Population modelling in drug development

Lewis Sheiner and Jon Wakefield

Stat Methods Med Res 1999 8: 183
DOI: 10.1177/096228029900800302

Statistical Methods in Medical Research 1999; **8**: 183–193

Population modelling in drug development

Lewis Sheiner Departments of Laboratory Medicine, Biopharmaceutical Sciences, and Medicine, University of California, San Fransisco, California, USA and **Jon Wakefield** Department of Epidemiology and Public Health, Imperial College of Medicine at St Mary's, London, UK

In this paper we discuss the vital role that population (hierarchical) modelling can play within the drug development process. Specifically, population pharmacokinetic/pharmacodynamic models can provide reliable predictions of an individualized dose–exposure-response relationship. A predictive model of this kind can be used to simulate and hence design clinical trials, find initial dosage regimens satisfying an optimality criterion on the population distribution of responses, and individualized regimens satisfying such a criterion conditional on individual features, such as sex, age, etc. Throughout we emphasize prediction and advocate mechanistic as opposed to empirical modelling, and argue that the Bayesian approach is particularly natural in this setting.

1 Introduction

It is the aim of clinical drug development (first-in-man to approval) to turn chemicals into drugs; that is, to provide the 'user's manual' required for their safe and efficacious use. In this paper we discuss the vital role that population modelling may take within this development, emphasizing a mechanistic approach to modelling. Such an approach is desirable since it allows predictions to be made for conditions that have not so far been encountered. For example, outcomes conditional on so far unseen dosage schedules, and types of individuals, may be envisaged. The population approach involves both pharmacokinetic (PK) and pharmacodynamic (PD) modelling. Pharmacokinetics may be defined as what the body does to a drug, and pharmacodynamics as what the drug does to the body. More precisely:

- *Pharmacokinetics* describes the relationship between drug inflow (a more general term than 'dose') and drug concentration(s) at various body sites, notably the so-called bio-phases(s), or sites of drug action. Sub-processes (sub-models) for drug absorption, distribution, metabolism and elimination determine the relationship.
- *Pharmacodyanmics* describes the relationship between drug concentrations and pharmacological effects (sometimes called surrogate effects but more precisely called bio-responses), and the relationship, in turn, of these responses to clinical outcomes.

Population PK/PD models can provide a vital aid to the drug development process by providing reliable predictions of the individualized dose–exposure-effect relationship (where 'effect' refers to both efficacy and toxicity), which is key to successful therapy. Our use of 'population' makes explicit the fact that we wish to gain an understanding of the dose–exposure-effect relationship across different populations of

Address for correspondence: J Wakefield, Department of Epidemiology and Public Health, Imperial College of Medicine at St Mary's, London W2 1PG, UK. E-mail: j.c.wakefield@ic.ac.uk

 0962-2802(99)SM184RA

individuals as defined by demographic variables (such as age, sex, weight), biological information (such as the values of biological markers), genetic information, co-medications, environmental factors, and disease states. Such an understanding is, in effect, the 'user's manual' (drug label) previously alluded to.

Here we take a very broad view of 'exposure'. It can be the complete concentration versus time profile, or a summary measure such as the area under the concentration/time curve (AUC) or the maximum concentration. The 'effect' may be a pharmacological marker, an index of efficacy, or a measure of safety. Just what summary measure of concentration versus time should be used for 'exposure', and just what index of effect should be used for efficacy/safety is a matter for domain experts in medicine and pharmacology; further discussion is beyond the scope of this paper.

As has been discussed previously,[1,2] drug development is an information gathering process that can be thought of as two successive 'learning-confirmation' cycles. The first cycle (traditional phase 1 and phase 2a) addresses the question of whether benefit in terms of efficacy and safety can reasonably be expected over existing therapies. It involves 'learning' (phase 1) what is the largest short-term dose that may be administered to humans without causing harm, and then testing (phase 2a), whether that dose induces some measurable short-term benefit in patients for whom the drug is intended to be therapeutic. An affirmative answer at this first cycle provides the justification for a more elaborate second cycle (traditionally phase 2b and phase 3). The aim of this second cycle is to first learn (phase 2b) what is a good, if not optimal, dosage regimen to achieve useful clinical value (i.e. an acceptable benefit–risk ratio), and then to perform several formal clinical trials (in phase 3) of that regimen versus a comparator. At the confirmatory stage the most credible analyses are those that make as few assumptions as possible. From a Bayesian perspective, it has been argued,[3] that subjective prior distributions may be used for nonconfirmatory analyses (i.e. the learning stages in our terminology), while for confirmatory analyses objective priors are more appropriate.

