Journal of Pharmacokinetics and Biopharmaceutics, Vol. 19, No. 3 (June Supplement 1991)

PHARMACOSTATISTICAL MODELING FOR OBSERVATIONAL DATA

Thaddeus H. Grasela, Jr., Pharm.D.[1] and Lewis B. Sheiner, M.D.[2]

[1] Departments of Pharmacy and Social and Preventive Medicine
Schools of Pharmacy and Medicine
State University of New York
Buffalo, New York
[2] Departments of Laboratory Medicine and Medicine
School of Medicine
University of California, San Francisco
San Francisco, California

INTRODUCTION

The process of analyzing the individual concentration time data obtained from each subject enrolled in an experimental pharmacokinetic study is very familiar to pharmacokineticists. With the aid of available computer programs, the data are analyzed to obtain estimates of pharmacokinetic parameters for each subject, and in a subsequent statistical analysis, summary statistics such as the mean and standard deviation are calculated for the patient population. The statistical considerations involved in this data analysis process are relatively straightforward and violations of the underlying statistical model are usually not severe and can be handled by the use of appropriate weighting procedures. The ease of data analysis in the experimental setting is due, in large part, to the quality and quantity of data obtained for each subject in the experiment. Blood samples for determination of serum drug concentrations are obtained over a wide range of concentrations, at carefully selected time points, and experimental conditions are maintained across the study population. Thus, conditions are optimized so that accurate and precise parameter estimates for each subject may be obtained.

Observational data, on the other hand, is characterized by a lack of control over

0090-466X/91/0600-025S$06.50/0

the frequency and timing of sampling. The data may arise during the routine clinical care of patients or during Phase III or Phase IV trials so that a limited number of samples are available. In some cases only one or two samples are obtained, and the timing of samples may not be optimal for yielding precise parameter estimates in each individual. Moreover, in the event of only one sample per patient, individual estimates of pharmacokinetic parameters are not attainable. Because of these limitations, the analysis of observational data must be approached cautiously and specialized data analysis methodologies must be employed.

In Chapter 1, the three types of population parameters of interest were discussed: fixed effect parameters (θ) describe the relationship between various independent variables and pharmacokinetic parameters, and the random effect parameters Ω and σ^2 provide estimates of the typical magnitude of interindividual variability in a pharmacokinetic parameter across the members of a population and the typical magnitude of residual variability consisting of combined intraindividual and measurement error variability, respectively.

One approach to the analysis of observational data involves the use of a mixed effect model, a model that mixes the influence of fixed effects and random effects on the observed plasma concentration, allowing certain parameters describing these influences to be simultaneously quantified. NONMEM is a Fortran 77 computer program that analyzes a general mixed effect model; specific adaptations for analysis of concentration time data are also available. NONMEM allows a general, correlated error structure with varying error magnitudes, for both interindividual and residual variability, that can be a function of the data and the fixed effect parameters as well as random effect parameters. The analysis of data using NONMEM requires the development of a pharmacostatistical model for each of the three types of population parameters to be estimated. This chapter provides an overview of the modeling process required for analyzing observational pharmacokinetic data using NONMEM.

PHARMACOSTATISTICAL MODELING

A complete pharmacostatistical model must be specified explicitly prior to analysis using a mixed effect model. This pharmacostatistical model includes a pharmacokinetic model for generating predicted drug concentrations as a function of individual pharmacokinetic parameters, and additional models for the mean and magnitude of variability of the pharmacokinetic parameters and the magnitude of residual variability. These models depend on population parameters, θ, Ω, and σ^2. The models can be divided into two basic types: structural models, including the pharmacokinetic model and regression formulae for investigating the effect of various fixed effects (coefficients = θ) on pharmacokinetic parameters; and statistical models for variability, including both interindividual variability (variance = Ω) and residual variability (variance = σ^2).

Pharmacokinetic Models

Pharmacokinetic models describe concentration time data. Such models generate predictions of concentrations at appropriate times as a function of their (pharmacokinetic) parameters, doses, times of observation, etc. The specification of the pharmacokinetic model employed in the analysis of experimental data is relatively straightforward since the method of drug administration is standardized and concentrations are generally obtained following a single dose, or over a single dosing interval while the patient is at a steady-state condition. An example of this process is illustrated in Figure 1 in which the concentrations obtained following a single intravenous bolus dose have been fit to the one compartment model described in Equation 1,

$$\hat{C}p_t = \frac{D}{V} e^{-K \cdot t} \qquad (1)$$

where, $\hat{C}p_t$ is the expected plasma concentration at time, t, following the administration of a single intravenous bolus dose, D, and K and V are the true kinetic parameters for the individual. In the process of fitting the data, of course, values for the parameters, volume of distribution, V, and the first order elimination rate

