THE ASSESSMENT OF PHARMACODYNAMIC VARIABILITY

ASSOCIATED WITH CATEGORICAL RESPONSE MEASURES

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy in the Faculty of Science

1999

Gordon Graham

School of Pharmacy and Pharmaceutical Sciences

1

ProQuest Number: 10758064

![ProQuest logo]

ProQuest 10758064

Contents

Abstract

The analysis of population pharmacokinetic/pharmacodynamic data is carried out to determine a model that describes the population and the individual in terms of the plasma concentration-time profile and the pharmacological effect-concentration relationship as well as describing the variability between subjects in both the pharmacokinetics and pharmacodynamics. Moreover, explaining the variability from both data sets and obtaining a model that links both pharmacokinetics and pharmacodynamics in a simultaneous analysis is one of the main goals.

Pharmacodynamics, unlike pharmacokinetics, can be described by a wide range of variables that represent the different effects of the particular drug. As well as there being many markers to represent the effect of the drug, the type of data that are observed can be either continuous, count, categorical or censored. The purpose of the current work is to consider categorical pharmacodynamic data and how such data can be analysed. One of the main aims of analysing categorical pharmacodynamic data is to describe the central tendency of the data and quantify and explain the interindividual variability.

Two toxicokinetic data sets for two different compounds which were administered in groups of animals for the evaluation of pharmacokinetics and toxicity of the drugs were considered. The first analysis demonstrated the use of the proportional odds model and mixed effects modelling when applied to categorical pharmacodynamic data and considered the effects of including pharmacokinetic information through the exposure to the drug by using estimates of individual AUC values. Model comparisons were also investigated using Bayes factors. This method allows the comparison of non-nested models which is usually difficult to carry out. The second toxicokinetic study comprised 9 data sets from 3 animal species. Scaling of the pharmacodynamics as well as the pharmacokinetics was considered across the 3 species. The analysis of the data was carried out in BUGS and NONMEM.

Categorical data was collected from a phase II clinical trial on sumatriptan where the pharmacodynamic variable, migraine pain relief, was measure on a 4 point scale, none, mild, moderate and severe pain. It was required to characterise the pharmacokinetic and pharmacodynamic data in terms of a population model and quantify the interindividual and residual variability and to consider the incorporation of pharmacokinetic information to describe the categorical pharmacodynamic response with the use of concentration and AUC predictors. In another study, oxybutynin data was available from a phase III trial where one of the aims was to determine an optimum dose based on efficacy and adverse effect data. The efficacy of oxybutynin was measured as the number of urinary urge incontinence episodes in a week and the adverse effect was defined as a categorical score on the degree of dry mouth. To determine the optimum dose, a utility function was developed and optimised in a Bayesian framework.

Before categorical data can be collected from clinical trials, it is important that the trial be designed appropriately so that the information collected is relevant to the objectives being considered. D-optimal designs were investigated for the 3 category proportional odds fixed effects model with one independent variable. Results obtained showed that for a 3 category model, there were typically 3 distinct design points but varied according to the total number of observations required.

6

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for the degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to thank the following people:

Leon Aarons, as supervisor, for getting me to this stage and for all the help and encouragement he has given me.

Eliane Fuseau, Malcolm Young and Glaxo Wellcome for providing data, money and a couple of trips to London.

Steve Duffull for asking too many questions.

Dad, Andrew, Lawrence and Margaret for always letting me make my own decisions and supporting me through all my years at university.

Mum, for the same reasons as above but I'm sure she would have been very proud had she still been with me now.

Finally, to the most important person in my life, Hannah who has been so supportive in my most 'interesting' year yet.

1 Introduction

1.1 Pharmacokinetics

Pharmacokinetics can be described as the study of drug and metabolite concentration-time profiles in the body as affected by absorption, distribution and elimination. The pharmacokinetics of a drug is studied during the different phases of the drug's development. These studies can range from early preclinical studies where the compound is administered to animals to large phase III clinical trials where the drug is given to individuals for whom the drug is intended. The data collected from such studies can be analysed in a variety of ways, but often, one of the main aims is to describe the data in terms of a model. Pharmacokinetic data has been modelled in many different ways to try and gain more accurate information on the drug and its pharmacokinetic parameters. In terms of modelling pharmacokinetic data, as well as describing the data in a population in terms of an average response, it is important to quantify the variability in the data. Variability in pharmacokinetics is an important factor to take into consideration when developing a drug as this can lead to important differences in how the drug should be administered to different sub-groups of the population. These subgroups can be distinguished by certain subject specific features such as age, weight, gender, race and other important covariates that differentiate one individual from another. The analysis of pharmacokinetic data where the mean response and variability between individuals is quantified is called population pharmacokinetics.

1.2 Pharmacodynamics

Pharmacodynamics can be defined as the study of drug effect related to some function of time, dose or pharmacokinetic measure. Pharmacodynamics is often not as well understood as pharmacokinetics, but being able to describe the pharmacodynamics is equally, if not more important, to describing the pharmacokinetics. This is certainly true in the opinion of the patient as it is the effect of the drug that is important to them rather than how the drug is absorbed, distributed and eliminated from the body. How the drug attains an effect is of less interest to the patient but having an understanding of how the drug effect is reached is important for the development of models describing the data and hence, for example making sensible inferences about treatment of the condition for the which the drug is designed. The mechanisms for how the drug effect is attained are generally harder to determine than the mechanisms that control the pharmacokinetics of the drug. This has been in part due to the ability to measure pharmacodynamic variables *in vivo* which has lagged behind the ability to measure plasma drug concentrations (Levy (1985)). As well as the ability to measure the pharmacodynamic variables, there is also the problem of which markers to measure. The data that arises from pharmacokinetic/pharmacodynamic studies often arise from complex dynamic models. They are typically of the same format as that of the pharmacokinetics, which are short times series of repeated measures from a number of individuals. Unlike pharmacokinetic data which are of a continuous nature, pharmacodynamic data can be either continuous, categorical, count or censored data. This makes the planning and analysis of pharmacodynamic studies that much more difficult than that of pharmacokinetic studies. The ability to select a supposedly appropriate model for pharmacokinetic modelling is not quite so easy for

10

pharmacodynamic data as ideas such as compartmental modelling do not necessarily apply. This means that other models need to be considered and these are often of an empirical nature. The most frequent model for pharmacodynamic data is a variant of the sigmoid E_{max} model described in section 2.2. This is still an empirical model but with a pharmacological interpretation and can be found in many pharmacodynamic analyses.

1.3 Preclinical and Clinical Trials

Drug development describes the whole process of taking a newly discovered entity through regulatory approval and to the point of market introduction. The process of developing a drug is sequential where the information concerning the compound is built up through preclinical studies in animals, then to the first application in man in phase I studies where the drug is administered to healthy volunteers and then into larger population studies in phase II/III where the drug is given to individuals for whom the drug is intended. It is at these stages of drug development that data pertaining to the pharmacokinetic and pharmacodynamics of the drug are collected.

1.3.1   Preclinical Studies

The first time a drug is studied *in vivo* is in animals. These studies are performed to consider a whole range of aspects of the drug which would not be so convenient and ethical to do in humans. Primarily, they are to check that the drug in consideration does not have toxicities that would possibly be harmful when administered in humans. Such

toxicology studies are not only for this purpose, but many other types of data are collected that might be of relevance for the particular drug. Also of importance in preclinical trials is to gain an understanding of the pharmacokinetics of the drug in animals for helping in designing future clinical trials. Preclinical studies are usually performed in homogeneous groups of animals (such as rats, dogs or rabbits). Often the studies are designed so that, for example, in a toxicokinetic study, one plasma concentration sample is obtained from each animal at a certain time and for a certain dose after the animal has been terminated. Such destructively obtained data allows the collection of other information that might be time and dose varying such as concentration data in other tissues in the body..

1.3.2 Phase I Clinical Trials

Phase I clinical trials is when the drug is first administered in humans which is usually in healthy male volunteers. The main aim of such trials is to obtain information on the pharmacokinetics of the drug in (healthy) humans. Healthy volunteers are selected because a homogeneous human population is required to study the pharmacokinetics without introducing too many confounding factors. The size of the phase I population is usually small and in a well controlled environment which also allows any variability between and within an individual to be reduced. As the volunteer is in a well controlled environment, this permits for the collection of many plasma drug concentrations over time. As this produces a rich data setting, it is possible to use classical regression techniques for the analysis of each individual's concentration-time data. Individual estimates can be obtained in such a setting but population modelling techniques can also be used to study the data, discussed in section 1.4. Other considerations in phase I

clinical trials include the assessment of linearity of the pharmacokinetics (dose proportionality) and the comparison of different formulations. Information collected at phase I can then be used to plan phase II clinical trials.

1.3.3 Phase II Clinical Trials

Phase II clinical trials are designed for the first administration of the drug into patients for whom the drug is intended. The population of patients involved is often only a little larger than that in phase I. The sample from the population is still in a well controlled setting but some variability is allowed to be introduced so that covariates that may influence the pharmacokinetics can be examined. These covariates must be studied carefully so that the covariate effect is not confounded with that of the drug effect which can be checked by comparing results in phase II to the results in phase I. One of the main goals in phase II clinical trials is to define a therapeutic window for the drug and determine the optimal dosage regimen. The therapeutic window is defined as the range of plasma concentrations from the minimum concentration to obtain an effect to the concentration that produces the maximum tolerable toxic effect. This can usually be determined by the analysis of efficacy and toxicity data also collected at this stage of the drug development by comparing the efficacy of the drug to the toxic effects of the drug. An optimal dose is obtained when a balanced level of hopefully high efficacy and low toxicity is reached. As well as the linearity of the pharmacokinetics being studied in healthy volunteers, it also needs to be studied in the patients for whom the drug was intended, as the linearity may not be the same for the two studied groups. From the results of the phase II clinical trials, phase III trials can be planned so that as an optimal treatment is administered.

1.3.4 Phase III Clinical Trials

Phase III clinical trials are designed so that they are as similar to the eventual setting of administering the drug to the population of individuals who will actually receive the drug once the drug has been approved. This is so that the drug can be assessed to see how well the drug would perform in a 'real' setting. These studies are still planned in advance but tend to be observational rather than experimental studies as nearly all of the factors involved in assessing an individual's pharmacokinetic behaviour are not controlled in any way. Phase III trials are often run with a large number of patients, as the population is heterogeneous. This allows the assessment of interindividual variability in terms of subject specific covariates that may have an effect on the individual's pharmacokinetic behaviour. As the collection is not as controlled as in phase I and II clinical trials and the plasma concentrations are taken when patients visit a trial centre, fewer concentrations tend to be collected from each patient. It is normal to have as few as one or two concentration measurements from each individual in phase III clinical trials which does not allow for individual model fitting to be carried out. Instead, complex statistical techniques are employed which allow both population and individual pharmacokinetic parameter estimates to be obtained as well as quantify the variability in the data. This type of approach is known as the population approach.

1.4 Population Pharmacokinetics/Pharmacodynamics

The population approach to the analysis of pharmacokinetic and pharmacodynamic data from large scale clinical trials has increasingly become a standard tool in the

14

pharmaceutical industry (Vozeh (1997)). For practical and ethical reasons, it is not possible to collect extensive amounts of data from patients in phase III clinical trials as in phase I and II trials. In this case, there is limited individual data but data is collected from many patients giving a large data base of pharmacokinetic and pharmacodynamic information from the population of interest. Such data has come under considerable interest over the years and the methods that have generally been applied to analysing such data have come to be known as population pharmacokinetic/pharmacodynamic modelling.

Population pharmacokinetic modelling was first proposed by Sheiner *et al* (1977). In that paper, a general data analysis technique for estimating population average parameters and their interindividual variability from routine clinical pharmacokinetic data was described. This approach did not require the intermediate estimation of individual pharmacokinetic parameters. The method was applied in a setting where patients were sampled on a few occasions only and other routinely assessed variables were available. This is the setting encountered when pharmacokinetic and pharmacodynamic data are collected from phase III clinical trials and hence succinctly described the general problem and one possible solution.

The models used in pharmacokinetics are generally nonlinear and hence nonlinear regression techniques are required to obtain parameter estimates. As well as obtaining the population average parameter estimates, quantifying the interindividual and residual variability are also required. To obtain estimates of these measures of variability, nonlinear mixed effects regression analysis needs to be carried out. The 'mixed effects' corresponds to the fixed effects representing the population average parameters and

random effects representing the variance components. Fixed and random effects are described in more detail in section 2.3.2. How these parameters are estimated has resulted in a considerable amount of research.

As one of the main aims of population pharmacokinetics and pharmacodynamics is to quantify and hopefully explain the variability between subjects, it is important to consider modelling of subject specific covariates. It is usually desirable to estimate the variability as a component of the pharmacokinetic and pharmacodynamic parameters as it is often the parameters that contain much of the information about the compound. To try and explain how the parameters might vary between individuals, the individual estimates of the parameters can be regressed on as a function of the subject specific covariates. If there are any significant relationships between the parameter of interest (clearance and volume of distribution for example) and the covariates, then these can reduce the unexplained parameter variability between subjects. What variability is not defined by subject specific information in the population model is generally termed residual variability. Such covariate modelling is described by Wakefield and Bennett (1996) and Mandema *et al* (1992).

There is a wide range of methods for determining relevant population characteristics. Probably the most basic technique for the analysis of population data is the naïve pooling approach which does not allow for the estimation of interindividual variability. This method, as it is called, pools all the data together and then estimates population average parameters. This approach tends to produce biased estimates as it does not take into account the influence of different individual's data on the parameter estimates and the combination of variance components. There has been a wide range of two step

16

approaches proposed, some of which are described by Steimer *et al* (1985). The basic idea of the two stage approach is to firstly estimate the individual parameters and then by some method, estimate the population parameter estimates and the interindividual variability based on the individual estimates. A concern with this method is that there might be a problem in estimating the individual parameter values and the combination of the population parameter estimates is often an *ad hoc* procedure. Other more sophisticated forms of the two step approach have been proposed, such as that by Mentre and Gomeni (1995). A more general approach is that of maximum likelihood which is essentially carried out by NONMEM (Beal and Sheiner (1992)). This is described in more detail in section 2.6.6 but, generally, this method allows a function known as the likelihood to be defined that takes into account all the population parameters and the interindividual variability simultaneously. In NONMEM the method used is known as extended least squares which is very similar to other methods based on distributional assumptions, also described in section 2.6. Other least squares approaches that have been used are generalised least squares and weighted least squares which are more or less general than extended least squares respectively (Davidian and Giltinan (1995)). As well as maximum likelihood techniques that are based on parameter estimation from the data alone, Bayesian methods allow the inclusion of prior knowledge (Racine-Poon and Wakefield (1998)). This is more general than the method of maximum likelihood as by simply taking out the prior distributions, a maximum likelihood formulation will usually be obtained. The inclusion of prior beliefs and knowledge is a useful step that can be implemented as it allows the passing of information from one phase of a drug development program to another. As well as parametric approaches, there are nonparametric approaches that do not make explicit assumptions about the way in which the data are distributed. Nonparametric methods

can be considered in terms of maximum likelihood and Bayesian ideas. Mallet *et al* (1988) and Schumitzky (1991) report on nonparametric maximum likelihood methods and Wakefield and Walker (1997) report on a Bayesian nonparametric method.

The general ideas of population modelling apply equally to pharmacokinetics as to pharmacodynamics. Most of the research carried out has been in the area of population pharmacokinetics but this has probably been due to the greater understanding of pharmacokinetics than of pharmacodynamics. More interest is growing in applying the ideas to pharmacodynamic data as it is just as relevant to understand the variability in the pharmacodynamics as the pharmacokinetics. Understanding the link between the pharmacokinetics and the pharmacodynamics is also of utmost importance to describing the variability in a particular individual and across a population in terms of the pharmacokinetic and pharmacodynamic models.

Since the introduction of the population approach to analyse routine clinical pharmacokinetic data, there have been many analyses carried out and reported in the literature. A few examples of these are for digoxin (Sheiner *et al* (1977)), quinidine (Fattinger *et al* (1991) and Davidian and Gallant (1992)), cyclosporine (Mallet *et al* (1988)), paclitaxel (Karlsson *et al* (1998)) and tobramycin (Aarons *et al* (1989)). Papers corresponding to the population analysis of pharmacodynamic data are ivabradine (Rageneau *et al* (1998)), glibenclamide (Rydberg *et al* (1997)), ketorolac (Mandema and Stanski (1997)) and bromfenac (Sheiner (1994)).

1.5 Categorical Data and its Analysis

Categorical data is described in section 2.4.1 and recently there has been considerable interest in techniques appropriate for modelling such data from pharmacodynamic studies. The first paper to focus on the problem of analysing such data was Sheiner (1994) when considering the best way of analysing analgesic drug clinical trials. The paper criticised the idea of simply hypothesis testing the different levels of response between the different dose groups from the number of observations in each category. The proportional odds model (McCullagh (1980)) was proposed by Sheiner (1994) to model the longitudinal aspect of the pharmacodynamic data and quantify the variability between subjects. The use of the proportional odds model for categorical data was not new in the statistical literature but it was for the analysis of pharmacodynamic categorical data. A more mathematical and statistical treatment of the same problem is given in Sheiner *et al* (1997).

The analysis of categorical data (with more than two categories) in terms of regression has only been considered over the past 20 years. One of the first papers to look at the analysis of categorical data was that of Koch *et al* (1977). Since then there has been an increasing amount of research into categorical data analysis. Some of the work that can be found in the literature is by Korn and Whittemore (1979), Anderson (1984), Chuang and Agresti (1986), Armstrong and Sloan (1988), Cox (1988), Conaway (1989), Senn (1991), Kahn and Raftery (1996) and Albert *et al* (1997). The methods used in these papers are wide ranging but all look at the development of models for the description of categorical data.

1.6 Experimental and Optimal Design

As well as studying the data that arises from clinical trials at all phases of drug development, it is necessary that the studies which will produce the data are designed well and appropriately. Before a clinical trial is designed there are certain objectives for the trial. The clinical trial must be designed so that the objectives can be met and then move on and plan the next trial. When one of the criteria of the clinical trial is for example to assess the safety and efficacy of the drug then this will involve the collection of pharmacodynamic data which in certain circumstances will be categorical. There is a considerable amount of work in the statistical literature on designing experiments, for example simple comparisons of treatment and block effects, as well as the optimal design of studies requiring regression analysis (Federov (1977)). In pharmacokinetics, the main requirement is to find the optimal times at which to take plasma concentrations. This can also be the case for the collection of pharmacodynamic data but the optimal design could be for determining the optimal dose to be administered to patients.

1.7 Aims and Objectives

One of the aims of this work is to look at general methods for analysing categorical data obtained from a range of preclinical and clinical studies. The analysis of categorical data has gained more importance in recent years in the analysis of clinical trial data. This has been in part due to the growing interest in the analysis of pharmacodynamic data and the need for models to describe the time profile of the pharmacodynamics and

relate these to the pharmacokinetics of the drug. A consequence of the modelling of categorical pharmacodynamic data is the opportunity to look at comparing the efficacy and toxicity of the compound of interest, as this is one of the main goals of phase III clinical trials. Much of the work on categorical data analysis is given in Agresti (1990) and one of the first papers to look at the analysis of analgesic drug trials where a categorical variable was to be modelled was given by Sheiner *et al* (1997) where the proportional odds model (McCullagh (1980)) was used. As part of the modelling of categorical data, of particular interest is to study different components of variability in the data. This is to be carried out using mixed effects models, and in particular the proportional odds mixed effects model where interindividual variability can be studied. Interindividual variability in drug response is of considerable interest as hopefully it can eventually lead to individualisation of dosage regimens.

As well as considering the description of the data by the proportional odds mixed effects model, it is also of interest to look at different methods of analysing the data. The two methods considered here will be to look at the use of maximum likelihood and Bayesian methods which will utilise the computer packages NONMEM and BUGS respectively.

Finally, the optimal design of clinical trials, where the analysis of categorical pharmacodynamic data is of interest, will be investigated. Optimal design has been considered in the literature for pharmacokinetic data and to a small degree for the analysis of pharmacodynamic data but only for continuous responses. Optimal design has also been studied in the case of logistic regression where the outcome is binary but these ideas have not been extended to categorical data. The objective is to consider how

categorical data should be collected for the aim of determining the parameters as accurately as possible (D-optimality) when the proportional odds model is proposed to model the data.

The plan of this thesis is as follows. Chapter 2 is a review of the methodology that is either used in this thesis or is applicable to the field of population pharmacokinetics/ pharmacodynamics. Chapters 3 and 4 deal with the analysis of toxicokinetic and preclinical trials. Chapter 5 contains work on population modelling of intranasal sumatriptan data obtained from a phase II clinical trial. Chapter 6 is on oxybutynin, a drug for the treatment of urinary incontinence and deals with the analysis of efficacy and adverse effects data collected from a phase III clinical trial. Based on the results of the analysis, it is required to define an optimal dose determined by a general decision analysis procedure. Categorical optimal design is considered in chapter 7 and a general discussion is given in chapter 8.

2 Methods

2.1 Population Pharmacokinetics

Population pharmacokinetics can be divided into two sections: a population structural model and a variance model explaining the variability of the population data around the mean population response. Together, the structural model and the variance model should be able to describe both the population data and the individual response.

The population structural model can be defined in several ways, for example empirically or using compartmental methods. Both methods will be used but generally compartmental methods will be employed. Empirical models are models that describe the data but do not have any mechanistic basis. In most cases, these empirical models will be the sum of exponentials of the form given in equation (2.1) where the A_i's and λ_i's are coefficients to be determined.

$$y(t) = \sum_{i=1}^{c} A_i e^{-\lambda_i t} \qquad (2.1)$$

This empirical model also describes a c compartment model after a bolus dose administration. The coefficients A_i's and λ_i's can be re-expressed as pharmacokinetic parameters which can be interpreted in a practical way whereas the coefficients A_i and λ_i, i=1,…,c cannot be interpreted easily. The compartmental parameterisation involves compartmental volumes of distribution and intercompartmental clearances.

Compartmental models are derived under assumptions about the way the body can be modelled to express the kinetics of the passage of the drug path through the body by a

set of differential equations. In the case of a bolus administration first-order two compartment open model, the concentration of the drug in the two different compartments can be summarised by the system of two differential equations and initial conditions given in equation (2.2).

$$\frac{dC_1}{dt} = k_{21}C_2 - (k_{12} + k_{10})C_1$$

$$\frac{dC_2}{dt} = k_{12}C_1 - k_{21}C_2 \qquad (2.2)$$

$$C_1(0) = \frac{Dose}{V}, C_2(0) = 0$$

$\frac{dC}{dt}$ represents the rate of change of concentration in the compartment over time, the k's represent the first-order rate constants that can be reparameterised into clearance and volume terms and V is the volume of distribution. This system of differential equations can be easily solved and the structural model in equation (2.3) is obtained.

$$C_1(t) = \frac{Dose}{V(\alpha - \beta)} [(k_{21} - \beta)e^{-\beta t} - (k_{21} - \alpha)e^{-\alpha t}]$$

$$\alpha = \frac{1}{2}[(k_{12} + k_{21} + k_{10}) + \sqrt{(k_{12} + k_{21} + k_{10})^2 - 4k_{21}k_{10}}] \qquad (2.3)$$

$$\beta = \frac{1}{2}[(k_{12} + k_{21} + k_{10}) - \sqrt{(k_{12} + k_{21} + k_{10})^2 - 4k_{21}k_{10}}]$$

This system of differential equations assumes that the system acts linearly, i.e. the differential equations are a linear combination of the variables.

The variance model is defined in order to try and explain the variability around the population structural model. As with any model, a functional form can be considered based on logical reasoning of how the variability between subjects and within a subject might occur. Usually the variance function is determined by model fitting. Two common variance functions are those with a constant variance and those with a constant coefficient of variation. These will be discussed in more detail later.

2.2 Population Pharmacodynamics

The ideas behind modelling and analysing population pharmacodynamic data are the same as those for modelling and analysing population pharmacokinetic data. The requirement of defining a structural model and a variance function is still needed but extra steps have to be considered.

The effect of a drug is often thought of as being achieved by a system of linear dynamic processes. The first chain of events is associated with the pharmacokinetics of the compound and this can be described by a series of differential equations describing the absorption, distribution and elimination of the drug as for example in equation (2.2). The concentration of the drug in the plasma can be determined by convolving these linear dynamic processes together and attaining a model for the concentration of the drug in the plasma. Another linear dynamic process can be introduced and that is for the effect of the drug. It is postulated that to obtain an effect, a concentration at an hypothetical 'effect site' must be achieved. This linear dynamic process can then be convolved with that of the pharmacokinetic component to obtain a model for the effect site concentration as a function of time and relevant pharmacokinetic/pharmacodynamic parameters.

Effect site concentrations are in general not available as this would require the collection of some tissue or termination which is not desirable. Instead of correlating effect site concentration to pharmacological effect of the drug, empirical models are most frequently used. The empirical approach involves information obtained previously which in most cases will be the plasma concentration data. The plasma concentration is

then used as the predictor for the pharmacologic activity of the drug. The most common empirical models used for the effect of the drug are the nonlinear E_{max}, sigmoid E_{max} and linear models as given in equation (2.4).

$$E_1(X) = \frac{E_{max}X}{EC_{50} + X} + E_0$$

$$E_2(X) = \frac{E_{max}X^N}{EC_{50}{}^N + X^N} + E_0 \qquad (2.4)$$

$$E_3(X) = \frac{E_{max}}{EC_{50}}X + E_0$$

The E_{max} model (E_1) is a function of a predictor variable, such as plasma concentration or effect site concentration and three pharmacodynamic parameters. E_{max} is the maximum effect that can be achieved (this may be positive or negative), EC_{50} is the concentration that must be achieved to attain 50 % of the maximum effect and E_0 is the baseline effect, i.e. when X is zero. E_0 is not always included in the model as it is often known that this parameter can be set to zero. The sigmoid E_{max} model (E_2) is a generalisation of the E_{max} model as it includes the parameter N which describes steeper (N>1) or shallower (0≤N<1) curves than the E_{max} model (N=1). The linear dynamic model (E_1) can be obtained from the E_{max} model when the data is in the early approximately linear part of the curve. The model can be collapsed so that the gradient is expressed as a fraction, E_{max}/EC_{50}.

These are the most frequently used pharmacodynamic models but more complex models such as multiple ligand models exist, however these will not be considered here. These models are not always appropriate and empirical models may need to be explored based on how well they describe the pharmacodynamic data. Splines or parametric models can be used for this purpose.

2.3 Nonlinear Mixed Effects Modelling

The analysis of repeated measurement data has progressed considerably in the last 20 years. The data that is obtained from pharmacokinetic/pharmacodynamic studies are usually given by the following setting.

A response variable is measured repeatedly in a particular cluster (human, animal or group). The repeated measurements are usually taken over time so the data are longitudinal. The observed response variable is denoted by y and the random variable by Y. There are repeated measurements within the clusters and there are multiple clusters observed so the response data can be denoted by y_{ij} where $i=1,\ldots,I$ indexes the cluster and $j=1,\ldots,n_i$ indexes the number of observations within a cluster. Note that it is not necessary that there be the same number of observations within each cluster. Similarly we can denote the random variable by Y_{ij}. Associated with the response variable is a set of q predictor variables or covariates. The set of covariates are denoted by a q×1 vector $\underline{x}_{ij}$ where the indexing is the same as the above.

To generalise notation slightly, a vector can be specified for each cluster of observed responses and a matrix for each cluster's covariates. This is denoted by an n_i×1 vector $\underline{y}_i$ for the ith cluster's observed responses and a n_i×q matrix X_i for the ith cluster's repeated covariates. A p×1 vector corresponding to the parameters is defined as β. When there is no index on the parameter β, then this corresponds to population parameters but when indexed by i, this corresponds to individual cluster parameters, β_i.

2.3.1 Linear Mixed Effects Models

With the data specified as above, a variety of models can be examined. The simplest example that can be applied is a general linear model which is defined in equation (2.5).

$$E(Y_{ij}) = \underline{\beta}\underline{x}_{ij} \tag{2.5}$$

$E(.)$ represents the expected response based on a linear function of the covariates summed over all observations. In this case, $p=q+1$ where the additional parameter is a constant term. Associated with this model is an error between the observed and expected response. The model can then be written in the form of equation (2.6).

$$y_{ij} = \underline{\beta}\underline{x}_{ij} + \varepsilon_{ij} \tag{2.6}$$

The difference between the observed and expected response is denoted by ε_{ij}. Linear models of this kind have assumptions corresponding to the way in which the errors are distributed. The standard assumption to make about the general linear regression model is that the errors are homoscedastic (additive), independent and identically distributed (iid) with mean zero and standard deviation σ and the covariance between errors is zero. This is denoted by equation (2.7) where N represents the normal distribution.

$$\varepsilon_{ij} \sim N(0,\sigma^2), i = 1,...,I, j = 1,...,n_i \tag{2.7}$$

This model allows the estimation of population average parameters known as fixed effects but does not allow the estimation of individual parameter values which will be referred to as random effects. To obtain both population and individual estimates of the parameters, the general linear regression model must be generalised to incorporate the ability to estimate the individual parameters. The standard form of the linear mixed effects model is given in equation (2.8).

$$\underline{y}_i = X_i\underline{\beta} + Z_i\underline{b}_i + \underline{\varepsilon}_i \tag{2.8}$$

Here, $\underline{y}_i$, X_i and $\underline{\beta}$ are defined as above. Z_i is an $n_i \times r$ subset matrix of X_i corresponding to parameters for which individual estimates are to be obtained. The $r \times 1$ random effect vector $\underline{b}_i$ is the offset from the population parameter estimates so that the individual parameter value is given by $\beta_i = \beta + \underline{b}_i$ assuming that the vectors are of the same dimension. The residual (error) vector $\underline{\varepsilon}_i$ is the difference between the observed vector and the expected vector of responses.

2.3.2 Nonlinear Mixed Effects Models

Linear mixed effects models are a flexible tool for analysing data but in many situations, linear combinations of factors do not describe the data satisfactorily. In pharmacokinetics, the models that usually best describe the data are nonlinear and require more sophisticated techniques. As there are many different nonlinear models, notation must be generalised to include any model that describes the data. In equation (2.9), f denotes a general function with y_{ij}, β, $\underline{x}_{ij}$ and ε_{ij} defined above.

$$y_{ij} = f(\underline{x}_{ij}, \underline{\beta}) + \varepsilon_{ij} \qquad (2.9)$$

In a pharmacokinetic model, f could be a one compartment open model with bolus dose, where dose and time are the independent variables, $\underline{x}_{ij}$ and β is the vector of pharmacokinetic parameters, clearance and volume of distribution. In a pharmacodynamic model, f could be the E_{max} model with concentration being the independent variable and the pharmacodynamic parameters being E_{max} and EC_{50}. The error term in equation (2.9) assumes that the difference between the observed and expected response is the same for all levels of $\underline{x}_{ij}$. The homoscedastic (additive) variance function is not necessarily appropriate for all nonlinear regression models which will be discussed in section (2.3.2).

The nonlinear model in equation (2.9) is formulated in terms of population parameters so no individual effects can be estimated. In population pharmacokinetic/ pharmacodynamic studies it is necessary to be able to estimate individual parameter values, so including random effects into nonlinear models is necessary. The inclusion of random effects can be achieved in a variety of ways. A simple method to include the random effects component would be to add a random effects term into the nonlinear fixed effects component in the same way as for the linear mixed effects model. This is given in equation (2.10).

$$\underline{y}_i = f(X_i, \underline{\beta}) + Z_i \underline{b}_i + \underline{\varepsilon}_i \tag{2.10}$$

The $n_i \times 1$ random effect vector is denoted by $\underline{b}_i$ and is additive on the model. This form of random effects model can be useful as it acts as a shift for each individual from the population average. However, it is also restrictive because it does not allow the estimation of individual parameter estimates. To get round this problem, the random effects can be directly associated with the population parameters. The model can then be defined as in equation (2.11).

$$\underline{y}_i = f(X_i, \underline{\phi}_i) + \underline{\varepsilon}_i$$
$$\underline{\phi}_i = A_i \underline{\beta} + B_i \underline{b}_i \tag{2.11}$$

The nonlinear mixed effects model has now in effect two levels, one corresponding to the individual level and the other corresponding to the population level. The random effect is again additive on the parameter which might not necessarily be appropriate but this will be discussed in section (2.3.2). The parameterisation of nonlinear models as in equation (2.11), unlike linear models, are analytically intractable and therefore must be subject to an analytical approximation, numerical or MCMC method (discussed in section 2.8).

2.3.3 Distributional Assumptions and Error Models

2.3.3.1 Residual Error Models

So far, only the types of structural and variance models that are appropriate for analysing repeated measurement data have been discussed. Choosing the relevant model to fit to the data is only one aspect of any population analysis and understanding the underlying assumptions is important in determining how to interpret the models that are chosen. As well as choosing a structural model for the pharmacokinetic/ pharmacodynamic data, it has already been mentioned that it is necessary to describe how the data varies around the population average. This is carried out by checking the variance models for the inter-subject variability and the residual variability. The inter-subject random effects model describes how each cluster's parameters vary from the population parameters and the residual error model describes how each observation varies from the cluster's expected value. Residual variability represents all the variability that has not been accounted for and can correspond to assay error, intraindividual error, measurement error and so on.

Pharmacokinetic data is continuous data and bounded below by zero. It is often the case that continuous data is assumed to be from a normal distribution with mean $E(Y_{ij})$ and variance $var(Y_{ij})$. In equation (2.6) and (2.7), an additive residual variance function was described for a simple linear regression model and the classic assumptions associated with such an error model were given. In pharmacokinetic data, an additive residual error function is not always appropriate as it might not describe the way the variability is distributed around the population average. A common type of residual

31

error model for pharmacokinetic data is the constant coefficient of variation (C.V.) (proportional error) model defined in equation (2.12).

$$Var(Y_{ij}) = \sigma^2 f(\underline{x}_{ij}, \underline{\beta})^2$$
$$\sigma^2 = C.V.^2$$

(2.12)

In this residual error model, the standard deviation at a particular level of $\underline{x}_{ij}$ varies proportionally to f, i.e. as concentration (y_{ij}) increases then so does the standard deviation but the coefficient of variation remains the same. This is a commonly observed variance function in assays and is referred to as a heteroscedastic residual error model as the variance is not constant over $\underline{x}_{ij}$.

A slightly more complicated error model is a combination of both the additive and proportional error model defined in equation (2.13).

$$Var(Y_{ij}) = \xi^2 + \sigma^2 f(\underline{x}_{ij}, \underline{\beta})^2$$

(2.13)

When the concentrations are high then the proportional error term, σ^2 describes most of the variability but when the concentrations are low then the additive error term, ξ^2 describes the variability as baseline noise.

These are only three possible examples of residual error models for pharmacokinetic data but there are others that can be used, see for example Karlsson *et al* (1995). To generalise notation for residual error models, we can specify a structural model and a variance model as in equation (2.14). The function g is a general variance function, depending on the mean response, the predictor variables and specific variance parameters.

$$E(Y_{ij}) = f(\underline{x}_{ij}, \underline{\beta}_i), \text{var}(Y_{ij}) = \sigma^2 g(\mu_{ij}, \underline{x}_{ij}, \underline{\varphi}_i), \mu_{ij} = E(Y_{ij}) \qquad (2.14)$$

Equation (2.14) specifies the mean and variance of the normal distribution and with this, the likelihood function for the pharmacokinetic data can be obtained.

Another frequently used residual error model involves taking the logarithm of both the model and data so that the model is normally distributed on the log scale as specified in equation (2.15).

$$\log(y_{ij}) = \log(f_{ij}) + \varepsilon_{ij}$$
$$\varepsilon_{ij} \sim N(0, \sigma^2) \qquad (2.15)$$

With this particular model specification, predicted concentrations can not be less than zero, whereas with the additive error model, predicted concentrations can be less than zero. Data of this form are assumed to follow a log-normal distribution and the error model on the original scale is called an exponential residual error model.

2.3.3.2 Interindividual Error Models

As with the residual error models, similar models can be specified for the variability between subjects of particular parameters. In equation (2.11), an additive error model was defined for the parameters. Pharmacokinetic parameters are always defined on a positive range so it is quite normal to use an exponential error model on the pharmacokinetic parameters, e.g. in a one compartment model, the between subject variability would be defined as in equation (2.16). Assuming Cl (clearance) and V (volume of distribution) are constrained to be greater than zero, then all individual values will also be greater than zero, where b_{i1} and b_{i2} are the random effects.

$$Cl_i = Cl \exp(b_{i1}), V_i = V \exp(b_{i2}) \qquad (2.16)$$

With the model specification in equation (2.11), it is normal to make the distributional assumption that $\underline{b}_i$ is distributed normally with mean vector $\underline{0}$ and variance-covariance matrix Ω.

2.4 Generalised Linear Mixed Effects Models (GLMMs)

2.4.1 Non-Continuous Data

In pharmacokinetic studies, the analysis of concentration-time data is the same generally from study to study except for choosing the appropriate models. In pharmacodynamics, a wide range of response variables can be observed which means that the type of analysis for pharmacokinetic data, e.g. assumptions about normality and independent errors, is not necessarily correct for the pharmacodynamic data obtained. As well as continuous measures of drug effect, pharmacodynamic variables can be measured on a categorical, count or censored scale. These types of variables should not be modelled as though they are continuous data and therefore need more general techniques for the data to be analysed. As well as making theoretical decisions on how the data should be analysed, it is still required that the models should tell us something about the pharmacodynamics of the drug and should be able to describe the variability in the data. Much of the work analysing non-continuous data has been published in the case where random effects are not included by McCullagh and Nelder (1989). More recently, work has been published on generalised linear mixed effects models for both maximum likelihood and Bayesian techniques (Zeger and Karim (1991), Booth and Hobert

(1999)). For more information on analysing non-continuous data, see McCullagh and Nelder (1989).

Categorical data is one type of non-continuous data and is the main emphasis in this thesis. The definition of a categorical variable is where there is a discrete number of possible outcomes that can be chosen (greater than 1). An example of categorical data is where an event can take only two possible courses: the event happens or does not happen. This situation could be encountered by observing some toxic response when a particular dose of a drug is given or whether some predefined level of drug effectiveness has been reached after a certain time. An example of a greater than two categorical response is in analgesic clinical trials where pain relief is measured on a 5 point scale where 1=no pain relief, 2=little pain relief, 3=moderate relief, 4=lot of relief and 5=complete pain relief.

There is an important distinction within categorical data about the different types of data that can be observed. The two most important types of categorical data are nominal and ordinal data. Nominal data is where there is no natural ordering of the categories. A simple example of this is gender (male/female) where it is not of importance which order the categories are but a category is chosen as a reference point from which to compare other categories. Another example with more than two categories is race (European, Asian or African). This is the type of categorical response that is usually associated with predictor variables rather than response variables although nominal categorical responses can be observed. Ordinal data is where there is a particular ordering to the categories which is more informative than nominal categorical data. An example is the pain relief response described above or in determining levels of toxicity

(none, low, medium, high). Nominal data and ordinal data are not analysed in the same way and therefore need careful consideration. When the data are dichotomous (two categories), the nominal and ordinal data can be treated in the same way. The work of this thesis will only involve ordinal categorical response variables.

2.4.2 Generalised Linear Models

In this section, it will be assumed that the data is from the following design. A response variable has been independently measured n times so that there is a set of response data $y_1, ..., y_n$. Associated with these response data is a $q \times 1$ vector of covariates denoted by $\underline{x}_i$, $i=1,...,n$.

To generalise from classical linear models, we can specify three components to the model:

1) The random component: the components of the random variable $\underline{Y}$ have independent normal distributions with $E(\underline{Y})=\underline{\mu}$ and constant variance σ^2.

2) The systematic component: covariates x_{ij}, $j=1,...,q$ produce a linear predictor η given by equation (2.17).

$$\eta_i = \sum_{j=1}^{q} x_{ij} \beta_j \qquad (2.17)$$

3) The link between the random and systematic components is given in equation (2.18).

$$\eta_i = g(\mu_i) \qquad (2.18)$$

2.4.2.1 Likelihood Functions for Generalised Linear Models

Assume that the random variable $\underline{Y}$ is from the exponential family of the form

$$L(\theta,\phi;y_i) = \exp\{(y_i\theta - b(\theta))/a(\phi) + c(y_i,\phi)\} \qquad (2.19)$$

for some specific functions $a(.)$, $b(.)$ and $c(.)$. For example, for the normal likelihood, $\theta=\mu$, $\phi=\sigma^2$, $a(\phi)=\phi$, $b(\theta)=\theta^2/2$ and $c(y_i,\phi)=-1/2\{y_i^2/\sigma^2+\log(2\pi\sigma^2)\}$. The mean and variance of the exponential family can be easily derived and is given by equation (2.20).

$$E(Y_i) = \frac{db(\theta)}{d\theta}$$

$$\qquad (2.20)$$

$$\mathrm{var}(Y_i) = \frac{d^2 b(\theta)}{d\theta^2}a(\phi)$$

Quite often, ϕ (known as the dispersion parameter) is a scalar so the variance function is only a function of the mean. For the normal distribution, the dispersion parameter is simply the variance σ^2.

2.4.2.2 Link Functions

The link function relates the linear transformed predictor η to the expected response value μ. In classical linear regression, there is no need to think about a link function because the models are already linear and for normally distributed data, the link function is the identity link. If the data is assumed to be another member of the exponential family then a link function needs to be considered other than the identity link function. If, for example, the data is on the domain [0,1], i.e. binary or binomial then a link function that transforms this scale onto the whole real line is required. The link function has the effect of transforming non-normal data onto a continuous range

and enabling the assumptions of independent, additive and uncorrelated errors to be made. These are discussed in more detail in later sections with specific examples.

2.4.2.3 Binary Data and Logistic Regression

One of the most frequently seen types of data that is non-normal is binary or binomial data so there is a considerable amount of work in the literature dedicated to the analysis of binary data. Collett (1991) is an excellent text on the basic ideas of binary data analysis.