The structure of this paper is as follows. In Section 2 we establish some notation and provide a statistical framework for the analysis of population PK/PD data. As briefly indicated above, we emphasize the predictive approach to modelling and in Section 3 we describe the distinction between mechanistic and empirical models. In Section 4 we discuss some specific issues that are relevant to population modelling. Section 5 contains a concluding discussion.

2 Statistical framework

We first establish notation. Let y and z denote, respectively, generic drug concentrations and effect measures of an individual. Each of these quantities may be multivariate, for example, the concentrations may be available for drug and metabolite and the effects may include measures of both efficacy and toxicity. Examples of effect measures include clotting times, pain scores and survival times. We let x denote the covariates of the individual, for example age and sex. The ostensibly controllable factors of the trial (the design) will be denoted by d. This quantity includes, for example, the numbers of individuals and the covariate-defined populations to be

studied, the treatments to be administered, and the outcome measurements to be made (including, but not limited to y and z), along with the (nominal) schedule of those measurements. We note that each of the quantities defined by d are *nominal*, and in practice deviations from this design will almost always be seen. For example, the actual times of dose administration and blood sample collection will differ from the nominal times. The actual times may or may not be recorded; we return to this aspect in Section 4.5. Similarly the treatment assigned will not always correspond to the treatment received. To ease the exposition we will not distinguish between the nominal and the actual characteristics of the trial, though this is certainly possible. The aim then is to build a model for y and z, conditional on x and d.

Once the design has been established we obtain concentrations and effect measures y_{ij}, z_{ij} from $i = 1, \ldots, N$ individuals at times t_{ij}, $j = 1, \ldots, n_i$. This notation implies that concentration and effect measures are obtained simultaneously; in fact, following the predictive approach that we advocate, this is not necessary, but for convenience we will ignore this aspect. Individual i has associated covariates x_i. In fact, these covariates may change over time but again for convenience we assume constancy of covariates.

Viewed from the dose–exposure-effect relationship it is natural to first model the concentrations given the dose, and then to model the effect as a function of exposure. We let θ denote a vector of PK parameters, ϕ a vector of PD parameters and $p(\cdot)$ a generic probability density function. The following factorization is then suggested

$$p(y, z|\theta, \phi, d) = p(y|\theta, d) \times p(z|\theta, \phi, d) \tag{2.1}$$

where, recall that d denotes the design and, in particular, includes the dosage regimen assigned/received, as well as the time points at which concentrations/effects are measured. Here, we have made the assumption that z is conditionally independent of y, given θ (and d). In particular this means that errors in y and z are unrelated and that we have a realistic PK model; if not then there may be additional information in y beyond that in the fitted values (say). The full probability model for each of the responses (concentration and effect) will also depend on variance–covariance parameters that describe the stochastic part of the model. For notational convenience (and because they are, in general, nuisance parameters) we have not included these parameters in (2.1).

We note here that it has been assumed (for convenience) that the PK and PD parameters are constant over time; this assumption is clearly false and the model can be elaborated to deal with this complication (for example, via the modelling of between-occasion variability[4,5]) if necessary.

It is natural to model the data from multiple individuals hierarchically. This modelling allows the variability in concentrations/effects to be separated into within-individual and between-individual components. At the first stage of the hierarchy the data of each individual are modelled conditional on a set of individual-specific PK and PD parameters. The fundamental assumption here is that the biological processes that are acting on each of the individuals are similar and so the functional forms of the concentration/effect profiles are identical. Differences between individual profiles arise because each individual has his or her own set of PK and PD parameters. These

differences are modelled at the second stage of the model by assuming that, given the covariates x of a particular individual, the PK and PD parameters may be viewed as realizations from probability density functions.

