



# The Bayesian approach to population pharmacokinetic/pharmacodynamic modeling

**Jon Wakefield**  
**Leon Aarons**  
**Amy Racine-Poon**

**ABSTRACT** It is one of the principal aims of drug development to discover, for a particular agent, the relationship between dose administered, drug concentrations in the body and efficacy/toxicity. Understanding this relationship leads to the determination of doses which are both effective and safe. Population pharmacokinetic/pharmacodynamic models provide an important aid to this understanding.

Pharmacokinetics considers the absorption, distribution and elimination over time of a drug and its metabolites. Pharmacokinetic data consist of drug concentrations along with (typically) known sampling times and known dosage regimens. A dosage regimen is defined by a route of administration and the sizes and timings of the doses. Pharmacodynamics considers the action of a drug on the body. Pharmacodynamic data consist of a response measure, for example blood pressure, a pain score or a clotting time, and a known dosage regimen. Population data arise when these quantities are measured on a group of individuals, along with subject-specific characteristics (covariates) such as age, sex or the level of a biological marker. When identical doses are administered to a group of individuals large between-individual variability in responses is frequently observed. The mechanisms which cause this variability are complex and include between-individual differences in both pharmacokinetic and pharmacodynamic parameters. The general aim of population studies then is to isolate and quantify the within- and between-individual sources of variability. The explanation of between-individual sources of variability in terms of known covariates is important as it has implications for the determination of dosage regimens for particular covariate-defined subpopulations.

In this chapter we describe the drug development process from a population pharmacokinetic/pharmacodynamic perspective. In particular we describe how the nature of the statistical analysis and the models that are used are modified as the type of data and the aims of the study change through the various phases of development. The Bayesian approach to population modeling is particularly appealing from a biological perspective as it allows informative prior distributions to be incorporated. These priors may arise from previous studies and/or from medical/biological considera-

tions. From an estimation standpoint a Bayesian approach is preferable because of the difficulties which a classical approach encounters due to the large numbers of parameters, the nonlinearity of the subject-specific models which are typically used and the large numbers of variance parameters.

We illustrate the population approach to drug development by describing a number of studies which were carried out by Ciba for a particular anti-clotting agent. We also present a detailed analysis for one of the studies.

## 1 Introduction

It is the aim of drug development to discover safe and efficacious doses for clinical use. Berry (1990) provides a description of the drug development process, and in particular the types of trial that are undertaken. Population pharmacokinetic/pharmacodynamic (PK/PD) models provide a valuable aid to the drug development process by identifying sources of, and quantifying the remaining, variability in drug concentrations and response measures. 'Population' here makes explicit the fact that we wish to gain an understanding of the dose/concentration/response relationship across different groups as defined by covariates such as sex, age and weight. Statistically, population PK/PD models fit within the framework of nonlinear hierarchical models. These models contain many layers of assumptions and it is difficult to check the appropriateness of these assumptions, particularly those which concern unobservable quantities. Consequently it is essential that the modeling be informed by medical/pharmacological information, both in the form of the deterministic/stochastic parts of the model and also via priors for model parameters. In this paper we aim to describe how population PK/PD modeling is carried out in practice. We illustrate the approach using a study of the anti-coagulant drug REVASC<sup>TM</sup>. This study was carried out in Phase II of the drug's development; we also describe five other studies which were carried out in Phase I and informed the analysis of the main study.

The structure of this chapter is as follows. In Section 2 we provide an overview of the population approach and drug development. In Section 3 we outline the role of PK/PD studies in drug development and in Section 4 describe the statistical models which are used for the PK/PD data of a single individual. In Section 5 we explicitly consider population models from a drug development perspective and in Section 6 describe a three stage hierarchical model which may be used for the analysis of population data. In Section 7 we describe a number of population PK/PD studies which were carried out for REVASC<sup>TM</sup> and present a detailed analysis of one of the studies in particular. We finish by describing those areas of population PK/PD for which research is still required.

## 2 Overview of the population approach and drug development

The aim of a clinical development program for a new drug is to provide relevant information on the safety and efficacy of the compound so enabling the prescribing physician to optimally treat individual patients. Frequently, however, the dosing recommendations that emerge from such studies are found inappropriate and when individual dose adjustment is needed, the recommendations provided may be insufficiently informative to allow the adjustment to be undertaken in an optimal manner.

Pharmacokinetics has been defined as what the body does to a drug, and pharmacodynamics as what a drug does to the body. A PK model of a drug attempts to relate drug dosage to drug concentration, usually measured in blood or plasma, or to drug excretion, usually in urine. A pharmacodynamic model attempts to relate drug concentration, ideally at the site of action of the drug but more usually in blood, to some pharmacological effect. The aim of PK/PD modeling therefore is to combine the PK model of a drug with a pharmacodynamic model in order to relate dose to effect in a quantitative manner. To this end PK/PD modeling allows the separation of the factors that influence the inter-individual variability in pharmacokinetics from those that influence pharmacodynamics. Specifically the identification of these influential individual-specific variables (covariates) provides the basis for individual dose selection and hence better therapy. Covariates which are likely to affect PK/PD parameters include demographic variables (such as age and sex), biological information (such as the values of physiological markers), genetic information (phenotype and genotype), comedications, environmental factors and disease states. Figure 1 shows a schematic representation of the relationship between dose, concentration and response.

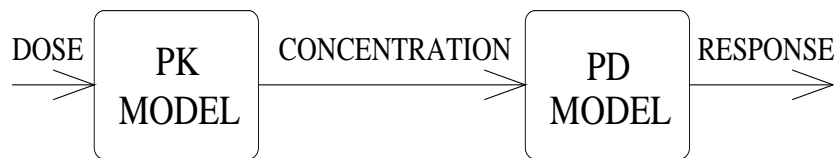


FIGURE 1. Schematic representation of the dose/concentration/response relationship.

An understanding of the dose-effect relationship is basic to the process of identifying those doses to be used in phase III clinical trials and in clinical practice. It is derived in a piece-wise fashion from the various phases of drug development. In establishing a dose-effect relationship, it is essential to also define the concentration-effect relationship. Appropriate mathematical models and the methodology to estimate the model parameters and to identify relevant covariates, that influence inter-individual variability in drug response, are required for both dose-concentration and concentration-effect modeling. In this sense dose-effect and PK/PD modeling are important tools in drug development. The size of

study necessary to detect important covariate effects currently has not been established and will depend on the variability and magnitude of the parameter-covariate relationship, on the sampling design and attained compliance levels.

In early drug development, the use of clinical outcomes, especially in healthy volunteers, is often not feasible and intermediate or surrogate outcomes have to be chosen to replace the real clinical endpoint. A primary motivation for the use of a surrogate endpoint is the possible reduction in sample size or trial duration that can be expected when a rare or distal outcome is replaced by a more frequent or proximate outcome. Such reductions have important cost implications, and in some cases may be pivotal in regard to the trial feasibility. Other motivations include the possibility that true outcome measurement may be unduly invasive, uncomfortable or expensive. Furthermore, outcome events close in time to the treatment or intervention activities under study may be more readily interpreted than are more distal outcomes such as study subject death, which may be also confounded by secondary treatments or competing risks. This motivation, however, seems to relate more to the choice of principal endpoint than to the issue of defining a replacement, or surrogate, for a selected endpoint.

The history of drug development contains many examples of drugs that entered the market with suboptimal dosing regimens, usually erring on the side of initially recommended doses being considerably higher than necessary (Temple, 1989). So far as suboptimal dosing regimens are concerned an important issue is the needless over-exposure of patients to drug, with avoidable dose-dependent side-effects and hazard to patients. The potential exists, when initially-selected doses are too high, that dose-dependent adverse reactions might trigger the conclusion that a potentially valuable drug has too high a risk-benefit ratio to remain in the marketplace. Loss of a potentially valuable therapeutic agent can have adverse public health consequences. Drug development today is an international business and in particular, drug registration is being organized increasingly at a pan-national level. Furthermore, drug development is a multidisciplinary activity and consequently needs input from specialists from widely differing backgrounds: for example, clinical medicine, clinical pharmacology, pharmaceutical science and statistics.

The benefit to drug producers of more efficient clinical research is reduced development costs. In addition, potentially more, or at least higher quality information will be obtained on the drug. This will have direct effect on the consumer and the community as end-costs may well be reduced and the patient is more likely to obtain optimum therapy. Therefore there is a potential knock-on effect, in that quality of life is likely to be improved and time in hospital may be reduced. Consequently the efficient design and performance of clinical trials during the development of a new drug is in the interest of both the developer and consumer. Therefore society as a whole benefits, in that medicines are designed in an optimal manner and patients should achieve the maximum benefit from a drug with the minimum risk.

### 3 Pharmacokinetics/pharmacodynamics in drug development

In Section 2 it was stated that the aim of PK/PD modeling was to relate dose to effect in a quantitative manner. Broadly speaking drug development is divided into a preclinical phase and a clinical phase. The preclinical phase includes drug discovery and early development, mainly involving safety assessment studies in animals. Pharmacokinetic studies are carried out in at least two animal species, typically rat and dog. These studies will involve the development of assay methodology and the assessment of absorption, distribution and elimination of the drug in the animal. Preliminary formulation work is undertaken to define the formulations to be used in man. Extensive toxicology testing is initiated – studies which may be still ongoing during the clinical development phase – and the metabolic fate of the drug is defined from *in vitro* and *in vivo* studies. One of the major aims of the pre-clinical program is to ensure safety before first administration to man of the IND (Investigational New Drug) and indeed to define the doses to be administered in the first clinical studies.

Although the definitions vary from company to company and depend on the therapeutic area, the clinical development program is divided into three broad phases – I, II and III. Phase I studies involve the first administration of the IND to man and can be defined as tolerability studies. In general these studies will be carried out in normal, healthy (usually male) young volunteers. However in fields like oncology, because of the toxicity of the agents, initial studies will be in patients. Initial pharmacokinetic studies will involve single rising dose protocols followed by multiple dose studies. Bioavailability assessment and studies investigating the influence of food, comedication, gender and age may also be performed. These studies are usually performed in small groups (6 to 24) of individuals under careful experimental control. Typically, between 12 to 15 blood samples will be taken per subject per study, depending on the total amount of blood drawn per subject during the whole study.

Phase II studies are usually the first patient studies and one of their primary goals is to define the dose-response relationship. The response may be a surrogate marker, as clinical outcome, for example survival time, may be too difficult to implement in study designs at this stage. As with phase I studies, phase II studies tend to be in small groups of patients and are carried out under carefully controlled experimental conditions. Pharmacokinetic studies will again involve single and multiple dose administration, primarily to assess differences between the patient population and a normal, young volunteer population. In addition the effect of disease states, such as renal and hepatic insufficiency are studied at this stage. Other pharmacokinetic and pharmacodynamic studies performed in phase II include drug interaction studies, which are increasingly being directed from pre-clinical information, and comparative studies with competitive products. Often a Phase II study will collect sparse concentration/response data on all patients. Alternatively the majority of the centres participating in the study will collect sparse

data on patients whilst the remainder will collect full profiles. This is because of logistical constraints on the centres.

Phase III studies are large scale clinical trials in the target population designed to demonstrate efficacy and tolerability of the drug. The dynamic marker is now a clinically relevant measure such as mortality or morbidity. Given that the primary purpose of these studies is the demonstration of efficacy, in the past relative little pharmacokinetic and pharmacodynamic information has been gathered during this phase. However now there is a growing realisation that it is important to collect pharmacokinetic and pharmacodynamic information during phase III. Logistically, however, it is difficult to obtain a large number of blood samples or dynamic measures in each patient. Consequently relatively sparse pharmacokinetic and pharmacodynamic data - sometimes only one sample per subject - is obtained during phase III studies. In addition phase III studies are frequently not carried out under the same degree of experimental control as phase I and II studies, and consequently the quality of data obtained during phase III may not be of the same standard as that obtained earlier in the development program, though the assay may have been improved. At the end of the phase III program, the company will apply to a regulatory authority for a NDA (New Drug Application) which will allow the drug to be marketed. Some data are collected post-marketing in what are called phase IV studies but in general any pharmacokinetic or pharmacodynamic studies are carried out by clinical investigation sites at this stage, rather than the company.

## 4 Mathematical models for individual PK/PD data

### 4.1 *Compartmental models for PK data*

In this section we describe the class of models that are used to model individual drug concentrations and responses. We let  $y(t)$  denote the concentration of an individual at time  $t$ . The drug is introduced into the body via a particular route of administration. Common routes include intravenous bolus, intravenous infusion, oral or subcutaneous. An *intravenous bolus* is an instantaneous introduction of drug directly into the blood stream via an injection, an *intravenous infusion* is a constant introduction of drug directly into the blood stream over some specified period, and a *subcutaneous dose* is an injection beneath the skin and is, consequently, not directly into the bloodstream. After introduction the drug undergoes the processes of absorption, distribution and elimination. These processes are assumed to give rise to concentrations which vary with time in such a way that they may be modeled via a sum of exponentials form. A conceptual way of viewing the way in which these models arise is by considering the body as being modeled as a series of homogeneous compartments or pools. Gibaldi and Perrier (1982) describe the rationale behind this modeling and the types of model which may be appropriate for different drugs/routes of administration. For example a three-compartment model may nominally consist of blood, soft tissue and deep muscle

compartments. Within each of these compartments the drug is assumed to have identical kinetic behaviour. The flow between compartments is typically described by a series of linear differential equations. For a  $p$ -compartment system denote the amount of drug in compartment  $i$  by  $x_i(t)$ ,  $i = 1, \dots, p$ . Then the rate of change of drug in compartment  $i$  may be modeled by

$$\frac{dx_i}{dt} = \sum_{j=0}^p (k_{ji}x_j - k_{ij}x_i) \quad (1)$$

where  $k_{ij}$  denotes the rate constant associated with flow from compartment  $i$  to compartment  $j$ . Compartment 0 denotes the outside environment from where the drug is administered and to where drug is eliminated. Under this model the rate of change of drug is assumed to be proportional to the amount of drug in the donor compartment. In general the rate of change of drug may be modeled via some other form. For example the rate of change may be taken to be independent of the amount of material in the donor compartment to give a zero-order equation. Alternatively the rate of flow from compartment  $i$  to compartment  $j$  may be described by a Michaelis-Menten type relationship:

$$\frac{-k_{ij1}x_i}{k_{ij2} + x_i}.$$

Note that for small  $x_i$  this equation is approximately linear whilst for larger values the rate of flow approaches a constant value. An advantage of a linear set of equations such as (1) is that it is straightforward to obtain expressions for the amounts of drug in the different compartments as a function of time. For other forms for which analytical solutions do not exist numerical integration is required.

For general PK modeling the compartmental system is a convenient visualisation but the compartments have no physiological interpretation. For physiological analyses, which typically use a large number of compartments to model the various organs and tissues of the body and only obtain data on a small number of individuals, greater interpretation is possible. However, parameter identification is limited due to the fact that samples can, in general, only be taken from one or two compartments. Gelman, Bois and Jiang (1996) consider the Bayesian approach to physiological modeling.

We now describe in some detail the simple two-compartment model illustrated in Figure 2. Compartments one and two may nominally be thought to represent blood and tissue, respectively. We suppose that a single intravenous bolus of drug of size  $D$  is given at time zero. We then have

$$\frac{dx_1}{dt} = k_{21}x_2 - k_{12}x_1 - k_{10}x_1 \quad (2)$$

and

$$\frac{dx_2}{dt} = k_{12}x_1 - k_{21}x_2. \quad (3)$$

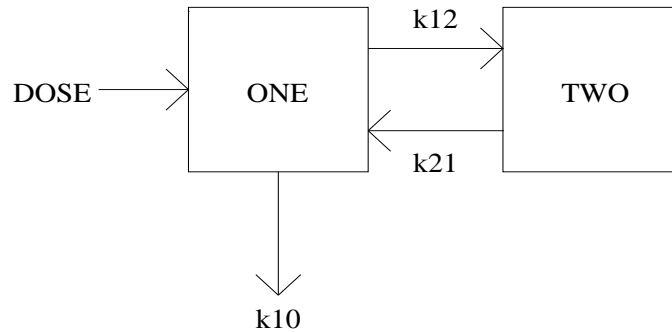


FIGURE 2. The two-compartment model with dose introduced into the plasma compartment.

Solving these two equations subject to  $x(0) = D$  gives

$$x_1(t) = D \{A \exp(-\lambda_1 t) + (1 - A) \exp(-\lambda_2 t)\}$$

and

$$x_2(t) = \frac{Dk_{12}}{(\lambda_2 - \lambda_1)} \{\exp(-\lambda_1 t) - \exp(-\lambda_2 t)\}$$

where

$$\lambda_1, \lambda_2 = (k_{10} + k_{12} + k_{21} \pm [(k_{10} + k_{12} + k_{21})^2 - 4k_{10}k_{21}]^{1/2})/2$$

and  $A = (\lambda_1 - k_{21})/(\lambda_1 - \lambda_2)$ . Figure 3 shows the concentrations of drug in each of the compartments versus time for particular choices of  $(k_{10}, k_{12}, k_{21})$ . We note that the amount of drug initially increases in compartment two during the distribution phase before elimination becomes the dominant process. Absorption into compartment one is immediate here because the drug was introduced as a bolus.