Continuing with the data arrangement specified in section 2.4.2, it is now assumed that the response data are a set of 0's and 1's with an associated set of covariates. The standard likelihood assumption for binary data is to employ the Bernoulli distribution defined in equation (2.21.

$$L(\theta \mid y_i) = \theta^{y_i} (1 - \theta)^{(1 - y_i)} \tag{2.21}$$

The parameter θ is the probability of an event occurring, i.e. $\Pr(Y_i=1)$ where 1 represents an event occurring and 0 is the event not occurring. The log-likelihood is given by equation (2.22) for the set of observations:

$$l(\theta \mid \underline{y}) = \log(L(\theta \mid \underline{y})) = \sum_{i=1}^{n} \{y_i \log\left[\frac{\theta}{1-\theta}\right] + \log(1 - \theta)\} \tag{2.22}$$

This log-likelihood can be re-expressed so that it is in the form of the logarithm of the exponential family and the parameters and functions are then given by equation (2.23).

$$\phi = 1$$
$$b(\theta_i) = \log(1 + e^{\theta_i})$$
$$c(y_i, \phi) = 1 \qquad\qquad (2.23)$$
$$\mu(\theta_i) = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$$
$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$$

The link function for binary or binomial data is known as the logit link function and is the natural link function for binary data. The logit link function is easily interpretable as the log odds ratio of an event occurring to an event not occurring. Other link functions are used for binary data such as the probit, complementary log-log, log-log and arcsin link functions but it is usual to use the logit link function. The probit link function is based on the inverse of the standard normal cumulative density function which makes the assumption that the errors between the observed and expected will be normally distributed on the probit scale. Results obtained from using the logit and probit link functions are normally very similar and there is a scaling factor of approximately 1.6-2 in the difference of parameter estimates if both the link functions work well (Collett (1991)). The probit and logit links are symmetric but the complementary log-log and log-log link functions are not. There are tests for assessing goodness of link (Pregibon (1980)) but it is generally assumed that the relevant link function is known.

When binary data are modelled as a generalised linear regression model then the resulting model is called a logistic regression model. The model is often specified on the logit link function transformed scale or the probability scale as in equation (2.24).

$$\log\left(\frac{E(Y_i)}{1-E(Y_i)}\right) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \sum_{j=1}^{q} x_{ij}\beta_j$$

$$E(Y_i) = \theta_i = \frac{\exp(\sum_{j=1}^{q} x_{ij}\beta_j)}{1 + \exp(\sum_{j=1}^{q} x_{ij}\beta_j)} \qquad (2.24)$$

The first part in (2.24) gives the logit transformed model on a linear scale and the second part is the inverse of the logit transformation which is on the probability scale and is known as the logistic function.

2.4.3 Fixed Effects (Marginal) Models

Returning to the situation where the data is as described in section 2.3, deciding on how the data is to be analysed is not as straightforward as in the population pharmacokinetic data scenario. Whether it is the population average model that is required or obtaining individual fits to the data, different approaches to the analysis can lead to different interpretations of the regression coefficients (Diggle *et al.* (1994)).

If the data is analysed as though there is no distinction between individuals then this corresponds to the case in population pharmacokinetics of the so-called naïve pooling method (Steimer *et al* (1985)). This method does not allow the inclusion of between-subject random effects in the model as there is no way of knowing from which individual the data came. This can also lead to bias of population parameter estimates as individual data are not influencing the population parameter estimates in necessarily the correct way. Even though this is not necessarily the best way of analysing repeated measurement data, it acts as a useful starting point.

Another possible explanation of marginal models is where between-subject random effects have been integrated out of the likelihood (Diggle *et al* (1994)). Although the resulting model does not include individual random effects, the population parameter estimates have an interpretation that would be more in line with that of population pharmacokinetic models.

2.4.4 Mixed Effects Models

Mixed effects models enable estimation of both the individual and population parameter values of a set of repeated measurement data. Assuming the generalised linear mixed effects (GLMM) model is defined as in equation (2.25), then the population parameter model can be found by integrating out the random effects leaving just the population parameters. The distribution F is the random effects distribution.

$$
\begin{aligned}
E(\underline{Y}_i \mid \underline{b}_i) &= g^{-1}(X_i \underline{\beta} + Z_i \underline{b}_i) \\
E(\underline{Y}_i) &= \int g^{-1}(X_i \underline{\beta} + Z_i \underline{b}_i) dF(\underline{b}_i)
\end{aligned}
\tag{2.25}
$$

For categorical data, the model parameters from marginal models and random effects models are interpreted differently (Diggle *et al* (1994)). For the marginal model, the parameters describe the ratio of population odds and for the random effects model, the ratio of individual odds. It has also been noted that the absolute value for marginal models is smaller than the parameter estimates for random effects models.

2.4.5 Transition (Markov) Models

Transition (Markov) models, are used when the current value of the response variable explicitly depends on the previous response values. The previous response variables

can be treated as though they are predictors for the current response. A transition model can be specified as in equation (2.26).

$$g(E(Y_{ij} \mid Y_{ij-1}, Y_{ij-2}, ..., Y_{i1})) = \underline{x}_{ij} \underline{\beta} + \alpha y_{ij-1} \tag{2.26}$$

The expected response (left hand side) is explicitly shown to depend on the previous response measurements but the model (right hand side) shows that there is only dependence on the previous response. The model can depend on any number of previous responses but this can be tested to see how much these predictors can describe the data. One important point is that the spacing (temporal or spatial) between the response measurements must be the same or the fitting of transition models can become complicated.

Another way of thinking about transition models is in terms of Markov chains. Take, for example a binary response, then the probability of being in a particular state/category (event occurred or did not occur) depends on the previous response observed by the Markov chain in equation (2.27). The previous state/category which is known is given by the row and the probability of 'jumping' to the next state/category is given by the probabilities in the cells of the matrix.

$$
\begin{array}{c}
\phantom{y_{i,j-1}}\quad y_{ij} \\
\phantom{y_{i,j-1}}\quad 0 \qquad\qquad\qquad 1 \\
y_{i,j-1}\;\begin{array}{c} 0 \\ 1 \end{array}
\begin{bmatrix}
\dfrac{1}{1 + \exp(\underline{x}_{ij}\underline{\beta})} & \dfrac{\exp(\underline{x}_{ij}\underline{\beta})}{1 + \exp(\underline{x}_{ij}\underline{\beta} + \alpha)} \\[3mm]
\dfrac{1}{1 + \exp(\underline{x}_{ij}\underline{\beta} + \alpha)} & \dfrac{\exp(\underline{x}_{ij}\underline{\beta} + \alpha)}{1 + \exp(\underline{x}_{ij}\underline{\beta} + \alpha)}
\end{bmatrix}
\end{array}
\tag{2.27}
$$

When the event occurred (y_{ij-1}=1), the model includes the parameter α, so as to indicate y_{ij-1}=1. As the matrix also depends on covariates, then the transition probabilities can vary across individuals.

2.4.6 Proportional Odds Model

Having discussed some of the models that are available in the general setting of generalised linear mixed effects models, the main purpose in this thesis is the analysis of categorical data in pharmacokinetic/pharmacodynamic studies. So far, binary data has been briefly examined but it is necessary to have a general framework in which to analyse any data of a categorical nature if the analyses are not going to be arbitrary. Agresti (1990) and McCullagh and Nelder (1989) report most of the theoretical work. One of the most important early papers on the proportional odds model is by McCullagh (1980).

Assuming that the data are of the same arrangement as that specified in section 2.3, we can make the additional assumption that the response variable is categorical such that Y_{ij} can only take the values $\{1,2,...,C\}$. As these numbers are only representations of the categories and represent the actual response qualitatively, it seems sensible to use another measure that can represent the response quantitatively. The easiest measures to use are probabilities and log odds ratios. These are used in logistic regression and allow the extension of methods to categorical responses of more than two categories. In logistic regression, the probability of being in a particular category is used as the measure we are trying to model, so a similar strategy can be used for polychotomous data. As the data are known to be ordered, then it is reasonable to use the ordering of

the categories as additional piece of information. This can be achieved in several ways by defining the odds ratios in specific ways. Firstly, we need to specify that the data can be described in terms of cumulative categories and cumulative probabilities as given in equation (2.28). R and γ denote the cumulative category observation and cumulative probabilities respectively.

$$
\begin{aligned}
R_{ij1} &= Y_{ij1} \\
R_{ijk} &= Y_{ij1} + \ldots Y_{ijk}, k = 2,\ldots,C-1 \\
R_{ijC} &= 1 \\
\gamma_{ij1} &= \Pr(R_{ij1} = 1) = \Pr(Y_{ij1} = 1) \\
\gamma_{ijk} &= \Pr(R_{ijk} = 1) = \Pr(Y_{ij1}) + \ldots + \Pr(Y_{ijk}), k = 2,\ldots,C-1 \\
\gamma_{ijC} &= \Pr(R_{ijC} = 1) = 1
\end{aligned}
\tag{2.28}
$$

With this parameterisation, the odds ratio can be defined for which the modelling is carried out. A commonly defined odds ratio used in ordinal categorical data is the continuation odds ratio defined as in equation (2.29) where OR denotes odds ratio.

$$
OR_{ijk} = \frac{\gamma_{ijk+1} - \gamma_{ijk}}{\gamma_{ijk}}, k = 1,\ldots,C-1
\tag{2.29}
$$

The continuation odds ratio uses the ordering of the categories but does not include all the information available as all categories from $k+2$ to C are ignored. Another odds ratio that uses all the information available is the cumulative odds ratio defined in equation (2.30).

$$
OR_{ijk} = \frac{\gamma_{ijk}}{1 - \gamma_{ijk}}, k = 1,\ldots,C-1
\tag{2.30}
$$

This odds ratio in effect uses a dichotomisation of all the ordered categories. This is similar to constraining the categories to be in one of two categories which then can be treated like binary data and so logistic regression can be used. For C categories, there are C-1 cumulative odds ratios (or cumulative probabilities on the probability scale), so there are C-1 logistic regression curves. The proportional odds model which is in effect

a constrained model of C-1 logistic models where each model has the same covariate parameter values but the intercepts or 'cut points' are allowed to vary as defined in equation (2.31).

$$\log it(\Pr(Y_{ij} \le k)) = \log\left(\frac{\Pr(Y_{ij} \le k)}{\Pr(Y_{ij} > k)}\right) = \theta_k + \underline{x}_{ij}\underline{\beta}, k = 1,...,C-1 \quad (2.31)$$

In equation (2.31), the intercept component is specified by θ_k. The parameters θ_k, $k=1,...,C$-1 represent the cumulative probabilities when no covariates are measured. The reason for the name of proportional odds model is because the odds ratio between any two specified covariate values is independent of what category is observed.

Another way of thinking about the proportional odds models is that there is an underlying latent variable which can only be observed as a particular number of categories. The cut points of the proportional odds model refer to the dividing lines of the continuous latent response but between the cut points, the only information available is from the covariates and how these influence the probability of lying between the cut points. The idea of latent variables is how Albert and Chib (1993) deal with analysing categorical data in a Bayesian setting.

As with generalised linear models, random effects can be included in the model to allow for subject specific parameter values. The mixed effects model can be written in the same way as for generalised linear mixed effects models and the model of the form given in equation (2.32) would be obtained.

$$\log it(\Pr(Y_{ij} \le k \mid \underline{b}_i)) = \theta_k + \underline{x}_{ij}\underline{\beta} + \underline{z}_{ij}\underline{b}_i, k = 1,...,C-1 \quad (2.32)$$

Although the main effects vector $\underline{\beta}$ does not include an intercept term, the random effects vector $\underline{z}_{ij}$ can include a random effect on the cut points.

The distribution for categorical data is usually specified to be the multinomial distribution (McCullagh (1980)). Depending on how the data are collected, depends on what distributional assumption is made but typically it is either multinomial, independent multinomial or poisson (Agresti (1990)). As the proportional odds model is in terms of cumulative probabilities, the multinomial used in later analyses is a reparameterisation of the multinomial distribution in terms of cumulative probabilities. This is given in equation (2.33) where R is the set of cumulative categories and H is the set of cumulative probabilities.

$$L(R \mid H) = \frac{n!}{R_{1i}! R_{1i}! \dots R_{ki}!} \prod_{k=1}^{C-1} \left\{ \frac{\gamma_{ki}}{\gamma_{(k+1)i}} \right\}^{R_{ki}} \left\{ \frac{\gamma_{(k+1)i} - \gamma_{ki}}{\gamma_{(k+1)i}} \right\}^{R_{(k+1)i} - R_{ki}} \qquad (2.33)$$

2.5 Generalised Nonlinear Mixed Effects Models

So far, for data from the exponential family, only linear models have been considered. Analogous to the situation with continuous data, a linear model on the link transformed scale might not be adequate in describing the relationship between the predictors and the response variable. An example from Sheiner (1994), Sheiner, *et al* (1997) and Mandema and Stanski (1997) use the E_{max} model on the logit transformed scale to model the pain relief scores from analgesic drug clinical trials. With this model on the logit transformed scale, there is an upper limit to what the probability of being in a cumulative set of categories as the E_{max} asymptotes to a maximum value so the probability of being in the particular cumulative set of categories can never reach 1. This might be a reasonable assumption in some situations where it is known that there is never a certain chance of being in a particular category.

2.6 Modelling Methods

The models specified in the above sections all need to be fitted to data to find out the parameter estimates for each particular model. As well as there being many different models with which to explain the data, there are a variety of methods that can be used for the actual model fitting process. Different methods of model fitting are more appropriate under different situations, for different types of data or for different models. Two distinct types of analyses, classical and Bayesian will be taken in future data analyses. The first general ideology, frequentist or classical statistics is where the parameter estimates are based solely on the data and specified assumptions. Classical modelling techniques are least squares methods, maximum likelihood and nonparametric maximum likelihood methods. The second ideology is where both the data and prior information about aspects of the model such as the parameter estimates are mathematically included in the model fitting. This is known as Bayesian statistics. When there are not any strong prior beliefs then the results of Bayesian method converge to those of frequentist methods as all information arises from the data.

The next few sections describe some of the frequentist methods that are used commonly. The Bayesian methods are left to section 2.7.

2.6.1 Weighted Least Squares

Weighted least squares (WLS) can be used for the fitting of any function to a set of data. The idea is to minimise an objective function which will then give the best fit of the model to the data based on the criteria defined. The weighted least squares objective

function is defined by equation (2.34), where the indexing in j is ignored for the moment.

$$Q_{WLS} = \sum_i w_i (y_i - f(\underline{x}_i, \underline{\beta}))^2 \qquad (2.34)$$

In addition to the model f and the data y_i, there is associated with each data point a weight w_i. This weight has to be specified before the analysis then examined afterwards through weighted residual plots to see if it is an adequate weighting scheme. In the situation of pharmacokinetic data with a constant coefficient of variation residual error model, the weights would be the inverse of the variance for each data point. Therefore the weights would be the reciprocal of the square of the model prediction multiplied by the standard deviation. The standard deviation is estimated after the regression parameters have been estimated making the interpretation of the standard deviation straight forward as being a measure of everything that has not been explained by the model and the chosen weighting scheme.

When no weighting scheme is used in the weighted least squares objective function then this reduces to ordinary least squares (OLS).

2.6.2 Extended Least Squares

Extended least squares is a generalisation of weighted least squares in that it allows the estimation of parameters associated with the weighting scheme (variance function) which is not possible in WLS or OLS (Peck *et al* (1984)). By choosing a function for the variance that will allow a variety of ways of describing the variation around the population average, the choice of a weighting scheme should no longer be a problem. However things are not normally as simple as this and models usually have to be

48

refitted with different variance functions. The extended least squares objective function is defined by equation (2.35).

$$Q_{ELS} = \sum_i \left(\frac{(y_i - f(\underline{x}_i, \underline{\beta}))^2}{\sigma^2 g(\underline{x}_i, \underline{\beta}, \underline{\varphi})} + \log(\sigma^2 g(\underline{x}_i, \underline{\beta}, \underline{\varphi})) \right) \qquad (2.35)$$

The extra term log(.) acts as a penalty so that the variance does not become too big in the first part of the objective function.

The extended least squares objective function is implemented in the program NONMEM which will be described in section (2.6.6).

2.6.3   Generalised Least Squares

In weighted least squares the weight function is fixed and assumed known whereas in extended least squares, the parameters of the (variance) weight function can be estimated but the structure can not. In generalised least squares, the variance function can be estimated as in extended least squares but it is not done simultaneously. The algorithm for generalised least squares is that the model parameters and variance parameters are estimated at separate steps. The procedure is as follows:

1. First fit the data using ordinary least squares and obtain the parameter estimates $\underline{\beta}^{(0)}$.

2. Use residuals from the preliminary fit to estimate σ^2 and φ. Create estimated weights based on the estimates of the variance parameters to form the weight function described in equation (2.36).

$$\hat{w}_i = g^{-1}(\underline{x}_i, \underline{\hat{\beta}}^{(0)}, \hat{\varphi}) \qquad (2.36)$$

3. Using the estimated weights from step 2, re-estimate β by minimising the weighted least squares objective function. Treating the new estimates as preliminary estimates return to step 2.

Generalised least squares is a robust method of estimating the parameters of the structural model and the variance model and has been recommended for analysing pharmacokinetic data (Davidian and Giltinan (1995)). Despite this extended least squares is often the choice of analysis as it is implemented in commonly available software, e.g. NONMEM (Beal and Sheiner (1992)). Extended least squares is adequate in most cases but has been shown to have some undesirable statistical properties when the wrong variance function is assumed (Houwelingen (1988)).

2.6.4 Maximum Likelihood

The method of estimation by maximum likelihood is a widely used method in parameter estimation problems in general but does not tend to get used in pharmacokinetic analyses. The reason for this is partly due to the development of the NONMEM package at an early stage in the development of population pharmacokinetic methods. Maximum likelihood theory is based on distributional assumptions about the data and then the likelihood is maximised with respect to the model parameters. When the mean and variance function are specified, the results obtained by maximum likelihood techniques are the same as those obtained from extended least squares as the extended least squares objective function has the same structure as the log-likelihood for normally distributed data. Generalised least squares has been shown to be more robust to variance component estimation than maximum likelihood but restricted maximum

likelihood improves estimation of variance components. More information on maximum likelihood and least squares estimation can be found in Searle *et al* (1992).

2.6.5 Nonparametric Methods

The nonparametric maximum likelihood approach puts no restriction on the population distribution of the parameters (Racine-Poon and Wakefield (1998)). A structural model is still defined but no assumptions about how the individual parameters may vary across the population are specified. Mallet (1986) introduced the theory for NPML and showed that the discrete distribution obtained can be related to D-optimal design theory. This method requires the full specification of the residual error model which if specified incorrectly will lead to biases in the parameter estimates. Also, no standard errors of the parameter estimates are obtained. Schumitzky (1991) proposes a similar method using the EM (Expectation-Maximisation) algorithm called NPEM. Other nonparametric programs and methods are briefly described in Aarons (1999).

2.6.6 NONMEM

Because NONMEM is such a widely used program in the analysis of pharmacokinetic and pharmacodynamic data, it is important to understand some of the methods it employs. The data specification is the same as in section 2.3.

The population likelihood is specified as in equation (2.37).

$$L(\underline{\beta}, \Sigma, \sigma^2) = \prod_{i=1}^{I} \int p_1(\underline{y}_i \mid \underline{\beta}, \underline{b}_i, \sigma^2) p_2(\underline{b}_i \mid \Sigma) d\underline{b}_i \tag{2.37}$$

The two distributions, p_1 (individual likelihood) and p_2 (random effects distribution) are usually specified to be normal or log-normal. In the case where the distributions are

normal and the model is linear, the integral in (2.37) is analytically tractable but if the model is nonlinear as is usually the case with pharmacokinetic data then the integral can not be solved analytically. With the linear case, the integral is tractable but the resulting equation can not be analytically maximised to obtain the parameter estimates so a numerical technique is required such as the Newton-Raphson algorithm (Lindstrom and Bates (1988)) to solve for the parameters. With a nonlinear model, before any maximisation takes place, an approximation for the intractable integral must take place. NONMEM uses an analytical approximation to the integral of which there are several. Once the approximation is carried out then the objective function that is numerically minimised is the objective function given in equation (2.38) where V_i is the variance function and J is the matrix of first partial derivatives of size ($p \times n$).

$$
\begin{aligned}
Q_{NONMEM} &= \sum_{i=1}^{I} \{ (\underline{y}_i - f(X_i, \underline{\beta}_i))^T V_i^{-1} (\underline{y}_i - f(X_i, \underline{\beta}_i)) + \log(V_i) \} \\
&= \sum_{i=1} \{ (\underline{y}_i - f(X_i, \underline{\beta}_i))^T V_i^{-1} (\underline{y}_i - f(X_i, \underline{\beta}_i)) + \log(V_i) \}
\end{aligned}
\tag{2.38}
$$

The objective function is numerically minimised with respect to the population parameters by the use of a derivative free quasi-Newton type algorithm.

The analytical approximations to the likelihood are based on Taylor series expansions about the random effects, $\underline{b}_i$. The most widely used approximations are first-order and are known as the First Order (FO) and First Order Conditional Estimation (FOCE) methods. The FOCE method is an expansion around the current estimate of $\underline{b}_i$ obtained at each iteration of the algorithm. It is called "conditional estimation" because the estimate of $\underline{b}_i$ is derived conditional on the current estimates of the random effects. The FO method is a simplification of the FOCE method with the $\underline{b}_i$ are set to zero. There exists a second-order approximation called the Laplacian method (Tierney and Kadane

(1986)) and is essentially the same as the first-order methods but gives a better approximation to the likelihood.

Th latest version of NONMEM (Version V) also allows the specification of a user defined likelihood. This is useful in the modelling of non-continuous or non-normal data as the approximations and assumptions for normally distributed data will not necessarily be correct. Any approximations are then taken from the user defined likelihood and will give better estimates. In this case, a residual error term is not estimated as this acts as a switch between whether the data are to be analysed using in-built distributions or a user defined likelihood.

2.7 Bayesian Statistics

The methods so far described are based solely on the data and assumptions about the way in which the data are distributed. Sometimes, the assumptions made with classical methods are not fully realised and it is important to understand the constraints being made as this affects the way the data will be interpreted. As well as making the standard assumptions about the way the data are distributed and modelled, Bayesian statistics uses an extra level of 'assumptions' to specify the beliefs about aspects of the data, the model or the variance and so on. These prior beliefs are included in the analysis and have an effect on the results. Depending on how strong these beliefs are will determine how much influence they have on the final parameter estimates. In the case where there is little information in the prior distribution then Bayesian results converge to those of classical likelihood methods. These prior beliefs are formulated in

terms of a prior distribution, the information supplied by the data is given in the likelihood and the combination of these two sets of information gives the posterior distribution. These three distributions are linked by Bayes' theorem as specified in equation (2.39) where $p(.|.)$ is a conditional distribution and $p(.)$ is a marginal distribution.

$$p(X \mid Y) = \frac{p(Y \mid X)p(X)}{\int p(Y \mid X)p(X)} = \frac{p(Y \mid X)p(X)}{p(Y)} \qquad (2.39)$$

This is an equation of conditional probability where $p(X|Y)$ is the posterior distribution, $p(Y|X)$ is the likelihood and $p(X)$ is the prior distribution.

Since the introduction of Monte Carlo Markov Chain (MCMC) techniques, Bayesian methodology has been used more and more and there has been an increase in the use of Bayesian methods for the analysis of population pharmacokinetic and pharmacodynamic data, see for example Wakefield (1996), Lunn and Aarons (1998) and Racine-Poon and Wakefield (1995).

2.7.1 Prior, Likelihood and Posterior

The main feature of a Bayesian analysis that distinguishes it from classical analyses is the use of prior distributions. The type of distribution before the use of computers was a particular concern because due to the intractability of the integral in the denominator of equation (2.39). Without an easy way of obtaining this integral, Bayesian methods were seen as very difficult and only specific types of distributions were used with specific likelihood functions. For example, in linear models a normal likelihood and a normal prior results in a normal posterior distribution. These types of distributions are known as conjugate distributions as the prior is of the same form of distribution as the posterior

given a particular likelihood. Other prior distributions that resulted in intractable integrals were not considered. With MCMC methods, there are no longer problems in what prior distributions are chosen, as the posterior will generally only be reported as a summary of the first two moments even though it is more informative to display the complete distribution.

The prior will primarily contain information about the data and model before the data set to be analysed is actually known. If nothing is known before the data are collected then low information priors can be specified. If information is available about certain parameters before a set of data is modelled then this can be included in the prior with low variability to show that the value is *a priori* known to a reasonable degree.

The likelihood is part of the modelling process that does not change from Bayesian to classical analyses. The most frequently used likelihood in a parametric setting for pharmacokinetic data is normal or log-normal. The only time this changes is in a nonparametric analysis where the likelihood is not defined explicitly.

The combination of the prior distribution and the likelihood via Bayes' theorem gives the posterior distribution. Using conjugate analysis then the posterior might be of a known form but more likely, the posterior will not be a standard distribution. In nonlinear pharmacokinetic models, the posterior distribution is not of a known form but often asymptotic theory is used to approximate the posterior by a normal distribution. In pharmacodynamic studies where the data are non-normal then a complicated posterior distribution will result unless a conjugate prior is used. Again, this is not a

problem as the distribution is often summarised by the first two moments. A comprehensive text on Bayesian theory is that by Bernardo and Smith (1994).

2.7.2 Hierarchical Models

Data arising from a population study falls naturally into the setting of a hierarchical model. A hierarchical model defines multiple levels of distributions at which each level, the parameters depend on the next level of distributions. This is particularly applicable in population pharmacokinetic and pharmacodynamic settings where the goal is to determine population and individual parameters and variability. The hierarchies in a Bayesian analysis are referred to as stages. The first stage is the level at which the individual models and parameters are defined, moving to the population level and then to the level of prior distributions. A standard hierarchical setting will have three stages but can have as many as desired, for example in a data set with interoccasion measurements, then four stages are used (Lunn and Aarons (1998)). Hierarchical models are also used in classical analysis and Lee and Nelder (1996) have developed a method for hierarchical generalised linear models using h-likelihoods.

The next three sections will define the three stages using a hierarchical pharmacokinetic model as an example. Details of this analysis can be found in Wakefield (1996) and Davidian and Giltinan (1995). A similar strategy could be used for generalised linear mixed effects models for pharmacodynamic data but using the relevant distributions in the particular case.

2.7.2.1 Stage 1

Assume that the data is as defined in section 2.3.

At the first stage of the hierarchical model, the probability distribution for the individual is defined in equation (2.40).

$$y_{ij} \mid \underline{\beta}_i, \tau \sim N(f(\underline{x}_{ij}, \underline{\beta}_i), \tau^{-1} g(\underline{x}_{ij}, \mu_{ij}, \underline{\varphi}_i))$$
$$\tau^{-1} = \sigma^2, \mu_{ij} = f(\underline{x}_{ij}, \underline{\beta}_i)$$

(2.40)

The reason for parameterising in terms of τ instead of σ^2 is because the inverse of the variance when distributed as a gamma distribution is a conjugate distribution for a normal likelihood. The assumption that the data are independent allows the likelihood to be defined as in equation (2.41). Assume for the moment that there are no variance parameters φ_i to be estimated.

$$y \mid \beta, \tau \sim \prod_{i=1}^{I} \prod_{j=1}^{n_i} N(f(\underline{x}_{ij}, \underline{\beta}_i), \tau^{-1} g(\underline{x}_{ij}, \mu_{ij}))$$
$$y \mid \beta, \tau \sim \prod_{i=1}^{I} \prod_{j=1}^{n_i} N(f(\underline{x}_{ij}, \underline{\beta}_i), \tau^{-1} g(\underline{x}_{ij}, \mu_{ij}))$$

(2.41)

The likelihood for the $\quad y = (\underline{y}_1, ..., \underline{y}_I), \beta = (\underline{\beta}_1, ..., \underline{\beta}_I) \quad$ onsideration of how the individual parameters are distributed in the population can be addressed.

2.7.2.2 Stage 2

The individual parameters are often assumed to arise from a multivariate normal distribution. Whether it is the parameter or some transformation of the parameters that are normally distributed is not a problem as it is usual to take a log transformation to ensure positivity of the parameters and also other constraints to ensure identifiability such as ordering the rate constants (Wakefield (1995)).

$$\underline{\beta}_i \mid \underline{\beta}, \Sigma^{-1} \sim N(\underline{\beta}, \Sigma^{-1}) \qquad\qquad (2.42)$$

The individual parameter vectors are specified as being realisations from a population with a common mean $\underline{\beta}$ and precision matrix Σ^{-1} (inverse of the variance-covariance matrix). At this level, covariate models on the parameters can be introduced. The variance-covariance matrix, Σ represents the interindividual variability of the population parameters or their transformations.

Another approach for the second stage probability distribution is to use a Student t distribution. This distribution allows for the inclusion of outliers as it is heavier in the tails and does not give such low probabilities to these observations. The model would then be specified as in equation (2.43). The degrees of freedom v are usually set to a low value, e.g. $v=4$ as this corresponds to a heavy tailed distribution.

$$\underline{\beta}_i \mid \underline{\beta}, \Sigma^{-1} \sim St_v(\underline{\beta}, \Sigma^{-1}) \qquad\qquad (2.43)$$

To complete the model specification for the second stage, a probability distribution is specified for the precision parameter τ. This distribution is usually taken to be gamma as defined by equation (2.44).

$$\tau \mid v_0, \tau_0 \sim G(0.5 v_0, 0.5 v_0 \tau_0) \qquad\qquad (2.44)$$

The values are usually set at this stage for the precision parameter distribution and the parameters specifying the distribution are given values that specify a low information prior.

2.7.2.3 Stage 3

The third stage of the hierarchical model defines the prior distributions on the population parameters. The standard distributional assumptions to make are typically conjugate, although this is not a requirement and are given in equation (2.45).

$$\underline{\beta} \sim N(\underline{\eta}, C), \Sigma^{-1} \sim W((\rho R)^{-1}, \rho) \tag{2.45}$$

$W(.,.)$ denotes a Wishart distribution which is a multivariate gamma distribution. It is at this stage that the distribution is fully specified to give prior information on how much is known about the parameters. It is usually the case that little information is known so the values of $\underline{\eta}$, C, ρ and R are chosen so that they reflect this lack of information. The value of $\underline{\eta}$ is set to the initial estimate of $\underline{\beta}$ and C is often large as to make the distribution flat and of low information. R is chosen to be the initial estimate of Σ with ρ being set equal to the dimension of the matrix.

2.7.2.4 Joint Posterior Model

To obtain the joint distribution of the parameters given the data, Bayes' theorem must be used. The posterior distribution is given in equation (2.46).

$$p(\underline{\beta}, \tau, \underline{\eta}, \Sigma^{-1} \mid \underline{y}) = \frac{p(\underline{y} \mid \underline{\beta}, \tau, \underline{\eta}, \Sigma^{-1}) p(\underline{\beta}, \tau, \underline{\eta}, \Sigma^{-1})}{p(\underline{y})} \tag{2.46}$$

The second term in the numerator is a combination of the second and third stages of the hierarchical model which can be rewritten to show this as equation (2.47).

$$p(\underline{\beta}, \tau, \underline{\eta}, \Sigma^{-1} \mid \underline{y}) = \frac{p(\underline{y} \mid \underline{\beta}, \tau) p(\underline{\beta}, \tau \mid \underline{\eta}, \Sigma^{-1}) p(\underline{\eta}, \Sigma^{-1})}{p(\underline{y})} \tag{2.47}$$

Conditional on β and τ, the distribution of $\underline{y}$ is independent of η and Σ^{-1}. Lastly the assumption is made that β and τ are conditionally independent and so are each of the prior distributions giving equation (2.48).

$$p(\underline{\beta},\tau,\underline{\eta},\Sigma^{-1}) = \frac{p(\underline{y} \mid \underline{\beta},\tau)p(\underline{\beta} \mid \underline{\eta},\Sigma^{-1})p(\underline{\eta})p(\tau)p(\Sigma^{-1})}{p(\underline{y})} \qquad (2.48)$$

Each of the distributions in equation (2.48) has been specified in sections 2.7.2.1-2.7.2.3 so the posterior distribution can be written as in equation (2.49).

$$p(\underline{\beta},\tau,\underline{\eta},\Sigma^{-1} \mid \underline{y}) \propto \prod_{i=1}^{I} \prod_{j=1}^{n_i} N(f(\underline{x}_{ij},\underline{\beta}_i),\tau^{-1}g(\underline{x}_{ij},\mu_{ij})) \times$$

$$\prod_{i=1}^{I} N(\underline{\beta},\Sigma^{-1}) \times G(0.5\nu_0,0.5\tau_0\nu_0) \times \qquad (2.49)$$

$$N(\underline{\eta},C) \times W((\rho R)^{-1},\rho)$$

Even though conjugate families are chosen for most pharmacokinetic models (i.e. nonlinear models), the solution of the Bayesian model is intractable hence numerical integration techniques, e.g. MCMC methods are required.

2.8 Monte Carlo Markov Chain (MCMC) Methods

The growth in the use of MCMC methods in the 1990's has seen the use of Bayesian data analysis increase considerably. The advances in computer power and speed have made MCMC techniques an additional method for modelling pharmacokinetic and pharmacodynamic data. These sampling methods are now a standard option in analysing any data and are likely to be used more as computer software such as WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs) is used and computing power increases.

Markov chain Monte Carlo methods can be described in two parts. Firstly, Monte Carlo integration refers to simulations for the evaluation of $E(f(X))$ with respect to some distribution $\pi(X)$. By drawing samples $\{X_t, t=1,\dots,n\}$ where n is the sample size from $\pi(X)$ and then approximating

$$E(f(X)) \approx \frac{1}{n} \sum_{t=1}^{n} f(X_t) \qquad (2.50)$$

So the population mean of $f(X)$ is estimated by a sample mean. When the samples are independent, laws of large numbers ensure that the approximation can be made as accurate as possible by increasing the sample size. Drawing independent samples is not always possible in practice so any method can be used where the samples can be made to be independent in some way. Using Markov chains is one way of sampling where $\pi(.)$ is the required stationary distribution.

Secondly, Markov chains define the process by which the random samples are chosen. Suppose a sequence of random variable is generated, $\{X_0, X_1, \dots\}$, such that at each time $t \geq 0$, the next state X_{t+1} is sampled from a distribution $P(X_{t+1}|X_t)$ which depends only on the current state of the chain X_t. This sequence is called a first-order Markov chain and $P(.|.)$ is called the transition distribution of the chain. Another feature of Markov chains is that as the random variables are sampled, the chain will gradually 'forget' its initial state and converge to the stationary distribution that does not depend on X_0 or t.

2.8.1 The Gibbs Sampling Algorithm

The Gibbs sampling algorithm was first introduced by Geman and Geman (1984) and was then applied to Bayesian analysis by Gelfand and Smith (1990). Wakefield *et al*

(1994) applied Gibbs sampling to population pharmacokinetic data and Wakefield and Racine-Poon (1995) applied it to population pharmacokinetic and pharmacodynamic data.

The Gibbs sampler requires that the joint distribution can be factored into full conditional distributions. The notation of Wakefield *et al* (1994) will be used to explain the Gibbs sampler.

Assume there is a joint distribution of random variables $\{X_1, X_2, ..., X_k\}$ given by $p(X_1, X_2, ..., X_k)$ and each of the conditional distributions $p(X_s | X_r, r \neq s)$ can be identified at least up to proportionality. Given arbitrary starting values $X^{(0)} = (X_1^{(0)}, ..., X_k^{(0)})$ for the k random variables, we then generate the following random variables:

$$
\begin{aligned}
X_1^{(1)} &\sim p(X_1^{(0)} \mid X_2^{(0)}, ..., X_k^{(0)}) \\
X_2^{(1)} &\sim p(X_2^{(0)} \mid X_1^{(1)}, ..., X_k^{(0)}) \\
&\,... \\
X_k^{(1)} &\sim p(X_k^{(0)} \mid X_1^{(1)}, ..., X_{k-1}^{(1)})
\end{aligned}
\tag{2.51}
$$

This completes one iteration of the sampling scheme giving $X^{(1)} = (X_1^{(1)}, ..., X_k^{(1)})$. After t such iterations we have $X^{(t)} = (X_1^{(t)}, ..., X_k^{(t)})$. The Markov chain just generated has equilibrium distribution $p(X_1, X_2, ..., X_k)$, the joint distribution of the random variables. Smith and Roberts (1993) showed that under weak regularity conditions, it follows that as t→∞, $X^{(t)}$ tends in distribution to a drawing from $p(X_1, X_2, ..., X_k)$, and the ergodic average is a consistent estimator of the expected value of any integrable function $g(X)$. The ergodic average is in essence a "moving average" and is defined by equation (2.52).

$$
\overline{g(X)} = \frac{1}{t} \sum_{j=1}^{t} g(X^{(j)})
\tag{2.52}
$$

To obtain a random sample from the joint distribution, either a long run is performed collecting suitably spaced realisations or by performing parallel independent runs of the chain and collecting the final realisation from each.

For population pharmacokinetic data, the full conditional distributions are given in Wakefield (1996).

2.8.2 The Griddy Gibbs Sampling Algorithm

As noted in the previous section, the full conditionals need to be known to be able to sample from. Also, in many situations, the conditional distribution is univariate. When the full conditional distribution in the Gibbs sampling algorithm is hard to sample from or is not of a known distributional form, then the idea of the Griddy-Gibbs sampler is to form a simple approximation to the inverse cumulative distribution function on a grid of points (Ritter and Tanner (1992)). The algorithm is as follows:

1. Evaluate $p(X_s|X_r, r{\neq}s)$ at $X_s{=}x_1, x_2, ..., x_k$ to obtain $w_1, w_2, ..., w_k$. The weights w are assigned to the vector $(X_1, X_2, ..., X_k)$ that has been drawn from the current approximation to the joint distribution g_i via

$$w = \frac{q(X_1, ..., X_k)}{g_i(X_1, ..., X_k)} \qquad (2.53)$$

where q is proportional to p, the distribution of interest.

2. Use $w_1, w_2, ..., w_k$ to obtain an approximation to the inverse cumulative distribution function of $p(X_s|X_r, r{\neq}s)$.

3. Sample a uniform (0,1), deviate and transform the observation via the approximate inverse cumulative distribution function.

The conditional probability function p need only be known up to proportionality (q) because the normalisation can be worked out directly from the weights. The grid does not need to be uniform, as a grid that puts more mass in neighbourhoods of greater density is more efficient. Also, the number of points on the grid does not have to remain constant throughout. As the sampling scheme is run then the grid can become finer as more information is built up.

This algorithm is employed in version 0.6 of the BUGS program as a way of estimating parameters in nonlinear models.

2.8.3   The Metropolis Algorithm

Metropolis *et al* (1953) introduced the Metropolis algorithm into the field of calculating chemical properties. It is implemented in the case where the probability density function $p(X)$ is known only up to proportionality and in its current form can not be sampled from directly. The Metropolis algorithm is implemented to obtain realisations from $p(X)/\int p(X)dX$. Firstly, a proposal distribution, $q(X|Y)$ must be defined such that random variates from this distribution are readily obtained and lie in the same range of interest. For the Metropolis algorithm, the proposal distribution must be symmetric such that $q(X|Y)=q(Y|X)$. The algorithm for the ith iteration is the following:

1. Generate Z from $q(X_i|Y)$.

2. Define:

$$\alpha = \min\left\{1, \frac{p(Z)}{p(X_i)}\right\} \qquad (2.54)$$

3. Generate U from a uniform distribution on (0,1).

4. Let $X_{i+1}=Z$ if $\alpha \leq U$, otherwise $X_{i+1}=X_i$.

Then as the number iterations, $t \to \infty$ the realisations will converge to those from the desired distribution.

The Metropolis algorithm is easy to implement within the Gibbs sampler when the conditional distribution is not of a known form. This algorithm is implemented in WinBUGS.

2.8.4 The Metropolis-Hastings Algorithm

Hastings (1970) generalised the method of Metropolis to allow for the proposal distribution not to be symmetric and allow more freedom in the choice of this distribution. The generalisation comes at step 2 of the Metropolis algorithm and is replaced by the following step:

2. Define:

$$\alpha = \min\left\{1, \frac{p(Z)q(Z \mid X_i)}{p(X_i)q(X_i \mid Z)}\right\} \tag{2.55}$$

The rejection step must now take into account that the proposal distribution is not symmetric and does not factor out. This generalisation allows for more distributions as proposal distributions will hopefully mean that less iterations are rejected.

2.8.5 The Rejection Sampling Algorithm

The rejection sampling algorithm is presented in Ripley (1987) and is a method for drawing independent samples from a distribution proportional to $p(X)$. Rejection

sampling is not a MCMC method as each sample is independent of the last so it does not have the Markov property but is a Monte Carlo method. To be able to sample from $p(X)$, it is required to be able to sample readily from $q(X)$ for which there is a finite constant m such that $mq(X) \geq p(X)$ over the whole range of X. For practical purposes it is important to be able to calculate m easily. The ith step in the algorithm is the following:

1. Sample X_i from $q(X)$.

2. Sample U from uniform distribution on $(0,1)$.

3. If $p(X_i)/mq(X_i) < U$ then go back to step 1, else set $X_R = X_i$.

4. Return X_R.

Rejection sampling does not require the evaluation of the integration of the distribution over its domain. This is very convenient for sampling from full conditional distributions when they are known only up to a constant of proportionality.

One of the main problems of rejection sampling is the acceptance rate can be very low meaning that it can take a long time for an adequate number of independent samples to be collected. The acceptance rate is dependent on m and the higher the value implies greater numbers of iterations being rejected.

2.8.6 The Adaptive Rejection Sampling Algorithm

Gilks and Wild (1992) introduced the adaptive rejection sampling algorithm as a way of getting round the problem of poor acceptance rates in rejection sampling. This is accomplished by updating the sampling density $q(X)$ at each iteration so as to reduce the number of rejections and speeding up the time to collecting an adequate number of

samples. This improvement is made by incorporating information about $p(X)$ in $q(X)$ so it approximates closer, the true distribution.

