POPULATION PHARMACOKINETIC AND PHARMACODYNAMIC MODELING

Roger Jelliffe, Alan Schumitzky, Aida Bustad, Michael Van Guilder, Xin Wang, and Robert Leary*

1. INTRODUCTION

As we acquire experience with the clinical and pharmacokinetic behavior of a drug, it is very desirable to capture this experience and its related information in the form of a population pharmacokinetic model, and then to relate the behavior of the model to the clinical effects of the drug or to a linked pharmacodynamic model. The purpose of population modeling is thus to describe and capture our experience with the behavior of a drug in a certain group or population of patients or subjects in a manner that will be useful for the treatment of future patients. This is true not only for clinical patient care, but also to optimize each step of drug development, to develop the optimal understanding of drug behavior, so that the next step can be taken most intelligently.

The traditional method of Naive Pooling has been used for population modeling when experiments are performed on animals, for example, which must be sacrificed to obtain a single data point per subject. Data from all subjects is then pooled as if it came from one single subject. One then can estimate pharmacokinetic parameter values, but cannot estimate any of the variability that exists between the various subjects making up the population. This method has generally been supplanted by the more informative methods described below.

2. PARAMETRIC POPULATION MODELING METHODS

A variety of parametric population modeling methods exist and have been very well described in [1]. They obtain means and standard deviations (SD's) for the pharmacokinetic parameters and correlations (and covariances) between them. Only a few of these will be described in this chapter, and quite briefly.

*Roger Jelliffe, Alan Schumitzky, Aida Bustad, Michael Van Guilder, Xin Wang, Laboratory of Applied Pharmacokinetics, USC School of Medicine, Los Angeles CA 90033. Robert Leary, San Diego Supercomputer Center, UCSD, San Diego CA 92093.

2.1. The Standard Two-Stage (S2S) approach

This approach involves first using a method such as weighted nonlinear least squares to obtain pharmacokinetic model parameter estimates for each individual patient. Correlations between the parameters may also be obtained, based on the individual parameter values in the various subjects. In the second and final step, the population means, SD's, and correlation coefficients in the sample of people studied are then computed for each of the pharmacokinetic parameters. This method usually requires at least one serum concentration data point for each parameter to be estimated.

One can also examine the frequency distributions of the individual parameter values to see if they are Gaussian or not. In this latter setting, the S2S method can also be regarded as being, in a sense, nonparametric as well, since no assumptions need to be made concerning the shape of the frequency distribution of the various individual parameter values. The method is basically parametric, however, as it gathers together the individual results into Gaussian summary parameter values of means, SD's, and correlations for each of the model parameters. This is what is meant by parametric population modeling - the frequency distributions of parameters in the model are described in terms of the parameters of an assumed function (with its specific distribution parameters such as means and SDs) that describes the assumed shape/class of the model parameter distributions. In this case, since the parameter distributions are usually assumed to be Gaussian or lognormal, these other distribution parameters are the means, SD's, variances and covariances, of the various pharmacokinetic – pharmacodynamic (PK / PD) parameters of the structural PK / PD model employed.

2.2. The Iterative Two-stage Bayesian (IT2B) method

This method can start by using the S2S mean parameter values as obtained above, and their SD's. On the other hand, one can set up any reasonable initial estimate of the population mean parameter values and their SD's. In the IT2B method, one uses these initial selected parameter means and SD's as the Bayesian priors, and then examines the individual patient data to obtain each patient's maximum a posteriori probability (MAP) Bayesian posterior parameter values, using the MAP Bayesian procedure in current wide use [2]. It uses the First Order Conditional Expectation (FOCE) approximation to calculate the log–likelihood of the population parameter values given the population raw data and the weighting scheme employed in analyzing the data.

With this method, one can iteratively recompute the population means and SD's of the parameter values found. For example, one can use these S2S summary population parameter values (see above) once again, now as initial Bayesian population priors, for another MAP Bayesian analysis of the data. One can once again obtain each patient's new MAP Bayesian values. This process can then continue iteratively indefinitely. The procedure ends when a convergence criterion is reached. The IT2B method is less subject to the problems of local minima often found when fitting data by least squares. In addition, it does not require as many serum concentration data points per patient (as few as only one per patient), and so is much more efficient in this respect. The Global Two Stage (G2S) method is a further refinement of the S2S and the

IT2B in which the covariances and correlations between the parameters are considered during the process of parameter estimation.

2.3. The Parametric EM method

This method is also an iterative method. The letters EM stand for the two steps in each iteration of 1) computing a conditional expectation (E) and 2) the maximization (M) of a conditional likelihood, resulting in a set of parameter values which are more likely than those in the previous iteration. The process continues until a convergence criterion is met. The results with the parametric EM method for the various pharmacokinetic parameter distributions are again given in terms of the model parameter means, SD's, and correlations, or means, variances, and covariances [3, 4]. As is the case with the IT2B method, an approximation such as FOCE is used to compute the conditional likelihoods to avoid computationally intensive numerical integrations.

2.4. NONMEM

True population modeling began with the Nonlinear Mixed Effects Model with first-order approximation (NONMEM) of Beal and Sheiner [5-7]. The overall population, even if it has only 1 data point per subject, almost always supplies enough data for this approach, if the various data points are spread throughout the dosage interval so that dynamic information about the behavior of the drug can be obtained. The NONMEM method estimates means, SD's, and covariances of population parameter values. However, it has sometimes given different answers from other methods. It is also a parametric method, and gives its results in terms of parameter means and variances.

This method was the first true population modeling program, as it eliminated the need for having at least one data point for each patient for each parameter to be estimated. It estimates both fixed effects (those containing only a single point value for a parameter, such as a parameter mean), and those containing random distributions , such as the random variability of a model parameter about its mean. This random variability is characterized by SD's, and covariances or correlations. The name "mixed" is used because the method estimates both types of model parameters, fixed and random. The method can function with as few samples as one per patient. While this method is in wide use, it lacks the desirable property of mathematical consistency [8-10]. Earlier (FO) versions of this method have at times given results which differed considerably from those of other methods [11,12]. Subsequent First Order Conditional Expectation (FOCE) versions of this method have shown improved behavior. This will be discussed further toward the end of this chapter. Other variations on this approach are those of Lindsdtrom and Bates [13], and Vonesh and Carter [14].

3. ANALYZING ASSAY AND ENVIRONMENTAL SOURCES OF ERROR

3.1 Determining the Assay Error Polynomial

In analyzing any data set, it is useful to assign a measure of credibility to each data point to be fitted or analyzed. In the IT2B program of the USC*PACK collection [16,42], for example, one is encouraged first to determine the error pattern of the assay quite specifically, by determining several representative assay measurements in at least quadruplicate, and to find the standard deviation (SD) of each of these points. One can measure, in at least quadruplicate, a blank sample, a low one, an intermediate one, a high one, and a very high one. One can then fit the relationship between the serum concentration (or other response) and the SD with which it has been measured, with a polynomial of up to third order, so that one can then compute the Fisher information, for example, as a useful measure of the credibility of each serum concentration data point [15,16]. One can then express the relationship as

$$SD = A_0 + A_1C + A_2C^2 + A_3C^3$$

where SD is the assay SD, A_0 through A_3 are the coefficients of the polynomial, C is the measured concentration, C^2 is the concentration squared, and C^3 is the concentration cubed. A representative plot of such a relationship, using a second order polynomial to describe the error pattern of an EMIT assay of gentamicin, is shown in Figure 1.

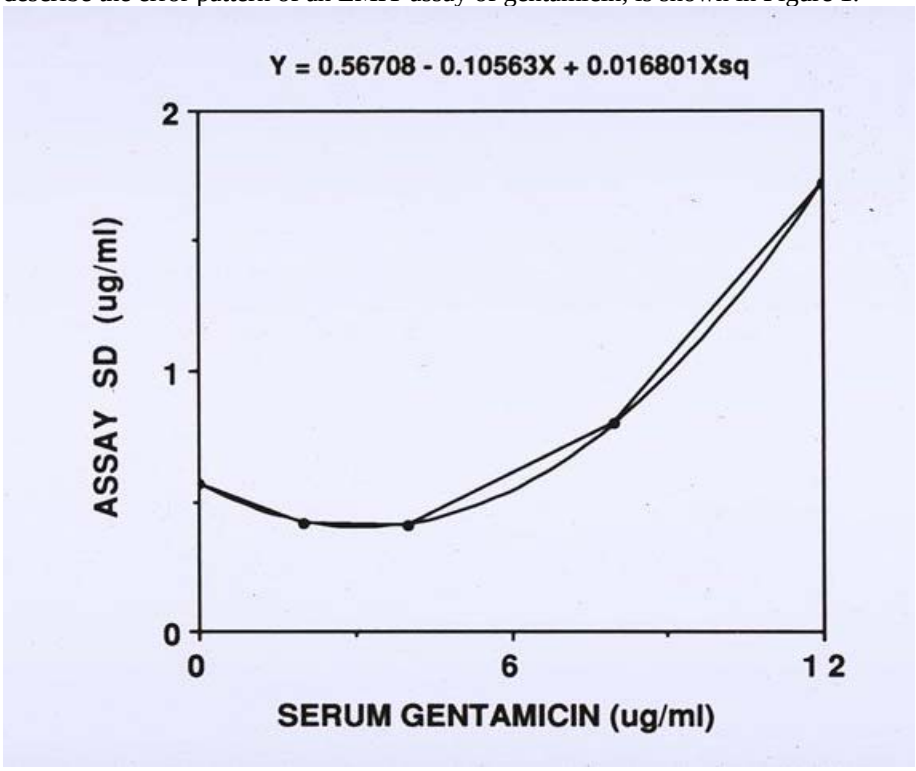

Figure 1. Graph of the relationship between serum Gentamicin concentrations, measured by Emit assay in at least quadruplicate (the dots) and the standard deviations (SD's) of the measurements. The relationship is captured by the polynomial equation shown at the top. Y = assay SD, X = measured serum concentration, Xsq = square of serum concentration.

3.2 There is no "Lower Limit of Quantification" (LOQ) for Pharmacokinetic Modeling

When a serum sample is assayed to determine the concentration of drug present in it, the issue often arises as to what constitutes the lower limit of quantification (LOQ) of the assay. This is usually taken as being a certain amount above the result of a blank determination, such as 2 standard deviations (SD's) above the blank, for example. A problem arises when a result is in the gray zone, below the LOQ but a little above the blank, and there has been a great deal of discussion about what the best thing is to do about this problem. Some sat it should be set to zero, others say perhaps it should be set to halfway between the blank and the LOQ, for example. Commonly, laboratories have reported the result simply as being "less than" whatever the LOQ is.

This is fine for purposes of toxicology, where the sample itself is the only source of information about the drug concentration. The question being asked here is whether or not the drug is present in the sample or not. This question cannot be answered satisfactorily unless the measured concentration is above the LOQ. If not, the answer is in doubt.

However, when doing therapeutic drug monitoring or any pharmacokinetic/dynamic modeling, this is a most unsatisfactory result. It simply cannot be used in any procedure to fit data or to make a population model.

