NONMEM Users Guide - Part VII

Conditional Estimation Methods

March 1998

by

Stuart L. Beal

Lewis B. Sheiner

NONMEM Project Group
C255
University of California at San Francisco
San Francisco, CA 94143

Table of Contents

I. Introduction 1
II. Methods 5
A. Estimation Methods 5
A.1 The Laplacian Method 5
A.2 The FOCE Method 5
A.3 The FO Method 6
A.4 The Hybrid Method 6
A.5 The Centering Methods 6
A.6 The Centering FOCE Method with the First-Order Model 7
B. Mixture Models 8
III. Usage Guidelines 10
A. Background 10
B. Model Form 11
C. Role of the FO Method 12
D. Role of Centering Methods 13
E. Role of the Hybrid Method 14
F. Problems 15
Figures 17
References 20

I. Introduction

This document gives a brief description of the estimation methods for population type data that can be used with NONMEM Version V. These include, in particular, a few methods that are new with this version, the centered and hybrid methods. The more important changes from the earlier edition published in 1992, but not all changes, are highlighted with the use of vertical bars in the right margin. This document contains no information about how to communicate with the NONMEM program.

To read this document it may be helpful to have some familiarity with the notation used with the representation of statistical models for the NONMEM program. See discussions of models in NONMEM Users Guide - Part I, but if one's interest is only in using NONMEM with PREDPP, see discussions of models in NONMEM Users Guides - Parts V and VI. Particular notation used in this Guide VII is given next.

The jth observation from the ith individual is denoted y_{ij}. Each individual may have a different number of observations. Each observation may be measured on a different scale: continuous, categorical, ordered categorical, discrete-ordinal.† An individual can have multivariate observations, each of different lengths. However, the multivariate nature of an observation is suppressed, as this is not relevant to the descriptions given in this document, and so the separate (scalar-valued) observations comprising the multivariate observations are all separately indexed by j. Each multivariate observation may have a different length. The vector of all the observations from the ith individual is denoted y_i.

It is assumed that there exists a separate statistical model for each y_i. This model is called the intraindividual model , or the individual model for the ith individual. It is parameterized by ψ, a (vector-valued) parameter common to all the separate intraindividual models, and η_i, a (vector-valued) parameter specific to the intraindividual model for y_i. Under this model, the likelihood of η_i for the data y_i (conditional on ψ) is denoted by $l_i(\eta_i;\psi)$, the dependence on y_i being supressed in the notation. This likelihood is called here the conditional likelihood of η_i .

When all the elements of y_i are measured on a continuous scale, an often-used intraindividual model is given by the multivariate normal model with mean $E_i(\eta_i;\theta)$ and variance-covariance matrix $C_i(\eta_i;\psi)$ (usually, ψ is comprised of parameters θ which are the only ones affecting E_i, and other parameters which, along with θ, affect C_i).†† This type of model shall be referred to as the mean-variance model . It is usually expressed in terms of a multivariate normal vector ε with mean 0 and variance-covariance matrix Σ. In the notation used here, the parameter ψ includes Σ (ignoring the matrix structure of Σ). For example,

$$y_{ij} = f_{ij}(\eta_i;\theta) + f_{ij}(\eta_i;\theta)\varepsilon_{ij}$$

where ε_{ij} is an instance of a univariate normal variable ε with variance $\Sigma = \sigma^2$. (When ε is multivariate, the observation y_{ij} is modeled in terms of a single instance of this multivariate random vector. A few other observations as well may be modeled in terms of this *same* instance, and thus under the model, all such observations are correlated and comprise a multivariate observation.) In this example, $E_{ij}(\eta_i;\theta)$ is $f_{ij}(\eta_i;\theta)$ (the mean of y_{ij}), and $C_{ijj}(\eta_i;\psi)$ is $f_{ij}(\eta_i;\theta)^2\sigma^2$ (the variance of y_{ij}). Since the ratio of the standard deviation of y_{ij} to the mean of y_{ij} is the constant σ, this particular model is called the constant coefficient of variation model.

---

† This document provides a description of estimation methods that can be used with observations of the same or different type. However, essentially, it neither contains any specific information about how to analyze observations of particular types, nor any information about how to communicate with NONMEM in order to do this.

†† Here and elsewhere in this section an explicit assumption concerning the normal probablility distribution is made. This is done primarily to keep the discussion simple. To various degrees in different situations the normality assumption does not play as important a role as our formally making the assumption might indicate.

The dependence of C_i on η_i is often a consequence of the intraindividual variance depending on the mean function, as with the above example, which in turn depends on η_i. This dependence represents an interaction between η_i and ε. With the (homoscedastic) model expressed by

$$y_{ij} = f_{ij}(\eta_i;\theta) + \varepsilon_{ij}$$

there is no such interaction; $C_{ijj}(\eta_i;\psi)$ is just σ^2. There are two variants of the first-order conditional estimation method described in chapter II, one that takes this interaction into account and another that ignores it.

When an intraindividual model involving ε is presented to NM-TRAN (the "front-end" of the NONMEM system), the model is automatically transformed. A linearization of the right side of the equation is used: a first-order approximation in ε_{ij} about 0, the mean value of ε_{ij}. Since the approximate model is linear in ε_{ij}, it is a mean-variance model. Clearly, if the given model is itself a mean-variance model, the transformed model is identical to the given model. Consider, for example, an intraindividual model where the elements of y_i are regarded as lognormally distributed (because the normally distributed ε_{ij} appear as logarithms):

$$y_{ij} = f_{ij}(\eta_i;\theta)\exp(\varepsilon_{ij})$$

In this case the transformed model is the constant cv model given above. (Therefore, no matter whether the given intraindividual model or the constant cv model is presented to NM-TRAN, the results of the analysis will be the same.)

Alternatively, the user might be able to transform the data so that a mean-variance model applies to the transformed data, which can then be presented directly to NM-TRAN. With the above example, and using the log transformation on the data y_{ij}, an appropriate mean-variance model to present to NM-TRAN would be

$$y_{ij} = \log f_{ij}(\eta_i;\theta) + \varepsilon_{ij}$$

(Actually, NM-TRAN allows one to explicitly accomplish the log transformation of both the data and the f_{ij}.) The results of the analysis differ depending on whether or not the log transformation is used. Without the log transformation, the values of the f_{ij} are regarded as arithmetic means (under the approximate model obtained by linearizing), and with the log transformation, these values are regarded as geometric means. Use of the log transformation (when this can be done; when there are no y_{ij} or f_{ij} with value 0) can often lead to a better analysis.

It is also assumed that as individuals are sampled randomly from the population, the η_i are also being sampled randomly (and statistically independently), although these values are not observable. The value η_i is called the random interindividual effect for y_i. It is assumed that the η_i are instances of the random vector η, normally distributed with mean 0 and variance-covariance matrix Ω. The density function of this distribution (at η) is denoted by $h(\eta;\Omega)$.

Often, some quantity P (viewed as a function of values of the covariates and the η_i) is common to different intraindividual models. For example, a clearance parameter may be common to different intraindividual models, but its value differs between different intraindividual models because the values of the covariates and the η_i differ. The randomness of the η_i in the population induces randomness in P. The quantity P is said to be a randomly dispersed parameter . When speaking of its distribution, we are imagining that the values of the covariates are fixed, so that indeed, there is a unique distribution in question.