We typically observe drug *concentrations* of drug in compartment one and so we introduce a further parameter  $V_1$  which is the *volume* of compartment one. Note that  $V_1$  is a nominal volume in that its value may be much greater than the volume of blood in the body. This may be due to the binding of the drug to plasma proteins and/or compartment one containing well-perfused tissues, such as the heart and lungs. For this reason this parameter is often referred to as the *apparent volume of distribution*. We then have  $y(t) = x_1(t)/V_1$  to give

$$y(t) = \frac{D}{V_1} \{A \exp(-\lambda_1 t) + (1 - A) \exp(-\lambda_2 t)\}.$$

It is clear from the form of  $y(t)$  that it is possible to work with many parameterizations. For example one may choose to work with the rate constants and volume as used above. There are certain parameters which are of fundamental interest

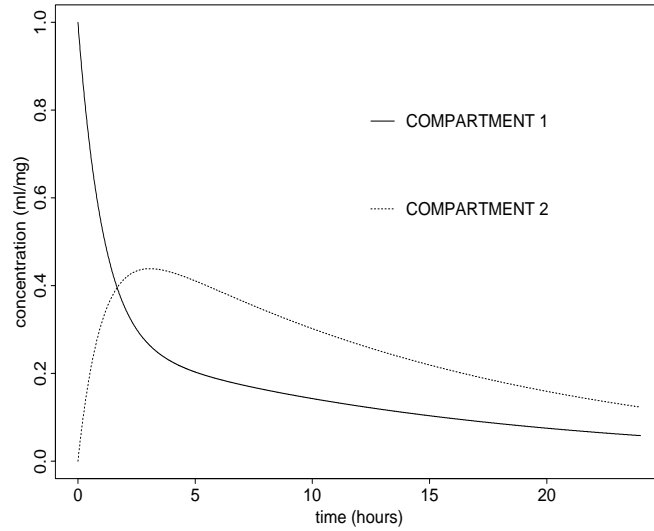


FIGURE 3. Concentration/time relationship for the system represented in Figure 2 with  $D = 1$ ,  $V = 1$ ,  $k_{10} = 0.2$ ,  $k_{12} = 0.5$  and  $k_{21} = 0.3$ .

and for which pharmacokineticists have prior knowledge. In particular *clearance* parameters define the volume of a compartment which is cleared of drug in unit time. For the model defined by (2) and (3) there are two clearance parameters, the clearance from the central compartment  $Cl = V_1 k_{10}$  and the distribution clearance  $Cl_d = k_{12} V_1 = k_{21} V_2$ . The parameter  $Cl$  is defined in general (that is for any compartmental system) as  $D/AUC$  where  $AUC$  denotes the area under the concentration/time curve. Hence a *pharmacokinetic* parameter set would be given by  $(V_1, V_2, Cl, Cl_d)$  where each of the parameters is greater than zero.

The most obvious parameterization from a *statistical* perspective would consider the set  $(A_1, A_2, \lambda_1, \lambda_2)$  where

$$y(t) = D[A_1 \exp(-\lambda_1 t) + A_2 \exp(-\lambda_2 t)]$$

with  $A_1, A_2 > 0$  to ensure positive concentrations. Identifiability is imposed by assuming  $\lambda_1 > \lambda_2 > 0$ . The disadvantage of using this parameterization is that the parameters which the pharmacokineticist has prior information on, namely  $Cl$  and  $V_1$ , are nonlinear transforms of the mathematical set and hence it is not so straightforward to carry out covariate modeling (see Section 6.1).

The identifiability issues become worse as the number of compartments increases. It is in phase I studies that more complex PK modeling is carried out, and in these studies, as described in Section 3, covariate modeling is rarely carried out due to the homogeneity of the individuals. Therefore the mathematical parameterization is appealing here. Godfrey (1983) describes in detail a number of issues, including identifiability, relating to compartmental systems (not just for

PK but for electrical engineering applications also). If interest focuses on the PK parameters then these may be easily found as functions of the mathematical set.

Regardless of the compartmental system used there are a number of fundamental parameters which are of interest to the pharmacokineticists as they *characterise* the drug. The *clearance*  $Cl$  and apparent volume of distribution  $V_1$  have already been discussed. The terminal half-life  $t_{1/2}$  is given by  $\log 2/\lambda_2$  and is used to guide the design of dosage regimens and, under certain circumstances, it is a measure of the time required to attain steady-state. Hence knowledge of  $t_{1/2}$  allows simple dose-determination. One of the major reasons for the importance of  $Cl$  and  $V_1$  is that often information is available on the relationship between these quantities and individual-specific covariates such as weight, gender and serum creatinine concentration (a measure of kidney function).

We have so far only considered the concentration-time profile following a single dose. Particularly during the later phases of drug development multiple doses will be given. The concentration/time profile obtained following multiple dosing can be predicted from the contribution of each individual dose (the principle of superposition, Gibaldi and Perrier, 1982). However, the superposition principle is only valid for linear pharmacokinetic systems which arise when the model parameters are independent of dose. If, for example, the elimination from the central compartment is described by a Michaelis-Menten relationship then this principle does not hold. If PK parameters change with either dose or time, the concentration-time profile at steady-state cannot simply be predicted by superposition and derived quantities such as  $AUC$  and maximum concentration increase either less than or greater than proportionally with dose. Phenytoin provides an example of a drug for which the  $AUC$  and maximum concentration relationship with dose increases greater than linearity. As described in Section 3 it is one of the aims of Phase I of drug development to explore the ‘dose proportionality’ issue. Therapeutically, drugs which display nonlinear pharmacokinetics can be difficult to prescribe.

We shall let  $\theta$  denote the  $p_\theta$ -dimensional vector of PK parameters, whether this be the pharmacokinetic or the mathematical parameterization. We assume that each of the elements of  $\theta$  is defined on the whole real line. So for example for the two compartmental model described above with the mathematical parameterization we will have  $\theta = (\log A_1, \log A_2, \log(\lambda_1 - \lambda_2), \log \lambda_2)$ . The model for the concentration as a function of time  $t$  arising from a particular compartmental system and dosage regimen (times and sizes) will be denoted  $f_1(D, \theta, t)$ . The subscript 1 acknowledges the hierarchy we shall develop in Section 6: here we are modeling a first-stage variable, namely the concentration.

Both the compartmental system and differential equations described above are obviously huge simplification of the body and the processes which act on a drug molecule. However, the resultant concentration-time models have been found empirically to mimic the true concentrations seen.

As mentioned in Section 3, in Phase I of drug development a small number of (usually) healthy volunteers, typically less than 20, are given single doses and subsequently a large number (relative to later phases) of blood samples are taken

in order to determine drug concentrations. Traditional, or non-compartmental PK analyses have used simple numerical methods such as the trapezium rule to estimate the  $AUC$  (and hence  $CI$ ). Other summary measures are similarly computed. During this phase the individuals in general form a very homogeneous group. Due to the abundance of data per individual the statistical modeling at this stage can often utilise two- or three-compartment models. The number of compartments is generally chosen as the maximum number which can be fitted. As drug development proceeds the individuals become more heterogeneous and the data per individual becomes sparser. Consequently, traditionally, simpler compartmental systems are assumed (though see the discussion of Sections 4.3 and 7.3).

#### 4.2 PK error models

We wish to model the differences between the observed concentrations  $y$  and the modeled concentrations  $f_1$ . These error terms, which describe the difference, will represent not only assay precision but also model misspecification. Particularly at the later stages of drug development this latter quantity is likely to be substantial. The compartmental systems described above are large simplifications. Also the environment within which the trials are performed is important. Early trials will be carried out within the drug company and so recorded sampling times and administered doses are likely to be accurate, in a hospital environment this is far less likely. Since the model misspecification is likely to vary smoothly as a function of time one might expect the errors to be correlated but such models have rarely been fitted due to the sparsity of data. Davidian and Giltinan (1995) describe such models.

Within the PK literature a frequently-used error model is

$$y = f_1(D, \theta, t) + \epsilon^y$$

with the error terms  $\epsilon^y$  taken as independent, zero mean normal random variables with variance  $f_1(D, \theta, t)^\gamma \sigma_y^2$  with  $\gamma \geq 0$ . Wakefield (1996a) uses such a model and treats  $\gamma$  as an unknown parameter. Davidian and Giltinan (1993) discuss various error models. The case  $\gamma = 2$  results in a constant coefficient of variation which often mimics assay precision. An alternative to this model which produces an approximate constant coefficient of variation, and is both statistically and biologically appealing, is given by

$$\log y = \log f_1(D, \theta, t) + \epsilon^y$$

with the  $\epsilon^y$  now taken as independent zero mean normal random variables with variance  $\sigma_y^2$ .

It is standard practice in analytical chemistry to set a lower limit of quantification below which concentrations are discarded and treated as missing. One common strategy is to accept only concentrations which have a coefficient of variation (based on a calibration curve) below 20%. Note that although zeros are recorded

for those concentrations below the lower limit of quantification this limit is typically known and so these observations may be treated as censored in the conventional way (Wakefield and Racine-Poon, 1995). Ignoring these zeros leads to estimation bias since the observations are clearly not missing at random.

### 4.3 PD models

Pharmacodynamic responses can either be quantal or continuous. An example of a quantal response is pain score recorded in an analgesic trial. For example Sheiner, Beal and Dunne (1997) analyse data recorded on a four-point scale using a mixed-effects approach with a cumulative logit model.

Here we shall concentrate on continuous responses since these are relevant to the application which we shall describe in Section 7. Let  $z(t)$  denote the response (or some transform of the response such as the logarithm) measured at time  $t$ . A general model is then given by

$$z = g_1(\theta, \phi, t) + \epsilon^z \quad (4)$$

with the  $\epsilon^z$ 's taken as independent zero mean normal random variables with variance  $\sigma_z^2$ . So the modeled response depends not only on a  $p_\phi$ -dimensional vector of PD parameters  $\phi$ , but also on the PK parameters  $\theta$  since the concentration of drug is assumed to influence the resultant response. Here and throughout the paper we use  $p(\cdot)$  as a generic notation for a probability density function. The distribution of  $y$  and  $z$  is given by

$$p(y, z|\theta, \phi) = p(y|\theta)p(z|\theta, \phi).$$

We are therefore making the assumption that  $z$  is conditionally independent of  $y$ , given  $\theta$ . In particular this means that the errors in concentrations and responses at the same time point are independent.

Equation (4) represents an ideal analysis in which inference for the PK parameters will also be influenced by the response data. Often when PK/PD data are jointly modeled this relationship is simplified. Suppose for example that the response at time  $t$  depends directly on the true concentration at time  $t$ ,  $f_1(D, \theta, t)$ . The easiest way to model this relationship would be to simply plug in the observed concentrations. This approach does not acknowledge that the concentrations are measured with error and so we are faced with an errors-in-variables problem. If it were assumed that the relationship between response and concentration were linear then, from standard errors-in-variables results (Carroll, Ruppert and Stefanski, 1995), one might expect the coefficient describing the strength of this relationship to be *attenuated*. If the response/concentration relationship is nonlinear then it is not clear what the effect of the errors-in-variables will be. A more refined procedure would be to obtain estimates of the PK parameters  $\hat{\theta}$  and then regress the response on  $f_1(D, \hat{\theta}, t)$ . The disadvantage of this is that the PD data will not inform the estimation of  $\theta$  and the variability in the PD parameters will

be underestimated since the variability in  $\hat{\theta}$  has not been acknowledged. One possibility here would be to allow fluctuations about the observed concentration  $y$  via a simple errors-in-variables approach, without explicitly considering a PK model. In some instances one of these more simplistic approaches may be followed when we *a priori* know that our PK and PK/PD models have been greatly simplified due to sparsity of data. For example when we use a one-compartment PK model for a two-compartment system we know that the high peak concentrations will be systematically underestimated whereas concentrations near the end of the administration period (trough concentrations) will be systematically overestimated. In this case fitting the full PK/PD model simultaneously will lead to bias in the PD parameter estimation. Similarly when the PK/PD link or PD model is oversimplified bias will be expected when the joint PK/PD model is used.

Another consequence of taking one of the simpler approaches is that it is not so straightforward to determine the response consequences of specific doses since we have not carried out joint estimation.

Often the pharmacological response is not directly related to concentration but lags behind the concentration/time profile. This can occur either when the drug acts as a precursor which gives rise to the response or when the site of action of the drug is not blood but a tissue into which the drug must distribute before eliciting an effect. A discussion of models used to handle this situation is given in Holford and Sheiner (1981).

## 5 Population PK/PD

Population pharmacokinetics can be defined as the study of the variability in outcome between individuals when standard dosage regimens are administered, the outcome usually being the plasma concentration-time profile. It is of interest to both quantify the variability of this response within the population and to account for that variability in terms of patient specific variables, such as age, sex, weight, disease state, etc. The current interest in population pharmacokinetics stems from the concern that the pharmacokinetics of new drugs are not studied in relevant populations, that is, patients likely to receive the drug, at an early enough stage in the drug development program. In particular the United States Food and Drug Administration (FDA) (Temple, 1983; Temple, 1985), and others (Abernathy and Azarnoff, 1990), are concerned that the pharmacokinetics of a new drug should be studied in elderly populations ‘so that physicians will have sufficient information to use drugs properly in their older patients’ (Food and Drug Administration, 1989). The obvious time to collect PK information on the target population is during large-scale clinical trials carried out during Phase III of the drug development program. However, because of logistic and ethical reasons, it is improbable that intensive experimentation can be carried out on each and every patient. At best one could hope for one or two blood samples per patient. Therefore traditional PK analysis, which involves the determination of an individual’s PK parameters,

is untenable (see Section 4.1). Instead data analysis techniques that focus on the central tendency of the PK information and are capable of utilizing very sparse data have to be employed. Population pharmacokinetics has come to mean the design, execution and analysis of PK studies involving sparse data, although the data analysis techniques can be applied to data obtained from conventional PK studies. The label *population pharmacokinetics* is perhaps unfortunate but it does convey the sentiment that interest is focused on the population rather than the individual.

The implementation of a population approach within the drug development program is the subject of much debate (Colburn, 1989). It has been suggested that a 'population screen' be employed in which blood samples are taken from a wide range of individuals so that, essentially, the concentration-time profile is covered within the population (Sheiner and Benet, 1985). The advantages of such an approach are that data are collected in the target population, an assessment of the variability within the population is obtained and, hopefully, the factors that control that variability may be discovered. Although the goals are indisputable, much concern has been expressed about the logistics of implementation of a population approach during Phase III of the drug development program. A common statement that is made is 'garbage in, garbage out'. It should be pointed out that this criticism can be made of any poorly designed or executed study, not just population studies. However there are particular problems associated with Phase III studies due to the fact that they are in general multicentre and in many cases in an outpatient setting. Compliance and accurate timing of both dosing and sampling are clearly critical issues. At present there are virtually no guidelines on experimental design, both in terms of sample timing and subject numbers, particularly within subgroups. Similarly, we have no idea of the power of the approach to detect important inter-subject differences and overall there is no hard data on the cost-to-benefit ratio. At present we are still on the learning curve.

To date most population modeling in phase III has been concerned with pharmacokinetic data, although some mixed effects modeling has been carried out on continuous response variables obtained from small scale controlled clinical trials (Pitsiu, Parker, Aarons and Rowland, 1993). Sheiner, Beal and Dunne (1997) provide an elegant application of mixed effects modeling to pain score data obtained from an analgesic study. The current status of population PK/PD is simply a reflection of the lack of PK/PD modeling in drug development. However with the growing realization of the importance of PK/PD modeling to drug development, we can expect to see an increase in the activity of population PK/PD and population PD.

We finally note that information obtained from population analyses during drug development may also be used after the drug has been marketed for the *individualization* of dosage regimens. Such on-line therapeutic drug monitoring is often carried out in a hospital environment when patients with an acute condition require careful tuning of doses. Typically only sparse data are available and so prior information, in the form of estimates of the parameters of the population distribution, are required. Vozeh and Steimer (1985) describe this technique and Wakefield (1994, 1996b) Bayesian sampling-based approaches.

## 6 Statistical aspects of population PK/PD

### 6.1 A three stage hierarchical model

It is natural to model population PK/PD data hierarchically since this allows the variability in concentrations/responses to be separated into within-individual and between-individual components. The joint PK/PD aspect is important since one can see how much of the variability in responses is due to variability in concentrations. At the first stage of the hierarchy *within-individual* modeling of the PK/PD data are carried out. The data of each individual are assumed to follow the same underlying functional form (that is  $f_1$  and  $g_1$ ) but the parameters of these forms take different values for different individuals. At the second stage the *between-individual* differences are modeled by assuming that the individual PK and PD parameters arise, after accounting for individual-specific covariates, from a common probability distributions. An excellent review of the statistical aspects of population PK/PD modeling is given by Davidian and Giltinan (1995), Yuh *et al* (1994) provide a bibliography of population PK applications and methodological approaches, and Steimer *et al* (1994) give an introduction to population PK/PD ideas in drug development.

#### First stage model

Let  $y_{ij}$  denote the concentration (in blood, plasma or urine) of individual  $i$  at time  $t_{ij}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ . Then the model for concentration is given by

$$\log y_{ij} = \log f_1(D_i, \theta_{ij}, t_{ij}) + \epsilon_{ij}^y$$

where  $\epsilon_{ij}^y$  are independent and identically distributed as  $N(0, \sigma_y^2)$  and  $\theta_{ij}$  represent the PK parameters of individual  $i$  at time  $t_{ij}$ . We index by both  $i$  and  $j$  because at the second stage of the hierarchy we will model the PK parameters as functions of individual covariates which may be time-varying.

Let  $z_{ij}$  denote the response of individual  $i$  at time  $t_{ij}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ . For notational simplicity, and because it is the case in the studies described in the next section it is assumed that the PK and PD data were collected from the same individuals and at the same time points. In general this is not a requirement, PK and PD data on the same individual need not be collected at the same time points. In fact some individuals may just contribute PK or PD data though to learn about the PK/PD relationship there must be some individuals with both types of data. We may also have PD data which is a cumulative response over some time period.

The response model is given by

$$z_{ij} = g_1(\theta, \phi_{ij}, t_{ij}) + \epsilon_{ij}^z$$

where  $\epsilon_{ij}^z$  are independent and identically distributed as  $N(0, \sigma_z^2)$ .

*Second stage model*

Let  $\mu_\theta$  and  $\Sigma_\theta$  represent location and scale parameters for the PK parameters and  $\mu_\phi$  and  $\Sigma_\phi$  represent location and scale parameters for the PD parameters. In the most general case  $\mu_\theta$  and  $\mu_\phi$  are vectors of length  $p_\theta$  and  $p_\phi$ , respectively and  $\Sigma_\theta$  and  $\Sigma_\phi$  are matrices of dimension  $p_\theta \times p_\theta$  and  $p_\phi \times p_\phi$ , respectively. In some cases, particularly with sparse data, some elements of  $\theta$  and/or  $\phi$  may be taken to be fixed effects, thereby reducing the dimensionality of  $\Sigma_\theta$  and/or  $\Sigma_\phi$ . In other instances the off-diagonal elements may be taken to be zero. A priori, however, enough empirical evidence has been gathered to follow the most general model – simplifications result when the data do not inform about particular elements of  $\theta$  and  $\phi$ . In such cases a Bayesian may stay with the general model and place informative priors on parameters for which there is little information in the data. In Section 7 such a model is used.