In doing either therapeutic drug monitoring or population modeling, there is always more information present than just what the measured sample concentration happens to be. For example, we usually know fairly well when the serum sample was obtained, and also when the various doses, especially the last dose, was given. Because of this, and because the great majority of drugs have half-times of decay in their concentrations, we actually know that the drug is actually still present in the body when the sample is obtained, and the real question is not whether or not it is present, but rather how much drug is present. This is a totally different question, and by using the assay error polynomial described above, one can give correct weight to any measured concentration, all the way down to and including a blank. The only thing required is that the meter recording the assay result have the ability to give either positive or negative results close to the blank, to permit the SD of the blank to be correctly determined. Then, if a result of zero or a negative result is obtained, it is probably a concentration of essentially zero, and the correct credibility (the assay SD at the measured concentration) can be given to that data point in the fitting procedure. Because of this, we feel that a clinical laboratory should not receive credit (or payment) for reporting a result as "less than" the LOQ when this that sample was obtained for purposes of therapeutic drug monitoring or pharmacokinetic modeling.

It is easy to do better than this, and to make both the toxicologists and the pharmacokineticists happy at the same time, by reporting the result both ways. For example, a gentamicin sample might be reported as having a measured concentration of "0.2 ug/ml, below our usual LOQ of 0.5 ug/ml". Then both parties can have what they

need for their work. The toxicologists know that the concentration is quite low, and when the sample is all the information they have, that it is not clear that any significant concentration of drug is present. On the other hand, those doing therapeutic drug monitoring have exactly what they need, a measured concentration, which can be given its appropriate weight in the fitting procedure.

It is interesting that so much attention is paid to determining the error of assays, when once the assay is shown to be acceptably precise, the error is simply forgotten or neglected. For example, many error models simply use the reciprocal of the assay variance, and forget the actual error of the assay. This is usually done because it is assumed that the assay variance in only a small part of the overall error variance, due to the many other significant remaining environmental sources of error.

3.3 The SD is the key, not the Coefficient of Variation

Most laboratories are accustomed to describing their assay errors in terms of the coefficient of variation (CV). They point out that as the concentration approaches zero, the CV becomes infinite, and say that "the signal becomes lost in the noise". This is actually not so. While the CV may become infinite, the SD never does. It always remains finite. Because of this, the SD is the key here, not the CV, and the sample result can always be given its correct weight as determined by the assay error polynomial. This is clearly borne out by the plot in Figure 1, for example. When assay errors are described by such polynomials, then, there clearly is no LOQ for pharmacokinetic analyses of therapeutic drug monitoring. There is no need for clinical laboratories to report a serum digoxin concentration as "less than 0.2 ng/ml', for example, and insurance companies should not be asked to pay when such a result is reported.

3.4 Determining the Remaining Environmental Error

In addition, a parameter which we have called gamma, a further measure of all the other environmental sources of intra-individual variability, can also be computed. It is used in the USC*PACK IT2B program as a multiplier of each of the coefficients of the assay error polynomial as described above. The nominal value of gamma is 1.0, indicating that there is no other source of variability that the assay error pattern itself. Gamma is therefore usually greater than 1.0. It includes not only the various environmental errors such as those in preparing and administering the doses, recording the times at which the doses were given, and recording the times at which the serum samples were obtained, but also the errors in which the structural model used fails to describe the true events completely (model misspecification), and also any possible changes in the model parameter values over time, due to the changing status of the patient during the period of data analysis. Gamma is thus an overall measure of all the other sources of intraindividual variability besides the assay error. In this way, one can calculate how much of the total SD is due to the assay SD, and how much is due to the remaining overall environmental SD.

Determining gamma will help to explain the environmental variability found in any fit. If gamma is small, it suggests that the sum of the environmental sources of noise is small. If it is large, it suggests that the overall environmental noise, the total effect of all the other factors mentioned above, is large.

However, most of these other sources are not really sources of measurement noise, but are rather due to noise in the differential equations describing the behavior of the drug. The environmental sources are most correctly described as sources of process noise rather than measurement noise. The problem is that it is difficult to estimate process noise, as it requires stochastic differential equations, which contain these other noise terms. However, no software for estimating process noise in pharmacokinetic models exists at present, to our knowledge.

The IT2B program can also be used to compute estimates of the various combined assay error and environmental polynomial error coefficients, if one has no knowledge of what the assay error pattern is, or if the measurement is one which is impossible to replicate to determine its error. In this case, gamma is not determined separately, but is included in the various other polynomial coefficients.

3.5 Evaluating the Degree of Therapeutic Environmental Noise

Using the assay error polynomial first, and then determining the remaining environmental error term as gamma now permits us to know just what fraction of the overall error is due to the assay (unavoidable) and what fraction is due to environmental noise factors discussed above. Gamma often is in the range of 2 to 4 when good quality clinical studies are done, showing that the overall environmental noise is 2 to 4 times that of the assay. If gamma is 10, however, that might well suggest that significant environmental noise is present in the study, and that attention should be given to these factors. In addition, when used with appropriate skepticism, the value of gamma can be used as an index of the quality of care the patients have received in any particular clinical setting. For patients who are approximately stable, comparisons can be made in this regard. Different ward services can be evaluated with regard to the precision of the therapeutic environment they provide for the patients under their care.

4. MAKING A PARAMETRIC (IT2B) POPULATION MODEL

The following results are taken from a representative run of the IT2B program. The original patient data files were made using the USC*PACK clinical software. The following illustrative results are taken from data obtained by Dr. Dmiter Terziivanov in Sofia, Bulgaria [17], on 17 patients who received intramuscular Amikacin, 1000 mg, every 24 hours for 5 or 6 days. For each patient, two clusters of serum concentrations were measured, one on the first day and the other on the 5th or 6th day, approximately 5 samples in each cluster. Creatinine clearance (CCr) was estimated from data of age,

gender, serum creatinine, height and weight [18]. Serum concentrations were measured by a bioassay method. The assay error pattern was described by a polynomial in which the assay SD = 0.12834 + 0.045645C, where C is the serum concentration. The assay SD of a blank was therefore 0.12834 ug/ml, and the subsequent coefficient of variation was 4.5645%. In this particular analysis, gamma was found to be 3.2158, showing that the SD of the environmental noise was about 3.2 times that of the assay SD, or conversely, that the assay SD was about 1/3 of the total noise SD.

The initial (very first) parameter estimates, and their SD's, were set at: Ka (the absorption rate constant) = 3.0 ± 3.0 hr^{-1}, Ks (the increment of elimination rate constant per unit of creatinine clearance in ml/min / 1.73M^2) = 0.004 ± 0.004 hr^{-1}, and Vs1 (the apparent central volume of distribution) = 0.3 ± 0.3 l/kg. The nonrenal intercept of the elimination rate constant (Ki) was held fixed at 0.0069315 hr^{-1}, so that the elimination rate constant = Ki + Ks1 x creatinine clearance, and the serum half-time, when CCr = 0, is fixed at 100 hours.

The following results were obtained with the USC*PACK IT2B program. The IT2B program converged, on this data set, on the 1053th iteration. The population mean values for the parameters Ka, Ks1, and Vs1 found were 1.349 hr^{-1}, 0.00326 hr^{-1}, and 0.2579 L/kg respectively. The medians were 1.352 hr^{-1}, 0.00327hr^{-1}, and 0.2591 L/kg respectively. The population parameter standard deviations were 0.062 hr^{-1}, 0.000485 hr^{-1}, and 0.0350 L/kg respectively, yielding coefficients of variation of 4.55, 14.83, and 13.86 percent respectively.

The individual MAP Bayesian distributions of Ka, Ks1, and Vs1 are shown in Figures 2 through 4. While the distributions of Ka and Vs1 are fairly Gaussian, that of Ks1 is skewed to the left. The joint distribution of Ks and Vs in shown in Figure 5, which shows an extremely high positive correlation between the two parameters, consistent with their population parameter correlation coefficient of +0.991. That between Ka and Ks1 was similar, +0.924, and that between Ka and Vs1 was also very high at +0.950. These probably spuriously high correlations are similar to thoise found by Leary [27] which are discussed in section 11 below.

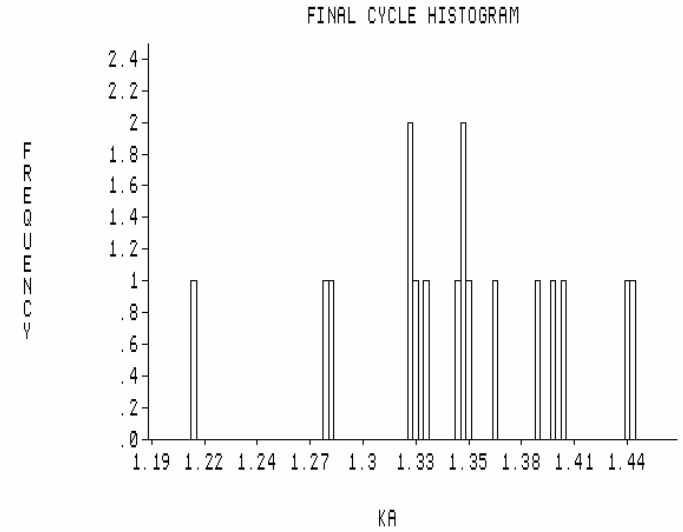

Figure 2. Graph of the marginal frequency of population parameter Ka. The plot is divided into 100 cells over the range from 1.19 to 1.47 (horizontal axis). The frequency of the patient parameter values in each cell is shown on the vertical. See text for discussion.

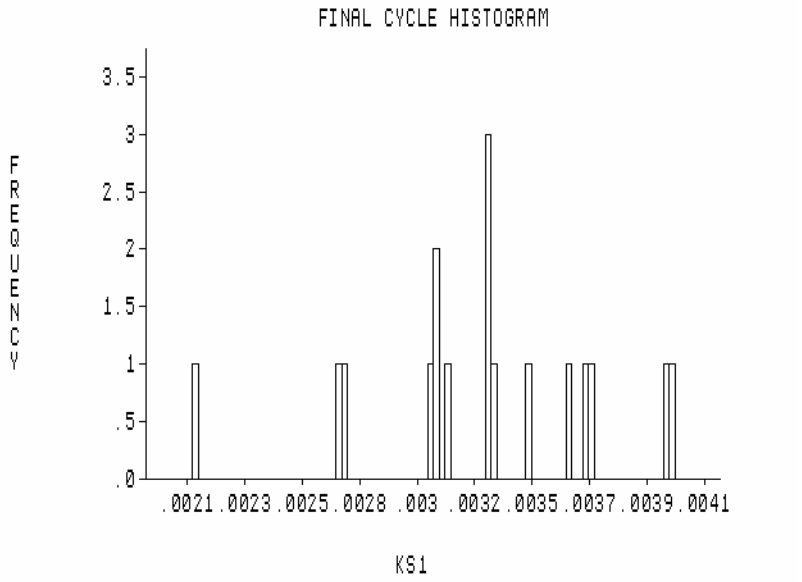

Figure 3. Graph of the marginal frequency of population parameter Ks1. The plot is divided into 100 cells over the range from 0.0019 to 0.0041 (horizontal axis). The frequency of the patient parameter values in each cell is shown on the vertical. See text for discussion.

