\boldsymbol{\theta}_{(k)}),h(\boldsymbol{\theta}_{(k)}))}{\sum_{k=1}^{r} W(l(\boldsymbol{\theta}_{(k)}),h(\boldsymbol{\theta}_{(k)}))} \tag{28}$$

Theoretical Expression Practical Evaluation

and

$$\bar{\mathbf{B}}_i = \frac{\int_{-\infty}^{\infty} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_i)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_i)' p(\boldsymbol{y}_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \sigma) d\,\boldsymbol{\theta}}{\int_{-\infty}^{\infty} p(\boldsymbol{y}_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \sigma) d\,\boldsymbol{\theta}}$$

Theoretical Expression

$$\Rightarrow \frac{\sum_{k=1}^{r} (\boldsymbol{\theta}_{(k)} - \bar{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_{(k)} - \bar{\boldsymbol{\theta}}_i)' W(l(\boldsymbol{\theta}_{(k)}), h(\boldsymbol{\theta}_{(k)}))}{\sum_{k=1}^{r} W(l(\boldsymbol{\theta}_{(k)}), h(\boldsymbol{\theta}_{(k)}))} \quad (29)$$

Practical Evaluation

where the weight w depends on the parameter density $h(\boldsymbol{\theta}_{(k)})$ (suppressing dependence on $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$), data density, $l(\boldsymbol{\theta}_{(k)})$ (suppressing dependence on y_i) and the method of Monte-Carlo used, for r randomly generated vectors of $\boldsymbol{\theta}_{(k)}$. Alternatively, samples may be accepted or rejected using a suitable testing function that depends on w, and the accepted samples would then be simply averaged.

Several Monte Carlo methods have been used in population PK/PD problems for evaluating the conditional means and variances, among them direct sampling,[8,29] importance sampling,[8] and stochastic approximation EM with importance sampling and Markov Chain Monte Carlo.[15,30]

Linearized Approximation EM or Iterative Two-Stage (ITS) Method

As mentioned earlier, the distribution of θ under the individual likelihood

$$p(\boldsymbol{y}_i, \boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \sigma) = l(\boldsymbol{y}_i \mid \boldsymbol{\theta}, \sigma) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) \quad (30)$$

is often non-normally distributed, the non-normality (and hence often statistical nonlinearity) imparted to it by the nonlinear PK/PD model function f. If the residual error is fairly small, and/or a rich collection of data are obtained for an individual, then the likelihood may often be approximated as a normal distribution with respect to $\boldsymbol{\theta}$. Normal distributions are characterized as symmetric about the mean, with the mean value equivalent to the most probable value, or the mode of the distribution. In this case the mode of $\boldsymbol{\theta}$ may be used in place of the more difficult to obtain mean of $\boldsymbol{\theta}$. As the mode of the individual density is easily obtained by finding the maximum of Equation 30, a suitable linearized approximation using the EM method would be to perform individual fits as is done in the FOCE method, to obtain $\hat{\theta}_i$ and $\hat{B}_i$, which serve as the expectation step, followed by the following maximization step:

$$\boldsymbol{\mu} = \frac{1}{m} \sum_{i=1}^{m} \hat{\boldsymbol{\theta}}_i \quad (31)$$

$$\boldsymbol{\Omega} = \frac{1}{m} \sum_{i=1}^{m} \hat{\boldsymbol{\Omega}}_i \quad (32)$$

This method is called the iterative 2-stage method (ITS), first suggested by Steimer et al,[31] popularized by others,[32,33] and implemented in Kinetica/P-PHARM[28,34,35], and Popkinetics.[36]

While the expectation step of ITS is equivalent to the linearized integration step of FOCE, the ITS method performs the simpler EM method to update $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, rather than the quasi-Newton search method of FOCE. The ITS algorithm therefore does not lead to the minimization of L_N, although it nearly minimizes it to the extent that the linearization does not unreasonably deviate from the true likelihood.

The ITS method approaches the minimum from even rather poor initial population parameter settings at a very rapid pace, and then slows down considerably in its convergence rate. When data are relatively rich (at least as many data points as there are parameters), the ITS method can yield results similar to NONMEM's FOCE method. However, when data are sparse (such as one data point per subject, and fitting a 1- or 2-parameter PK model), the ITS and FOCE methods yield different answers, both of which can be biased.[8,21,23] Their different answers result because of the different "maximization" step each method uses. We shall see this in example 1 below. Their bias results because both methods use a simplified, linearized method to perform the expectation step when the density $p(y_i,\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})$ with respect to $\boldsymbol{\theta}$ is highly non-normal.

Three-Stage Hierarchical/Bayesian Method

The EM and NONMEM FOCE methods are called 2-stage hierarchical maximum likelihood methods, the first stage being the individual likelihood of the data $l(y_i|\boldsymbol{\theta},\boldsymbol{\sigma})$, given a set of model parameters and residual error coefficients $\boldsymbol{\sigma}$ for that individual, and the second stage the likelihood of the set of model parameters, given some knowledge or assumptions of the distribution parameters among the population $h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})$, parameterized by the population means $\boldsymbol{\mu}$, and intersubject variance $\boldsymbol{\Omega}$. A third hierarchical stage may be introduced and incorporated into the total likelihood to take into consideration the uncertainty of the knowledge of the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, to the population distributions.[9,10,17] This consideration is particularly appropriate if one has empirical knowledge of μ, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, based on a previous population analysis which one would like to incorporate into a new analysis.[37,38]

In addition, Bayesian methods of analysis do not maximize the likelihood. Rather, a series of possible $\boldsymbol{\mu}$'s, $\boldsymbol{\Omega}$'s, and $\boldsymbol{\sigma}$'s are collected, with a frequency that is based on their likelihood of explaining the data using the following probability[10,17]:

$$p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma} / \boldsymbol{y}, \mathbf{q}, \mathbf{H}, \mathbf{W}, \boldsymbol{\tau}) = p(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\sigma} / \mathbf{q}, \mathbf{H}, \mathbf{W}, \boldsymbol{\tau}) / p(\boldsymbol{y} / \mathbf{q}, \mathbf{H}, \mathbf{W}, \boldsymbol{\tau}) \quad (33)$$

where

$$p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) = p(\boldsymbol{y}/\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})\pi(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) = \quad (34)$$

$$\prod_{i=1}^{m}\left\{\int_{-\infty}^{\infty} p(\boldsymbol{y}_i,\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})d\boldsymbol{\theta}\right\}\pi(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) = \quad (35)$$

$$\prod_{i=1}^{m}\left\{\int_{-\infty}^{\infty} /(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})d\boldsymbol{\theta}\right\}\pi(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma})/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) \quad (36)$$

So:

$$p(\boldsymbol{y}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) = \int p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})d\boldsymbol{\mu}d\boldsymbol{\Omega}d\boldsymbol{\sigma} \quad (37)$$

and therefore:

$$p(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\boldsymbol{y},\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau}) = \frac{\left\{\prod_{i=1}^{m}\int_{-\infty}^{\infty} /(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})d\boldsymbol{\theta}\right\}\pi(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})}{\int_{-\infty}^{+\infty}\left\{\prod_{i=1}^{m}\int_{-\infty}^{\infty} /(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})h(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Omega})d\boldsymbol{\theta}\right\}\pi(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})d\boldsymbol{\mu}d\boldsymbol{\Omega}d\boldsymbol{\sigma}} \quad (38)$$

The probability $\boldsymbol{\pi}(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})$ is the distribution of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\mu}$, based on prior knowledge/data analysis. Typically, the distribution of $\boldsymbol{\mu}$ is modeled as a normal distribution with mean $\mathbf{q}$, and variance $\mathbf{H}$. The $\mathbf{q}$ and $\mathbf{H}$ could be based on the population mean $\boldsymbol{\mu}$ and its standard error (not intersubject variance, but the uncertainty of knowledge of $\boldsymbol{\mu}$) from a previous analysis. The $\boldsymbol{\Omega}$ intersubject variance is modeled as a Wishart distribution with parameters $\mathbf{W}$, and the residual variance $\boldsymbol{\sigma}$ is modeled as a gamma distribution with parameter $\boldsymbol{\tau}$. A poor man's process may be conceived of by randomly selecting a particular $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ from their respective distributions, then randomly selecting the $\boldsymbol{\theta}$'s based on the $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and then evaluate $l(\boldsymbol{y}_i|\boldsymbol{\theta},\sigma)$ based on the $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$. Many parameters may be randomly selected, their $l(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})$ evaluated, and summed together for a given $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$. These sums of $l(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})$ represent empirical assessments of integration over all possible $\boldsymbol{\theta}$, and is therefore an empirical assessment of $p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})$. These subsums are then divided by the total sum over all randomly selected $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ where this total sum represents the empirical assessment of $p(\boldsymbol{y}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})$. What results is a collection of thousands of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, with their frequencies $p(\boldsymbol{y}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})$. From these, weighted averages of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ may be obtained, as well as their variances. Also, one may identify the maximum likelihood estimate of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ by choosing the set with the highest frequency. In this sense, a complete distribution profile of the population parameters is obtained, rather than a single estimate of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ at the maximum likelihood.

If no prior knowledge is available of the distribution of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$, say, from a previous analysis, so that their distribution $\boldsymbol{\pi}(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\sigma}/\mathbf{q},\mathbf{H},\mathbf{W},\boldsymbol{\tau})$ is unknown, one can set the $\mathbf{H}$ matrix to very large values, and τ to a very small value, to represent this lack of knowledge, and to prevent a constraint of the analysis based on untenable preconceived notions. This is called an uninformative prior, and the frequency distribution of $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ will be completely determined by the observed data $\boldsymbol{y}$.