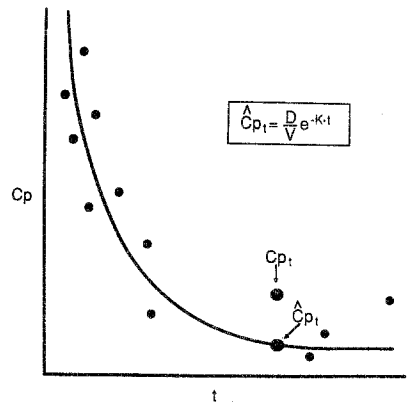

Figure 1. Fit obtained using a one compartment model to fit concentration time data observed following intravenous bolus administration of a drug. The Cp designates the actual measured concentrations and $\hat{C}p$ represents the concentrations predicted by the "true" pharmacokinetic model.

constant, K, would be selected to minimize the difference between the observed value and that predicted by Equation 1.

One of the challenges arising in the analysis of observational data has been the specification of the pharmacokinetic model. Observational data is collected under a variety of circumstances; data may be collected at steady-state and non-steady-state conditions, concentrations may be obtained over several dosing intervals, dosing intervals may be irregular and missing doses are a common problem in the clinical setting. As a result, the pharmacokinetic model is cumbersome and difficult to formulate. The model must be written in a recursive format so that the solution can be advanced from one dosing or blood level event to the next, in time order, so as to generate predictions of drug concentrations at times of corresponding measured values. In the past, the data analyst was responsible for providing the program subroutine, PRED, which specified the pharmacokinetic model to be used for the NONMEM analysis. This task represented a significant obstacle to widespread use of the program. Recently, this task has been made considerably easier with the release of PREDPP, a library of pharmacokinetic models for use with NONMEM.[1] PREDPP contains subroutines for analyzing data with either one or two compartment linear models using either first order or zero order drug administration. These models encompass the majority of situations encountered in pharmacokinetic studies. In addition, general linear and non-linear models are available for more complex data analysis tasks.

Regression Models

The primary objective of the analysis of either experimental or observational data is to determine the typical (mean) values for pharmacokinetic parameters in a patient population and identify independent variables affecting these parameters to a clinically significant degree. This evaluation has traditionally been carried out via regression of individual parameter estimates on independent variables such as age, weight, creatinine clearance, gender, and so forth.

This process can be illustrated by considering an experiment designed to investigate the relationship between renal function, as estimated by creatinine clearance, and the total clearance of a drug. Drug clearance and creatinine clearance would be determined for each subject in the experimental group using an appropriate experimental methodology. The resultant data would appear as in Figure 2, and linear regression analysis would be performed using Equation 2.

$$\tilde{Cl} = \theta_1 + \theta_2 \cdot CrCl \qquad (2)$$

where, CrCl is the observed creatinine clearance, the intercept, θ_1, is the drug clearance expected in a patient with no renal function, and θ_2 is the proportionality constant relating creatinine clearance to expected drug clearance, $\tilde{Cl}$.

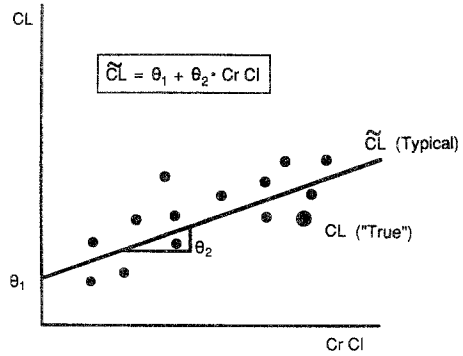

Figure 2. Linear regression analysis of drug clearance versus creatinine clearance (CrCl). Typical values of drug clearance are generated for an individual or group of individuals with a given creatinine clearance. The discrepancy between the true value for drug clearance (Cl) and the typical value $\tilde{C}l$ necessitates the use of a statistical model for interindividual variability.

The influence of multiple independent variables on pharmacokinetic parameters can also be determined by incorporating additional terms into the regression formula to yield, for example, Equation 3.

$$\tilde{C}l = \theta_1 + \theta_2 \cdot CrCl + \theta_3 \cdot AGE + \theta_4 \cdot WT \quad (3)$$

Likewise, categorical variables can be incorporated into the regression formula to evaluate the effect of all-or-none phenomena as in Equation 4.

$$\tilde{C}l = (\theta_1 + \theta_2 \cdot CrCl + \theta_3 \cdot AGE + \theta_4 \cdot WT) \cdot (1\text{-}CHF \cdot \theta_5) \quad (4)$$

where, CHF is an indicator variable which has the value unity if the patient has congestive heart failure and zero if the patient does not have the disease, and θ_5 represents the fractional increase or decrease in clearance associated with the presence of congestive heart failure. Indicator variables are useful for factors that are categorized, e.g., "present" versus "absent." While this does not usually present a problem in evaluating the effect of gender, it can present problems in evaluating the

effect of disease states with a range of severity. Careful consideration is required in determining the severity of disease that must be present before the patient is classified as having the disease "present."