An important condition for adaptive rejection sampling is that the density function be log-concave. The definition of this is given by the following where D is the domain of the distribution p, which is not necessarily continuously differentiable: Log-concavity ensures that envelopes can be easily constructed for the distribution in question.

$$\log(p(a)) - 2\log(p(b)) + \log(p(c)) < 0, \forall a,b,c \in D, a < b < c \qquad (2.56)$$

For the adaptive rejection sampling algorithm, choose a set of points $S_n = \{x_i; i=0, ..., n+1\}$ in ascending order. For $1 \leq i \leq j \leq n$ let $L_{ij}(x; S_n)$ denote the straight line through points $[x_i, ln(p(x_i))]$ and $[x_j, ln(p(x_j))]$, and for other (i,j) let $L_{ij}(x; S_n)$ be undefined. Define a piecewise linear function $h_n(x) = \min[L_{i-1,i}(x; S_n), L_{i+1,i+2}(x; S_n)]$, $x_i \leq x < x_{i+1}$ where the notational dependence of $h_n(x)$ on S_n is suppressed. $h_n(x)$ is an envelope everywhere for $log(p(x))$, i.e. $h_n(x) \geq log(p(x))$. Adaptive rejection sampling can now be performed with the sampling distribution given by equation (2.57).

$$q_n(x) = \frac{1}{m_n} \exp(h_n(x))$$
$$\text{where} \qquad (2.57)$$
$$m_n = \int \exp(h_n(x)) dx$$

The adaptive rejection sampling is as follows:

1. Initialise n and S_n.

2. Sample X from $q_n(x)$.

3. Sample U from uniform distribution $(0,1)$.

4. If $p(X)/exp(h_n(X)) < U$ then (rejection step) set $S_{n+1} = S_n \cup \{X\}$, relabel points in S_{n+1} in ascending order, increment n and go back to step 2, else (acceptance step) set $X_A = X$.

5. Return X_A.

At each rejection, the number of points in S is increased by one, thereby reducing the probability of rejection at the next step.

2.8.7 The Adaptive Rejection Metropolis Algorithm

Gilks *et al* (1995) introduced a method where sampling can take place from non-log-concave distributions. The Metropolis algorithm is capable of this but to avoid many rejections during sampling, the proposal distribution can be updated as in adaptive rejection sampling. The mixture of these two sampling algorithms allows the development of a good proposal distribution from which to sample. Unlike adaptive rejection sampling which produces independent samples, the adaptive rejection Metropolis algorithm produces correlated samples due to the Markov property of the algorithm. Details of this algorithm are given in Gilks *et al* (1995).

2.8.8 BUGS and WinBUGS

BUGS (Bayesian inference Using Gibbs Sampling) was originally developed at the University of Cambridge by Gilks *et al* (1994). The purpose of the original program was to allow Gibbs sampling to be applied to complex statistical models that when tackled using Bayesian methods previously, would require programs to be written from scratch. Since its original development, BUGS has progressed from a DOS based program to a Windows program with the introduction of WinBUGS. Earlier work in this thesis was carried out using BUGS but some of the later analyses were carried out using WinBUGS.

BUGS (referring to both BUGS and WinBUGS) assumes a full Bayesian probability model, in which all quantities are treated as random variables. The *model* consists of a defined joint distribution over all unobserved (parameters and missing data) and observed quantities (the data). The data is then conditioned on to obtain a posterior distribution over the parameters and unobserved data. The model can also be defined by a directed acyclic graph (DAG) which gives a graphical view of how the full distribution model is put together and each part is connected to the parent nodes.

BUGS 0.5 only allowed for the estimation of linear models, generalised linear models, survival models, and other simple models but with complicated distributional structures. This was due to the fact that within the Gibbs sampler, there was not a routine for non-log-concave distributions to be sampled from and so nonlinear models could not be considered. BUGS 0.6 included a Griddy-Gibbs sampler to allow for the estimation of nonlinear models and then the introduction of WinBUGS allowed for general nonlinear models to be considered as an adaptive Metropolis algorithm was included. With the latest version being released, WinBUGS 1.2, the program should be more general and allow for an even wider class of models to be analysed. The current program already allows for many of the types of models that are found in pharmacokinetic and pharmacodynamic analyses to be estimated.

2.9 Model Checking and Diagnostics

The inspection of standardised residuals has become one of the main methods for determining goodness of fit in normally distributed data. The use of the objective

function in NONMEM and its asymptotic χ^2 properties has also provided a method for comparing nested models. These methods are not necessarily appropriate for non-normal data and model checking must be given some consideration. A range of methods are possible as there is some work in the literature on model checking for categorical data such as the work by Albert and Chib (1995), Lemenshow and Hosmer (1982), Farrington (1996) and Pregibon (1981). A few of the standard methods will be reviewed now as well as a Bayesian method called Bayes Factors which can be used for model comparison for any types of models.

2.9.1 Goodness of Fit Statistics

In fitting models to data, it is important to show that a particular model fits the data well or is superior to other models. In population pharmacokinetic analysis, the ideas behind determining how well a model fits the data are well understood. In categorical data analysis, it is not quite so obvious how to compare models or to see how well a model fits the data. One method for nested models without random effects is to use the deviance statistic. Assume that there are two models called the current model and the saturated model. The current model is being tested to see how well it fits the data. The saturated model is where the model predicted data coincides with the observed data. The deviance statistic is defined as in equation (2.58).

$$D = -2\log\left(\frac{\hat{L}_{current}}{\hat{L}_{saturated}}\right) = -2[\log(\hat{L}_{current}) - \log(\hat{L}_{saturated})] \tag{2.58}$$

where L is the likelihood value. The deviance statistic compares the current model to the saturated model by equation (2.58). Large values of D are encountered when the

current model is a poor fit to the data and small values when it is a good fit. For binary data, the deviance statistic is given in equation (2.59).

$$D = 2\sum_i \left\{ y_i \log\left(\frac{p_{i,saturated}}{p_{i,current}}\right) + (1 - y_i)\log\left(\frac{1 - p_{i,saturated}}{1 - p_{i,current}}\right) \right\}$$
(2.59)

For binary data, the log of the full model likelihood is zero so the deviance can be simplified to equation (2.60). As can be seen, the deviance does not depend on the data, so the deviance statistic cannot be used for the special case of binary data (Collett (1991)).

$$D = -2\sum_i \left\{ p_{i,current} \log\left(\frac{p_{i,current}}{1 - p_{i,current}}\right) + \log(1 - p_{i,current}) \right\}$$
(2.60)

To compare models, usual large sample theory is used to show that the deviance statistic is distributed as a χ^2 with (n-p) degrees of freedom where n is the number of observations and p is the number of parameters in the current model. To compare models, the difference in the deviance statistics is compared to the χ^2 on the (p-q) degrees of freedom which is the difference in the number of parameters between the two models.

The Pearson statistic is another statistic that can be used for model checking but it does not allow for the comparison of models. Lemeshow and Hosmer (1982) discuss several other statistics for logistic regression modelling in the field of epidemiologic research.

71

2.9.2 Residual Analysis

Residual analysis plays a large part in determining whether models for population pharmacokinetic data are adequate. The standardised residuals method used in population pharmacokinetic analysis is not appropriate in population pharmacodynamic analysis where the response is categorical. Several papers have dealt with the topic of residuals for binary and binomial data such as Albert and Chib (1995), Jennings (1986), Landwehr *et al* (1984) and Pregibon (1981).

The types of residuals usually defined for binary or binomial data are those that are linked to the statistics defined in section 2.9.1. A common type of residual to use is the deviance residual defined in equation (2.61) for binary data

$$d_i = \text{sgn}(y_i - \hat{p}_i)\sqrt{-2[y_i \log \hat{p}_i + (1 - y_i)\log(1 - \hat{p}_i)]} \qquad (2.61)$$

where sgn(.) is +1 when $y_i \geq p_i$ and -1 otherwise. By summing over the square of the deviance residual, the deviance statistic is obtained. This residual is more like an ordinary residual for the normally distributed data case.

Another common type of residual is the Pearson residual also connected to the Pearson statistic and is defined in equation (2.62) for binary data.

$$X_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}} \qquad (2.62)$$

Anscombe specified another type of residual in 1953 called the Anscombe residual. The problem with the interpretation of residuals is helped if their distributional assumptions to the fitted model are known. Since the exact distribution of the residuals

defined above are not known exactly, this suggests that if some function of the data can be found that is approximately normally distributed, then a standardised residual defined to approximate unit variance can be defined. The Anscombe residual is defined in equation (2.63).

$$r_{Ai} = \frac{A(y_i) - A(\hat{y}_i)}{s.e.\{A(y_i) - A(\hat{y}_i)\}} \tag{2.63}$$

This residual has rarely been used in practice because finding the function $A(.)$ which makes the data approximately normally distributed is difficult.

Landwehr *et al* (1984) suggested a series of diagnostic plots for analysing binary and binomial data fitted using logistic regression. Local mean deviance plots are suggested which 'lump' near binary points together and use the estimated proportion of success for the deviance statistic. An empirical probability plot was also suggested where the residuals are plotted against simulated quantiles to check for model adequacy by looking for a straight line. Another residual is the partial residual which is used to see whether there are any missing predictors or whether a current covariate should be transformed to improve the fit of the model.

All of these residuals are not particularly satisfactory as the distributions for residuals are intractable which does not make the interpretation easy. Residual plots can not be interpreted in the same way as for normally distributed data as the nature of binary data can mean that correlation can occur in residual plots even if the model fits the data well. For this purpose, it is believed something based on simulations will give a better idea of how good the model fits the data as is the case for Bayes factors described in the next section.

2.9.3 Bayes Factors

The Bayesian approach to hypothesis testing was first developed by Jeffreys (1935,1961) as reported by Kass and Raftery (1995). Rather than being related to the testing of hypotheses, Bayes factors deal with the comparison of two or more competing models. Much work has been carried out on Bayes factors (e.g. Carlin and Chib (1995), Smith and Spiegelhalter (1981) and Aitkin (1991)). One of the advantages of Bayes factors is that it deals with prediction of the response variable rather than a comparison of models based on asymptotic theory. Therefore any models can be compared whether they are nested or not.

Carlin and Chib (1995) described the problem of model choice with the use of Bayes factors. Suppose the problem is to choose between K models for an observed set of data, y. Corresponding to each model there is a distinct parameter vector θ_j, j=1,..,K. Assume an integer subscripts the model, then interest lies in p(M=j|y) and p(θ_j|y), the probability of getting model j given the data and the posterior distribution of the parameter vectors given the data respectively. The Bayes factor for the comparison of two models is given in equation (2.64). It is defined as the ratio of posterior to prior odds in favour of model 2. By Bayes' theorem, it is the ratio of the observed marginal densities for the two models.

$$B_{21} = \frac{p(y \mid M = 2)}{p(y \mid M = 1)} \tag{2.64}$$

The Bayes factor is derived by starting from Bayes theorem as defined in equation (2.65).

$$\Pr(M_k \mid \underline{y}) = \frac{\Pr(\underline{y} \mid M_k)\Pr(M_k)}{\Pr(\underline{y} \mid M_1)\Pr(M_1) + \Pr(\underline{y} \mid M_2)\Pr(M_2)} \tag{2.65}$$

$$k = 1,2$$

The posterior odds are given by equation (2.66) and so the posterior odds can be obtained by simply multiplying the prior odds by the Bayes factor as defined in equation (2.64). As the value of the Bayes factor increases in value then there is more support for model 2 against model 1.

$$\frac{\Pr(M_1 \mid \underline{y})}{\Pr(M_2 \mid \underline{y})} = \frac{\Pr(\underline{y} \mid M_1)}{\Pr(\underline{y} \mid M_2)} \frac{\Pr(M_1)}{\Pr(M_2)} \tag{2.66}$$

3 Toxicokinetic Data Set I

3.1 The Study and The Data

For toxicokinetic evaluation, 4 groups of rats were studied at different doses of the drug. Group one was the placebo group and contained 5 rats per gender. Group 2 was given a dose between 0 and 30 mg/kg but was unknown and 3 rats per gender were assigned to the group. Group 3 received 30 mg/kg and there were 3 rats per gender randomly assigned to this treatment group. Finally, group 4 received 200 mg/kg of which 5 rats per gender were assigned to this group. The rats were given the drug (which was also unknown) orally once a day for 4 weeks. On day 28 of the study, plasma samples were obtained at 0.5, 1, 2, 4, 8 and 24 hours post-dose (assumed steady state). Rats in the control group had samples taken at 1 hour post-dose. The number of concentrations (ng/mL) obtained from dose groups 3 and 4 are given in table 3.1. Group 2 was not considered as the dose was unknown.

Table 3.1. *Number of concentration measurements for groups 3 and 4.*

	30 mg/kg	(n)	200 mg/kg	(n)	Total	(n)
Male	18	(3)	30	(5)	48	(8)
Female	18	(3)	30	(5)	48	(8)
Total	36	(6)	60	(10)	96	(16)

(n) number of rats.

At the end of day 28, the rats were sacrificed and were dissected to observe the severity of lesions within the rats conjectured to be caused by the drug. In this case, three rats per gender per group were sacrificed and observed for lesions. Data was only available on the 30 and 200 mg/kg groups. The severity of the lesions were used as the

pharmacodynamic measure of the drug. Severity was defined on a five category scale, where 0=no lesion, 1=very slight lesion, 2=slight lesion, 3=moderate lesion and 4=marked lesion. Lesions were observed in several physiological areas of the rat such as skin, subcutis and fascia, eyes, lacrimal gland, lymphoid tissues, pyloric stomach, intestinal tract, liver, bone marrow and prostate gland. Within these areas, subsections were observed for lesions. In total there were a possible 25 lesion scores for each rat. Data is available for the 30 and 200 mg/kg groups. The number of lesions observed for each group is given in table 3.2.

Table 3.2. *Number of lesions in 30 and 200 mg/kg groups.*

	30 mg/kg	200 mg/kg	Total
Male	60	75	135
Female	59	71	130
Total	119	146	265

3.2 Pharmacokinetic Analysis

To construct a pharmacokinetic/pharmacodynamic model, it is first necessary to be able to describe the pharmacokinetics. The data was initially viewed graphically to check what model might be appropriate. After examining the data, a 2 compartment first order absorption model at steady state was assumed as the drug was known to be given orally. The model was analysed in NONMEM version IV on a Hewlett Packard workstation using the UNIX operating system. The 2 compartment model is given in equation (3.1).

$$E(C) = \lambda_1 \frac{e^{-\alpha t}}{1 - e^{-\alpha \tau}} + \lambda_2 \frac{e^{-\beta t}}{1 - e^{-\beta \tau}} + \lambda_3 \frac{e^{-k_a t}}{1 - e^{-k_a \tau}} \tag{3.1}$$

where

$$\alpha = \frac{1}{2}\{(k_{10} + k_{12} + k_{21}) + \sqrt{(k_{10} + k_{12} + k_{21})^2 - 4k_{10}k_{21}}\}$$

$$\beta = \frac{1}{2}\{(k_{10} + k_{12} + k_{21}) - \sqrt{(k_{10} + k_{12} + k_{21})^2 - 4k_{10}k_{21}}\}$$

$$\lambda_1 = \frac{(\alpha - k_{21})}{(k_a - \alpha)(\alpha - \beta)V_1}$$

$$\lambda_2 = \frac{(k_{21} - \alpha)}{(k_a - \beta)(\alpha - \beta)V_1}$$

$$\lambda_3 = \frac{(k_{21} - k_a)}{(\alpha - k_a)(\beta - k_a)V_1}$$

V_1 = volume of distribution of the central compartment

k_{12} and k_{21} = intercompartmental rate constants

k_{10} = elimination rate constant

k_a = absorption rate constant

$k_{12} = Cl_d/V_1$

$k_{21} = Cl_d/V_2$

$k_{10} = Cl/V_1$

Figure 3.1. *Plot of dose normalised data.*

Time(hours)
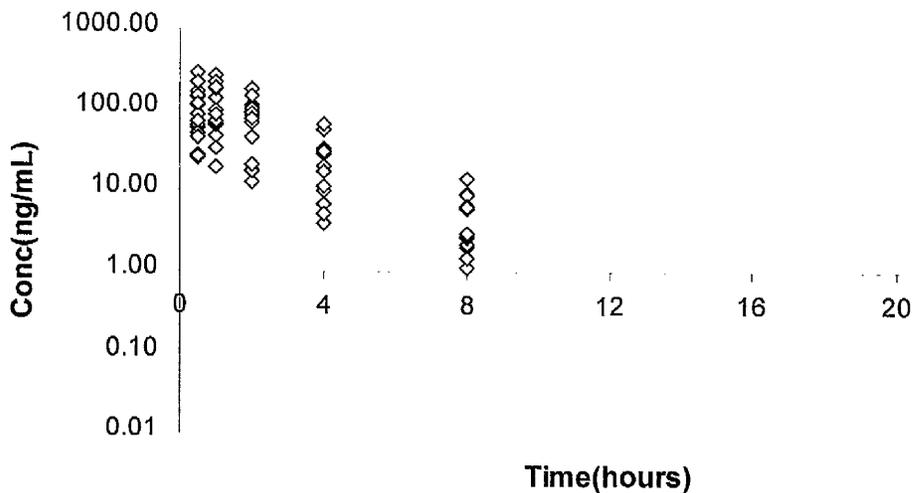
In this model, k_a was set to a value of 10 hr^{-1}. Viewing the data in figure 3.1 shows that the drug appears to be completely absorbed by the time the first concentration measurements are taken at 0.5 hours. The value of 10 hr^{-1} corresponds to a very rapid absorption phase. An exponential error model was used for the interindividual variance estimates on the parameters Cl/F and V_1/F and on the residual error term. The population parameter estimates are given in table 3.3 and the individual estimates of apparent clearance and apparent central compartment volume of distribution are given in table 3.4. The residual and weighted residual plots and two individual plots are given in figures 3.2-3.5. A one compartment model and a variety of error models were also tried but these did not fit the data adequately.

Table 3.3. *Population parameter estimates for two compartment first order absorption model.*

	Mean	Standard Deviation
Cl/F(L/hr)	2.55	0.334
V_1/F(L)	14	4.87
V_2/F(L)	11.4	10.8
Cl_d/F(L/hr)	0.883	0.467
ω_{Cl}^2	0.0824	0.0612
ω_V^2	2.15	1.69
σ^2	0.394	0.169

Table 3.4. *Individual estimates of apparent clearance and central compartment volume of distribution. 3 rats per gender in 200 mg/kg group corresponding to the rats with pharmacodynamic data available.*

Rat i.d.	Dose(mg/kg)	Gender	Cl/F	V_1/F
496	30	Male	2.513	7.18
497			2.531	4.562
498			2.15	4.769
515		Female	2.705	8.168
516			1.921	3.317
522			1.666	2.643
499	200	Male	2.423	9.780
501			2.483	10.915
503			2.563	8.683
518		Female	3.088	9.143
519			3.079	6.318
520			2.592	10.901

Figure 3.2. *Residual versus predicted plot.*
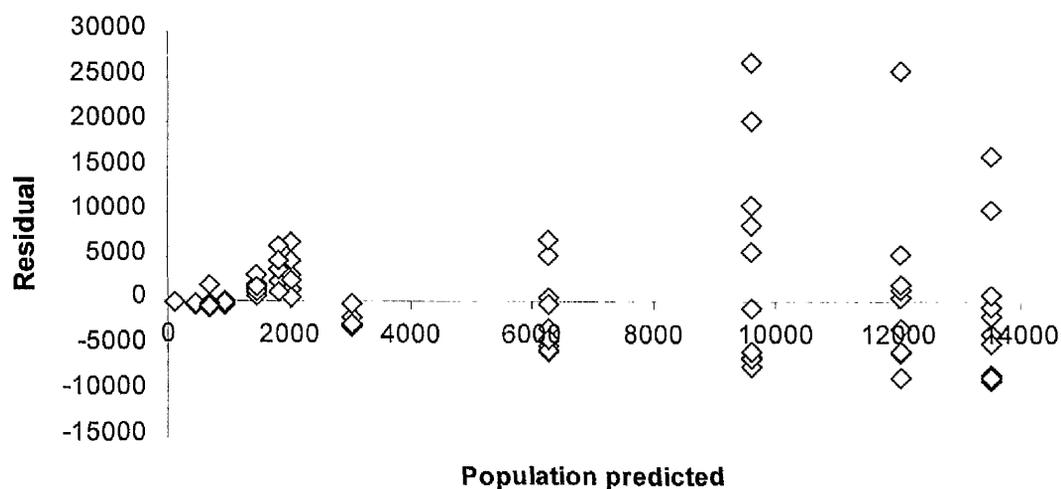
Figure 3.3. *Weighted residual versus predicted plot.*
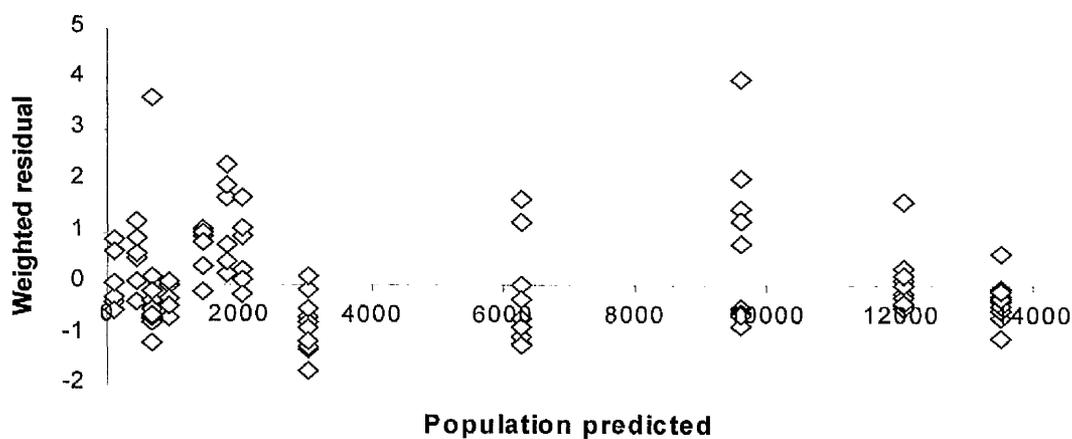
Figure 3.4. *Rat 496.*
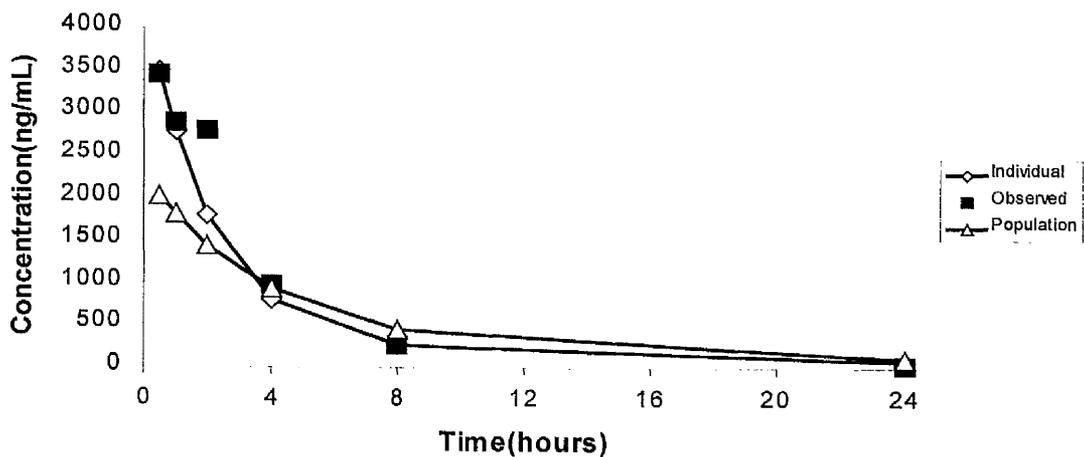
Figure 3.5. *Rat 519.*
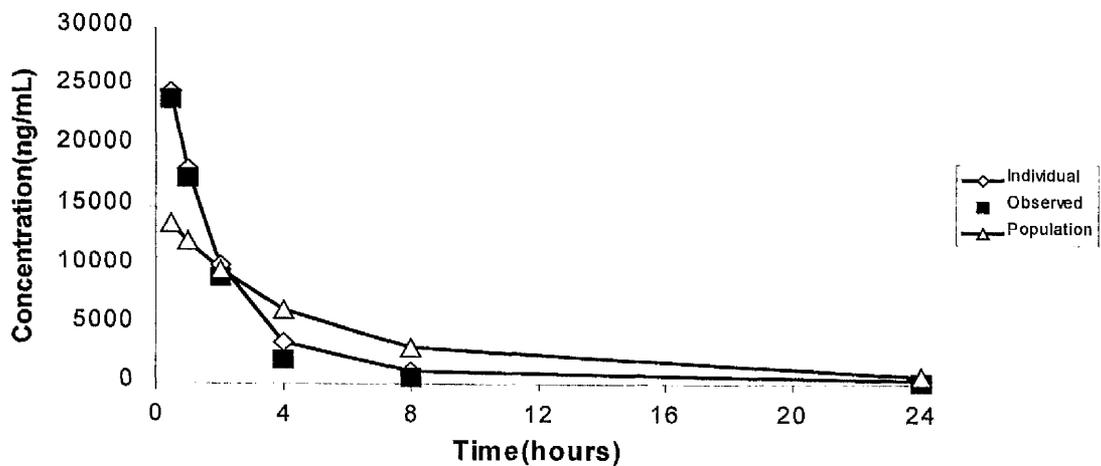
3.3 Pharmacodynamic Analysis

The pharmacodynamic data is in the form of 25 different responses which correspond to the different areas of the dissected rat. To analyse these data fully would require a multivariate logistic regression analysis where the response vector for a particular rat and covariate pattern would be of dimension 25×1 which would lead to a complicated analysis. The data could have been analysed where each response component (each observed area of the rat) was analysed separately but there might not have been enough information in the separate components of the rat to enable this. The data were analysed by pooling the observations in each category for each rat. These data are given in table 3.5. This is not necessarily the best way of analysing these data as it does not include any marker to indicate what part of the rat the data came from but will give an overall measure of the degree of lesions in the rat. An analysis of counts from each main category (skin, subcutis and fascia, eyes, lacrimal gland, lymphoid tissues, pyloric stomach, intestinal tract, liver, bone marrow and prostate gland) of the dissected rat could have been carried out and could have been reasonable at describing the data but this was not done.

Table 3.5. *Counts of lesion scores for each rat.*

Category				0	1	2	3	4
Rat	Dose	Gender	AUC					
496	30	Male	11.94	13	4	3	0	0
497			11.85	17	1	2	0	0
498			13.95	16	4	0	0	0
515		Female	11.09	18	2	0	0	0
516			15.62	8	3	7	2	0
522			18.01	11	4	4	0	0
499	200	Male	82.35	2	2	11	7	3
501			80.56	0	7	8	7	3
503			78.04	4	8	10	2	1
518		Female	64.76	4	12	6	1	0
519			64.95	13	6	5	0	0
520			77.17	8	7	6	3	0

The AUC estimates in table 3.5 are based on *F.Dose/Cl* where the *Cl/F* estimates are those in table 3.4. The numbers in table 3.5 are the counts of each score for each rat where the total number of observations in each rat is found by adding the number of scores.

From this table, certain characteristic probabilities can be obtained such as the probability having no lesions by the end of the study for either the 30 or 200 mg/kg dose group or the difference in probabilities of being in a certain category between male and female rats. These probabilities (percentages) for each rat can be seen easily in figure 3.6. It can be seen that the first six rats (corresponding to dose group 30 mg/kg) has higher percentages in the lower categories compared to the last six rats (corresponding to the 200 mg/kg dose group).

Figure 3.6. *Empirical probabilities of being in each category for each rat. Numbers 1-12 correspond to same order 496-520 as in table 3.5.*
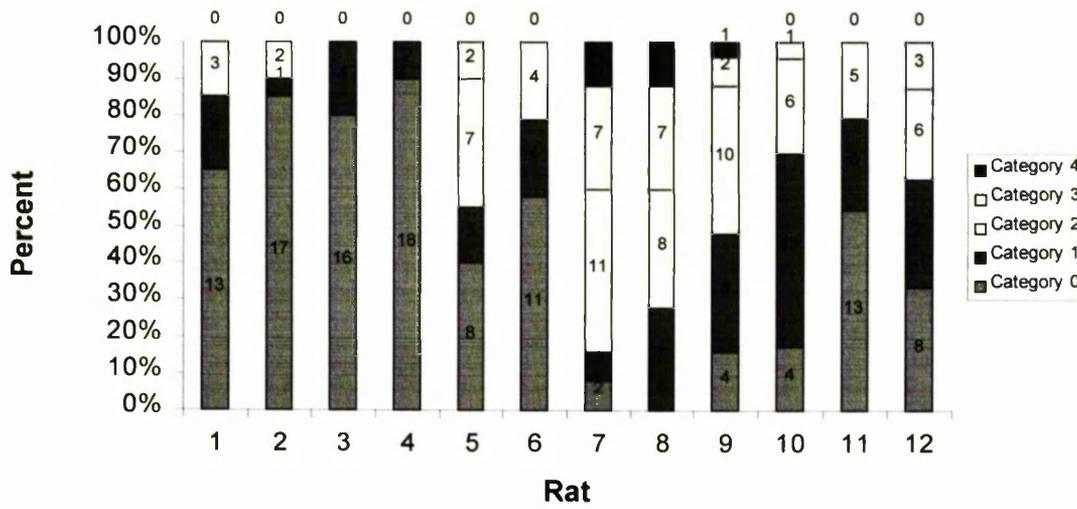
Another possible difference between the number of counts of each rat could be due to the pharmacokinetics of the drug which is taken into account in the estimation of the AUCs for each rat. Figure 3.7 shows how the counts change as a function of AUC. The figure is in the form of a scatter plot rather than a percentage bar plot as in figure 3.6 because the predictor, AUC, is continuous rather than categorical (although dose can be treated as continuous).

Figure 3.7. *Counts as a function of AUC.*
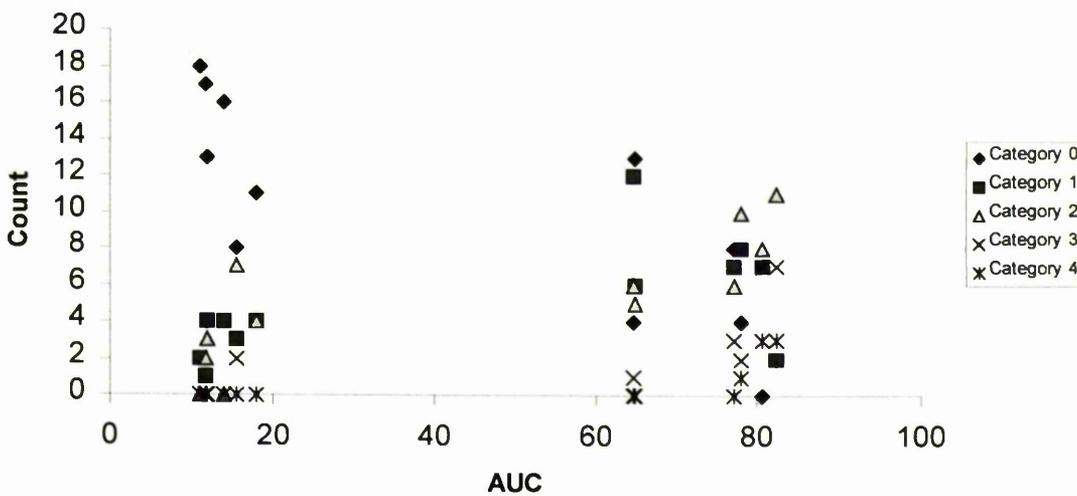
Instead of following a route of hypothesis testing, e.g. the difference between probabilities from different groups of the data, the purpose of this analysis is to find a model that describes the data. Hypothesis testing will not allow prediction to new dose ranges or different levels of exposure to the drug through the AUC value. Whether these predictions are any good would need to be tested but modelling still allows this possibility. There is also the possibility of designing new studies based on the model obtained from the data analysis.

For data of an independent multinomial sampling form (Agresti (1990)) as in table 3.5, one of the appropriate models to use is the proportional odds model as described in section 2.4.6. A proportional odds model with five categories would have four cut points corresponding to the baseline cumulative probabilities. The models that were tried for these data correspond to the general form given in equation (3.2).

$$\text{log}it(\text{Pr}(Y_i \leq k \mid \underline{b}_i)) = \theta_k + \underline{x}_i\underline{\beta} + \underline{z}_i\underline{b}_i, k = 0,1,2,3 \tag{3.2}$$

This model allows for the inclusion of random effects components.

The program used to estimate any parameter estimates was BUGS version 0.5. This is one of the earlier versions that did not allow the estimation of nonlinear parameter models, such as pharmacokinetic models. Generalised linear models could be estimated with the use of the adaptive rejection sampling algorithm. BUGS was used because it allows the specification of distribution and model characteristics easily. All model specifications and distributions will be defined for each model.

The first model to be considered was a linear model in all factors other than AUC as given in equation (3.3). An interaction term was included between dose and gender and

only linear components were considered as there were only two doses and so polynomials in dose would not have added anything.

$$\text{logit}(\Pr(Y_i \le k \mid b_i)) = \theta_k + \beta_1 dose_i + \beta_2 gender_i + \beta_3 dose_i gender_i + b_i, k = 0,1,2,3$$
$$Y_i \sim Mult(\gamma_0, \gamma_1, \gamma_2, \gamma_3, n_i)$$
$$\theta_k \sim N(0,1 \times 10^5), k = 0,1,2,3$$
$$\beta_j \sim N(0,1 \times 10^5), j = 1,2,3$$
$$b_i \sim N(0, \tau^{-1})$$
$$\tau \sim G(0.001, 0.001)$$

(3.3)

Gender was binary with 0 representing male and 1 for female. The gamma distribution in BUGS is parameterised so that $Y \sim G(a,b)$, $E(Y) = a/b$ and $var(Y) = a/b^2$. The gamma distribution for the precision (reciprocal of the variance) is the standard conjugate distribution when a normal distribution is put on the random effect b_i. The results for this model are given in table 3.6.

Table 3.6. *Results for BUGS estimation of linear proportional odds model* $\text{logit}(\Pr(Y_i \le k \mid b_i)) = \theta_k + \beta_1 dose_i + \beta_2 gender_i + \beta_3 dose_i gender_i + b_i, k = 0,1,2,3$.

	Mean	Standard Deviation	95% Credible Interval
θ_0	2.076	0.618	(0.779,3.171)
θ_1	3.467	0.633	(2.158,4.575)
θ_2	5.439	0.667	(4.082,6.627)
θ_3	7.239	0.758	(5.784,8.677)
β_1	-0.0229	0.00412	(-0.0306,-0.02316)
β_2	-1.498	0.868	(-3.097,0.4013)
β_3	0.0171	0.00566	(0.00487,0.0303)
σ	0.845	0.326	(0.3209,1.598)

These parameter estimates are based upon a burn in of 500 iterations then a sample of 2000 was saved for parameter estimation. The mean is the arithmetic average of the sample and the standard deviation is the sample standard deviation.

The cut points correspond to baseline probabilities of $\Pr(Y_i \leq 0) = 0.88$, $\Pr(Y_i \leq 1) = 0.97$, $\Pr(Y_i \leq 2) = 0.996$ and $\Pr(Y_i \leq 3) = 0.999$. It is easily seen that the model predicts that there is greatest probability of having no lesions ($Y_i = 0$) when the dose is 0 mg/kg and the rat is male. Significance tests are not usually performed in Bayesian statistics but one way of looking at the importance of certain components of the model is by looking at the 95% credible interval (the middle 95% of the samples obtained from Gibbs sampling). The 95% credible interval for the interaction between dose and sex dose not include 0 as a sample so it can be interpreted that there is a 95% probability that the parameter value is different from 0. This means that there is no need to remove any parameters from this model as the interaction term is 'significant'. The value of σ is the standard deviation for the variability around the model on the logit scale. This acts as a shift for the individual from the population average. This shift occurs on the logit scale so that probabilities remain on the probability scale [0,1]. The magnitude of the interaction term is small and so linear models without the interaction term and without the interaction and gender terms were considered for completeness. The model equations are given in equation (3.4) and the results are given in table 3.7. The prior distributions are the same as those in equation (3.3) where any parameters not in the model, the priors are excluded.

$$
\begin{aligned}
\text{log}it(\Pr(Y_i \leq k \mid b_i)) &= \theta_k + \beta_1 dose_i + \beta_2 gender_i + b_i, k = 0,1,2,3 \\
\text{log}it(\Pr(Y_i \leq k \mid b_i)) &= \theta_k + \beta_1 dose_i + b_i, k = 0,1,2,3
\end{aligned}
\tag{3.4}
$$

Table 3.7. *Results of linear models in dose and dose+gender.*

	Dose+gender model			Dose model		
	Mean	S.D.	95% C.I.	Mean	S.D.	95% C.I.
θ_0	0.782	0.651	(-0.0441,2.055)	1.328	0.696	(0.0326,2.751)
θ_1	2.153	0.663	(0.902,3.48)	2.697	0.715	(1.331,4.142)
θ_2	4.103	0.687	(2.828,5.481)	4.634	0.753	(3.176,6.171)
θ_3	5.888	0.763	(4.487,7.484)	6.431	0.841	(4.9,8.1)
β_1	-0.0133	0.00437	(-0.022,-0.00452)	-0.0144	0.00456	(-0.0231,-0.0058)
β_2	0.848	0.717	(-0.436,2.477)	-	-	-
σ	1.158	0.35	(0.6125,1.978)	1.192	0.352	(0.671,2.051)

S.D. Standard deviation

In the dose and gender model, the 95% credible interval for the gender term includes 0 and so it appears that between the dose and dose+gender model, there is not a significant difference in model fit. A plot of the dose model is given in figure 3.8. Pearson and deviance residual plots are given in figures 3.9a-b and 3.10a-b.

Figure 3.8. *Proportional odds model for linear dose.*

88
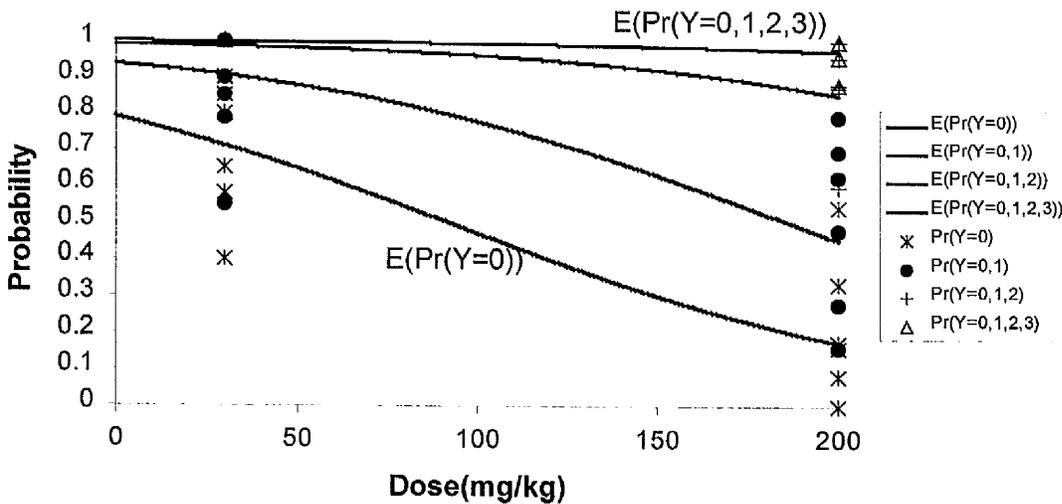
Figure 3.9a. *Pearson residuals versus dose for linear dose model.*
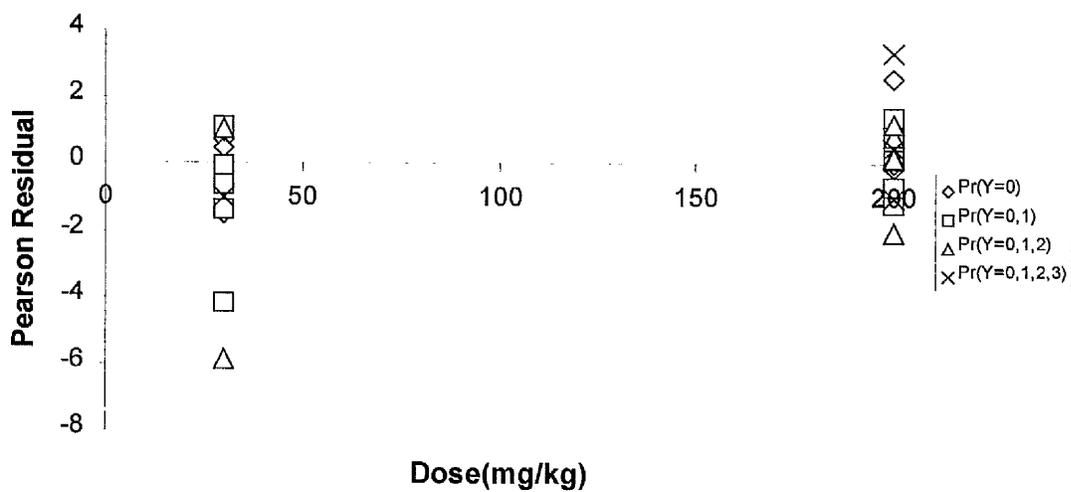
Figure 3.9b. *Pearson residuals versus predicted count for linear dose model.*
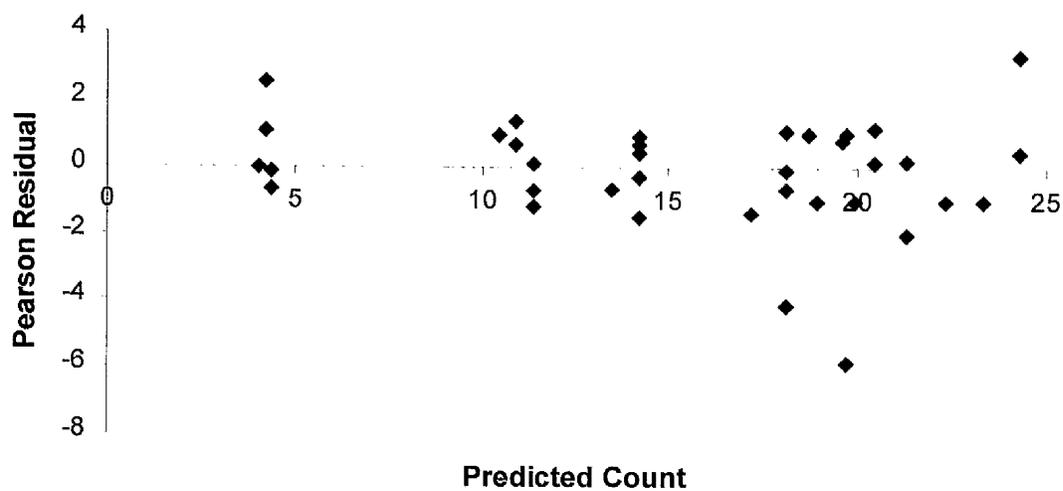
Figure 3.10a. *Deviance residuals versus dose for linear dose model*
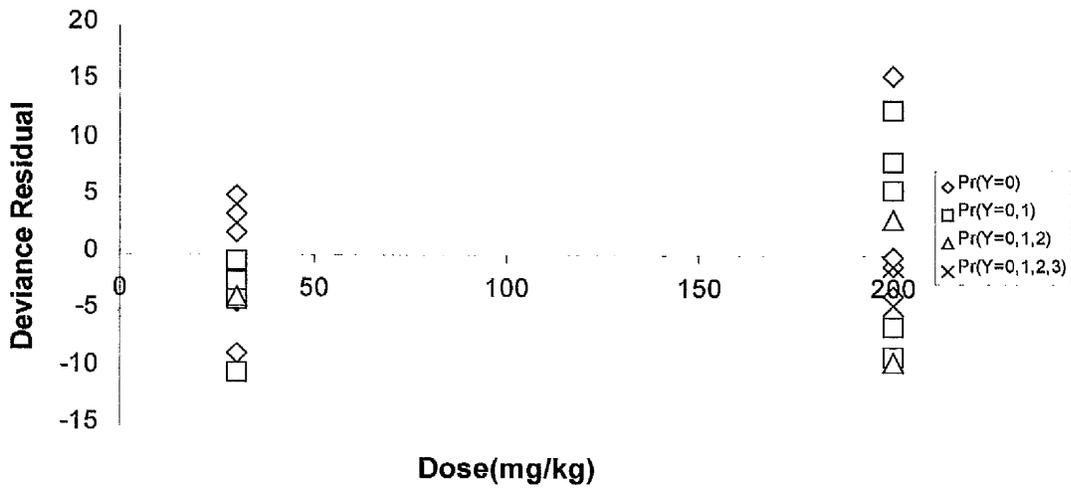
Figure 3.10b. *Deviance residuals versus predicted count for linear dose model.*
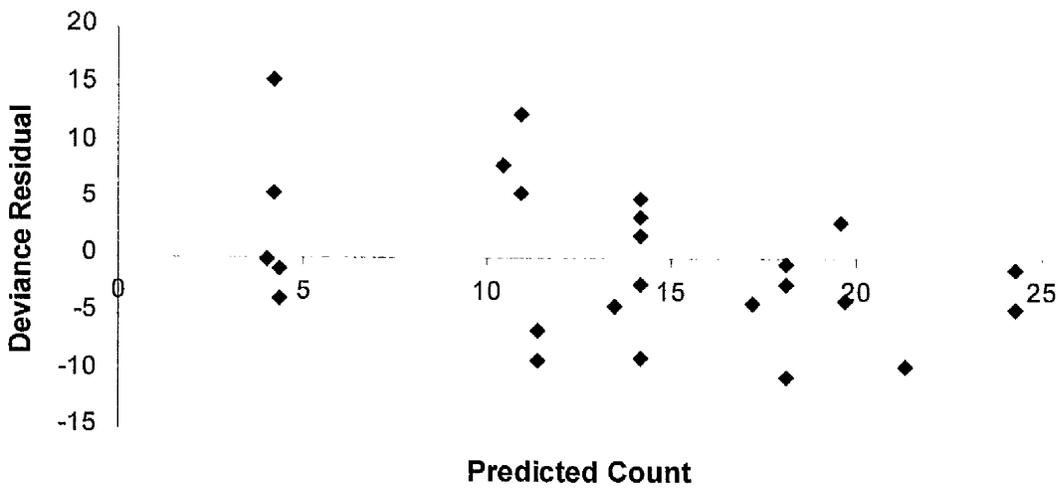
As well as dose being a possible predictor of response, AUC is also commonly used as a correlate with pharmacodynamic responses. The first model considered in terms of AUC is given in equation (3.5).

$$\text{logit}(\text{Pr}(Y_i \le k \mid b_i)) = \theta_k + \beta_1 AUC_i + \beta_2 gender_i + \beta_3 AUC_i gender_i + b_i, k = 0,1,2,3$$
$$Y_i \sim Mult(\gamma_0, \gamma_1, \gamma_2, \gamma_3, n_i)$$
$$\theta_k \sim N(0,1 \times 10^5), k = 0,1,2,3$$
$$\beta_j \sim N(0,1 \times 10^5), j = 1,2,3$$
$$b_i \sim N(0, \tau^{-1})$$
$$\tau \sim G(0.001,0.001)$$

$$(3.5)$$

As the AUC values are individual rat estimates and are continuous, a wider range of models could have been considered but instead, only simple linear models were considered. The results of this model are given in table 3.8.