In practice, if the direct sampling Monte Carlo approach is used in the manner described above, millions of random vectors of parameters at the $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\sigma}$ level must be selected in order to obtain enough samples that have sufficiently high probability of explaining the data, and thereby provide an accurate collection of parameters. The more uninformative the prior, the greater the search region, and the more random vectors that must be selected and tried. In addition, if the PK/PD model consists of numerically integrating differential equations, then each $l(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{\sigma})$ that must be evaluated is computationally expensive.

One approach is to increase the efficiency of the Monte-Carlo method used. The Markov-Chain Monte Carlo method (MCMC) has been shown to be a suitable approach for 3-stage hierarchical/Bayesian methods.[39-41] The MCMC method can be implemented such that the conditional integration under $\boldsymbol{\theta}$ is performed in an efficient manner as part of the Bayesian analysis.[10] One caveat in using this sophisticated Monte Carlo technique is that because it is a conditional probability method so that the random vectors selected are not statistically independent of each other, there is a danger of the algorithm to reduce its search area to a subdimensional space for extended series of random samples, and provide biased estimates. As we shall see, simple PK models can be executed fairly quickly. With more complex PK/PD models requiring integration of ordinary differential equations, the computation time can be much greater than that of the 2-stage hierarchical methods. Nonetheless, advanced PK/PD modeling methods using ordinary differential equations have been implemented recently using WinBUGS.[37,42]

The 3-stage hierarchical/Bayesian approach provides a comprehensive analysis of the population data through the PK/PD model selected, and provides the ability to study the profile of an entire set of likely population parameters. In this sense, it is typically superior to 2-stage hierarchical maximum likelihood methods, but requires greater computation time.

Software Evaluation Using 4 Examples

For this survey, we sought to assess the general reliability of the estimates as well as their standard errors provided by several population analysis programs. As only one data set was created for each of the problems, this was done by testing if

the estimated parameters differed from the reference values by more than 2 or 3 estimated standard errors. This test was selected because as a first approximation the PK/PD modeler typically considers that the estimates have a 95% probability of being within 2 standard errors, or to be more conservative considering that the analysis is not linear, 3 standard errors, from the true values. This method therefore serves as an informal assessment of parameter estimation for a single data set, as would be applied in typical PK/PD modeling sessions. If the standard error is accurately assessed, it would represent the precision of the parameter estimate that one should expect, given the number of subjects, number of data points per subject, sampling strategy, and dosing regimen in the simulated data set. So that assessment of accuracy of estimated parameters would not be penalized because of inaccurate assessment of standard errors, it was duly noted wherever the parameters exceeded 2 or 3 standard errors because the standard errors were undervalued by the program, or the estimates were nonetheless within 10% of the reference values.

These examples have not been analyzed in as exhaustive a manner that would be required to make more formal statistical assessments. A formal statistical analysis would require creating at least 30 simulated data sets for each problem, and performing statistical significance tests on the 30 sets of results compared with the reference values. It is statistically probable that a single simulated data set could be randomly created such that one or more software programs yield results that are unusually far from the reference values. However, the data sets created here were made large enough in number of subjects and/or in amount of data per subject, and were sufficiently balanced in terms of distribution of sampling times and/or covariate characteristics, so that the assessment of the parameters would be fairly unambiguous, and we should only occasionally expect a large deviation of 1- or 2-parameter estimates from their reference values.

We were also interested in testing the general robustness of each program using one arbitrary initial parameter setting that was very poorly informed. We typically used initial values where all of the parameters had the same value, to provide as little "fore-knowledge" as possible, while still being numerically sensible. For all of the programs and examples, when alternative initial settings were tried, they reached similar positions as those reported in the Tables. The reason again is that the data sets were simulated in a fairly balanced manner, with sufficient information for all of the varied parameters, with a resulting likelihood profile that provided a single minimum. Such balance and informativeness for all of the parameters may not be obtainable from empirical data, so in practice some trial and error in selecting initial values and fixing certain values is necessary. Additional tests such as posterior predictive checks and other bootstrap techniques should also be applied in practice, but these require more time and effort.

Five population analysis programs are described: NONMEM, PDx-MCPEM, S-ADAPT, MONOLIX, and WinBUGS/PK-Bugs. S-ADAPT and PDx-MCPEM provide an exact EM algorithm with a mature environment for advanced PK/PD modeling. MONOLIX provides an exact EM algorithm in the MATLAB environment, but leaves it up to the user to provide all of the code necessary to evaluate the model function, so the user must supply the dose parsing routine, and differential equation solver, if needed for the model. The authors of MONOLIX plan on adding these essential tools for the PK/PD modeler in the future. For each program we describe its general interface, ease of use, abilities, and how it performs on 4 simulated data sets. PK/PD parameters were simulated based on a log-normal distribution among subjects, and data were simulated based on a proportionate error model.

Analyses were performed on a Dell Pentium 4 3.20-Ghz computer, with 1 gigabyte non-ECC 400-MHz DDR2 memory, and 80-GB SATA 7200-rpm hard drive with Data Burst Cache. The operating system was Windows XP, and the NONMEM, S-ADAPT, and PDx-MCPEM software packages were compiled using Intel Fortran 9.1.

Simulated Data Sets for the 4 Examples

The first data set was simulated from a 2-compartment PK model with 2 data points per subject, 1000 subjects, to determine parameter assessment in a sparse data sampling design. The model parameters consist of

CL: clearance

V1: volume of distribution of the central compartment

Q: distribution rate to peripheral compartment

V2: volume of distribution of the central compartment

For each subject, a parameter set was randomly selected from a log-normal multivariate distribution, and simulations were performed with an intravenous bolus dose of 100 units. From this random parameter set, data with residual error were simulated at 2 sampling times from a discrete set of times: 0.1, 0.2, 0.4, 0.7, 1, 2, 4, 7, 10, 20, 40, and 70 times units. All possible pairs of times were equally represented among the subjects. As there are $(12 \times 11)/2=66$ combinations, there were $1000/66=15$ subjects for each sample time combination. This method of distributing sample times provided a balanced data set, providing sufficient information to obtain estimates of interindividual variability among parameters as well as a residual error coefficient of variation, despite the sparse amount of data per subject.

A second problem is the same 2-compartment model, but the population means of CL and V1 are modeled to be dependent on each subject's sex and age:

If male:

$$CL_i = CL_m age^{CL_{m\_age}} \exp(\eta_i) \quad (39)$$

$$V1_i = V1_m age^{V1_{m\_age}} \exp(\eta_i) \quad (40)$$

If female:

$$CL_i = CL_i age^{CL_{f\_age}} \exp(\eta_i) \quad (41)$$

$$V1_i = V1_{f^{age}}^{v1_{f\_age}} \exp(\eta_i) \quad (42)$$

The data set consisted of 200 male and 200 female subjects, each subject simulated with an intravenous bolus dose of 100 units. Each subject had data simulated at 5 time points, using 1 of the 3 combinations: (0.1, 0.4, 0.7, 4, 20); (0.2, 1, 4, 7, 40); or (0.4, 2, 10, 40, 70) time units. This problem was designed to assess the program's ability to perform covariate estimation.

A third problem is a single bolus 2-compartment PK model with an E-max PD model, requiring 8 total parameters to describe the model. Only analytical functions are used to describe the model, as follows:

$$C(t) = Ae^{-\alpha t} + Be^{-\beta t} \quad (43)$$

$$C_e(t) = \frac{Dk_{eo}}{Vc}(E_1 + E_2 + E_3) \quad (44)$$

$$E(t) = \frac{E_{\max} C_e(t)^Y}{C_e(t)^Y + C_{e50}} \quad (45)$$

$$\alpha = \frac{(k_{12} + k_{21} + k_{10}) + \sqrt{(k_{12} + k_{21} + k_{10})^2 - 4k_{21}k_{10}}}{2} \quad (46)$$

$$\beta = \frac{(k_{12} + k_{21} + k_{10}) - \sqrt{(k_{12} + k_{21} + k_{10})^2 - 4k_{21}k_{10}}}{2} \quad (47)$$

$$A = \frac{D}{Vc}\frac{\alpha - k_{21}}{\alpha - \beta} \quad (48)$$

$$B = -\frac{D}{Vc}\frac{\beta - k_{21}}{\alpha - \beta} \quad (49)$$

$$E_1 = \frac{k_{21} - \alpha}{(\beta - \alpha)(k_{eo} - \alpha)} e^{-\alpha t} \quad (50)$$

$$E_2 = \frac{k_{21} - \beta}{(\alpha - \beta)(k_{eo} - \beta)} e^{-\beta t} \quad (51)$$

$$E_3 = \frac{k_{21} - k_{eo}}{(\beta - k_{eo})(\beta - k_{eo})} e^{-k_{eo} t} \quad (52)$$

where:

D = intravenous bolus dose (500 units); V_c, k_{10}, k_{12}, and k_{21} are PK parameters to a 2-compartment model; k_{eo} is the link effect parameter; E_{max}, C_{e50}, and γ are PD parameters to the Emax model; C(t) is the concentration of the drug in plasma; $C_e(t)$ is the concentration of drug in the effect compartment; E(t) is the effect as it changes with time. For each of 500 subjects, data were simulated at 3 or 4 PK data points, and 3 or 4 PD data points, with 1 of 3 combinations of time points: (0.03, 1, 20, 65), (0.1, 3, 35), or (0.3, 10, 50) time units, and the same time points were used for both PK and PD sampling in each subject. Residual error coefficients were fixed to their known values. PK and PD data were analyzed simultaneously in the population analysis. This problem is designed to test the ability of programs to efficiently analyze a population problem with many parameters. An earlier comparison of NONMEM V FOCE with S-ADAPT was made using this problem.[8]

The fourth problem is a 1-compartment PK kinetic model with first-order elimination and saturable elimination, plus an indirect response PD model, requiring numerical integration of a set of differential equations:

$$\frac{dX_1}{dt} = -k_{10} * X_1 - V_M * X_1 X_2 / (X_1 + K_M) \quad (53)$$

$$\frac{dX_2}{dt} = -V_M X_1 X_2 / (X_1 + K_M) - \mathrm{k}_{20} X_2 + \mathrm{k}_{02} \quad (54)$$

This problem is designed to test the ability of programs to efficiently analyze a population problem for a model requiring numerical integration. Each of 25 subjects received an intravenous bolus of 100 units, followed by an infusion of 1000 dose units over 1 time unit at time 7. For each subject a rich set of data were simulated, with PK and PD times at 0.05, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 7, 7.125, 7.25, 8, 10, 12, 14, 16, 18, 20, 22, 26, and 28 units. PK and PD data were analyzed simultaneously in the population analysis.