Hence we have the following two-stage hierarchy:

- *First-stage model.* Model for within-individual variability:

$$p(y_{ij}|\theta_i, d) \quad \text{and} \quad p(z_{ij}|\theta_i, \phi_i, d), \quad i = 1, ..., N;\ j = 1, ..., n_i$$

- *Second-stage model.* Model for between-individual variability:

$$p(\theta_i|\mu_1, \Sigma_1, x_i) \quad \text{and} \quad p(\phi_i|\mu_2, \Sigma_2, x_i), \quad i = 1, ..., N$$

  Here μ_1 and Σ_1 (μ_2 and Σ_2) denote the population mean and variance–covariance parameters of the PK (PD) parameters. Note that we have assumed that the distributions of the PK and PD parameters are independent given x.

A Bayesian analysis would add a third stage to the model containing prior distributions on the population parameters.

Let x denote the totality of covariate information, and y and z the totality of PK and PD data, and θ and ϕ the collection of all of the individuals PK and PD parameters. The probability model for y and z is then given by

$$\begin{aligned} p(y, z|\theta, \phi, x, \mu_1, \Sigma_1, \mu_2, \Sigma_2, d) = \Pi_i \Pi_j p(y_{ij}|\theta_i, d) p(z_{ij}|\theta_i, \phi_i, d) \\ \times \Pi_i p(\theta_i|\mu_1, \Sigma_1, x_i) p(\phi_i|\mu_2, \Sigma_2, x_i) \end{aligned} \tag{2.2}$$

To illustrate the above model we now describe a specific modelling structure. This structure is widely used for population PK, and PK/PD analyses (see Sheiner and Ludden[6] for a review, or for a specific example, Wakefield *et al.*[7]).

First-stage model

Suppose each individual provides a set of univariate drug concentrations and/or effect measures. We let $f_{PK}(\theta_i, d)$ and $f_{PD}(\theta_i, \phi_i, d)$ denote the functional forms of the PK and PD models, respectively (recall that the sampling times are contained in d). It may then be assumed that

$$y_{ij} = f_{PK}(\theta_i, d) + \epsilon_{ij}^{PK}$$

where the error terms ϵ_{ij}^{PK} are independently and identically distributed normal random variables with mean zero and a variance that may depend on the mean concentration f_{PK}. For example, a common model takes the standard deviation of the concentration to be proportional to $f_{PK}(\theta_i, d)$; alternatively a transformation of the concentration (usually the logarithm) may be taken to produce constant variance. The PD response may similarly be modelled as

$$z_{ij} = f_{PD}(\theta_i, \phi_i, d) + \epsilon_{ij}^{PD}$$

with, again, independent and identically distributed error terms ϵ_{ij}^{PD} whose variance

may depend on f_{PD}. The forms of f_{PK} and f_{PD} are obviously drug and effect specific (see Section 3). It is common to make normality assumptions for both the PK and the PD error terms.

Second-stage model

It is usual (but not universal) to model the PK and PD parameters (after each have been transformed onto the whole of the real line) via:

$$\theta_i = x_i'\mu_1 + \delta_i^{PK} \quad \text{and} \quad \phi_i = x_i''\mu_2 + \delta_i^{PD} \tag{2.3}$$

where the error terms δ_i^{PK} and δ_i^{PD} are independent and identically distributed and are assumed to have zero mean and variance–covariance matrices Σ_1 and Σ_2, respectively. The covariate vectors x_i' and x_i'' are constructed as subsets of the original x_i. We discuss some of the issues relating to this modelling in Section 4.4 below. The most usual choice of second-stage distribution is the multivariate normal distribution, although we discuss this choice in more detail in Section 4.2.

From a simulation viewpoint we may generate a set of PK/PD data by: (a) selecting a design d; (b) simulating specific individual covariates x, conditional on the design d, and (c) simulating individual data y and z. In (b) the study population (e.g. healthy volunteers or patients) will have been defined by the design, but there will still be some uncertainty in the actual values of x that are realized. Steps (b) and (c) both, therefore, require probability models. Once the data have been obtained then for analysis the design d and the covariates x are conditioned upon. Models for the distribution of x are thus important for simulating clinical trials, but are less important for other uses of modelling in drug development, since, as just noted, they are conditioned upon in the analysis stage. Similarly, when the assumption that doses assigned are doses received is abandoned, a model for actual design conditional on x and the nominal design d will also be needed for simulation, but not for analysis. This is because, in the absence of actual design observations, the analysis is carried out according to the intention-to-treat principle. In the presence of actual design observations, they are usually conditioned upon.