Let  $X_{ij}$  denote a vector of individual-specific covariates measured at time  $t_{ij}$ . The relationship between the PK parameters and covariates is then modeled as

$$\theta_{ij} = f_2(\mu_\theta, X_{ij}) + \delta_i^\theta.$$

The ‘error’ terms  $\delta_i^\theta$  model the difference between the PK parameters of individual  $i$  at time  $t_{ij}$  and that predicted by the covariate model  $f_2$ . Note that  $\delta_i$  is independent of time, that is  $j$ . In general the function  $f_2$  is taken to be linear, that is  $f_2(\mu_\theta, X_{ij}) = X_{ij}\mu_\theta$ , see for example Davidian and Gallant (1993) and Wakefield (1996a).

The choice of which components of  $X$  to include for each element of  $\theta$  is a very difficult problem. It is essentially a multiple regression in which the dependent variable  $\theta$  is multivariate and unobserved. Covariate selection in population PK analyses has often proceeded via a forward selection procedure, with graphical plots driving the inclusion of covariates ( Maitre, Buhner, Thomson and Stanski ,1991), and analytical approximations being used to evaluate the likelihood of the population parameters. Using these plots to drive a forward selection procedure is hazardous in several ways. It is difficult to decide on which of a discrete or a continuous covariate is to be included first. Adding more than one covariate at a time is also dangerous since the covariates themselves are correlated. Choosing whether a particular covariate should be included for different elements of the  $\theta$  vector is also difficult. A covariate may be needed for volume only (say) but may appear to be necessary for clearance also because of correlation in the random effects distribution. The problems of forward selection are well-documented, Miller (1990). The approach described above, which is typical of those taken in population PK analyses, has a number of additional disadvantages. The procedure is computationally expensive as each of the analyses requires a numerical minimization to be carried out. The test statistic has only an asymptotic distribution under the null hypothesis and the statistic itself is being calculated via an analytic approximation. It is difficult to assess the impact of these approximations for a particular dataset. Simulating the test statistic under the null hypothesis in order to obtain a Monte Carlo significance level is computationally prohibitive. Most importantly, the significance of the covariates is only being judged in a *statistical*,

and not in a *clinical*, sense. For example suppose the statistical significance levels of two regressors are of similar size. This does not imply that each has equal clinical significance. The regression coefficient for a particular covariate may be statistically significant but may have little effect on the clinical aim. Conversely a regressor which does not attain statistical significance may have a clinical impact. Wakefield and Bennett (1996) consider this problem in a specific application when the design of a dosage regimen was the objective.

We argue that the choice of covariates is driven by the aim of the analysis and prior knowledge, both from previous studies with the same or similar drugs and from pharmacokineticists. To this end it is relevant to recall the meaning of the components of  $\theta$ . Although the mathematical parameterization is the most convenient for overcoming identifiability problems, it is not the parameterization which is natural for covariate modeling. In particular, pharmacokineticists frequently have prior beliefs concerning the effect of the covariates on clearance and volume. In general the logarithms of clearance and volume are not linear functions of the mathematical parameters and so a general strategy is difficult to determine.

The relationship between the PD parameters and covariates is similarly modeled as

$$\phi_{ij} = g_2(\mu_\phi, X_{ij}) + \delta_i^\phi.$$

The interpretation/difficulties with this modeling are similar to those for the PK modeling.

We now consider the joint distribution of  $\delta^\theta, \delta^\phi$ . The basic assumption we make is that

$$p(\delta^\theta, \delta^\phi | \mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi) = p(\delta^\theta | \mu_\theta, \Sigma_\theta) p(\delta^\phi | \mu_\phi, \Sigma_\phi).$$

A number of approaches have been suggested for the modeling of  $p(\delta^\theta | \mu_\theta, \Sigma_\theta)$ . The choice of this distribution is far less obvious than the choice of the distribution at the first stage for which assay data are available. At the second stage the distribution of the PK parameters in their untransformed form, after conditioning on covariates, has often been found empirically to be skewed. The obvious choice is a  $p_\theta$ -dimensional lognormal distribution. This was used implicitly by Beal and Sheiner (1982) and explicitly by Lindstrom and Bates (1990). Wakefield, Smith, Racine-Poon and Gelfand (1994) used a log student distribution in an attempt to robustify the modeling.

Figure 4 represents a directed graph of the joint PK/PD model of a single individual.

There is always the possibility that a covariate which is an important predictor has not been measured, however. This will result in a mixture distribution. Consequently a number of authors have avoided a specific parametric family. Mallet (1986) proposed to allow the distribution to be completely general and obtained a nonparametric maximum likelihood (NPML) estimate (which is discrete). The NPML method suffers from a number of disadvantages. The first stage variance,  $\sigma_y^2$ , is assumed known and no standard errors are produced. No estimates of uncertainty are produced by the method and so it is not clear whether interesting

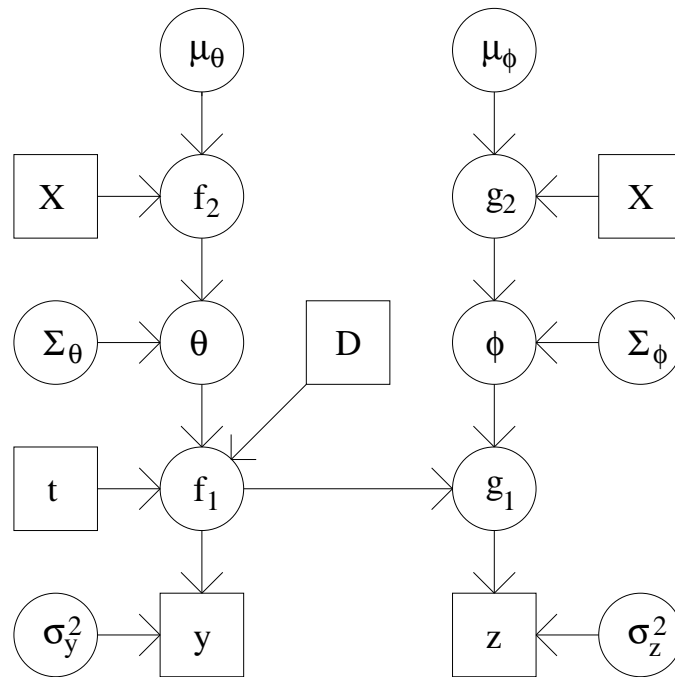


FIGURE 4. Directed graph for the population PK/PD model. Observable quantities are placed in square boxes,  $X$  represents individual-specific covariates,  $D$  dose,  $t$  time,  $y$  concentration and  $z$  response.

features such as bimodalities are real or merely artifacts of the data. Mentré and Mallet (1994) extended the original method to allow a completely general specification at this stage and obtain a nonparametric maximum likelihood estimate of the relationship between  $\theta$  and  $X$  by assuming that the joint distribution of  $(\theta, X)$  is being estimated. The difficulty with this is that a probability distribution of high dimension is being estimated with very sparse data.

Davidian and Gallant (1993) propose a semiparametric method in which the assumption of smoothness is made at the second stage and the random effects are modeled as arising from a distribution that is represented as a normal times a polynomial.

Wakefield and Walker (1997), Muller and Rosner (1997), and Walker and Wakefield (1998) take a Bayesian nonparametric approach using Dirichlet process priors.

### Third stage model

At the third stage of the hierarchy priors are specified for the population parameters  $\mu_\theta$ ,  $\mu_\phi$ ,  $\Sigma_\theta$ ,  $\Sigma_\phi$ ,  $\sigma_y^2$  and  $\sigma_z^2$ . We first note that those elements of  $\mu_\theta$ ,  $\mu_\phi$  that correspond to nonlinear parameters at the first stage, proper priors are required in order to guarantee propriety of the posterior distribution. For the same reason we also need proper priors for  $\Sigma_\theta$  and  $\Sigma_\phi$ . For computational convenience (see next Section) normal prior distributions  $N(c_\theta, C_\theta)$ ,  $N(c_\phi, C_\phi)$  are assumed for  $\mu_\theta$ ,  $\mu_\phi$ , respectively; Wishart distributions  $W(r_\theta, (r_\theta R_\theta)^{-1})$ ,  $W(r_\phi, (r_\phi R_\phi)^{-1})$  are assumed for  $\Sigma_\theta^{-1}$ ,  $\Sigma_\phi^{-1}$ , respectively; and gamma distributions  $Ga(a_\theta^2/A_\theta, a_\theta/A_\theta)$ ,  $Ga(a_\phi^2/A_\phi, a_\phi/A_\phi)$ , for  $\sigma_y^{-2}$  and  $\sigma_z^{-2}$ , respectively. The values  $c_\theta$ ,  $c_\phi$ ,  $R_\theta$ ,  $R_\phi$ ,  $a_\theta$  and  $a_\phi$  represent prior guesses for the relevant parameters and  $C_\theta$ ,  $C_\phi$ ,  $r_\theta$ ,  $r_\phi$ ,  $A_\theta$  and  $A_\phi$  represent the precision of these estimates.

## 6.2 Implementation details

We let  $y_i = (y_{i1}, \dots, y_{in_i})$ ,  $z_i = (z_{i1}, \dots, z_{in_i})$ ,  $y = (y_1, \dots, y_N)$ ,  $z = (z_1, \dots, z_N)$ ,  $\delta^\theta = (\delta_1^\theta, \dots, \delta_N^\theta)$  and  $\delta^\phi = (\delta_1^\phi, \dots, \delta_N^\phi)$ . In general interest will focus upon the population parameters  $\mu_\theta$ ,  $\Sigma_\theta$ ,  $\mu_\phi$ ,  $\Sigma_\phi$ . The posterior for these quantities is given by

$$p(\mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi \mid y, z) = \int p(\mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi, \sigma_y^2, \sigma_z^2 \mid y, z) d\sigma_y^2 d\sigma_z^2$$

where

$$\begin{aligned} p(\mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi, \sigma_y^2, \sigma_z^2 \mid y, z) &\propto p(y, z \mid \mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi, \sigma_y^2, \sigma_z^2) \\ &\times p(\mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi, \sigma_y^2, \sigma_z^2) \end{aligned} \quad (1)$$

and, utilizing the conditional independencies discussed previously and represented in Figure 4,

$$p(y, z \mid \mu_\theta, \Sigma_\theta, \mu_\phi, \Sigma_\phi, \sigma_y^2, \sigma_z^2) = \prod_{i=1}^N \int p(y_i \mid \delta_i^\theta, \mu_\theta, \sigma_y) \times$$

$$p(z_i | \delta_i^\theta, \mu_\theta, \delta_i^\phi, \mu_\phi, \sigma_z) \times p(\delta_i^\theta | \Sigma_\theta) p(\delta_i^\phi | \Sigma_\phi) d\delta_i^\theta d\delta_i^\phi.$$

Due to the nonlinear nature of the first stage PK and PD models these integrals are analytically intractable. Racine-Poon and Wakefield (1998) provide a review of estimation techniques for population PK modeling. Within the pharmaceutical industry the package NONMEM (Beal and Sheiner, 1993) is often used. Within this package various analytical approximations are available for calculating the required integrals and the resultant likelihood is then numerically maximised. In particular the First Order (FO), First Order Conditional Estimation (FOCE) and Laplacian methods may be used. The FO method takes a linearisation about the population mean whilst the FOCE is essentially the method of Lindstrom and Bates (1990). Pinheiro and Bates (1995) compare various analytical and numerical approximations. Inference is made through asymptotic arguments. Unfortunately it is, in general, difficult to assess the adequacy of the approximation and the appropriateness of the asymptotics.

Mallet's NPML method is computationally fast but the numerical maximisation is sensitive to the starting ranges specified for the parameters. The parameters of the normal times polynomial form of Davidian and Gallant (1993) are estimated by maximum likelihood. Again the numerical maximization is difficult and it is recommended that a 'wave of starting values be used'. The likelihood is evaluated using numerical or Monte Carlo integration. These and other techniques are described in Davidian and Giltinan (1995) and Vonesh and Chinchilli (1997). Wakefield and Walker (1997) compare the methods of Mallet and Davidian and Gallant with a nonparametric Bayesian method. For the latter convergence and prior sensitivity are important issues.

The first Bayesian approach to population PK modeling is due to Racine-Poon (1985) using an approximation to the first stage model that allows an EM-type algorithm to be used. Markov chain Monte Carlo (Smith and Roberts, 1993) was first used in the context of population PK models by Wakefield, Smith, Racine-Poon and Gelfand (1994) who used a Gibbs sampler. For the priors described in the previous section and the first stage models outlined in Section 4 all of the required conditional distributions assume known forms apart from the random effects. Wakefield, Smith, Racine-Poon and Gelfand (1994) used the ratio-of-uniforms method (Wakefield, Gelfand and Smith, 1991) for generating from these distributions whilst Gilks, Best and Tan (1995) used the Metropolis Adaptive Rejection Sampling (MARS) algorithm (see also Gilks, Neil, Best and Tan, 1997). Both of these methods are only practically feasible for univariate conditional distributions which can lead to slow mixing. Subsequently a Hastings-Metropolis (Metropolis *et al*, 1953, Hastings, 1970) step has been used to generate from the multivariate conditional (Bennett, Racine-Poon and Wakefield, 1996). This is the approach that was used for the analysis in Section 7.

## 7 Population PK/PD of Recombinant Hirudin

### 7.1 Pharmacology of hirudin

REVASC<sup>TM</sup> (recombinant desulfatohirudin, produced in yeast by recombinant DNA technology), is a selective thrombin inhibitor nearly identical in protein structure to the natural leech anticoagulant, hirudin variant 1. Preclinical studies showed REVASC<sup>TM</sup> to be a highly specific thrombin inhibitor and a potent anticoagulant. REVASC<sup>TM</sup> demonstrated positive results in various preclinical models of venous, arterial and foreign surface thrombosis, as well as arterial thrombolysis. In vitro, the dose-response curve is shallower than that of unfractionated heparin (a competitor drug) enabling it to be used over a much greater concentration range (more than two orders of magnitude) and to higher levels of anticoagulation (up to three times control APTT (activated partial thromboplastin time)). Based upon these results, REVASC<sup>TM</sup> was recommended for clinical development for prevention of venous and arterial thrombosis and reduction of reocclusion during and after thrombolysis. In acute and chronic toxicological studies, REVASC<sup>TM</sup> displayed no evidence of systemic toxicity. Dose is limited by the extension of the pharmacological activity, that is inhibition of blood coagulation.

In a biotransformation study in six healthy male volunteers following a single intravenous infusion of 3mg/kg/h for 6 hours, two metabolites were detected in trace amounts in the urine. They represented only about 7% of the administered dose. No other metabolites were detected. The observed data suggested that REVASC<sup>TM</sup> is eliminated and metabolized in the kidney. Total unitary excretion of intact REVASC<sup>TM</sup> amounted to 40–50% of the total dose. The apparent renal clearance was approximately 80 mL/min. The difference between the renal clearance and the creatinine clearance (120 mL/min) may be accounted for by resorption. The non-renal clearance, about 20% of the administered dose, was essentially due to this resorption process and an additional undefined non-renal clearance. After intravenous bolus administration, the steady-state volume of distribution of REVASC<sup>TM</sup> was about 18L. It may be concluded from this small steady-state volume that REVASC<sup>TM</sup> does not enter cells but stays in the circulation and extracellular space. The central volume of distribution is about 5L, approximately the plasma volume. This also suggested that a minimum of two compartments is required to model the pharmacokinetics of REVASC<sup>TM</sup>. The pharmacological effect of REVASC<sup>TM</sup> is based on its binding to thrombin. The thrombin-hirudin complex, which is much too large to be cleared by renal filtration, is present in trace amounts in plasma in healthy volunteers. The binding to thrombin is therefore not a major route of elimination for healthy volunteers.

Human studies demonstrated that after both intravenous and subcutaneous administration, areas under the plasma concentration curves are dose proportional. Mean terminal elimination half-lives were similar (2–2.5 hr), and dose-independent. Clearance is primarily renal and dose-independent (2.1–2.5 mL/min/kg). Similarly, the steady state volume of distribution is dose-independent (0.25 L/kg). Following intravenous administration, plasma concentrations are best

described by a three exponential model. The absolute bioavailability of subcutaneous doses is approximately 100% and half-lives were similar for the subcutaneous and intravenous routes. Total REVASC<sup>TM</sup> plasma clearance is reduced in the elderly (1.62 ml/min/kg) compared to young subjects. REVASC<sup>TM</sup> is assayed by a sandwich ELISA method, based on the recognition of the analyte by a monoclonal capture antibody and a polyclonal signal antibody. The assay measures free drug, *i.e.* not drug bound to thrombin.

Phase I clinical trials in normal volunteers and patients demonstrated that the APTT response to REVASC<sup>TM</sup> was dose-dependent and consistent after intravenous or subcutaneous dosing. The details of the studies which demonstrated these findings are given in the next section. APTT is strongly correlated with REVASC<sup>TM</sup> plasma levels. There is no evidence for delay in the onset of anticoagulant effect or from an extended duration of activity beyond the actual presence of REVASC<sup>TM</sup> in plasma.

We first describe each of the study objectives and designs, and then the models that were used for analysis.

## 7.2 Study description

In this section we describe five different phase I studies that were carried out for REVASC<sup>TM</sup>. The first study is reported in Wakefield and Racine-Poon (1995) whilst studies 2–5 are described in Racine-Poon and Wakefield (1996). We then describe a Phase II study upon which we shall concentrate.

### Study 1

This study consisted of four groups of four volunteers, each of whom received on day 1 of the study an intravenous bolus injection of REVASC<sup>TM</sup>. Each group received one of the doses 0.01, 0.03, 0.05 and 0.1 mg/kg. Approximately 28 hours later they received an intravenous bolus followed by a constant rate intravenous infusion of heparin over 24 hours during which they received a second intravenous bolus injection of REVASC<sup>TM</sup> at the same doses as given on day 1. Hence the concentration of heparin was held constant over the second administration of REVASC<sup>TM</sup>. Heparin also inhibits blood clotting forming and it was envisaged that the two drugs may be co-administered and so it was of interest to see whether there was any interaction between them. Blood samples were taken at 0 hours, immediately before the bolus of drug, and between 0.08 hours and 24 hours subsequently. Sixteen blood samples were taken in total from each individual on each of days 1 and 3. For this study, and each of the studies we now describe, the plasma concentration of REVASC<sup>TM</sup> of each sample was determined along with the clotting measure APTT.