These 4 problems represent several of the types of challenges facing today's population PK/PD modelers.

NONMEM

NONMEM, distributed by ICON/Globomax (Ellicott, MD),[12] is open-source and written in Fortran 77, and requires a minimum yearly fee for support. The user compiles the source code using a compiler suitable for the operating system and hardware that the user would obtain separately (free Fortran compilers from the Web are available). The program uses the most basic terminal and file input and output features provided by Fortran compilers, and is therefore very portable. It is thus not interactive, but the NMTRAN pre-processor program allows the user to create a control stream file, with instructions given by the user on how to read the data file, what model to use, and what type of analysis to perform. The program is then compiled and linked, and is executed in batch mode. Users may select from predefined models, or write their own by

inserting Fortran-type code in the control stream file. The code may describe analytical model functions, or differential equations, which NONMEM then presents to its numerical integration routine. Furthermore, the population parameters may be modeled as a function of covariates such as gender, age, creatinine clearance, and so forth, using a syntax that represents the statistical basis of the model. The NONMEM program provides one of the most versatile modeling environments available for population PK/PD. Another advantage to this noninteractive method is that the entire problem is documentable, and therefore reproducible. By merely storing the control stream file and data file, the entire problem can be run again, and identical results will be obtained. Once the control stream file is written, the file may be modified and various versions may be created and executed.

A graphical user interface called PDx-POP is available that allows user-friendly set-up and data and parameter input. From this interface, the user may define the model, and enter initial model parameters. The PDx-POP program creates an NMTRAN control stream file based on these inputs, and runs NONMEM in the background. Various statistical diagnostic and result plots may be created as well to allow easy checking of data set integrity, residual error analysis, and other goodness-of-fit assessments. Additional features that may be helpful for the advanced user include convenient set-ups for multiple problem processing, simulations, and model checking. Furthermore, interfacing of input and output files with popular data processing programs such as Excel and S-Plus are also included.

The NONMEM FO method results for example 1 are listed in Table 1, and was completed in approximately 1 minute. The CL, V1, Q, and V2 values were more than 2 to 3 standard errors from the reference values, although CL and V1 were within 10% of the reference values. Log-transforming the data and using a homoscedastic residual error model did not improve the model values. The NONMEM FOCE analysis was completed in approximately 3 minutes, providing population results but not standard errors due to the R matrix being non-positive definite. Because there were a large number of subjects (1000), the standard errors were evaluated from the S Matrix instead, with the off-diagonal intersubject variances set to 0, which was allowable since this was the way the data were simulated. Most of the population parameters for the FOCE method were within 2 standard errors of the reference values. The CL parameter was approximately 5 standard errors from the reference, and V1 was 3 standard errors from the reference. However, the estimates of CL and V1 differ from the reference values by less than 10%, the estimation facilitated by the large number of subjects in the data set. The Laplace conditional method in NONMEM provides a second-order approximation to the individual integration step, so this method was tried. The Laplace method failed to complete when the initial values were far from the true values, so the analysis was begun at the FOCE result values. The analysis completed, but standard errors could not be obtained even with the S matrix method. Nonetheless, the final population values from the Laplace method were closer to the reference values than those of the FOCE method.

In the second example, 5 data points per subject were sampled, and the residual errors were small (Table 2). Thus, as expected, the FOCE method, with its linearized approximation of the residual error, yielded population values that were within 2 standard errors of the reference value, and completed the problem in approximately 3 minutes. The Laplace method required an additional 17 minutes starting near the FOCE derived parameter estimates, and provided similar results, but without standard errors. Because the intersubject variances were also small (~10%), the NONMEM FO method also yielded reasonable results, particularly when data were log-transformed and residual error was modeled homoscedastically.

In the third example, a linear kinetic 2-compartment PK model and sigmoidal E-max PD model were used to fit PK and PD data. The NONMEM FOCE method executed for approximately 20 minutes, beginning at poor initial values, and then terminated abnormally. The outputted values were used as new initial parameters for an additional run for 10 minutes, and the analysis completed successfully. However, the standard errors could not be evaluated because of lack of positive-definiteness in the information matrix, even when using the S matrix. The data were simulated with intersubject correlations between parameters, so they could not be set to 0 to facilitate the standard error evaluation. Nonetheless, the resulting values were similar to the reference values (Table 3), with only the Ce50, Gamma, and its variance being somewhat undervalued. The Laplace method required an additional 18 minutes starting near the FOCE-derived parameter estimates, and provided similar results, but without standard errors. The NONMEM FO method could not perform the problem unless a constant error model was used on log-transformed data, and also required 2 restarts, and the total analysis was completed in 7 minutes. The FO results were very inaccurate for k10, and the PD parameters Ce50, Emax, and Gamma. Starting the analysis near the reference values did not improve the final results.

For the fourth example, NONMEM's ability to perform an analysis expressed in the form of differential equations was assessed. The PK model was a highly nonlinear 1-compartment model with parallel first-order and Michaelis-Menten elimination, coupled with an indirect response PD model. Each subject was richly represented in data regarding the PK and PD components. The NONMEM FOCE method was able to advance the analysis considerably beginning at poor initial values, but terminated with round-off errors after 6 hours.

Table 1. Results of Example 1. Data Simulated by NONMEM

Parameter	Reference	Initial Values	NONMEM FO	NONMEM FO Loc(c)	NONMEM FOCE	NONMEM Laplace
CL	4.96	2	4.50†‡	5.66†	5.30†‡	4.90
			(0.0786)	(0.0998)	(0.0873)	
V1	5.06	2	5.52†‡	5.41*‡	5.53†‡	4.98
			(0.123)	(0.127)	(0.116)	
Q	1.99	2	2.20*	2.45†	2.03	2.01
			(0.0911)	(0.108)	(0.0761)	
V2	9.83	2	14.7†	10.9*	10.1	9.58
			(0.422)	(0.374)	(0.300)	
Var(CL)	0.163	0.8	0.208*	0.195*	0.190*	0.189*
			(0.0160)	(0.0154)	(0.011)	
Var(V1)	0.154	0.8	0.155	0.108	0.175	0.137
			(0.0251)	(0.0247)	(0.0188)	
Var(Q)	0.154	0.8	0.242	0.348*	0.230*	0.156
			(0.0682)	(0.0934)	(0.0368)	
Var(V2)	0.147	0.8	0.153	0.277*	0.124	0.191
			(0.0561)	(0.0530)	(0.0179)	
Sigma	0.25	0.25	0.273	0.278	0.205†	0.221*
			(.0190)	(0.0209)	(0.00937)	
–2LL			–300.347	1021.656	–2630.7	–2603.832
Computation Time			1 min	1 min	3 min	6 min

A resumption of the analysis moved the objective function an additional 0.1 unit, required 2 hours, and completed with a success statement. Standard errors were not determined because the R matrix was not positive definite, but the final parameters were similar to the reference values (Table 4). The S matrix version could not be used because there were few subjects for this data set. Attempts with NONMEM Laplace method were unsuccessful. NONMEM FO completed much more quickly, within 3 minutes (including 1 restart), but the resulting parameters were several-fold in error from the reference values. Even log-transforming the data so that a constant residual error model could be used did not improve the answers. Starting the analysis near the reference values did not improve the final results.

PDx-MCPEM

PDx-MCPEM is a program written by Serge Guzy and distributed by ICON/Globomax (Ellicott, MD)[13] that provides population analysis using direct and importance sampling Monte-Carlo Expectation-Maximization methods. A user-friendly interface is provided that is similar to PDx-POP for NONMEM, from which the user may select predefined PK and PD models. In the next version, the user may create his own PK/PD models. While a Fortran compiler is not needed for the predefined models, the user-defined Fortran model file must be compiled by Intel Fortran, Compaq Fortran, or gfortran, a freeware compiler supplied with PDx-MCPEM. With the user-defined feature, the versatility of the types of PK and PD models one can build in PDx-MCPEM is extended considerably over its first version, including models defined by differential equations. Diagnostic goodness-of-fit plots may be easily obtained, and the progress of the analysis can be monitored by a running plot of the objective function versus iteration number. The intersubject variance may be defined as a normal or log-normal distribution.

For example 1, importance sampling was performed with 1000 random samples evaluated per subject for the expectation step for 100 iterations for a total of 5 minutes. The final results were within 2 standard errors of the reference values (Table 1).