3 Types of models

As pointed out in the introduction, prediction is of utmost importance in clinical drug development. In particular we wish to construct an input-outcome model that relates drug inflow (which is, for the purposes of this paper, as noted above, part of the design d) to clinically important outcomes across all relevant covariate groups. To achieve this a model must interpolate between and extrapolate beyond, the value of the conditions of the actual study or studies available. The need for a predictive model directly drives the types of models that can be considered as useful learning model candidates; most importantly, to provide credible extrapolation, they must be mechanistic (i.e. embody relevant scientific knowledge) as opposed to empirical.[2,6]

We consider first the dose–concentration relationship (the PK model); a more complete discussion is provided, for example, by Rowland and Tozer.[8] After introduction into the body the drug undergoes absorption, distribution, metabolism

and elimination. Models are assumed for each of these processes, often defined in terms of differential equations that describe the rate of flow of drug between a series of compartments that are nominally assumed to model the body. These differential equations frequently lead to a concentration–time profile that is of sums-of-exponential form. The parameters of this form, or whatever form is applicable, are contained in θ. In terms of the mechanistic framework that we have been advocating, it is important that these models are valid for a range of administration routes and drug input rates. In this way predictions may be made for dosing scenarios not already investigated, and hence, desirable doses; i.e. those fulfilling population-level performance objectives with specified probability, can be discovered for covariate classes (old, young, male, female, etc.).

The exposure–effect relationship (the PD model) can take a far wider range of forms since there are numerous outcome measures that may be considered.[9] For example, clotting measures are continuous and may be modelled using the $E_{\max}$ model[8,9] while pain scores assume one of a discrete number of outcomes and may be modelled using cumulative logit models.[10]

For both the PK and the PD models a *subject-specific*, as opposed to a *population-averaged*, approach (see, for example, Diggle *et al*.[11]) to modelling is preferable. This is because: (a) it allows predictions to be made for specific individuals and it is individuals for whom dose schedules must ultimately be produced; (b) mechanistic models for subject-average responses rarely exist: physiology (mechanisms) directly govern subject-specific models, not population averages; and (c) conditional modelling using random effects allows examination of specific components of within- and between-individual variability, which is often of great interest in itself.

A major difficulty in later-phase clinical trials is noncompliance, which can be (a) failure to adhere faithfully to the assigned dosage regimen, and (b) an extreme form of (a), drop-out. There is a growing literature on these topics (see, for example Urquhart,[12] Angrist *et al*.,[13] Sheiner and Rubin[14] and Frangakis and Rubin[15]). Regarding (a), when actual dosage and nominal dosage do not coincide (which is the usual case in later phases), confounding becomes a serious issue: using an intention-to-treat analysis (i.e. proceeding as we have here by assuming the two dosages are identical) is not a valid solution if a predictive model is a goal.[14] There is also the technical complication of having to account for a different 'design' for each individual. Regarding (b), we note here that, in some instances, bias will be introduced if the data of patients who drop-out is treated in an identical fashion to those that complete the study. A number of strategies for dealing with this problem have been suggested (see, for example, Little[16]), and because it has bearing on the selection of models, we discuss them briefly now. For simplicity, consider a single individual and PK data Y only, and let S denote a random variable that denotes the time of dropout. The aim is to provide a joint model for Y and S. Two modelling strategies are then suggested: the *selection model* approach considers $p(y,s) = p(y) \times p(s|y)$ while the *pattern mixture model* approach considers $p(y,s) = p(s) \times p(y|s)$. From a predictive standpoint the selection model approach is far more appealing since the dropout is treated as a function of the response y (although one may imagine situations in which the pattern-mixture model is appropriate, see Verbyla's discussion of Diggle and Kenward[17]). Unfortunately the

selection model approach requires assumptions to be made that may not be testable from the data alone; again subject-matter information, and a Bayesian approach, may aid in this difficulty.

We will not discuss the relative merits of the likelihood and Bayesian approaches to inference here. We note, however, that under the Bayesian approach prediction is far more natural, and the incorporation of prior information is wholly consistent with the learning process that we are advocating.

4 Statistical issues

In this section we discuss some of the outstanding statistical issues in population modelling.