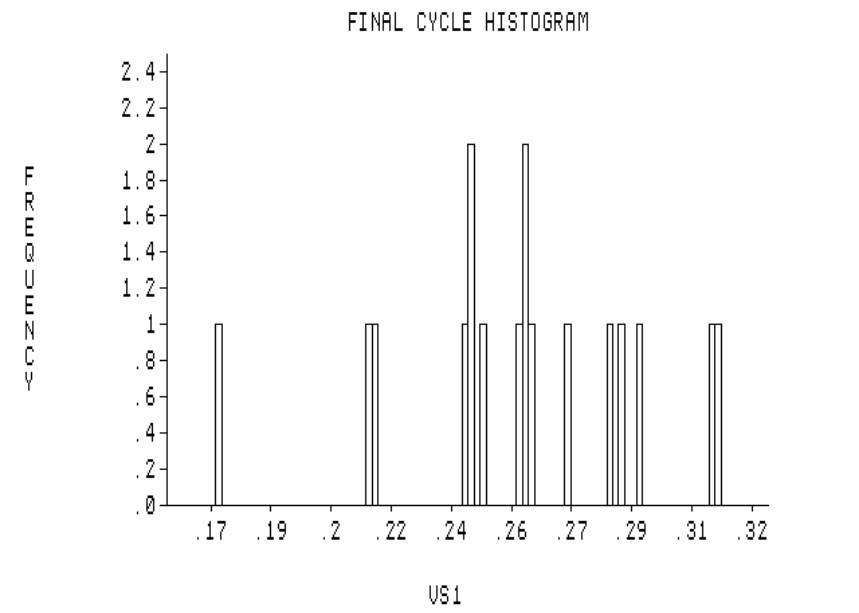

Figure 4. Graph of the marginal frequency of population parameter VS1. The plot is divided into 100 cells over the range from 0.15 to 0.32 (horizontal axis). The frequency of the patient parameter values in each cell is shown on the vertical. See text for discussion.

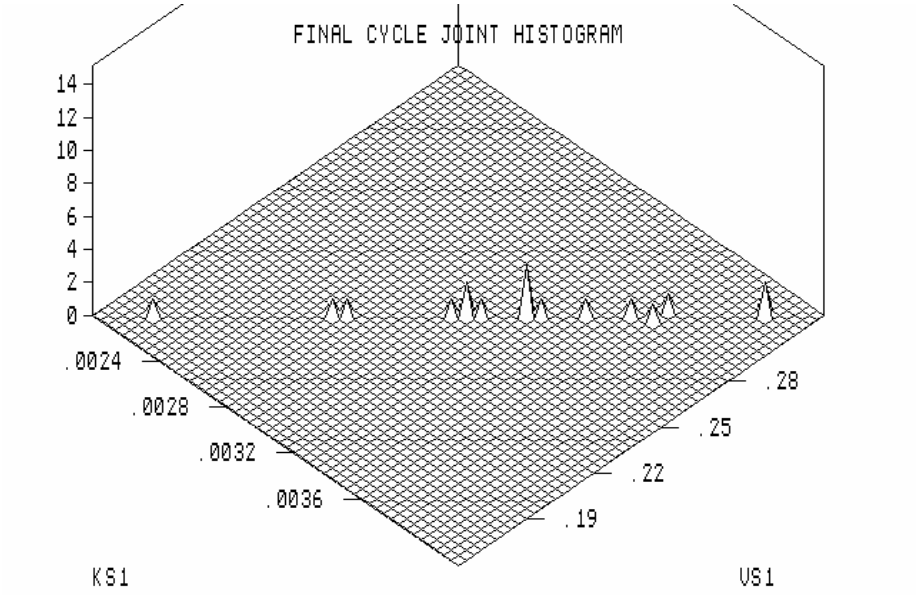

Figure 5. Graph of the joint marginal frequency of population parameter Ks1 and Vs1. The plot is divided into 50 by 50 cells over the ranges stated in Figures 2 - 4. The frequency of the patient parameter values in each cell is shown on the vertical. Note the extremely high (probably spurious) correlation between the parameters. The correlation coefficient was 0.991. See text for discussion.

Figures 6 and 7 are scattergrams of predicted versus measured serum concentrations. Figure 6 shows the predictions based on the population parameter medians and the doses each subject received. In contrast, Figure 7 shows the predictions made using each subject's individual MAP Bayesian posterior parameter values to predict only his/her own measured serum concentrations. The improved predictions in Figure 7 are due to the removal of the population inter-individual variability, as perceived by the IT2B program. The remaining smaller scatter is due to the intraindividual variability resulting not only from the assay error, but also to the other sources of noise in the system, such as the errors in preparation and administration of the various doses, errors in recording the times the doses were given and the serum samples drawn, and the mis-specification of the pharmacokinetic model used. The results shown in Figure 7 show that the study was done with reasonable precision.

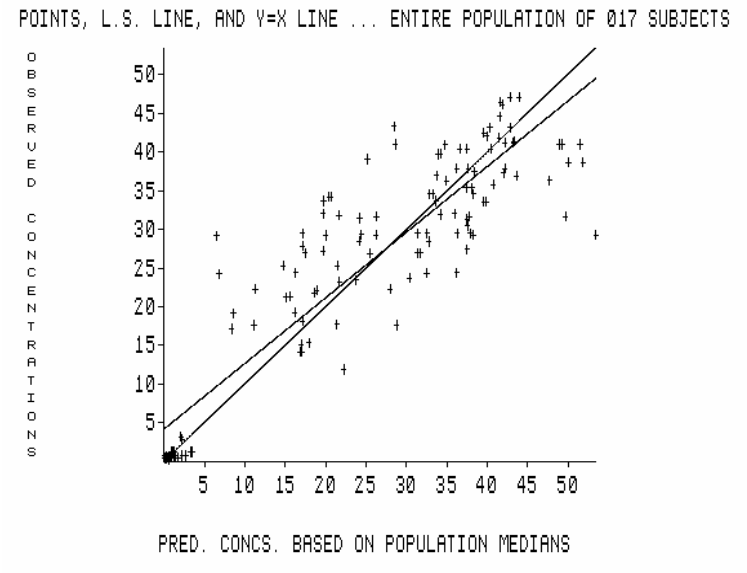

Figure 6. Scattergram of relationship between predicted serum concentrations (horizontal) and measured ones (vertical), based on median population parameter values.

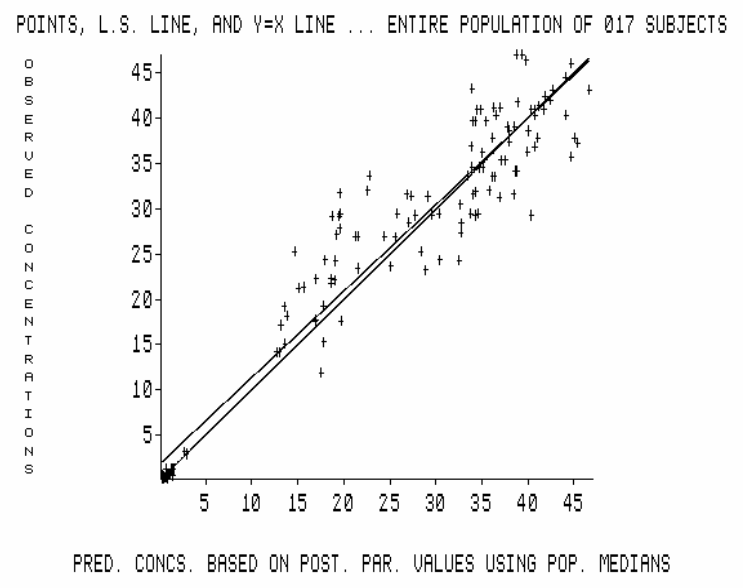

Figure 7. Plot of relationship between predicted serum concentrations (horizontal) and measured ones (vertical), based on each subject's own maximum aposteriori probability (MAP) Bayesian posterior parameter values, where each subject predicts only his/her own measured concentrations.

The IT2B method of population modeling is a useful one, and is based on the widely used and robust strategy of MAP Bayesian individualization of pharmacokinetic models. Its weaknesses, like those of any parametric method, are that it only perceives population parameter values in terms of their means, medians, variances, and correlations. The actual parameter distributions are usually not of this type. Lognormal assumptions have often been made, but the actual parameter distributions are frequently not of that form either. In addition, the true log-likelihood (not the FOCE approximation) of the distribution of the individual subjects' parameter values in the population was also computed exactly, just as if the discrete collection of individual parameter values had come from an NPEM or NPAG analysis (see below) for direct comparison with their results. It was found, for this IT2B analysis, to be -389.548 (see Table 1, below).

5. LARGER AND NONLINEAR IT2B POPULATION MODELS.

Similar software for IT2B population modeling of large and nonlinear PK/PD models has been implemented on the IBM "Blue Horizon" parallel computer at the San Diego Supercomputer Center (SDSC), as a research resource for such work, on a Linux cluster of Dell machines in our laboratory, and on a larger Linux cluster at USC. The user uses a PC program on his machine in the USC*PACK collection to specify the data files to be analyzed and the instructions for the analysis. This becomes an input file for the parallel machine or cluster. One also either writes the linear or nonlinear ordinary differential equations for the specific structural PK/PD model to be used, or

employs the BOXES program in the USC*PACK collection [42], placing boxes on the screen for the compartments and connecting them with arrows to represent the various types of pathways involved. The differential equations of the PK/PD model are then generated automatically and stored in a model file.

These two files are then sent to the parallel machine or cluster via a secure protocol, over the Web. The model file, which is in Fortran source code, is compiled and linked. The analysis is performed using the desired number of processors. A differential equation solver (VODE) is employed. The results are then sent back by email to the user's PC where they are examined just as in Figures 2 through 7 above. Thus one can now make large and nonlinear models of a drug, with multiple responses such as serum concentrations and various drug effects [20].

6. STRENGTHS AND WEAKNESSES OF PARAMETRIC POPULATION MODELING

The major strength of the parametric population modeling approaches has been their ability to separate inter-individual variability in the population from intra-individual variability in the individual subjects (gamma, for example), and also from variability due to the assay error itself. Because of this, it still seems best, for the present, to begin making a population model of a drug by using a parametric method such as IT2B. First, though, one should estimate the assay error pattern explicitly, obtaining the assay error polynomial as described above [15,16]. Then, having that assay error polynomial, one can use a parametric method such as IT2B to find gamma, to determine the overall intraindividual variability, and to know what fraction of that is actually due to the assay itself, and what is due to the environmental uncertainty. Then, one is in a position to overcome the weaknesses of these approaches (see below) by using this information in making a nonparametric population (NP) model.

One weakness of the parametric methods is that they generally have lacked the desirable property of mathematical consistency, which is a real strength of the nonparametric methods discussed later on [21-23]. In addition, the parametric methods make parametric assumptions about the shape of the parameter distributions, and do not take into account the entire actual shape of the distributions, as the nonparametric methods do [24,25]. Further, they give only single point summary parameter estimates such as the mean, median, or mode of each overall parameter distribution. Much discussion has taken place about which of these is the better estimate.

The major clinical weaknesses of the most widely used FOCE parametric approaches are that they only obtain the single point estimates of the parameter distributions, and that they are not statistically consistent, have poor statistical convergence, and poor statistical efficiency [27]. This will be discussed further on in more detail later on.

Further, when one uses such a model to develop a dosage regimen for a patient to achieve a desired target goal at a desired target time, the regimen is simply the one which should hit the desired target exactly. There is no method to estimate in advance the degree to which the regimen will fail to hit the target, as there is only a single

model, with each parameter consisting of only a single point estimate. Because of this, with parametric population models, the only course of action that can be taken (the next dosage regimen) based on that model, to hit a desired target goal, is based only on the central tendencies of the various parameter distributions, and not on the entire distribution itself. Such action (the dosage regimen) is therefore not designed to achieve the target goals optimally.