### Study 2

The aim of this study was to investigate the absorption characteristics of the subcutaneous administration route. To this end it is of interest to determine whether

the PK model and the PK/PD relationship were altered by the administration route. In terms of the pharmacokinetics it was of interest to see whether the distribution and elimination phases are the same for the two administration routes and also to estimate the *bioavailability*. This quantity is defined as the proportion of the administered dose which is absorbed into the bloodstream and is obviously of importance when decisions concerning the size of the administered dose are being considered. The bioavailability can only be estimated when concentration data are available from the intravenous route *and* from the alternative route of administration (subcutaneous here). This is because we do not measure absolute levels of drug and from concentrations alone the total amount of drug that has been absorbed cannot be determined. Sixteen young healthy male volunteers were administered each of the intravenous and subcutaneous doses in a two-period randomized crossover open (that is not 'blind') study. Note that to carry out a blind trial here two injections would have to be given at each administration time, one intravenous and one subcutaneous. Eight of the volunteers received, on separate days, doses of 0.3mg/kg of body weight, one intravenous and one subcutaneous, while the remaining eight volunteers each received 0.5mg/kg in both forms. Notice, therefore, that the doses depend on the weight of the individual. On each of the days of administration blood samples were taken during a 24-hour period following administration. Figure 5 shows the concentration/time data for four of the individuals and both routes of administration.

#### *Study 3*

The purpose of this study was to investigate the possible modification of the PK profile and the PK/PD relationship by the covariate age. The study was carried out with 12 elderly volunteers who had a normal range of renal function. Each of the six volunteers received a single subcutaneous application. Six of the volunteers received a dose of 0.3mg/kg, and the remaining volunteers received 0.5mg/kg. Blood samples were taken for a 24-hour period after application.

#### *Study 4*

The purpose of this study was to investigate the PK and PK/PD relationship after repeated subcutaneous dosing. Eight healthy volunteers received subcutaneous dosing twice daily for six consecutive days. Four of the volunteers received 0.3mg/kg every 12 hours, whereas the remaining four received 0.5mg/kg every 12 hours. Two blood samples were taken in each administration period, one immediately before administration (trough) and one 3 hours after administration (peak). On the seventh day three samples were collected following the morning dose. This study was reported in Verstraete *et al* (1993).

#### *Study 5*

The purpose of this study was to investigate the PK and PK/PD relationship at equilibrium. If an intravenous infusion is continued indefinitely, a constant concentration/time profile results as the amount of drug entering the body at any instant of time is equal to the amount being eliminated from the body. In prac-

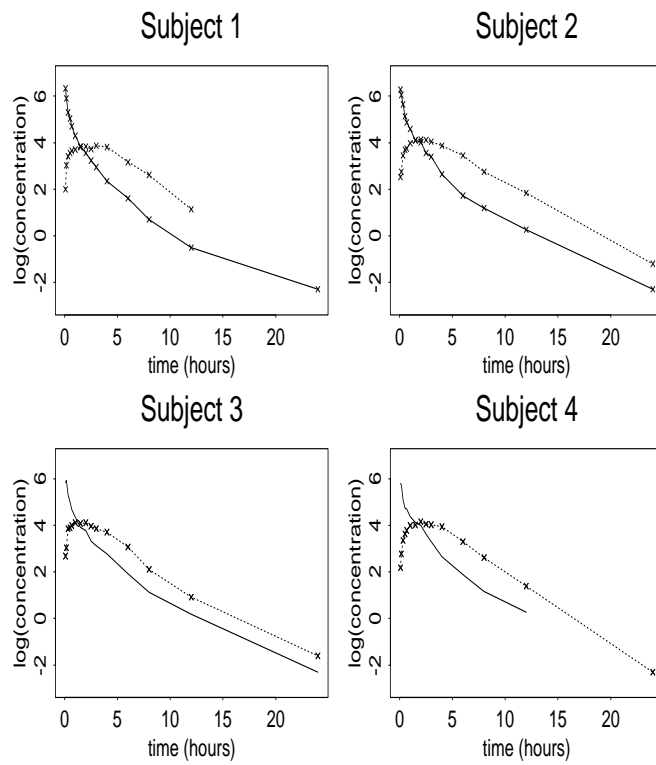


FIGURE 5. Concentration/time data for four individuals from Study 2 who received doses of 0.3mg/kg. The solid line denotes concentrations from the intravenous bolus route, the dashed from the subcutaneous.

tice, equilibrium is reached after some finite time that depends on the elimination rate constant. Eight young male volunteers received a constant infusion over 72 hours, four at a rate of 0.2mg/kg per hour and four at 0.3mg/kg per hour. Blood samples were taken during the infusion period and for 12 hours after the infusion was complete.

### Study 6

In this phase II dose-finding trial, 301 orthopaedic patients undergoing total hip replacement therapy were administered subcutaneous doses twice daily, following the operation, with REVASC<sup>TM</sup>. More details of this study may be found in

Eriksson *et al* (1996). The principal aim of the trial was to find an appropriate dose for later trials. Consequently each of the patients received repeated administration of one of the doses 10, 15 or 20mg. Patients in another arm of the trial received unfractionated heparin as an active comparator treatment; we do not consider these patients in this analysis. Four days after the operation two blood samples were taken after the morning dose and the concentration of REVASC<sup>TM</sup> was determined. By this time *steady-state* concentration levels had been attained. This means that the amount of drug absorbed daily is equal to the amount eliminated. In the trial protocol it was specified that the samples should be taken approximately 2 and 10 hours after the dose. Due to the hospital environment the actual times, which were recorded, varied quite considerably about these nominal, that is protocol pre-determined, values. The nominal 12 hour doses ranged between 10 and 14 hours. The 2 and 12 hour measurements correspond, approximately, to peak and trough concentrations. Figure 6 shows all of the (log) concentration data plotted versus time. We can see the spread in the sampling times around 2 and 12 hours. We see that there is a tendency for the concentrations corresponding to the 20mg/10mg doses to be high/low, but there is a large amount of variability.

At each of the sampling times the APTT was also determined, along with an additional baseline measure before the randomization. These data are plotted in Figure 7.

Patient specific covariates were also recorded. These were: dose (mg), weight (kg), height (cm), age (years), gender, smoking and serum creatinine concentration (mg/dl).

For this study the clinical outcomes of primary interest were incidence of deep vein thrombosis and safety, not the clotting measure APTT though the latter was measured. APTT is linked to bleeding, which is one potential well-known complication of anticoagulant drugs such as heparins and REVASC<sup>TM</sup>, and to efficacy and so to illustrate PK/PD modeling we shall investigate the relationship between individual PK and PD parameters and covariates and the links to safety/efficacy.

Table 1 summarises the 6 studies which we consider here and illustrates how the study design (numbers of subjects and observations) changes through drug development.

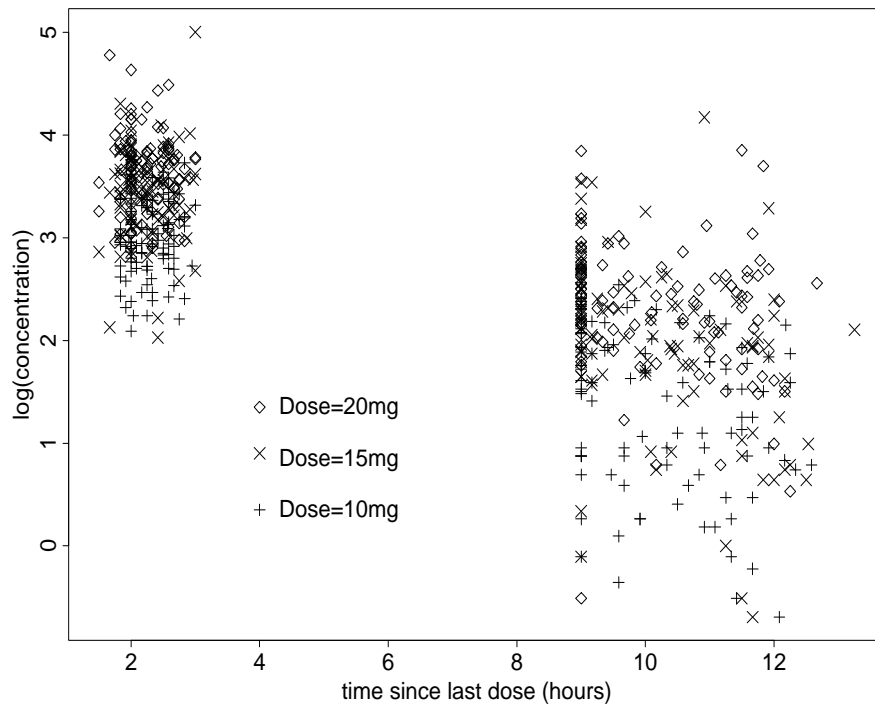


FIGURE 6. Concentration/time data for the 301 individuals of study 6.

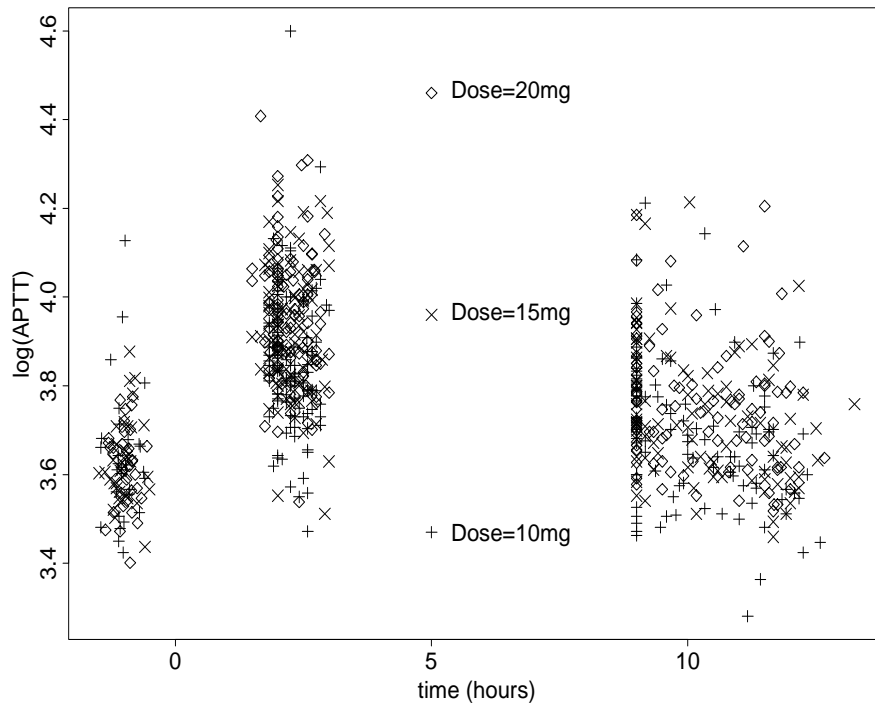


FIGURE 7. APTT/time data for the 301 individuals of Study 6. The times of the APTT measurements before time zero are nominal, that is planned, times, they are pre-dose measurements.

Study	Study population	No. of subjects/ samples	Aim of study, effect of:
1	Healthy volunteers	16 and 16/16	Coadministration (Single dose IV bolus)
2	Healthy volunteers	16 and 16/16	Absorption route (Single dose IV/SC)
3	Elderly volunteers	12/16	Age (Single dose SC)
4	Healthy volunteers	8/27	Repeated doses (SC)
5	Healthy volunteers	8/18	Steady-state dosing (IV Infusion)
6	Patients	301/2 <sup>a</sup> , 3 <sup>b</sup>	Covariates (SC repeated doses)

TABLE 1: Six studies of REVASC<sup>TM</sup>, IV and SC denote intravenous and subcutaneous, respectively. <sup>a</sup> PK samples, <sup>b</sup> PD samples.

### 7.3 Models

#### Phase I

##### *First stage kinetic models*

Based on information from kinetics specialists (see Section 7.1), the kinetics profile after a single intravenous dose of size  $D$  can be described by a sum of three exponential terms,

$$f_1(D, \theta, t) = D \times [A_1 \exp(-\alpha_1 t) + A_2 \exp(-\alpha_2 t) + A_3 \exp(-\alpha_3 t)] \quad (2)$$

where  $\theta = (\log A_1, \log A_2, \log A_3, \log(\alpha_1 - \alpha_2), \log(\alpha_2 - \alpha_3), \log \alpha_3)$ . This parameterization ensures identifiability via the constraint  $\alpha_1 > \alpha_2 > \alpha_3 > 0$ , and positive predicted concentrations via  $A_1, A_2, A_3 > 0$ . Equation (2) implies dose proportionality (see Section 3), an assumption that was validated in previous studies in which different doses were administered to the same volunteers in a randomized order.

For the infusion route of administration which was used for Study 5 the disposition model described by (2) was convolved with a constant rate of input to produce predicted concentrations (Gibaldi and Perrier, 1982). No additional parameters are required in this case.

A first-order absorption model was assumed for the subcutaneous data. The concentration after a single dose is therefore given by:

$$f_1(D, \theta, t) = F_a D k_a \times \left[ A_1 \frac{\exp(-\alpha_1 t)}{(k_a - \alpha_1)} + A_2 \frac{\exp(-\alpha_2 t)}{(k_a - \alpha_2)} + A_3 \frac{\exp(-\alpha_3 t)}{(k_a - \alpha_3)} - B \exp(-k_a t) \right] \quad (3)$$

where  $0 < F_a < 1$  is the fraction of the dose absorbed,  $k_a > 0$  is the absorption rate constant and

$$B = \frac{A_1}{(k_a - \alpha_1)} + \frac{A_2}{(k_a - \alpha_2)} + \frac{A_3}{(k_a - \alpha_3)}$$

For this model we have  $\theta = (\text{logit} F_a, \log k_a, \log A_1, \log A_2, \log A_3, \log(\alpha_1 - \alpha_2), \log(\alpha_2 - \alpha_3), \log \alpha_3)$ .

Note that  $F_a$  can only be estimated when both intravenous and subcutaneous doses are given to the same individual, which is true for Study 2. Note also that the data from a subcutaneous dose alone would have been insufficient to estimate all three distribution phases due to the confounding with the absorption phase. From Figure 5 it is possible from the intravenous data to identify three straight lines corresponding to the three phases of distribution; for the subcutaneous data this is not possible, however. With the additional intravenous data all three phases may be estimated, however.

For the repeated dosing study (Study 4) the principle of superposition was used to sum the contributions of single doses of the form (3), so the predicted concentrations were of the form

$$\sum_{s=0}^m f_1(D, \theta, t + s\tau)$$

where  $m + 1$  doses are administered prior to time  $t$  and  $\tau$  is the dosing interval.

Let  $y_{ijk}$  denote the  $j$ th concentration on the  $i$ th individual with  $k = 1$  representing the intravenous experiments and  $k = 2$  the subcutaneous experiment. The error models were taken to be of the form

$$\log y_{ijk} = \log f_1(D_i, \theta_i, t_{ijk}) + \epsilon_{ijk}^y$$

where  $\epsilon_{ijk}^y$  are independent and identically distributed as  $N(0, \sigma_{yk}^2)$ . We use different error variances because although the same assay technique is used (see Section 7.1), we would expect greater model misspecification for the subcutaneous experiment as the assumption of first order absorption is only approximately true.

#### First stage dynamic models

To aid in the identification of the effect/concentration relationship the observed APTT measurements were plotted against drug concentrations. The consecutive

observations were joined to identify whether *hysteresis loops* were evident (Holford and Sheiner, 1981). Hysteresis essentially relates to the constancy over time of the effect/concentration relationship. The absence of hysteresis indicates that regardless of the time the effect corresponding to a given concentration is constant. Such loops may occur for a number of reasons, for example when the effect lags behind the concentration which could occur when the site of action is not the bloodstream. In this case an effect compartment model (Holford and Sheiner, 1981) may be assumed. No such loops were evident and an instantaneous relationship between drug concentration and dynamic effect was therefore assumed (see Racine-Poon and Wakefield, 1996, Figure 2). In Wakefield and Racine-Poon (1995) a linear relationship was assumed between APTT and concentration. This was later refined and the reciprocal of APTT was modeled via an inhibition sigmoid Emax model (Holford and Sheiner, 1981). This model is given by:

$$g_1(\theta, \phi, t) = APPT_0 \left[ \frac{f_1(d, \theta, t)^{1/2} + IC_{50}^{1/2}}{IC_{50}^{1/2}} \right] \quad (4)$$

where  $APPT_0 > 0$  is the baseline APTT (that is, the APTT when no drug is present) and  $IC_{50} > 0$  is a parameter that corresponds to the concentration which would be required to produce 50% inhibition. We take  $\phi = (\log APPT_0, \log IC_{50})$ . Initially the Hill coefficient (the power within equation 4) was estimated for the 16 individuals of Study 2 using maximum likelihood. Its value was found to be close to 0.5 and subsequently its value was fixed at this value. The fit of this model to the data of Studies 3–5 confirmed that this was reasonable.

Let  $z_{ij}$  denote the  $j$ th APTT measurement on the  $i$ th individual. The error model was taken to be of the form

$$\log z_{ij} = \log g_1(\theta_i, \phi_i, t_{ijk}) + \epsilon_{ij}^z$$

where  $\epsilon_{ijk}^z$  are independent and identically distributed as  $N(0, \sigma_z^2)$ .

#### *Second stage kinetic model*

The initial analyses did not include covariates. Once the analysis of Study 3 had been completed, estimates of the random effects of the PK parameters were plotted with plotting symbols representing the young/old covariate (the only covariate for these phase I studies). No differences between the two groups were detected. A t distribution was taken as the population distribution.

#### *Second stage dynamic model*

Again a no-covariate analysis was carried out. For Study 3 when the random effects were plotted it was found that the  $\log APPT_0$  and  $\log IC_{50}$  parameters were both much lower for the elderly volunteers. Consequently we allowed different locations for the population distributions of the young and old. A normal

distribution was taken as the population distribution.

### Third stage priors

For both the PK and the PD parameters relatively non-informative priors were assumed for all of the population parameters.