4.1 Implementation

Both likelihood and Bayesian approaches have been suggested for inference. Neither are trivial to implement due to the nonlinear models that are typically assumed for the first stage PK and PD models. In particular, the likelihood for the population parameters is given by

$$l(\mu_1, \Sigma_1, \mu_2, \Sigma_2) = \Pi_i \int \int \Pi_j p(y_{ij}|\theta_i, d) p(z_{ij}|\theta_i, \phi_i, d) \times p(\theta_i|\mu_1, \Sigma_1, x_i) p(\phi_i|\mu_2, \Sigma_2, x_i) d\theta_i d\phi_i$$

These integrals are analytically intractable and hence maximization of the likelihood, or examination of the posterior that is obtained by combining $l(\mu_1, \Sigma_1, \mu_2, \Sigma_2)$ with a prior distribution is not straightforward. We do not provide a review of approaches to implementation here; see Davidian and Giltinan[18] and Racine-Poon and Wakefield.[19]

4.2 Parametric vs nonparametric

At the first stage of a population PK model mechanistic models are the rule, and usually simplified versions of these (e.g. compartmental models) have been found to provide an adequate description of the concentration–time profile. Similarly, the rather more empirical assumption of normal errors has been found to be adequate, often based on assay-validation data. Consequently, there is no need, in general, to consider nonparametric or empirical submodels at the first stage of the PK model, and indeed there are good reasons to eschew them: they do not in general incorporate subject matter knowledge, and they are less parsimonious (in a Bayesian context) than mechanistic models that do. In general, the PD model is less subject-matter motivated and so there may be some need to consider empirical or nonparametric approaches here, but even so, forms must be chosen that obey realistic subject matter constraints (e.g. positivity and monotonicity are usual).

A more robust alternative to the normal distribution that has been used in a population pharmacokinetic/pharmacodynamic context is the Student t-distribution.[7] Although this distribution may accommodate outlying individuals more readily than the normal distribution, it may still be somewhat restrictive, for example, the Student t-distribution is unimodal and symmetric.

A number of approaches to the nonparametric estimation of the second stage distribution have been suggested in the context of population PK modelling (the seminal work is that of Mallet;[20] see Racine-Poon and Wakefield[19] for a review of various likelihood and Bayesian approaches). In general, there is less certainty in the second-stage distribution simply because the modelling concerns unobservable quantities. In the PK/PD context there are particular reasons to believe that multi-modalities, for example, may be present in the population distribution when important covariates such as genetic polymorphism of drug-metabolizing enzymes are unmeasured. Consequently, flexible modelling is required, in particular to identify vulnerable subgroups, although semi-mechanistic models for at least the mean structure of the relationship between parameters and covariates are often available (see Section 4), and should not be neglected.

4.3 Diagnostics

The multi-stage hierarchical models that are typically used for population modelling contain many layers of assumptions and currently there is very little guidance as to the checking of these assumptions. This is not restricted to population modelling but is true of all hierarchical modelling, see Hodges[21] and the accompanying discussion for the current position.

4.4 Covariate modelling

The choice of which elements to include at the second stage for the modelling of θ_i and ϕ_i (i.e. in equations (2.3)) is difficult. Fortunately, scientific understanding usually exists to guide the choice. From a purely statistical viewpoint, the difficulty is that one is confronted with a multiple regression in which the dependent variables θ and ϕ are multivariate and unobserved. Covariate selection has often proceeded via a forward selection procedure, with graphical plots driving the inclusion of covariates (first proposed in the PK/PD literature by Maitre *et al*.,[22] and later elaborated upon by Mandema *et al*.;[23] see Davidian and Giltinan[18] for further discussion). Using such plots to drive a forward selection procedure is hazardous in several ways. It is not easy to decide on which of a discrete or a continuous covariate to include first; adding more than one covariate at a time is also dangerous as the covariates are correlated; and choosing which element of θ or ϕ for which a particular covariate should be included is also difficult. There is no straightforward solution to these difficulties except to try to incorporate as much subject-matter information as possible, both as to which covariates are likely to be important, and to the functional form taken. This is again vital if extrapolation outside of the range of the observed covariate space is considered.