For example 2, first PDx-MCPEM was used assuming all parameters have both a fixed and a random effect. Importance sampling was performed with 1000 random deviates per subject during the expectation step, and 100 iterations led to reasonable estimates of the fixed effects. The covariate regression coefficients were then entered as fixed in the second step procedure where only the parameters exhibiting both fixed and random effects were estimates. All parameters were within 2 standard errors of the reference values,

Table 1. *Continued*

S-ADAPT ITS	S-ADAPT ITS-Laplace	S-ADAPT MCPEM	PDx-MCPEM	MONOLIX	WinBUGS
4.94	5.00	4.90	4.91	4.89	4.90
(0.0849)	(0.0848)	(0.0829)	(0.078)	(0.0766)	(0.0853)
5.04	5.22	5.16	5.13	5.14	5.16
(0.116)	(0.118)	(0.120)	(0.12)	(0.0957)	(0.123)
1.58†	1.55†	1.97	1.94	1.96	1.97
(0.0822)	(0.0783)	(0.0857)	(0.082)	(0.0478)	(0.0922)
8.36†	8.29†	9.50	9.45	9.42	9.50
(0.343)	(0.336)	(0.321)	(0.22)	(0.183)§	(0.336)
0.181	0.187	0.184	0.18	0.183	0.185
(0.0126)	(0.0121)	(0.0128)	(0.012)	(0.0102)	(0.0127)
0.225†	0.228†	0.138	0.144	0.137	0.141
(0.0214)	(0.0209)	(0.0129)	(0.025)	(0.0112)	(0.0197)
0.373†	0.222	0.173	0.158	0.179	0.167
(0.0726)	(0.0609)	(0.0481)	(0.052)	(0.0156)	(0.0497)
0.240	0.193	0.146	0.142	0.128	0.139
(0.0523)	(0.0478)	(0.0269)	(0.035)	(0.0109)	(0.0325)
0.173†	0.182†	0.243	0.237	0.244	0.246
(0.00277)	(0.00293)	(0.0120)		(0.00691)	(0.0129)
–2530.6	–2451.6	–2729.5	–2719.2	–2716.97	
2 h	2 h	14 min	10 min	11 min	41 min

Standard errors could not be obtained for NONMEM Laplace method, so those of WinBUGS were used for assessing the relative deviation from reference. The standard errors for NONMEM FOCE could only be obtained by fixing the intersubject covariances to 0, and selecting the S matrix feature.

*Estimated value more than 2 SE from reference.

†Estimated value more than 3 SE from reference.

‡Estimated value less than 10% different from reference value, despite being 2 or 3 SEs from reference.

§Estimated value is more than 2 SE from reference, but this is due to standard error being under-valued. Using standard error from Winbugs would make this value within 2 SE of reference.

Values in () are standard errors of the reported means.

In the second example, 5 data points per subject were sampled, and the residual errors were small (Table 2). Thus, as expected, the FOCE method, with its linearized approximation of the residual error, yielded population values that were within 2 standard errors of the reference value, and completed the problem in approximately 3 minutes. The Laplace method required an additional 17 minutes starting near the FOCE derived parameter estimates, and provided similar results, but without standard errors. Because the intersubject variances were also small (~10%), the NONMEM FO method also yielded reasonable results, particularly when data were log-transformed and residual error was modeled homoscedastically.

except for V1m, due to the standard error being undervalued. If the standard error from one of the other programs is used, the V1 is within 2 standard errors of the reference value. The total cpu time was 20 minutes.

For example 3, 1000 random samples were evaluated per subject for 100 iterations. For this analysis, the importance sampling method was facilitated by imposing user-defined constraints on the conditional variance of the proposal density. The final results were also within 2 standard errors of the reference values except for the parameter gamma (Table 3). The total cpu time was 40 minutes.

For example 4, 1000 random samples were evaluated per subject for the expectation step for 300 iterations for a total of 30 minutes. The importance sampling method was facilitated by imposing user-defined constraints on the conditional variance of the proposal density. PDx-MC-PEM computes the standard errors using a first-order linearization approach and therefore could not estimate reliable standard errors for the variance components since the sample size was only 25. However, all of the parameters were close to the reference values (Table 4).

S-ADAPT

S-ADAPT is a Fortran 95 open-source, free program[14] distributed by the University of Southern California,

Table 2. Results of Example 2. Data Simulated by S-ADAPT

Parameter	Reference	Initial values	NONMEM FO	NONMEM FO Log(c)	NONMEM FOCE
CL_m	26.9	2	27.0 (0.901)	27.5 (0.897)	27.3 (0.890)
Cl_{m_age}	–.609	2	–0.611 (0.00960)	–0.613 (0.00951)	–0.612 (0.00952)
CL_f	25.4	2	25.1 (1.65)	25.9 (0.745)	26.0 (0.745)
CL_{f_age}	–.204	2	–0.205 (0.0197)	–0.205 (0.00836)	–0.208 (0.00831)
$V1_m$	1.98	2	2.18* (0.0789)	2.12 (0.0804)	2.09 (0.0816)
$V1_{m\_age}$	0.352	2	0.326*‡ (0.0109)	0.333 (0.0110)	0.336 (0.0113)
$V1_f$	3.02	2	3.23 (0.137)	3.21 (0.117)	3.13 (0.112)
$V1_{f\_age}$	0.201	2	0.190 (0.0381)	0.188 (0.0105)	0.192 (0.0104)
Q	1.99	2	1.89*‡ (0.0381)	2.03 (0.0215)	2.00 (0.0209)
V2	9.93	5	10.5†‡ (0.103)	10.1 (0.0878)	9.99 (0.0856)
Var(CL)	0.0106	0.2	0.010 (0.000933)	0.0103 (0.000986)	0.0103 (0.000968)
Var(V1)	0.00941	0.2	0.00645* (0.00146)	0.00783 (0.00145)	0.00795 (0.00138)
Var(Q)	0.00972	0.2	0.0136 (0.00414)	0.0102 (0.00131)	0.0100 (0.00305)
Var(V2)	0.00881	0.2	0.00689 (0.00140)	0.0102 (0.00147)	0.00967 (0.00193)
Sigma	0.10	0.3	0.116† (0.00340)	0.103 (0.00313)	0.100 (0.00283)
–2LL			–9976.137	–4635.410	–10772.107
Computation time			1 min, could not be started at poor initial values	1min	3 min

Biomedical Simulations Resource department (USC, BMSR). The S-ADAPT program was developed by R. Bauer as an extension of the ADAPT II PK/PD modeling software[43] provided by USC, BMSR, and has been successfully used to analyze clinical data for Raptiva, consisting of 6 differential equations and 16 model parameters.[44] The S-ADAPT program is designed for the advanced PK/PD modeler who likes to have an interface with complete access to all variables and actions at run-time. It also has extensive simulation tools.

The program provides several interface types: interactive command line, interactive menu, and script-controlled interface. Because it uses advanced interactive input/output features, the program is heavily operating system and hardware dependent, and does not have the portability of NONMEM. The open-source code may be compiled by Intel Visual Fortran compiler or Compaq Visual Fortran compiler running under the Windows 98 or higher operating system on an Intel-based computer. A free g95 compiler is also provided, which creates a program with a console window environment, so that purchasing a Fortran compiler is not necessary. However, an S-ADAPT program compiled by Intel Fortran provides the fastest computing environment.

Table 2. *Continued*

NONMEM Laplace	S-ADAPT ITS	S-ADAPT-MCPEM	PDx-MCPEM	MONOLIX	WinBUGS
27.1	27.2	27.1	27.1	27.1	27.1
	(1.04)	(0.891)	(0.22)	(0.857)	(0.902)
–0.612	–0.611	–0.612	–0.612	–0.612	–0.612
	(0.0111)	(0.00958)	(0.0369)	(0.0092)	(0.00963)
25.9	26.0	25.9	25.3	26.0	25.9
	(0.748)	(0.746)	(0.18)	(0.70)	(0.756)
–0.208	–0.208	–0.209	–0.201	–0.209	–0.209
	(0.00841)	(0.00835)	(0.044)	(0.00794)	(0.00835)
2.08	2.08	2.09	2.03	2.08	2.08
	(0.102)	(0.0843)	(0.014)§	(0.0724)	(0.0847)
0.335	0.337	0.335	0.342	0.336	0.335
	(0.0136)	(0.0117)	(0.1)	(0.0101)	(0.0117)
3.12	3.12	3.13	3.11	3.13	3.13
	(0.129)	(0.114)	(0.075)	(0.0933)	(0.114)
0.192	0.191	0.191	0.192	0.191	0.191
	(0.0119)	(0.0105)	(0.08)	(0.00876)	(0.0105)
1.99	2.00	2.00	1.99	2.00	1.99
	(0.0218)	(0.0208)	(0.014)	(0.0128)	(0.0233)
9.96	9.97	9.97	9.95	9.96	9.95
	(0.0913)	(0.0850)	(0.075)	(0.0618)	(0.0947)
0.0102	0.0105	0.0105	0.0102	0.00984	0.0103
	(0.00103)	(0.000965)	(0.003)	(0.000778)	(0.000933)
0.00802	0.00825	0.00830	0.0087	0.00774	0.00838
	(0.00150)	(0.00140)	(0.0022)	(0.000941)	(0.00136)
0.00939	0.0127	0.0123	0.0103	0.00657	0.00868
	(0.00294)	(0.00252)	(0.0022)	(0.000867)§	(0.00253)
0.00937	0.0111	0.0109	0.00937	0.00867	0.00945
	(0.00207)	(0.00176)	(0.0039)	(0.000834)	(0.00171)
0.100	0.0989	0.0989	0.100	0.102	0.101
	(0.00163)	(0.00262)		(0.0107)	(0.00274)
–10772.157	–10770.1	–10780	–10772	–10777.2	
17 min started near FOCE result)	6.8 min	6 min	20 min	21 min	22 min