From the above assumptions, the (marginal) likelihood of ψ and Ω for the data y_i is given by

$$L_i(\psi,\Omega) = \int l_i(\eta;\psi)h(\eta;\Omega)d\eta \qquad (1)$$

In general, this integral is difficult to compute exactly. The likelihood for all the data is given by

$$L(\psi,\Omega) = \prod_i L_i(\psi,\Omega) \qquad (2)$$

The first-order estimation method was the first population estimation method available with NONMEM. This method produces estimates of the population parameters ψ and Ω, but it does not produce estimates of the random interindividual effects. An estimate of η_i is nonetheless obtainable, conditional on the first-order estimates for ψ and Ω (or on any other values for these parameters), by maximizing the empirical Bayes posterior density of η_i, given y_i: $[l_i(\eta;\psi)h(\eta;\Omega)]/L_i(\psi,\Omega)$, with respect to η. In other words, the estimate is the mode of the posterior distribution. Since this estimate is obtained after values for ψ and Ω are obtained, it is called the posthoc estimate . When a mean-variance model is used, and a request is put to NONMEM to compute a posthoc estimate, by default this estimate is computed using $C_i(\psi,0)$. In other words, the intraindividual variance-covariance is assumed to be the same as that for the mean individual , the hypothetical individual having the mean interindividual effect, 0, and sharing the same values of the covariates as has the ith individual). However, it is also possible to obtain the posterior mode without this assumption.

The posterior density can be maximized using any given values for ψ and Ω. Since the resulting estimate for η_i is obtained conditionally on these values, it is sometimes called a conditional estimate at these values, to emphasize its conditional nature.

In contrast with the first-order method, the conditional estimation methods to be described produce estimates of the population parameters and, *simultaneously*, estimates of the random interindividual effects. With each different method, a different approximation to the likelihood function (1) is used, and (2) is maximized with respect to ψ and Ω. The approximation to (1) at the values ψ and Ω depends on an estimate $\hat{\eta}_i$, and as this estimate itself depends on the values ψ and Ω, the approximation gives rise to a further dependence of L_i on the values of ψ and Ω, one not expressed in (1). Consequently, as different values ψ and Ω are tried, different estimates $\hat{\eta}_i$ are obtained *as a part of* the maximization of (2). The estimates $\hat{\eta}_i$ at the values ψ and Ω that maximize (2) constitute the estimates of the random interindividual effects produced by the method (except for the hybrid method†). The estimate $\hat{\eta}_i$ also depends on y_i, and so, the approximation gives rise to a further dependence of L_i on y_i, one also not expressed in (1).

In contrast with the first-order method, a conditional estimation method involves multiple maximizations within a maximization. The estimate $\hat{\eta}_i$ is the value of η_i that maximizes the posterior distribution of η_i given y_i (except for the hybrid method††). For each different value of ψ and Ω that is tried by the maximization algorithm used to maximize (2), first a value $\hat{\eta}_1$ is found that maximizes the posterior distribution given y_1, then a value $\hat{\eta}_2$ is found that maximizes the posterior distribution given y_2, etc. Therefore, maximizing (2) is a very difficult and CPU intensive task. The numerical methods by which this is accomplished are not described in this document.

Fortunately, it often suffices to use the first-order method; a conditional estimation method is not needed, or if it is, sometimes it is needed only minimally during the course of a data analysis. Some guidance is given in chapter III. Briefly, the need for a conditional estimation method increases with the degree to which the intraindividual models are nonlinear in the η_i. Population pharmacokinetic models are often actually rather linear in this respect, although the degree of nonlinearity increases with the degree of multiple dosing. Population pharmacodynamic models are more nonlinear. The potential for a conditional estimation method to produce results different from those obtained with the first-order estimation method decreases as the amount of data per individual decreases, since a conditional estimation method uses conditional estimates of the η_i, which are all shrunken to 0, and the shrinkage is greater the less the amount of data per individual. Many population analyses involve little amounts of data per individual.

---

† After obtaining the population parameter estimates with the hybrid method (see chapter II), NONMEM ignores the estimates of the η_i that have been produced simultaneously with the population parameter estimates, and as with the first-order method, the posthoc estimates (described above) are the ones reported as the estimates of the random interindividual effects.

†† With the hybrid method, a constrained maximum is computed.

The conditional estimation methods that are available with NONMEM and which are described in chapter II are: the first-order conditional estimation method (with and without interaction when mean-variance models are used, and with or without centering), the Laplacian method (with and without centering), and the hybrid method (a hybrid between the first-order and first-order conditional estimation methods). For purposes of description here and in other NONMEM Users Guides, the term conditional estimation methods refers not only to these population estimation methods, but also to methods for obtaining conditional estimates themselves.

To summarize, each of the (population) conditional estimation methods involves maximizing (2), but each uses a different approximation to (1). Actually, $-2\log L$ is minimized with respect to ψ and Ω. This is called the objective function . Its minimum value serves as a useful statistic for comparing models. Standard errors for the estimates (indeed, an estimated asymptotic variance-covariance matrix for all the estimates) is obtained by computing derivatives of the objective function.

II. Methods

A. Estimation Methods

A.1. The Laplacian Method

Let $\Phi_i(\eta)$ be $-2\log l_i(\psi,\eta)$, and let $\Gamma_i(\eta)$ and $\Delta_i(\eta)$ be the gradient (column) vector and hessian matrix, respectively, of Φ_i evaluated at η. An approximation to $-2\log L_i$ is given by

$$\hat{\Phi}_i + \log|\Omega| + \log|\Omega^{-1} + \tfrac{1}{2}\hat{\Delta}_i| + \hat{\eta}_i'\Omega^{-1}\hat{\eta}_i - (\tfrac{1}{2}\hat{\Gamma}_i + \Omega^{-1}\hat{\eta}_i)'(\Omega^{-1} + \tfrac{1}{2}\hat{\Delta}_i)^{-1}(\tfrac{1}{2}\hat{\Gamma}_i + \Omega^{-1}\hat{\eta}_i)$$

where $\hat{\eta}_i$ is some estimate of η_i, and $\hat{\Phi}_i$, $\hat{\Gamma}_i$, and $\hat{\Delta}_i$ are Φ_i, Γ_i, and Δ_i all evaluated at $\hat{\eta}_i$. This results from applying a general approximation approach to integrals, attributable to the French mathematician Laplace, and described by De Bruijn (1961). With $\hat{\eta}_i$ equal to the conditional estimate obtained by maximizing the posterior density of η_i (in an unconstrained manner) - call this the unconstrained conditional estimate , this particular approximation has been used by others (Lindley, (1980); Mosteller and Wallace (1964)), although not with a function l_i that is as complicated as that which often arises in population pharmacokinetic and pharmacodynamic analyses. See also: Tierny and Kadane (1986). In this particular case, the last term of the approximation is 0. In general, the approximation can produce reasonable results as long the posterior distribution of η_i is dominated by a single mode. On occasion, a randomly dispersed parameter seems to have a multimodal distribution. See the discussion in section B concerning mixture models for a way to address this issue.

Each of the estimation methods uses a different variant of this approximation. However, with whatever variant is used, when in particular, the $\hat{\eta}_i$ are taken to be conditional estimates of the η_i at ψ and Ω, the general method described in chapter I becomes what we call a conditional estimation method. When the approximation is used just as it is stated above, and when the $\hat{\eta}_i$ are taken to be the unconstrained conditional estimates, the method is called the Laplacian estimation method , to honor the individual whose approximation plays such an essential role. However, the method itself involves an idea which is peculiar to NONMEM implementation. Namely, the approximation to L (the likelihood function of ψ and Ω), resulting from using the Laplacian approximation, is maximized.

When mean-variance models are used, the assumption can be made that each intraindividual variance-covariance matrix $C_i(\eta_i;\psi)$ is actually given by $C_i(0;\psi)$, the matrix for the mean individual. With this particular assumption, there is said to be no η,ε-interaction ; see chapter I. The Φ_i are computed differently, depending on whether an η,ε-interaction is assumed, as are the posterior modes. With mean-variance models, by default, NONMEM implements the Laplacian method assuming that there is no η,ε-interaction. With the currently distributed NONMEM code it is possible to apply the Laplacian method when there is an η,ε-interaction, but this code and its usage are not supported by the NONMEM Project Group.