7. THE SEPARATION PRINCIPLE

The separation or heuristic certainty equivalence principle states that whenever the behavior of a system is controlled by separating the control process into:

1. Getting the best single point parameter estimates, and then,
2. Using those single point estimates to control the system,

the task of control is usually done suboptimally [26], as there is no specific performance criterion that is optimized. This is the major weakness of using parametric population models for designing drug dosage regimens. The parameter distributions are often neither Gaussian or symmetrical, and measures of central tendency are not optimal in computing the dosage regimens, and may occasionally be dangerous, as when the volume of distribution, for example, may be at the 70^{th} percentile of the distribution. Other thing being equal, that would result in 70% of the concentrations being above a desired target goal, and only 30% being below. Because of this, when single point parameter estimates are used to compute dosage regimens, one cannot estimate the expected precision with which a given dosage regimen will hit the target, and this is usually results in a suboptimal (less than maximally precise) dosage regimen. The resulting regimen is based only on the central tendencies of the various parameter distributions, and not on the shapes of the entire distributions themselves, which are often non Gaussian and multimodal. This problem will be discussed more fully in the chapter on clinical applications.

One may ask why we make models – to simply capture such single point estimates, or to take some useful and practical action based on the information obtained from the modeling process? It is useful and practical to supplement the knowledge of the assay error, empirically determined before starting the modeling, with information about the intra-individual variability (gamma, above) obtained from a parametric IT2B population model. Having this information, one can then proceed to make a nonparametric population model which can overcome the difficulties presented by the separation principle stated above. This will enhance the process of drug development, as it will help to optimize the design of dosage regimens for various populations of patients, to achieve the desired goals most precisely. It will also optimize the process of designing dosage regimens for various clinical trials, or designing the next step in drug development optimally. Again, this will be discussed in the chapter on clinical applications.

8. NONPARAMETRIC POPULATION MODELING

There is great variability among patients with regard to their pharmacokinetic parameter values. Nevertheless, we have become accustomed to using selected single numbers to summarize such diverse behavior. For example, we have usually used the population mean or median parameter values as the best single number to describe the central tendencies of their distributions, and the standard deviation (SD) to describe the dispersion of values about the central tendency. It has been customary to focus on such single numbers as summaries of experience with subjects or patients, rather than to consider the entire collection of our varied experiences with each individual patient. We will now examine newer nonparametric methods which can give us richer and more likely information from the raw population data.

What is meant here by the word nonparametric? Most of the time, when we gather data and summarize it statistically, we have been accustomed to obtaining a single parameter value to summarize the central tendency of a distribution such as the mean, median, or mode, and another single parameter value to describe the dispersion about this central tendency, such as the standard deviation (SD). The usual reason for this is that many events in statistics have a normal or Gaussian distribution, and that the mean and the SD are the two parameters in the equation which describe the shape of a Gaussian distribution explicitly. Because of this, describing a distribution parametrically, in terms of its mean and SD, is very common. Indeed, the entire concept of analysis of variance is based on the assumption that the shape of the parameter distributions in the system are Gaussian, and are therefore best described by means, SD's, and covariances. A great body of experience has been brought to bear to describe pharmacokinetic models parametrically, in this way. Much of this is described in an excellent review given in [1].

On the other hand, if the structural PK/PD model could be exactly known, if each individual subject's pharmacokinetic parameter values in a given population could also somehow be truly and exactly known, and if, for example, we were examining two such typical parameters such as volume of distribution (V), and elimination rate constant (K), then the truly optimal joint population distribution of these parameter values would be the entire collection of each individual patient's exactly known parameter values. All subpopulations, and those in between, would be truly known as well (perhaps not yet explicitly recognized or classified), but nevertheless located and quantified.

Obviously though, the individual subject parameter values can never be known exactly. They must be estimated from the data of doses given and serum concentrations measured, and in the setting of environmental uncertainty, as described above.

In the nonparametric approach, the maximum likelihood parameter distributions one obtains are discrete spikes, up to one for each subject studied in the population [24,25,27]. The location of each spike (support point) reflects its set of estimated parameter values. The height of the spike represents the estimated probability of that individual set of estimated parameter values. The likelihood of the entire collection of support points can be computed and compared under similar conditions (see Table 1 further on for an example). No summary parameters such as mean or SD will be any more likely, given the data of dosage and serum concentrations, than the

actual collection of all the estimated individual discrete points, each one having certain parameter values such as V and K, for example, and the estimated probability associated with each combined point of V and K [24,25,27].

This is what is meant by the word nonparametric in this sense. It is not to be confused with the noncompartmental modeling approach based on statistical moments, which is often also called nonparametric. The NP approach always has a specific structural model. In addition, as will be seen further below, the NP estimates of the parameter means SD's, and correlations are at least as reliable as those obtained by parametric methods, and not infrequently are significantly better than those obtained using the FOCE approximation [27]. The actual shape of the discrete NP parameter distribution is totally determined by the raw data of the subjects studied in the population and the error pattern used, and not by any assumed equation describing the assumed shape (Gaussian, lognormal, etc.) of the parameter distributions. The NP methods estimate the collection of support points and their probabilities which is most likely (maximum likelihood), that which specifically maximizes the likelihood or log-likelihood function. It is interesting that in a great number of papers which have described parametric population models, the likelihood of the results obtained (as opposed to common indices of "goodness of fit") is usually not reported.

Many patient populations actually are made up of genetically determined clusters or subpopulations. For example, there may be fast and slow metabolizers of a drug. The relative proportions of fast, in between, and slow subjects may vary from one population (Caucasian people, for example) to another (Asian people, for example) [28] Describing such a distribution of clusters optimally is not possible with a normal or lognormal distribution.

Since it is not possible to know each patient's values exactly in real life, we study a sample of patients requiring therapy with a drug (the most relevant population sample) by giving the drug and measuring serum concentrations and/or other responses. Lindsay [29] and Mallet [24] were the first to show that the optimal solution to the population modeling problem is actually a discrete (not continuous), spiky (not smooth) probability distribution in which no preconceived parametric assumptions (such as Gaussian, lognormal, multimodal, or other) have to be made about its shape. The nonparametric maximum likelihood (NPML) estimate of the population joint parameter density or distribution (analogous to the entire collection of each patient's exactly known parameter values described above), whatever its shape or distribution turns out to be, is supported by up to N discrete points for the N patients in the population studied. Each such support point is a collection of estimated single numbered parameter values, one for each parameter such as V, K, etc., along with an estimate of the probability associated with each such combination. The probabilities of all the various points add up to 1.0. The NPML methods of Mallet [24], like the parametric NONMEM method, can also function with only one sample per patient. However, as mentioned, the nonparametric parameter distributions may have any shape, and this depends only on the actual subject data. The means, SD's, and other common statistical summary parameters can easily be obtained as well, from the entire discrete distribution. The only assumption made about the shape of the discrete parameter distributions is that, for each model parameter, the shape, whatever it is, is the same for

all subjects in the population. Because of this, the method is capable of discovering unsuspected subpopulations of subjects such as fast and slow metabolizers, without recourse to other descriptors or covariates, and without recourse to individual Bayesian posterior parameter estimates [25].

A similar nonparametric EM (NPEM) method was developed by Schumitzky [25,30]. It is an iterative EM method like the parametric EM method, but is nonparametric. Like the NPML method, it also can function with only one sample per patient. Like the NPML method, it also does not have to make any parametric assumptions about the shape of the joint probability distribution. It also computes the entire discrete joint density or distribution of points. In contrast to the NPML method, though, the NPEM method obtains a continuous (although very spiky) distribution. This distribution becomes discrete in the limit, after an infinite number of iterations. Within each iteration, the NPEM method examines the patient data and develops a more and more spiky (and more likely) joint distribution. In the limit, the spikes become up to one discrete support point for each subject studied, just as with the NPML method. As with the NPML method, the NPEM population joint distribution also becomes a collection of discrete support points, each of which contains a set of parameter values, and each of which has a certain probability. Both the NPML and the NPEM methods have been shown to converge to essentially the same results [31]. Both the NPML and the NPEM methods are proven under suitable hypotheses to have the desirable property of mathematical consistency [8,23,27,32].

Figures 8 through 10 illustrate the ability of the nonparametric approach, as shown by the NPEM algorithm, to discover unsuspected subpopulations [25]. The NPML method of Mallet has similar capability. Figure 8 shows a carefully constructed simulated population of patients, which is actually a bimodal distribution consisting of two subpopulations. Half were "fast" and half were "slow" metabolizers of a drug. While they all had the same volume of distribution, they had two different elimination rate constants. There was no correlation between the two parameters.

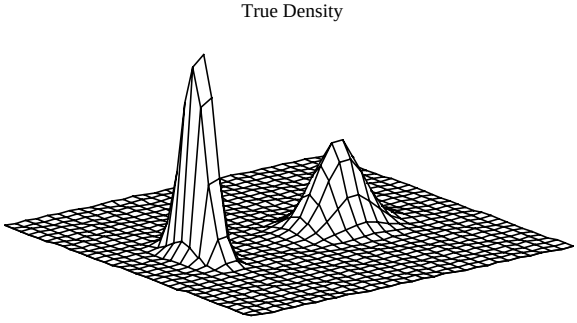

Figure 8. The true pharmacokinetic population joint density from which the 20 samples were taken. If the bottom (or 0.0,0.0) corner is "Home plate", then the axis toward third base is that of the volume of distribution V, while that toward first base is the elimination rate constant K. The vertical axis is the relative probability of each parameter pair. Note that there are actually two subpopulations, with two clusters of distributions for K. V and K are uncorrelated.

From this simulated population, twenty hypothetical patients were sampled at random. Their parameter values were therefore known exactly. Figure 9 shows these sampled patients' exactly known parameter values as they appear when smoothed and graphed in the same manner as shown in Figure 8. Figure 9 therefore shows the true population parameter distribution that any population modeling method should now discover.

These twenty hypothetical patients then each were "given" a simulated single dose of a drug having one compartment behavior, and five simulated serum samples were drawn at uniformly spaced times after the very short intravenous infusion. The simulated assay SD was ± 0.4 concentration units. The simulated data were then presented to the NPEM algorithm and computer program [25,30], as well as to a parametric population modeling method such as NONMEM or IT2B.

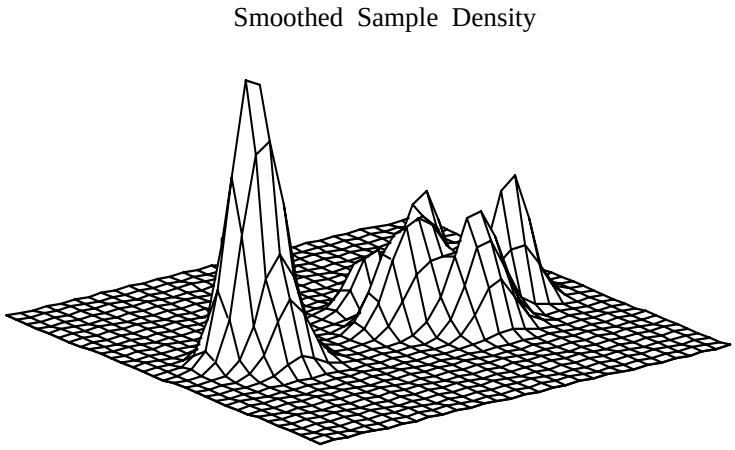

Figure 9. Graph, smoothed as in Figure 8, of the actual parameter values in the twenty sampled patients. The axes are as in Figure 8. This is the true population distribution that NPEM, or any other method, should now discover.