Table 3.8. *Proportional odds model in AUC.*

	Mean	Standard Deviation	95% Credible Interval
θ_0	1.979	0.619	(0.799,3.384)
θ_1	3.362	0.638	(2.215,4.802)
θ_2	5.333	0.687	(4.101,6.838)
θ_3	7.128	0.786	(5.671,8.759)
β_1	-0.0544	0.0118	(-0.0776,-0.0313)
β_2	-1.192	0.893	(-2.757,0.6638)
β_3	0.0348	0.018	(-0.00573,0.0657)
σ	0.7586	0.306	(0.302,1.453)

The 95% credible interval for the interaction term does includes 0 so a new model was run with no interaction term. The results are in table 3.9 and the 95 % credible interval for the gender term includes 0. A new model was run with only AUC as a predictor and the results are also in table 3.9. A plot of the proportional odds model is given in figure 3.11 and individual category probabilities in figure 3.12.

Table 3.9. *Results of AUC+gender and AUC model.*

	AUC+gender model			AUC model		
	Mean	S.D.	95% C.I.	Mean	S.D.	95% C.I.
θ_0	1.216	0.825	(-0.833,3.037)	1.643	0.463	(0.815,2.695)
θ_1	2.581	0.836	(0.496,4.479)	3.008	0.485	(2.136,4.074)
θ_2	4.514	0.864	(2.458,6.506)	4.938	0.556	(3.984,6.181)
θ_3	6.302	0.945	(4.187,8.377)	6.711	0.684	(5.5,8.153)
β_1	-0.0392	0.0124	(-0.0838,-0.0107)	-0.0431	0.00793	(-0.0589,-0.0286)
β_2	0.426	0.607	(-0.663,1.791)	-	-	-
σ	0.989	0.348	(0.475,1.809)	0.9276	0.285	(0.48,1.574)

Figure 3.11. *Plot of AUC proportional odds model.*
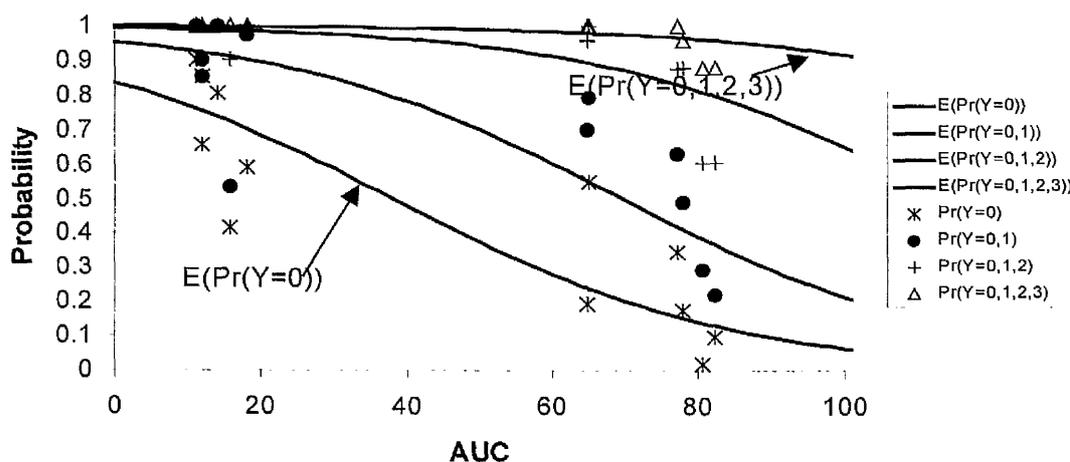
Figure 3.12. *AUC model in terms of individual category probabilities.*
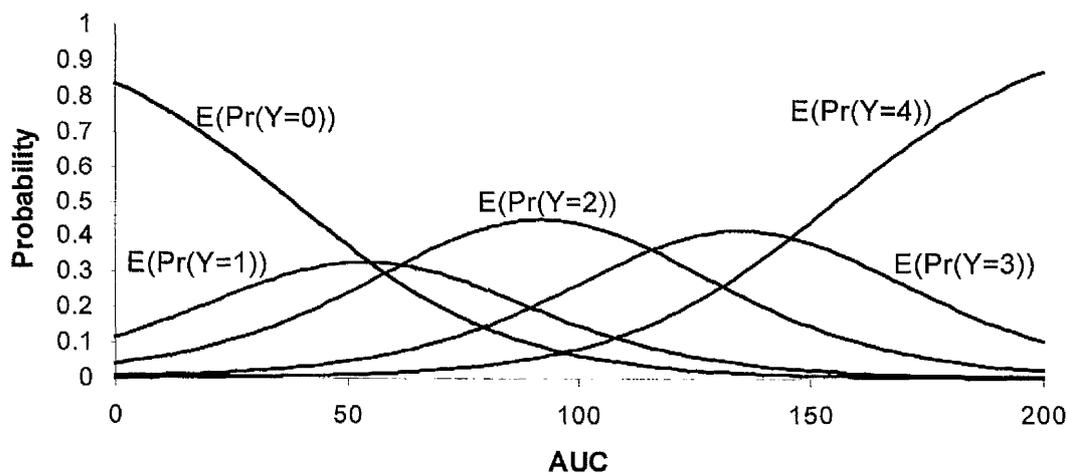
The cut points for the models given in equations (3.4) and (3.5) (dose and AUC models with interaction terms) are very similar as they correspond to the logit of the baseline cumulative probabilities. The models with dose+gender and AUC+gender have a difference of approximately 0.3 between the corresponding cut points and comparing the dose and AUC models, the difference is approximately 0.45.

As a Bayesian analysis was used, one important aspect of the analysis that can be checked is the sensitivity of the final parameter estimates to the prior distribution specification and random effects distributions. Only one aspect of sensitivity was checked and that was for the distribution of the random effects. In the previous models, the random effect was assumed to be normally distributed with mean 0 and variance τ^{-1}. A common alternative to this assumption is that the distribution has heavier tails due to possible out lying points so a Student-t distribution was specified, such as in population pharmacokinetic studies (Wakefield *et al* (1994)). As there are only 12 rats, a range of distributional assumptions might work but this was the only alternative that was tried. As the Student-t distribution requires an additional parameter, the degrees of freedom υ, a value was not explicitly specified but a discrete prior distribution was assigned to the parameter.

The model defined in equation (3.5) can be adapted to include a Student-t distribution on the random effect as defined in equation (3.6).

$$\text{logit}(\Pr(Y_i \le k \mid b_i)) = \theta_k + \beta_1 AUC_i + \beta_2 gender_i + \beta_3 AUC_i gender_i + b_i, k = 0,1,2,3$$

$$Y_i \sim Mult(\gamma_0, \gamma_1, \gamma_2, \gamma_3, n_i)$$

$$\theta_k \sim N(0, 1 \times 10^5), k = 0,1,2,3$$

$$\beta_j \sim N(0, 1 \times 10^5), j = 1,2,3$$

$$b_i \sim St_\upsilon(0, \tau^{-1})$$

$$\tau \sim G(0.001, 0.001)$$

$$\upsilon \sim \frac{i}{10}, where \quad i = 2,4,6,8,10,12,15,20,30,50$$

(3.6)

The distribution for υ is discrete in BUGS because of the difficulty in sampling from the conditional distribution for this parameter. The discrete nature of the prior is like sampling with the Griddy Gibbs sampler but in version 0.5 of BUGS, this was not implemented so it had to be specified explicitly. The results of this model are given in table 3.10 where the burn in was for 1500 iterations and the sample was 5000 iterations. Also in table 3.10 are the results for a similar model where AUC is replaced by dose.

Table 3.10. *Results of AUC and dose models with Student-t distribution on the random effect.*

	AUC+gender+AUC×gender model			dose+gender+dose×gender model		
	Mean	S.D.	95% C.I.	Mean	S.D.	95% C.I.
θ_0	1.929	0.653	(0.615,3.325)	1.82	0.619	(0.722,3.156)
θ_1	3.305	0.677	(1.975,4.779)	3.195	0.64	(2.75,4.554)
θ_2	5.268	0.723	(3.836,6.837)	5.149	0.685	(3.954,6.64)
θ_3	7.069	0.821	(5.471,8.761)	6.933	0.78	(8.597,6.887)
β_1	-0.0541	0.0112	(-0.08,-0.0336)	-0.0212	0.00436	(-0.0315,-0.0136)
β_2	-1.054	1.005	(-3.104,1.047)	-1.282	0.995	(-3.406,0.587)
β_3	0.0331	0.0176	$(-8.5 \times 10^{-4}, 0.071)$	0.0158	0.00645	(0.00349,0.0291)
σ	0.698	0.316	(0.238,1.461)	0.741	0.314	(0.26,1.501)
υ	13.9	9.042	(2,30)	10.94	7.925	(2,30)

The parameter estimates have not changed greatly by changing the random effects distribution from a normal to a Student-t distribution. The change in random effects

distribution has caused the parameter values to decrease in magnitude. The standard deviation of the random effects component has been reduced with the use of the Student-t distribution but whether this is the same reduction as the parameter values or because of a better description of the model is not known. The degrees of freedom parameter, υ was between 10 and 14 as well as in other analyses not reported here which corresponds to a distribution with heavy tails.

The logit link function is not the only link function that can be used for categorical data models. Two other common types of link function are the probit link function and the complementary log-log link function. These link functions are defined in equation (3.8).

$$complementary \log - \log = \log(-\log(1 - p))$$
$$probit = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} \exp(-\frac{1}{2}u^2)du \qquad (3.8)$$

For the probit link function, the integral in (3.8) is the distribution function of the standard normal random variable, U, so $p=\Pr(U \leq \xi)$ for some probability p. The probit and complementary log-log link functions were run with AUC as a linear predictor. The results compared to the logit link function are given in table 3.11 and figure 3.13. The model specification is the same as that in equation (3.5) but omitting the gender and interaction terms.

Table 3.11. *Logit, probit and complementary log-log models with AUC as a linear predictor.*

	Logit model		Probit model		Comp. log-log model	
	Mean (S.E.)	95% C.I.	Mean (S.E.)	95% C.I.	Mean (S.E.)	95% C.I.
θ_0	1.643 (0.463)	(0.815, 2.695)	0.924 (0.325)	(0.301, 1.642)	0.393 (0.271)	(-0.123, 0.995)
θ_1	3.008 (0.485)	(2.136, 4.074)	1.729 (0.334)	(1.084, 2.476)	1.25 (0.273)	(0.733, 1.885)
θ_2	4.938 (0.556)	(3.984, 6.181)	2.849 (0.357)	(2.159, 3.638)	2.341 (0.294)	(1.792, 3.014)
θ_3	6.711 (0.684)	(5.5, 8.153)	3.8 (0.394)	(3.05, 4.631)	3.136 (0.325)	(2.514, 3.876)
β_1	-0.0431 (0.00793)	(-0.0589, -0.0286)	-0.0246 (0.00602)	(-0.0368, -0.0127)	-0.023 (0.00621)	(-0.0357, -0.00986)
σ	0.928 (0.285)	(0.48, 1.574)	0.562 (0.169)	(0.313, 0.968)	0.575 (0.172)	(3.045, 0.98)

Figure 3.13. *Plot of logit, probit and complementary log-log model with linear AUC predictor for Pr(Y=0).*
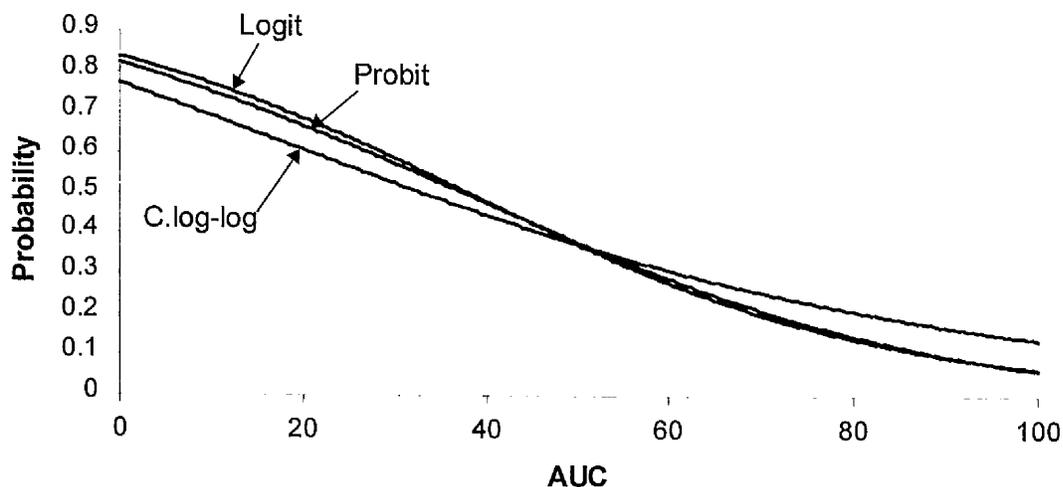
As can be seen from figure 3.13, there is virtually no difference between the logit and probit models. This is quite common when both link functions work well in fitting the data. It is common to get a scaling factor of between 1.6 and 2 when both the logit and

probit models fit the data. In this case, the scaling factor is approximately 1.75 between the parameters of the two models.

3.4 Comparison of Models

As discussed in section 2.9.3, one way of comparing non-nested models (or any models) is with the use of Bayes factors. To see which of these models best describe these data, a method such as Bayes factors is needed to show which model has the ability to predict response best. The first comparison was made between the best of the AUC models and the best of the dose models. This is a check to see whether including pharmacokinetic information through the AUC values is better at describing the pharmacodynamic variability than models not including such information. The dose+gender+dose×gender model was compared to the AUC model in BUGS using the method described by Carlin and Chib (1995). The model specification in BUGS requires a prior to be put on the model so that the sampling algorithm can choose which model to sample from. The results of the BUGS run to compare the two proportional odds models are given in table 3.12.

Table 3.12. *Results of model comparison between the dose+gender+dose×gender and AUC models.*

	Prior Probabilities	B_{21}
M_1: logit(dose+gender+dose×gender)	0.999	10126
M_2: logit(AUC)	0.001	

Kass and Raftery (1995) reported a scale for the Bayes factor that is on the same scale as the deviance and likelihood ratio test statistics. The scale is based on twice the

natural logarithm of the Bayes factor. The levels of model comparison are given in table 3.13.

Table 3.13. *Levels of evidence based on Bayes factors.*

$2\log_e(B_{21})$	(B_{21})	Evidence against model 1
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

The Bayes factor in table 3.12 is over 10,000 so there is very strong evidence that model 2 (AUC model) describes the data better than model 1 (dose+gender+dose×gender model).

Although there are such things as goodness of link tests (Pregibon (1980)), another way of comparing models with different link functions, is with the use of Bayes factors as again, non-nested models are being compared. As the previous comparison in table 3.12 showed, the AUC model is better so only models with an AUC linear term were compared across link functions. The Bayes factors for pairwise comparison between link functions are given in tables 3.14-4.16.

Table 3.14. *Results of model comparison between probit and logit model linear in AUC.*

	Prior Probabilities	B_{21}
M_1: probit(AUC)	0.5	13.61
M_2: logit(AUC)	0.5	

Table 3.15. *Results of model comparison between complementary log-log and logit model linear in AUC.*

	Prior Probabilities	B_{21}
M_1: cloglog(AUC)	0.999	4990
M_2: logit(AUC)	0.001	

Table 3.16. *Results of model comparison between complementary log-log and probit model linear in AUC.*

	Prior Probabilities	B_{21}
M_1: cloglog(AUC)	0.5	383.6
M_2: probit(AUC)	0.5	

As can be seen, the logit model is superior to the probit and complementary log-log models. The difference between logit and probit models is small but there is positive evidence in favour of the logit model which was suspected previously.

3.5 Discussion

The emphasis of this data analysis was to show how pharmacodynamic data of a categorical nature could be modelled. The pharmacodynamic and pharmacokinetic data allowed a model to be developed for the description of the drug's concentration-effect profile. As there was virtually no information on the drug's pharmacokinetics and pharmacodynamics, the value in this analysis was not in gaining a specific description of the pharmacokinetic/pharmacodynamic model for this drug, but how such a model might be obtained for categorical pharmacodynamic data with pharmacokinetic information. The pharmacodynamic data was not longitudinal but cross sectional, whereas the pharmacokinetic information was considered at steady state.

The modelling of the pharmacokinetics was of secondary importance but it was still important to be able to describe this data. This was required for predicting the pharmacodynamics through a parameter or summary measure of the pharmacokinetics, in this case the AUC measure (based on apparent clearance and dose). Whether the pharmacokinetic information was useful as a predictor for the pharmacodynamic data was examined in the comparison of the AUC and dose models in section 3.4. The two compartment first order absorption model at steady state used to describe the pharmacokinetic data was an adequate description of the data. As there was so little information in the absorption phase, the intravenous bolus administration model could have been used but instead, the absorption rate was fixed to 10 hr^{-1}. As this chapter was an exercise in how to analyse pharmacodynamic data, the model for the pharmacokinetic data was chosen to be the most suitable simple model.

The pharmacokinetic/pharmacodynamic modelling ideas used for the analysis of the data are not new. The belief that the use of pharmacokinetic information in the prediction of the pharmacodynamics will give a better correlation with the pharmacodynamic data than the use of dose alone has been the general idea for a long time (Levy (1985)). The use of dose as a predictor for the pharmacodynamics is like using some averaged measure of the pharmacokinetics over the population. The variability between individuals' pharmacokinetic information is not accounted for in dose alone so this metric is not expected to do as well as an individualised measure. The models used for the description of the data showed that the model using individual estimates of the AUCs appeared to be a better predictor of the pharmacodynamics than dose and gender. Similar types of analyses have been reported where the pharmacodynamics can be better described by the inclusion of individualised

pharmacokinetic parameters rather than dose such as in Moore and Theissen (1992) and Danhof *et al* (1992). Plasma concentration at the effect site was not computed as at steady state, the concentration in the plasma and at the effect site would be in equilibrium and no advantage would be gained in obtaining such data, but this is only for infusions.

The software used was NONMEM version IV and BUGS version 0.5. NONMEM was used in the estimation of the individual and population pharmacokinetic parameters. With the parameter values of apparent clearance obtained, the individual AUC estimates were put into BUGS to then model the pharmacodynamics. The pharmacokinetic parameters could have been estimated in BUGS but this would have required a discrete prior distribution to be put on the individual parameters which probably would have resulted in inadequate parameter estimates. NONMEM could have been used for the estimation of the pharmacodynamic model parameters but the distributions for categorical data are more easily specified in BUGS. As NONMEM is virtually a maximum likelihood package and BUGS a Bayesian package, the use of both methods to obtain a pharmacokinetic/pharmacodynamic model could be questioned on grounds that the two different methods might have produced very different answers had either of the packages been used for the whole analysis.

The use of the proportional odds model for the analysis of the categorical pharmacodynamic data has been described by Sheiner *et al* (1997), Mandema and Stanski (1996) and Sheiner (1994). The proportional odds model is a useful model for the description of categorical data but has not been used to any great extent yet in the development of pharmacokinetic/pharmacodynamic models with categorical data. The

101

assumption that odds are proportional between categories was not tested but not using such a constraint can lead to complicated models. If a category specific fixed effect was assigned to each gradient parameter, then this can lead to negative probabilities being obtained for being in a particular category. This can be seen graphically when the cumulative logistic regression curves cross. If this happens in the region of interest (such as between two dose groups) then this implies the model is nonsense. If the curves cross at a level that was outside the region of interest, then this problem can be ignored. The models considered were linear in the predictor variables, which made the analysis easier. For this particular data set, they probably were appropriate, especially for the dose model where there were only two dose groups available, but had more data been available then more complicated models could have been considered such as nonlinear models on the logit scale. Another approach that could have been considered was that of Karlsson *et al* (1998). The approach used in their paper was to describe an extra model between the observed pharmacokinetic and pharmacodynamic models. The extra model was a model based on an unobserved effect linking the pharmacokinetics and the observed pharmacodynamics. The model for the unobserved effect would have to be postulated and a possible range of models tested to see what model was appropriate. This could have been useful for this data set where the unobserved model could have been the E_{max} model as a function of concentration and then the area under the indirect model used as a predictor for the observed effect. This would have brought some physiological interpretation to the model, as a model that is well understood would have linked the observed pharmacokinetics and pharmacodynamics.

All the models used included a random effect term. One random effect was assigned to each model and it was always additive on the cut points on the logit scale. The reason

for the random effect being included on the logit scale is so that the predicted probabilities remain on the scale [0,1]. The random effect acts as a shift for the individual from the population average. When on the probability scale, the random effect acts like a shift left and right determining where the logistic curve crosses the probability axis. The estimates of the standard deviation for the dose and AUC models show that there is a reduction in the variability when AUC was used as a predictor (1.192 and 0.9276). If this is due to using the individual estimates of AUC then this shows that more variability is accounted for when using the pharmacokinetics to predict the pharmacodynamics.

The use of the Bayes factors to determine which model fits the data best is a general method for model comparison. Within BUGS, it is an easy method to implement and can be used for the comparison of any models to describe a particular data set. Although, this technique was performed using a Bayesian package, the idea of simulation to determining model choice and adequacy is something that can be applied to any form of modelling paradigm. Similar methods have been used in the frequentist literature but have been given the name of posterior predictive checks. These methods do not require any asymptotic theory as samples are being obtained from the posterior distribution through the full conditional distributions. As well as being applied to pharmacodynamic models, there is no reason why this method can not be applied to any area of population pharmacokinetic/pharmacodynamic modelling.

4 Toxicokinetic Data Set II

4.1 Introduction

The data analysed in this chapter comprises 10 data sets: 9 preclinical studies and 1 phase I study. The compound never progressed any further than the first phase I study due to toxicity.

The compound, PD-142676 is an acetylcholinesterase inhibitor and potential cognition activator. The compound had a proposed use in the treatment of Alzheimer's disease. The chemical structure is 1,3-dichloro-6,7,8,9,10,12-hexahydro-azepino[2,1-b]quinazoline monohydrochloride and a molecular weight of 305.637.

The 10 studies are given two research report numbers, one corresponding to the plasma concentration data collected and the other research report number for all other data collected, such as clinical signs, hematology, biochemistry, urinalysis, pathology and histopathology. The research report numbers are given in table 4.1 for reference.

The data for each species were analysed separately. Within each species, the pharmacokinetic analyses of each data set will be reported and then the pharmacodynamic analyses.

All of the analyses were carried out using NONMEM Version V on a Hewlet Packard work station using the UNIX operating system.

Table 4.1. *Research report numbers for studies on PD-142676.*

Species	Study Duration	PK Data	Other Data
Wistar Rat	2 Weeks	764-01891	745-02051
	4 Weeks	764-01978	250-01686
	4 Weeks	764-02167	745-02222
	13 Weeks	764-02041	250-01696
Beagle Dog	2 Weeks	764-02152	745-02250
	4 Weeks	764-02185	745-02251
Cynomolgus Monkey	2 Weeks	764-01936	745-02083
	4 Weeks	764-02162	745-02236
	13 Weeks	764-02064	250-01694
Human	Single Dose	744-00357	-

4.2 Rat Studies

There are four Wistar rat studies, ranging from a 2 week to a 13 week oral toxicity study. In studies 764-01891, 764-01978 and 764-02167, the compound PD-142676 was given orally, as this was the intended route of administration for humans. The drug was administered by oral gavage using a stainless steel oral dosing cannula or infant feeding tube attached to a hypodermic syringe. The drug was administered as a suspension (in 0.5% aqueous methylcellulose) daily on a mg active drug moiety/kg body weight/day basis. The dose volume was 20 mL/kg. Control group animals received vehicle only at the same dose volume as the treated animals. Study 764-02167 was different to the other three rat studies in that PD-142676 was administered as a food admixture.

Dose selection criteria and rat numbers will be given in each study section as these altered from study to study.

4.2.1 Pharmacokinetic Analyses of Rat Data Sets

4.2.1.1 Study RR-764-01891 – 2 Week Study

4.2.1.1.1 Study Design

For toxicokinetic evaluation, 3 different dosing groups of rats per gender were given PD-142676 as an oral dose solution once daily for 2 weeks. The doses were selected based on an acute toxicology study in rats. In 2 female rats, single oral doses of PD-142676 were administered 30, 50, 70, 90, 100 and 300 mg/kg. Deaths occurred within 180 minutes post-dose in 1 animal at 70 mg/kg, 1 animal at 100 mg/kg and 2 animals at 300 mg/kg. In each case, neurologic signs (tremor, hyperactivity, hypertonia and convulsion) preceded death. No effects were noted at 30 and 50 mg/kg. Subsequently, 3 males and 3 females were given 90 mg/kg. Salivation was the only sign elicited in all males and 1 female. The other 2 females died within 180 minutes post-dose following tremors and convulsions. Based on this information, 60 mg/kg administered for 2 weeks was expected to produce neurologic signs and was chosen as the high dose. A low dose of 15 mg/kg was expected to produce no clinical signs of toxicity. A mid-dose of 30 mg/kg was expected to be intermediate in toxicity. The numbers in each dosing group was 30 per group per gender and an additional 5 rats per gender in the placebo group.

On Day 14, plasma samples were obtained from 5 rats per gender in each dose group at each of 6 time intervals: 2, 4, 8, 12, 18 and 24 hours post-dose. This study resulted in just one plasma sample per rat. Plasma concentrations were determined with a validated HPLC method. The limit of quantification was 0.01 µg/mL.

4.2.1.1.2 Pharmacokinetic Analysis

The total number of concentrations above and below the limit of quantification (BLQ) are given in table 4.2.

Table 4.2. *Concentration measurements for 2 week toxicity study in Wistar rats.*

	0 mg/kg	15 mg/kg	30 mg/kg	60 mg/kg	Total
Male	0(5)	14(16)	18(12)	19(8)	51(41)
Female	0(5)	11(19)	16(14)	20(10)	47(48)
Total	0(10)	25(35)	34(26)	39(18)	98(89)

(.) BLQ

Out of a total of 187 possible plasma concentrations, there were 98 above the limit of quantification. Only values above the limit of quantification were used in the modelling procedure.

Preliminary analyses indicated that the half life, $t_{1/2}$, ranged from 2 to 6 hours and increased with dose. This is characteristic of drugs that exhibit nonlinear pharmacokinetics. As there is so little data (one observation from each rat and a total of 98 observations), it was considered unlikely that it would be possible to fit a complicated pharmacokinetic model to these data. To simplify the analysis and to allow

reasonable parameter estimation, a one compartment, first-order absorption model at steady state was considered. As the half life was between 2 and 6 hours, after 14 days, the concentrations were assumed to be at steady state. The structural model equation is given in equation (4.1).

$$E(C) = \frac{F.D.k_a}{Cl - Vk_a} \left(\frac{e^{-k_a t}}{(1 - e^{-k_a \tau})} - \frac{e^{-\frac{Cl}{V}t}}{(1 - e^{-\frac{Cl}{V}\tau})} \right)$$

(4.1)

$E(C)$ is the expected concentration where C denotes concentration. F is the bioavailability of the compound, D is the dose, V is volume of distribution, Cl is clearance and k_a is the absorption rate constant. The duration between doses is given by τ and time is t. The first error model considered for these data was the exponential error model given on the log scale in equation (4.2).

$$\log(C) = E(\log(C)) + \eta$$
$$\mathrm{var}(\eta) = \omega^2$$

(4.2)

As there were no intravenous data, the bioavailability term F can not be estimated so clearance and volume estimates become apparent parameter estimates, i.e. normalised by bioavailability to become Cl/F and V/F. There is only one data point per rat and so the residual term in equation (4.2) is an interindividual error term as there are no repeated measurements within a rat. The results obtained using this model specification and using the FO method are given in table 4.3 as model (1). If the standard errors are not shown in any of the tables, then they were unable to be estimated.

The value of k_a is very high at 45.9 virtually corresponding to a bolus administration of PD-142676. This parameter is probably being estimated very poorly due to the lack of data immediately after dosing. Also the correlation between the estimates of Cl/F and k_a is 0.986 suggesting some identifiability problems. This could be due to not using a

more appropriate model specification. To see if the error model fitting could be improved, an additive interindividual error model was considered and the results are in table 4.3 as model (2). The objective function was increased by approximately 50 points suggesting a worse description of the variability around the mean pharmacokinetic model. The comparison of the error models by finding the difference in the objective function values is an empirical way of comparing models as they are not based on asymptotic distributional theory. The values of Cl/F and V/F did not change much but the value of k_a was reduced by an order of magnitude. This again is probably due to the lack of information in the time immediately after the drug is given. The estimates of the half life for model (1) and (2) are 3.8 and 3.04 hours respectively which is in the region expected from the preliminary analyses.

The situation of one sample per animal in preclinical studies has been discussed by Ette *et al* (1995) and McArthur (1988) and they have shown that the estimation of random effects in such models are a problem as the interindividual and residual variability can not be partitioned. As it is not possible to estimate both interindividual and residual variability at the same time in such settings, then fitting a model with the random effects on the parameters themselves may prove a better option for describing the data. This gives a measure of the variability of the pharmacokinetic parameters. Keeping the same structural model, the error model is now assigned to the parameters rather than the model. The error model is now given in equation (4.3).

$$Cl = E(Cl)e^{\eta_{Cl}}, V = E(V)e^{\eta_V}$$
$$\text{var}(\eta_{Cl}) = \omega_{Cl}^2, \text{var}(\eta_V) = \omega_V^2$$

(4.3)

There is no random effect on the absorption constant as estimating this parameter has proved difficult even in the simplest of settings. The results are given in table 4.3 as

model (3). This model reduced the objective function value by approximately 20 points when compared to the previous best model (model (1)). The interindividual variability for Cl/F has an approximate coefficient of variation of 0.6% while the coefficient of variation for V/F is approximately 732%. These values would appear to be inaccurate, as there is not much information in the data and probably cannot be estimated well. To try and improve the fit of the model to the data, an additive error model was tried on the parameters (model (4)). This resulted in an approximate 110 increase in the value of the objective function from model (3) and hence is much worse at fitting the data. From these NONMEM runs, it appears that a proportional error model best describes the variability around the one compartment first-order absorption model.

It was noticed in the preliminary analyses that there could be nonlinear pharmacokinetics as clearance appeared to be dose dependent. Instead of fitting a Michaelis-Menton type equation that would be computationally difficult, a fixed effect model was used, such that at each dose level separate estimates of Cl/F, V/F and k_a were obtained for each dose. A proportional error model was used and the results are given in table 4.3 as model (5). The results show that apparent clearance is decreasing with increasing dose implying nonlinear pharmacokinetics. This model reduced the objective function value by 35 points from the previous best model (3). A nonlinear pharmacokinetic model was not tried due to the lack of information in the data.

The models used until now have involved the same structural model but different random effects models. A two compartment model is unlikely to improve the fit but a different absorption model was considered. Although PD-142676 was given by oral gavage, a zero-order absorption model was considered as an alternative to the first-order

absorption model. The one compartment zero-order absorption model at steady state is given in equation (4.4).

$$E(C) = \begin{cases} \dfrac{F.D}{T_{inf}.Cl}[(1-e^{-\frac{Cl}{V}T_{inf}})e^{-\frac{Cl}{V}(t-T_{inf})}\dfrac{e^{-\frac{Cl}{V}T_{inf}}}{(1-e^{-\frac{Cl}{V}T_{inf}})}+(1-e^{-\frac{Cl}{V}t})], t \leq T_{inf} \\[4ex] \dfrac{F.D}{T_{inf}.Cl}[(1-e^{-\frac{Cl}{V}T_{inf}})\dfrac{e^{-\frac{Cl}{V}(t-T_{inf})}}{(1-e^{-\frac{Cl}{V}T_{inf}})}], t > T_{inf} \end{cases} \qquad (4.4)$$

When this model is fitted to the data (model (6)) with exponential random effects on the parameters, *Cl/F* and *V/F*, the objective function is increased by 35 points as compared to the same model with first-order absorption. It appears in this case that the first-order absorption model is better at describing the data. Due to the possible nonlinearity of the pharmacokinetics, a fixed effect was used at each dose level for each parameter (model (7)) in the zero-order absorption model. This resulted in the clearance values once again increasing as the dose decreased. The objective function value was virtually identical to that of the first-order absorption model version (model (5)).

From these NONMEM runs, it appears that the model best describing these data is the fixed effects model on the parameters at each dose level. Whether a zero or first-order absorption model is more appropriate is impossible to distinguish at this stage as the true error model could be a mixture of both absorption processes. Plots of the model, residuals and weighted residuals are given in figures 4.1-4.5.

Figure 4.1. *One compartment first-order absorption model for dose 15 mg/kg.*
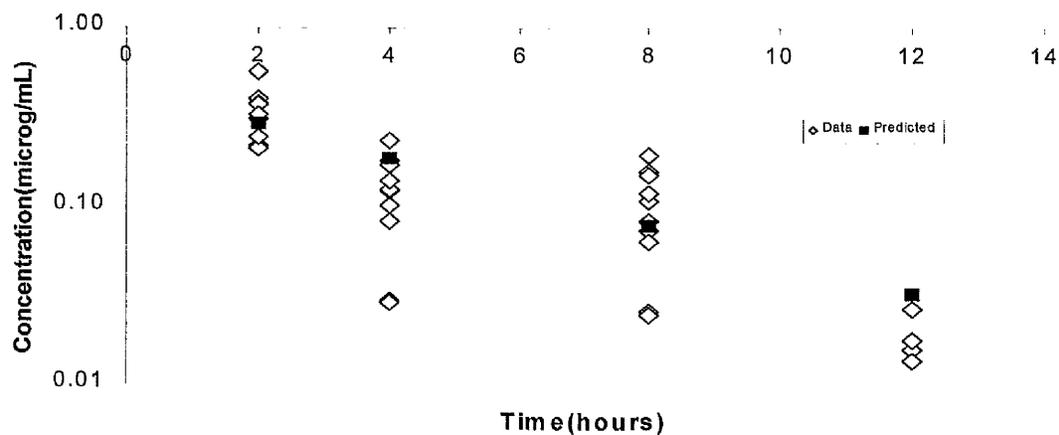
Figure 4.2. *One compartment first-order absorption model for dose 30 mg/kg.*
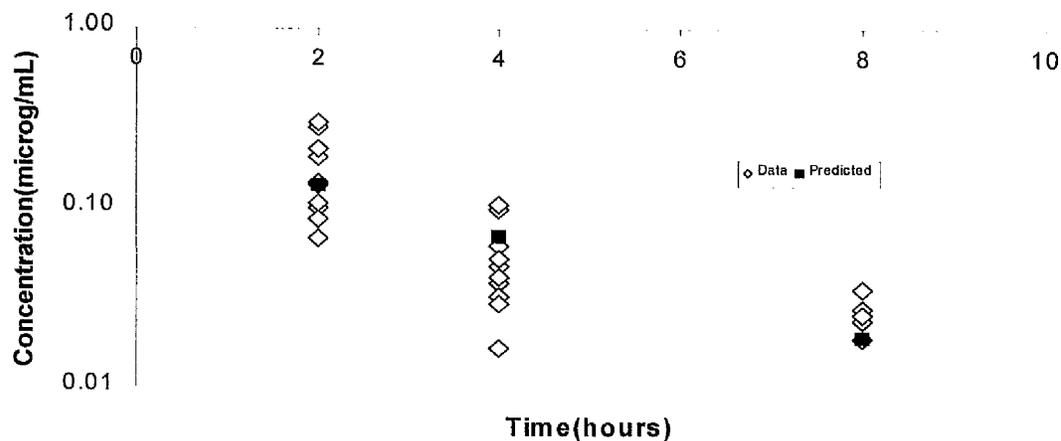
Figure 4.3. *One compartment first-order absorption model for dose 60 mg/kg.*
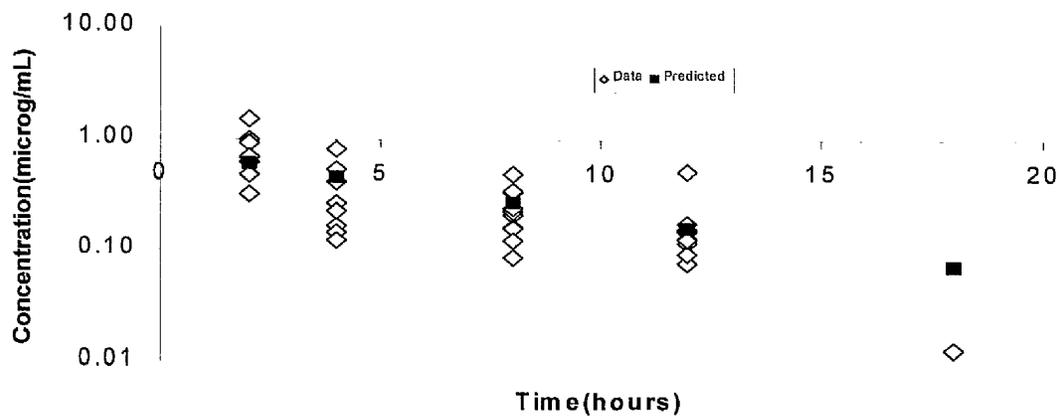
Figure 4.4. *Residual versus predicted plot for model (5) in 2 week rat study 764-01891.*
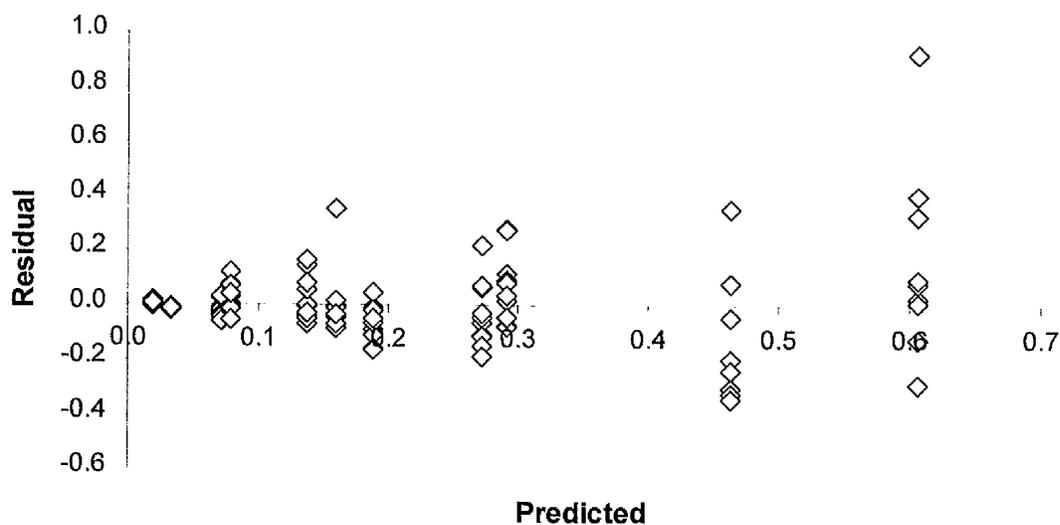
Figure 4.5. *Weighted residual versus predicted plot for model (5) in 2 week rat study 764-01891.*
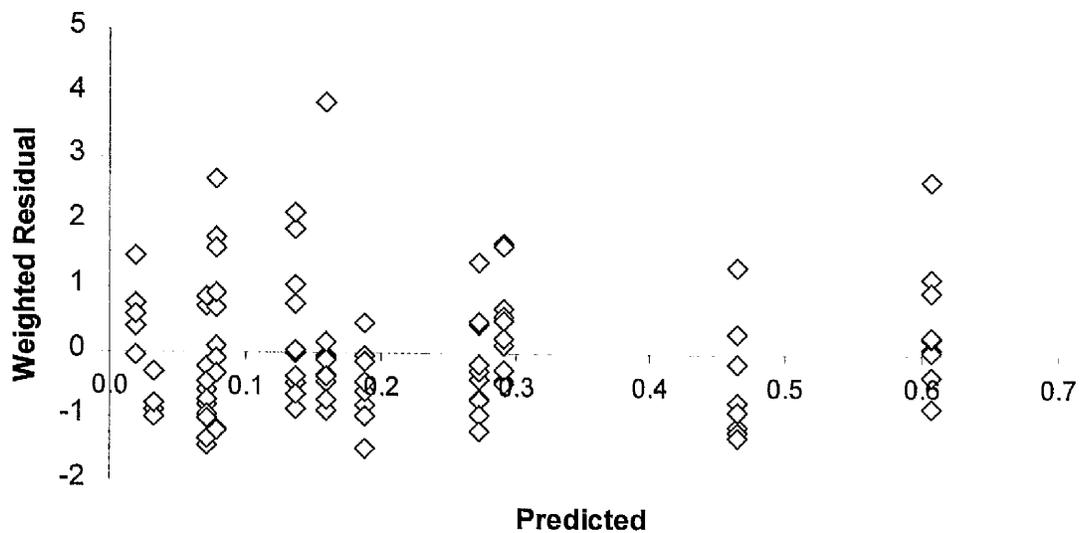
Table 4.3. *Results of NONMEM analyses for 2 week Wistar rat study 764-01891.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	T_{inf} (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	-315.858	10.7 (0.819)	59.2 (5.92)	45.9 (2.26)	-	0.3[a]	-
(2)	-266.941	13.2 (1.33)	57.8 (12.5)	4.34 (0.061)	-	0.0242[b] (0.008)	-
(3)	-335.12	15.4 (2.01)	10.6 (2.1)	0.146 (0.061)	-	4.3×10^{-5}[c] (1.7×10^{-5})	53.6[d] (56.8)
(4)	-246.638	6.54	139	0.0101	-	27.2[e]	1.1×10^{-4}[f]
(5)	-371.345	[15]18.7 [30]14.7 [60]10.7	[15]0.476 [30]0.194 [60]0.133	[15]0.324 [30]0.219 [60]0.133	-	0.324[a]	-
(6)	-301.083	14.3 (1.08)	46.7 (4.97)	-	2.4 (0.22)	0.449[c] (0.102)	0.226[d] (0.0597)
(7)	-371.345	[15]26.1 (2.5) [30]17.5 (0.806) [60]11.9 (1.13)	[15]80.5 (12.7) [30]79.8 (8.85) [60]89.4 (16.1)	-	[15]2 (4×10^{-4}) [30]1.58 (0.96) [60]1.83 (2.18)	0.324[a] (0.0476)	-

[a] exponential error on the model.
[b] additive error on the model.
[c] proportional error on Cl/F.
[d] proportional error on V/F.
[e] additive error on Cl/F.
[f] additive error on V/F.
[.] Dose group.