A.2. The FOCE Method

The matrix Δ_i can be approximated by another matrix. Suppose given η_i, y_i is comprised of statistically independent subvectors $y_{i(1)}$, $y_{i(2)}$, etc., so that Φ_i can be written as a sum over terms $\Phi_{i(1)}$, $\Phi_{i(2)}$, etc. Then each of Γ_i and Δ_i can be written as a sum over terms $\Gamma_{i(1)}$, $\Gamma_{i(2)}$, etc. and $\Delta_{i(1)}$, $\Delta_{i(2)}$, etc., respectively. An approximation Λ_i to Δ_i is obtained by replacing each $\Delta_{i(k)}$ in the sum for Δ_i by $\Lambda_{i(k)} = \tfrac{1}{2}\Gamma_{i(k)}'\Gamma_{(i)k)}$. This is a type of first-order approximation; terms involving second derivatives have been dropped. It is called the first-order approximation . With this approximation, and when all the $\hat{\eta}_i$ are taken to be equal to the unconstrained conditional estimates of the η_i, the method is called the first-order conditional estimation (FOCE) method .

Actually, NONMEM allows the implementation of several versions of this method.

- When a mean-variance intraindividual model is used, by default, $\Lambda_{i(k)}$ is replaced by $\frac{1}{2}E(\Gamma_{i(k)}'\Gamma_{(i)k})$, where E represents the expectation over y_i under the intraindividual model. With the currently distributed NONMEM code it is possible to use the FOCE method without doing this, but this code and its usage are not supported by the NONMEM Project Group.
- The first-order conditional estimation method without interaction is the FOCE method applied with intraindividual mean-variance models and assuming no η,ε-interaction. When the intraindividual variance is assumed to be homoscedastic, and moreover, to be the same across individuals, then there is no η,ε-interaction, and in this case it may be shown that the FOCE method (without interaction) often produces results similar to those obtained with a method described by Lindstrom and Bates (1990). The first-order conditional estimation method with interaction is the FOCE method applied with intraindividual mean-variance models, but without the no interaction assumption. FOCE with and without interaction are both supported. With the currently distributed NONMEM code it is possible to apply the FOCE method with intraindividual models that are not mean-variance models, but this code and its usage are not supported by the NONMEM Project Group.

A.3. The FO Method

When the first-order approximation is used (with $\Lambda_{i(k)}$ replaced by $\frac{1}{2}E(\Gamma_{i(k)}'\Gamma_{(i)k})$), but when all $\hat{\eta}_i$ are taken to be 0 (the population mean value of η), the method is called the first-order (FO) estimation method . With the first-order method, the terms $\hat{\eta}_i'\Omega^{-1}\hat{\eta}_i$ and $\Omega^{-1}\hat{\eta}_i$ in the Laplacian approximation are 0. Note that since conditional estimates are not used, the first-order method is not a conditional estimation method.

It can be shown that when intraindividual mean-variance models are used, the method is equivalent to the first-order method as described, for example, in NONMEM Users Guide - Part I (also see e.g., Beal and Sheiner (1985)). Such an earlier description is also given below in section A.6. These earlier descriptions of the method apply only to mean-variance models. With the currently distributed NONMEM code it is possible to apply the FO method *as defined above* with intraindividual models that are not mean- variance models, but this usage is not recommended, and the code is not supported by the NONMEM Project Group.

A.4. The Hybrid Method

Suppose certain (but not all) elements of η are chosen to be in a set κ, that the elements of $\hat{\eta}_i$ corresponding to the elements of κ are taken to be 0, and that the remaining elements of $\hat{\eta}_i$ are taken to be those given by the Bayes posterior mode of η_i *under the restriction that all elements of* η *in* κ *are 0*. The conditional estimate thus defined is an example of a constrained conditional estimate. Suppose also that the first-order approximation is made. Then the method is a hybrid between the first-order method and the FOCE method. Accordingly, this conditional estimation method is called the hybrid method . Note that with the definition of the $\hat{\eta}_i$ used with this method, in contrast with the definition used with the FOCE and Laplacian methods, the last term in the Laplacian approximation is not 0.

A hybrid method can be considered that uses a weaker version of the first-order approximation. Consider using the first-order approximation, but only for the submatrix of Δ_i consisting of just those partial second derivatives such that the two variables with respect to which the differentiation occurs are in κ. This method is not supported with the currently distributed NONMEM code.

When the intraindividual models are statistical linear models (linear in the parameters η_i), the first-order, first-order conditional, hybrid, and Laplacian methods are all the same method, the classical maximum likelihood method.

A.5. The Centering Methods

The η_i are assumed to be distributed in the population with mean 0. When the *population* model fits the data well, this will be reflected by the average, $\bar{\eta}$, of the conditional estimates of the η_i across the sam-

pled individuals (at the values of the population parameters given by the model) being close to 0. (The converse does not necessarily hold.) When $\bar{\eta}$ is close to 0, the fit will be called centered . There is nothing about the methods defined above that insures that the fit will be centered. There are infrequently arising situations where the average is "far" from 0, where the model does not fit well (as judged e.g. by the differences $y_{ij}-f_{ij}(0,\hat{\theta})$ with mean-variance intraindividual models) and where a method that is designed to better center the fit might be tried (*do* see chapter III for some guidance). With a centering estimation method , the $\hat{\eta}_i$ are taken to be the unconstrained conditional estimates, and the approximation to $-2\log L_i$ is given by

$$-2\log l_i(\psi,\hat{\eta}_i-\bar{\eta}) + \log|\Omega| + \log|\Omega^{-1} + \tfrac{1}{2}\hat{\Delta}_i| + (\hat{\eta}_i'-\bar{\eta}')\Omega^{-1}(\hat{\eta}_i-\bar{\eta})$$

With NONMEM, there are centering FOCE and Laplacian estimation methods (with no η,ε-interaction). A centering hybrid method is not implemented in NONMEM.

A.6. The Centering FOCE Method with the First-Order Model

The first-order model is the population model which results when for all i, the ith given intraindividual model is a mean-variance model with mean $E_i(\eta_i;\theta)$ and variance-covariance matrix $C_i(\eta_i;\psi)$, and this model is replaced by another such model with mean

$$E_i(0;\theta) + \frac{\partial E_i}{\partial \eta_i}(0;\theta)\eta_i$$

and variance-covariance matrix $C_i(0;\psi)$.

The linearity of the η_i under this model implies that the population expectation of y_{ij} is $f_{ij}(0;\theta)$, the prediction obtained by taking η_i to be 0, its population mean. With mean-variance models, the FO estimation method is sometimes described as the application of the maximum likelihood method to the first-order model that results from the given model, and when using this method, it is usual to judge goodness of fit by the differences $y_{ij}-f_{ij}(0;\hat{\theta})$. When a conditional estimation method is used instead of the FO method, a centered fit may result, confirming that the population mean of the η_i is 0. However, the given intraindividual models are used, and they may be nonlinear in the η_i. Therefore, conceivably, $f_{ij}(0;\theta)$ may be a poor approximation to the population expectation of y_{ij}, and for this reason alone, an apparent bias in the fit may result. Experience suggests, though, that this should not be a major concern (perhaps because the nonlinear effect is small relative to the size of intraindividual variability in the residuals). If one is concerned, there are a couple of strategies one might use.

First, the NONMEM program allows the expectation of the y_{ij} to be estimated by means of a couple different types of actual integration (and not just when the intraindividual models are of mean-variance kind); see NONMEM Users Guide - Part VIII. Second, when the intraindividual models are mean-variance models, NONMEM allows the first-order model to be obtained automatically from the given model and used with the centering FOCE method. (If the first-order model is used with the noncentering FOCE method, the result is the same as that obtained with the FO method.) When a conditional estimation method is needed (see chapter III), application of the centering FOCE method to the first-order model that results from the given model may yield adequate results, and of course, the expectation of y_{ij} under the first-order model is simply given by $f_{ij}(0;\theta)$. Moreover, due to the linearity of the intraindividual models (of the first-order model) in the η_i, the computational requirement is substantially less than that incurred with application of the (centering or noncentering) FOCE method to the given model. The savings in CPU time is achieved at the expense of possibly using too simple a model (and, of course is still not as great a savings as is achieved with the FO method).