Standard errors could not be obtained for NONMEM Laplace method, so those of WinBUGS were used for assessing the relative deviation from reference.
*Estimated value more than 2 SE from reference.
†Estimated value more than 3 SE from reference.
‡Estimated value less than 10% different from reference value, despite being 2 or 3 SE's from reference.
§Estimated value is more than 3 SE from reference, but this is due to standard error being under-valued. Using standard error from Winbugs would make this value within 2 SE of reference.
Values in () are standard errors of the reported means.
In the third example, a linear kinetic 2-compartment PK model and sigmoidal E-max PD model were used to fit PK and PD data. The NONMEM FOCE method executed for approximately 20 minutes, beginning at poor initial values, and then terminated abnormally. The outputted values were used as new initial parameters for an additional run for 10 minutes, and the analysis completed successfully. However, the standard errors could not be evaluated because of lack of positive-definiteness in the information matrix, even when using the S matrix. The data were simulated with intersubject correlations between parameters, so they could not be set to 0 to facilitate the standard error evaluation. Nonetheless, the resulting values were similar to the reference values (Table 3), with only the Ce50, Gamma, and its variance being somewhat undervalued. The Laplace method required an additional 18 minutes starting near the FOCE-derived parameter estimates, and provided similar results, but without standard errors. The NONMEM FO method could not perform the problem unless a constant error model was used on log-transformed data, and also required 2 restarts, and the total analysis was completed in 7 minutes. The FO results were very inaccurate for k10, and the PD parameters Ce50, Emax, and Gamma. Starting the analysis near the reference values did not improve the final results.

Table 3. Results of Example 3. Data simulated by S-ADAPT

Parameter	Reference	Initial Values	NONMEM FO (log(C)	NONMEM FOCE
VC	50.0	1.1	53.5*‡ (1.59)	53.3*‡
K12	0.956	1.1	1.03 (0.0486)	0.977
K21	0.319	1.1	0.348*‡ (0.0129)	0.320
K10	1.39	1.1	1.64† (0.0654)	1.37
Ke0	0.100	1.1	0.124* (0.00838)	0.0991
Ce50	0.0205	1.1	0.581† (0.0314)	0.0175†
Emax	12.4	1.1	35.7† (5.20)	12.4
Gamma	1.59	1.1	0.387† (0.0324)	1.19†
Var(VC)	0.234	3	0.249 (0.0279)	0.253
Var(K12)	0.0822	3	0.114 (0.0405)	0.0653
Var(K21)	0.309	3	0.314 (0.0406)	0.286
Var(K10)	0.528	3	0.491 (0.0642)	0.566
Var(Ke0)	0.421	3	2.02† (0.512)	0.426
Var(Ce50)	0.563	3	290* (114)	0.624
Var(Emax)	0.571	3	13.1* (4.25)	0.560
Var(Gamma)	0.493	3	4.37† (0.945)	0.260†
–2LL			9452.240	–5699.472
Computation time			7 min, required 2 restarts	30 min, 1 restart

Standard errors could not be obtained for NONMEM Laplace and NONMEM FOCE methods, so those of WinBUGS were used for assesing the relative deviation from reference.

*Estimated value more than 2 SE from reference.

†Estimated value more than 3 SE from reference.

‡Estimated value less than 10% different from reference value, despite being 2 or 3 SE's from reference.

S-ADAPT adds to the modeling environment provided by ADAPT II. The user fills out a Fortran model file from a template, providing the code necessary to describe the PK and PD model functions, residual error functions, parameter transformations, covariate models for the population parameters, and differential equations (which are evaluated at run-time by the ADAPT II numerical integration solver), as needed. A series of model files are available, named after the various Advan/Trans models of NONMEM, so that the user may use any of these models as a starting template on which to add complexity, or to simply use them as is. Thus, filling out an S-ADAPT template file is no more complicated

Table 3. *Continued*

NONMEM Laplace	S-ADAPT ITS	S-ADAPT-MCPEM	PDx-MCPEM	WinBUGS
54.3*‡	52.6	51.2	51.4	51.4
	(1.52)	(1.43)	(1.1)	(1.48)
0.975	0.907	0.941	0.914	0.948
	(0.0382)	(0.0347)	(0.037)	(0.0366)
0.322	0.310	0.316	0.311	0.317
	(0.00913)	(0.00895)	(0.0093)	(0.00903)
1.33	1.38	1.36	1.36	1.36
	(0.0534)	(0.0505)	(0.041)	(0.0511)
0.103	0.0915†‡	0.0978	0.0976	0.0994
	(0.00283)	(0.00330)	(0.0049)	(0.00351)
0.0181*	0.0217	0.0210	0.0193	0.0198
	(0.00106)	(0.00103)	(0.0032)	(0.000996)
12.5	11.9	12.6	12.3	12.5
	(0.422)	(0.437)	(0.524)	(0.436)
1.22†	1.99†	1.36†	1.42†	1.37*
	(0.124)	(0.0735)	(0.03)	(0.0788)
0.233	0.235	0.245	0.212	0.266
	(0.0250)	(0.0237)	(0.03)	(0.0236)
0.0651	0.0786	0.0635	0.048	0.135*
	(0.0345)	(0.0258)	(0.051)	(0.0240)
0.290	0.282	0.291	0.247	0.314
	(0.0260)	(0.0230)	(0.033)	(0.0238)
0.520	0.528	0.544	0.498	0.546
	(0.0437)	(0.0424)	(0.028)	(0.0427)
0.468	0.279†	0.422	0.4046	0.440
	(0.0266)	(0.0377)	(0.038)	(0.0400)
0.601	0.434*	0.544	0.613	0.584
	(0.0508)	(0.0535)	(0.032)	(0.0609)
0.491*	0.573	0.565	0.57	0.568
	(0.0420)	(0.0373)	(0.066)	(0.0383)
0.217†	0.345§	0.549	0.343	0.426
	(0.0602)	(0.0721)	(0.078)	(0.0723)
–5596.874	–5555.8	–5784	–5652	
18 min, started from FOCE position	15.5 min	22 min	40 min	31 min

§Estimated value is more than 2 SE from reference., but this is due to standard error being under-valued. Using standard error from Winbugs would make this value within 2 SE of reference.

Values in () are standard errors of the reported means.

For the fourth example, NONMEM's ability to perform an analysis expressed in the form of differential equations was assessed. The PK model was a highly nonlinear 1-compartment model with parallel first-order and Michaelis-Menten elimination, coupled with an indirect response PD model. Each subject was richly represented in data regarding the PK and PD components. The NONMEM FOCE method was able to advance the analysis considerably beginning at poor initial values, but terminated with round-off errors after 6 hours. A resumption of the analysis moved the objective function an additional 0.1 unit, required 2 hours, and completed with a success statement. Standard errors were not determined because the R matrix was not positive definite, but the final parameters were similar to the reference values (Table 4). The S matrix version could not be used because there were few subjects for this data set. Attempts with NONMEM Laplace method were unsuccessful. NONMEM FO completed much more quickly, within 3 minutes (including 1 restart), but the resulting parameters were several-fold in error from the reference values. Even log-transforming the data so that a constant residual error model could be used did not improve the answers. Starting the analysis near the reference values did not improve the final results.

Table 4. Results of Example 4. Data Simulated by NONMEM

Parameter	Reference	Initial Values	NONMEM FO	NONMEM FO Log(c)
Vc	47.7	2	260.0†	110†
K10	0.0943	2	0.00299†	0.0468†
Vm	9.40	2	49.4†	6.91†
Kmc	1.12	2	9.96†	1.16
K02	37.9	2	59.3†	57.8†
K20	0.507	2	0.693†	0.504
SigmaPK	0.1	0.3	0.104	0.817†
SigmaPD	0.15	0.3	0.261†	0.314†
Var(VC)	0.101	2	38.5†	0.667†
Var(K10)	0.0795	2	28100†	4.49†
Var(Vm)	0.133	2	6830†	0.0748
Var(Kmc)	0.0725	2	8050†	0.841†
Var(K02)	0.130	2	141†	0.414†
Var(K20)	0.0932	2	348†	0.119
–2LL			–1278.644	–96.732
Computation time			3 min,1 restart	3 min,1 restart

than developing a NONMEM control file. Once this model file is completed, the user builds the program, and then executes it. When the program is loaded, a command window is displayed. The user then enters various commands in this window, providing initial values, inputting data files, and commands to begin the population analysis. Alternatively, these initial value settings and commands may be written into a script file, and then executed at run-time. Once the script file has completed, control returns to the program's command window, awaiting additional commands. By using a script file, S-ADAPT could be run completely in batch mode and controlled by other programs. Whereas in NONMEM a single control stream file provides a complete instruction set for creating the model, inputting data, and actions to take, in S-ADAPT the Fortran model file provides the mathematical model description, while a separate run-time script file provides the instructions for execution at run-time. If S-ADAPT is used in this way, then the Fortran model file, the script file, and the data file form the complete, documented, and therefore reproducible, analysis.