The specific regression formulas used for investigating the influence of patient factors can be as complicated as necessary to appropriately model the effect of the factor on pharmacokinetic parameters. The models do not necessarily have to be linear or additive; further illustrations of regression modeling in population analyses are available.[1] The interested reader is referred to the NONMEM documentation for additional examples and implementations within the NOMEM program.[1]

Statistical Models

When data collected from an experimental pharmacokinetic study is to be fit to a pharmacokinetic model, a number of statistical assumptions are implicit. These assumptions concern the nature of the errors that arise between the measured drug concentrations and the predicted drug concentrations generated by the pharmacokinetic model. In using ordinary least squares to fit data, it is assumed that the errors are additive, independent, and the same typical size. Although these assumptions are frequently violated, particularly the assumption regarding the typical size of the errors, the quality and quantity of data available for each individual usually allows reasonable estimates of pharmacokinetic parameters. When the typical size of the errors varies in a known way, weighted least squares can be used to adjust for this. Unlike experimental data, however, observational data lack design restrictions regarding data quality, quantity, and organization (See Chapter 1 for further discussion). The lack of these characteristics produce violations of the above assumptions to the extent that traditional weighting schemes may be inadequate.[2]

The statistical model required for a NONMEM analysis of observational data allows explicit statements regarding the underlying distribution of the random error terms for interindividual and residual variability. In order to illustrate these components of the statistical model we can consider this problem from the perspective of the individual analyses discussed above. In the case of interindividual variability, and using Figure 2 as an example, the model focuses on the errors between "observed" values of clearance in each subject and the expected value of clearance for an individual obtained from his creatinine clearance and the regression formula in Equation 2. In the case of residual variability, and using Figure 1 as an example, the statistical model focuses on the errors between observed drug concentrations and the corresponding concentrations predicted by the pharmacokinetic model given the individual's pharmacokinetic parameters.

The requirement for an explicit statistical model for both interindividual and residual variability can be appreciated if one considers the differences between the analysis of individual data and data analysis from a population perspective. In fitting data from an individual, the parameter values of the pharmacokinetic model are chosen such that the deviations between the observed concentrations in the individual

and those predicted by the model are minimized. This is the process illustrated in Figure 1 where $(C_p-\hat{C}_p)^2$ is the quantity minimized, and $\hat{C}_p$ is understood to be the value of Cp when the individual's true parameters are used in Equation 1.

When data are analyzed from a population viewpoint using a mixed effect model, the concentration predicted by the pharmacokinetic model is generated using the typical values of pharmacokinetic parameters for the population. These are the values generated by the regression formulae, e.g., Equation 2, given the values of the independent variables. As a result, the discrepancy between the measured and predicted value has two components. The first component is the difference between the measured drug concentration in an individual and the predicted drug concentration that would be obtained if the individual's true pharmacokinetic parameters were known and used in the pharmacokinetic model to generate the predictions. This represents residual variability and is the difference minimized in the case of fitting individual data. The second component is the difference between the concentrations obtained using the individual's true parameters and those obtained using the pharmacokinetic parameters of the regression formulae, such as Equation 2. This difference represent interindividual variability and is unique to population analysis. The two components of the residual differences encountered in population analysis are graphically illustrated in Figure 3. The statistical models to be specified for MEM analysis can be thought of as models for residual variability and inter-individual variability.

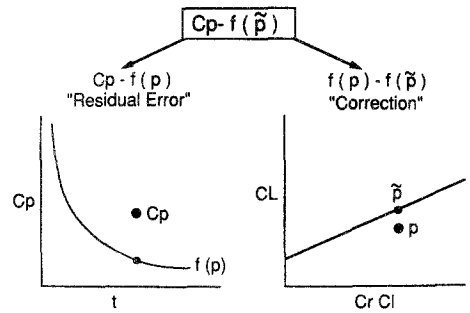

Figure 3. In a population analysis, the discrepancy between the measured and predicted drug concentrations can be dissected into two distinct components: residual variability, the difference between the observed, Cp, and expected concentrations, f(P), where P is the individual's true pharmacokinetic parameter; and interindividual variability, the difference between expected concentrations obtained using true parameters (P) versus typical values ($\tilde{P}$) from regression formulae.