The first-order model may be used with the centering FOCE method, but not with the centering Laplacian method (because due to the linearity, the result would be the same as that obtained with the centering FOCE method). Be aware that when this model is used with the centering FOCE method, the conditional estimates produced by the method are based on the first-order intraindividual models (unlike

whenever the noncentering FOCE method is used, where the conditional estimates are based on the *given* intraindividual models). It is possible nonetheless to obtain posthoc estimates based on the given intraindividual models, at the population estimates obtained from using the centering FOCE method with the first-order model. A centering hybrid method is not implemented in NONMEM.

B. Mixture Models

On occasion, a model may need to incorporate a randomly dispersed parameter that has a possibly multimodal distribution. In this case a mixture model may be useful. This is a model where for each i, there are several possible intraindividual models, M_1, M_2, ..., M_r for y_i, and it is assumed that the particular model that actually describes y_i is one of these, but it is not known which one. It is assumed that the probability that it is M_k is p_k, where $p_1 + p_2 + \cdots + p_r = 1$. Loosely put, the ith individual is chosen randomly from a population divided into r subpopulations, their relative sizes either being known or unknown. The subpopulation of which the individual is a given member is not observable, but for each subpopulation, a model for data from an individual from the subpopulation is available. The mixing probabilities p_k correspond to the sizes of the subpopulations and are usually treated as parameters whose values are unknown and are estimated. With NONMEM, these probabilities can be modeled, i.e. related to covariables, and therefore, can vary between individuals. The parameters of these relationships can be estimated; they are included in ψ. To indicate this generality, the p_k may be written $p_{ik}(\psi)$ (the kth mixing probability for the ith individual).

Suppose, for example, that a clearance parameter of a pharmacokinetic model may be bimodally distributed in the population. Here is how this may be expressed with a population model. One may consider a mixture model with two intraindividual models for each individual: for the ith individual, one where the individual's clearance is given by

$$CL_i = \theta_1 \exp(\eta_{i1}) \tag{3}$$

and another where it is given by

$$CL_i = \theta_2 \exp(\eta_{i2}) \tag{4}$$

(The parameters η_{i1} and η_{i2} are the first two elements of η_i.) For each i, the value η_i arises randomly (see chapter I). For each i, a choice between the two intraindividual models is also viewed as one being made in a random fashion, according to probabilities p_1 and p_2 ($p_1 + p_2 = 1$). As a result of this choice, a value η_i^*, which is either η_{i1} or η_{i2}, is also "chosen". (Consequently, if *after* η_{i1}, say, is chosen, the value of η_{i2} does not influence the data.) From the point of view of not knowing what choices between intraindividual models were actually made, the distribution of the η_i^* across individuals is a mixture of two normal distributions, and the distribution of the CL_i is a mixture of two lognormal distributions.

The first two elements of the random variable η may have the same or different variances, i.e. Ω_{11} may or may not equal Ω_{22}. If these variances are sufficiently small, while the parameters θ_1 and θ_2 are sufficiently far apart, and if both probabilities p_1 and p_2 are sufficiently large (however in this regard, the variances, the θ's, and the probabilities must actually be considered altogether), the distribution of CL_i is bimodal. Often, the data may not allow all of the different variances between mixture components, such as Ω_{11} and Ω_{22}, to be well estimated, in which case the assumption might be made that these variances are the same (a homoscedastic assumption). With NONMEM, this can be done explicitly, or alternatively, the "same η" can be used with both mixture components, e.g. η_{i1} can be used in (3) and also in (4), instead of η_{i2}. NONMEM will understand that η_{i1} is symbolizing two "different η's", each having the same variance.†

† With NONMEM Version IV, the same η can also be used, and NONMEM will understand that it is symbolizing two different η's with the same variance, *provided the first-order estimation method is used*.

Other examples of mixture models may be given. See NONMEM Users Guide - Part VI, section III.L.2 for an example where the mixture model describes a mixture of two joint lognormal distributions for clearance and volume, *but which is not a bimodal distribution*. The differences between the models M_k need not be differences concerning parameters; they could be differences in model form. They can be any set of differences whatsoever.

The likelihood for y_i under a mixture model is

$$L_i(\psi,\Omega) = \sum_{k=1}^{r} p_{ik}(\psi)L_{ik}(\psi,\Omega)$$

where L_{ik} is the likelihood function for y_i under the the kth possible intraindividual model for individual i. With a mixture model, any of the estimation methods described in section A uses the defining approximation for the method with each of the L_{ik}, k=1, ..., r.

With a set of values for the population parameters ψ and Ω, NONMEM classifies each individual into one of the r subpopulations. The classification gives the most probable subpopulation of which the individual is a member. For each k, the empirical Bayes (marginal) posterior probability that y_i is described by M_k, given y_i, is computed by $[p_{ik}(\psi)L_{ik}(\psi,\Omega)]/L_i(\psi,\Omega)$. The individual is classified into the kth subpopulation if the kth probability is the largest among these r values.

III. Usage Guidelines

A. Background

Many data sets (real and simulated) have been examined using the first-order (FO) estimation method and, more recently, the conditional estimation methods. With many population pharmacokinetic data sets, the FO method works fairly well. It requires far less CPU time than does a conditional estimation method. However, from the time of its earliest usage there has been a small number of examples where the method has not worked adequately. Evidence suggesting that the method may not be adequate with a particular data set can be readily obtained with the goodness-of-fit scatterplot: with mean-variance intraindividual models, a plot of observations versus (population) predictions. Consider two such scatterplots in Figures 1 and 2. The one in Figure 1, resulting from use of the FO method, shows a clear bias in the fit. The data result from single oral bolus doses being given to a number of subjects; the data are modeled with a two compartment linear model with first-order absorption from a drug depot. The scatterplot in Figure 2 results from use of the FOCE method without interaction. Much of the bias is eliminated with the use of this method. In this situation, the benefit from the extra expenditure of computer time that is needed with the method is substantial.

The Laplacian method can use considerably more computer time than the FOCE method, depending on the complexity of the computations for obtaining needed second derivatives. In this example, the extra expenditure of computer time needed with the Laplacian method is not much, but the benefit is also not much. The scatterplot resulting from using the Laplacian method is very similar to that of Figure 2.

The Laplacian method should perform no worse than the FOCE method (the former avoids the first-order approximation). The FOCE method should perform no worse than the FO method (the adaquacy of the first-order approximation is better when the Λ_i are evaluated at the conditional estimates, rather than at 0). Similarly, the hybrid method should also perform no worse than the FO method, but perhaps not as well as the FOCE method. (See e.g. Figure 3, which is the goodness-of-fit plot for the same data described above, using the hybrid method (with two out of four η's "zeroed".)) This defines a type of hierarchy to the methods.

The need to proceed up the hierarchy from the FO method increases as the degree to which the intraindividual models are nonlinear in the η_i increases. The need to use the Laplacian method increases because as the degree of nonlinearity increases, the adequacy of the first-order approximation decreases. The need to use the FOCE method increases because as the degree of nonlinearity increases, the adequacy of the first-order approximation depends more on the values at which the Λ_i are evaluated.

Population (structurally) linear pharmacokinetic models are often rather linear (as just defined), although the degree of nonlinearity increases with the degree of multiple dosing. With these models the Laplacian method is rarely, if ever, needed. With simple bolus dosing, the FOCE method is often not needed, although the example cited above serves as a reminder not to interpret this last assertion too optimistically. On the other hand, population nonlinear pharmacokinetic models (e.g. models with Michaelis-Menten elimination) can be quite nonlinear. Population pharmacodynamic models also can be quite nonlinear, and especially with models for categorical- and discrete-ordinal-type observations, the Laplacian method is invariably the best choice.