4.2.1.2 Study RR-764-01978 – 4 Week Study

4.2.1.2.1 Study Design

For toxicokinetic evaluation, 3 groups of 10 rats per gender were given PD-142676 as an oral dose solution once daily for 4 weeks. 5 rats per gender were assigned to a placebo group. The doses to be administered were 0, 30, 60 and 90 mg/kg and were based on preliminary results from the 2 week oral toxicity study 764-01891. The 30 mg/kg dose was expected to produce transient clinical signs in some animals. The 60

mg/kg dose was expected to produce moderate and probably transient clinical signs in most animals but no deaths. The 90 mg/kg dose was expected to produce moderate to severe clinical signs in most animals and possibly some deaths. The rats were randomised to one of the dose groups and then randomised within each dose group.

On day 6 of week 4, blood samples were obtained from 5 rats per gender in the 30 and 60 mg/kg dose groups and 2 male and 3 female rats in the 90 mg/kg dose group at 0 (predose), 1 and 4 hours post-dose. Another group of 5 rats per gender in the 30 and 60 mg/kg dose groups and 2 male and 3 female rats in the 90 mg/kg dose group were sampled at 0.5, 2 and 6 hours post-dose. Blood samples were collected from control rats at one hour post-dose. Plasma concentrations were determined with a validated HPLC method. The minimum limit of quantification was 0.01 µg/mL.

4.2.1.2.2 Pharmacokinetic Analysis

The total number of concentrations above and below the limit of quantification (BLQ) are given in table 4.4.

Table 4.4. *Concentration measurements for 4 week toxicity study 764-01978.*

	0 mg/kg	30 mg/kg	60 mg/kg	90 mg/kg	Total
Male	0(5)	26(4)	25(2)	11(1)	62(15)
Female	0(5)	26(4)	28(2)	12(2){3}[1]	66(13){3}[1]
Total	0(10)	52(8)	53(7)	23(3){3}[1]	128(28){3}[1]

(.) BLQ
{.} Rat died
[.] missing due to insufficient sample

Out of a possible 160 concentration measurements, 128 are above the limit of quantification. In the female 90 mg/kg group, three rats had excessively high concentrations possibly due to an incorrect dose being administered. These three rats died and their concentration measurements were withdrawn from the analysis which left 121 concentration measurements.

As there are repeated measurements in each rat, it is possible to distinguish between interindividual and residual variability. The first model to be fitted to the data was a 1 compartment first-order absorption model at steady state with exponential error models on the parameters CL/F and V/F and an exponential residual error model. The FO method was used. The results are given in table 4.5 as model (1). The interindividual variance values are very close to zero which would mean that any empirical Bayes estimates of the individual parameters would be virtually identical to the population values. It appears that all of the variability is being estimated as part of the residual variance. To see if any other error models would describe the variability better, three more models were run in NONMEM. Model (2) specified additive random effects on both the parameters and the residual term, model (3) has additive errors on the parameters and exponential error on the residual term and model (4) has exponential errors on the parameters and additive error on the residual term. The results are given in table 4.5. The only significant difference is due to changing the interindividual error model. An additive error model results in an increase of the NONMEM objective function value by approximately 6 points. In all cases, the interindividual error terms are poorly estimated, probably due to the lack of data.

To see if a different structural model could improve the fit, a one compartment zero-order absorption model at steady state was considered. Models (5) and (6) both have exponential error models on the parameters and an exponential and additive error model for the residual term respectively. The objective function values did change appreciably from the first-order absorption model.

To try and gain a better estimate of the interindividual variability, a combination of the data from studies 764-01891 (2 week) and 764-01978 (4 week) were used. The total number of concentration measurements available were 223 from 146 rats. Even though the first study corresponded to a 2 week study and the second to a 4 week study, the profiles were assumed to be at steady state so the data were pooled as though the data had been collected under the same conditions.

Table 4.5. *Results of NONMEM analyses for 4 week Wistar rat study 764-01978.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	T_{inf} (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)	σ^2 (S.E.)
(1)	-141.28	5.17 (0.98)	102 (35.5)	10 (1.5×10^{-6})	-	2.4×10^{-10}[a] (1.2×10^{-9})	0.0049[a] (0.187)	0.323[a] (0.0427)
(2)	-135.11	4.05 (0.246)	392 (100)	9.91 (0.17)	-	5.1×10^{-8}[b] (6.6×10^{-7})	2340[b] (8×10^4)	0.112[b] (0.0203)
(3)	-141.49	4.5 (1.07)	100 (70.6)	0.44 (0.151)	-	2.2×10^{-4}[b] (0.00113)	3560[b] (1×10^4)	0.312[a] (0.0452)
(4)	-135.18	3.94	400	0.444	-	7.3×10^{-10}[a]	7×10^{-4}[a]	0.117[b]
(5)	-141.48	5.25	102	-	0.499	5.5×10^{-11}[a]	8.17[a]	0.328[a]
(6)	-136.28	3.75	3.75	-	4	3.34×10^{-7}[a]	1.71[a]	0.113[b]

[a] Exponential error model.
[b] Additive error model.

Table 4.6. *Results of NONMEM analyses for combined 2 and 4 week Wistar rat studies.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	σ^2 (S.E.)
(1)	-462.163	9.64 (0.532)	70 (4.8)	19.2 (17.7)	0.0153 (0.0267)	2.3×10^{-12} (2.3×10^{-11})	0.442 (0.0457)
(2)	-396.183	10.1 (0.523)	49.2 (3.31)	1(fixed)	0.0238 (0.0356)	0.301 (0.131)	0.467 (0.0559)
(3)	-381.709	9.58 (0.505)	49.4 (3.6)	1(fixed)	0.0851 (0.0189)	-	0.541 (0.0736)

All error models are exponential.

Only the first-order absorption model was considered for this set of data as the zero-order model did not improve the description of the absorption phase. Model (1) in table 4.6 is a one compartment model at steady state with exponential errors on both the parameters and residual term. This model appears to give a better estimate of the interindividual variability on apparent clearance but not on apparent volume. The absorption rate constant at 19.2-hr^{-1} virtually corresponds to a bolus dose. The absorption rate constant was next held fixed at 1-hr^{-1} (model (2)) as this would hopefully allow the other parameters to be estimated with more confidence even though there is no information at the absorption stage. Model (2) resulted in an increase of 66 points in the objective function. By removing the variance term on apparent volume the objective function increased by a further 15 points but allowed the interindividual variance to be estimated better. Residual plots are given in figures 4.6-4.8 for model (3) of table 4.6. The dose effect observed in 2 week study 764-01891 was not included in the combined model. As the nonlinearity of the 2 week model was not accounted for in the combined model, this would have lead to incorrect estimates of the parameters. In table 4.6, apparent clearance seems independent of holding k_a fixed but apparent volume appears to be positively correlated giving a much reduced estimate of apparent volume.

The estimates of apparent clearance given in table 4.6 is approximately a doubling of the estimates obtained in table 4.6. The estimates of apparent volume based on only the 4 week data are poorly estimated whereas the estimates obtained from the combined data set are better estimated given the correlation with k_a.

Figure 4.6. *Residual versus time plot for 2 and 4 week combined Wistar rat data.*

Time(hours)
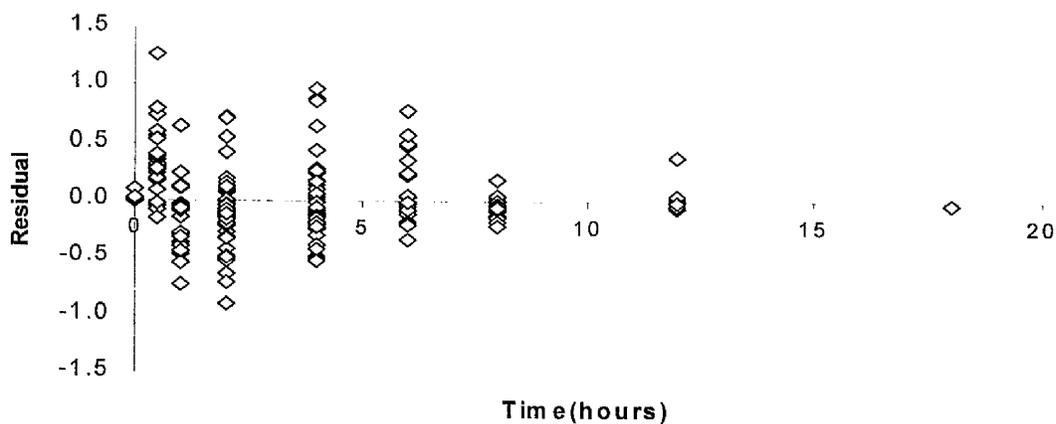
Figure 4.7. *Residual versus predicted plot for 2 and 4 week combined Wistar rat data.*

Population Predicted
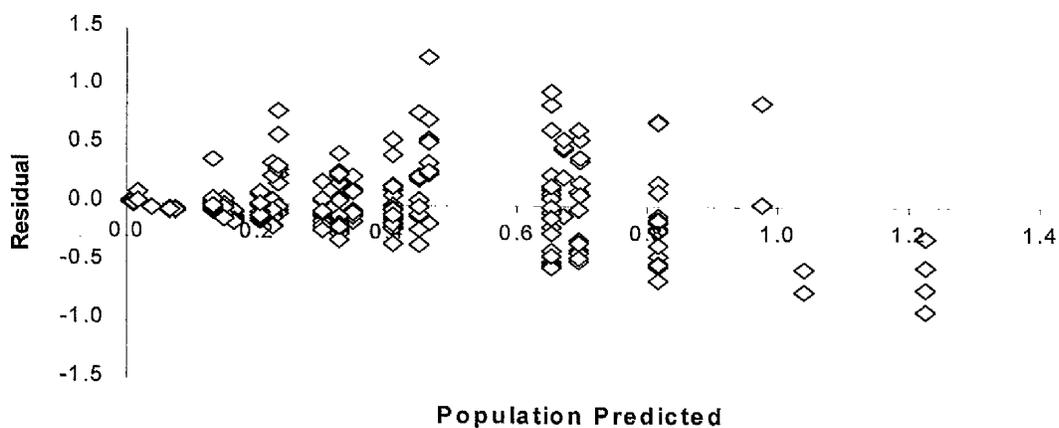
Figure 4.8. Weighted *Residual versus predicted plot for 2 and 4 week combined Wistar rat data.*
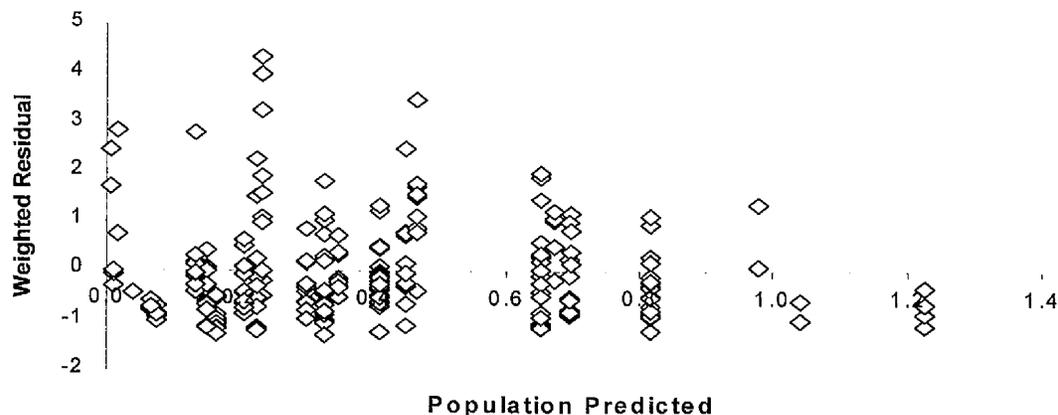
4.2.1.3 Study RR-764-02041 – 13 Week Study

4.2.1.3.1 Study Design

For toxicokinetic evaluation, 3 groups of 10 rats per gender were given PD-142676 as an oral dose solution once daily for 13 weeks. 5 rats per gender were assigned to a placebo group. The doses to be administered were 0, 5, 15 and 30 mg/kg and were based on preliminary results from the 2 week and 4 week oral toxicity studies. The 5 mg/kg dose was expected to be a no-effect dose. The 15 mg/kg dose was expected to produce moderate transient clinical signs in most animals. The 30 mg/kg dose was expected to produce moderate to severe clinical signs in most animals and possibly some deaths. The rats were randomised to one of the dose groups and then randomised within each dose group.

During week 13, blood samples were obtained from 5 rats per gender in each dose group at 0.5, 2 and 6 hours post-dose. Another group of 5 rats per gender in each dose group were sampled at 1, 4 and 24 hours post-dose. Blood samples were collected from placebo control rats at 1 hour post-dose. Plasma concentrations were determined with a validated HPLC method. The limit of quantification was 0.01 μg/mL.

4.2.1.3.2 Pharmacokinetic Analysis

The total number of concentrations above and below the limit of quantification (BLQ) are given in table 4.7.

Table 4.7. *Concentration measurements for 13 week toxicity study 764-02041.*

	0 mg/kg	5 mg/kg	15 mg/kg	30 mg/kg	Total
Male	0(5)	25(5)	25(5)	19(5)	69(15)
Female	0(5)	23(7)	25(5)	25(5)	73(17)
Total	0(10)	48(12)	50(10)	44(10)	142(32)

(.) BLQ

Out of a possible 174 concentration measurements, 142 are above the limit of quantification. There were repeated measurement data so individual parameters could be estimated. A one compartment first-order absorption model was fitted to the data with a range of error models. The results are given in table 4.8 (models (1)-(4)). The best fitting model out of the models (1)-(4) was that with exponential errors on the parameters Cl/F and V/F and the residual term. Specifying a fixed effect model on the dose level (models (5) and (6)) did not seem to improve the fit of the model appreciably as the objective function value decreased by only 10 points. Model (6) shows that there

was a decrease in apparent clearance as dose increased. Although this suggests nonlinear pharmacokinetics, the model fit was not significantly better in terms of the objective function than model (2). The apparent volume estimates appeared poorly determined. There did not appear to be any correlation between apparent clearance and dose and so linear pharmacokinetics could be assumed for this data set. This could be explained by the difference in the dosing levels between the two studies, 764-01891 and 764-02041, and nonlinear pharmacokinetics only becoming apparent at higher doses. A one compartment model with zero-order absorption was also considered (model (7)) with exponential error terms but this did not improve the fit to the data at all.

Table 4.8. *Results of NONMEM analyses for 13 week Wistar rat study 764-02041.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	T_{inf} (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	σ^2 (S.E.)
(1)	-271.54	4.68 (1.85)	0.452 (0.214)	0.0617 (0.059)	-	0.141[a] (1.4)	1130[a] (1970)	0.016[b] (0.006)
(2)	-421.21	6.19 (5.02)	56.4 (32.6)	18.2 (18.8)	-	0.43[a] (1.4)	0.151[a] (0.495)	0.412[a] (0.372)
(3)	-421.18	6.09	56.7	46.2	-	15[b]	476[b]	0.417[a]
(4)	-342.96	6.67 (1.14)	58 (9.41)	7.21 (9.41)	-	25.5[b] (28.1)	1200[b] (1210)	0.0212[b] (0.009)
(5)	-431.7	[5]3.96 (2.44) [15]3.63 (2.48) [30]3.76 (2.72)	[5]75.1 (87.4) [15]0.312 (0.175) [30]0.504 (0.251)	[5]6.13 (1.95) [15]0.013 (0.0746) [30]0.169 (0.112)	-	4×10^{-4}[b] (4×10^{-4})	-	0.49[b] (0.0264)
(6)	-431.3	[5]7.17 (2.27) [15]4.31 (1.05) [30]3.76 (2.72)	[5]0.336 (0.11) [15]0.348 (0.336) [30]0.249 (0.0753)	[5]0.144 (0.07) [15]0.062 (0.0369) [30]0.015 (0.0802)	-	0.043[a] (0.0407)	-	0.481[a] (0.0786)
(7)	-420.96	8.32	86.6	-	42.3	1×10^{-5a}	0.0303[a]	0.499[a]

[a] Exponential error model.
[b] Additive error model.
[.] Dose.

To try and improve the estimates of the interindividual variance parameters, all three data sets were combined (studies 764-01891, 764-01978 and 764-02041). This combined data set has 365 concentration measurements from 204 rats. The models considered were the one compartment first-order absorption model at steady state ((1)-(4)) and a one compartment zero-order absorption model at steady state (5). The results are given in table 4.9. The best of these models was the first-order absorption model with exponential error model on the interindividual and residual terms. Residual plots for model (1) are given in figures 4.9-4.11.

Table 4.9. *Results of NONMEM analyses for combined 2, 4 and 13 week Wistar rat studies.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	T_{inf} (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	σ^2 (S.E.)
(1)	-809.203	8.45 (0.61)	58.4 (4.41)	50 (7.6×10^{-8})	-	0.135^a (0.264)	$6 \times 10^{-11 a}$ (5×10^{-10})	0.503^a (0.113)
(2)	-671.03	7.75 (0.7)	3.12 (0.61)	0.103 (0.021)	-	0.126^a (0.0372)	4.71^a (2.17)	0.0339^b (0.0078)
(3)	-645.91	9.43 (2.44)	63.8 (12.7)	15.4 (11.3)	-	30.1^b (47.5)	1020^b (179)	0.0402^b (0.0161)
(4)	-795.192	10.6 (3.52)	64 (11)	10.6 (1.98)	-	75.9^b (162)	1070^b (2060)	0.46^a (0.845)
(5)	-780.77	7.24	97.2	-	0.501	0.256^a	1.99^a	0.519^a

[a] Exponential error model.
[b] Additive error model.

Figure 4.9. *Residual versus time plot for 2, 4 and 13 week combined Wistar rat data.*
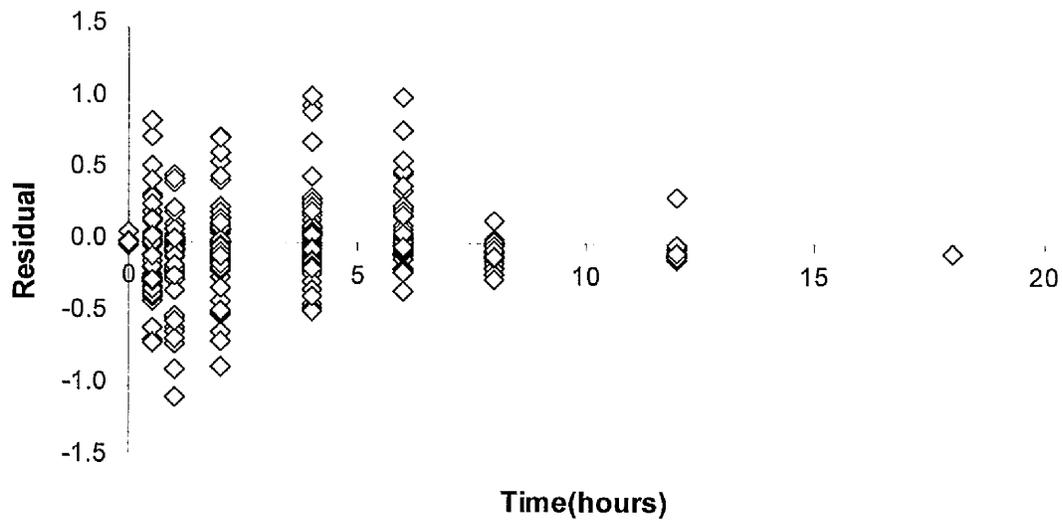
Figure 4.10. *Residual versus predicted plot for 2, 4 and 13 week combined Wistar rat data.*
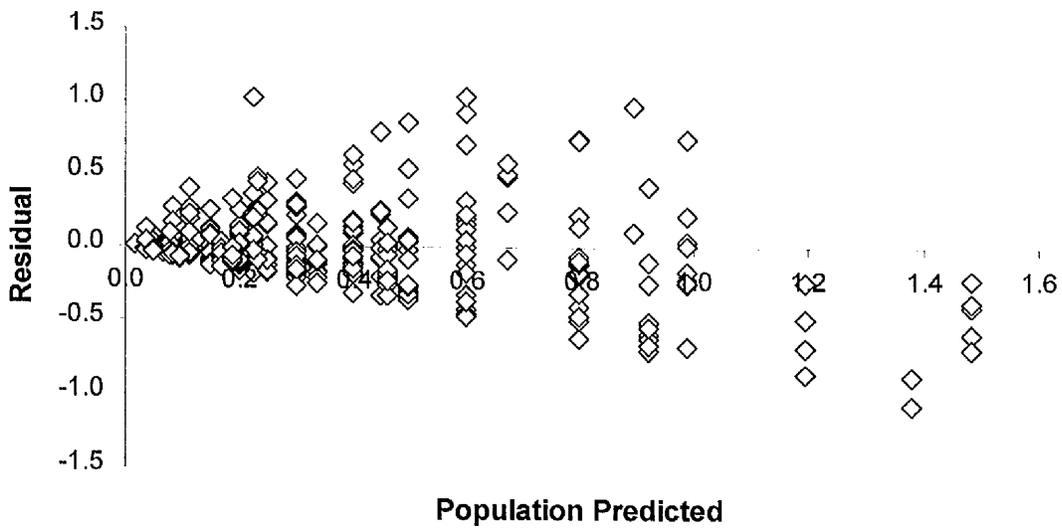
Figure 4.11. *Weighted residual versus predicted plot for 2, 4 and 13 week combined Wistar rat data.*
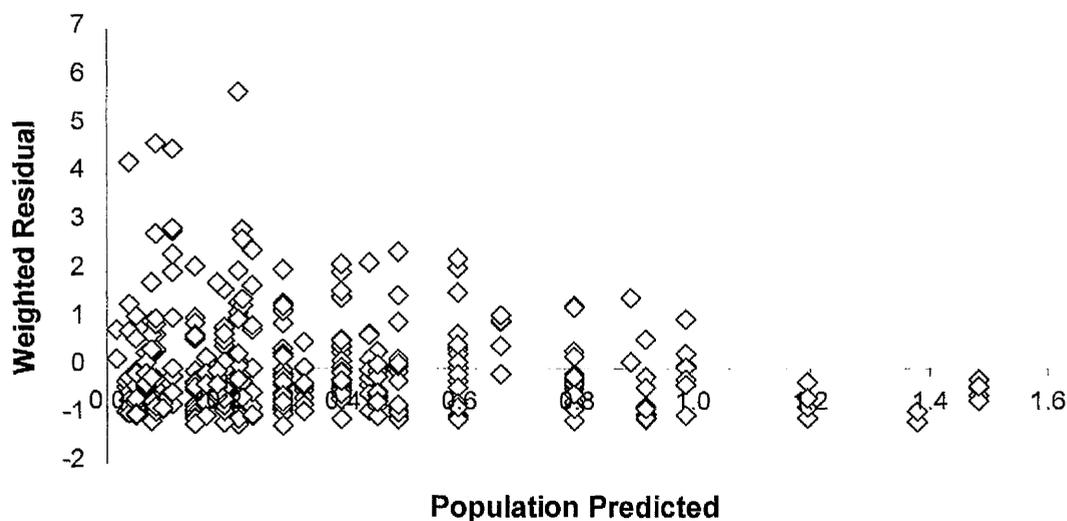
4.2.1.4 Study RR-764-02167 – 4 Week Study With Dietary Admixture Administration

4.2.1.4.1 Study Design

For toxicokinetic evaluation, 3 groups of 30 rats per gender were given PD-142676 as a drug-diet admixture on a mg/kg body weight basis. 5 rats per gender were assigned to a placebo group. The doses to be administered were 0, 15, 30 and 90 mg/kg.

Plasma samples were collected from 5 rats representing each gender in the 15, 30 and 90 mg/kg dose groups into heparinised tubes at 1, 2, 4, 8, 12, and 24 hours from the beginning of week 4. Only one sample was obtained from each rat.

4.2.1.4.2 Pharmacokinetic Analysis

The number of concentration measurements are given in table 4.10.

Table 4.10. *Concentration measurements for 4 week toxicity study with dietary admixture 764-02167.*

	0 mg/kg	15 mg/kg	30 mg/kg	90 mg/kg	Total
Male	0(5)	22(8)	23(7)	30(0)	75(20)
Female	0(5)	27(3)	28(2)	25(1)	80(11)
Total	0(10)	49(11)	51(9)	55(1)	155(31)

(.) BLQ

The results of NONMEM runs are given table 4.11. As the route of administration was different from all the other studies, the results are just reported for completeness and are not considered further. For future pharmacodynamic analyses, no clinical signs of interest were reported so no analysis could be carried out.

Models:

(1) 1 compartment first-order absorption – exponential error model on parameters and residual.

(2) 1 compartment first-order absorption – exponential on parameters and additive residual.

(3) 1 compartment first-order absorption – additive on parameters and exponential residual.

(4) 1 compartment first-order absorption – additive error model on parameters and residual.

(5) 1 compartment zero-order absorption – exponential error model on parameters and residual.

(6) 1 compartment first-order absorption – fixed effects on dose levels with exponential error on model for interindividual and residual terms.

Table 4.11. *Results of NONMEM analyses for 4 week Wistar rat study 764-02167.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	T_{inf} (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	σ^2 (S.E.)
(1)	-702.722	12.5 (0.518)	92.5 (8.63)	0.132 (0.0106)	-	8×10^{-9a} (2×10^{-8})	0.365^a (0.412)	0.232^a (0.041)
(2)	-692.774	13.2 (0.738)	58.5 (10.5)	0.0991 (0.0135)	-	0.287^a (0.0971)	0.866^a (0.574)	1×10^{-10b} (3×10^{-10})
(3)	-687.879	13.8	46	0.0775	-	83^b	12.6^b	4×10^{-10a}
(4)	-701.722	12.5 (0.518)	92.5 (8.63)	0.132 (0.0106)	-	1×10^{-7b} (3×10^{-7})	3120^b (3530)	0.232^b (0.041)
(5)	-698.414	11.6 (0.586)	139 (26.3)	-	9.14 (1.73)	0.0092^a (0.0337)	1.5×10^{-6a} (4×10^{-6})	0.261^a (0.0561)
(6)	-644.858	[15]11 [30]13.2 [90]10.9	[15]99.6 [30]99.9 [90]90.6	[15]1.16 [30]0.13 [90]0.13	-	0.179^a	-	0.114^a

[a] Exponential error model.
[b] Additive error model.
[.] Dose.

4.2.2 Pharmacodynamic Analyses of Rat Data Sets

4.2.2.1 Study RR-745-02051 – 2 Week Study

4.2.2.1.1 Pharmacodynamic Data

Although clinical data were collected on a daily basis, the observations were summarised and reported on a weekly basis. This meant that there were a maximum of 2 observations per rat. If the rat died in one of the two weeks then the clinical signs

127

until death were reported. If the death occurred in week one then that rat would have one pharmacodynamic measurement whereas if it died in week two then that rat would have two pharmacodynamic measurements. Even if a rat was observed to have more than one convulsion a week, it was still reported only as having a convulsion in that week. The response variable is therefore binary with 0=no convulsion and 1=convulsion. Out of a total of 190 rats, there were 379 pharmacodynamic measurements as one male rat died in week one in dose group 60 mg/kg.

The type of clinical signs and symptoms reported included alopecia, salivation, convulsion, soreness of skin, tremors, and urine discharges. Instead of taking an overall measure of clinical outcome, it was decided that convulsion would be used as the univariate measure of clinical outcome. This will be used throughout all studies in all animals as the pharmacodynamic measure.

4.2.2.1.2 Pharmacodynamic Analysis

The number of convulsions observed in each dose group and at each time is given in table 4.12.

Table 4.12. *Number of convulsions in 2 week Wistar rat study 745-02051.*

Gender	Week	0 mg/kg	15 mg/kg	30 mg/kg	60 mg/kg	Total
Male	1	0(5)	0(30)	0(30)	1(30)	1(95)
	2	0(5)	0(30)	0(30)	1(29)	1(94)
Female	1	0(5)	0(30)	0(30)	0(30)	0(95)
	2	0(5)	0(30)	0(30)	0(30)	0(95)
Total		0(20)	0(120)	0(120)	2(119)	2(379)

(.) Total number of possible observations.

Had this been the only data set and as there are only two convulsions, usually there would be nothing done in terms of a statistical analysis of such data. However as there were other data sets with similar data but with more convulsions, an attempt at finding a model for these data was carried out.

In the equations below, i indexes the rat and j indexes the week $\{1,2\}$.

(1) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \begin{cases} \beta_1 + b_{1i}, Dose = 0 \\ \beta_1 + \beta_2 + b_{1i}, Dose = 15 \\ \beta_1 + \beta_3 + b_{1i}, Dose = 30 \\ \beta_1 + \beta_4 + b_{1i}, Dose = 60 \end{cases}$

(2) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \begin{cases} \beta_1 + \beta_5 time_{ij} + b_{1i}, Dose = 0 \\ \beta_1 + \beta_2 + \beta_5 time_{ij} + b_{1i}, Dose = 15 \\ \beta_1 + \beta_3 + \beta_5 time_{ij} + b_{1i}, Dose = 30 \\ \beta_1 + \beta_4 + \beta_5 time_{ij} + b_{1i}, Dose = 60 \end{cases}$

(3) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \begin{cases} -50 + \beta_2 time_{ij} + b_{1i}, Dose = 0 \\ -50 + \beta_2 time_{ij} + b_{1i}, Dose = 15 \\ -50 + \beta_2 time_{ij} + b_{1i}, Dose = 30 \\ -50 + \beta_1 + \beta_2 time_{ij} + b_{1i}, Dose = 60 \end{cases}$

(4) $\log it(E(\Pr(Y_{ij}=1\mid b_i))) = \beta_1 + \beta_2 AUC + b_i$

(5) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \beta_1 + \beta_2 AUC_i + \beta_3 time_{ij} + b_{1i}$

(6) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid \underline{b}_i))) = (\beta_1 + b_{1i}) + \beta_2 AUC_i + (\beta_3 + b_{2i})time_{ij}$

(7) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \dfrac{\beta_1 AUC_i}{\beta_2 + AUC_i} + b_{1i}$

(8) $\operatorname{logit}(E(\Pr(Y_{ij}=1\mid b_{1i}))) = \beta_1 + \beta_2 time_{ij} + b_{1i}$

Models (3) have a baseline logit of –50 as there were no convulsions observed in these dose groups so the placebo logit was arbitrarily set to a small value.

All of these models are random effects models as they are conditional on the random effect component, $\underline{b}_i$. The low value of the NONMEM objective function value can be attributed to the fact that the vast majority of observations are zero.

Table 4.13. *Results of NONMEM pharmacodynamic analyses for 2 week Wistar rat study 745-02051.*

Model	O.F.	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	β_5 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	12.033	-16.6	-36.7	-36.7	5.07	-	245	-
(2)	7.093	-63.9	-1.82	-1.81	17	15.7	7980	-
(3)	20.309	45.9	0.0156	-	-	-	0.0562	-
(4)	12.093	-15.5	0.613	-	-	-	260	-
(5)	7.101	-53.1 (4.72)	0.955 (0.471)	15.8 (0.865)	-	-	8490 (3270)	-
(6)	6.800	-27.6	0.561	3.82	-	-	1.61	3450
(7)	12.284	-13.6 (1.78)	2.88 (0.169)	-	-	-	312 (323)	-
(8)	7.203	-52.8 (28.7)	17.5 (10.7)	-	-	-	18800 (1×10^5)	-

The models in table 4.13 involving dose as a covariate ((1)-(3)) are defined so that there is a dose specific intercept. Of the dose specific models, model (2) is the best (based on which model has the lowest objective function value).

The models with *AUC* as a covariate ((4)-(8)) are based on estimates of *AUC=Dose/Cl*. The clearance estimates, are individual estimates from the one compartment first-order absorption model at steady state fitted to the combined pharmacokinetic data from the week 2 and 4 data sets. Individual estimates could not be obtained from the 2 week data set alone as the data had only one sample per rat. The best model involving *AUC* as a covariate was model (5) which has linear terms in *AUC* and *time* and an additive random effect on the logit scale. The objective function is not quite as low as that of model (6) which has a random effect on the time parameter (7.115 versus 6.8) but since

there is virtually no variability in the pharmacodynamic response, then the random effects would not be expected to explain much. The model that would best describe this data set is that which has neither *dose* nor *AUC* as a covariate, model (8). This model which includes only time as a covariate does no worse than any of the *dose* or *AUC* models because there is very little information in the data. An objective function value of 7.203 is only slightly worse than the AUC models which probably describe the data just as well. The implication of this model will be discussed in section 4.5.

Although a model has been selected as being best in terms of parsimony and objective function value for this data, no inferences should be considered at this stage, as the data in this study, by themselves, do not produce much information.

4.2.2.2 Study RR-250-01686 – 4 Week Study

4.2.2.2.1   Pharmacodynamic Data

All rats were observed on a daily basis but the observations were recorded as being summarised and reported at the end of each week. The same types of clinical signs were observed in this study as were observed in the 2 week rat study. Convulsion was used as the pharmacodynamic measure. The number of convulsions are given in table 4.14.

Table 4.14. *Number of convulsions in 4 week Wistar rat study 250-01686.*

	Week	0 mg/kg	30 mg/kg	60 mg/kg	90 mg/kg	Total
Male	1	0(20)	0(25)	0(25)	7(25)	7(95)
	2	0(20)	0(25)	4(25)	4(23)	8(93)
	3	0(20)	0(24)	5(23)	7(20)	12(87)
	4	0(20)	0(24)	2(19)	9(15)	11(78)
Female	1	0(20)	0(25)	2(25)	6(25)	8(95)
	2	0(20)	0(25)	3(25)	9(23)	12(93)
	3	0(20)	0(24)	7(23)	9(20)	16(87)
	4	0(20)	3(24)	9(22)	7(11)	19(77)
Total		0(160)	3(196)	32(187)	58(162)	93(705)

(.) Total number of observations.

It can be seen in table 4.14 that as dose increases then the number of convulsions increases. Also as time and dose increases, the mortality increases. For example, the marginal number of convulsions and total number of observations for each dose group over time and gender are 0(160), 2(196), 12(185) and 36(162) for dose groups 0, 30, 60and 90 mg/kg respectively. To make this trend more apparent, table 4.15 gives the number of convulsions with gender not included as a factor. In table 4.16, the number of deaths or being withdrawn from the study in terms of gender, week and dose is presented. Rats were withdrawn by the investigators if it was required to carry out a reversibility study or because the animals were suffering unnecessarily. It can be seen that by the end of the study, 52 rats had dropped out of a possible 190 rats. Table 4.17 shows the number of dropouts with gender not included as a factor. The drop-outs are

recorded as being in the week they occurred which is different to the convention where drop-outs are recorded as the time up until the drop-out occurred.

Table 4.15. *Number of convulsions without gender as a factor in 4 week Wistar rat study 250-01686.*

Week	0 mg/kg	30 mg/kg	60 mg/kg	90 mg/kg	Total
1	0(40)	0(50)	2(50)	13(50)	15(190)
2	0(40)	0(50)	7(50)	13(46)	20(186)
3	0(40)	0(48)	12(46)	16(40)	28(174)
4	0(40)	3(48)	11(41)	16(26)	30(155)
Total	0(160)	3(196)	32(187)	58(162)	93(705)

(.) Total number of observations.

Table 4.16. *Number of dropouts in 4 week Wistar rat study 250-01686.*

	Week	0 mg/kg	30 mg/kg	60 mg/kg	90 mg/kg	Total
Male	1	0(20)	0(25)	0(25)	2(25)	2(95)
	2	0(20)	1(25)	3(25)	3(23)	7(93)
	3	0(20)	0(24)	4(22)	5(20)	9(86)
	4	0(20)	0(24)	0(18)	9(15)	9(77)
Female	1	0(20)	0(25)	0(25)	2(25)	2(95)
	2	0(20)	1(25)	2(25)	3(23)	6(93)
	3	0(20)	0(24)	1(23)	9(20)	10(87)
	4	0(20)	0(24)	2(22)	5(11)	7(77)
Total		0(160)	2(196)	12(185)	38(162)	52(703)

(.) Total number of possible observations at start of week.

Table 4.17. *Number of dropouts without gender as a factor in 4 week Wistar rat study 250-01686.*

Week	0 mg/kg	30 mg/kg	60 mg/kg	90 mg/kg	Total
1	0(40)	0(50)	0(50)	4(50)	4(190)
2	0(40)	2(50)	5(50)	6(46)	13(186)
3	0(40)	0(48)	5(45)	14(40)	19(173)
4	0(40)	0(48)	2(40)	14(26)	16(154)
Total	0(160)	2(196)	12(185)	38(162)	52(703)

(.) Total number of possible observations at start of week.

4.2.2.2.2 Pharmacodynamic Analysis

The data that was obtained in this toxicokinetic/toxicodynamic study is found in many settings. Longitudinal data with missing observations is an important area of research and only recently have methods been devised to analyse such data. Some early papers for analysis of longitudinal data with missing values include Schluchter (1992) and Diggle and Kenward (1994) for continuous variables and Conaway *et al* (1992), Kenward *et al* (1994), Follmann and Wu (1995) and Fitzmaurice *et al* (1995) for categorical data. Sheiner *et al* (1997) described a general analysis for data collected from analgesic drug clinical trials where the probability of dropping out of the clinical trial was dependent on the level of analgesia. Work by Ten Have *et al* (1998) and Pulkstenis *et al* (1998) described similar settings for longitudinal data with missing data. The last three papers describe data settings similar to the rat data observed in study 250-01686. The data is observed repeatedly over time in each rat but some rats do not have full pharmacodynamic profiles (data up to the end of the study) due to censoring. The

way in which these data are statistically assumed missing can vary. The three standard assumptions of how clusters can drop out of studies and create missing data are the following:

1.) Completely random drop-out (CRD) – the dropout and measurement processes are independent.