## Phase II

### First stage kinetic model

The sparsity of the data here lead to the fitting of a simple one compartment model with first-order absorption and elimination (Section 4.1); we discuss the implications of this simplification later in this section. Under the principle of superposition the predicted concentration is the sum of the contributions from the previous two doses which are of size  $D$ . The first stage form is therefore given by

$$f_1(D, \theta, t) = \frac{Dk_a}{V(k_a - Cl/V)} \sum_{l=1}^2 \left( \frac{\exp(-Cl/V(t - t_l))}{1 - \exp(-Cl\Delta_l/V)} \right) \quad (5)$$

$$- \frac{\exp(-k_a(t - t_l))}{1 - \exp(-k_a\Delta_l)} \quad (6)$$

where  $t_l$  is the time since dose  $l$  was given,  $l = 1, 2$ ,  $\Delta_l$  is the dosing interval for the two doses and  $\theta = (\log Cl, \log V, \log k_a)$ . The planned dosing intervals were  $\Delta_1 = 10$  hours and  $\Delta_2 = 14$  hours though the actual times varied for different patients. As with the Phase I data lognormal errors were assumed.

One of the primary objectives here is to obtain accurate estimates of clearance and volume for the patient population. One aim then is to investigate the relationship between these parameters and the covariates. Clearance and volume are also important because they may be used to determine maintenance and loading doses for a patient. When multiple doses are given the loading dose is given by

$$\text{Loading Dose} = \frac{\bar{C} \times V}{F} \quad (7)$$

where  $\bar{C}$  represents the desired average concentration,  $V$  the volume and  $F$  the bioavailability. The dosing rate as a function of clearance is given by  $\text{DR}(Cl) = \bar{C} \times Cl/F$  where  $Cl$  is the clearance. Then, for a known dosing interval  $\Delta$ , we have

$$\text{Maintenance Dose} = \text{DR}(Cl) \times \Delta. \quad (8)$$

Given the importance of clearance and volume we wished to examine the effect on the estimation of these quantities of using a simplified model (recall for the phase I data a three compartment model was used). To this end ten simulations of the following kind were carried out. Using the three compartment model and population parameter estimates from Studies 1–5 we simulated data of the same design as Study 6 (301 individuals with two time points each). The ‘true’ clearance

and volume are then functions of the seven parameters of the three compartment model. These were compared with the estimates obtained from the simulated data when analysed using a one compartment model. It was found that both the clearance and volume were well-estimated but the peak and trough concentrations were under- and over-estimated, respectively. This is reflected in Figure 8 which shows a plot of the residuals versus fitted concentrations for the Study 6 data. We see the effect of assuming a simple one compartment model – low (trough) concentrations are over-estimated whilst high (peak) concentrations are under-estimated. Figure 9 shows the (unstandardised) residuals from the PK analysis. These and all other residuals were calculated by simply substituting the posterior means of the parameters into  $f_1(D, \theta, t)$ , strictly we should evaluate the posterior distribution of the residual itself.

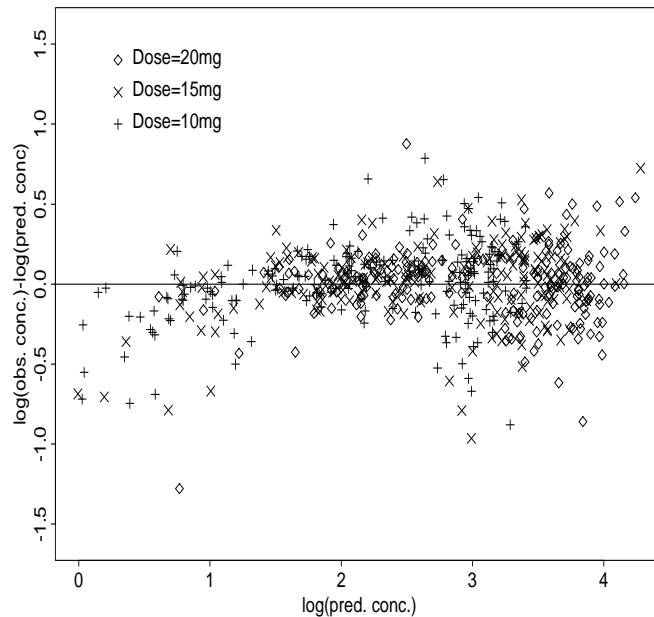


FIGURE 8. Residuals versus predicted concentrations for the 301 individuals of study 6.

#### *First stage dynamic model*

The model (4) was used to model the APTT though now we substitute *fitted* concentrations into the relationship, i.e. we do not account for the uncertainty in these fitted values. We do this because the concentration data are very sparse and we believe that the PD parameter estimation could be biased due to the feedback from the PK data. If we had complete faith in our PK/PD link model this would

not be a problem but here the model is empirically-derived. Again lognormal errors were assumed.

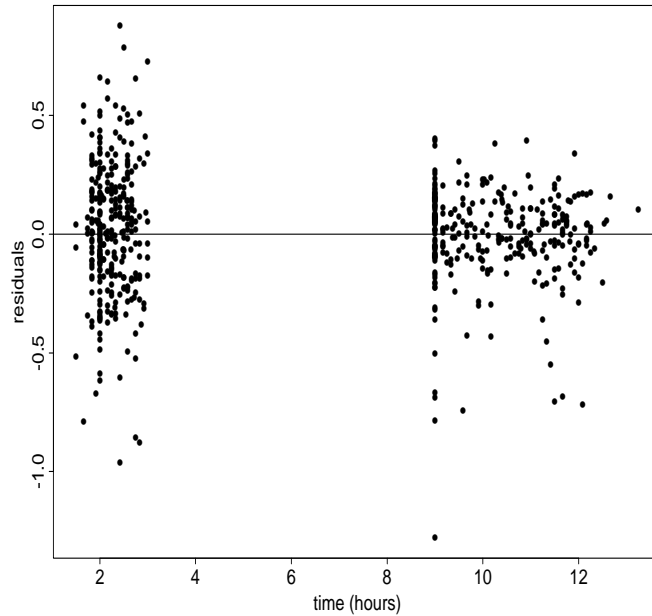


FIGURE 9. Residuals versus time for the population PK analysis of Study 6.

#### *Second stage kinetic model*

Recall from Section 6 that we have

$$\theta_i = f_2(\mu_\theta, X_i) + \delta_i^\theta.$$

where here  $\theta_i = (\log k_{ai}, \log Cl, \log V_i)$  and we have suppressed the subscript  $j$  because none of the variables were time-varying. As noted in Section 6 it is very difficult to carry out covariate modeling, in particular because the dependent variable  $\theta_i$  is multivariate and unobserved. For these data we have two random effects and six covariates and hence  $(2^6)^2 = 4096$  potential models, even before we consider transforms of the covariates such as body surface area, which is a function of height and weight and is sometimes used as a predictor of volume.

To form the covariate model for  $\theta_i$  we use biological information as far as possible. Here we describe an initial model, in the next Section further relationships will be investigated using graphical techniques. We know that REVASC<sup>TM</sup> is eliminated primarily through the kidneys via filtration (Section 7.1). In fact we know from studies in healthy volunteers in which urine is collected that the drug

is 98% renally eliminated. Hence if we could obtain the filtration rate of the kidneys, as determined by the creatinine concentration in urine this would give us a highly informative covariate for the clearance. Unfortunately this covariate is logistically difficult to measure so instead the serum creatinine concentration is measured from plasma samples. From this the creatinine clearance can be estimated via the empirically-derived formula (Rowland and Tozer, 1995):

$$\text{Creatinine Clearance} = \begin{cases} \frac{(140 - \text{age}) \times \text{weight}}{72 \times \text{serum creatinine}} & \text{for males} \\ \frac{(140 - \text{age}) \times \text{weight}}{85 \times \text{serum creatinine}} & \text{for females.} \end{cases}$$

An estimate of the creatinine clearance can therefore be derived for a patient from their measured age, weight and serum creatinine. We let  $X_{1i}$  denote the creatinine clearance of patient  $i$  centred on 80ml/min which is approximately the average in the population as a whole. We then model the clearance via:

$$\log Cl_i = \mu_{\theta 0} + \mu_{\theta 1} X_{1i} + \delta_{1i}^{\theta} \quad (9)$$

where  $\delta_{1i}^{\theta}$  is the  $i$ -th patients 'adjustment' to the log clearance which is predicted by the model. From the earlier discussion we would expect the clearance to increase with the creatinine clearance which is why we explicitly consider this parameter. We do not have an exact form to model this relationship and choose the loglinear form in (9) for mathematical convenience and numerical stability. Note that  $\mu_{\theta 0}$  represents the log  $Cl$  of a patient with an average creatinine clearance. Alternative models to (9) are given by:

$$Cl_i = \mu_{\theta 0} + \mu_{\theta 1} X_{1i} + \delta_{1i}^{\theta}$$

and

$$Cl_i = \mu_{\theta 0} \times X_{1i}^{\mu_{\theta 1}} + \delta_{1i}^{\theta},$$

both of which have been used in the PK literature. Within the observed limit of concentration/creatinine clearance one cannot assess which is the better model from a purely statistical perspective. This choice will be essential, however, if one wishes to extrapolate beyond the range of the observed data. Hence if such extrapolation is required the choice of model must be based on pharmacological/biological information. In the patient population it is not the case that 98% is renally eliminated. More thrombin is produced, hirudin binds to thrombin, and is then eliminated through the liver (hepatic elimination). Hence we have a significant non-renal route of elimination. We will return to this in the analysis stage when we address sensitivity.

We turn now to the volume parameter. On simple physiological grounds we expect the volume to increase with weight or body mass. We denote the weight of the  $i$ -th patient, centred by 80kg, by  $X_{2i}$  and again fit the simple loglinear model

$$\log V_i = \mu_{\theta 2} + \mu_{\theta 3} X_{2i} + \delta_{2i}^{\theta} \quad (10)$$

where  $\delta_{2i}^{\theta}$  is the  $i$ -th patients ‘adjustment’ to the log volume which is predicted by the model.

For the third parameter  $k_a$  there is very little information in the absorption phase since there were no early sampling times. An initial model

$$\log k_{ai} = \mu_{\theta 4} + \delta_{3i}^{\theta}$$

was assumed with very tight priors being placed on  $\mu_{\theta 4}$  and the variance of the  $\delta_{3i}^{\theta}$ . The variance on the prior for  $\mu_{\theta 4}$  corresponded to a change in  $k_a$  of  $\pm 5\%$ . The posterior for  $\mu_{\theta 4}$  from this analysis was located at an unreasonably high value, however, because even with a strong prior the data were too sparse to discount the intravenous model which corresponds to an infinite  $k_a$ . The same behaviour occurred when  $k_a$  was treated as a fixed effect, again with a tight prior. Two final analyses were carried out. In the first of these  $k_a$  was allowed to take the single value  $\exp(-1.2)$ . This latter was chosen from Studies 2–4 where the posterior mean of the population  $\log k_a$  was  $-1.2$ . In the second analysis  $k_a$  was allowed to take one of five discrete values, centred on  $\exp(-1.2)$  and with a spread of  $\pm 5\%$ .

We assume that the pair corresponding to volume and clearance  $\delta_i^{\theta} = (\delta_{1i}^{\theta}, \delta_{2i}^{\theta})$  arise from the zero mean bivariate normal distribution with variance-covariance matrix  $\Sigma^{\theta}$ .

#### Second stage dynamic model

We initially assume no covariate relationships, ie

$$\begin{aligned} \log APTT_{0i} &= \mu_{\phi 0} + \delta_{1i}^{\phi} \\ \log IC_{50i} &= \mu_{\phi 1} + \delta_{2i}^{\phi} \end{aligned} \quad (11)$$

with  $\delta_i^{\phi} = (\delta_{1i}, \delta_{2i}) \sim N(0, \Sigma^{\phi})$ . Although age was used as a covariate in earlier studies, for those studies we did not have measures of creatinine clearance (which are functions of age). Consequently the need for age in the covariate model for the PD parameters could have been due to the fact that we did not account for creatinine clearance in the PK second stage model.

#### Third stage kinetic model

Here we specify prior distributions for  $\mu_{\theta} = (\mu_{\theta 0}, \mu_{\theta 1}, \mu_{\theta 2}, \mu_{\theta 3})$ ,  $\Sigma^{\theta}$  and  $\sigma_y^2$ .

Studies 1–5 were all in healthy volunteers and a three compartment model was used. Extrapolation from healthy volunteers to patients is a risky enterprise. REVASC<sup>TM</sup> binds to thrombin and after an operation more thrombin is produced by the body. This leads to less free drug being present in the plasma and so increases both the apparent volume of distribution and the clearance.

If we have a compartment model with more than one compartment then there are two volume parameters, the volume of the central compartment and the volume at steady-state. At steady-state all of the compartments are in equilibrium and the volume relates total drug in the body to the concentration in the central compartment (which is the same as the concentration in all other compartments since we are at equilibrium). With a three compartment model there is a specific formula to evaluate the volume at steady state, Gibaldi and Perrier, (1982), p. 215. The clearance can also be readily calculated as the dose divided by the area under the concentration/time curve.

On the basis of the studies using the three compartment model we can obtain prior distributions for  $\mu_{\theta 0}$  (clearance intercept) and  $\mu_{\theta 2}$  (volume intercept) based on predictions from the Phase I studies. There were also three additional studies which we have not described in which weight and creatinine clearance were measured. Hence prior estimates of the regressors describing these relationships in (9) and (10) were also obtained. This was done by first simulating a large number of individuals from the posterior distribution of the population parameters. For each of these individuals we calculate the clearance and the volume at steady-state and the mean and the variance/covariance matrices of these quantities was then evaluated to give  $c_{\theta} = (2.41, 0.0101, 3.56, 0.095)$ . The variance of the prior  $C_{\theta}$  was taken to be a diagonal matrix with diagonal elements 0.09. The prior estimate for  $\mu_{\theta 0}$  corresponds to a clearance value of 11litres/hour for the patient population. The prior estimate of  $\Sigma_{\theta}$ ,  $R_{\theta}$  was taken to be a diagonal matrix with diagonal elements 0.04. These values correspond to a coefficient of variation on clearance and volume of 20% which is typical for studies such as these. We take  $r_{\theta} = 2$ , which is the smallest value that gives a proper prior, and choose  $A_{\theta} = a_{\theta} = 0$ , with 602 observations there are sufficient information in the data to estimate  $\sigma_y^2$ .

#### *Third stage dynamic model*

As a prior mean for  $\mu_{\delta}$  we use studies 1–5 to obtain the value  $c_{\delta} = (3.5, 5.0)$  with  $C_{\delta}$  diagonal with variances 1.0. We take  $R_{\delta}$  as diagonal with diagonal elements 0.04 which again corresponds to 20% coefficient of variation for  $APTT_0$  and  $IC_{50}$ .

### *7.4 Summary of analyses of studies 1–5*

Studies 1–5 provided various information which was subsequently used in the analysis of Study 6. These studies were therefore vital to the analysis of Study 6 because the data in that study were very sparse. In particular it was found that the bioavailability was equal to one and there was dose proportionality so the principle of superposition could be assumed. The instantaneous PK/PD relationship was also found to be appropriate for the three compartment model. As described above the third stage priors for Study 6 were also informed by the earlier studies.

### 7.5 Analysis of Study 6

The ratio of APTT post-administration to baseline APTT may be taken as a surrogate for the possibility of a bleed. In particular it was desirable to keep this ratio to less than 1.5 and in the observed data of this study this was always the case.

We first examine the adequacy of the model via various diagnostics. Figure 10 displays the first stage residuals,  $\log z_{ij} - \log g_1(\theta, \phi, t)$ , plotted versus time. We see that residuals at the first time are slightly more positive whilst at the final time point they are more negative. One possible reason for this is the bias in the predicted concentrations. To investigate this we carried out an analysis in which the observed concentrations were used to regress the response on but this made little difference to the residual plots or the posterior distributions of the population parameters. Another possibility to explain this lack of fit would be to allow the Hill coefficient to be an unknown parameter rather than the value of 0.5 which was that used for the healthy volunteer Studies 2–5.

Figure 11 plots the first stage PD residuals versus the individual covariates – patterns in this plot might indicate that we have missed a second stage covariate relationship. Here there appear to be no patterns.

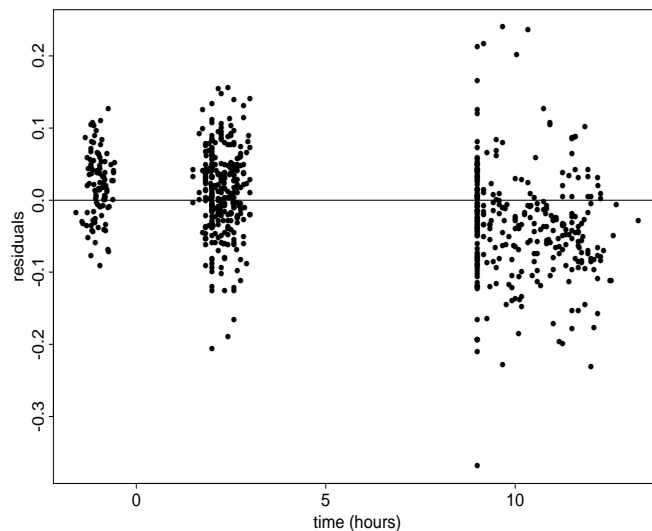


FIGURE 10. First stage PD residuals (unstandardized) plotted versus time.

Figures 12 and 13 display the posterior means of the PK random effects  $\delta_{i1}^\theta$ ,  $\delta_{i2}^\theta$ ,  $i = 1, \dots, 301$  versus the individual covariates. There appears to be an association between age and both the clearance and volume random effects. There is negative correlation (posterior median -0.71, 90% highest posterior interval -0.44, -0.85) in the population distribution of clearance and volume, however, so inclusion of age for one of these may lead to a disappearance of the apparent as-

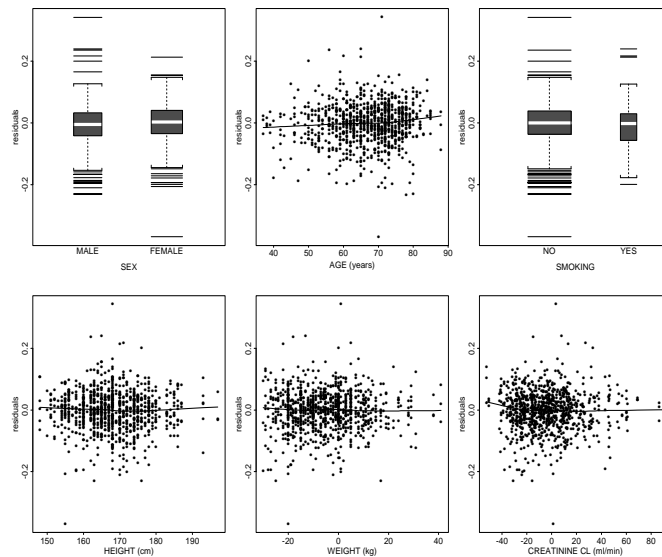


FIGURE 11. First stage PD residuals (unstandardized) plotted versus individual specific covariates.

sociation in the other. Figure 14 shows a scatterplot of the posterior means of the clearance and volume random effects. From this alone the high negative correlation is not apparent but this plot ignores the uncertainty in each of the random effects and so can be deceptive. After the operation patients received blood transfusions which results in drug concentrations falling which could lead to an apparent increase in volume and it is likely that more blood transfusions are carried out for the elderly. We consulted pharmacokineticists with regard to this point and they indicated that this effect would not be large, however. We carried out an analysis with age as a covariate in a loglinear model for volume but found that the posterior distribution was centred close to zero and so did not include this relationship in our model.