Interindividual Variability. The regression formula in Equation 2 and the estimates for the intercept, θ_1, and the slope, θ_2, provide a mechanism for predicting the typical value of clearance in a patient or group of patients with a given creatinine clearance. In any specific individual, the typical value of drug clearance predicted for that individual using his creatinine clearance will not exactly equal his observed drug clearance. This discrepancy can be expressed as:

$$Cl = \tilde{Cl} + \eta \tag{5}$$

where Cl represents the true value of clearance for the individual, $\tilde{Cl}$ is, as noted above, the typical value predicted by Equation 2 for a given creatinine clearance, and η represents the persistent difference between these values. The distribution of the η's, for various individuals represents the distribution of their true clearances about the typical (predicted) value, and the mean squared value of η in the population, ω^2, is the variance of η and represents the magnitude of interindividual variability in clearance that has not been explained by the regression formula in Equation 2.

In addition to the known patient factors, such as age, weight, creatinine clearance, etc., which can affect a pharmacokinetic parameter one can consider a number of unknown factors, such as diet, genetic influences and environmental exposures which unpredictably affect drug clearance in the individual and result in the discrepancy between the true and typical values. In formulating the statistical model for interindividual variability, the discrepancy, η, is assumed to be a random variable that has a symmetric distribution with a mean of zero and variance, ω^2 (see Figure 4). For every individual in the population a different η arises from this distribution, interacts with the typical value, and generates the true value. It is this interaction between η and the typical value that must be modeled.

One of two basic statistical models are almost always used to model the interaction between η and the typical value. Equation 5 represents an additive model in which η simply adds to the typical value, and the variance of the parameter remains constant over the range of the independent variable. This presumes that the distribution of true values around the typical value has a constant degree of variability. Alternatively, one can write a model that has the variance about the typical value increasing with increasing values of the parameter. An example is the constant coefficient of variation model described in Equation 6:

$$Cl = \tilde{Cl} + \tilde{Cl} \cdot \eta \tag{6}$$

In this model the variance of Cl increases with increasing values of $\tilde{Cl}$ (it is $\tilde{Cl}^2 \cdot \omega^2$ to be exact). These models and the relationship between the variance of the parameter and its typical value implied by their use are illustrated in Figure 5.

The selection of a model for interindividual variability for a specific data

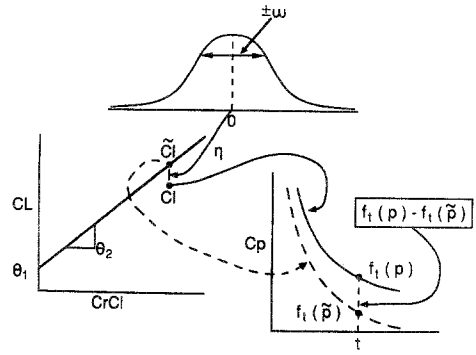

Figure 4. For each individual in the population, an η arises from a symmetric distribution with a mean of zero and variance of ω^2. This η interacts with the typical value to generate the true value and the nature of this interaction must be included in the pharmacostatistical model.

analysis problem should be based on the expected pattern of variability. If the population is reasonably homogeneous, an additive model is probably appropriate. If, however, there is a large degree of variability, either in the range of the independent variables affecting a parameter or in the true parameter values themselves, the use of the constant coefficient of variation model is probably more appropriate. It is also possible to formally compare the fit obtained with each model.[3] These basic models for interindividual variability can be further modified to account for differences in the magnitude of variability for different patient groups. For example, the interindividual variability in clearance may be greater in patients in an intensive care unit than in patients treated on a general medical service. The interested reader is referred to the NONMEM documentation for additional examples and implementation within the NONMEM program.[1]

Residual Variability. In the analysis of concentration time data from an individual, the measured drug concentrations will vary around the predicted concentration time curve as shown in Figure 1. The concentration of a drug measured at a given time following a dose is a function of the individual's pharmacokinetic parameters, such as clearance and volume of distribution, and a number of factors which can be considered to act in a random manner. These unknown and unpredictable influences include the variability introduced by the drug

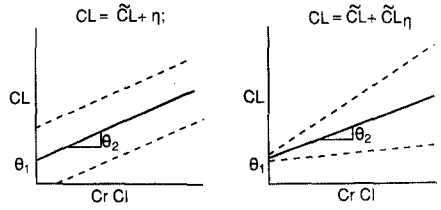

Figure 5. Basic statistical models for interindividual variability.