Figure 10 shows the results of the NPEM analysis, again smoothed and graphed as in Figure 8. The NPEM program clearly detected and located the two subpopulations of patients. Figure 10 is similar in shape to the known original population joint distribution shown in Figure 9.

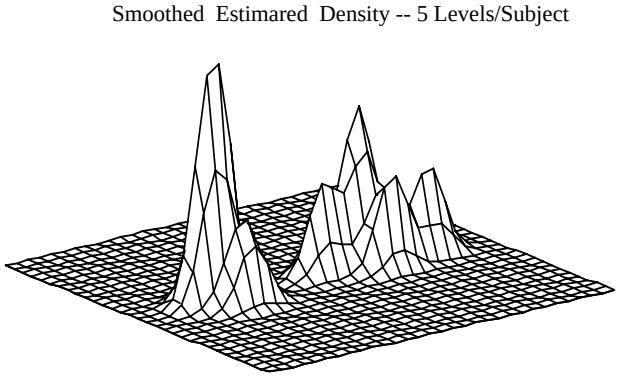

Figure 10. Smoothed estimated population joint density obtained with NPEM, using all five serum concentrations. Axes as in Figure 8. Compare this figure with Figure 9.

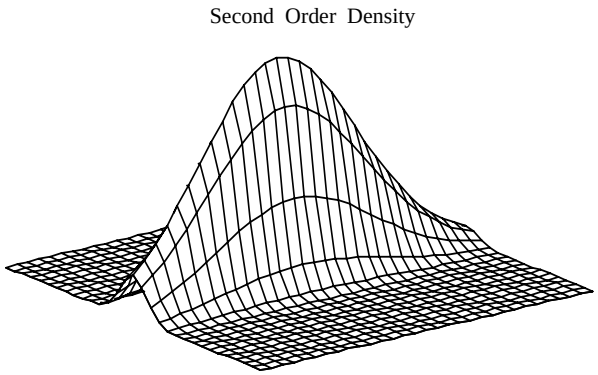

Figure 11. Plot of the population density (the second order density) as perceived by a theoretically optimal parametric method. Axes as in Figure 8. Compare this figure with Figure 9. The two subpopulations are not discovered at all. The true parameter distributions are perceived with great error.

In contrast, Figure 11 shows how a parametric method such as IT2B, the parametric EM, or NONMEM perceives the population shown in Figure 9. The second order density in the figure is the one obtained by a parametric method, which obtains means, variances, and covariances or correlations. Note that the mean is actually where there are no subjects at all. Parametric methods thus cannot discover subpopulations without additional aid. They give an entirely different impression of the population behavior of the drug. One also gets an impression in Figure 11 of much greater

variability between patients than actually exists among the two fairly tightly grouped subpopulations, which is shown in Figures 9 and 10.

9. MAKING A NONPARAMETRIC POPULATION MODEL

The same data of the 17 patients receiving intramuscular Amikacin, described earlier in "Making a Parametric Population Model" above, was also analyzed using both the NPEM and the nonparametric adaptive grid (NPAG, a later version of NPEM with improved speed, precision, and convergence rate) software [27]. As before, the parameters were Ka, the absorptive rate constant from the intramuscular injection site, Vs, the volume of distribution in L/kg, and Ks, the increment of elimination rate constant per unit of creatinine clearance. Initial ranges for these parameters were set at 0 to 6 hr-1 for Ka, 0 to 0.6 l/kg for Vs, and 0 to 0.008 for Ks. Gamma was set at the value previously obtained with the IT2B analysis.

The results are summarized in Table 1, below, where they are also compared with the previous results from the IT2B program.

		IT2B	NPEM	NPAG
Mean	Ka	1.349	1.408	1.380
	VS1	0.258	0.259	0.258
	KS1	0.003258	0.003271	0.003275
Median/CV%	Ka	1.352/4.55	1.363/20.42	1.333/21.24
	VS1	0.2591/13.86	0.2488/17.44	0.2537/17.38
	KS1	0.003273/14.83	0.003371/15.53	0.003183/15.76
Log – Likelihood		-389.548	-374.790	-374.326

Table 1. Parameter values (mean, median, percent coefficient of variation -CV%, and log likelihood obtained with the IT2B, the NPEM, and the NPAG programs. Ka = absorptive rate constant (hr^{-1}), VS1, apparent central volume of distribution (L/kg), KS1, increment of elimination rate constant (hr^{-1} per unit of creatinine clearance). CV% is less with IT2B, but so it the Log likelihood, which is better with NPEM and NPAG.

When comparing the results of these IT2B, NPEM, and NPAG analyses, at first glance there seems to be little to choose between them. The parameter values are all quite similar. In fact, the population percent coefficients of variation (CV%) are clearly least with the IT2B program. This might suggest that the population parameters are estimated with greater precision with that program. However, the likelihood of the results is clearly least with the IT2B program, as shown in Table 1. It is greater with NPEM, and just a bit better still with NPAG. In addition, as shown in Figures 12 through 14, very high correlations were found between all parameters with IT2B. This was not the case with the two nonparametric methods. Because of this, the very high correlations between all pairs of parameter seen with IT2B is probably spurious. The general finding is that the likelihood is significantly greater with the nonparametric methods, and the smaller population CV% found with IT2B is therefore probably due to

its constraining assumption that the parameters must have Gaussian distributions. The price paid for this Gaussian assumption is seen in the lower likelihood of the results obtained. In contrast, NPEM and NPAG, because they are not restricted to the assumption of Gaussian parameter distributions, were better able to detect the full diversity in the population parameter distributions, and were able to obtain more likely results. Notice also that the parameter distributions in Figures 12-14 are not Gaussian, but skewed. The specific marginal distributions of Ka, Ks1 and Vs1 obtained with NPEM and NPAG are not shown because of space considerations.

Many papers using parametric population modeling have not reported the likelihood of their results, but have restricted themselves to reporting only the common indices of "goodness of fit". The reason for this is that IT2B and other methods such as NONMEM, which use the FOCE (first order, conditional expectation) approximation, compute only an approximate value of the likelihood. This is usually not reported. In contrast, the nonparametric methods compute the likelihood value exactly.

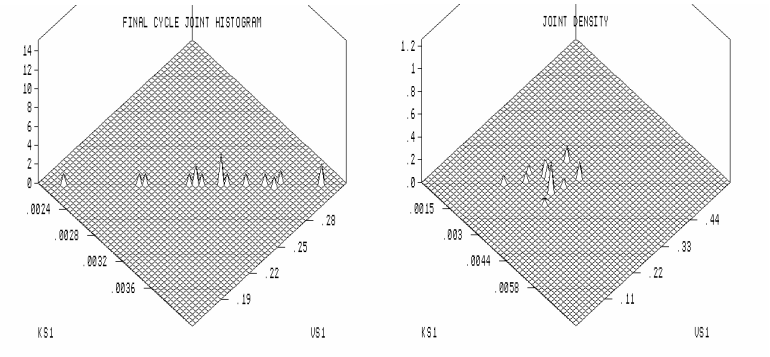

Figure 12. Joint marginal density obtained for KS1 and VS1 with IT2B, left, and NPAG, right. Note the very high, and probably incorrect, correlation between the parameters seen with the IT2B program.

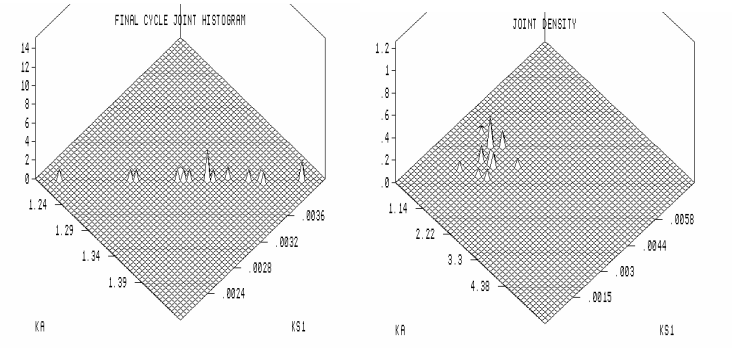

Figure 13. Joint marginal density obtained for KA and KS1 with IT2B, left, and NPAG, right. Note the very high, and probably incorrect, correlation between the parameters seen with the IT2B program.

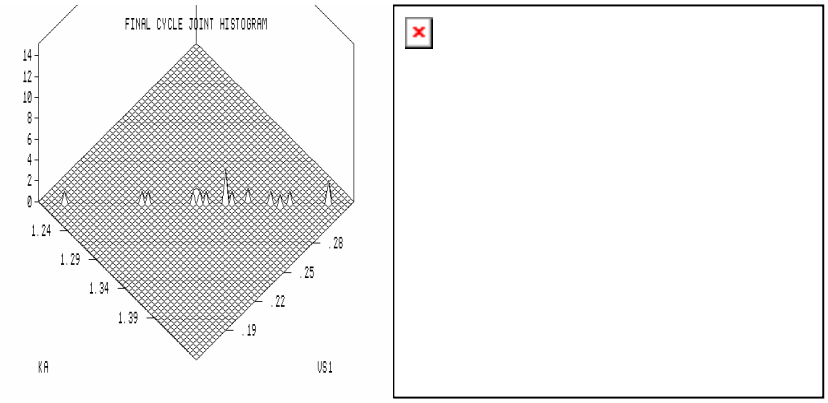

Figure 14. Joint marginal density obtained for KA and VS1 with IT2B, left, and NPAG, right. Note the very high, and probably incorrect, correlation between the parameters seen with the IT2B program.

Figures 15 through 18 show the scatterplots of the above analysis of such fits using the IT2B and the NPAG programs.

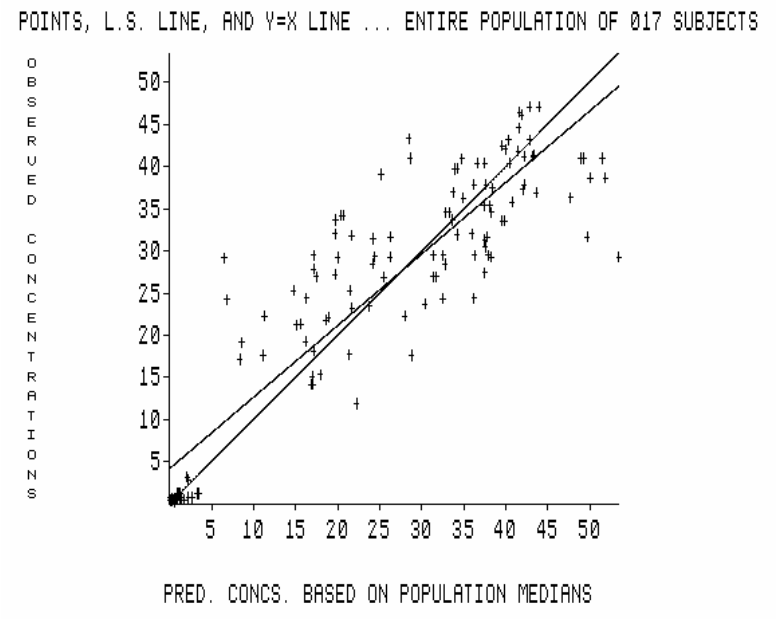

Figure 15. Comparison of predicted and measured serum concentrations based on the population median parameter values, using the IT2B program.