Alternatively, the user may activate a menu, which guides the user through the most often used series of actions needed to perform a complete population analysis. The menu system is helpful particularly for beginning users, but is not as user-friendly and graphical as PDx-POP or PDx-MCPEM. For example, the menu items lead you to enter an analysis name, input a data file, enter initial parameters, set analysis switches, and finally, the command to begin the population analysis itself. While the menu items are sorted in the logical order suitable for a straightforward analysis, the user may repeat any of the actions in the menu at will, in any order. For example, one may enter new initial parameters, and then jump

Table 4. *Continued*

NONMEM FOCE	S-ADAPT ITS	S-ADAPT-MCPEM	PDx-MCPEM	WinBUGS
	47.4	47.9	47.7	47.8
47.8	(3.41)	(2.99)	(1.3)	(3.44)
	0.0959	0.0943	0.997	0.0960
0.0955	(0.00583)	(0.00543)	(0.006)	(0.00666)
	9.28	9.33	9.2	9.21
9.30	(0.699)	(0.689)	(0.88)	(0.757)
	1.09	1.11	1.10	1.09
1.10	(0.0698)	(0.0647)	(0.10)	(0.0765)
	37.7	38.1	37.5	37.6
37.8	(2.81)	(2.67)	(3.1)	(2.80)
	0.507	0.513	0.505	0.506
0.510	(0.0341)	(0.0311)	(1.2)	(0.0335)
	0.0940	0.0956		0.0964
0.0957	(0.00356)	(0.0363)	0.096	(0.00359)
	0.148	0.147		0.147
0.147	(0.00490)	(0.00449)	0.147	(0.00452)
	0.0970	0.0963		0.127
0.0924	(0.126)	(0.0271)	0.096	(0.0393)
	0.0798	0.0774		0.118
0.0758	(0.286)	(0.0227)	0.071	(0.0367)
	0.129	0.131		0.166
0.122	(0.325)	(0.0380)	0.122	(0.0526)
	0.0756	0.0721		0.112
0.0710	(0.0973)	(0.0231)	0.070	(0.0371)
	0.116	0.118		0.137
0.109	(0.466)	(0.0339)	0.103	(0.0426)
	0.0829	0.0845		0.107
0.0808	(0.366)	(0.0251)	0.080	(0.0334)
–4028.374	–4027.4	–4027.1	–4013	
8 hr, 1 restart	5.5 min	12 min	30 min	9.8 h

Standard errors could not be obtained for NONMEM FO and NONMEM FOCE methods, so those of WinBUGs were used for assessing the relative deviation from reference.

*Estimated value more than 2 SE from reference.

†Estimated value more than 3 SE from reference.

‡Estimated value less than 10% different from reference value, despite being 2 or 3 SE's from reference.

Values in () are standard errors of the reported means.

right to repeating the analysis, without having to reenter analysis conditions. Various forms are also available to assist the user to enter parameter values, plotting variables, population analysis conditions, and so forth. Once the population analysis is completed, additional menu items are available for performing standard error analyses, post-hoc analyses, and graphical viewing of post-hoc results. All results are stored in S-ADAPT tables, the contents of which may be exported to comma delimited files that can be read by most programs. Other than post-hoc graphs, however, additional post-analysis statistical and graphical diagnostic tools are not easily available in S-ADAPT, so the user must typically export the results to programs such as S-Plus, SAS, or Matlab.

For example 1, the importance sampling method was used, with 100 random samples evaluated per subject for the expectation step for 150 iterations, and 1000 random samples for 20 iterations, for a total of 14 minutes. The final results were well within 2 standard errors of the reference values.

For example 2, 300 random samples were evaluated per subject for the expectation step for 20 iterations, followed

by 1000 random samples for 10 iterations, before parameters were visually assessed to no longer change in a directional manner, for a total time of 6 minutes. For this particular problem, the first 20 iterations were performed most efficiently by using the maximum a priori estimation method (MAP) to obtain an improved importance sampling proposal density for each iteration, which was then used for Monte-Carlo sampling. The final results were well within 2 standard errors of the reference values.

For example 3, 300 random samples were evaluated per subject for the expectation step for 30 iterations, and 1000 random samples for 70 iterations, before parameters were visually assessed to no longer change in a directional manner, for a total of 22 minutes. The first 30 iterations were performed most efficiently by using the MAP estimation method to obtain an improved importance sampling proposal density for each iteration, which was used for Monte-Carlo sampling. The final results were well within 2 standard errors of the reference values, except for gamma, which was a little larger than 3 standard errors of the reference value.

For example 4, 1000 random samples were evaluated per subject for the expectation step for 20 iterations, for a total of 12 minutes. The final results were well within 2 standard errors of the reference values. For this problem also, the analysis was performed most efficiently by using the MAP estimation method to obtain an improved importance sampling proposal density for each iteration, which was then used for Monte-Carlo sampling.

MONOLIX

MONOLIX Version 1.1. is an open-source free program developed for population analysis using nonlinear mixed effect models. The current version is prototype software implemented in the Matlab environment, and distributed by the MONOLIX Group, directed by France Mentre (INSERM and University Paris) and Marc Lavielle (University René Descartes and University Paris-Sud).[15] MONOLIX implements an algorithm that combined the stochastic approximation EM (SAEM) with a Markov Chain Monte Carlo procedure for maximum likelihood estimation of the PK/PD parameters in nonlinear mixed effect models without any linearization techniques. The stochastic EM method generates 1 to 5 random samples per subject per iteration to allow for an efficient and rapid convergence toward the solution, which is called the burn in period, requiring typically 200 iterations. Thereafter, the program accumulates random sample results among the next 300 iterations, to obtain a more precise estimate of the population means and intersubject variances.

In the MONOLIX program, the user supplies a user-defined PK/PD model and data file for analysis. Because the current version of MONOLIX is implemented in the Matlab environment, the user defines the PK/PD model using a Matlab programming language. A user-friendly interface is provided to link the user-defined PK/PD model and data file for the data analysis.

For example 1, the best results were obtained after 1000 iterations to obtain the parameter estimates, and ran for a total of 11 minutes. Most of the population parameters were within 2 standard errors of the references values. The population mean value for V2 was greater than 2 standard errors from the reference value, but this is because of a slight under-valuing of the standard error for this parameter.

For example 2, a new transformed covariate (cov_S) and the following reparameterization for CL needed for the MONOLIX program:

$$\begin{aligned} &\text{cov\_S} = \text{Gender} * \log(\text{age}) \\ &\log(\text{CLi}) = \log(\text{CLm}) + \theta_1 * \text{Gender} + \\ &\qquad \text{CLm\_age} * \log(\text{age}) + \theta_2 * \text{cov\_S} + \eta_i \end{aligned}$$

where (55)

$$\begin{aligned} &\log(\text{CLf}) = \log(\text{CLm}) + \theta_1 * \text{Gender} \\ &\text{CLf\_age} = \text{CLm\_age} + \theta_2 \end{aligned}$$

The same approach was used for the V1. The standard errors for the CLf, CLf_age, V1f, and V1f_age were obtained from the variance-covariance matrix using the principle of propagation of errors. The results with 2000 iterations, requiring 21 minutes, are presented in Table 2. Most of the population mean parameters were within 2 standard errors of the reference values.

The estimated value for the intersubject variance of Q was more than 3 SE from reference, but this is because of the standard error being under-valued. Using the standard error results from one of the other programs would make this value within 2 SE of reference.

The current version of MONOLIX only supports single-dose studies with 1 output function, therefore, example 3 with simultaneous PK/PD analysis cannot be done using MONOLIX without modification of the source code. Furthermore, the program did not have the ordinary differential equation solver that supports PKPD analysis using differential equations, and hence, no results were reported for example 4 for MONOLIX. The next version of MONOLIX (version 2.1) will support different PKPD libraries, multiple-dose studies, model differential equations, and NONMEM data format.

Iterative 2-Stage Method

S-ADAPT's iterative 2-stage mode was used. The results from these analyses are listed in Tables 1, 2, 3, and 4. S-ADAPT's ITS analyses were generally quite satisfactory

for problems 2 to 4. For the sparse data problem 1, the values were farther from the reference values compared with the MCPEM method, as expected for a linearized EM method. Using a second-order evaluation of the expectation step (ITS-Laplace) did not improve the values.

WinBUGS

WinBUGS is a general software for performing any type of multistage hierarchical problem using various types of distributions, not just univariate or multi-normal.[9,10] It is enjoying increasing popularity in the PK/PD population analysis field.[37,38,42,45,46] The program is downloadable from the Web, and is at present free of charge. The user creates a document window, or imports an already existing model file into the document window, and enters his model in Modulo-2 language, and has at his disposal the use of a handful of mathematical functions, and distributions. Following this, the user enters data and initial values for the problem into a document window. Again, data may be imported from a preexisting data file. Once all of the information is available in one or more document windows within WinBUGS, The user selects a series of commands from the menus, such as "check model" for syntax errors, "load data," "compile model," each step returning a success/failure diagnostic. The cursor must be pointed at the beginning of the model text when selecting "check model," and at the beginning of the data listing when "load data" is selected. Thus, one is committed to point-and-clicking quite a bit before analysis can begin. The output of WinBUGS can be extensive, consisting of thousands of individual sample values of population parameters and individual parameters, and provides a statistical summary, which can be summarized in the form of mean, standard deviation (equivalent to standard error in 2-stage methods), and quantile values. WinBUGS also provides graphical outputs for these results. Alternatively, the individual sample values can be exported for further statistical exploration in other programs.