The ability of a conditional estimation method to produce results different from those obtained with the FO method decreases as the degree of random interindividual variablity, i.e. "the size" of Ω decreases. This is because the conditional methods use conditional estimates of the η_i, which are all shrunken to 0, and the shrinkage is greater the smaller the size of Ω. The value 0 is the value used for the $\hat{\eta}_i$ with the FO method. Similarly, the ability of FOCE to produce results different from those obtained with the hybrid method decreases as the degree of random interindividual variablity, i.e. "the size" of Ω decreases. In fact, suppose one tries to use the FOCE method and finds that some estimates of interindividual variances are rather large compared to others. Then using the hybrid method where those elements of η with small variance are "zeroed", may well result in a fit about as good as that using FOCE (in contrast to that

shown in Figure 3), and if the number of elements of η that are zeroed is large relative to the total number of elements, CPU time may be significantly reduced.

The ability of a conditional estimation method to produce results different from those obtained with the FO method decreases as the amount of data per individual decreases. This is because the conditional methods use conditional estimates of the η_i, which are all shrunken to 0, and the shrinkage is greater the less the amount of data per individual. Actually, the amount of data from the ith individual should be measured relative to the number of parameters in the model for the individual, i.e. the number of elements of η_i upon which the model really depends. As the number of parameters increases, the amount of data decreases, and can "approach 0". Also, strictly speaking, the amount of data might be understood as being relative to the "data design" (the poorer the design, the less useful the data) and the magnitude of intraindividual error (the more error, the less useful the data).

With intraindividual mean-variance models where it may appear theoretically plausible that there is an η,ε-interaction, it might seem more appropriate to use the FOCE method with interaction than to use the FOCE method without interaction. However, when the amount of (true) intraindividual variance is large (though the intraindividual models may be structurally well-specified), or the amount of data per individual is small, it will be difficult for the data to support an η,ε interaction, in which case the FOCE method with interaction may produce no improvement over the FOCE method without interaction. Otherwise, and especially when intraindividual variance is small for some observations, but not for others due to structural model misspecification, and when there is considerable interindividual variability, the FOCE method with no interaction can lead to a noticeably biased fit (as can the FO method).

There seems to be no consistent relationship between the value of the objective function using one method and the value of the objective function using another method. Therefore, objective function values should not be compared across methods. However, objective function values (in conjunction with graphical output) can provide a very useful way to compare alternative models, as long as the values are obtained using the same method.

B. Model Form

Unless interindividual variability is small, use of a random interindividual effect in the model should be such that quantities that depend on the effect are always computed with physically meaningful values. For example, rather than model a clearance parameter by

$$CL_i = \theta_1 + \eta_{i1} ,$$

it is better to use

$$CL_i = \theta_1 \exp(\eta_{i1})$$

since clearance should always be positive. With the FO method, use of either model produces essentially the same results. (The formulas for clearance and for the derivatives of clearance with respect to η_{i1} are computed only with the value $\eta_{i1} = 0$.) However, with a conditional estimation method, different values of η_{i1} are tried. A negative value for CL_i can result with the first model, especially when Ω_{11} is large and large negative values of η_{i1} are tried.

To take another example: Suppose that with the one compartment linear model with first-order absorption from a drug depot, it is assumed that pharmacokinetically, for all individuals, the rate constant of absorption exceeds the rate constant of elimination, i.e. $KA_i > KE_i$. Instead of

$$KA_i = \theta_1 \exp(\eta_{i1}) ,$$

$$KE_i = \theta_2 \exp(\eta_{i2})$$

one should use

$$(KA-KE)_i = \theta_1 \exp(\eta_{i1})$$

$$KE_i = \theta_2 \exp(\eta_{i2})$$

and constrain both θ_1 and θ_2 to be positive. Again, with the FO method, use of either model produces essentially the same results. The problem with the first model is that when using a conditional estimation method, as η_{i1} and η_{i2} vary, the value of KE_i can exceed KA_i, due to "flip-flop". As this can happen, or not, from one individual to the next, if it happens at all, the conditional estimation method will "become confused" and fail. The conditional estimation method by itself has no way of knowing that it has been assumed that KE_i will not exceed KA_i, and it cannot distinguish flip-flop from this possibility. (If pharmacokinetically, KE_i may exceed KA_i, and vice versa, then if flip-flop occurs, again the conditional estimation method will become confused, not being able to distinguish flip-flop from these possibilities, but in this case, a modification of the model will not help.)

Consider again the simple model for a clearance parameter,

$$CL_i = \theta_1 \exp(\eta_{i1})$$

With the FO method, all derivatives with respect to η are evaluated at 0. Consequently, in effect, a transformed model for CL_i is used: a first-order approximation in η_i, of the right side of the equation,

$$CL_i = \theta_1 + \theta_1 \eta_{1i}$$

This is a constant cv type model. With the FO method, no matter whether the given model or the transformed model is "used", the results of the analysis will be the same. The same is true even if covariates are involved. However, when a population conditional estimation method is used, the results of the analysis will differ between the two models, as derivatives with respect to η are evaluated at conditional estimates.

C. Role of the FO Method

The following general guidelines are offered so that conditional estimation methods are used only when necessary, and thus unnecessary expenditure of computer time and other difficulties that sometimes arise with conditional estimation methods (see section D) are avoided. They are based on impressions, rather than systematic study. Clearly, there will arise situations where alternative approaches might be tried.

If the model is of a very nonlinear kind (see section A), then from the outset, a conditional method might be used instead of the FO method. Indeed, with models for categorical- and discrete-ordinal type observations, the Laplacian method should always be used, and the remainder of this discussion concerns the use of conditional estimation methods with models for continuous outcomes (more precisely, the intraindividual models are of mean-variance type).

When analyzing a new data set and/or using a very new model with the data set, it is a good practice to use the FO method with at least the first one or two NONMEM runs, in order to simply check the data set and control stream. The Estimation Steps with these runs should terminate successfully, although if a conditional estimation method is really needed, the results themselves may not be entirely satisfactory. At this very early stage of data analysis, the user needs to be able to detect elementary errors that may have been introduced into the data set or control stream, and to be able to detect significant modeling difficulties. This cannot be done easily if other unrelated problems that can arise with conditional estimation methods interfere.

One might do well to begin to develop a complete model, incorporating the covariates, etc., using the FO method. Decisions regarding the effects of covariates on randomly dispersed parameters are aided by examining scatterplots of conditional estimates versus particular covariates. When the FO method is used, the posthoc estimates are the conditional estimates that are used for this purpose. After it appears that the model can be developed no further, there nonetheless exists appreciable bias in the final fit, think about how this bias might be well-explained by model misspecification that has not been possible to ad-

dress (e.g. there is a known covariate effect, and the covariate has not been measured). The use of an estimation method cannot really compensate for bias due to model misspecification, and one should not imagine that a conditional estimation method is any different.

After model development is complete using the FO method, if there seems to be no bias in the fit, you might simply want to do one run with FOCE to check this impression. If after this, the fit does not significantly improve, you can stop. After model development is complete using the FO method, if there seems to be no bias in the fit, consider doing one run with FOCE to obtain the best possible estimates of variance-covariance components. The variance-covariance components are often estimated better using FOCE (but realize that sometimes, they may be estimated very similarly by FO - see discussion in section A), and when these estimates are important to you, it can therefore be worthwhile investing the time needed with the additional FOCE run. It is not necessary to use FOCE to sharpen the estimates of variance-covariance components until after an adequate model is developed using the FO method.

After model development is complete using the FO method, if appreciable unexplainable bias remains, do try using FOCE. Indeed, do not hesitate to try FOCE before model development is complete when a number of initial concientious attempts to improve your model using FO have resulted in *considerable* bias, and when conditions are such that *a priori*, the FO and FOCE results are not expected to be very similar (see background section). When the intraindividual models you are using permit the possibility of an η,ε-interaction that the data may be rich enough to support, try FOCE with interaction. If the use of FOCE significantly reduces the bias, continue to develop the model using FOCE. Or, before embarking on continued model development, first experiment with the hybrid method to see whether this produces as much bias reduction as does FOCE, along with significant improvement in run time over FOCE. Continued model development may entail repeating much of the work already done with the FO method. In particular, try adding covariates rejected when using the FO method, and reconsider alternative ways that the covariates already accepted can enter the model. As a result of the cost involved in possibly needing to repeat work already undertaken with the FO method, the question of how soon one begins to try FOCE is not clearly answerable. Surely, increased computational times must be considered, and usually one wants to delay using a conditional estimation method until use of such a method seems to be clearly indicated.