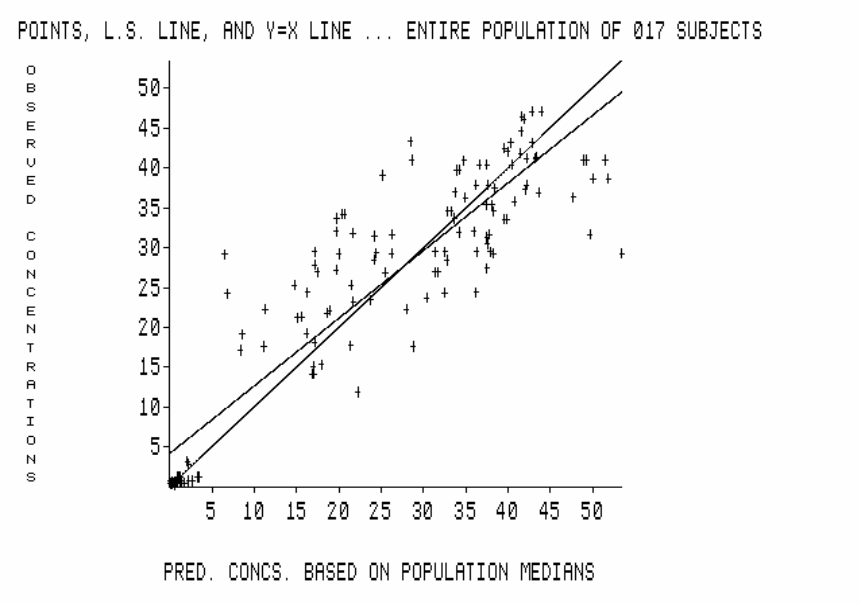

Figure 16. Comparison of predicted and measured serum concentrations based on the population median parameter values, using the NPAG program.

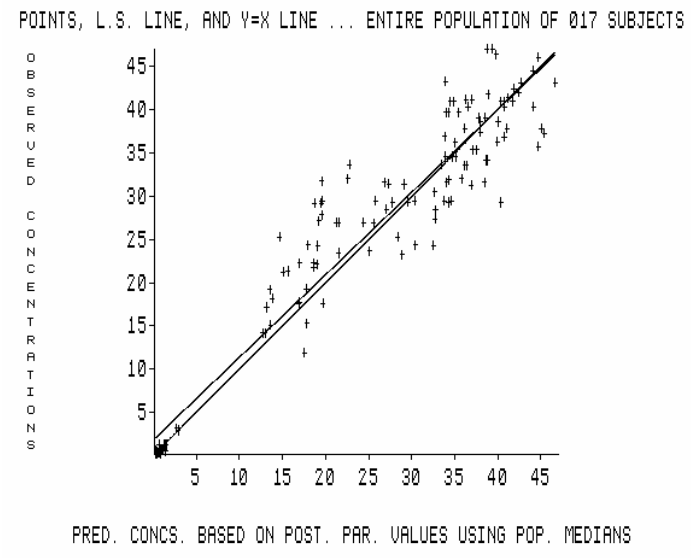

Figure 17. Comparison of predicted and measured serum concentrations based on each subject's individual maximum aposteriori probability (MAP) Bayesian posterior parameter values, predicting only his/her own serum data, using the IT2B program.

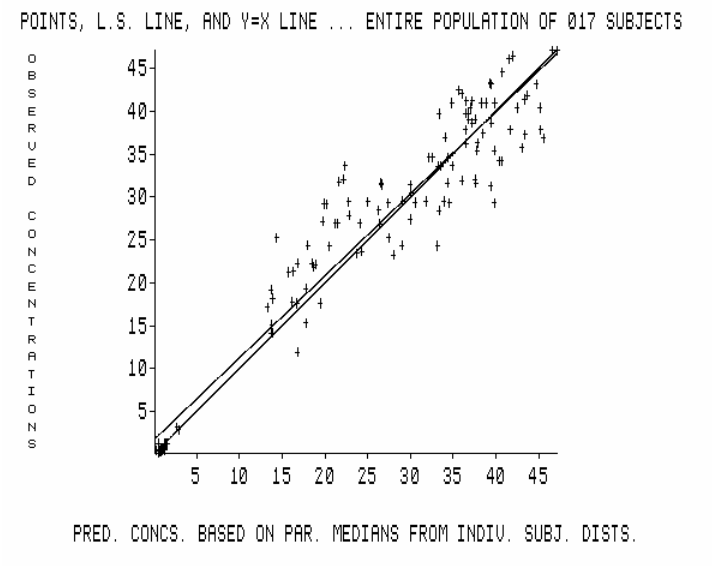

Figure 18. Comparison of predicted and measured serum concentrations based on the median value of each subject's individual nonparametric Bayesian posterior parameter distributions, predicting only his/her own serum data, using the NPAG program.

An analysis of the scatterplots of estimated versus measured serum concentrations, and also those of the NPEM program (not shown), reveals that R^2, the coefficient of the determination (the fraction of the variance explained by the regression relationship between the estimated and the measured data), was greatest with the nonparametric programs. The mean error was the least, and the mean squared error was also the least, showing that the scatterplots actually were more correlated, less biased, and more precise with the two nonparametric modeling methods.

	IT2B	NPEM	NPAG
R^2 =	0.814	0.879	0.880
ME =	-0.575	-0.751	0.169
MSE =	48.69	29.01	29.70

Table 2. Analysis of estimated versus measured serum concentrations based on population median parameter values, using the IT2B, NPEM, and NPAG programs. R^2: square of the correlation coefficient, ME: mean error, MSE: mean squared error.

9.1 Large and Nonlinear Nonparametric PK/PD Population Models

Large and nonlinear nonparametric population modeling software for combined pharmacokinetic and pharmacodynamic modeling has now been implemented on the large IBM "Blue Horizon" parallel machine at the San Diego Supercomputer Center (SDSC), as a research resource for such work. The user describes the specific structural model and pathways using PC software now in the USC*PACK collection [42]. The model file, the patient data files, and the instructions

are sent to the SDSC machine, to a 3-cpu Linux cluster of Dell PC's in our laboratory, or to a larger cluster at USC, over the Web. The analysis is done, and the results are sent back by email to the user's PC, where they are examined just as with the NPEM program described above [39]. More recently, many of these analyses can be done on a single cpu machine such as any PC, though this may require more computer time.

10. NEW DEVELOPMENTS IN NONPARAMETRIC POPULATION MODELING

A significant improvement in nonparametric modeling has recently been made by Leary [27]. He pointed out that the NPEM strategy, based on a large grid covering the parameter space, is computationally quite intensive. He showed that for an analysis of a 5 parameter model, for example, the number of grid points required for a given percent resolution grows greatly with the desired degree of precision, as shown in the table below.

RESOLUTION	GRID POINTS
10%	100,000
5%	3.2 million
2%	310 million
1%	10 billion

Table 3. Relationship between percent resolution for each parameter of a 5 parameter model, and the number of fixed grid points needed to achieve that resolution.

Leary also showed that for a 5 parameter model with only 8 subjects, the likelihood of the results correlated strongly with the number of grid points used in the computations. Using the IBM "Blue Horizon" parallel computer with 1152 processors at the San Diego Supercomputer Center the following relations in Table 4 were obtained between the grid size, computer time, memory required, and the log-likelihood of the results [27].

Grid Size	CPU hours	Memory (Mbytes)	Log-likelihood
10,000 points	0.1	0.7	-7221.2
40,000	0.5	2.5	-572.8
160,000	2.0	10	-543.7
640,000	7.9	40	-506.2
2,560,000	30.8	160	-462.1
10,280,000	121.8	640	-454.5
40.960,000	501.2	2500	-437.3
164,000,000	2037.4	10000	-433.1

Table 4. Relationship between Grid Size, Computer time (CPU hours), Memory required, and Log likelihood of the results obtained with the NPEM program.

The quality of the result (the log-likelihood) thus depends in large part on the number of grid points used in the analysis. A powerful machine and much computer time are needed to obtain good results with NPEM.

In contrast, Leary has now developed a new nonparametric "adaptive grid" (NPAG) procedure, which combined with an interior point algorithm, has made significant advances in the quality, speed and memory requirements for the analysis [27]. The method begins with a much smaller and coarser grid, often as low as 5000 grid points for a problem similar to that above. After this is solved, the grid is refined by adding perturbations (extra grid points, about 10 for each previous solution support point), near them. Then the problem is solved again. Once again, new grid points are placed near the previous solution points. This process then continues iteratively, using decreasing perturbations, adaptively obtaining finer and finer resolution of the grid, until a convergence criterion is met.

The outcome has been much improvement in quality of the results, with far less overall computational time and effort. For example, the above problem using NPEM, using 256 processors, took 2037 processor-hours on the IBM Blue Horizon (at the time the fastest non-classified computer in the world), used 10000 Mbytes of memory, and achieved a likelihood of -433.1. In contrast, NPAG, running on a single processor 833 MHz Dell PC, used only 6 Mbytes of memory, took only 1.7 processor hours, and obtained a result with a likelihood of -433.0. NPAG thus greatly reduces the computational time and memory requirements compared to NPEM, and now permits many tasks to be done on a notebook PC that used to require the large parallel machine using NPEM. It also permits still more complex analyses to be done on the larger parallel machine, tasks that previously would have been impossible using NPEM, such as a large 15 parameter model of an antibiotic dosage regimen and its ability to prevent the emergence of resistant organisms. Furthermore, gamma, the measure of intra-individual variability other than the assay error polynomial, can also now be computed directly in the NPAG software. A new version of the NPAG program with this feature is currently being implemented.

11. STRENGTHS AND WEAKNESSES OF PARAMETRIC AND NONPARAMETRIC POPULATION MODELING METHODS

The main strength of the parametric approaches has been that their ability currently to separate inter-individual from intra-individual variability in the population. In the IT2B program, this has been extended to the determination on the environmental noise by first using an explicit polynomial that reflects the relationship between the assay SD and the measured concentration. Then the remaining intraindividual variability can be ascribed to the other sources of environmental noise as discussed earlier.

There has been a general consensus that nonparametric methods are better when it is known that the model parameter distributions are non-Gaussian. However, there has also been a general impression that the nonparametric methods may be less efficient, especially when the parameter distributions are truly Gaussian, and that

parameter means and variances may be more reliable for such Gaussian distributions, when parametric modeling methods are used.

This question has recently been examined by Leary [27]. He did a careful study on a simulated population in which the parameter distributions were carefully determined to be Gaussian. A one-compartment, two-parameter model was used, with parameters apparent volume of distribution V and elimination rate constant K. The mean V was 1.1, with standard deviation 0.25. Mean K was 1.0, also with SD = 0.25. The correlation coefficient was set at three different values in three different simulated scenarios: -0.6, 0.0, and +0.6. A single unit intravenous bolus dose was given at time zero, and two simulated serum concentrations were used, each with a 10 % observational error. Populations ranging in size from 25 to 800 simulated subjects were studied.

The study objectives were to evaluate the consistency, efficiency, and asymptotic convergence rate of the NPAG program, and to compare it with an approximate parametric method using the FOCE approximation (the IT2B program) and with another parametric EM (PEM) method developed by Schumitzky [4], but using the more recent Faure low discrepancy sequence integration method [27]. Over 1000 replications were done to evaluate bias and efficiency.