An interface called PKBUGS that operates within WinBUGS has been created for 1-, 2-, or 3-compartment PK models, to automate creating the model text, entering basic covariate relationships, and constructing the data structures.[10] However, to perform more complex models, the user must create his own model, either by modifying the model file and data structures created by PKBUGS, or creating one from scratch. Furthermore, for the program to run more efficiently, the user should use the BlackBox Pascal compiler component, also downloadable from the Web, free of charge at present. This requires writing the most computationally intense part of the model in Pascal, and compiling it in Pascal, in addition to the main WinBUGS Model file written in Modulo-2 code. Fortunately, WinBUG's integration into the BlackBox environment is very straightforward.

Example 1 was simple enough to allow the PKBUGS interface to perform the analysis, using an available precompiled Pascal module called pk.model. The analysis was run for 40 000 iterations, requiring 10 minutes to complete. All values were well within 2 standard errors of the reference values (Table 1). With this and the other examples, statistics on the last 10 000 values of population parameters were performed to obtain their mean and standard error.

Example 2 was set up using the PKBUGS interface, and the covariate equations were further modified to match the manner in which the data were simulated. The analysis was allowed to run for 20 000 iterations, requiring 22 minutes to complete. All parameters were within 2 standard errors of the reference values (Table 2).

Example 3 was a PK/PD model that could not be evaluated with PKBUGS's pk.model routine; therefore, the equations were programmed in the WinBUGS Modulo-2 language. When this was done, the analysis required 8 hours to perform 10000 iterations, often with floating point errors occurring, requiring frequent re-starts. As this was very inefficient, another version of the model was created, with the majority of the PK/PD equations programmed in Component Pascal in the BlackBox environment, and then compiled, using modules made available by the WBDev package, which can be downloaded at present free of charge. In addition, the equations were constructed in such a way as to avoid floating point overflow. With this version of the program, analysis required 30 000 iterations to come to steady-state, executed in 31 minutes. Most parameters were within 2 standard errors of the reference, and all were within 3 standard errors.

Example 4 required the integration of differential equations, a computationally expensive process, so a Pascal component BlackBox module was created, which in turn used ordinary differential equation (ODE) modules available in the WBDiff package, an additional, free-of-charge, component of WinBUGS. While the parameters were close to steady state by approximately 30 000 iterations, another 40 000 iterations were required to be satisfied that there was no persistent drifting of all parameters. Thus, the program required 9.8 hours for 70 000 iterations, although the results after 4.2 hours were acceptable. All parameter estimates were within 2 standard errors of the reference values.

General Conclusions and Recommendations

The 4 examples used above offer a good representation of the types of problems that are often considered in population

PK/PD modeling. Selected population PK/PD programs that are available to the public and have or are expected to have a mature PK/PD modeling environment were tested in their ability to analyze the data with reasonable accuracy (within 2 or 3 SE of the reference values), the approximate amount of time, and with the degree of intervention required once the analysis run was started.

Certain types of population problems have not been investigated here, such as working with ordinal or some other non-normally distributed data,[26] and population mixture models. In this regard, the latest or soon-to-be-released versions of NONMEM, S-ADAPT, and PDx-MCPEM include ways to perform population mixture modeling, and allow the user to define his or her own individual likelihood function, so that non-normally distributed data can be modeled. While initial tests using the EM software show reasonable results, a full assessment of bias and precision for these types of problems has not been performed.

Among the 2-stage hierarchical methods, NONMEM FO method, which is the simplest methodology, performed very well in terms of speed and accuracy with a straightforward 2-compartment model (example 2), if the residual and interindividual variances were small (~10%). Accuracy was improved when the data were log transformed so that the residual error could be modeled homoscedastically. In contrast, when residual errors or interindividual errors were large, the FO method was very inaccurate, even for a simple 2-compartment model, as shown in example 1. The NONMEM FOCE method requires more computation time than FO, but generally gives accurate population values regardless of size of interindividual errors. The NONMEM FOCE, version VI, can complete the problem with reasonable efficiency if the PK/PD model is described in closed form, a great improvement over NONMEM FOCE V's performance.[8] On occasion, even with NONMEM VI, the problem terminates prematurely, and may need to be restarted at renewed initial values, but much fewer restarts are required than with NOMMEM FOCE V, and far fewer searches for the "right" new initial values are required. However, if data are very sparse (fewer than p data points per subject, where p is the number of PK/PD model parameters), the NONMEM FOCE method can lead to inaccurate values, in which case the Laplace method can at times provide more accurate results. Because the Laplace method uses a more complicated analysis method, however, it is not very robust, and one should progress to this method after employing the FOCE method. Also, NONMEM VI's differential equation solver in combination with the optimization search algorithm is still inefficient, and can lead to extensive computation times for such problems, as demonstrated with example 4. It is therefore recommended that the NONMEM FOCE method be used when the data are reasonably rich, the PK/PD problem can be described in analytical form, and when extensive covariate modeling and hypothesis testing needs to be performed. The NONMEM FO method at one time served a useful purpose when computers were slower, but with the improved speed of computers, and the increased efficiency of NONMEM FOCE in version VI, NONMEM FOCE should now always be used for future analyses, and NONMEM FO should NEVER be used for final analysis.

The constraint of making the intersubject variance matrix to be diagonal to improve stability in NONMEM should no longer be necessary, except if there are difficulties estimating the standard errors. In such cases, the full intersubject variance derived from the result may be analyzed for patterns of block diagonal correlations, and these constraints can then be imposed in order to evaluate the standard errors. In this regard, the failure by NONMEM to provide estimates of standard errors is seen as a positive feature by many statisticians in big Pharma. This failure usually signals some problem with the parameter surface at the apparent minimum and this should be taken into consideration when one is trying to develop a parsimonious and stable model. One of the more frequent causes of the failure of the Covariance Step in NONMEM is the user attempting to estimate intersubject covariances that are either very small or very large. Restructuring the intersubject variance-covariance matrix can often obviate this problem. All of the examples in this review were executed with full intersubject variance-covariance matrix assessment unless otherwise stated for purposes of standard error assessment, although the tables only list the diagonal variances for simplicity.

The exact EM methods have the advantage of greater stability in population analyzing complex PK/PD models, and additionally have reduced bias in assessing sparse or rich data.[47] For simple 2-compartment PK models with fairly rich individual data, and for complex covariate assessment, NONMEM FOCE performs more quickly and is sufficiently stable. Surprisingly, when PK/PD models are more complex, such as 2-compartment PK with Emax sigmoidal PD, or when requiring numerical integration of differential equations, the exact EM methods perform more quickly and stably than NONMEM FOCE.[8] This has been demonstrated again here. The computational expense of using Monte Carlo methods to obtain the results is more than offset by this increased stability, when the number of model parameters is 8 or greater. The MCPEM methodologies provided accurate assessments of the parameters, except for one parameter in example 3, which was a little over 3 standard errors deviant from the reference value. Furthermore, the MCPEM methods can always provide standard error assessments, even when modeling full intersubject variances. Because of this, the significance of an intersubject covariance between 2 parameters should be carefully assessed based on the

relative size of its standard error, since the algorithm is not likely to provide a failure signal as would occur in NONMEM. The S-ADAPT standard error assessment was found to be most consistent with the sampled standard error of WinBUGS. Also, both S-ADAPT and the second version of PDx-MCPEM, to be issued soon, provide mature user-definable PK/PD modeling environments. Both programs also have efficient differential equation solvers for nonclosed form PK/PD problems. S-ADAPT has the added advantage of incorporating ADAPT II's Nelder-Mead search algorithm along with Monte Carlo integration during the first several iterations to quickly move the population means toward the answer. Hence, S-ADAPT was particularly efficient in solving problem 4. MONOLIX is also expected to have a mature PK/PD modeling environment in its next version. While S-ADAPT is capable of modeling elaborate covariate models, the coding of modeling the PK/PD parameters to covariate data are not as straightforward as in NONMEM. Also, for MCPEM methods in general, a "trick" is needed to fit fixed effects that are shared among individuals with no intersubject variances associated with them.

WinBUGS, which uses a 3-hierarchical stage method for population analysis, provided accurate assessments of most of the population parameters. WinBUGS provided standard errors as well 95% quantile values (not reported here). WinBUGS uses the Gibbs sampling and Metropolis-Hastings methods,[9,10] and is the most sophisticated of the analysis methods among the programs tested. Therefore, it took typically twice as long as S-ADAPT, and 4 to 5 times as long as NONMEM FOCE, for the first 2 examples. However, like the MCPEM methods, it quickly caught up in efficiency when solving example 3, the closed form complex PK/PD model, and took no longer than NONMEM FOCE. Because WinBUGS provides a sampling distribution of values for the population parameters, more complete statistical analysis is provided than with any other program. However, like NONMEM FOCE, PK/PD problems modeled using differential equations such as example 4 took many hours, compared with the MCPEM methods that required 30 minutes or less. As mentioned earlier, WinBUGS is not ready made for PK/PD modeling, lacking in modules that would perform general packaging of dosing information and data that can then be submitted to any user-defined model. Furthermore, to obtain reasonable computation times, the user-defined model must be coded in Component Pascal, by the program BlackBox, which is an extra step in the model-development process. While the entire modeling environment together is free, 4 program components from separate Web pages had to be downloaded and installed (WinBUGS 1.41, BlackBox, WBDiff, and WBDev). The module PKBUGS makes available only 1-, 2-, or 3-compartment ready-made PK models, and no PD models. At this stage, WinBUGS does not have an extended interface available for the casual PK/PD modeler. However, an important improvement in Version 1.4 is the addition of a scripting language, allowing the user to automate the importing of data files and execution of steps, saving of results, even to the point of executing the WinBUGS (actually BlackBox) program completely as a batch process from the operating system command line, with subsequent program termination.