The model might be very nonlinear, in which case try the Laplacian method. If after using the FOCE and Laplacian conditional estimation methods, an appropriate goodness-of-fit plot is unsatisfactory, then there is very likely a modeling difficulty, and one must seriously acknowledge this.

D. Role of Centering Methods

If after concientious modeling using the appropriate (noncentering) conditional estimation method(s), a model results with which substantial bias still appears in the fit, there is probably a model-related explanation for this, though it may allude one. In these circumstances, one may want to proceed to obtain the best possible fit with the model in hand. The fit that has been obtained using the noncentering conditional estimation method is not necessarily the best fit that may be obtained with the misspecified model.

The bias may be reflected by an uncentered fit. When a population conditional estimation method is used, the average conditional estimate for each element of η is given in NONMEM output (the conditional estimates being averaged are those produced by the method), along with a P-value that can be used to help assess whether this average is sufficiently close to 0 (the null hypothesis). The occurence of at least one small P-value (less than 0.05, though when the P-value is small, it can be much less than 0.05) indicates an uncentered fit.

A centering method might be tried. Using centering FOCE or centering Laplacian, one should notice that the P-values are somewhat larger (although perhaps some are still less than 0.05), and often one will also notice considerable improvement in the fit to the data themselves. When it is necessary to use a centering method, the population parameter estimates (at least those identified with the misspecified part of the model) are themselves of little interest; population predictions under the fitted model are what is of

interest. Also, because the model is misspecified, one should anticipate possible problems with model validation and model applications involving extrapolation.

Although it may be that (at least in certain specifiable situations) fits with centering methods are in general no worse than those obtained with appropriate noncentering methods, this idea is not yet well enough tested. Moreover, routine use of centering methods will mask modeling problems. Centering methods should be used only when, after concientious modeling, bias in fit seems unavoidable. CENTERING METHODS SHOULD NOT BE ROUTINELY USED. When the model is well-specified, it seems unlikely that when using an appropriate noncentering method, bias in fit will result, and there should be no expectation that any further improvement can be gained with a centering method.

Even when the fit is centered, it may be possible (though rare) that the fit to the data themselves still shows bias (see remarks in chapter II). One might then also use centering FOCE with the first-order model, subject to the same cautions given above. (Recall that in this case, the conditional estimates of the η_i resulting from the centering method are based on linear intraindividual models. When centering is actually needed, these conditional estimates should probably be adequate for whatever purposes conditional estimates might be used. It is possible nonetheless to obtain posthoc estimates based on the given intraindividual models.)

Even when a model is well-specified, it may be so complicated (e.g. it uses difficult differential equations) that to use it with a conditional estimation method requires a great amount of computer time. In this case, if indeed a conditional estimation method is needed, one might use centering FOCE with the first-order model, even though centering per se may not be needed. In this situation, use of centering, along with the first-order model, is just a device allowing a conditional estimation method to be implemented with less of a computational burden. A compromise is achieved; the fit should appear to be an improvement over that obtained with the FO fit, but it may not be quite as good as one obtained with the noncentering FOCE or Laplacian methods. Because the first-order model is automatically obtained from the given model, the final form of the given model (obtained after completing model development) is readily available, and with this model, one might try to implement one run using either the noncentering FOCE or Laplacian methods and compare results.

E. Role of the Hybrid Method

As already noted in section A, use of the hybrid method may require appreciably less computer time than the FOCE method and yet result in as good a fit. There is another important use of this method.

A change-point parameter of the ith intraindividual model is a parameter of the model such that for any value of y_i, the derivative of l_i with respect to this parameter, evaluated at some value of the parameter (a change-point value), is undefined. An example of this is an absorption lagtime parameter A of a pharmacokinetic model for blood concentrations Cp. If a dose is given at time 0, then the derivative of the pharmacokinetic expression for Cp at time t with respect to A, evaluated at $A = t$ is undefined. So if moreover, an observation y_{ij} occurs at time t (so that the expression for Cp must be evaluated at this time), then the derivative of l_i evaluated at $A = t$ is undefined (for any value of y_{ij} or for any of the other observations of y_i). Therefore under these circumstances, if the change-point parameter is randomly dispersed, and η_i may assume a value at which $A = t$, then Γ_i is undefined at this value, and, strictly speaking, all estimation methods described in chapter II are undefined. But practically speaking, a method will fail only when, during the search to minimize $-2\log L$, a value of $\hat{\eta}_i$ at which $A = t$ cannot be avoided. A symptom that this is happening, when there is a randomly dispersed change-point parameter, is a search terminization with a large gradient, i.e. some gradient elements are 10^4 or larger. Often, a lagtime is estimated to be very near the time of the first observation within an individual record, and so the problem described here can be a very real problem. One remedy is to delete observations at times that are too close to estimated lag times. However, aside from entailing the deletion of legitimate data, there can also be implementation problems with this strategy.

If the hybrid method is used, and the element(s) of η associated with the change-point parameter - denote this by ζ - is zeroed, this reduces the number (across individuals) of values $\hat{\zeta}_i$ at which Γ_i could possibly be undefined in the computation, as only the value of the change-point parameter for the typical subject is needed in the computation. Indeed, unless the change-point parameter itself depends on a covariate, only at the value $\hat{\zeta}_i = 0$ can Γ_i possibly be undefined in the computation. Thus, the chance of the problem occurring is reduced (but not eliminated).†

F. Problems

A conditional estimation method can demonstrate somewhat more sensitivity to rounding error problems during the Estmation Step than will the FO method. When the search for parameter estimates terminates with rounding error problems, oftentimes intermediate output from the Estimation Step will indicate the accuracy with which each of the final parameter estimates has been obtained. For example, 3 significant digits may be requested for each estimate, but for some estimates, less than 3 digits is actually obtained. If only a little less than 3 digits is obtained (e.g. 2.7-2.9), and if the gradient vector of the objective function with the final parameter estimates is small (e.g. no element is greater than 5 in absolute value), then this degree of accuracy is probably acceptable. If much less accuracy is obtained, but only with those estimates where this might be expected and where this is tolerable (e.g. estimates of Ω elements), then again, one might regard the Estimation Step as having terminated successfully. (The order of the parameter estimates printed in the iteration summaries is: the θ's in their subscripted order, followed by the (unconstrained) Ω elements, followed by the (unconstrained) Σ elements. Note though, that these estimates are those of the scaled transformed parameters (STP), rather than the original parameters; see NONMEM Users Guide - Part I, section C.3.5.1.)

With a conditional estimation method (in contrast with the FO method), NONMEM can more readily terminate during the Estimation Step with a PRED error message indicating e.g. that a nonallowable value for a parameter has been computed in PRED code, perhaps a negative value for a rate constant.†† This is because a parameter may be randomly dispersed, and with a conditional estimation method, values of η different from 0 are tried, as well as are different values of θ, and some of these values might result in a nonallowable value of the parameter. If such a termination occurs, then, if not already doing so, consider modeling the parameter in a way that prevents it from assuming a nonallowable value, e.g if the parameter cannot be negative, consider using a model such as $P = \theta_1 \exp(\eta_1)$ (see section B). Sometimes this cannot completely solve the problem, e.g. if the parameter cannot also be 0, the model just given will not insure this (η_1 can be very large and negative). So, a termination may still occur. The next step is to try to include the `NOABORT` option on the $ESTIMATION record (see NONMEM Users Guide - Part IV, section IV.G.2). However, doing so will have no effect if the termination occurs during the 0th iteration.††† The `NOABORT` option activates one type of PRED error-recovery (THETA-recovery), and the other type (ETA-recovery) is always activated, without using the option. So the option may not need to be used initially, and if PREDPP is being used, to have used the option before a termination has actually occured has the detrimental effect that this can mask the occurrence of an error detected by PREDPP, of which the user needs to be informed. With PREDPP, *never use the* `NOABORT` *option until* you have had an opportunity (i) to see what happens when you do not use it, i.e. to see the contents of PRED error messages that might arise when you do not use the option, (ii) to respond, if possible, to these messages in a sensible way (other than using the option), and (iii) to see what happens after you have done this.