Figures 15 and 16 show the covariate plots for the PD random effects. There is some suggestion of an association between the  $APTT$  random effects and height but otherwise no obvious patterns.

Figure 17 shows normal plots of the individual PD random effects for  $APTT_0$  (left) and  $IC_{50}$  (right) Figure 18 shows a bivariate plot of the posterior means of these same quantities. The normality assumption appears viable although there is some evidence of skewness in the second stage distribution of the  $APTT_0$  random effects. These plots should be interpreted with caution, however, as estimates and not observations are being plotted (Lange and Ryan, 1989).

Figure 19 shows the posterior distributions of the PK population mean parameters. Each of these distributions are well-behaved apart from  $\mu_{\theta_3}$  which is highly skewed. We found that the Markov chain was slow mixing and a large number

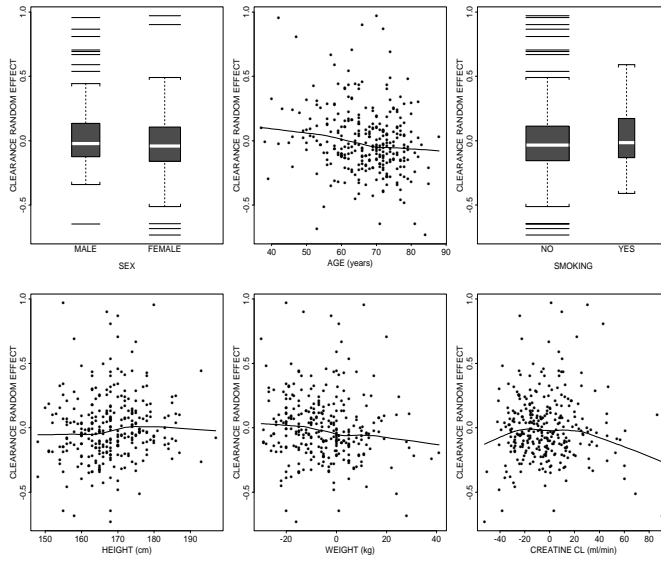


FIGURE 12. Second stage residuals (random effects) for clearance plotted versus individual specific covariates.

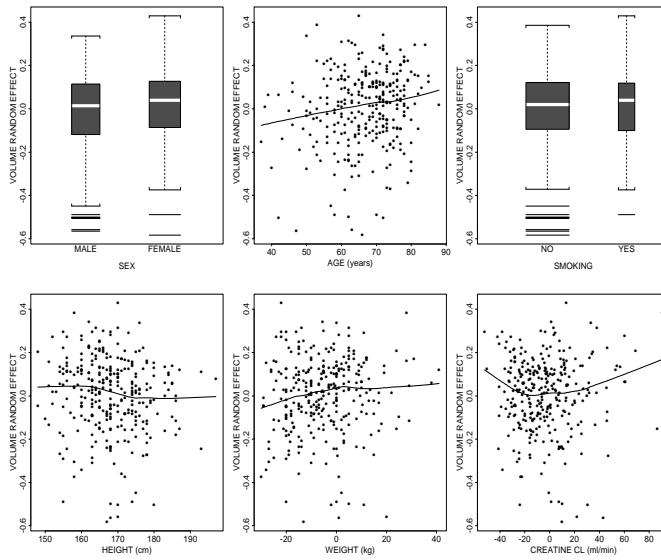


FIGURE 13. Second stage residuals (random effects) for volume plotted versus individual specific covariates.

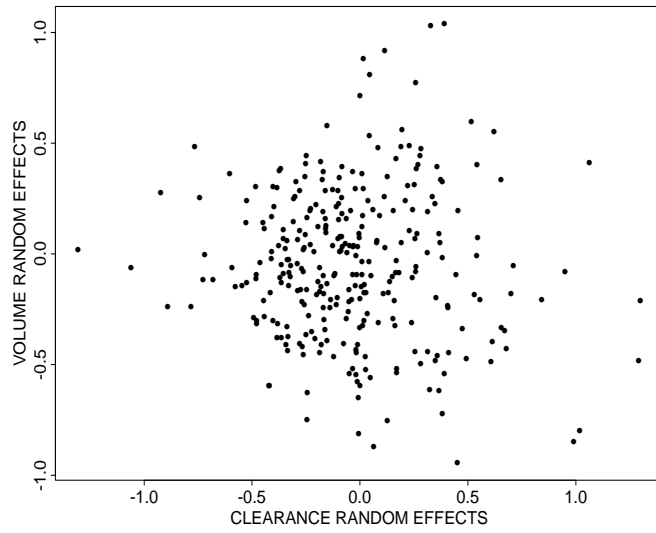


FIGURE 14. Posterior means of clearance random effects plotted versus posterior means of volume random effects.

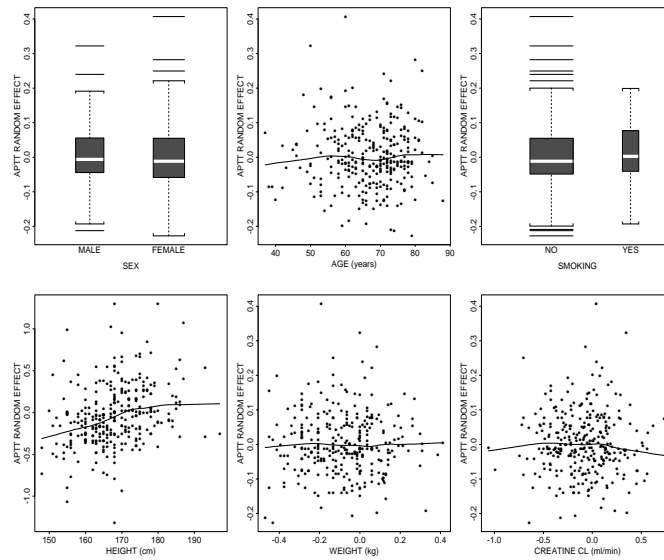


FIGURE 15. Second stage residuals (random effects) for  $APTT_0$  plotted versus individual specific covariates.

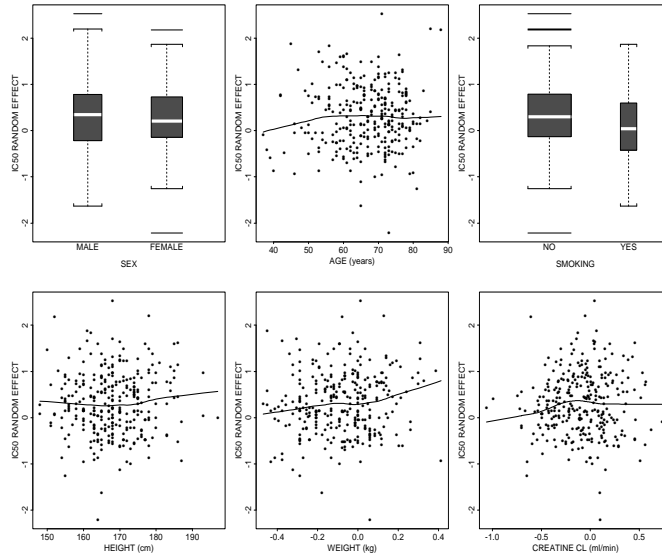


FIGURE 16. Second stage residuals (random effects) for  $IC_{50}$  plotted versus individual specific covariates.

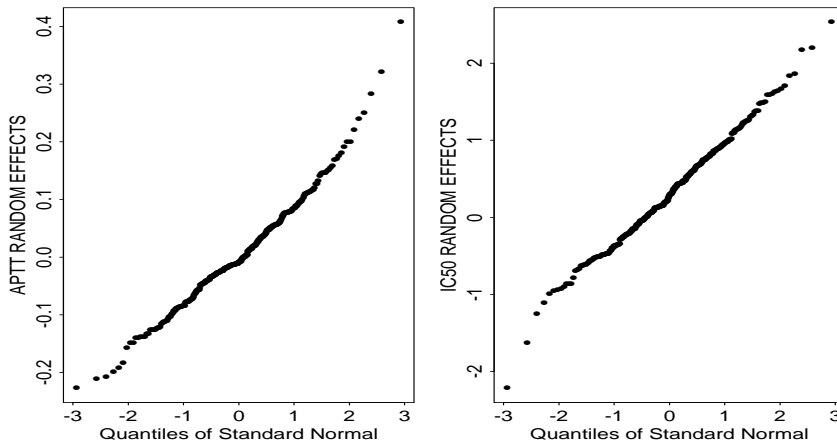


FIGURE 17. Normal scores plots for  $APTT_0$  random effects (left) and  $IC_{50}$  random effects (right).

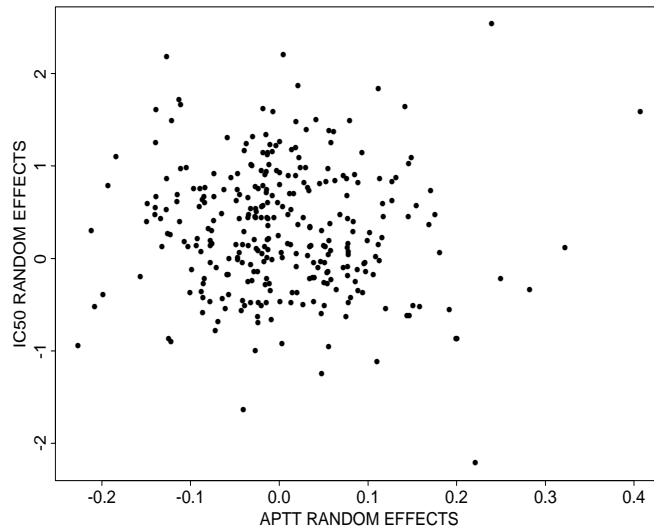


FIGURE 18. Posterior means of  $APTT_0$  random effects plotted versus posterior means of  $IC_{50}$  random effects.

of iterations were required to obtain reliable inference. The posterior means of these four parameters can be used within a simple plug-in approach for formulas (7) and (8) in order to determine loading and maintenance doses for new patients with measured weight and creatinine clearance.

Various sensitivity analyses were carried out. The above results were all for a fixed value of  $k_a = \exp(-1.2)$ . When we allowed this parameter to take one of five possible values between  $\exp(-1.1)$  and  $\exp(-1.3)$  we found that the volume increased with increasing  $k_a$  and the clearance decreased. The substantive conclusions were unchanged in this range, however.

The prior that was used for  $\mu_{\theta_0}$  (clearance intercept) reflected the patient population. A prior which was more consistent with the healthy volunteer studies (lower clearance) was also used but the posterior distribution was again unchanged.

### Postscript

Following Study 6, the dose of 15mg b.i.d. was chosen for the Phase III pivotal studies because the efficacy (i.e. the rate of thromboembolic events) of the 15 mg dose was similar to that of the 20mg dose and superior to the 10 mg dose, while the observed amount of bleeding was similar in all groups and comparable to that of the unfractionated heparin group. A second confirmatory trial carried out in 400 patients in Scandinavia (Eriksson *et al*, 1997a) confirmed the efficacy and the safety of the 15mg dose regimen. Finally, another large-scale study performed in 2000 patients (, 1997b) successfully demonstrated the efficacy and the safety of the REVASC<sup>TM</sup> 15mg b.i.d. compared to a low-molecular-weight heparin.

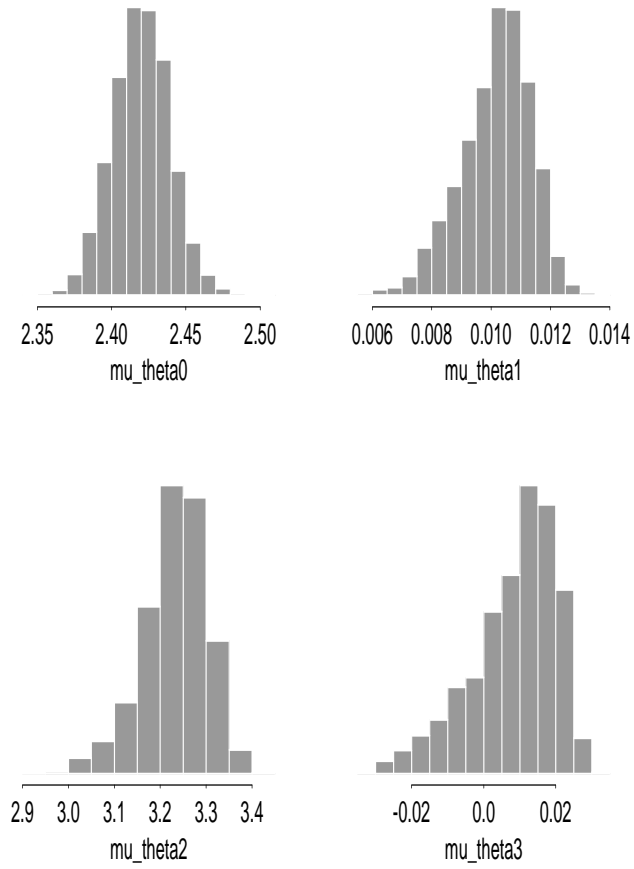


FIGURE 19. Posterior distribution for PK population mean parameters,  $\mu_{\theta_0}$  is the clearance intercept,  $\mu_{\theta_1}$  is the regressor for clearance/creatinine clearance relationship,  $\mu_{\theta_2}$  is the volume intercept and  $\mu_{\theta_3}$  is the regressor for volume/weight.

REVASC<sup>TM</sup> is now approved in the European Community.

## 8 Future directions

There are a number of outstanding methodological problems associated with population PK/PD analyses.

There is an extensive literature on general optimal experimental design but because of complexity and logistical considerations little attention has been paid to the design of population PK/PD studies in phase III clinical trials (Aarons *et al.*, 1996). In general, the choice of the number and nature of the subjects in a phase III clinical trial is made in relation to the primary goal of the study, which is usually concerned with the demonstration of efficacy and assessment of safety. In fact since a population PK/PD analysis is only one of the objectives of a phase III clinical trial, it should not compromise the other (major) objectives. Computer simulation and optimal design measures (regression) have been used to plan the timing of measurements and the idea of a sampling window (that is, a range of times rather than a particular time) has been widely used, as it helps to structure the sampling process and ensure that an adequate description of the PK/PD profile is obtained.

Some prior knowledge of the PK and PD models and covariate relationships is necessary for the analysis of sparse phase III data. Sample timing and dosing history is fundamental to pharmacokinetic/pharmacodynamic analysis. Consequently, good data and sample handling practices are essential to the successful application of the population approach to phase III clinical studies.

In phase I there is little borrowing of strength because of the abundance of data per individual. Hence it would appear that there is some room for designing trials with fewer sampling times and fully exploiting the hierarchy in the analysis.

Another area which is likely to grow in importance is physiological models in which the whole body is described more realistically by a complex system of compartments.

There are a number of components of population PK/PD models for which it may be more appropriate to use errors-in-variables modeling. The sampling times, particularly for Phase II/III studies are often nominal times – the actual scheduled times are reported rather than the actual times (see the number of observations reported at 9 hours in Figure 9) so a Berkson model (Carroll, Ruppert and Stefanski, 1995) may be appropriate. Wang and Davidian (1996) have examined this issue. Many second stage covariates are measured with error or are the result of formulas such as that for creatinine clearance in Section 7.3. Finally when a reliable joint PK/PD model cannot be obtained it may be preferable to use the observed concentrations and acknowledge the uncertainty in these values using an errors-in-variables approach.

The multi-stage hierarchical models that are used for population data contain many layers of assumptions and there has been little work on assessing their ade-

quacy (see Hodges, 1998). Flexible models such as the nonparametric and semi-parametric techniques described in Section 6.2 are important. Experience using these techniques is required but their use, at least as exploratory tools, is likely to be valuable though any physiological information that is known should obviously be incorporated.

Covariate selection remains an important problem as the detection of associations between PK and PD parameters and individual characteristics has important implications for the determination of dosage regimens.

For population studies the important principle underlying each of design, prior specification, assessment of model adequacy and model selection is that each of these procedures must be informed by biological/pharmacological information and not by statistical methodology alone.

## References

- Aarons, L., Balant, L.P., Mentré, F., Morselli, P.L., Rowland, M., Steimer, J.-L. and Vozeh, S. (1996). Practical experience and issues in designing and performing population pharmacokinetic/pharmacodynamic studies, *Eur. J. Clin. Pharmacol.*, **49**, 251–254.
- Abernathy, D.R. and Azarnoff, D.L. (1990). Pharmacokinetic investigations in elderly patients. Clinical and ethical considerations. *Clin. Pharmacokinet.* **19**, 89–93.
- Beal, S.L. and Sheiner, L.B. (1982). Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, **8**, 195–222.
- Beal, S.L. and Sheiner, L.B. (1993). *NONMEM User's Guide*, University of California, San Francisco.
- Bennett, J.E., Racine-Poon, A. and Wakefield, J.C. (1996). MCMC for nonlinear hierarchical models. In *Markov Chain Monte Carlo Methods in Practice* (eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter), 339–357. London: Chapman and Hall.
- Berry, D.A. (1990). Basic principles in designing and analyzing clinical studies. In *Statistical Methodology in the Pharmaceutical Sciences* (ed. D.A. Berry), 1–55. Marcel-Dekker, Inc. New York and Basel.
- Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Non-linear Models*. Chapman and Hall, London.
- Colburn, W.A. (1989). Controversy IV: population pharmacokinetics, NONMEM and the pharmacokinetic screen; academic, industrial and regulatory perspectives. *J.Clin.Pharmacol.*, **29**, 1–6.
- Davidian, M. and Gallant, A. (1993). The non-linear mixed effects model with a smooth random effects density. *Biometrika*, **80**, 475–88.