Because the 3-stage hierarchical Bayesian analysis is much more computationally expensive than 2-stage maximum likelihood methods, it is not practical to use the computing time to have WinBUGS find the answer from poor initial values, although it is capable of doing so. We found most efficient use by performing a population analysis in S-ADAPT, and then had the results and data exported into WinBUGS readable data files by a specially created module that is hooked into S-ADAPT. We also created template WinBUGS model files that would unpackage the dosing information from these data files, so that one could focus on developing the model equations themselves. Typically, having 20000 WinBUGS iterations performed around the S-ADAPT answer yielded very satisfactory steady-state results, without wasting WinBUGS iterations to move the answer to the proper location. In the not too distant future, we expect that population analysis would be typically performed in this way: first working out the model and performing a 2-stage heirarchichal analysis, followed by export into WinBUGS aided by data packaging routines, providing a final 3-stage hierarchical Bayesian analysis.

Acknowledgments

The authors thank Dr David D'Argenio of the University of Southern California for helpful suggestions on the manuscript, Dr Thomas Ludden of ICON for helpful suggestions on the manuscript and advice on the NONMEM analyses, Dr William Bachman of ICON for providing information on the PDx-POP capabilities, Dr Marc Lavielle of University of Paris-Sud for helpful advice on the MONOLIX analyses, and Dr David Lunn of Imperial College School of Medicine, London, for assisting us on the use of WBDev, WBDiff, and BlackBox.

References

1. Csajka C, Verotta D. Pharmacokinetic-pharmacodynamic modelling: history and perspectives. *J Pharmacokinet Pharmacodyn*. 2006;33:227-279.

2. 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.

3. Pillai G, Mentre F, Steimer JL. Non-linear mixed effects modeling—from methodology and software development to driving implementation in drug development science. *J Pharmacokinet Pharmacodyn*. 2005;32:161-183.

4. Mandema JW. Population pharmacokinetics and pharmacodynamics. In: Welling PG, Tse FLS, eds. *Pharmacokinetics.* vol. 67. 2nd ed. New York, NY: Marcel Dekker, Inc.; 1995:411-450.

5. Roe DJ. Comparison of population pharmacokinetic modeling methods using simulated data: results from the population modeling workgroup. *Stat Med*. 1997;16:1241-1262.

6. Aarons L. Software for population pharmacokinetics and pharmacodynamics. *Clin Pharmacokinet*. 1999;36:255-264.

7. Mentre F, Mallet A, Steiner JL. Hyperparameter estimation using stochastic approximation with application to population pharmacokinetics. *Biometrics*. 1988;44:673-683.

8. Bauer RJ, Guzy S. Monte Carlo parametric expectation maximization (MC-PEM) method for analyzing population pharmacokinetic/pharmacodynamic data. In: D'Argenio DZ, ed. *Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis.* vol. 3. Boston, MA: Kluwer Academic Publishers; 2004:135-163.

9. Best NG, Tan KK, Spiegelhalter DJ. Estimation of population pharmacokinetics using the Gibbs sampler. *J Pharmacokinet Biopharm*. 1995;23:407-435.

10. Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D. Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn*. 2002;29:271-307.

11. Beal SL, Sheiner LB. *NONMEM Users Guide—Part VII.* Hanover, MD: Globomax, Inc; 1992.

12. Users Guides NONMEM. [computer program]. Version V. Hanover, MD: Globomax, Inc; 1989-1998.

13. PDx-MCPEM Users Guide [computer program]. Version 1.0. Hanover, MD: Globomax, Inc; 2006.

14. S-ADAPT/MCPEM User's Guide [computer program]. Version 1.52. Berkeley, CA.; 2006.

15. Monolix Users Manual [computer program]. Version 1.1. Orsay, France: Laboratoire de Mathematiques, U. Paris-Sud; 2005.

16. PKBUGS software [computer program]. Version 2.0. Cambridge, UK: MRC Biostatistics Unit; 2006.

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

18. Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990;46:673-687.

19. Pinheiro JC, Bates DM. Approximations to the Log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Statist*. 1995;4:12-35.

20. Beal SL, Sheiner LB. Estimating population kinetics. *Crit Rev Biomed Eng*. 1982;8:195-222.

21. Ette EI, Kelman AW, Howie CA, Whiting B. Analysis of animal pharmacokinetic data: performance of the one point per animal design. *J Pharmacokinet Biopharm*. 1995;23:551-566.

22. Ette EI, Sun H, Ludden TM. Balanced designs in longitudinal population pharmacokinetic studies. *J Clin Pharmacol*. 1998;38:417-423.

23. Jones CD, Sun H, Ette EI. Designing cross-sectional population pharmacokinetic studies: implications for pediatric and animal studies. *Clin Res Pr Drug Regul Aff*. 1996;13:133-165.

24. White DB, Walawander CA, Tung Y, Grasela TH. An evaluation of point and interval estimates in population pharmacokinetics using NONMEM analysis. *J Pharmacokinet Biopharm*. 1991;19:87-112.

25. SAS online documentation, SAS/STAT Users Guide, NLMIXED procedure [computer program]. Version 8. Cary, NC: SAS Institute, Inc.; 2005.

26. Jonsson S, Kjellsson MC, Karlsson MO. Estimating bias in population parameters for some models for repeated measures ordinals data using NONMEM and NLMIXED. *J Pharmacokinet Pharmacodyn*. 2004;31:299-320.

27. Schumitzky A. EM algorithms and two stage methods in pharmacokinetics population analysis. In: D'Argenio DZ, ed. *Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis.* vol. 2. Boston, MA: Kluwer Academic Publishers; 1995:145-160.

28. Mentre F, Gomeni R. A two-step iterative algorithm for estimation in nonlinear mixed-effect models with an evaluation in population pharmcokinetics. *J Biopharm Stat*. 1995;5:141-158.

29. Walker S. An EM algorithm for nonlinear random effects models. *Biometrics*. 1996;52:934-944.

30. Lavielle M. SAEM in MATLAB: an alternative to linearization (software presentation). Presented at: PAGE Meeting; June 17-18, 2004; Uppsala, Sweden. Uppsala, Sweden: Population Approach Group Europe; 2004: Abstract 544.

31. Steimer JL, Mallet A, Golmard JL, Boisvieux JF. Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model. *Drug Metab Rev*. 1984;15:265-292.

32. Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc*. 1988;83:1014-1022.

33. Aarons L. The estimation of population pharmacokinetic parameters using an EM algorithm. *Comput Methods Programs Biomed*. 1993;41:9-16.

34. Thermo Kinetica software [computer program]. Version 4.4.1. Waltham, MA: Thermo Electron Corporation; 1997-2007.

35. Gomeni R, Pineau G, Mentre F. Population kinetics and conditional assessment of the optimal dosage regimen using the P-PHARM software package. *Anticancer Res*. 1994;14:2321-2326.

36. Popkinetics software [computer program]. Version 1.0. University of Washington Seattle; 2006.

37. Zhou Z, Rodman JH, Flynn PM, Robbins BL, Wilcox CK, D'Argenio DZ. Model for intracellular lamivudine metabolism in peripheral blood mononuclear cells ex vivo and in human immunodeficiency virus type 1-infected adolescents. *Antimicrob Agents Chemother*. 2006;50:2686-2694.

38. Dokoumetzidis A, Aarons L. Propagation of population pharmacokinetic information using a Bayesian approach: comparison with meta-analysis. *J Pharmacokinet Pharmacodyn*. 2005;32:401-418.

39. Gilks WR. Full conditional distributions. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov Chain Monte Carlo in Practice.* New York, NY: Chapman and Hall; 1996:75-88.

40. Gilks WR, Richardson S, Spiegelhalter DJ. Introducing Markov chain Monte Carlo. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov Chain Monte Carlo in Practice.* New York, NY: Chapman and Hall; 1996:1-19.

41. Bennett JE, Racine-Poon A, Wakefield AJ. MCMC for nonlinear heirarchical models. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov Chain Monte Carlo in Practice.* New York, NY: Chapman and Hall; 1996:339-357.

42. Gueroguieva I, Aarons L, Rowland M. Diazepam pharmacokinetics from preclinical to Phase 1 using a Bayesian population physiologically

based pharmacokinetics model with informative prior distributions in Winbugs. *J Pharmacokinet Pharmacodyn*. 2006;33:1-24.

43. ADAPT II Users Guide. Pharmacokinetic/Pharmacodynamic Systems Analysis Software. [computer program]. Version (Release) 4. Los Angeles, CA: Biomedical Simulations Resource, University of Southern California; 1997.

44. Ng CM, Joshi A, Dedrick R, Garovoy M, Bauer R. Pharmacokinetic-pharmacodynamic-efficacy analysis of efalizumab in patients with moderate to severe psoriasis. *Pharm Res*. 2005;22:1088-1100.

45. Mu S, Ludden TM. Estimation of population pharmacokinetic parameters in the presence of non-compliance. *J Pharmacokinet Pharmacodyn*. 2003;30:53-81.

46. Duffull SB, Kirkpatrick CJ, Green B, Holford NH. Analysis of population pharmcokinetic data using NONMEM and Winbugs. *J Biopharm Stat*. 2005;15:53-73.

47. Girard P, Mentre F. A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. Presented at: PAGE Meeting; June 16-17, 2005; Pamplona, Spain. Population Approach Group Europe; 2005: Abstract 834.