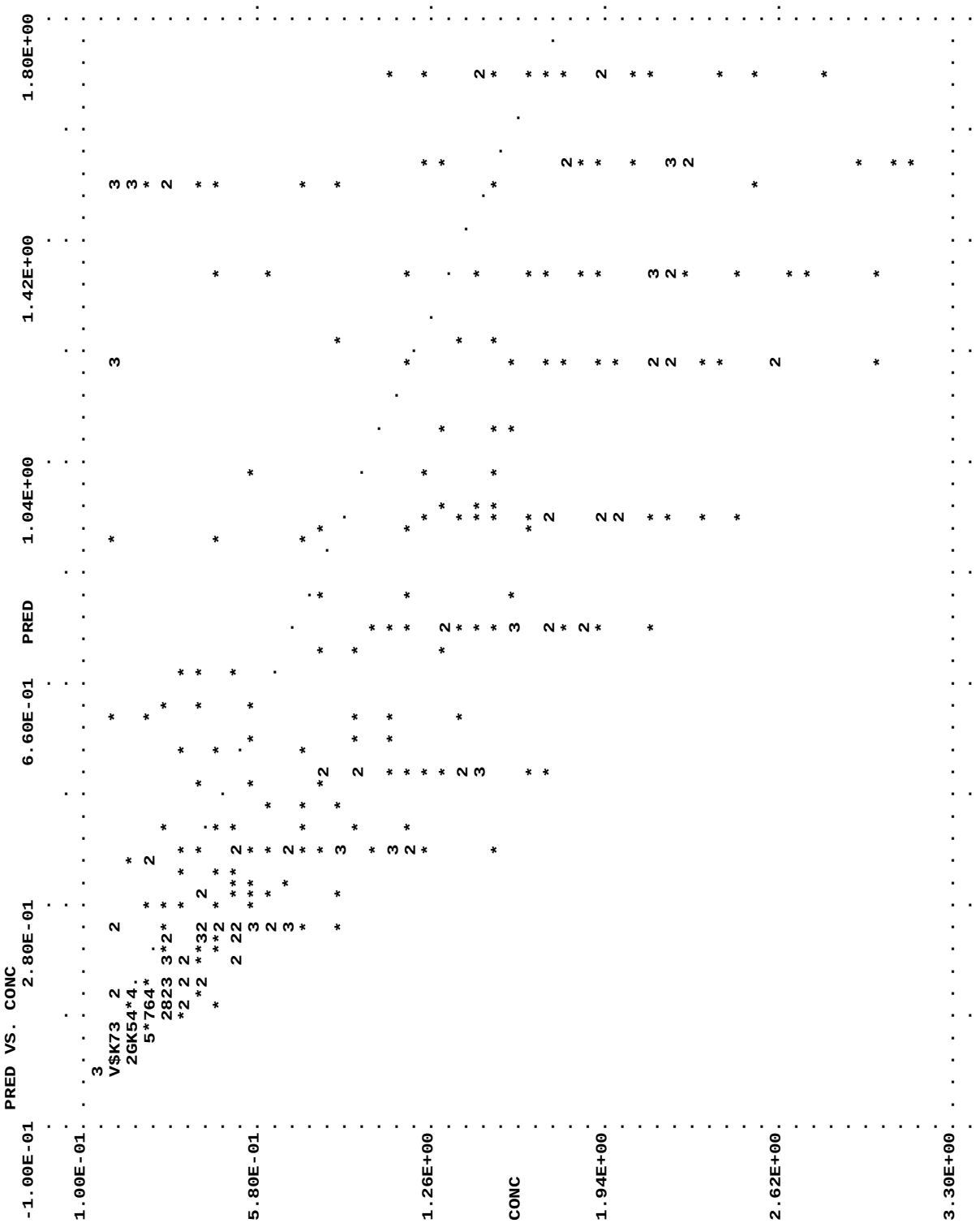
---

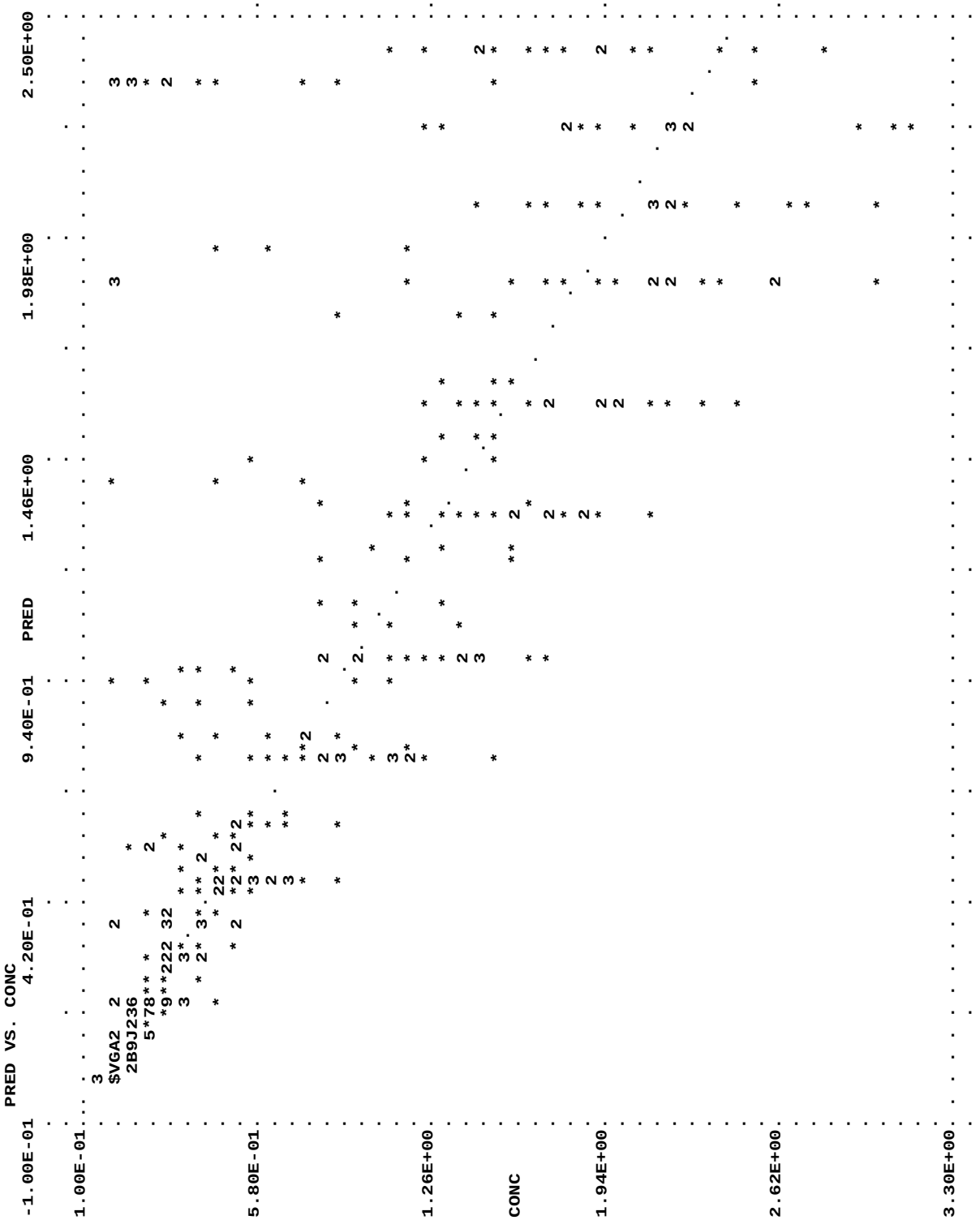
† Keep in mind that with the hybrid method, even though the elements of ζ are zeroed, an estimate of the magnitude of random interindividual variability in the parameter is still obtained.