- Davidian, M. and Giltinan, D.M. (1993). Some simple estimation methods for investigating intra-individual variability in nonlinear mixed effects models. *Biometrics*, **49**, 59–73.
- Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, London.
- Eriksson, B.I., Kalebo, P., Zachrisson, B., Ekman, S., Kerry, R. and Close, P. (1996). Prevention of deep vein thrombosis with recombinant hirudin CGP 39393 in hip prosthesis surgery. Evaluation of three dose levels of recombinant hirudin in comparison with unfractionated heparin. *Lancet*, **347**, 635–39.
- Eriksson, B.I., Ekman, S., Lindbratt, S., Baur, M., Torholm, C., Kalebo, P. and Close, P. (1997a). Prevention of deep vein thrombosis with recombinant hirudin – results of a double-blind multicenter trial comparing the efficacy of desirudin (Revasc) with that of unfractionated heparin in patients having a total hip replacement. *J. Bone J. Surg. (Am)*, **79A**, 326–33.
- Eriksson, B.I., Wille-Jorgensen, P., Kalebo, P., Mouret, P., Rosencher, N., Bosch, P., Baur, M., Ekman, S., Bach, D., Lindbratt, S. and Close, P. (1997b). Recombinant hirudin, desirudin, is more effective than a low-molecular-weight heparin, enoxaparin, as prophylaxis of major thromboembolic complications after primary total hip replacement. To appear in *New England Medical Journal*.
- Food and Drug Administration (1989). Guideline for the study of drugs likely to be used in the elderly. Washington, DC.
- Gelman, A., Bois, F.Y. and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, **91**, 1400–12.
- Gibaldi, M. and Perrier, D. (1982). *Drugs and the Pharmaceutical Sciences, Volume 15 Pharmacokinetics, Second Edition*. Marcel Dekker.
- Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.* **44**, 455–472.
- Gilks, W.R., Neal, R.M., Best, N.G. and Tan, K.K.C. (1997). Corrigendum to ‘Adaptive rejection metropolis sampling within Gibbs sampling’. *Applied Statistics*, **46**, 541–2.
- Godfrey, K.R. (1983). *Compartmental Models and their Applications*. London, Academic Press.
- Hastings, W. (1970). Monte Carlo sampling-based methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- Hodges, J.S. (1998). Some algebra geometry for hierarchical models, applied to diagnostics. To appear in *Journal of the Royal Statistical Society, Series B*, **60**, 497–536.

- Holford, N.H.G. and Sheiner, L.B. (1981). Understanding the dose-effect relationship: Clinical application of pharmacokinetic- pharmacodynamic models. *Clinical Pharmacokinetics*, **6**, 429–453.
- Lange, N. and Ryan, L. (1989). Assessing normality in random effects model. *Annals of Statistics*, **17**, 624–642.
- Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects model for repeated measures data. *Biometrics*, **46**, 673–87.
- Maitre, P., Buhner, M., Thomson, D., and Stanski, D. (1991). A three-step approach to combining Bayesian regression and NONMEM population analysis. *Journal of Pharmacokinetics and Biopharmaceutics*, **19**, 377–84.
- Mallet, A. (1986). A maximum likelihood estimation method for coefficient regression models. *Biometrika*, **73**, 645–656.
- Mentré, F. and Mallet, A. (1994). Handling covariates in population pharmacokinetics, *International Journal of Bio-Medical Computing*, **36**, 25–33.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chemical Physics*, **21**, 1087–91.
- Miller, A.J. (1990). *Subset Selection in Regression*, Chapman and Hall, London.
- Muller, P. and Rosner, G.L. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association*, **92**, 1279–92.
- Pinheiro, J. and Bates, D. (1995). Approximations to the loglikelihood function in the nonlinear mixed effects model. *Computational and Graphical Statistics*, **4**, 12–35.
- Pitsiu, M., Parker, E.M., Aarons, L. and Rowland, M. (1993). Population pharmacokinetics and pharmacodynamics of warfarin in healthy young adults, *Eur.J.Pharm.Sci.*, **1**, 151–157.
- Racine-Poon, A. (1985). A Bayesian approach to nonlinear random effects models. *Biometrics*, **41**, 1015–1024.
- Racine-Poon, A. and Wakefield, J.C. (1996). Bayesian analysis of population pharmacokinetic and instantaneous pharmacodynamic relationships. In *Bayesian biostatistics*, (ed. D. Berry and D. Stangl). Marcel-Dekker.
- Racine-Poon, A. and Wakefield, J.C. (1998). Statistical methods for population pharmacokinetic modelling. *Statistical Methods in Medical Research*, **7**, 63–84.
- Rowland, M. and Tozer, T.N. (1995). *Clinical Pharmacokinetics: Concepts and Applications, Third Edition*. Williams and Wilkins.
- Sheiner, L.B., Beal, S.L. and Dunne, A. (1997). Analysis of non-randomly censored ordered categorical longitudinal data from analgesic trials (with discussion). *Journal of the American Statistical Society*, **92**, 1235–55.

- Sheiner, L.B. and Benet, L.Z. (1985). Premarketing observational studies of population pharmacokinetics of new drugs. *Clin.Pharmacol.Ther.*, **38**, 481–487.
- Smith, A. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J Roy Statist Soc, Series B*, **55**, 3–23.
- Steimer, J., Vozeh, S., Racine-Poon, A., Holford, N., and O'Neill, R. (1994). The population approach: rationale, methods and applications in clinical pharmacology and drug development. In *Handbook of experimental pharmacology*, (eds. P. Welling and H. Balant). Springer Verlag.
- Temple, R. (1983). Discussion paper on the testing of drugs in the elderly. Washington, DC: Memorandum of the Food and Drug Administration of Department of Health and Human Service.
- Temple, R. (1985). Food and Drug Administration's guidelines for clinical testing of drugs in the elderly. *Drug Information Journal*, **19**, 483–486.
- Temple, R. (1989). Dose-response and registration of new drugs. In *Dose-response relationships in Clinical Pharmacology*. Eds. Lasagne, L., Emill, S. and Naranjo, C.A. Amsterdam, Elsevier, p145–167.
- Verstraete, M., Nurmohamed, M., Kienast, J. *et al* (1993). Biological effects of recombinant hirudin (GP 39393) in human volunteers. *J. Amer. Coll. Cardiol.*, **22**, 1080–1088.
- Vonesh, E.F. and Chinchilla, V.M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York, Dekker.
- Vozeh, S. and Steimer, J.J. (1985). Feedback control methods for drug dosage optimization: concepts, classifications and clinical applications. *Clinical Pharmacokinetics*, **10**, 457–476.
- Wakefield, J.C. (1994). An expected loss approach to the design of dosage regimens via sampling-based methods. *The Statistician*, **43**, 13–29.
- Wakefield, J.C. (1996a). The Bayesian analysis of population pharmacokinetic models. *J. Amer. Statist. Assoc.*, **91**, 62–75.
- Wakefield, J.C (1996b). Bayesian individualization via sampling based methods. *Journal of Pharmacokinetics and Biopharmaceutics*, **24**, 103–31.
- Wakefield, J.C. and Bennett, J.E. (1996). The Bayesian modeling of covariates for population pharmacokinetic models. *Journal of the American Statistical Association*, **91**, 917–927.
- Wakefield, J.C., Gelfand, A.E. and Smith, A.F.M. (1991). Efficient generation of random variates via the ratio-of-uniform method. *Statist. Comput.*, **1**, 129–133.
- Wakefield, J.C. and Racine-Poon, A. (1995). An application of Bayesian population pharmacokinetic/pharmacodynamic models to dose recommendation. *Statistics in Medicine*, **14**, 971–86.

- Wakefield, J.C., Smith, A.F.M., Racine-Poon, A. and Gelfand, A.E. (1994). Bayesian analysis of linear and non-linear population models using the Gibbs sampler. *Appl. Statist.*, **43**, 201–221.
- Wakefield, J.C. and Walker, S.G. (1997). Bayesian nonparametric population model: formulation and comparison with likelihood approaches. *Journal of Pharmacokinetics and Biopharmaceutics*, **25**, 235–53.
- Walker, S.G. and Wakefield, J.C. (1998). Population models with a nonparametric random coefficient distribution. To appear in *Sankhya, Series B*.
- Wang, N. and Davidian, M. (1996). A note on covariate measurement error in nonlinear mixed effects models. *Biometrika*, **83**, 801–812.
- Yuh, L., Beal, S., Davidian, M., Harrison, F., Hester, A., Kowalski, K., Vonesh, E., and Wolfinger, R. (1994). Population pharmacokinetic/pharmacodynamic methodology and applications: a bibliography. *Biometrics*, **50**, 566–675.

## Discussion

### Frédéric Y. Bois

*Lawrence Berkeley National Laboratory, Berkeley*

I would first like to congratulate the authors for a very clear and complete presentation of the state of the art in Bayesian population pharmacokinetics. I would warmly recommend its reading to statisticians interested in the subject matter and its idiosyncrasies, and to all pharmacologists curious to understand how real-life clinical data can be analyzed with pharmacokinetic models. Reviews of this quality are in my opinion essential tools for building interdisciplinary teams in academic, regulatory, or industrial environments. It may be no accident that it has itself been written by a multitalented group . . .

Incidentally, a major advantage of the Bayesian approach is that it is more natural to many clinicians. That may well be linked to the frequentation of patients: at the time of diagnosis we never wonder whether the subject is “statistically significantly” ill with pneumonia, but rather what are the chances of pneumonia versus competing possibilities. The development of numerical tools able to deal with the computational complexity of population pharmacokinetics is in this respect a major advance.

My only concerns are with the potential pitfalls of the modeling philosophy which consists in using simpler models as data per individual become sparser, while it is known that a complex model is in fact more scientifically reasonable. The authors have been careful in checking that no major bias (at least of clearance and volume of distribution) was introduced when going from a three-compartment to a one-compartment model. Yet, there may still remain more subtle problems with the proposed estimation. It seems also that hirudin is a nicely behaved drug for that matter. Things may not go that well with other compounds, and we should

be very cautious. The first problem with that expedient is potential estimation bias. Figure 1 shows what can happen when concentration kinetic data generated by an underlying two-compartment process are fitted with a one-compartment model (with first-order absorption). With only two data points, the solution is unique. The estimated peak concentration is 1.3 times the true peak. Other data points along the true two-compartment kinetic curve could have led to underestimation of the peak, or worse errors. I did not look for an exaggerated figure, and actual drugs can certainly exhibit worse behavior. Peak levels of drugs in the body tend to be associated with acute toxic effects and their correct estimation is a basic requirement of formulation quality control.

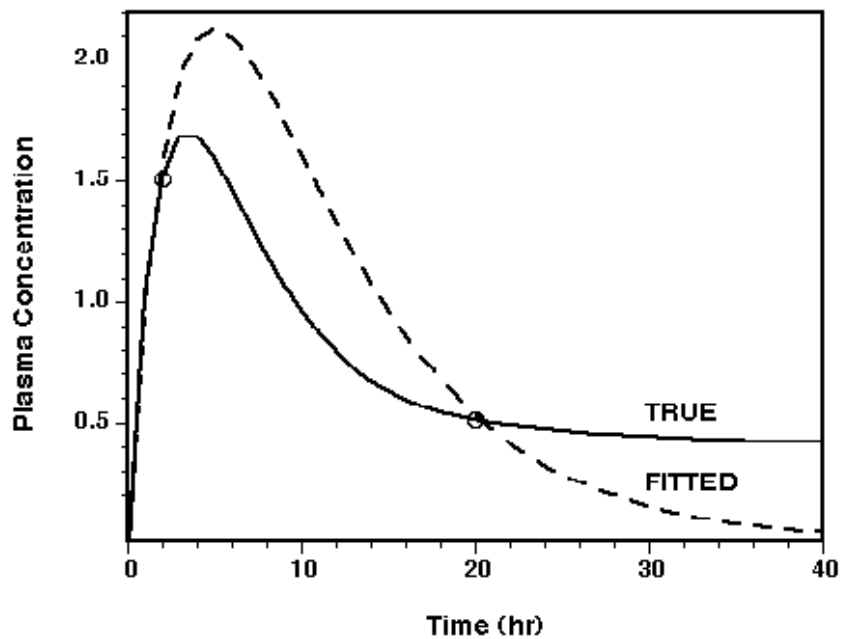


FIGURE 1: Potential estimation bias in the peak plasma concentration of a drug when a one-compartment model (dashed line) is fitted to data from an actual two-compartment kinetic process (solid line).

The second problem I see concerns scale rather than location. In some situations, population variability can be greatly overestimated if the wrong model is fitted. See Figure 2: for all four simulated subjects, exactly the same two-compartment kinetics actually apply for drug X; yet, because of different sampling schedules, very different peak concentrations are obtained when fitting a one-compartment model. The problem disappears if the same sampling schedule

is adopted for all subjects, but a particularity of most population studies (which can be construed as an advantage) is that samples can be obtained at different times in different subjects . . .

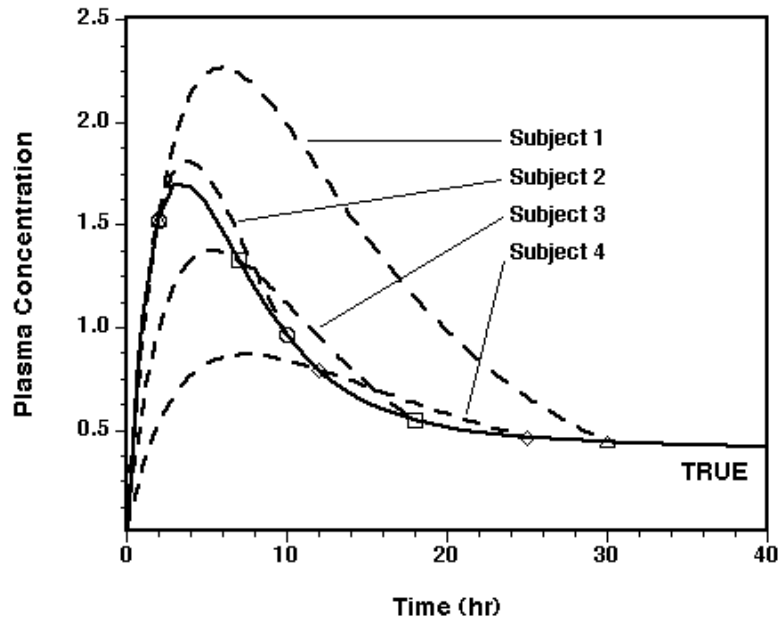


FIGURE 2: Estimation bias in the variance of peak plasma concentrations of a drug when a one-compartment model (dashed lines) is fitted to individual data arising from actually the same two-compartment kinetic process (solid line).

Further problems, not met by the authors, can arise when extrapolating the fitted model outside the data range. This is a standard pitfall, exemplified in Figure 3. The plasma concentration after 40 hours will be very badly predicted by the small, wrong, model. Indeed, one would never ever attempt such extrapolations, except maybe when looking at posterior predictions in a Bayesian context, e.g. for dosing adjustments.

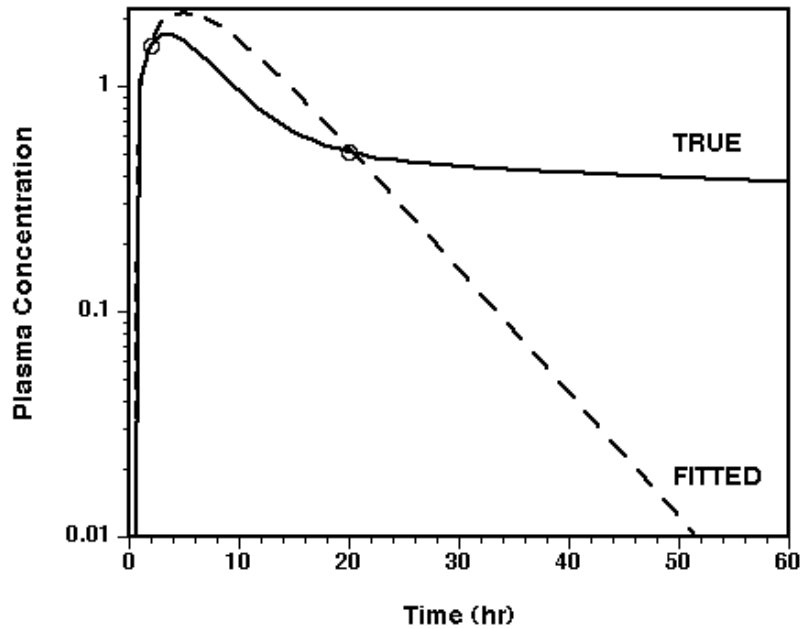


FIGURE 3: Error when extrapolating plasma concentration of a drug with a one-compartment model (dashed line), fitted to data arising from a two-compartment kinetic process (solid line).

More philosophically, when juggling models, I would be afraid of introducing some confusion in the booming field of hirudin pharmacokinetics. Are the kinetics mono-, bi-, or tri-compartmental? I am not sure myself. Maybe this is an arcane and hair-splitting issue of quality assurance in science. Yet, I can easily imagine a clinician stumbling, after a quick and dirty literature search, on a population study of hirudin and learning with delight that a one-compartment model has been “successfully” fitted to the data. S/he takes a pocket calculator and computes the plasma concentration remaining after 40 hours and . . . see Figure 3. We may maintain that models are application-specific, that they have no validity outside the particular purpose they were designed for, but that is a grim and pessimistic perspective for pharmacokinetics as a research activity. I would rather favor building trust in the models we use.

On another philosophical plane, I would argue that forgetting the past data and experience, when starting an analysis of hirudin population kinetics, is not an epitome of the Bayesian approach. I wish we could do better. Naively, assuming in particular that the US FDA did not exist, we might ask what the phase I pharmacokinetic studies are good for, if they don't inform much of the analy-

ses in phase II. The authors might advance two excuses for erring on the side of economy: they avoid identifiability problems, and they may not have in fact good prior information on their population of patients. These two issues are linked. If prior information about the parameters of a three-compartment model were strong, identifiability problems resulting from very sparse data would be much reduced (even if at the price of some reparameterizations). The problem rests mostly with the issue of extrapolating prior information. I would actually add that question as one of the challenges for future work in the area of population pharmacokinetics. The problem does not concern the extrapolation of the model structure; a three-compartment model, at least, is certainly adequate for hirudin kinetics in patients. The obstacle is the non-applicability of some of the parameter distributions derived from healthy individuals to patients. I would argue that part of the problem is due to the inability of the models used, even with three-compartments, to make full use of *a priori* physiological information. A more complete model could incorporate a better description of elimination processes, describe the fixation of hirudin to plasma proteins, and identify the various compartments. Physiological models can be quite refined in their description of drug kinetics. They are harder to compute but they allow the use of strong informative distributions and ease the extrapolations by modeling the source of the differences between subjects. Obviously, I have the easy part in just suggesting further work, but the authors themselves underline in their discussion the potentials of physiological modeling and do point to current research in the area. There are, from the practicing pharmacokineticist point of view, terrible difficulties in parameterizing and fitting a physiological model, "it just can't be done" as someone told me early in my Ph.D. work. I think that recent development in Bayesian statistics do give us a way to go forward, provided we learn to forget some of our classical habits.