2.) Random drop-out (RD) – the drop-out process depends on the observed measurements, i.e. those preceding drop-out.

3.) Informative drop-out (ID) – the drop-out process depends on the unobserved measurements as well as the observed measurements, i.e. those that would have been observed if the unit had not dropped out.

In the case of the rat data studied here, the assumption that the data are missing completely at random would be wrong as there is a definite trend in the numbers of drop-outs in terms of dose and probably time as seen in tables 4.16 and 4.17. The assumption that the data are missing at random is plausible in that rats dropped out due to the dose of the PD-142676 given previously. This would be a sensible assumption to make but informative drop-out could also be a possibility. The probability of dropping out depending on some unobserved variable is also possible as rats dying could be due to some other factor other than exposure to the drug.

The first method to be tried for 250-01686 was that of Sheiner *et al* (1997). In their work, they refer to using a 'selection model' where they model the longitudinal data and the drop-out data conditional on the longitudinal data independently rather than a 'pattern-mixture model' which models the drop-out data and the longitudinal data conditional on the drop-out data independently. NONMEM code was downloaded from

the NONMEM repository (Mandema and Stanski (1998)) which is of a very similar form to that of Sheiner *et al* (1997). The code was adapted from being appropriate for a 5 point categorical variable to being binary but this code failed to run in NONMEM and was abandoned.

The method of Ten Have *et al* (1998) was also tried for the analysis of study 250-01686. This method is used for the analysis of longitudinal binary response data to incorporate informative drop-outs. The model consists of observed longitudinal and missing response components that share random effects parameters. FORTRAN code was available from the author but unfortunately, would not work on the rat data set. The reason for not producing any results was due to technical problems with running the code.

The important part of the analysis of study 250-01686 is to be able to describe the probability of getting a toxic response, i.e. a convulsion. The drop-out is of secondary interest in this case but must be taken into account, as a wrong assumption about the way in which the data are missing might lead to bias of the parameter estimates in the logistic regression models.

In the following, a simple method of determining how the missing data might have affected the parameter estimates and the choice of model is given. Firstly, the binary convulsion data are analysed as though the mechanism for the missing data could be completely ignored. Secondly, the missing data are imputed with values last observed for that particular rat. Thirdly, the missing data was imputed as all zeros. Finally, the missing data was imputed as all ones. This method should in a sense put bounds on

what the parameter estimates would be at the extremes if the missing data were all zeros or all ones.

In the first case, a wide range of models were fitted to the data. They are all random effects models of the form $\log it(\Pr(Y_{ij} = 1 \mid \underline{b}_i)) = f(\underline{\beta}, \underline{b}_i, \underline{x}_{ij})$, where f is given below.

(1) $\beta_1 + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + b_{1i}$

(3) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(4) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 dose_i time_{ij} + b_{1i}$

(5) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij}$

(6) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(7) $(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) time_{ij}$

(8) $\beta_1 + \beta_2 auc_i + b_{1i}$

(9) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + b_{1i}$

(10) $\quad \beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + b_{1i}$

(11) $\quad \beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(12) $\quad \beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + \beta_5 sex_i + b_{1i}$

(13) $\quad \beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + \beta_5 sex_i + \beta_6 auc_i sex_i + b_{1i}$

(14) $\quad (\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) auc_i$

(15) $\quad (\beta_1 + b_{1i}) + (\beta_2 \exp(b_{2i})) auc_i$

(16) $\quad (\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i}) time_i$

(17) $\quad \beta_1 + \beta_2 auc_i + (\beta_3 + b_{1i}) time_i$

(18) $\quad \beta_1 + \beta_2 auc_i + (\beta_3 + b_{1i}) time_i + \beta_4 sex_i$

(19) $\quad \beta_1 + \beta_2 auc_i + \beta_3 auc_i^2 + b_{1i}$

(20) $\quad \dfrac{\beta_1\, auc_i}{\beta_2 + auc_i} + b_{1i}$

Table 4.18. *Results of NONMEM runs for 4 week rat study 250-01686 with unimputed*

data.

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	β_5 (S.E)	β_6 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	524	-2.56 (0.314)	-	-	-	-	-	2.57 (0.993)	-
(2)	419	-4.95	0.0491	-	-	-	-	0.425	-
(3)	378.4	-9.68 (1.83)	0.078 (0.0168)	0.95 (0.237)	-	-	-	2.34 (1.51)	-
(4)	400.7	-4.37	0.0271	-0.0156	0.0064	-	-	0.101	-
(5)	378.2	-9.75 (2.04)	0.0795 (0.0193)	0.934 (0.26)	-	-	-	2.08 (1.69)	0.0487 (0.165)
(6)	374.9	-9.98 (2.04)	0.0782 (0.0184)	0.919 (0.242)	0.749 (0.42)	-	-	2.6 (1.51)	-
(7)	429.8	-45.8 (7.2)	8.38 (1.69)	-	-	-	-	1500 (742)	159 (60.8)
(8)	420.3	-5.52 (0.522)	0.565 (0.0659)	-	-	-	-	0.653 (0.066)	-
(9)	380.8	-10.1 (1.38)	0.817 (0.124)	0.991 (0.205)	-	-	-	2.72 (1.24)	-
(10)	380.8	-10.2 (1.81)	0.839 (0.224)	1.03 (0.484)	-0.006 (0.0494)	-	-	2.7 (1.23)	-
(11)	377	-10.4 (1.4)	0.815 (0.121)	0.968 (0.199)	0.823 (0.431)	-	-	2.45 (1.1)	-
(12)	377	-10.5 (1.84)	0.83 (0.221)	1 (0.477)	-0.0046 (0.0683)	0.823 (0.43)	-	2.44 (1.12)	-
(13)	376.6	-11 (2.05)	0.908 (0.252)	0.994 (0.465)	-0.005 (0.0669)	1.74 (1.1)	-0.131 (0.155)	2.39 (1.1)	-
(14)	420.3	-5.24	0.565	-	-	-	-	0.652	5.84×10^{-10}
(15)	420	-5.24	0.528	-	-	-	-	2.71×10^{-9}	0.0389
(16)	380.6	-10.2 (1.49)	0.846 (0.146)	0.973 (0.219)	-	-	-	2.39 (1.41)	0.0724 (0.197)
(17)	384.7	-9.28 (1.32)	0.809 (0.139)	0.735 (0.159)	-	-	-	0.336 (0.23)	-
(18)	381.7	-9.39 (1.27)	0.784 (0.131)	0.726 (0.155)	0.673 (0.406)	-	-	0.271 (0.19)	-
(19)	412.9	-9.12	1.9	-0.106	-	-	-	0.322	-
(20)	607.6	-2.74	4.9×10^{-10}	-	-	-	-	6.93	-

To decide on a 'best' model, parsimony and the objective function value was used as a model choice criteria. Under this criteria, the best model would be model (6), which is a linear function in dose, time and sex which has an objective function value of 374.9. Model (5) is similar in objective function value but has one less fixed effect parameter and one more random effect (additive on time). The models involving AUC as a predictor do not perform any better than models involving dose as a predictor. Models (11) and (12) have objective function values of 377 with 3 and 4 fixed effects respectively.

For the models tried with the imputed data where the last observed value was carried forward for any missing data, there are three alternative models to those given in the list for the unimputed data set: these are replaced by the following in table 4.19. These models were replaced because they were either inappropriate or because the previous models did not run.

(7) $(\beta_1 + b_{1i}) + \beta_2 time_{ij}$

(15) $\beta_1 + \beta_2 dose_i + (\beta_3 + b_{1i})time_{ij}$

(18) $\beta_1 + (\beta_2 + b_{1i})dose_i + \beta_3 time_{ij}$

The best model that describes the imputed data set is between model (3), linear in dose and time and model (6), linear in dose, time and sex, both with one additive random effect. As model (6) only reduces the objective function value by just over 2 points, then model (3) would appear to be the best model. Once again the models including, AUC are not superior to the dose models in any way but models (10) and (11) come close in terms of the objective function value.

Table 4.19. *Results of NONMEM runs for 4 week rat study 250-01686 with imputed data as the same value as last observed value.*

Model	O.F.	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	β_5 (S.E.)	β_6 (S.E)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	606.68	-3.47 (1.05)	-	-	-	-	-	9.8 (5.92)	-
(2)	496.83	-6.74 (0.774)	0.0752 (0.01)	-	-	-	-	2.34 (0.803)	-
(3)	401.03	-14.1 (1.89)	0.114 (0.017)	1.52 (0.249)	-	-	-	7.37 (2.73)	-
(4)	467.55	-5.04	0.0253	0.0257	0.0117	-	-	0.103	-
(5)	401.02	-14.1 (1.85)	0.114 (0.0163)	1.52 (0.256)	-	-	-	7.4 (3.03)	0.0003 (0.108)
(6)	398.85	-14.6 (1.99)	0.115 (0.0172)	1.52 (0.252)	0.893 (0.603)	-	-	7.32 (2.76)	-
(7)	491.05	-14.9 (2.42)	2.13 (0.467)	-	-	-	-	94.7 (34.1)	-
(8)	500.04	-6.82 (0.721)	0.763 (0.0916)	-	-	-	-	2.53 (0.8)	-
(9)	404.19	-14.1 (1.84)	1.14 (0.161)	1.52 (0.247)	-	-	-	7.76 (2.77)	-
(10)	404.19	-14.1 (2.91)	1.15 (0.347)	1.53 (0.707)	-0.0006 (0.0931)	-	-	7.78 (2.81)	-
(11)	401.63	-14.7 (1.94)	1.16 (0.166)	1.52 (0.248)	0.983 (0.614)	-	-	7.67 (2.78)	-
(12)	401.63	-14.7 (2.98)	1.15 (0.347)	151 (0.702)	0.0002 (0.0929)	0.985 (0.62)	-	7.7 (2.78)	-
(13)	401.49	-15.2	1.21	1.53	1.66	-0.001	-0.09	7.66	-
(14)	500.04	-6.81 (0.719)	0.762 (0.0914)	-	-	-	-	2.52 (0.798)	1×10^{-7} (1×10^{-5}
(15)	404.53	-12.5	0.11	1.06	-	-	-	1.05	-
(16)	404.19	-14.1 (1.83)	1.14 (0.163)	1.52 (0.248)	-	-	-	7.75 (2.8)	3×10^{-5} (0.016)
(17)	407.44	-13.5 (1.77)	1.2 (0.176)	1.12 (0.219)	-	-	-	1.31 (0.593)	-
(18)	404.86	-11.8 (1.37)	0.09 (0.0128)	1.47 (0.228)	-	-	-	0.0012 (0.0004	-
(19)	495.04	-10.7 (2.14)	2.18 (0.73)	-0.113 (0.582)	-	-	-	2.09 (0.798)	-
(20)	674.8	-3.03 (0.779)	3×10^{-9} (3×10^{-6})	-	-	-	-	12.2 (4.59)	-

For the third scenario where the missing data are imputed as all 0's, less models were considered as some models seemed unlikely to be appropriate for this data set such as the E_{max} model on the logit scale. These models are given below.

(1) $\beta_1 + \beta_2 dose_i + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(3) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 dose_i time_{ij} + b_{1i}$

(4) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(5) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij}$

(6) $(\beta_1 + \beta_2 dose_i + (\beta_3 + b_{1i}) time_{ij}$

(7) $\beta_1 + (\beta_2 + b_{1i}) dose_i + \beta_3 time_{ij}$

(8) $\beta_1 + \beta_2 auc_i + b_{1i}$

(9) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + b_{1i}$

(10) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + b_{1i}$

(11) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(12) $(\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i}) time_{ij} + \beta_4 auc_i time_{ij}$

(13) $(\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i}) time_{ij}$

Table 4.20. *Results of NONMEM runs for 4 week rat study 250-01686 with imputed data being all 0's.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	458.207	-4.75	0.0439	-	-	1.5×10^{-4}	-
(2)	450.462	-5.41	0.0423	0.312	-	0.0067	-
(3)	452.064	-5.8	0.0528	0.335	-0.0028	0.108	-
(4)	448.665	-5.23	0.0372	0.288	0.463	0.0216	-
(5)	456.86	-4.61	0.0321	0.289	-	8.7×10^{-4}	8.1×10^{-4}
(6)	454.898	-4.76	0.034	0.288	-	1.4×10^{-4}	-
(7)	450.566	-5.38	0.0422	0.292	-	1.3×10^{-6}	-
(8)	460.366	-4.75 (0.357)	0.444 (0.0427)	-	-	2×10^{-11} (0.0086)	-
(9)	452.301	-5.6 (0.429)	0.45 (0.0422)	0.308 (0.116)	-	2×10^{-13} (0.0071)	-
(10)	449.648	-8.34	0.803	1.2	-0.116	8.6×10^{-4}	-
(11)	447.707	-5.91	0.453	0.311	0.521	2.6×10^{-7}	-
(12)	447.636	-5.97	0.462	0.303	0.533	1.8×10^{-8}	0.00883
(13)	452.201	-5.66 (0.435)	0.46 (0.0488)	0.298 (0.126)	-	2.6×10^{-8} (1×10^{-8})	0.0115 (0.05)

The lowest values of the objective function are associated with models involving AUC as a predictor ((10)-(12)) but these models have more parameters than model (2) which has three fixed effects parameters and one random effect parameter. The objective function value for model (2) (linear in dose and time) is 450.5, which is only 2.6 points greater than that for the AUC models (10)-(12). It would appear that the best model for this imputed data set is linear in dose and time.

The last set of imputed data has all missing values set to 1. The models considered are given below.

(1) $\beta_1 + \beta_2 dose_i + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(3) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 dose_i time_{ij} + b_{1i}$

(4) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(5) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij}$

(6) $(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) dose_i + (\beta_3 + b_{3i}) time_{ij}$

(7) $\beta_1 + \beta_2 auc_i + b_{1i}$

(8) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + b_{1i}$

(9) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + b_{1i}$

(10) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(11) $(\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i}) time_{ij}$

(12) $(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) auc_i + (\beta_3 + b_{3i}) time_{ij}$

(13) $(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) auc_i + \beta_3 time_{ij}$

Table 4.21. *Results of NONMEM runs for 4 week rat study 250-01686 with imputed data being all 1's.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)	ω_3^2 (S.E.)
(1)	529.58	-6.11 (0.703)	0.0683 (0.0092)	0.0683 (0.0092)	-	2.23 (0.76)	-	-
(2)	421.28	-13.8 (1.88)	0.109 (0.0166)	1.6 (0.26)	-	7.86 (2.94)	-	-
(3)	496.155	-4.21	0.0136	-0.0228	0.016	0.104	-	-
(4)	419.049	-14.3 (1.96)	0.11 (0.0167)	1.6 (0.26)	0.893 (0.59)	7.72 (2.9)	-	-
(5)	416.179	-15 (2.29)	0.124 (0.0203)	1.57 (0.306)	-	5.59 (3.07)	0.723 (0.533)	-
(6)	416.174	-15.1	1.74	-	-	5.66	1×10^{-18}	0.73
(7)	528.353	-6.27 (0.658)	0.708 (0.0848)	-	-	2.26 (0.714)	-	-
(8)	420.172	-13.9 (1.84)	1.11 (0.162)	1.6 (0.256)	-	7.74 (2.78)	-	-
(9)	420.158	-14.1 (2.78)	1.14 (0.331)	1.66 (0.65)	-0.008 (0.0863)	7.79 (2.83)	-	-
(10)	417.316	-14.5 (1.92)	1.12 (0.163)	1.6 (0.256)	1.02 (0.592)	7.62 (2.76)	-	-
(11)	415.514	-15.3 (2.2)	1.28 (0.199)	1.6 (0.293)	-	5.88 (2.89)	0.693 (0.529)	-
(12)	415.514	-15.3	1.28	1.6	-	5.88	4×10^{-11}	0.693
(13)	420.172	-13.9	1.11	1.6	-	7.75	5.9×10^{-8}	-

The best models appear to be linear in dose or AUC with time and sex (models (4) and (10)) or linear in dose with an additive random effect or AUC with time and additive random effects on the intercept and time parameter (models (5) and (11)). The models with sex as a predictor have slightly higher objective function values but one less random effect. In this case it was decided to choose the model with a random effect on time and the intercept and dose as a linear predictor.

The final chosen models for the 4 data sets are the following:

Data set A. Unimputed data.

(A) $\text{logit}(E(\Pr(Y_{ij} = 1 \mid \underline{b}_i))) = \beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

Data set B. Imputed data with last value carried forward.

(B) $\text{logit}(E(\Pr(Y_{ij} = 1 \mid \underline{b}_i))) = \beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

Data set C. Imputed data as all 0's.

(C) $\text{logit}(E(\Pr(Y_{ij} = 1 \mid \underline{b}_i))) = \beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

Data set D. Imputed data with all 1's.

(D) $\text{logit}(E(\Pr(Y_{ij} = 1 \mid \underline{b}_i))) = (\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij}$

To compare how the imputed data has an effect on the parameter estimates, the linear logistic model in dose and time with an additive random effect was chosen as the comparison model. The parameter estimates for these models are given in table 4.22.

Table 4.22. *Parameter values for the linear logistic model* $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_i$ *for the 4 different data sets in the rat 4 week study 250-01686.*

Data set	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	ω_1^2 (S.E.)
A	378.4	-9.68 (1.83)	0.078 (0.0168)	0.95 (0.237)	2.34 (1.51)
B	401.03	-14.1 (1.89)	0.114 (0.017)	1.52 (0.249)	7.37 (2.73)
C	450.462	-5.41	0.0423	0.312	0.0067
D	421.28	-13.8 (1.88)	0.109 (0.0166)	1.6 (0.26)	7.96 (2.94)

In figure 4.12 the models are compared graphically for probability as a function of time, where dose has been set to 30 mg/kg.

Figure 4.12. *Logistic regression plots of 4 data sets for model in dose=30mg/kg and time.*
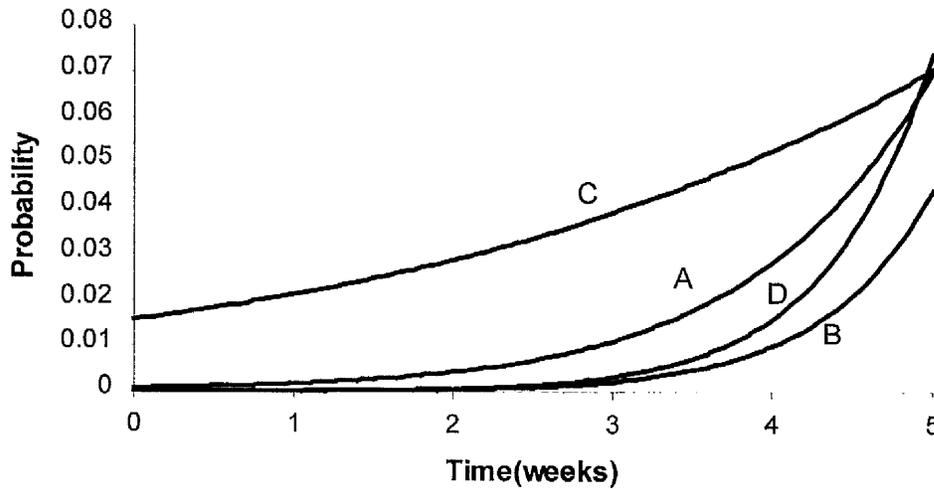
The probabilities here are very low, with the probability of a convulsion in week 4 being approximately 0.04 for data set A (unimputed data). The parameter estimates for data sets A and B are similar, but imputing the data as the last observed observation causes an increase in the magnitude of the parameter estimates by approximately 50%. As expected, imputing the missing values all as 0's (C) initially 'flattens' the curve by reducing the magnitude of the parameter estimates and imputing the missing data as all 1's (D) increases the parameter estimates magnitude making the logistic curve steeper.

4.2.2.3 Study RR-250-01696 – 13 Week Study

4.2.2.3.1 Pharmacodynamic Data

All rats were observed on a daily basis but the observations were recorded as only being from a particular week. The same types of clinical signs were observed in this study as

were observed in the 2 and 4 week rat studies. Convulsion was used as the pharmacodynamic measure. The number of convulsions are given in table 4.23.

Table 4.23. *Number of convulsions in 13 week Wistar rat study 250-01696.*

Sex	Week	0 mg/kg	5 mg/kg	15 mg/kg	30 mg/kg	Total
Male	1	0(20)	0(25)	0(25)	0(25)	0(95)
	2	0(20)	0(25)	0(25)	0(25)	0(95)
	3	0(20)	0(25)	0(25)	0(25)	0(95)
	4	0(20)	0(25)	0(25)	0(25)	0(95)
	5	0(20)	0(25)	0(25)	0(24)	0(94)
	6	0(20)	0(25)	0(25)	0(24)	0(94)
	7	0(20)	0(25)	0(25)	0(24)	0(94)
	8	0(20)	0(25)	0(25)	1(24)	1(94)
	9	0(19)	0(25)	0(24)	0(24)	0(92)
	10	0(19)	0(25)	0(24)	0(24)	0(92)
	11	0(19)	0(25)	0(24)	1(24)	1(92)
	12	0(19)	0(25)	0(24)	0(24)	0(92)
	13	0(19)	0(25)	0(24)	5(24)	5(92)
Female	1	0(20)	0(25)	0(25)	0(25)	0(95)
	2	0(20)	0(25)	0(25)	0(25)	0(95)
	3	0(20)	0(25)	0(25)	0(24)	0(94)
	4	0(20)	0(25)	1(25)	2(24)	3(94)
	5	0(20)	0(25)	0(25)	0(24)	0(94)
	6	0(20)	0(25)	0(25)	0(24)	0(94)
	7	0(20)	0(25)	0(25)	0(24)	0(94)
	8	0(20)	0(25)	0(25)	2(24)	2(94)
	9	0(20)	0(25)	0(25)	0(23)	0(93)
	10	0(20)	0(25)	0(25)	0(23)	0(93)
	11	0(20)	0(25)	1(25)	2(23)	3(93)
	12	0(20)	0(25)	1(25)	1(22)	2(92)
	13	1(20)	0(25)	2(25)	1(22)	4(92)
Total		1(515)	0(650)	5(645)	15(623)	21(2433)

() Total number of possible observations.

4.2.2.3.2 Pharmacodynamic Analysis

The models considered for the analysis of study 250-01696 are given below.

(1) $\beta_1 + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + b_{1i}$

(3) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(4) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij}$

(5) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i}) time_{ij} + \beta_4 sex_i$

(6) $\beta_1 + \beta_2 dose_i + (\beta_3 + b_{1i}) time_{ij}$

(7) $\beta_1 + \beta_2 AUC_i + b_{1i}$

(8) $\beta_1 + \beta_2 AUC_i + \beta_3 time_{ij} + b_{1i}$

(9) $(\beta_1 + b_{1i}) + \beta_2 AUC_i + (\beta_3 + b_{2i}) time_{ij}$

(10) $(\beta_1 + b_{1i}) + \beta_2 AUC_i + (\beta_3 + b_{2i}) time_{ij} + \beta_4 sex_i$

Table 4.24. *Results of NONMEM runs for 13 week rat study 250-01696.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	231.383	-8.47 (1.01)	-	-	-	21.1 (10.4)	-
(2)	220.825	-8.21	0.0941	-	-	5.26	-
(3)	201.565	-8.57	0.0543	0.181	-	4.12	-
(4)	182.893	-12.1	-0.0129	0.155	-	26.9	0.688
(5)	183.175	-12.1	-0.0183	0.0532	1.49	22.7	0.735
(6)	193.003	-7.36	0.507	-	-	0.77	-
(7)	226.194	-7.8 (1.69)	0.507 (0.193)	-	-	6.4 (7.73)	-
(8)	191.574	-14 (9.09)	0.619 (0.55)	0.446 (0.423)	-	22.6 (42.4)	-
(9)	181.072	-13.1	0.62	0.167	-	18.6	0.424
(10)	180.697	-12.5	0.539	0.035	1.6	12.3	0.487

Out of the models considered, the best model is either model (4) or (9). Both models have 3 fixed effects parameters and 2 random effects parameters on the intercept and time covariate. The difference between the two models is the use of either dose or AUC. The objective function values for the dose and AUC models are 182.893 and 181.072 respectively. As there is virtually nothing between the objective function values, the dose model is preferred as it does not require the computation of the individual AUC's.

4.3 Dog Studies

There were two Beagle dog studies, a 2 week and 4 week oral toxicity study. In studies 764-02152 and 764-02185, the compound PD-142676 was given orally. The drug was administered in gelatin capsules on a mg/kg body weight basis. Dose selection criteria and dog numbers will be given in each study section as these altered from study to study.

4.3.1 Pharmacokinetic Analyses of Dog Data Sets

4.3.1.1 Study RR-764-02152 – 2 Week Study

4.3.1.1.1 Study Design

For toxicokinetic evaluation, 8 beagle dogs per gender were assigned to 8 experimental groups of 2 dogs of the same gender/group. Dogs were administered PD-142676 at 0, 5,

20 and 60 mg/kg in a gelatin capsule once daily for two weeks. Plasma samples were collected either on day 1 or day 9 at 1, 2, 4, 6, 8 and 12 hours post-dose.

Escalating dose and 2 week repeated dose toxicity studies with PD-142676 in monkeys demonstrated that CNS effects occurred at oral doses between 20 and 75 mg/kg. Without prior experience in dogs, an initial dose of 60 mg/kg was used based on the lack of clinical signs in another study in monkeys. A dose of 20 mg/kg was subsequently selected based on clinical signs induced at 60 mg/kg, and 5 mg/kg was selected based on clinical signs at 20 mg/kg.

4.3.1.1.2 Pharmacokinetic Analysis

The number of plasma concentrations measured are given in table 4.25.

Table 4.25. *Number of concentration measurements for 2 week toxicity study in beagle dogs study 764-02152.*

	0 mg/kg	5 mg/kg	20 mg/kg	60 mg/kg	Total
Male	0(14)	5(9)	7(1)	8(0)	20(24)
Female	0(14)	0(14)	19(4)	8(0)	27(32)
Total	0(28)	5(23)	26(5)	16(0)	47(56)

(.) BLQ

There are only 47 concentration measurements above the limit of quantification from 16 dogs. An examination of the data showed that there is very little information and only simple models of the type used for the rat data were used. One compartment models

150

were used with either first-order absorption or bolus administration. The models examined are given below. The results are in table 4.26.

(1) One compartment first-order absorption, exponential random effects on apparent clearance and apparent volume, exponential and additive residual error model.

(2) One compartment first-order absorption, exponential random effects on apparent clearance and apparent volume and exponential residual error model.

(3) One compartment bolus administration, exponential random effects on apparent clearance and apparent volume, exponential and additive residual error model.

(4) One compartment bolus administration, exponential random effects on apparent clearance and apparent volume, exponential residual error model.

Table 4.26. *Results of NONMEM runs for 2 week beagle dog study 764-02152.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	σ_1^2 (S.E.)	σ_2^2 (S.E.)
(1)	559.02	5.02 (0.858)	35.6 (7.34)	2.83 (0.821)	0.141 (0.129)	0.759 (.303)	0.49 (0.189)	3×10^{-9} (0.277)
(2)	559.22	5.03 (0.862)	35.7 (7.37)	2.83 (0.819)	0.142 (0.129)	0.76 (0.305)	0.499 (0.19)	-
(3)	559.38	4.86	34.9	-	0.144 [0.097]	0.855	0.504	5×10^{-8}
(4)	559.66	4.86 (0.825)	34.8 (7.25)	-	0.126 (0.126)	0.781 (0.333)	0.516 (0.192)	-

[.] Covariance between apparent clearance and apparent volume.

As the objective function values are virtually the same, the simplest, model (4) would seem appropriate. Residual and weighted residual plots of model (4) are given in figures 4.13 and 4.14.

Figure 4.13. *Residual versus population predicted for model (4) for beagle dog 2 week study.*
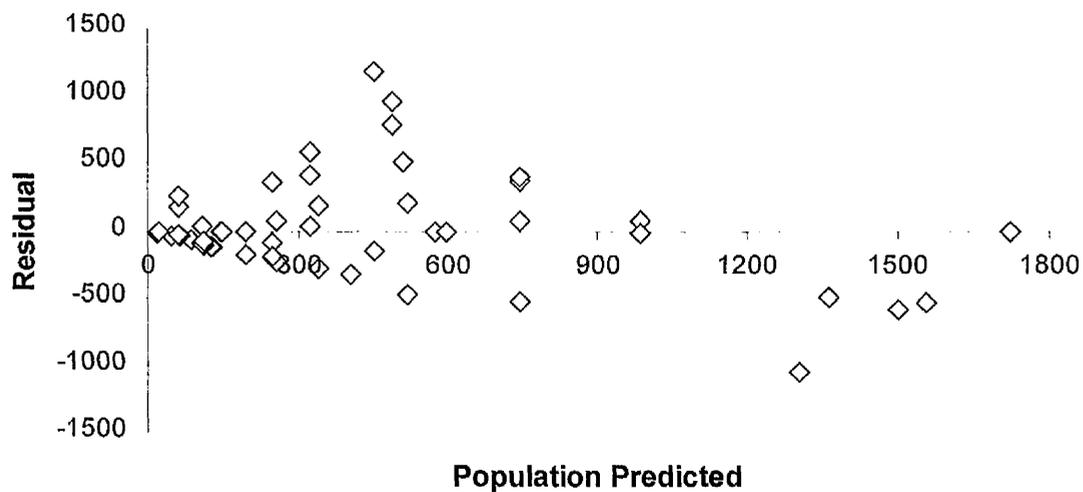
Figure 4.14. *Weighted residual versus population predicted for beagle dog 2 week study.*
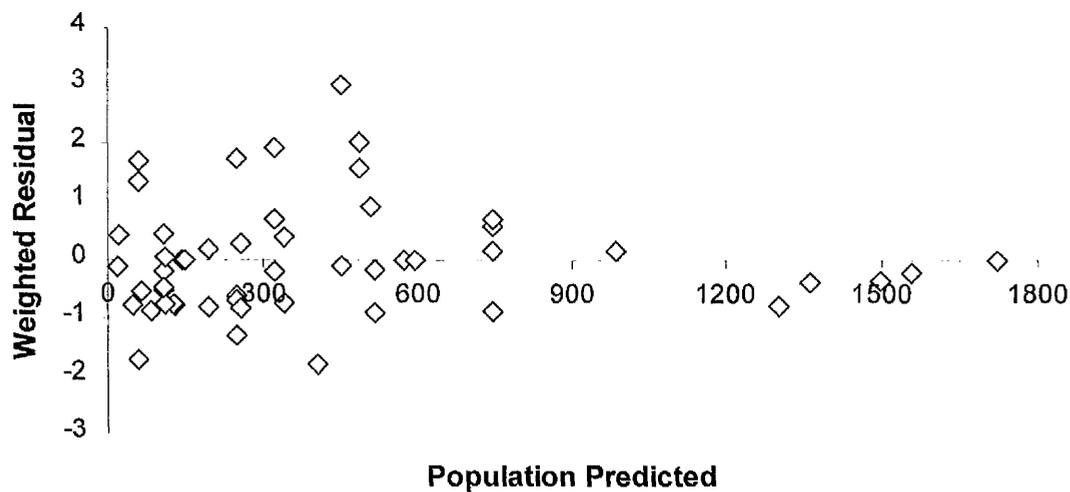
4.3.1.2 Study RR-764-02185 – 4 Week Study

4.3.1.2.1 Study Design

For toxicokinetic evaluation, 4 groups of 3 dogs per gender were given PD-142676 as an oral dose once daily for 4 weeks. Male and female beagle dogs were administered PD-142676 at 0, 5, 10 and 20 mg/kg/day. The 20 mg/kg/day group was divided into 10 mg/kg given 4 hours apart. The dosing interval for animals in the 20 mg/kg dose group was extended during the study because of adverse drug reactions. Doses were given 6 hours apart to 2 female dogs on days 3 and 4. Dosing interval was increased to 8 hours for all female dogs beginning on day 5 and for all male dogs beginning on day 7 of the study. On days 1, 14 and 22 of the study, plasma samples were collected from the dogs representing each gender and each dose group at 1, 2, 4, 6, 8, 12 and 24 hours post-dose. The 4 hour sample on day 1 and the 8 hour sample on days 14 and 22 were obtained prior to administration of the second dose for the 20 mg/kg dose group. A single sample was obtained 2 hours post-dose for the placebo group.

Doses were selected based on the previous 2 week oral toxicity study. All dogs convulsed after a single dose of 60 mg/kg and one animal died. At 20 mg/kg, both males convulsed after a single dose and dosing was continued. At 5 mg/kg, dogs exhibited no clinical signs with 2 weeks of dosing. Based on these data, doses of 5, 10 and 20 mg/kg were chosen for this study. The 20 mg/kg dose was split to reduce deaths and maintain exposure. As the 20 mg/kg/day dose group was divided into two 10 mg/kg doses, this suggests that it was considered that C_{max} might have been an important parameter but was not considered at the time of analysis.

4.3.1.2.2 Pharmacokinetic Analysis

The number of concentration measurements are given in table 4.27.

Table 4.27. *Number of concentration measurements for 4 week toxicity study in beagle dogs.*

	0 mg/kg	5 mg/kg	10 mg/kg	20 mg/kg	Total
Male	0(9)	30(33)	41(22)	44(5)	115(60)
Female	0(9)	33(30)	42(21)	54(9)	129(60)
Total	0(18)	63(63)	63(43)	16(0)	244(120)

(.) BLQ

There are 244 concentration measurements from 18 dogs. The measurements are not only repeated within a particular day but repeated on 3 days (1, 14 and 22). Such data comes under the name of multi-occasion data and requires extensions of the current methods of interindividual and residual variability modelling to include quantification of the variability due to observations being made at different occasions. Karlsson and Sheiner (1993) and Lunn and Aarons (1997) have reported on methodology for analysing multi-occasion data using NONMEM and MCMC methods respectively. This data set is not specifically in the form of multi-occasion data as the dosing is repeated daily. The natural way to think of multi-occasion data is where the drug is given on completely independent occasions. From the previous 2 week dog study, $t_{1/2}$ was estimated to be 5 hours, so by the end of 24 hours, approximately 95% of the drug would have been eliminated. This is approximately analogous to the drug being given on independent occasions.

To test that there was not a trend in the concentration profiles from occasion to occasion, individual AUC values were obtained by the trapezoidal rule and ANOVA was used to test for significant differences between the AUCs on different occasions for the dosing groups. No significant difference was seen, so there was no need to look for trends between occasions. The models considered for the 4 week dog study were:

(1) 1 compartment first-order absorption, exponential variability on Cl/F and V/F and residual terms.

(2) 1 compartment first-order absorption, exponential variability on Cl/F and V/F and combined error on residual term.

(3) 1 compartment first-order absorption, exponential variability and exponential interoccasion on Cl/F and V/F and combined error residual term.

(4) Same as (4) with exponential error on k_a.

(5) 1 compartment first-order absorption, exponential error and exponential interoccasion on Cl/F and V/F and exponential error residual term.

(6) 1 compartment first-order absorption, additive error on Cl/F, V/F and k_a and exponential residual term.

(7) 1 compartment first-order absorption, exponential error on Cl/F, V/F, k_a and residual terms.

(8) 2 compartment first-order absorption, exponential error on Cl/F and $V1/F$ and exponential residual term.

(9) 2 compartment first-order absorption, exponential error on Cl/F, $V1/F$ and k_a and exponential residual term.

The results of these models are given in table 4.28.

Table 4.28. *Results of NONMEM runs for 4 week beagle dog study 764-02185.*

Model	O.F.	Cl/F (S.E.)	V(1)/F (S.E.)	Q/F (S.E.)	V2/F (S.E.)	k_a (S.E.)	ω_{cl}^2 (S.E.)	ω_v^2 (S.E.)	ω_{ka}^2 (S.E.)	$\omega_{Cl,IOV}^2$ (S.E.)	$\omega_{v,IOV}^2$ (S.E.)	σ_1^2 (S.E.)	σ_2^2 (S.E.)
(1)	2216.43	15.7 (2.1)	9.73 (1.5)	-	-	0.142 (0.02)	0.238 (0.109)	19.9 (12.8)	-	-	-	1.36 (0.2810)	-
(2)	2204.72	16.7 (2.59)	96.9 (20.7)	-	-	3.49 (2.7)	0.253 (0.171)	0.552 (0.186)	-	-	-	1.47 (0.288)	13.9 (6.87)
(3)	2202.44	17.2 (2.63)	99.7 (27.7)	-	-	3.1 (1.62)	0.181 (0.163)	0.111 (0.11)	-	0.494 (0.219)	0.228 (0.414)	1.42 (0.3)	12.6 (10.2)
(4)	2198.57	16.9 (2.53)	107 (46.5)	-	-	2.65 (0.869)	0.162 (0.151)	0.135 (0.127)	0.714 (0.523)	0.351 (0.663)	22.6 (33.9)	1.25 (0.334)	4.34 (28.2)
(5)	2198.64	16.7 (2.39)	108 (6.3)	-	-	2.59 (0.734)	43.3 (43)	39.7 (37.8)	8660 (10400)	4520 (7800)	154 (2068)	1.23 (0.333)	-
(6)	2202.72	16 (2.18)	103 (26.5)	-	-	2.24 (0.55)	59.5 (47.7)	8880 (8040)	93.5 (106)	-	-	1.24 (0.306)	-
(7)	2202.74	16 (2.18)	103 (26.7)	-	-	2.24 (0.548)	0.233 (0.139)	0.834 (0.368)	18.7 (16.7)	-	-	1.24 (0.306)	-
(8)	2194.1	15.7 (2.19)	82.9 (39.9)	19.6 (11.2)	50.6 (37.4)	1.9 (0.651)	0.224 (0.134)	3.25 (3.91)	11.9 (24.5)	-	-	1.33 (0.316)	-
(9)	2196.3	16.3	104	2.42	300	4.71	8×10^{-9}	0.27	0.21	0.786	192	1.33	-

Inclusion of interoccasion variability on the parameters did not reduce the objective function by an appreciable amount and so it was decided that an interoccasion parameter was not needed to be estimated for this data set. A two compartment model with first-order absorption did not improve the fit and so the model chosen as best describing this data set was model (7), a one compartment first-order absorption with exponential errors on the parameters Cl/F, V/F, k_a and the residual term. Residual plots are given figures 4.15 and 4.16.

Figure 4.15. *Residual versus population predicted for 4 week beagle dog study.*
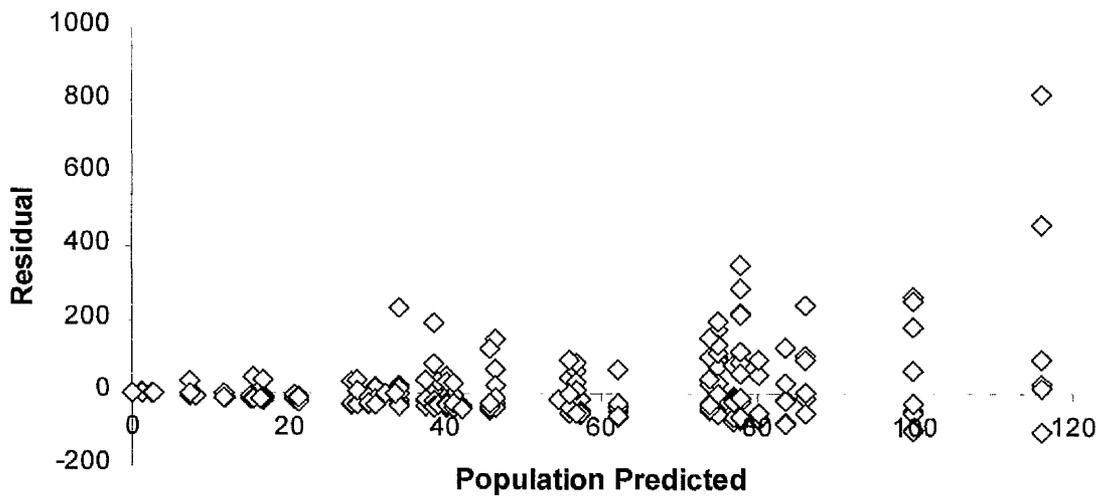
Figure 4.16. *Weighted residual versus population predicted for 4 week dog study.*
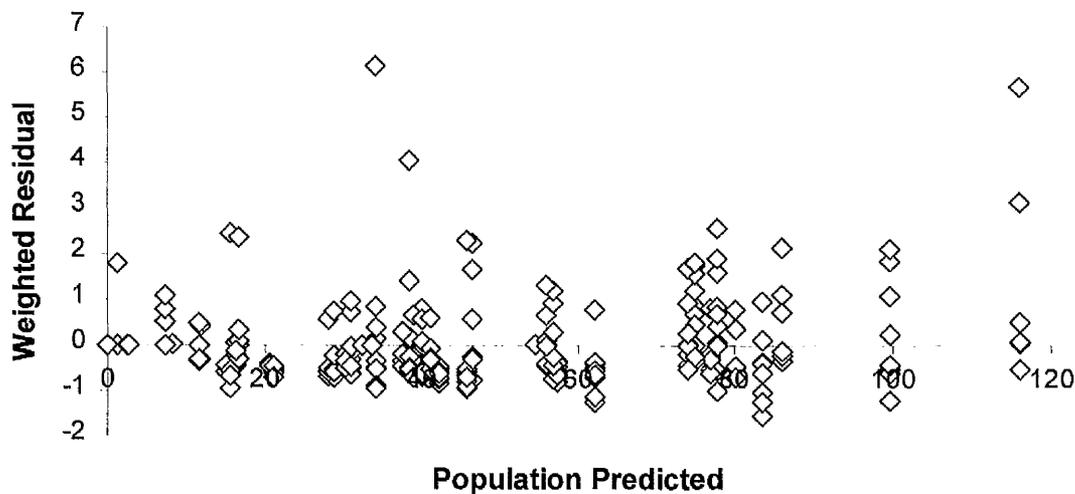
4.3.2   Pharmacodynamic Analyses of Dog Data Sets

4.3.2.1 Study RR-750-02250 – 2 Week Study

4.3.2.1.1   Pharmacodynamic Data

Clinical signs were observed on a daily basis and reported in 745-02250 on a daily basis. The type of clinical signs observed were emesis, convulsion, salivation, tremor, hyperthermia, diarrhoea, hyperemia, tachypnoea and ataxia. Convulsion was used as the measure of pharmacodynamic response.