As a final comment we note that the majority of standard procedures for covariate selection judge significance in a *statistical* and not a *clinical* sense. Wakefield and Bennett[24] consider this problem in a specific application in which the design of a dosage regimen was the objective.

4.5 Errors-in-variables modelling

There are a number of components of population PK/PD models for which it may be more appropriate to use errors-in-variables modelling, particularly in the later stages of drug development when the environment of data collection is less controlled.

The sampling times are often nominal times with the scheduled collection times being reported rather than the actual times. In this case a Berkson model[25] may be appropriate; Wang and Davidian[26] have examined this issue. Similarly the sizes and sampling times of doses may be inaccurate and may be treated the same way, although this does not deal with the problem of confounding. While covariate information such as age and sex may be accurately measured, other information will not. For example, biological measurements such as markers in the blood will be subject to measurement error and hence, strictly, should not be treated as fixed quantities. Even such variables as age and sex are more properly regarded as 'noisy' surrogates for the underlying physiological states that determine PK/PD, and analyses might profit from recognizing this.

Joint PK/PD data may be analysed in a number of ways. The simplest approach is to substitute the observed concentrations (or some summary) into the PD model though this approach neither acknowledges that the concentrations are measured with error, nor if the PD model is expressed as differential equations, or PD observations are made at times distinct from the PK ones, does it provide continuous-time PK as required. A slightly more refined approach is to first model the PK data and then to substitute in the fitted concentrations or the continuous time PK function (see Shi *et al*.[27] for an example). The difficulties here are that (a) the uncertainty in the fitted concentrations is not acknowledged and, (b) since there is no feedback in the model, the PD data will not aid in the estimation of the PK parameters. The joint modelling of the dose–exposure-effect relationship, as outlined in Section 2, is the ideal but may not always be convenient. Regarding (a), Bennett and Wakefield (Errors-in-variables in joint population pharmacokinetic/pharmacodynamic modelling, not yet published but available from the authors) showed that the use of an oversimplistic model can lead to bias in the estimation of the parameters of the PD model and instead used an errors-in-variables approach in which the uncertainty in the concentration data was acknowledged. Unfortunately this approach does not allow prediction from dose to effect because there is no mechanistic model connecting the dose to the exposure. Regarding (b), this may be more an advantage than a drawback in that PD models are usually far less certain than PK ones, and a fit that distorts the PK profile to accommodate PD model misspecification will obscure rather than illuminate PD model inadequacies.

4.6 Design

When the data are to be modelled by a nonlinear hierarchical model the design, that is the choice of sampling times, numbers of samples, and number of individuals, remains a challenge. The work of Mentre *et al*.[28] is one of the few attempts at solving this problem, see also Tod *et al*.[29] Although the general problem is very difficult, the design for an actual population study is theoretically more straightforward; the design space is often highly restricted due to logistical considerations, and there is typically a large amount of prior information concerning the form of PK/PD model and range of parameter values. An example of an approximate but practical approach is provided by the work of Bruno *et al*.[30] with Docetaxel.

5 Discussion

In this paper we have stressed the predictive approach to drug development and the analysis of clinical data. The unavoidable possibility of serious error despite using the best available scientific knowledge, when extrapolation is being carried out beyond conditions already seen has undoubtedly led many statisticians to eschew such an approach and to advocate primarily descriptive analyses with hypothesis tests based on the design-imposed sampling distribution of that data. While we do not intend in any way to minimize the possibility of serious error, we do take exception to an over-cautious approach. Extrapolations must be made: some dosage must be chosen for subsequent trials; physicians must daily make usage decisions under as yet unstudied circumstances. Sound public policy and medical ethics both dictate that statistical and subject-matter scientists cooperate to make such extrapolations as fully informed as possible.

References

1 Sheiner L. Learning versus confirming in clinical drug development. *Clinical Pharmacology and Therapeutics* 1997; **61**: 275–91.

2 Sheiner L, Steimer J-L. Pharmacokinetic/pharmacodynamic modelling in drug development. *Annual Review of Pharmacology and Toxicology* 2000 (in press).

3 Racine-Poon A, Grieve A, Fluhler H, Smith, AFM. Bayesian methods in practice: experiences in the pharmaceutical industry (with discussion). *Applied Statistics* 1986; **35**: 93–150.