assay during the determination of the drug concentration, variability introduced by errors in the recorded time of sampling, and pharmacokinetic model misspecification. This latter problem can arise from a number of sources including intraindividual variability in pharmacokinetics, and selection of an inappropriate or incomplete pharmacokinetic model. Intraindividual variability in pharmacokinetics can result from true day to day variability in drug elimination efficiency. A linear pharmacokinetic model may have been used to model the concentration time profile of a drug but elimination may be nonlinear with concentration dependent clearance. Finally, a simple one compartment model may be inadequate. Enterohepatic recirculation of drug may be occurring or samples may have been obtained during an unappreciated distribution phase following rapid intravenous administration of drug. Each of these factors will add to the variability in the observed drug concentrations and produce discrepancies between observed and predicted values. These discrepancies can be expressed as

$$Cp = \hat{Cp} + \varepsilon \tag{7}$$

where, Cp represents the measured drug concentration, $\hat{Cp}$ is the corresponding concentration predicted by the pharmacokinetic model, (using the true individual parameters) and ε represents the difference between these values. The variance of ε (its mean squared value), σ^2, provides an estimate of the typical magnitude of squared residual variability.

In specifying the model for ε it is assumed that the unknown influences, i.e., those not incorporated into the pharmacokinetic model, are random and that they arise from a symmetric probability distribution as illustrated in Figure 6. From this distribution, with a mean of zero and a variance of σ^2, a different value for ε arises for each measured concentration in each individual in the population. This value of ε interacts with the expected concentration to generate the measured concentration. The statistical model for residual variability describes the nature of the interaction between

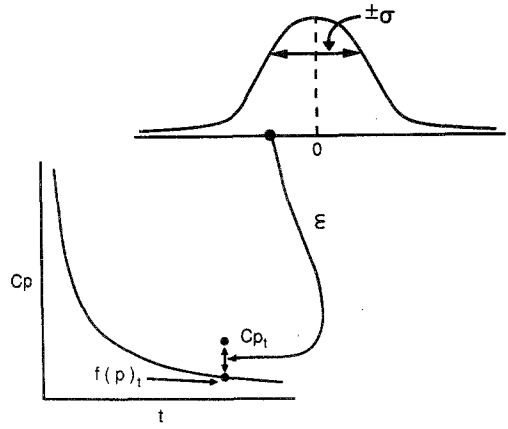

Figure 6. In modeling residual variability, errors between the measured and expected concentrations, obtained using the individual's true pharmacokinetic parameter, p, are assumed to arise from a symmetric distribution with a mean of zero and a variance of σ^2. The statistical model for residual variability defines the nature of the interaction between the error and the expected value to yield the observed value.

the error term ε and the expected drug concentration.

Similarly to interindividual variability, one of two basic models for ε usually suffices. Equation 7 represents an additive model in which ε adds to the expected drug concentration, and the variance of Cp, σ^2, remains constant over the range of expected concentrations. Alternatively, the variance of Cp may increase with increasing values of the expected concentration. Equation 8 is an example of a constant coefficient of variation model, applied to residual variability.

$$Cp = \hat{Cp} + \hat{Cp} \cdot \varepsilon \tag{8}$$

As in the model for η, the selection of a particular model for ε should be based on the expected pattern of errors. The relationship between the variance of Cp and the expected drug concentration implied by the use of basic models for residual variability are illustrated in Figure 7.

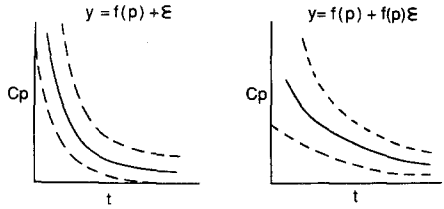

Figure 7. Basic statistical models for residual variability.

SUMMARY

The analysis of observational data using a mixed effect model is more complex than the traditional approach to pharmacokinetic parameter estimation. Structural and statistical models must be explicitly formulated and implemented within the computer program NONMEM. This presents some unique challenges in the selection and evaluation of alternative models for the evaluation of interindividual and residual variability.

REFERENCES

1. S.L. Beal and L.B. Sheiner, "NONMEM User's Guides," NONMEM Project Group, University of California, San Francisco (1989).
2. L.B. Sheiner and S.L. Beal, Analysis of nonexperimental pharmacokinetic data, *in*: "Drug Absorption and Disposition," K.S. Albert, ed., Washington, American Pharmaceutical Association, Chapt. 3, pp. 31-49, (1980).
3. D.R. Mungall, et al., Population pharmacokinetics of racemic warfarin in adult Patients, *J. Pharmacokinet. Biopharm.* 13:213 (1985).