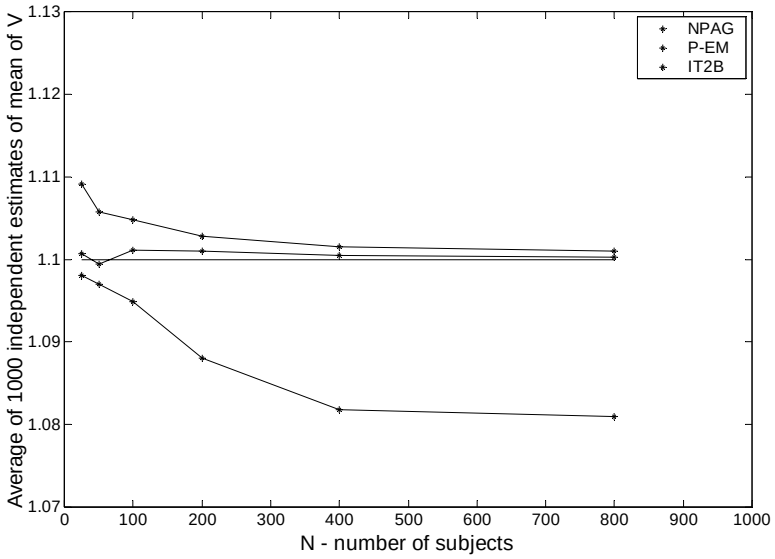

Figure 19. Consistency of estimators of mean of V. NPAG and PEM are consistent. The estimates approach the true value as the number of subjects increases. FOCE IT2B is not consistent.

Figure 19 shows that the NPAG and the PEM programs have consistent behavior. As the number of subjects in the simulated population increased from 25 to 800, the estimated value of the mean of V became closer and closer to the true value, while the IT2B program with the FOCE approximation did not

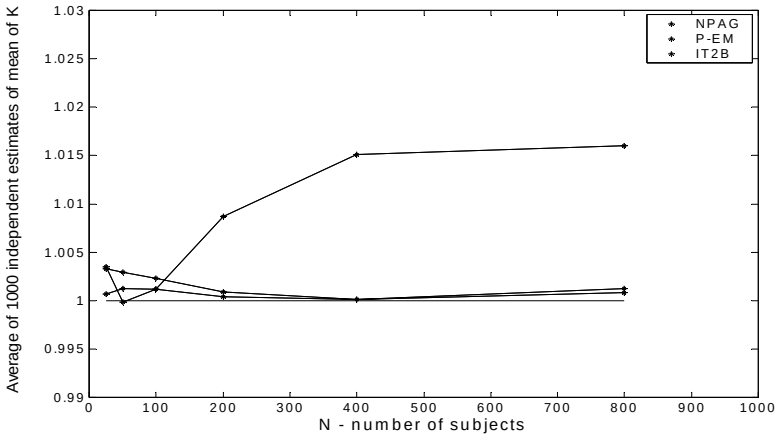

Figure 20. Consistency of estimators of mean of K. NPAG and PEM are consistent. True value of mean K = 1.0. Again, estimates with NPAG and PEM approach the true value as the number of subjects increases. The FOCE IT2B is not consistent.

Figure 20 shows that the same was true for the estimation of K. The estimates of K with NPAG and PEM approached the true value as the number of subjects in the population increased, while the IT2B FOCE estimates again drifted away from the true value.

Figure 21 shows the same behavior for the estimates of the SD of K. NPAG and PEM had consistent behavior whlie the FOCE IT2B again drifted away. Similar behavior was present in the estimation of the SD of V (not shown).

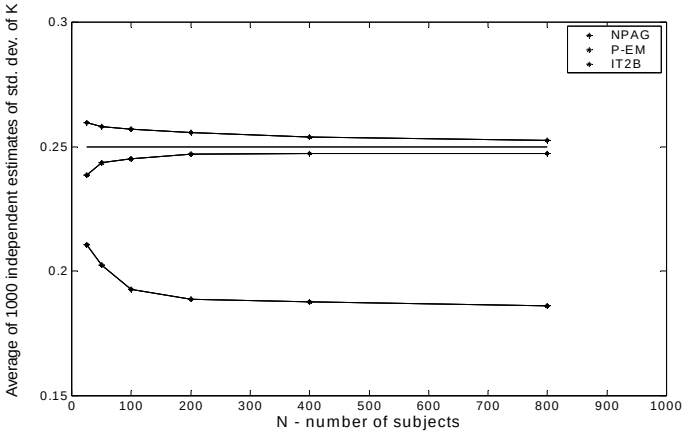

Figure 21. Consistency of estimators of SD of K. NPAG and PEM are consistent. Results approach the true value of SD of K = 0.25 as the number of subjects increases. FOCE IT2B is not consistent. Results actually drift away from the true value with increasing subjects.

Figure 22 shows that the estimation of the correlation coefficient between V and K with NPEM and NPAG was consistent, approaching the true value more and more closely as the number of subjects increased, while the FOCE IT2B started at about zero instead of the true –0.6, and then increased up to +0.2. Similar behavior was seen when the correlation coefficient was 0.0, and also when it was +0.6.

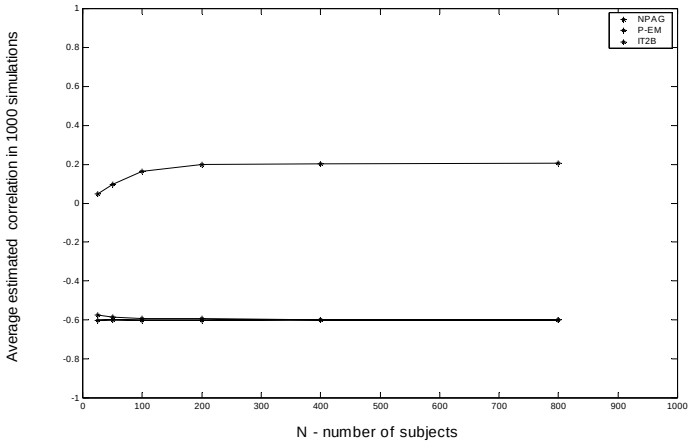

Figure 22. Consistency of estimators of correlation coefficient between V and K. NPAG and PEM are consistent. FOCE IT2B is not, and is severely biased.

In the further examination of statistical efficiency, Figure 23 shows that the efficiency with NPAG and PEM both were about 0.8 throughout, while that of the FOCE IT2B was much less, starting at about 0.4 with 25 subjects, decreasing to less than 0.1 for 800 subjects.

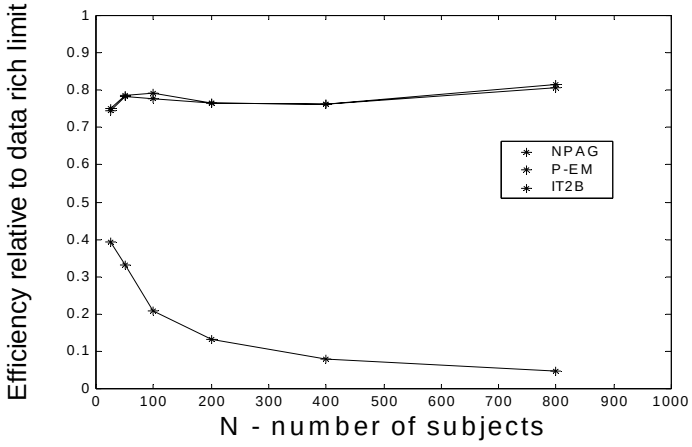

Figure 23. Statistical efficiencies of NPAG and PEM are nearly identical, and are much greater than that of the FOCE IT2B.

Furthermore, as shown in Figure 24, the asymptotic convergence rate was close to theoretical with NPEM and NPAG, while that of the FOCE IT2B was very much less. In order to decrease the SD of a parameter estimate by half, 4 times as many subjects were required by NPAG and PEM, as is consistent with asymptotic theory, while 16 times as many were required by the FOCE IT2B program.

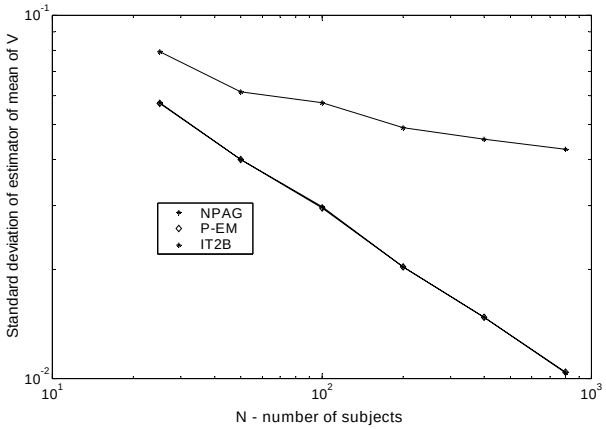

Figure 24. Asymptotic convergence rate of IT2B is much less than that of NPAG and PEM. For NPAG and PEM, the SD decreases by half with 4 times the number of subjects. For the FOCE IT2B, fully 16 times the number of subjects are required for the same SD decrease by half.

In summary, NPAG and PEM, which have exact or very low discrepancy computations of the likelihood, had quite consistent behavior, good efficiency, and good asymptotic convergence. In contrast, the IT2B program, which used the FOCE approximation, suffered a loss of consistency, with a small (1 – 2%) bias for the mean parameter values, moderate (20 – 30%) bias for the SD's, and severe bias for the correlation coefficients, as described above. Further, the NPAG and PEM programs were quite efficient, much more so than the FOCE IT2B, and had much better statistical convergence, close to theoretical.

One weakness of the nonparametric population modeling approach is that until recently there has been no feature to separate the various sources of variability into their respective components – the inter-individual variability due to the diversity among the subjects in the ways they handle the drug, and the intra-individual variability due to the assay and the environmental errors. Current nonparametric methods have not resolved these things. That, however, is what the parametric methods do very well, as described in other papers in this collection. Nevertheless, this deficiency has recently been overcome by Leary [27]. An improved version incorporating this feature is now being implemented.

Another weakness of current nonparametric approaches is that there is no method to obtain confidence limits for these nonparametric distributions. There is nothing analogous to the "standard error of the mean" in parametric distributions. However, since the NP parameter means and variances are at least as good as those

seen with P methods [27], confidence intervals based on these means and variances can be used with at least the same justification as those based on parametric modeling methods. However, to get confidence intervals for the entire nonparametric parameter distributions, one must use the much more computationally intensive bootstrap methods to obtain them.

The strengths of nonparametric approaches are many. First, they have the desirable properties of mathematical consistency, good statistical efficiency, and good asymptotic convergence [8,10,27,32]. Second, no assumptions about the shape of the parameter distributions need to be made. Because of this, nonparametric methods can detect, without additional aid from covariates or descriptors, previously unsuspected subpopulations of patients, as shown in Figures 8-11. Third, instead of obtaining only single-point parameter estimates, one gets multiple estimates, up to one for each subject studied. This is why the nonparametric approach is more informative. It comes the closest to the ideal of the best that could ever be done, namely to obtain the collection of each subject's exactly known parameter values. Fourth, the multiple sets of parameter values provide a tool to circumvent the separation principle [40] and to calculate and optimize the predicted precision with which any candidate dosage regimen is predicted to hit a desired target goal at a desired time. The nonparametric population models thus permit "multiple model" design of dosage regimens to optimize a specific performance criterion [37,40, 41]. The development of such optimally precise dosage regimens, using nonparametric population models, will be discussed in a separate chapter. Table 5 below summarizes the various current strengths and weaknesses of the parametric and nonparametric modeling approaches.