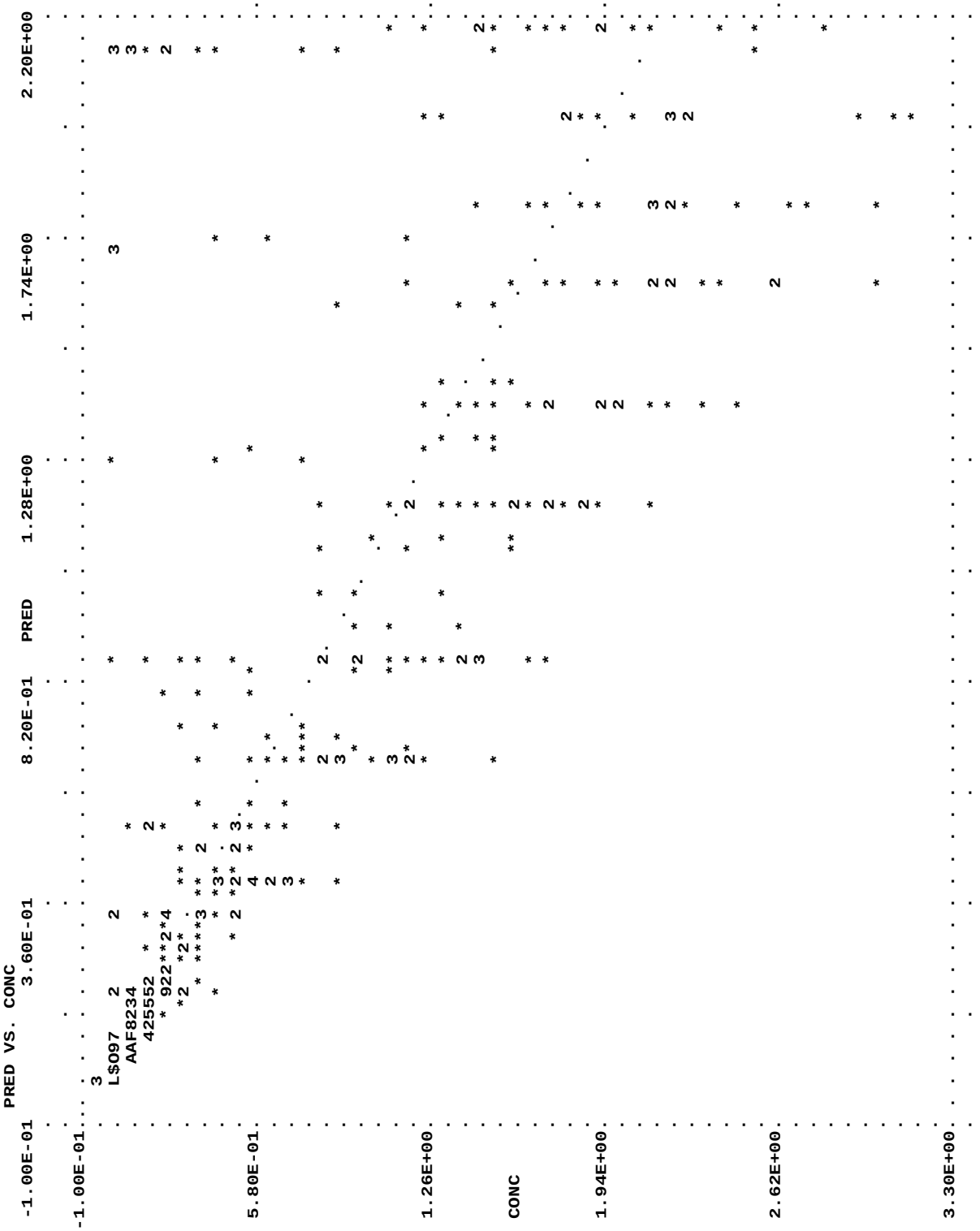
†† An PRED error message arises when PRED error-recovery (see NONMEM Users Guide - Part IV, section IV.G) is used in a user's PRED code, or if PREDPP is used, in a user's PK or ERROR code. PREDPP itself uses PRED error-recovery.

††† A termination during the 0th iteration can arise due to a problem with either the data set, user code, or control stream. Different initial estimates might be tried (perhaps smaller interindividual variances).

Perhaps the operating system, rather than NONMEM, terminates the program with a message indicating the occurrence of a floating point exception in a user-code. Again, this may be because a value η is tried which results in the exception when a value of a randomly dispersed parameter is computed. Underflows are ignorable, and terminations due to underflows should be disabled (see NONMEM Users Guide - Part III). With an operand error, or overflow, or zero-divide, the user needs to identify where the exception occurs in the code. For this purpose, the use of a debugger, or debugging print statements in the code, may be helpful. Then perhaps the exception may be avoided by using PRED error-recovery in the user-code, i.e. by using the EXIT statement with return code 1 (see NONMEM Users Guide - Part IV, section IV.G.2). Try this, and rerun. If with the earlier run, the termination occured after the 0th iteration, and if PREDPP is not being used, rerun the problem using the `NOABORT` option on the $ESTIMATION record. If the termination occured after the 0th iteration, and if PREDPP is being used, rerun, but *do not use the* `NOABORT` *option*. If the termination still occurs, then rerun a second time, this time using the `NOABORT` option. If the termination occurs during the 0th iteration, the `NOABORT` option has no effect. Such a termination can arise due to a problem with either the data set, user code, or control stream. Different initial estimates might be tried (perhaps smaller interindividual variances).


PRED VS. CONC
PRED
CONC
-1.00E-01
2.80E-01
6.60E-01
1.04E+00
1.42E+00
1.80E+00
-1.00E-01
5.80E-01
1.26E+00
1.94E+00
2.62E+00
3.30E+00


PRED VS. CONC
4.20E-01
9.40E-01
PRED
1.46E+00
1.98E+00
2.50E+00
-1.00E-01
-1.00E-01
5.80E-01
1.26E+00
CONC
1.94E+00
2.62E+00
3.30E+00


PRED VS. CONC
PRED
CONC
-1.00E-01
3.60E-01
8.20E-01
1.28E+00
1.74E+00
2.20E+00
-1.00E-01
5.80E-01
1.26E+00
1.94E+00
2.62E+00
3.30E+00
L$097
AAF8234

References

Beal, S.L., and Sheiner, L.B. (1985). "Methodology of population pharmacokinetics." Drug Fate and Metabolism: Methods and Techniques , Vol. 5, Garrett, E.R. and Hirtz, J. (Eds.), Chapter 4, 135-183. New York: Marcel Decker.

De Bruijn, N.G. (1961). Asymptotic Methods in Analysis . Amsterdam: North Holand.

Lindley, D.V. (1980). "Approximate Bayesian methods." Bayesian Statistics , J.M. Bernado, M.H. Degroot, D.V. Lindley, and A.M.F. Smith. (Eds.). Valencia, Spain: University Press.

Lindstrom, M.J. and Bates, D.M. (1990). "Nonlinear mixed effects models for repeated measures." Biometrics 46, 673-687.

Mosteller, F. and Wallace, D.L. (1964). Inference and Disputed Authorship: The Federalist . Reading, MA: Addison_Wesley.

Tierny, L. and Kadane, J.B. (1986). "Accurate approximatinos for posterior moments and marginal densities." JASA 81, 82-86.