## Discussion

### Marie Davidian

*North Carolina State University*

The authors are to be congratulated for a clear, well-written, and practically relevant exposition on a topic that has both generated much interest and engendered much confusion. Appreciation for the role of pharmacokinetics, pharmacodynamics, and their interplay in the drug development process among statisticians has been increasing. However, many of these statisticians familiar with PK/PD have acquired their knowledge through serendipitous experience or individual initiative. For the broader population of statisticians in industry and especially those working exclusively in Phase III clinical trials, my experience is that, although there is an appreciation that PK/PD is somehow important, knowledge of its role in the overall process is a bit vague. Compounding the problem is that there are few self-contained accounts that clarify the role of PK/PD in a way accessible

to the statistician. This article fulfills this need well; thus, regardless of religious preference (Bayesian vs. frequentist) these statisticians will find it to be a useful resource. I look forward to being able to hand it to my students interested in a realistic account of PK/PD analysis and associated statistical methods in drug development.

Praising the broad utility of the article is nice, but what of the statistical aspects? Unfortunately, those hoping for some anti-Bayesian polemics from one whose work in this area has been decidedly frequentist will be disappointed, as I believe that taking a Bayesian perspective in the current context is entirely sensible. Thus, the focus of my comments is not on statistical underpinnings but rather on the critical issues in population PK/PD analysis that I believe transcend the mere choice of statistical philosophy. I concentrate on this because I believe that, ultimately, one of the main contributions of the work of Drs Wakefield, Aarons, and Racine-Poon, henceforth WAR-P, is their fair and balanced focus on these issues.

As made abundantly clear by the authors, both statistical and subject-matter considerations in PK/PD analysis are characterized by a myriad of assumptions. On the subject-matter side, the choice of structural models is dictated by assumptions on the underlying physiological processes that are, mostly of necessity, gross simplifications. Depending on the sparsity of the data at hand, this choice may be subject to further simplifying assumptions in order to make fitting feasible. A prominent feature of the situation is that the processes of PK and PD occur at the *individual* level, while, for drug development, interest focuses mostly on the *population*. Thus, the natural statistical framework within which to represent data on PK/PD from several subjects is that of a hierarchical model. Tailoring this model to fit the needs of the particular situation involves a number of additional assumptions. At the very least, one must specify the nature of intra- and inter-subject variation. For studies involving identification of the sources of this variation, like the authors' Study 6, the analyst is faced with the additional challenge of incorporating subject-specific covariate information into the model; this involves assumptions about the functional form of the relationship between meaningful parameters like drug clearance and covariates, for which there is usually little scientific guidance. These parameters are of course unobservable, exacerbating the challenge.

Once the basic model framework has been established, the statistical paradigm within which implementation will be carried out is chosen. The most widely-used methods are "frequentist." The marginal likelihood is almost always analytically intractable due to the nonlinearity of the structural model, forcing numerical evaluation of the required integrals. The most popular inferential strategies seek to avoid this complication by appealing to a linearization that allows the likelihood to be approximated; variants of this approach are implemented in the package NONMEM (Beal and Sheiner, 1993), whose widespread availability and focus on PK/PD makes the methods favored by pharmacokineticists. Fitting proceeds treating the approximation as exact (an assumption), and inference on model parameters is carried out by invoking asymptotic approximations (a further assump-

tion). It is certainly possible to avoid approximations by carrying out the integration numerically (e.g. Davidian and Gallant, 1993); inference will still involve an appeal to large sample results. As demonstrated by WAR-P the Bayesian approach introduces a third “hyperprior” stage, which incorporates further distributional assumptions, possibly representing realistic prior information from previous studies. The necessary integrations may be carried out by appealing to MCMC techniques, thereby eliminating the need for linear approximation. However, this comes at the expense of possible sensitivity to the distributional assumptions that most often are fully specified at each model stage, including priors, thus, representing a source of concern to frequentists in the same way asymptotic approximations are to Bayesians. Luckily, my experience has been that, when the data contain sufficient information, both frequentist and Bayesian approaches yield similar conclusions, which is as it should be, as Dr Wakefield and I once encountered firsthand. While analyzing the same data set, we noted that our respective results exhibited an almost eerie correspondence given our different assumptions, techniques, and software, right down to comparison of my large-sample confidence intervals to the quantiles of his posteriors, that could not be attributed to a mutual hallucinatory experience induced by the previous night’s session at the local pub.

Before I stray into the details of such sessions, back to the point. To summarize the population PK/PD analysis endeavor, this type of modeling and inference in general is not an enterprise to be undertaken lightly. Before one even begins to contemplate the details of a statistical analysis, one must recognize that the basic model framework relies on numerous assumptions, some that may be justified by subject-matter considerations or on the basis of empirical evidence, others that may be a matter of convenience or the subjectivity of the analyst. That done, the inferential strategy chosen carries with it further, generally unverifiable assumptions. This choice may be dictated by the political leanings of the analyst; however, in practice, it is usually dominated by the availability of accessible software.

My main message is thus. For PK/PD problems, complex hierarchical models represent the appropriate framework for analysis, and I believe that the Bayesian perspective, allowing information from previous studies to be incorporated in a natural way throughout the development process, holds considerable promise. This modeling is of course not new; population PK/PD analysis via frequentist techniques has been carried out for almost two decades with success by pharmacokineticists, some of whom, although not trained formally in statistics, have an intuitive feel for how things work. Among these individuals, such as the pioneer of population PK/PD analysis, Lewis Sheiner, and his colleagues, the complexity of the modeling and limitations of what may be gleaned from real data are well-appreciated. However, regardless of whether frequentist methods, approximate or exact, or Bayesian methods are used, I am concerned that, as popularity of the approach increases, the complexity, the broad array of assumptions, both subject-matter and statistical, and the numerous pitfalls of implementation, may be under-recognized by the broader population of consumers, both statisticians and pharmacokineticists. The FDA has already issued a draft guidance on popu-

lation PK/PD analysis, encouraging more widespread undertaking of these analyses in the pharmaceutical industry. There is thus an urgent need to make these users aware of the limitations involved; this would be beneficial to all users, even those with practical experience. In their paper, WAR-P have done a commendable job of realistically warning the reader at each instance where an assumption has been made and of its possible consequences and highlighting the need for relevant sensitivity analyses. It is important that the considerable enthusiasm among statisticians and scientists be tempered by realism; this cautious stance stems not only from what I have encountered when fielding questions from individuals new to the area, but also from my own painful learning process.

To be a little more specific, I would like to point out briefly some pitfalls in greater detail. First, subject-matter issues aside, these models are *hard*. The non-linearity of the structural model is in itself a major complication, raising questions of parameterization, as noted by the authors. Incorporation of subject-level covariates in the second stage model leads to additional parameters to be fitted, and it is natural within such a complex framework to wonder whether lack of identifiability of some parameters may be an unintended consequence of this modeling, which may be impossible to determine by inspection. Although the classical Bayesian view would hold that identifiability is not a problem for a Bayesian analysis, in practical situations this is simply not true if one is interested in more than a simply hollow academic exercise. If one's goal is to get sensible, useful answers to real questions, it is a key issue in this application, regardless of paradigm. Lack of identifiability may rear its head through unstable implementation, or may lead to physically implausible solutions. *Near* lack of identifiability may be more insidious, and may lead to modeling assumptions that, although nonsensical from a biological perspective, may be required to achieve stability. For example, when the variation in one individual-specific PK/PD parameter is small relative to the others, this may dictate treating that parameter as if it is fixed in the population, which is certainly not realistic. Recognizing such situations is tricky, and determining how best to handle them while holding true to the science is trickier still. These issues and others like them may not be fully appreciated; whether this poses major consequences for the qualitative inferences drawn may or may not be serious in any given problem, but it is prudent to be concerned. Moreover, they may manifest themselves in or affect different methods in different ways, making validity of results method-dependent, a feature that will likely go unnoticed unless different methods are tried.

Speaking of implementation, both frequentist and Bayesian approaches are difficult. For the former, with which I am more familiar, complex, high-dimensional optimization or equation-solving is required, even if approximations are introduced. In my experience, one can never be too skeptical about the results; objective functions may be riddled with local maxima or minima, so finding the "true" solution may require appeal to numerous starting values. The validity of common approximations, particularly for the computation of standard errors and confidence intervals, may be poor. Even for exact likelihood methods, the relevance of asymptotic theory in finite samples is not guaranteed. How these factors

conspire with those mentioned above may be impossible to sort out or interpret. I have noted a tendency for fascination with some of the most complex procedures whose operating characteristics are the least understood, such as those involving less restrictive assumptions on the distribution of the random effects. I was once witness to a presentation of an analysis where a mixture of *three* normal distributions for the random effects (yes, the mixing proportions were estimated) was fitted to PK data on *seventeen* subjects; it is scary to contemplate how these various problems may have afflicted this endeavor...

A strength of the authors' account is their willingness to point out the corresponding issues in the Bayesian implementation. My comments here represent the impressions of an observer who has little experience with the nuts and bolts, so are of necessity somewhat superficial, but I believe the spirit is correct. The obvious comparison with computational issues for frequentist approaches is with implementation through the use of Markov chain Monte Carlo techniques. I say this somewhat tongue and cheek, but it is almost as though a religious movement has developed over the application of these methods. In this context, my fear is that, just as with frequentist approaches that allow flexible, nonparametric estimation of the random effects distribution, there is a widespread enthusiasm for MCMC among individuals who may not have studied them with same depth as the authors. This is no doubt in part a consequence of the facts that the premise seems (deceptively) simple and the implementation involves some pretty-fancy sounding stuff, a combination which tends both to fascinate and be unwittingly misunderstood. The possibility for confusion is indeed ripe; I am sure that Drs Wakefield and Racine-Poon the perhaps extreme case of one individual who was certain that the MCMC implementation was equivalent to the "EM algorithm."

It is clear that the performance of these methods in difficult problems such as population PK/PD are still not entirely understood (not that they are for frequentist methods, either). For example, as the authors remark with respect to their analysis of Study 6 in Section 7.5, "We found that the Markov chain was slow mixing and a large number of iterations were required to obtain reliable inference." This statement is a bit vague (how "large?" what constitutes "slow?" "reliable?"), but I do not fault the authors for that, as they were speaking to an audience with some familiarity with the methods. This statement does convey the message that, even among the experienced, the attitude is still cautious, and interpretation of the behavior observed not entirely clear. Another obvious frequentist target is sensitivity to prior specification; WAR-P are again to be commended for stressing the need for investigating investigation of this issue. A concern is again that, for less experienced users, these issues may not be given their full day in court, leading to the Bayesian version of analyses similar to the three-normal mixture affair above.

Although my tone comes off as pessimistic, in reality I have positive expectations for the future. There are other applications, such as analysis of HIV dynamic data from AIDS clinical trials, where hierarchical nonlinear methods are just beginning to be applied; thus, both in the population PK/PD arena and these other areas, there are new challenges, such as issues of measurement error in covariates, missing covariates, and informatively missing responses. The Bayesian

approach has a certain advantage over frequentist methods for these problems, as introduction of these complications into the analysis is more straightforward. As we continue to learn about performance in practice and as appreciation for what can and cannot be accomplished realistically with real data becomes more widespread, these analyses will hopefully one day be a standard component of the statistician's repertoire. Thankfully, researchers like WAR-P here to offer balanced guidance along the way.

### *References*

- Beal, S.L. and Sheiner, L.B. (1993). *NONMEM User's Guide*, University of California, San Francisco.
- Davidian, M. and Gallant, A.R. (1993). The nonlinear mixed effects model with a smooth random effects density, *Biometrika* **80**, 475-488.

## Discussion

### **S. Greenhouse**

*George Washington University*

Dr. Aarons gave us an excellent review of the long haul necessary to develop a new drug almost always taking a number of years. He refers to the large amount of information that becomes available at the time a phase three clinical trial is planned. Now the drug they were discussing, hirudin, is an anticoagulant and therefore must have some adverse effects of bleeding. A possible long term effect is stroke. Clearly it is difficult for the drug company to do long term studies but does anyone think about the issue of how to estimate potentially long term severe adverse effect? Somehow one is uneasy with the response that that is the reason we need to do phase three clinical trials.

## Rejoinder

We would first like to thank Professor's Bois and Davidian for their discussions. We consider their comments under a number of headings.

### *Simplified Model*

We essentially agree with the comments of Professor Bois concerning the use of a simplified model and find the examples he presents enlightening. Ultimately

the effects, and therefore the adequacy, of the simplified model will be scenario-specific and in particular will depend on the objectives of the study. In our case study, *clinically*, a one-compartment model is adequate for the safe and efficacious use of the drug. Clearance is not poorly estimated here and the clearance controls the dosing rate. From our simulations and Professor Bois' examples it is clear that the peak and trough levels are less-well predicted by the simple model. This was also found for the drug bismuth by Bennett, Wakefield and Lacey (1997) where, in addition, some systematic discrepancies were found for the clearance parameter.

We note that for extravascular administration, due to the finite time for absorption, the peak/trough swings are dampened down making the use of the simple model less of a clinical problem. Extrapolation is probably not as great a problem for chronic (i.e. multiple) dosing.

Phase I studies are tolerability studies and provide baseline information on the pharmacokinetics of a drug. They therefore guide the design of phase II studies by providing the basis of the PK model and parameter estimates. Phase II studies are generally carried out in a patient population which differs from the young, healthy population that is considered in phase I. Consequently the PK parameter values, but usually not the model, differ between phase I and phase II studies. Due to the sparsity of data in most phase II studies one must either assume a simplified model (as we did), or impose a strong prior distribution. The latter is difficult, however, because of the aforementioned differences between the phase I and II groups. Ideally, as suggested by Professor Bois, one would resort to the underlying physiology to predict how the parameters would change between the two groups. However the information necessary to make these predictions only becomes available during the phase II and III programs; the phase I population is too homogeneous to provide the necessary range in the covariates. Nevertheless it may be possible to utilize pre-clinical information and data on related drugs to facilitate this prediction in a Bayesian manner.

Finally it should be mentioned that there are many examples in clinical practise where a reduced model is used routinely and successfully to guide therapy. A good example is the antibiotic gentamicin for which a one-compartment model is used for dosing but strictly the model has at least two-compartments (Evans *et al*, 1980). We re-iterate that in practice it is always necessary to show that the simpler model is adequate for its purpose.

In general the combination of information from different studies/experiments is an outstanding problem. The perceived wisdom currently seems to be that unless the data are directly comparable, in which case one would combine the datasets anyway, then one should use the posterior distributions from previous studies as priors, but, in the sense of conservatism, with increased variances. Clearly further work is needed in this area.

In fact the reason that we did not use a joint PK/PD model was because we did not sufficiently believe in our first-stage PK model. Bennett and Wakefield (1998) consider in detail the data of Study 6 and compare the effect of: using the observed concentrations; using the fitted concentrations obtained from a one-compartment

model; jointly estimating the PK/PD relationship with a one-compartment model; and using the observed concentrations with the errors in these being acknowledged via an errors-in-variables model.

A closely-related problem is one of population meta-analysis. Initial work (Wakefield and Rahman, 1998) indicates that this is feasible but the exchangeability of studies is a very strong assumption which needs very careful consideration, particularly beyond phase I.

### *Identifiability*

We agree with Professor Davidian concerning the difficulties of parameterization in the face of model identifiability. Even in the case of the simple ‘flip-flop’ model (Gibaldi and Perrier, 1982) the usual approach is to ensure identifiability by assuming the parameterization  $(\log V, \log(k_a - k_e), \log Cl)$ , if it is believed that  $k_a > k_e$ , or parameterization  $(\log V, \log(k_e - k_a), \log Cl)$ , if it is believed that  $k_e > k_a$ . We note that already we face difficulties if we wish to then regress  $\log k_a$  on individual-specific covariates at the second stage. Real difficulties arise if we obtain data from a population within which  $k_a$  and  $k_e$  are of similar size. In this case a mixture distribution at the second stage may provide a solution.

### *Implementation*

Professor Davidian is right to point out that the implementation of the Bayesian approach is not without its difficulties. The BUGS software ( Spiegelhalter, Thomas, Best and Gilks, 1994) is currently being extended to handle population models within a menu-driven environment, but convergence issues remain, particularly if the model that is used is inadequate in some respect, or when one is faced with near-identifiability problems.

### *Safety*

There is no simple answer to Professor Greenhouse’s contribution, the possibility of long-term and infrequent adverse events are difficult to deal with. These possibilities must be considered case-case by ethics committees and regulatory bodies that oversee the conduct of clinical trials, such as the FDA. The experience with related compound is an important consideration. We also note that safety monitoring continues after the drug has been granted a marketing licence. Safeguards have improved but are not infallible.

## Additional References

- Bennett, J.E., Wakefield, J.C. and Lacey, L.F. (1997). Modeling of trough plasma bismuth concentrations. *Journal of Pharmacokinetics and Biopharmaceutics*, **25**, 79-106.

- Bennett, J.E. and Wakefield, J.C. (1998). Errors-in-variables in joint PK/PD modeling. *Manuscript under preparation*.
- Evans, W.E., Taylor, R.H., Feldman, S. Crom, W.R., Rivera, G., Yee, G.C. (1980). A model for dosing gentamicin in children and adolescents that adjusts for tissue accumulation with continuous dosing, *Clinical Pharmacokinetics*, **5**, 295-306.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (1994). *BUGS: Bayesian Inference Using Gibbs Sampling, Version 3.0*. Cambridge: Medical Research Council Biostatistics Unit.
- Wakefield, J.C. and Rahman, N.J. The combination of population pharmacokinetic studies. *Submitted for publication*.