Method	
Parametric	**Nonparametric**
Strengths	
• Get intraindividual variability.	• Get results with max likelihood.
• Get parameter confidence intervals.	• Don't need Gaussian assumptions.
	• Consistent, efficient, convergent.
	• Best quality means, SD's, etc.
	• Best suited for maximally precise dosage regimens.
	• Can use conf intervals from Gaussian theory just as well as parametric methods can.
Weaknesses	
• Not consistent, efficient, or convergent.	• No confidence intervals for the full nonparametric distributions.
• Constrained by Gaussian assumptions.	
• Not suited for optimally precise dosage regimens.	

Table 5. Strengths and Weaknesses of Parametric and Nonparametric Poulation Modeling Methods.

11.1 Current Optimal Strategies in Population Modeling

The optimal strategy for making clinically useful population PK/PD models, until the estimation of Gamma is implemented in NPAG, currently appears to be the following sequence of steps:

1. determine the assay error pattern explicitly and obtain the assay error polynomial.
2. use a parametric population modeling program such as IT2B, to obtain gamma.
3. having both of the above, use a nonparametric population modeling program such as NPAG to obtain the most likely entire discrete joint parameter distribution.

This sequence of steps currently appears to make optimal use of the information about the assay SD, often 1/3 to 1/2 of the overall environmental SD, and the raw data present in the population studied, to obtain the most probable parameter distributions. The approach also appears to provide optimal tools to develop dosage regimens to achieve desired target goals with maximum precision. This is useful in drug development for designing maximally precise dosage regimens for specific populations of patients, and clinically, for developing the most precise initial regimen for individual patients where the margin of safety for the drug in question is small and the dosage must be carefully individualized. The nonparametric population models, with their multiple support points, are extremely well suited for the new “multiple model” method of dosage design [35,41], which can specifically evaluate the weighted squared error of the failure of a dosage regimen to achieve a desired target goal, and optimize the regimen to minimize this error. This will be described in more detail in another chapter.

Acknowledgements

Supported by US Government grants LM 05401, RR 01629, RR11526, and GM65619.

References

1. Variability in Drug Therapy: Description, Estimation, and Control. Ed by Rowland M, Sheiner L, and Steimer JL. Raven Press, New York, 1985.

2. Sheiner L, Beal S, Rosenberg B, and Marathe V: Forecasting Individual Pharmacokinetics. Clin. Pharmacol. Therap. 26: 294-305, 1979.

3. Aarons L: The Estimation of Population Pharmacokinetic Parameters using an EM Algorithm. Comput. Methods and Programs in Biomed. 41: 9-16, 1993.

4. Schumitzky A: EM Algorithms and Two Stage Methods in Pharmacokinetic Population Analysis. In: Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis II. D. Z.D'Argenio, ed., Plenum Press, New York, 1995, pp. 145-160.

5. Beal S, and Sheiner L: NONMEM User's Guide I. Users Basic Guide. Division of Clinical Pharmacology, University of California, San Francisco, 1979.

6. Sheiner L: The population Approach to Pharmacokinetic Data Analysis: Rationale and Standard Data Analysis Methods. Drug Metab. Rev. 15: 153-171, 1984.

7. Beal S: Population Pharmacokinetic Data and Parameter Estimation Based on their First Two Statistical Moments. Drug Metab. Rev. 15: 173-193, 1984.

8. De Groot M: Probability and Statistics, 2nd edition, 1986, reprinted 1989, Addison-Wesley, Reading MA, pp. 334-336.

9. Spieler G and Schumitzky A: Asymptotic Properties of Extended Least Squares Estimators with Approximate Models. Technical Report 92-4, Laboratory of Applied Pharmacokinetics, University of Southern California School of Medicine, 1992.

10. Spieler G and Schumitzky A: Asymptotic Properties of Extended Least Squares Estimates with Application to Population Pharmacokinetics. Proceedings of the American Statistical Society, Biopharmaceutical Section, 1993, pp. 177-182.

11. Rodman J and Silverstein K: Comparison of Two Stage (TS) and First Order (FO) Methods for Estimation of Population Parameters in an Intensive Pharmacokinetic (PK) Study. Clin. Pharmacol. Therap. 47: 151, 1990.

12. Maire P, Barbaut X, Girard P, Mallet A, Jelliffe R, and Berod T: Preliminary results of three methods for population pharmacokinetic analysis (NONMEM, NPML, NPEM) of amikacin in geriatric and general medicine patients. Int. J. Biomed. Comput., 36: 139-141, 1994.

13. Lindstrom M and Bates D: Nonlinear Mixed-Effects Models for Repeated Measures Data. Biometrics, 46: 673-687, 1990.

14. Vonesh E and Carter R: Mixed Effects Nonlinear Regressions for Unbalanced Repeated Measures. Biometrics, 48: 1-17, 1992.

15. Jelliffe R: Explicit Determination of laboratory assay error patterns: a useful aid in therapeutic drug monitoring. No. DM 89-4 (DM56). Drug. Monit. Toxicol. 10: (4) 1-6, 1989.

16. Jelliffe R, Schumitzky A, Van Guilder M, Liu M, Hu L, Maire P, Gomis P, Barbaut X, and Tahani B: Individualizing Drug Dosage Regimens: Roles of Population Pharmacokinetic and Dynamic Models, Bayesian Fitting, and Adaptive Control. Therap. Drug Monit. 15: 380-393, 1993.

17. Bustad A, Jelliffe R, and Terziivanov D: A comparison of Parametric and Nonparametric Methods of Population Pharmacokinetic Modeling. A poster presentation at the Annual Meetings of the American Society for Clinical Pharmacology and Therapeutics, Atlanta, GA, March 26, 2002.

18. Jelliffe R: Estimation of Creatinine Clearance in Patients with Unstable Renal Function, without a Urine Specimen. Am. J. Nephrol. 22: 320-324, 2002.

19. Jazwinski A: Stochastic Processes and Filtering Theory. Academic Press, New York, 1970.

20. Van Guilder M, Leary R, Schumitzky A, Wang X, Vinks S, and Jelliffe R: Nonlinear Nonparametric Population Modeling on a Supercomputer. Presented at the 1997 ACM/IEEE SC97 Conference, San Jose CA, November 15-21, 1997.

21. De Groot M: Probability and Statistics, 2nd edition, 1986, reprinted 1989, Addison-Wesley, Reading MA, pp. 334-336.

22. Spieler G and Schumitzky A: Asymptotic Properties of Extended Least Squares Estimators with Approximate Models. Technical Report 92-4, Laboratory of Applied Pharmacokinetics, University of Southern California School of Medicine, 1992.

23. Spieler G and Schumitzky A: Asymptotic Properties of Extended Least Squares Estimates with Application to Population Pharmacokinetics. Proceedings of the American Statistical Society, Biopharmaceutical Section, 1993, pp. 177-182.

24. Mallet A: A Maximum Likelihood Estimation Method for Random Coefficient Regression Models. Biometrika. 73: 645-656, 1986.

25. Schumitzky A: The Nonparametric Maximum Likelihood Approach to Pharmacokinetic Population Analysis. Proceedings of the 1993 Western Simulation Multiconference - Simulation for Health Care. Society for Computer Simulation, 1993, pp 95-100.

26. Bertsekas D: Dynamic Programming: deterministic and stochastic models. Englewood Cliffs (NJ): Prentice-Hall, pp.144-146, 1987.

27. Leary R, Jelliffe R, Schumitzky A, and Van Guilder M: A Unified Parametric/Nonparametric Approach to Population PK/PD Modeling. Presented at the Annual Meeting of the Population Approach Group in Europe, Paris, France, June 6-7, 2002.

28. Bertilsson L: Geographic/Interracial Differences in Polymorphic Drug Oxidation. Clin. Pharmacokinet. 29: 192-209, 1995.

29. Lindsay B: The Geometry of Mixture Likelihoods: A General Theory. Ann. Statist. 11: 86-94, 1983.

30. Schumitzky A: Nonparametric EM Algorithms for Estimating Prior Distributions. App. Math. and Computation. 45: 143-157, 1991.

31. Maire P, Barbaut X, Girard P, Mallet A, Jelliffe R, and Berod T: Preliminary results of three methods for population pharmacokinetic analysis (NONMEM, NPML, NPEM) of amikacin in geriatric and general medicine patients. Int. J. Biomed. Comput., 36: 139-141, 1994.

32. Spieler G and Schumitzky A: Asymptotic Properties of Extended Least Squares Estimates with Application to Population Pharmacokinetics. Proceedings of the American Statistical Society, Biopharmaceutical Section, 1993, pp. 177-182.

33. Hurst A, Yoshinaga M, Mitani G, Foo K, Jelliffe R, and Harrison E: Application of a Bayesian Method to Monitor and Adjust Vancomycin Dosage Regimens. Antimicrob. Agents and Chemotherap. 34: 1165-1171, 1990.

34. Bayard D, Milman M, and Schumitzky A: Design of Dosage Regimens: A Multiple Model Stochastic Approach. Int. J. Biomed. Comput. 36: 103-115, 1994.

35. Bayard D, Jelliffe R, Schumitzky A, Milman M, and Van Guilder M: Precision Drug Dosage Regimens using Multiple Model Adaptive Control: Theory and Application to Simulated Vancomycin Therapy. in Selected Topics in Mathematical Physics, Professor R. Vasudevan Memorial Volume, ed. by Sridhar R, Srinavasa Rao K, and Vasudevan Lakshminarayanan, Allied Publishers Inc., Madras, 1995, pp. 407-426.

36. Mallet A, Mentre F, Giles J, Kelman A, Thompson A, Bryson S, and Whiting B: Handling Covariates in Population Pharmacokinetics with an Application to Gentamicin. Biomed. Meas. Infor. Contr. 2: 138-146, 1988.

37 Taright N, Mentre F, Mallet A, and Jouvent R: Nonparametric Estimation of Population Characteristics of the Kinetics of Lithium from Observational and Experimental Data: Individualization of Chronic Dosing Regimen Using a New Bayesian Approach. Therap. Drug Monit. 16: 258-269, 1994.

38. Jerling M: Population Kinetics of Antidepressant and Neuroleptic Drugs. Studies of Therapeutic Drug Monitoring data to Evaluate Kinetic Variability, Drug Interactions, Nonlinear Kinetics, and the Influence of Genetic Factors. Ph. D. Thesis, Division of Clinical Pharmacology, Department of

Medical Laboratory Sciences and Technology, Karolinska Institute at Huddinge University Hospital, Stockholm, Sweden, 1995, pp 28-29.

39. Van Guilder M, Leary R, Schumitzky A, Wang X, Vinks S, and Jelliffe R: Nonlinear Nonparametric Population Modeling on a Supercomputer. Presented at the 1997 ACM/IEEE SC97 Conference, San Jose CA, November 15-21, 1997.

40. Bertsekas D: Dynamic Programming: deterministic and stochastic models. Englewood Cliffs (NJ): Prentice-Hall, pp.144-146, 1987.

41. Jelliffe R, Schumitzky A, Bayard D, Milman M, Van Guilder M, Wang X, Jiang F, Barbaut X, and Maire P: Model-Based, Goal-Oriented, Individualised Drug Therapy: Linkage of Population Modelling, New "Multiple Model" Dosage Design, Bayesian Feedback and Individualised Target Goals. Clin. Pharmacokinet. 34: 57-77, 1998.

42. Jelliffe R, Schumitzky A, Van Guilder M, and Jiang F: User Manual for Version 10.7 of the USC*PACK Collection of PC Programs, USC Laboratory of Applied Pharmacokinetics, USC School of Medicine, Los Angeles, CA, December 1, 1995.