4.3.2.1.2   Pharmacodynamic Analysis

The number of convulsions are given in table 4.29

Table 4.29. *Number of convulsions in 2 week beagle dog study 745-02250.*

	0 mg/kg	5 mg/kg	20 mg/kg	60 mg/kg	Total
Male	0(28)	0(28)	2(2)	2(2)	4(60)
Female	0(28)	0(28)	2(23)	2(2)	4(81)
Total	0(56)	0(56)	4(25)	4(4)	8(141)

(.) Total number of possible observations.

All 4 dogs in the 60 mg/kg group were withdrawn after one day of treatment for a reversibility study due to severe toxicities after one dose. One female dog died after one day in the 60 mg/kg group. Both male dogs in the 20 mg/kg group were also withdrawn

due to severe toxicities after one dose but the female dogs continued but were sacrificed on days 11 and 12. The models considered for the analysis of study 745-02250 were the following.

(1) $\beta_1 + b_{1i}$

$$(2) \begin{cases} \beta_1 + b_{1i}, Dose = 0 \\ \beta_1 + b_{1i}, Dose = 5 \\ \beta_1 + \beta_2 + \beta_4 time_{ij} + b_{1i}, Dose = 20 \\ \beta_1 + \beta_3 + \beta_4 time_{ij} + b_{1i}, Dose = 60 \end{cases}$$

$$(3) \begin{cases} \beta_1 + b_{1i}, Dose = 0 \\ \beta_1 + b_{1i}, Dose = 5 \\ \beta_2 + \beta_3 time_{ij} + b_{1i}, Dose = 20 \\ \beta_4 + \beta_5 time_{ij} + b_{1i}, Dose = 60 \end{cases}$$

$$(4) \begin{cases} -50 + \beta_3 time_{ij} + b_{1i}, Dose = 0 \\ -50 + \beta_3 time_{ij} + b_{1i}, Dose = 5 \\ -50 + \beta_1 + \beta_3 time_{ij} + b_{1i}, Dose = 20 \\ -50 + \beta_2 + \beta_3 time_{ij} + b_{1i}, Dose = 60 \end{cases}$$

$$(5) \begin{cases} -50 + b_{1i}, Dose = 0 \\ -50 + b_{1i}, Dose = 5 \\ -50 + \beta_1 + \beta_3 time_{ij} + b_{1i}, Dose = 20 \\ -50 + \beta_2 + \beta_3 time_{ij} + b_{1i}, Dose = 60 \end{cases}$$

$$(6) \begin{cases} -50 + b_{1i}, Dose = 0 \\ -50 + b_{1i}, Dose = 5 \\ -50 + \beta_1 + b_{1i}, Dose = 20 \\ -50 + \beta_2 + b_{1i}, Dose = 60 \end{cases}$$

The results of these models are given in table 4.30.

Models linear in dose and AUC could not be estimated, probably due to the lack of information in the data. The models considered are all fixed effects models on the dose level as this was the only way to see if there was an effect by dose. Apart from model

159

(3) which has five parameters and model (1), the other models all have relatively the same fit to the data in terms of the objective function. As no convulsions were observed in the placebo and 5 mg/kg group, it is not sensible to estimate a baseline probability but better to fix it to a number corresponding to a low probability. This was set arbitrarily to -50 and the baseline values in the 20 and 60 mg/kg dose groups could be estimated with respect to this level. Model (6) was chosen because it did not require an estimate for the baseline probability in the lowest two dose groups and in total required the estimation of two other baseline probabilities for the two highest dose groups.

Table 4.30. *Results of NONMEM runs for 2 week beagle dog study 745-02250.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	β_5 (S.E.)	ω_1^2 (S.E.)
(1)	51.803	-8.45 (4)	-	-	-	-	110 (193)
(2)	33.52	-9.85	8.92	9.82	-0.224	-	0.004
(3)	17.974	-35.9	-2.31	0.052	22.1	-14.3	1.2×10^{-7}
(4)	33.541	49.2	50	-0.242	-	-	4.8×10^{-4}
(5)	33.689	49.3	50.2	-0.275	-	-	1.3×10^{-4}
(6)	34.771	48.1 (1.04)	50.7 (3.07)	-	-	-	3.37 (5.48)

4.3.2.2 Study RR-745-02251 – 4 Week Study

4.3.2.2.1 Pharmacodynamic Data

Clinical signs were reported on a daily basis in study 745-02251. The number of convulsions are reported in table 4.31.

Table 4.31. *Number of convulsions in 4 week beagle dog study 745-02251.*

	0 mg/kg	5 mg/kg	10 mg/kg	20 mg/kg	Total
Male	0(84)	0(84)	0(84)	2(67)	2(319)
Female	0(84)	0(84)	0(84)	3(84)	3(336)
Total	0(168)	0(168)	0(168)	4(151)	5(655)

(.) Total number of possible observations.

One male dog in the 20 mg/kg group had convulsions and was sacrificed on day 11. One female dog in the 20 mg/kg group had convulsions but survived to the end of the study.

4.3.2.2.2 Pharmacodynamic Analysis

The models considered for the analysis of the 4 week beagle dog study were:

(1) $\beta_1 + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + b_{1i}$

(3) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(4) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + \beta_4 sex_i + b_{1i}$

(5) $\beta_1 + \beta_2 time_{ij} + \beta_3 sex_i + b_{1i}$

(6) $\begin{cases} \beta_1 + b_{1i}, Dose = 0 \\ \beta_1 + \beta_2 + b_{1i}, Dose = 5 \\ \beta_1 + \beta_3 + b_{1i}, Dose = 10 \\ \beta_1 + \beta_4 + b_{1i}, Dose = 20 \end{cases}$

(7) $\beta_1 + \beta_2 auc_i + b_{1i}$

(8) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + b_{1i}$

(9) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + \beta_4 auc_i time_{ij} + b_{1i}$

(10) $(\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i})time_{ij}$

The results of these models are given in table 4.32.

Table 4.32. *Results of NONMEM runs for dog 4 week study 745-02251.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	42.476	-11.2 (1.31)	-	-	-	65.6 (21.5)	-
(2)	40.457	-22.4	0.876	-	-	5.53	-
(3)	29.601	-33.7	1.58	-0.412	-	4.99	-
(4)	29.569	-30.1	1.4	-0.408	0.382	4.6	-
(5)	31.652	-8.36	-0.447	0.13	-	62.7	-
(6)	40.452	-24.6	-13.4	-13.3	19.9	49.8	-
(7)	36.524	-10.8 (3.25)	3.54 (1.11)	-	-	1.02 (1.72)	-
(8)	27.016	-8.89 (3.07)	3.72 (1.27)	-0.426 (0.186)	-	2.84 (3.46)	-
(9)	27.126	-8.97	3.73	-0.418	-0.0118	3.36	-
(10)	27.015	-8.73	3.65	-0.423	-	2.85	8×10^{-7}

The models with the lowest objective function are those where AUC is a predictor but these are only slightly lower than those in dose. The best model appears to be model (3), the model involving dose and time as linear predictors. Unlike previous models in the rat data, the time coefficient is negative. This implies that as time increases, the probability of observing a convulsion reduces.

4.4 Monkey Studies

There were three Cynomolgus monkey studies, a 2, 4 and 13 week oral toxicity study. In studies 764-01936, 764-02162 and 764-02064, the compound PD-142676 was given orally as this was intended to be the route of administration when given in human clinical trials. The drug was administered by oral gavage as a suspension in 0.5% methylcellulose on a mg/kg body weight basis. Dose selection criteria and monkey numbers will be given in each study section as these altered from study to study.

4.4.1   Pharmacokinetic Analyses of Monkey Data Sets

4.4.1.1 Study RR-764-01936 – 2 Week Study

4.4.1.1.1   Study Design

Two male and 2 female monkeys were assigned at random to each of four dose groups, 0, 5, 10 or 20 mg/kg given orally daily for 2 weeks. On day 10, plasma samples were obtained at each of 7 time intervals post-dose: 0.5, 1, 2, 4, 8, 12 and 24 hours. In the control group, 1 sample was obtained from each monkey 24 hours post-dose.

Doses were selected based on a dose escalation study in which significant clinical effects were not seen until 20 mg/kg was reached. On this basis, the upper dose group was selected as 20 mg/kg with the lower dose groups as 5 and 10 mg/kg.

4.4.1.1.2 Pharmacokinetic Analysis

The number of concentration measurements are given in table 4.33.

Table 4.33. *Number of concentration measurements in monkey 2 week study 764-01936.*

	0 mg/kg	5 mg/kg	10 mg/kg	20 mg/kg	Total
Male	0(2)	14(0)	13(1)	13(1)	40(4)
Female	0(2)	8(6)	12(2)	14(0)	34(10)
Total	0(4)	22(6)	25(3)	27(1)	74(14)

(.) Number of concentrations BLQ.

There are 74 concentration measurements from 12 monkeys. A male monkey in the 5 mg/kg dose group had exceptionally high concentrations uncharacteristic of the other monkeys so this monkey was deleted from the analysis. Animals were deleted form the study if it was considered that an incorrect dose was administered or the data had been reported incorrectly. This reduced the number of concentrations to 67 from 11 monkeys. The models considered in the pharmacokinetic analysis of PD-142676 are given below.

(1) 1 compartment first-order absorption, exponential variability on CL/F, V/F and residual terms.

(2) 1 compartment first-order absorption, exponential variability on CL/F and V/F and combined residual error model.

(3) 1 compartment first-order absorption, full exponential interindividual variance matrix on CL/F and V/F and exponential residual error model.

(4) 1 compartment first-order absorption, exponential variability on CL/F, V/F, k_a and residual terms.

(5) 1 compartment first-order absorption, exponential variability on CL/F, V/F and k_a and combined residual error model.

(6) 2 compartment first-order absorption, exponential variability on CL/F, V/F and residual terms.

The results of these models are given in table 4.34.

Models (2) and (5) appear to show that the addition of an additive component in the residual error model cause downward bias in the estimates of the $V1/F$. From table 4.34, the best model is model (5), the one compartment first-order absorption model with exponential errors on all three parameters and a combined residual error model. Residual plots are shown in figures 4.17-4.19. Figures 4.20 and 4.21 give individual and population fits for two monkeys, one male from the 5 mg/kg group and the other, a female from the 20 mg/kg group.

Table 4.34. *Results of NONMEM runs for 2 week monkey study 764-01936.*

Model	O.F.	Cl/F	(S.E.)	V1/F	(S.E.)	Q/F	(S.E.)	V2/F	(S.E.)	k_a	(S.E.)	ω_{cl}^2	(S.E.)	ω_v^2	(S.E.)	ω_{ka}^2	(S.E.)	σ_1^2	(S.E.)	σ_2^2	(S.E.)
(1)	517.864	21.4	(2.91)	129	(23.7)	-	-	-	-	0.457	(0.122)	0.0849	(0.0243)	0.347	(0.145)	-		0.167	(0.0308)	-	
(2)	510.514	22.2	(3.56)	78.9	(21.7)	-	-	-	-	0.312	(0.058)	0.104	(0.034)	0.33	(0.152)	-		0.11	(0.0334)	29.6	(23.6)
(3)	517.845	21.2	(2.98)	129	(24.1)	-	-	-	-	0.456	(0.12)	0.0862	(0.0252)	0.351	(0.147)	-0.0115[*]	(0.0159)	0.164	(0.0286)	-	
(4)	515.212	21.6	(3.2)	116	(25.2)	-	-	-	-	0.414	(0.151)	0.0906	(0.0215)	0.286	(0.142)	0.169	(0.0934)	0.145	(0.024)	-	
(5)	504.461	23.1	(3.82)	59.4	(6.52)	-	-	-	-	0.253	(0.0526)	0.12	(0.0338)	0.0844	(0.0459)	0.207	(0.0637)	0.0953	(0.0196)	27.9	(20.1)
(6)	516.883	20.5	(3.22)	112	(26.4)	2.56	(4.47)	106	(369)	0.375	(0.114)	0.0791	(0.0269)	0.388	(0.138)	-		0.172	(0.0337)	-	

* Covariance between *Cl/F* and *V/F*.

Figure 4.17. *Residual versus time for model (2) from monkey 2 week study.*

Time(hours)
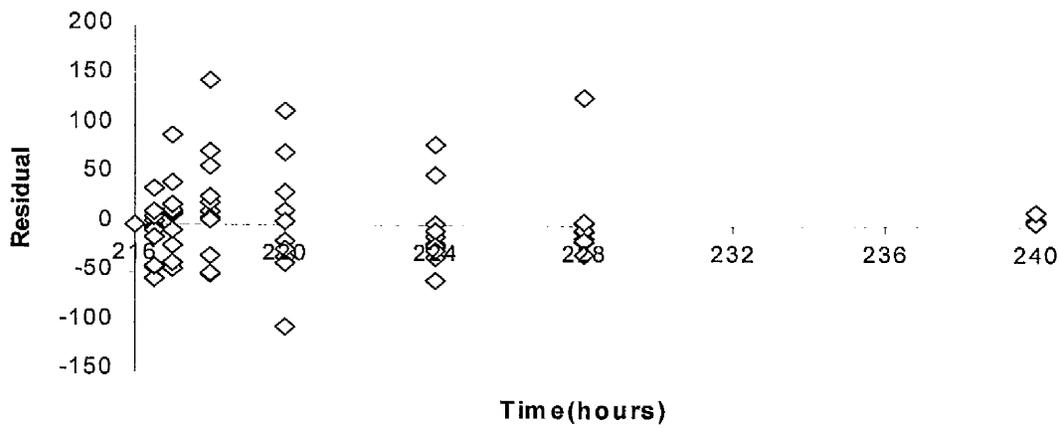
Figure 4.18. *Residual versus predicted for model (2) from monkey 2 week study.*

Population Predicted
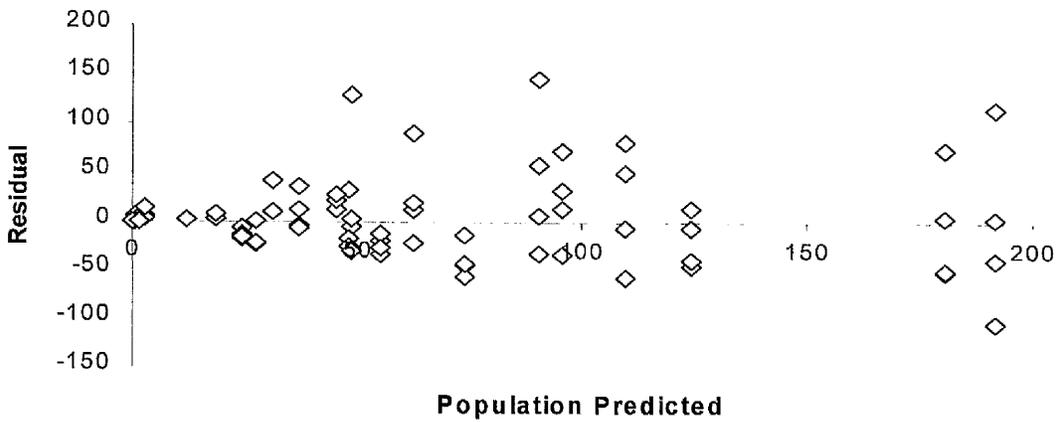
Figure 4.19. Weighted residual versus predicted for model (2) from monkey 2 week study.

Population Predicted

168
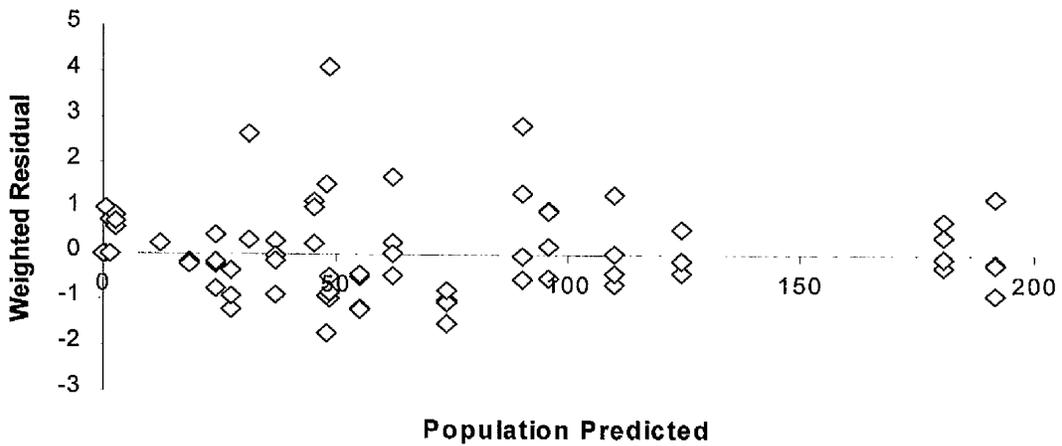
Figure 4.20. *Individual plot for male monkey 681 from 5 mg/kg dose group.*
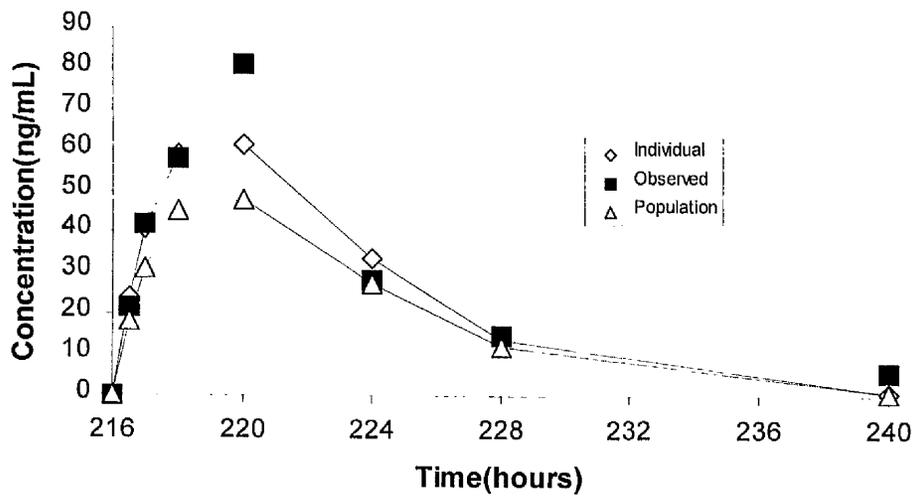
Figure 4.21. *Individual plot for female monkey 693 from 20 mg/kg dose group.*
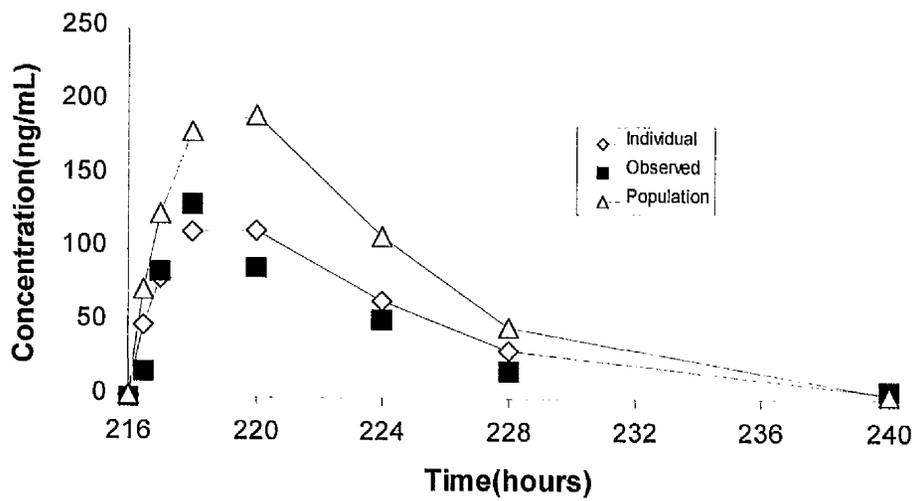
4.4.1.2 Study RR-764-02162 – 4 Week Study

4.4.1.2.1 Study Design

Four groups of 2 monkeys per gender were given PD-142676 as an oral dose suspension once daily for 4 weeks. Male and female monkeys were administered 25, 40, 60 and 120 mg/kg. An additional 2 monkeys per gender were given placebo. Plasma samples were collected on day 9 at 0, 1, 2, 4, 10 and 24 hours post-dose for the 25, 40 and 60 mg/kg dose groups. The placebo group had one sample per monkey taken at 4 hours post-dose. Animals in the 120 mg/kg group were sampled on day 3 at the same times as those in the other groups.

Doses were based on the previous 2 week toxicokinetic study. The initial doses for this study were intended to explore a higher range of doses than had been tested previously in repeated dose regimens.

4.4.1.2.2 Pharmacokinetic Analysis

The number of concentration measurements are given in table 4.35.

Table 4.35. *Number of concentration measurements in 4 week study 764 02162.*

	0 mg/kg	25 mg/kg	40 mg/kg	60 mg/kg	120 mg/kg	Total
Male	0(2)	10(2)	10(2)	6(0)	10(0)	36(6)
Female	0(2)	11(1)	10(2)	11(1)	11(1)	43(7)
Total	0(4)	21(3)	20(4)	17(1)	21(1)	79(13)

(.) BLQ

There are 79 concentrations above the limit of quantification from 15 monkeys. A male monkey was deleted from the study due to exceptionally high concentrations in the 120 mg/kg dose group which resulted in death for the monkey. This left 75 concentrations from 14 monkeys. The models considered were the following.

(1) 1 compartment first-order absorption, exponential variability on CL/F, V/F and residual terms.

(2) 1 compartment first-order absorption, exponential variability on CL/F, V/F and k_a and exponential residual error model.

(3) 1 compartment first-order absorption, exponential variability on CL/F, V/F and k_a and combined residual error model.

The results are given in table 4.36.

Table 4.36. *Results of NONMEM runs for monkey 4 week study 764-02162.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	ω_{ka}^2 (S.E.)	σ_1^2 (S.E.)	σ_2^2 (S.E.)
(1)	693.2	21.9 (2.91)	40.4 (9.27)	0.193 (0.011)	0.106 (0.053)	0.433 (0.361)	-	0.272 (0.049)	-
(2)	692.16	22.4 (3.13)	46 (11.2)	0.2 (0.01)	0.108 (0.057)	0.399 (0.259)	0.07 (0.14)	0.258 (0.049)	-
(3)	691.29	23 (3.22)	58.3 (16.5)	0.228 (0.033)	0.124 (0.064)	0.328 (0.201)	0.221 (0.295)	0.22 (0.068)	23.2 (35.7)

As all the objective function values are similar, model (1) was chosen as it was the simplest. Residual plots for model (1) are given in figures 4.22 and 4.23.

Figure 4.22. *Residual versus population predicted for model (1) in monkey 4 week study.*
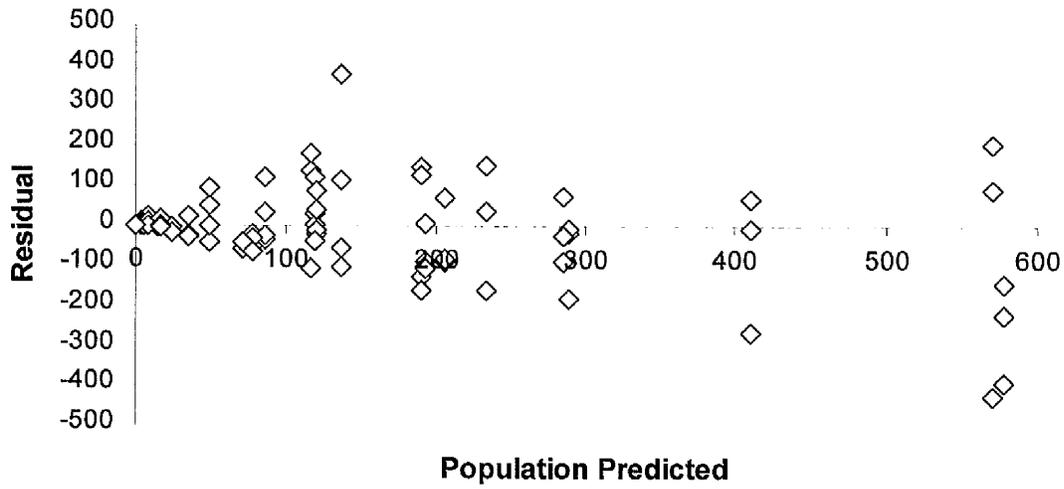
Figure 4.23. *Weighted residual versus population predicted for model (1) in monkey 4 week study.*
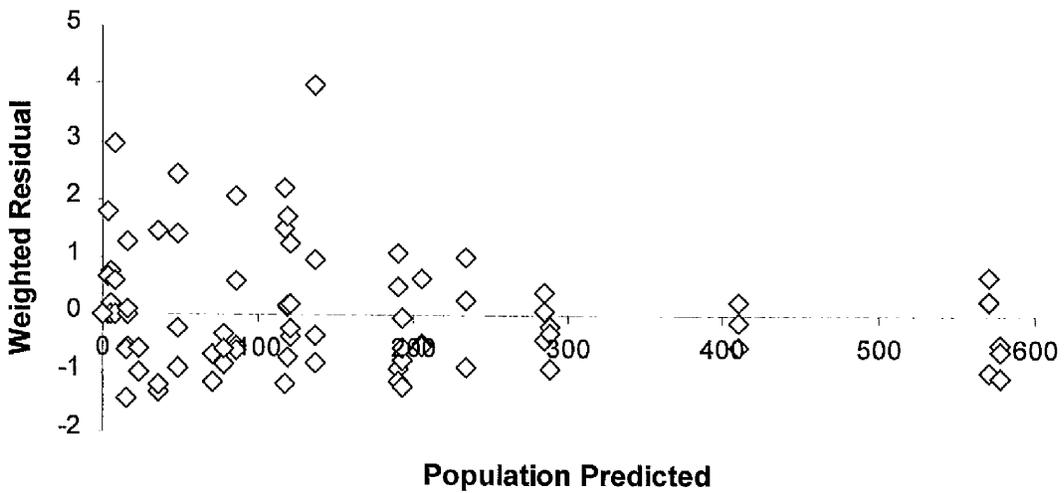
4.4.1.3 Study RR-764-02064 – 13 Week Study

4.4.1.3.1 Study Design

Three groups of 4 monkeys per gender were given PD-142676 as an oral dose solution once daily for 13 weeks. Male and female monkeys were administered PD-142676 at 1, 5 or 20 mg/kg. An additional 4 monkeys per gender were given dosing vehicle and used as controls. Plasma samples were collected at 0, 1, 2, 4, 10 and 24 hours post-dose.

The doses were selected based on the previous 2 and 4 week toxicokinetic studies. The high dose selected for this study, 20 mg/kg, was expected to cause significant clinical signs in at least some monkeys. The low dose,1 mg/kg, was expected to be a no-effect dose or to elicit mild clinical signs, 5 mg/kg was chosen as the middle dose.

4.4.1.3.2 Pharmacokinetic Analysis

The number of concentration measurements are given in table 4.37.

Table 4.37. *Number of concentrations in monkey 13 week study 764-02064.*

	1 mg/kg	5 mg/kg	20 mg/kg	Total
Male	4(20)	9(15)	20(4)	33(39)
Female	5(19)	11(13)	13(11)	29(43)
Total	9(39)	20(28)	33(15)	62(82)

(.) BLQ

There were 62 concentration measurements from 20 monkeys. The models considered for this data were the following.

(1) 1 compartment first-order absorption, exponential error on CL/F, V/F and residual terms.

(2) 1 compartment first-order absorption, exponential error on CL/F and V/F and combined residual error model.

(3) 1 compartment first-order absorption, full interindividual variance matrix on CL/F and V/F and combined residual error model.

(4) 1 compartment first-order absorption, exponential error on CL/F, V/F, k_a and residual terms.

The results of these models are given in table 4.38.

Table 4.38. *Results of NONMEM runs for monkey 13 week study.*

Model	O.F.	Cl/F (S.E.)	V/F (S.E.)	k_a (S.E.)	ω_{Cl}^2 (S.E.)	ω_V^2 (S.E.)	ω_{ka}^2 (S.E.)	σ_1^2 (S.E.)	σ_2^2 (S.E.)
(1)	395.34	28.9 (3.58)	99.5 (19.6)	0.49 (0.076)	0.16 (0.066)	1.73 (0.862)	-	0.101 (0.031)	-
(2)	390.0	32.8 (3.07)	197 (32.5)	0.401 (0.075)	0.0002 (0.035)	2.79 (1.74)	-	0.0782 (0.028)	53.2 (23.4)
(3)	395.9	53.2 (22.5)	77.4 (27.8)	0.325 (0.053)	1.02 (1.41)	0.184 (0.318)	0.113* (0.45)	0.059 (0.066)	52.8 (25.7)
(4)	404.76	33.2 (4.02)	32.4 (8.06)	0.372 (0.036)	0.244 (0.095)	0.048 (0.066)	1.08 (0.428)	0.06 (0.021)	-

* Covariance between CL/F and V/F.

The model best describing this data set is model (1). Although model (2), has an objective function value which is 5 points lower, there is an extra residual error term and does not add much to the description of the variability around the population mean. Residual plots are given in figures 4.24 and 4.25.

Figure 4.24. *Residual versus predicted plot for 13 week monkey study.*

Population Predicted
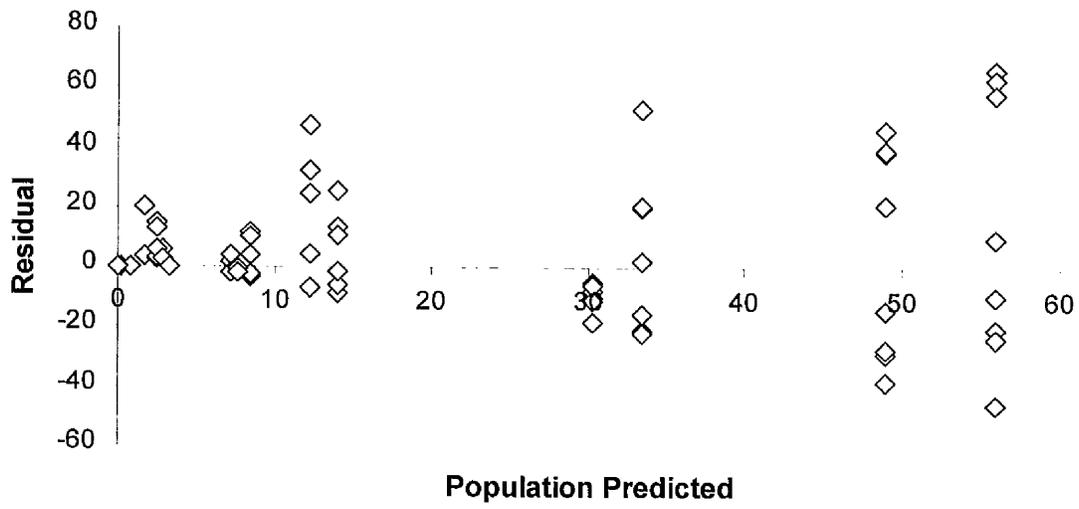
Figure 4.25. *Residual versus predicted plot for 13 week monkey study.*

Population Prediction
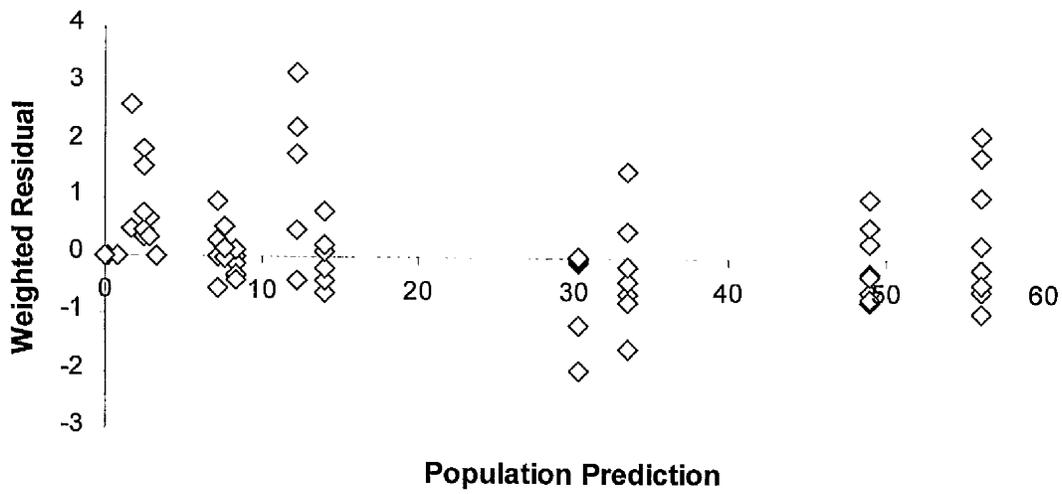
4.4.2 Pharmacodynamic Analyses of Monkey Data Sets

4.4.2.1 Study RR-745-02083 – 2 Week Study

4.4.2.1.1 Pharmacodynamic Data

Clinical signs were observed on a daily basis but only reported and summarised as being from a particular week. The types of clinical signs and symptoms observed in monkey were anorexia, emesis, alopecia, skin soreness, convulsions and comatose. Convulsions were not observed in any of the monkeys so there was no data to analyse in the 2 week study.

4.4.2.2 Study RR-745-02236 – 4 Week Study

4.4.2.2.1 Pharmacodynamic Data and Analysis

Clinical signs were reported on a daily basis in study 745-02236. The number of convulsions are given in table 4.39.

Table 4.39. *Number of convulsions in monkey 4 week study 745-02236.*

	0 mg/kg	25 mg/kg	40 mg/kg	60 mg/kg	120 mg/kg	Total
Male	0(56)	0(56)	0(56)	1(29)	1(8)	2(205)
Female	0(56)	0(56)	4(45)	1(48)	1(10)	6(215)
Total	0(112)	0(112)	4(101)	2(77)	2(18)	8(420)

(.) Total number of possible observations.

There were 420 observations, 8 of which are convulsions from 20 monkeys.

The models considered for this study were the following.

(1) $\beta_1 + b_{1i}$

(2) $\beta_1 + \beta_2 dose_i + \beta_3 time_{ij} + b_{1i}$

(3) $\beta_1 + \beta_2 time_{ij} + b_{1i}$

(4) $(\beta_1 + b_{1i}) + \beta_2 dose_i + (\beta_3 + b_{2i})time_{ij}$

(5) $\begin{cases} \beta_1 + b_{1i}, Dose = 0 \\ \beta_1 + b_{1i}, Dose = 25 \\ \beta_1 + \beta_2 + b_{1i}, Dose = 40 \\ \beta_1 + \beta_3 + b_{1i}, Dose = 60 \\ \beta_1 + \beta_4 + b_{1i}, Dose = 120 \end{cases}$

(6) $\begin{cases} \beta_1 + \beta_5 time_{ij} + b_{1i}, Dose = 0 \\ \beta_1 + \beta_5 time_{ij} + b_{1i}, Dose = 25 \\ \beta_1 + \beta_2 + \beta_5 time_{ij} + b_{1i}, Dose = 40 \\ \beta_1 + \beta_3 + \beta_5 time_{ij} + b_{1i}, Dose = 60 \\ \beta_1 + \beta_4 + \beta_5 time_{ij} + b_{1i}, Dose = 120 \end{cases}$

(7) $\beta_1 + \beta_2 auc_i + b_{1i}$

(8) $\beta_1 + \beta_2 auc_i + \beta_3 time_{ij} + b_{1i}$

(9) $(\beta_1 + b_{1i}) + \beta_2 auc_i + (\beta_3 + b_{2i})time_{ij}$

The results of these models are given in table 4.40.

The model of choice from the models shown is model (2) which is linear in dose and time. The same model in AUC has a slightly lower objective function value but requires the computation of the AUC. The models with two random effects also have

slightly lower values for the objective function but not enough to warrant the use of another random effect in the model.

Table 4.40. *Results of NONMEM runs for monkey 4 week study.*

Model	O.F	β_1 (S.E.)	β_2 (S.E.)	β_3 (S.E.)	β_4 (S.E.)	β_5 (S.E.)	ω_1^2 (S.E.)	ω_2^2 (S.E.)
(1)	40.175	-9.98 (2.22)	-	-	-	-	70.3 (80.2)	-
(2)	34.236	-21.8	0.0608	0.389	-	-	176	-
(3)	35.445	-19.2	0.326	-	-	-	280	-
(4)	32.786	-15.1	0.005	0.0213	-	-	216	2.09
(5)	36.933	-14.5	8.85	10.7	10.8	-	2.97	-
(6)	32.659	-42	4.9	6.44	26.1	0.949	648	-
(7)	39.875	-10.6 (2.53)	0.434 (0.26)	-	-	-	58.1 (79.4)	-
(8)	34.007	-38.4	0.9	0.908	-	-	916	-
(9)	32.147	-34.8	3.47	0.293	-	-	552	6.12

4.4.2.3 Study RR-745-01694 – 13 Week Study

4.4.2.3.1   Pharmacodynamic Data and Analysis

Convulsions were observed on a daily basis but were reported as being from a particular week. The number of convulsions are given in table 4.41.

Table 4.41. *Number of convulsions in monkey 13 week study 745-01694.*

	0 mg/kg	1 mg/kg	5 mg/kg	20 mg/kg	Total
Male	0(52)	0(52)	0(52)	0(52)	0(208)
Female	0(50)	1(52)	0(52)	0(52)	1(206)
Total	0(102)	1(104)	0(54)	0(104)	1(414)

(.) Total number of possible observations.

Out of 414 observations from 32 monkeys, there was only one convulsion from a female in the 1 mg/kg dose group. The ability to estimate any sort of model from this data is negligible so no data analysis was carried out.

4.5 Discussion

The analysis of the rat pharmacokinetic data sets did not prove to be an easy task. The rat 2 week data set was from a destructive sampling scheme with only one sample available per rat. This meant that it was only possible to obtain one level of random variability and that was on the inter-rat level. When inter-rat variability was estimated as a function of the model (one compartment first-order absorption at steady state), the constant coefficient of variation (exponential error model) was estimated as approximately 55% whereas the additive error model gave an estimate of the variability of 0.16 (standard deviation). The parameter estimates of the apparent clearance and volume were approximately the same for different error models, but the first-order absorption rate constant changed by an order of magnitude. This implied that the absorption phase was poorly determined as could be seen from an examination of the

data. When the inter-rat variability was estimated as a component of the apparent parameters, there was a lack of information as could be seen from the estimates of the variability as they altered markedly under different assumptions. Supposedly better estimates of the variability on the parameters were determined when a zero-order absorption model was chosen and proportional errors were assigned to the parameters Cl/F and V/F although this did not prove to be the best model of those considered. The best model considered was one where fixed effects were estimated for each dose level. This resulted in the lowest objective function value whether a first or zero-order absorption model was assumed. From the fixed effects model, as dose increased apparent clearance decreased implying nonlinear pharmacokinetics.

The rat 4 week data set had three concentration measurements from each rat that allowed the estimation of interindividual as well as residual variability. Again, one compartment models at steady state were used with first and zero-order absorption models. For this data set, there was considerable difficulty in estimating the apparent volume term. It appeared to be better to use an exponential error model on the parameters and residual term, but a combination of additive and exponential errors caused the volume term to vary considerably under different model assumptions. The absorption rate constant was difficult to estimate due to the absence of data in the absorption phase and the infusion time in the zero-order model was also difficult to estimate. Inter-rat variability was also difficult to determine with parameter estimates varying by 6 orders of magnitude under different assumptions. The best model appeared to be the first-order absorption model with exponential error model for the inter-rat and residual variability.

For the 13 week rat data set, there were multiple observations from each rat so both levels of variability were estimated. The same models were considered as in the two previous data sets with the fixed effects models on dose giving the lowest objective function values. As with the two week data set, there appeared to be an increase in clearance as dose decreased but this was only the case when an exponential error model was assigned to the inter-rat and residual variability. A one compartment first-order absorption mixed effects model resulted in an increase in the objective function by 10 points from the fixed effects models with a difference of 6 fixed effects parameters. It was decided that the best model of those considered was the one compartment first-order absorption model at steady state.

As none of the three rat data sets were particularly well estimated, all three sets of data were combined to see if this would improve estimation of the inter-rat variability in particular. As was the case with the three separate data sets, the best model was a one compartment first-order absorption model with exponential error terms. Even with this model specification and combined data, it was still difficult to estimate inter-rat variability for V/F and also difficult to estimate the absorption rate constant.

The 2 week dog study had plasma concentration data collected on two days which meant that a multiple dose model was needed. There was only a total of 47 concentration measurements taken and so there was not much data with which to develop a model. Only one compartment models were considered. The best model of the four considered was the bolus model with exponential error on the parameters and residual term. A simple model such as this was probably all that could be estimated with such limited data.

The 4 week dog study allowed the estimation of interoccasion variability as plasma samples were collected on three separate occasions although PD-142676 was administered once daily. From the data analysis, it was found that including an interoccasion component in the model did not improve the model an appreciable amount. Adding another compartment into the model did not improve the fit either and the best model of those considered was a one compartment first-order absorption model with exponential error on all three parameters and on the residual error term.

There were three monkey data sets. Similar results were found in these studies as for the rat and dog studies. For the 2 week monkey data set, the best model was a one compartment first-order absorption model with exponential error on all three parameters and a combined residual error model. For the 4 week data set, the best model was again a one compartment first-order absorption model with exponential error on apparent clearance and volume and exponential residual error. For the 13 week data set, the best model was of the same type as for the 4 week data set. Throughout all the toxicokinetic data sets, the model that systematically described adequately the different species data sets (except for dog 2 week data) was the one compartment first order absorption model. However, even though a first order absorption model was chosen, this was consistently difficult to estimate. Nothing more complicated could be discerned from the data, probably due to the low number of concentration measurements and high levels of variability in the data sets.