4 Karlsson MO, Sheiner LB. The importance of modelling inter-occasion variability in population pharmacokinetic analyses. *Journal of Pharmacokinetics and Biopharmaceutics* 1993; **21**: 735–50.

5 Lunn DJ, Aarons L. Markov chain Monte Carlo techniques for studying inter-occasion variability: application to population pharmacokinetic data. *Applied Statistics* 1997; **46**: 73–91.

6 Sheiner LB, Ludden TM. Population pharmacokinetics/pharmacodynamics. *Annual Review of Pharmacology and Toxicology* 1992; **32**: 185–209.

7 Wakefield JC, Aarons L, Racine-Poon A. The Bayesian approach to population pharmacokinetic/pharmacodynamic modelling. In: Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M eds. *Case studies in Bayesian statistics, Volume IV*, pp. 205–265, 1999.

8 Rowland M, Tozer TN. *Clinical pharmacokinetics: concepts and applications*, 3rd edn. Baltimore, MD: Williams & Wilkins, 1995.

9 Holford N, Sheiner L. Understanding the dose–effect relationship: clinical applications of pharmacokinetic/pharmacodynamic models. *Clinical Pharmacokinetics* 1991; **6**: 429–53.

10 Sheiner LB, Beal SL, Dunne A. Analysis of non-randomly censored ordered categorical longitudinal data from analgesic trials (with discussion). *Journal of the American Statistical Association* 1997; **92**: 1235–55.

11 Diggle PJ, Liang K-Y, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford Science Publications, 1994.

12 Urquhart J. Role of patient compliance in clinical pharmacokinetics: review of recent research. *Clinical Pharmacokinetics* 1994; **27**: 202–15.

13 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**: 444–72.

14 Sheiner LB, Rubin DB: Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* 1995; **57**: 6–15.

15 Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none

treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999; **86**: 365–79.
16 Little RJA. Modelling the drop-out mechanism in repeated-measured studies. *Journal of the American Statistical Association* 1995; **90**: 1112–21.
17 Diggle PJ, Kenward MG. Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* 1994; **43**: 49–93.
18 Davidian M, Giltinan DM. *Nonlinear models for repeated measures data*. London: Chapman & Hall, 1995.
19 Racine-Poon A, Wakefield JC. Statistical methods for population pharmacokinetic modelling. *Statistical Methods in Medical Research* 1998; **7**: 63–84.
20 Mallet A. A maximum likelihood estimation method for random coefficient regression models. *Biometrika* 1986; **73**: 645–56.
21 Hodges JS. Some algebra and geometry for hierarchical models, applied to diagnostics (with discussion). *Journal of the Royal Statistical Society, Series B* 1998; **60**: 497–536.
22 Maitre PO, Buhrer M, Thomson D, Stanski DR, A three-step approach combining bayesian regression and NONMEM population analysis: application to midazolam. *Journal of Pharmacokinetics and Biopharmaceutics* 1991; **19**: 377–84.
23 Mandema JW, Verotta D, Sheiner, LB. Building population pharmacokinetic-dynamic models. I. Models for covariate effects, *Journal of Pharmacokinetics and Biopharmaceutics* 1992; **20**: 511–28.
24 Wakefield JC, Bennett JE. The Bayesian modelling of covariates for population pharmacokinetic models. *Journal of the American Statistical Association* 1996; **91**: 917–27.
25 Carroll RJ, Ruppert D, Stefanski LA. *Measurement error in nonlinear models*. Chapman & Hall, London, 1995.
26 Wang N, Davidian M. A note on covariate measurement error in nonlinear mixed effects models. *Biometrika* 1996; **83**: 801–12.
27 Shi J, Benowitz NL, Denaro CP, Sheiner, LB. Pharmacokinetic-pharmacodynamic modeling of caffeine: tolerance to pressor effects. *Clinical Pharmacology and Therapeutics* 1993; **53**: 6–14.
28 Mentre F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika* 1997; **84**: 429–42.
29 Tod M, Mente F, Merle Y, Mallet A. Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics* 1998; **26**: 689–716.
30 Bruno R, Vivier N, Vergniol JC, De Phillips SL, Montay G, Sheiner LB: A population pharmacokinetic model for Docetaxel (Taxotere): Model building and validation. *Journal of Pharmacokinetics and Biopharmaceutics* 1996; **24**: 153–72.