In preclinical studies, one of the main aims is to accrue information about the pharmacokinetics and pharmacodynamics of the drug to aid in the study of the drug in humans. One way of doing this is by looking at relationships between species in terms

of body weight and then extrapolating to man (Cosson *et al* (1997)). As the human data was not considered, this was not possible but it was possible to look at the scaling between the three species analysed here. As the drug was given normalised to the animals weights, the apparent parameters estimated previously needed to be scaled to be independent of weight. The results are given in table 4.42.

Table 4.42. *Values of apparent clearance volume independent of body weight.*

Species	Time scale	Body Weight (kg)	Cl/F/kg	V/F/kg	Cl/F	V/F
Rat	2 week	0.268	10.7	59.2	2.87	15.87
	4 week	0.303	5.17	102	1.57	30.91
	13 week	0.411	6.19	56.4	2.54	23.18
Dog	2 week	11.577	4.86	34.8	56.26	402.88
	4 week	10.777	16	103	172.43	1110.03
Monkey	2 week	3.963	23.1	59.4	91.55	235.40
	4 week	3.707	21.9	40.5	81.18	150.13
	13 week	3.502	28.9	99.4	101.21	348.10

The values chosen for the apparent clearance and volume were from models which were analysed from each specific data set alone, rather than combined data. The values of the body weights in table 4.42 are the arithmetic means of the animals in the particular data sets. In figures 4.26 and 4.27 are log-log plots of the apparent clearance and volume estimates versus average body weight. As no information was available on bioavailability, absolute clearance and volume could not be estimated. As can be seen, the log-log plot for apparent volume shows very good linearity implying that a power function in terms of body weight would be an appropriate model for apparent volume. The log-log plot for apparent clearance was not quite as good but was still reasonable. It was surprising that such a good plot was found given bioavailability was not accounted for.

183

Figure 4.26. *Plot of apparent clearance versus average body weight.*
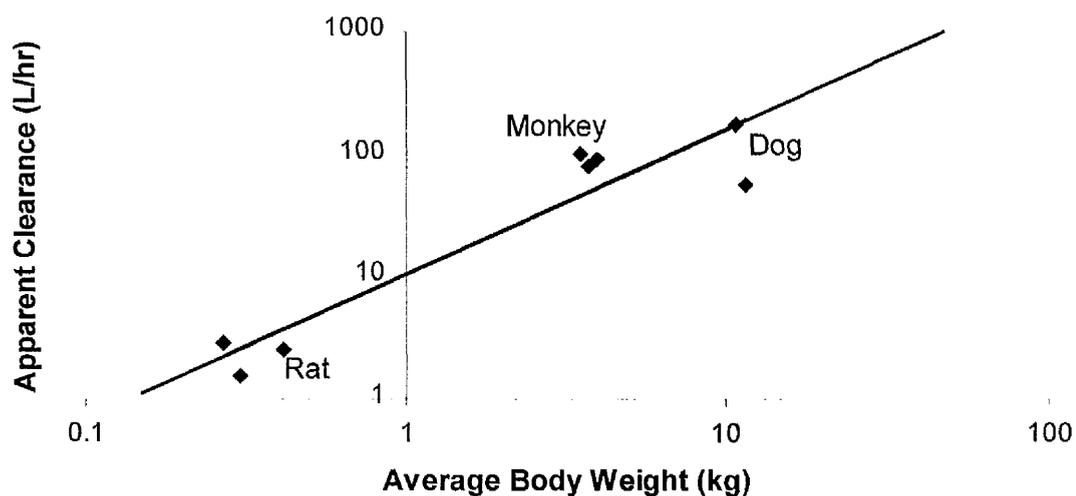

Figure 4.27. *Plot of apparent volume versus average body weight.*
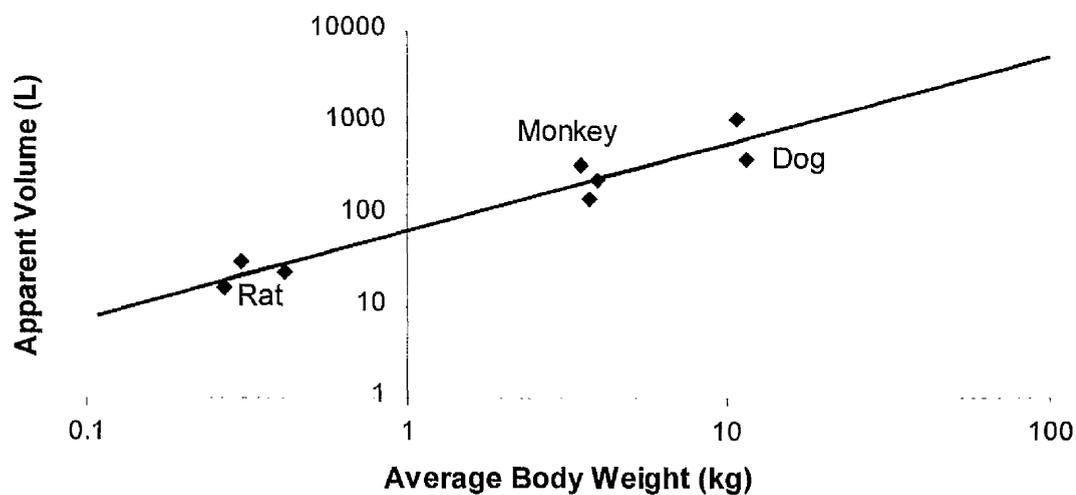
The results of the power function fitted to the data are given in table 4.43 and graphs are given in figures 4.26 and 4.27. These results were obtained from Splus 4.5.

Table 4.43. *Results of power function fitted to apparent parameters versus average body weight* $\log(Cl(V)/F) = \alpha + \beta \log(weight)$.

	Parameter	Value	Standard Error	t value	Pr(>|t|)	R^2
Apparent Clearance	α	2.351	0.279	8.440	0.0002	0.886
	β	1.180	0.173	6.826	0.0005	
Apparent Volume	α	4.198	0.174	24.132	0.000	0.928
	β	0.949	0.108	8.793	0.0001	

The pharmacodynamics of PD-142676 was difficult to analyse due to the sparseness of the data (lack of convulsions). As well as this, the data were binary corresponding to whether a convulsion was observed or not. In several of the data sets, there were only a few convulsions and in one data set, no convulsions were observed. In such a setting, it was highly unlikely that any model of interest would be found to describe the data and allow inferences to be made.

For the rat pharmacodynamic data, the quality of data varied considerably between studies. In the two week data set, there were only 2 observed convulsions from 379 observations and one of these rats died post convulsion. The model that was selected did not include dose or AUC as a predictor of the drug response demonstrating the limited information in the data. In this particular case, including pharmacokinetic parameters into the model would not have improved the modelling. The objective function values were all small in magnitude as nearly all the observations were zero and the model was adequate in describing such data. For the 13 week data set, there were 21 convulsions from a possible 2433 observations. This was better for modelling purposes as there were more convulsions to model but the probabilities of observing convulsions for different doses and times were very small. The best model chosen was a linear logistic model in terms of dose and time with additive subject specific random

effects on the intercept and time parameters. A similar model in terms of AUC had a lower objective function value (by 1 point) but was not chosen because it involved the estimation of individual AUC values. Including the random effect on the time parameter caused the dose parameter estimate to switch sign. This was worrying because it makes it difficult to interpret what the random effect in this particular case was adding.

The 4 week rat data set was more informative as 93 convulsions were observed out of a total of 705 observations. In this data set, there was a problem with missing data as many of the rats that had convulsions subsequently died. This meant that to see how the missing data affected the results of the modelling, a sensitivity analysis was carried out where different imputing methods were considered. By imputing the missing data as the last observation carried forward or as all 0's resulted in the graphs in figure 4.12, being similar to analysing the data as though the data were missing completely at random. The biggest change came by imputing the data as all 1's and produced a markedly different graph on the 4 week time scale. By imputing the data as all 1's resulted in all the parameters increasing in magnitude and imputing the missing data as 0's resulted in a decrease in the magnitude of all the parameters. By imputing the data as the last observation carried forward resulted in parameter values similar to those of imputing the missing data as all 1's because most of the missing data followed convulsions. From the four different data sets, the models were variations on linear dose and time. The unimputed data set included sex as a covariate whereas the imputed data sets did not. It appears that for whichever imputation method was chosen, approximately the same model was obtained.

For the 2 week dog pharmacodynamic data, there were 8 observed convulsions from a possible 141 observations. This data set proved difficult to analyse as models of the form from the 13 week rat data could not be estimated. The models that were chosen were dose specific models, and the model chosen as the best model out of the ones considered was where the intercept was fixed for the two lowest doses and for the 20 and 60 mg/kg dose groups, dose specific intercepts were estimated. Dogs were withdrawn for reversibility studies as 6 dogs had convulsions after just one dose. This meant that convulsions were observed at earlier but not later times and so there was very little information with which to define a model for such data. In the 4 week dog data, there were 5 observed convulsions from a total of 655 observations. A linear dose and time model was chosen but the time coefficient was negative. These convulsions occurred earlier in the study which probably caused the time parameter estimate to be negative. The standard errors could not be estimated implying that the parameters were probably not well identified and there was limited information in the data.

The monkey data had very few observed convulsions which made modelling difficult or even unnecessary. In the 2 week monkey data set, no convulsions were observed so no modelling was required. In the 4 week monkey data, 8 convulsions were observed out of a possible 420 observations. Again the model chosen was a model linear in dose and time. The 13 week monkey data set had 1 observed convulsion from a total of 414 observations and so it was considered worthless to model such data.

As well as scaling the pharmacokinetics between species, it was of interest to look at the scaling of the pharmacodynamics between species. This was more difficult because unlike the pharmacokinetics, the models were empirically based. No mechanism for the

action of PD-142676 was proposed for the modelling and so simple exploratory models were examined. A comparison of the dose or time parameter estimates could have been made but because of the different models, this would have been difficult to carry out and interpret. Instead, a method of looking at the graphical representation of the models was considered to see how well the graphs overlapped. This would allow a way of checking how well the models from shorter time studies were at predicting longer time studies in a particular species and across species. The logistic curves for the rat data are shown in figure 4.28. The 2 week data did not include dose as a predictor and so no dose is shown.

Figure 4.28. *Plot of 2 and 4 week rat models.*
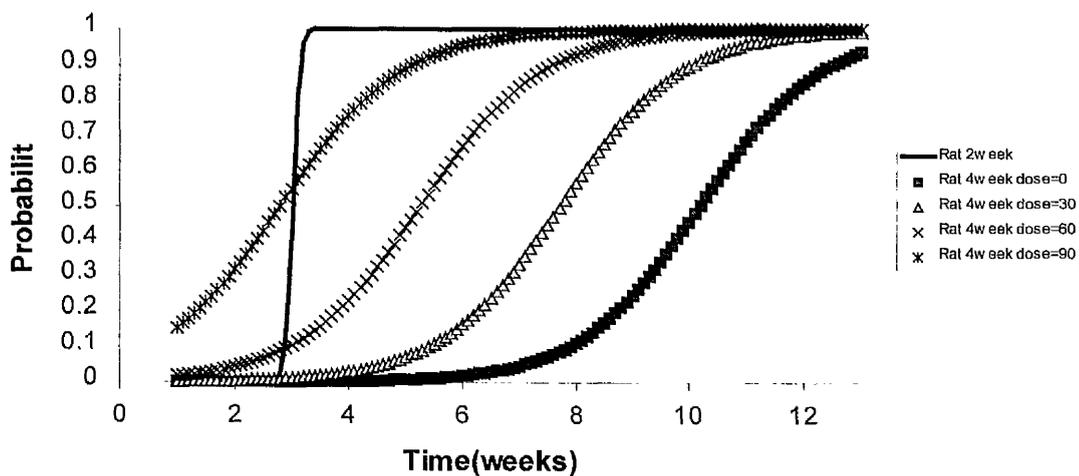
The time parameter for the rat 2 week data was large and so when time was greater than the length of the actual study, the model predicted that there was a certain chance (Pr(Y=1)=1) of observing a convulsion. The 4 week rat data only reached high probabilities of convulsion when a high dose was given and the end of the study was approached. The 13 week rat study produced very low probabilities as shown in figure

4.29 which would have resulted in a straight line in figure 4.28. The different rat data sets did not fit together, as can be seen from the two figures.

Figure 4.29. *Plot of 13 week rat study.*
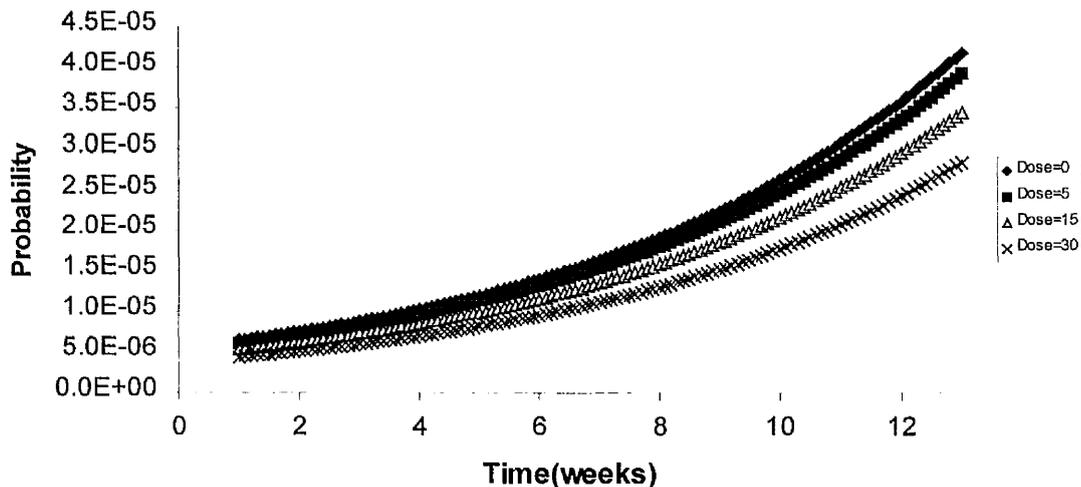
As the dog data produced negative time parameter estimates which was different to the other species studies, this would not allow scaling at all to the other species. Also, partly because of the difference in the number of convulsions in the two data sets, the models did not scale well between the two dog studies either.

The 2 week monkey pharmacodynamic data did not have any observed convulsions so no analysis took place. For the 13 week pharmacodynamic data, there was only one convulsion so again no modelling took place. For the 4 week pharmacodynamic data, a plot of the model is given in figure 4.30.

To see how well the logistic curves superimpose on each other, it was assumed a 20 mg/kg dose was given and then the curves were plotted over the longest study period to

see how well they would extrapolate and relate to each other. As can be seen, the curves are markedly different implying that the curves do not scale well.

Figure 4.30. *Plot of 4 week monkey pharmacodynamic model.*
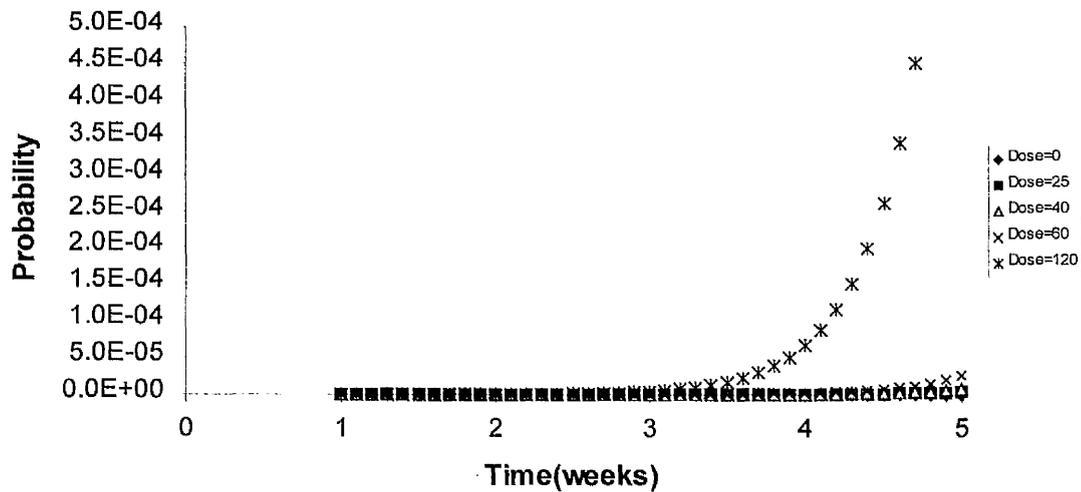
Figure 4.31. *Plot of 4 different studies based on given a 20 mg/kg dose.*
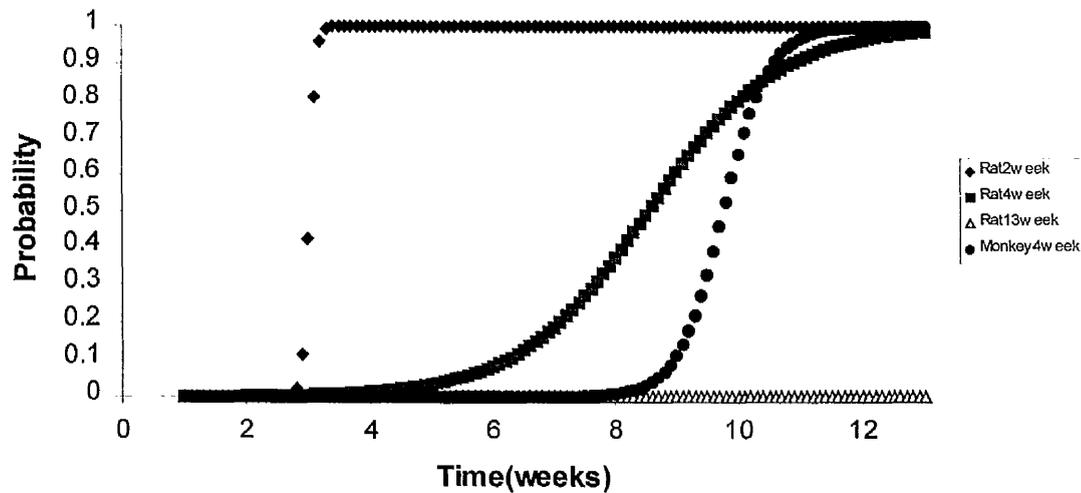
There are many possible reasons why the pharmacodynamic models did not scale well in this case. Firstly, one problem could be due to the data being binary. Categorical data contains less information than continuous data and trying to scale models that are

based on response variables that are binary is a difficult task and requires well designed and rich data settings.

Another problem with the modelling and extrapolating of the models was the use of different time scales for the time covariate. In some studies, time was measured in days and in others, time was in weeks. This meant that the parameter estimates were on different scales for the time effect which would have affected the results. Also, because the time effect was sometimes measured in weeks, this was less informative than if the time was in days because the time measure was split into smaller components giving more precise times. When time was measured in weeks, even if there had been more than one convulsion in a week, this would still have been considered as observing one convulsion in a week and no account was taken of the frequency in a particular week.

The different lengths of the studies gave the opportunity to study the extrapolation of the shorter studies to the longer studies but this proved not to be possible. Because of the difference in the number of convulsions in each study, there was a lot of variability in how well the models could be defined. Some models were difficult to estimate and not much confidence can be put in these models as they were poorly determined. Estimation of standard errors also seemed to be a serious problem. Extrapolation for these data should not have been considered due to the lack of quality of the data. The different number of animals in each study probably also contributed to a difficult task of scaling the logistic curves as different models were based on different numbers of observations.

Scaling pharmacodynamics is still possible but not from the approach taken here for such data. Only one pharmacodynamic measure was chosen but there were several measures that could have been taken as the response variable. Instead of taking a univariate approach to the problem, all the response variables could be analysed in a multivariate form which would hopefully allow more information for scaling purposes. Drugs do not have a singular effect but instead multiple effects and it seems sensible to analyse the pharmacodynamics in such a way. This would automatically get rid of the problem of choosing which variable to take as the response variable. Scaling of the pharmacodynamics would be very useful in planning phase I clinical trials and so suitable doses could be selected without severe toxic effects being encountered. The advantage of pharmacodynamic scaling is that it may be possible to predict the pharmacological activity in man from preclinical studies.

5. Sumatriptan – Phase II Data Set

5.1 Review of Sumatriptan

Sumatriptan is a selective agonist at a subtype of the 5-hydroxytryptamine-1 (5-HT1) receptor. Sumatriptan mediates vasoconstriction of cranial blood vessels which may form the basis of its efficacy in migraine and cluster headache. Sumatriptan is currently marketed as a subcutaneous injection and oral tablet with dose ranges 1-16 mg and 25-400 mg respectively. Migraine episodes are often associated with nausea and vomiting and patients suffering these symptoms may find it difficult to take an oral formulation. A subcutaneous injection is a possible alternative but injections are disliked by some people. A nasal spray of sumatriptan has been formulated as a way of bypassing the problems associated with the oral and subcutaneous formulations.

Duquesnoy *et al* (1998), report a comparative study on the clinical pharmacokinetics of single doses of sumatriptan following subcutaneous, oral, rectal and intranasal formulations in 23 patients. The pharmacokinetics of intranasal sumatriptan showed multiple peaks in most of the 23 subjects. The intranasal formulation performed similarly to that of the oral tablet in terms of absorption and excretion but this could possibly be due to a large amount of the drug being swallowed. The bioavailability for the intranasal and oral formulations relative to the subcutaneous was 15.8% and 14.2% respectively. C_{max} and t_{max} were similar for intranasal and oral. The pharmacokinetic analysis presented by Duquesnoy *et al* (1998) could not explain the observed higher efficacy in terms of migraine relief at early time points after intranasal administration

compared to oral. A deconvolution method was used to obtain an absorption rate-time profile for the formulations. From these profiles, it could be seen that the average profile for the intranasal spray peaked much earlier than that for the oral tablet. This could help explain why more subjects had greater migraine pain relief at earlier times when using the intranasal spray rather than the oral tablet.

The following work was based on a dose ranging study into intranasal sumatriptan. The purpose of the work was to carry out a population analysis of the pharmacokinetic and pharmacodynamic data. The computer package BUGS version 0.6 was used for the analysis. In this version of BUGS, the Griddy-Gibbs sampling algorithm was implemented, so nonlinear pharmacokinetic models could be analysed. The pharmacokinetic and pharmacodynamic analyses were carried out separately and an attempt at a simultaneous analysis of the pharmacokinetic/pharmacodynamic data was made. The pharmacokinetic and pharmacodynamic data were measured longitudinally up to 2 hours. For many patients, this corresponded to being in the absorption phase of the drug (in the study of Duquesnoy *et al* (1998), the range for t_{max} was 0.25-3 hours).

5.2 Study Design

The design of the study was a double-blind, placebo-controlled, parallel group, randomised study and was carried out on a multicentre, multinational basis. The dose levels considered were 0, 2.5, 5, 10 and 20 mg. The primary study objective was to compare the efficacy of intranasal sumatriptan to that of placebo in the acute treatment of migraine in terms of headache relief and resultant severity grades post-dose.

Secondary objectives were to compare the efficacy of intranasal sumatriptan with placebo in terms of resolution of nausea, vomiting, photophobia/phonophobia, reduction of patients' clinical disability scores and the time to meaningful headache pain relief. Data was only available for the primary objective and so the secondary objective was not considered here.

Based on the primary objective of the study reported in the trial protocol (to compare improvement rates at 120 minutes between 10 mg and placebo), assuming 70% improvement on 10 mg and 25% improvement on placebo from baseline migraine severity, 22 evaluable patients per group were required to detect this difference at the 5% level of significance, with 80% power. The actual number of patients in each arm of the study available for analysis were the following: 0 mg = 11 patients, 2.5 mg = 15 patients, 5 mg = 19 patients, 10 mg = 18 patients and 20 mg = 20 patients. The concentrations were measured in ng/mL with a limit of detection of 1 ng/mL.

The pharmacodynamic variable, pain severity, was measured on a categorical scale defined as the following: 0 = no pain, 1 = mild pain, 2 = moderate pain and 3 = severe pain.

The number of time points at which pharmacokinetic and pharmacodynamic measurements were taken differed for the two response variables. There were 8 possible concentration measurements and 6 possible migraine pain severity measurements per individual. The times (in minutes) at which these measurements were taken are given in table 5.1.

Table 5.1. *Times at which concentration and pain severity measurements were taken.*

Concentration times(mins)	0	10	-	20	30	45	60	90	120
Pain Severity times(mins)	0	-	15	-	30	-	60	90	120

It is not known what selection criterion was used to determine these time points for the pharmacokinetic and pharmacodynamic measurements. Patients were allowed to drop out at any time if it was felt it was to the detriment of the patient to continue the study or the patient reported inadequate migraine relief.

5.3 Pharmacokinetic Analysis

The total number of concentrations measured from the 83 patients not including 0 ng/mL concentrations (time = 0 minutes) or concentrations below the level of detection, were 364. The mean number of concentration measurements per individual not including the placebo group was 5.06 with a range of 0-7 observations per individual. Individuals with no concentrations (in active treatment groups) did not contribute to the parameter estimates.

The pharmacokinetic analysis was carried out using BUGS 0.6. As reported in Duquesnoy *et al* (1998), the time over which the data were collected (0-120 minutes) corresponds to the absorption phase in most individuals. For this reason and the fact that there was no information on the disposition of intranasal sumatriptan in these data, an empirical model was used. A standard model could have been used where the distribution and elimination parameters were fixed to published results on intranasal

sumatriptan but this was decided against because Cosson and Fuseau (1998) reported in a population analysis of a set of oral formulation data that the best model describing the data was a 2 compartment model with a first order absorption process followed by a zero order absorption process. This would have been a complicated model to estimate in BUGS without the Metropolis algorithm. Instead, the empirical model defined in equation (5.1) was used to describe the data.

$$E(C) = Dose \times e^{(\beta_1 + \beta_2 time)} (1 - e^{-\beta_3 time}) \qquad (5.1)$$

This model is sufficiently flexible to allow the concentration data observed in this data set to be modelled adequately. The model is shown in figures 5.1-5.3 where in each figure, one parameter is altered at a time to show the effect of such a change.

Figure 5.1. Plot of equation (5.1) with default parameter values given by $(\beta_1,\beta_2,\beta_3)$ = $(0,-0.005,0.1)$where β_1 changes value.
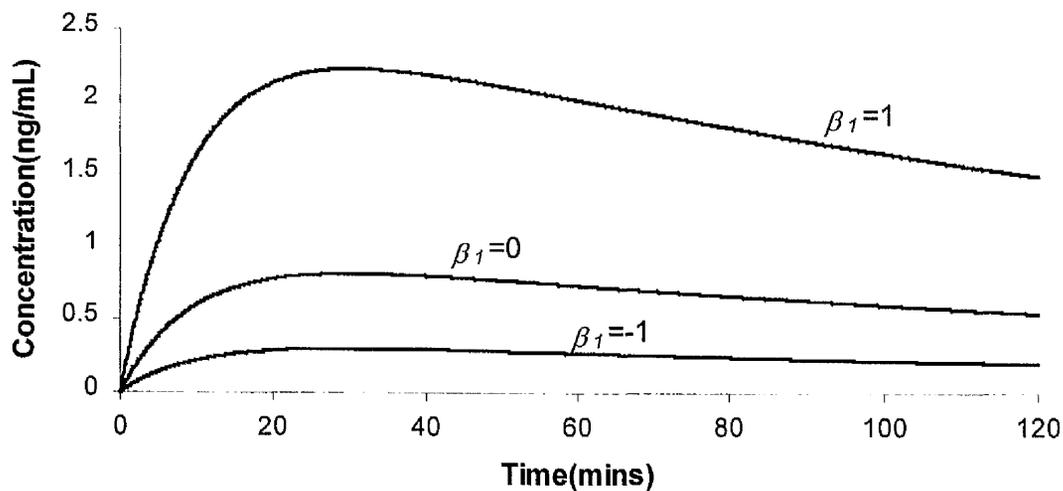
Figure 5.2. Plot of equation (5.1) with default parameter values given by $(\beta_1,\beta_2,\beta_3) =$ (0,-0.005,0.1)where β_2 changes value.
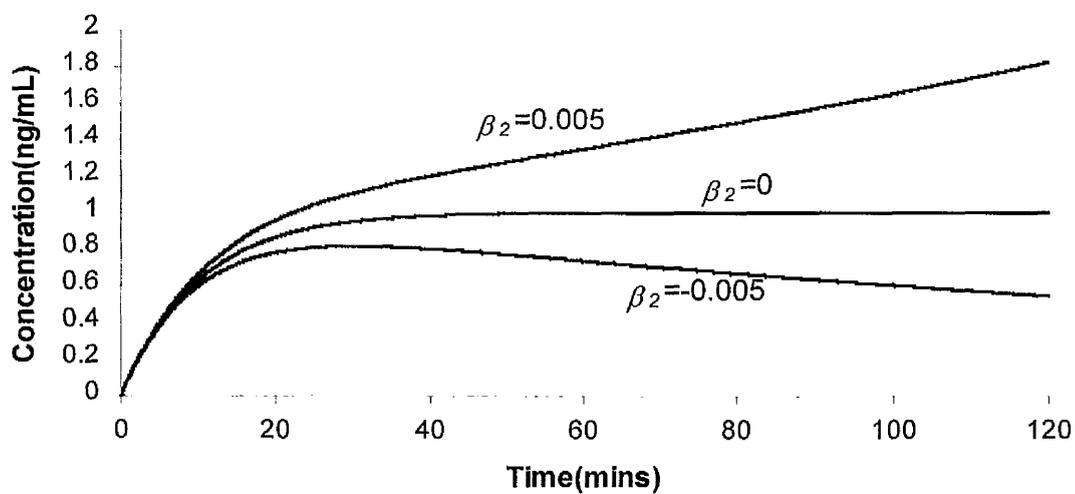
Figure 5.3. Plot of equation (5.1) with default parameter values given by $(\beta_1,\beta_2,\beta_3) =$ (0,-0.005,0.1)where β_3 changes value.
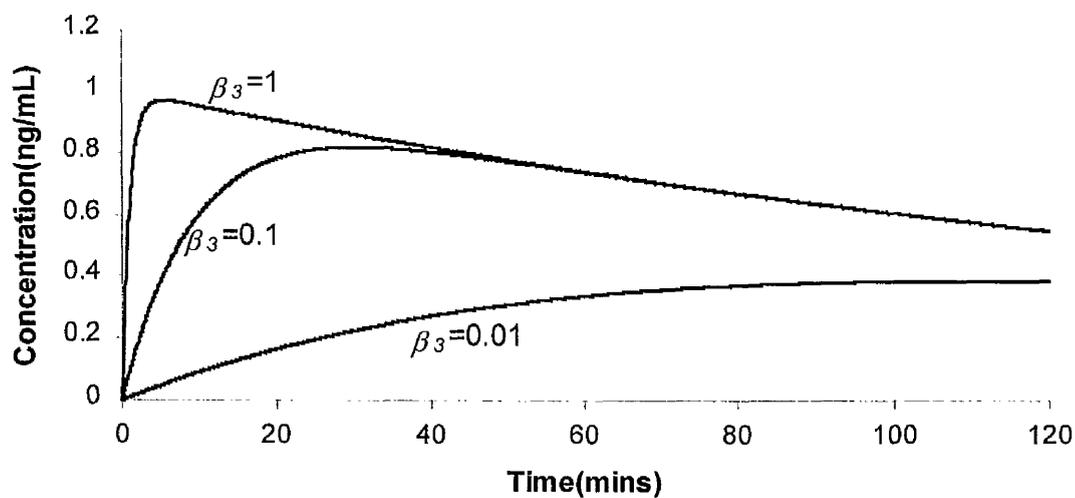
This empirical nonlinear model has three parameters that require estimating. With the parameterisation of the model shown in equation (5.1), all three parameters are nonlinear in that to sample a random variable from the full conditional for Gibbs sampling, a Metropolis step was required because each full conditional distribution is non-log-concave. As this would result in the use of the Griddy-Gibbs routine in BUGS and hence a much slower time to produce results due to more complicated sampling procedures, a log transformation of the model and data was performed. The transformed model is given in equation (5.2).

$$E(\log(C)) = \log(Dose) + \beta_1 + \beta_2 time + \log(1 - e^{-\beta_3 time}) \tag{5.2}$$

This model only requires the Griddy-Gibbs algorithm on one parameter, β_3. The other two parameters, β_1 and β_2, can be sampled directly by Gibbs sampling. The residual error model for these data was assumed to be homoscedastic on the log scale (exponential on the original scale).

The parameters β_1, β_2 and β_3 are not pharmacokinetic parameters such as clearance and volume. This function was chosen because only three parameters per individual needed to be estimated compared to a nonparametric model which might require as many parameters as data points. Most patients' data takes the form of a jump in concentrations from time 0 minutes to 10 or 20 minutes during which time the drug is rapidly absorbed and then the concentration profile flattens off until 120 minutes. Because of this characteristic of the data, it was difficult to estimate β_3 so all individual values of β_3 were set to the estimated population value which was estimated during the model fitting. Additive random effects were assigned to the parameters β_1 and β_2 so individual estimates could be obtained.

As a Bayesian analysis was carried out, it was important to state all distributional assumptions about the likelihood, population distributions and priors. At the first stage of the hierarchical model, the likelihood was assumed to be a log-normal distribution (or equivalently the logarithm is normally distributed) as defined in equation (5.3).

$$\log(Y_{ij}) \sim N(f(Dose_i, time_{ij}, \underline{\beta}_i), \tau^{-1}) \tag{5.3}$$

where $i=1,\ldots,I = 83$ is the total number of individuals in all groups and $j=1,\ldots,n_i$ is the number of observations for each individual not including time = 0 minutes. The structural model is denoted by $f(.)$ and τ^{-1} is the residual variance.

At the second stage of the hierarchical model, the population distribution of the parameters was established. As the parameter β_3 was fixed to be the same for every individual, the prior distribution for this parameter can be left to the third stage. The common assumption to make at this stage is that the parameters come from a normal or log-normal distribution. For the parameters β_1 and β_2 it was not necessary that they be positive so no transformation of the parameters was made and a bivariate normal distribution was assumed. This distribution is given in equation (5.4).

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta_1} & \sigma^2_{\beta_1\beta_2} \\ \sigma^2_{\beta_1\beta_2} & \sigma^2_{\beta_2} \end{bmatrix} \right) \tag{5.4}$$

At the third stage of the hierarchical model, information that was available before the data were collected could be used to form a distribution on the remaining unknown parameters. As no information was available given such a model formulation, low information priors were assigned to the population parameters and interindividual variance-covariance matrix as well as the residual variance. A bivariate normal distribution was assigned to the population parameters, β_1, and β_2 with mean vector set (μ_1,μ_2) to $(0,0)$ and variance covariance matrix C set to $100\times I$ where I is the 2×2 identity

matrix. The parameter β_3 was constrained to be positive by assigning a uniform distribution on the range (0.01,12). A bounded distribution has to be specified when the Griddy-Gibbs algorithm is implemented so a histogram can be constructed. The Wishart distribution was defined for the interindividual variance-covariance matrix where the matrix R is the prior estimate of the interindividual variance covariance matrix and ρ is the dimension of the matrix, 2. A gamma distribution was assigned to the inverse of the residual variance. The full model specification is given in equation (5.5).

$$\log(Y_{ij}) = \log(Dose_i) + \beta_{1i} + \beta_{2i} time_{ij} + \log(1 - e^{-\beta_3 time_{ij}}) + \varepsilon_{ij} = f(Dose_i, time_{ij}, \underline{\beta}_i) + \varepsilon_{ij}$$

$$\log(Y_{ij}) \sim N(f(Dose_i, time_{ij}, \underline{\beta}_i), \tau^{-1})$$

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta_1} & \sigma^2_{\beta_1 \beta_2} \\ \sigma^2_{\beta_1 \beta_2} & \sigma^2_{\beta_2} \end{bmatrix} \right)$$

$$\tau \sim G(0.001,0.001)$$

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \right)$$

$$\beta_3 \sim U(0.001,12)$$

$$\begin{bmatrix} \sigma^2_{\beta_1} & \sigma^2_{\beta_1 \beta_2} \\ \sigma^2_{\beta_1 \beta_2} & \sigma^2_{\beta_2} \end{bmatrix}^{-1} \sim Wish \left(\left(\rho \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \right)^{-1}, \rho \right)$$

$$(5.5)$$

This model was run in BUGS 0.6 and the results are given in table 5.2. The first 5000 iterations were discarded as the burn-in to allow the Markov chain to converge to the stationary (posterior) distribution and then every tenth iteration of the next 20000 iterations were saved. This gave a sample of 2000 iterations from which the parameter probability density functions could be formed.

Table 5.2. *Results of the model specified in equation (5.5).*

	Mean	Median	Standard Deviation	95% Credible Interval
β_1	-0.747	-0.747	0.0959	(-0.941,-0.555)
β_2	-3.321×10^{-4}	-7.679×10^{-4}	0.0115	(-0.0223,0.0217)
β_3	0.536	0.183	1.638	(0.1437,6.671)
$\sigma_{\beta1}^{2}$	0.457	0.443	0.092	(0.31,0.673)
$\sigma_{\beta1\,\beta2}^{2}$	-0.00351	-0.00345	0.00792	(-0.02,0.0122)
$\sigma_{\beta2}^{2}$	0.00824	0.00807	0.00152	(0.000583,0.0116)
σ	0.226	0.226	0.0104	(0.207,0.247)

The value of the covariance between β_1 and β_2 has a 95% credible interval of (-0.02, 0.0122) which suggests that the covariance between the two parameters was negligible and probably did not add anything to the model. Although the inclusion of the covariance term in this model has no adverse effect on parameter estimation, the model was re-run with the covariance term set to zero. To achieve this, instead of specifying a bivariate normal distribution on the individual parameters, β_{1i} and β_{2i}, two independent univariate normal distributions were specified with the same mean and variance as in equation (5.5). The 95% credible interval for β_2 includes zero but this does not mean that β_2 should be removed from the model as this parameter was allowed to range over zero. The 95% credible interval for β_1 did not include zero and was negative implying a normalising constant of between (0,1). The new model without the covariance term was run and the results are given in table 5.3.

Table 5.3. *Results of model where covariance term was fixed to zero.*

	Mean	Median	Standard Deviation	95% Credible Interval
β_1	-0.777	-0.778	0.0893	(-0.95,-0.601)
β_2	7.812×10^{-4}	8.156×10^{-4}	0.00126	(-0.00183,0.00316)
β_3	1.27	0.202	2.73	(0.151,10.3)
$\sigma_{\beta1}^{2}$	0.431	0.422	0.0806	(0.296,0.613)
$\sigma_{\beta2}^{2}$	7.384×10^{-5}	7.156×10^{-5}	1.515×10^{-5}	(5.147×10^{-5},1.094×10^{-4})
σ	0.227	0.227	0.0105	(0.208,0.248)

It can be seen that there is very little difference in the parameter estimates of β_1 and β_2 between table 5.2 and 5.3. β_3 is quite different but the Standard Deviation in both tables is large showing that this parameter is poorly determined. A plot of the model using the parameter estimates from table 5.2 is given figure in 5.4. The data correspond to the 20 mg dose group.

Figure 5.4. *Population predicted and data for 20 mg dose group.*
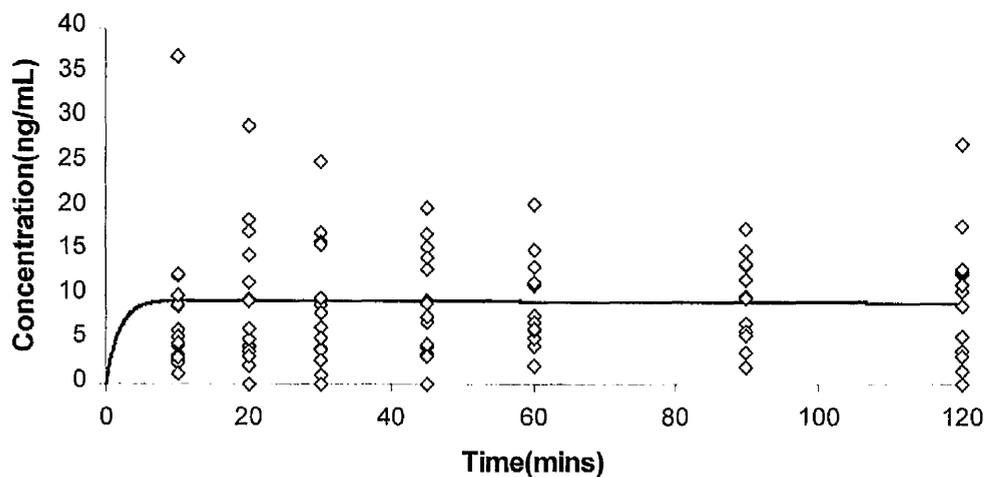
As can be seen from figure 5.4, the population model is not particularly useful for describing the data.

Figure 5.5. *Individual plot from 20 mg dose group.*
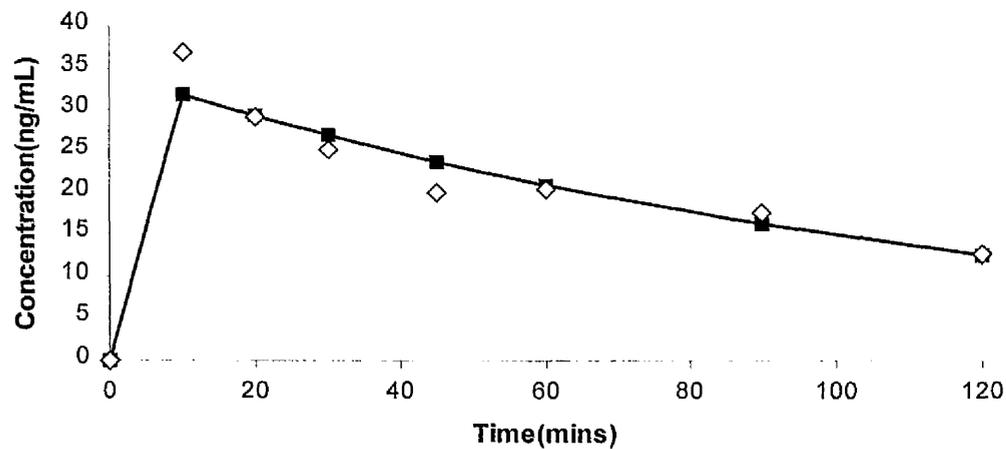
Figure 5.6. *Individual plot from 20 mg dose group.*
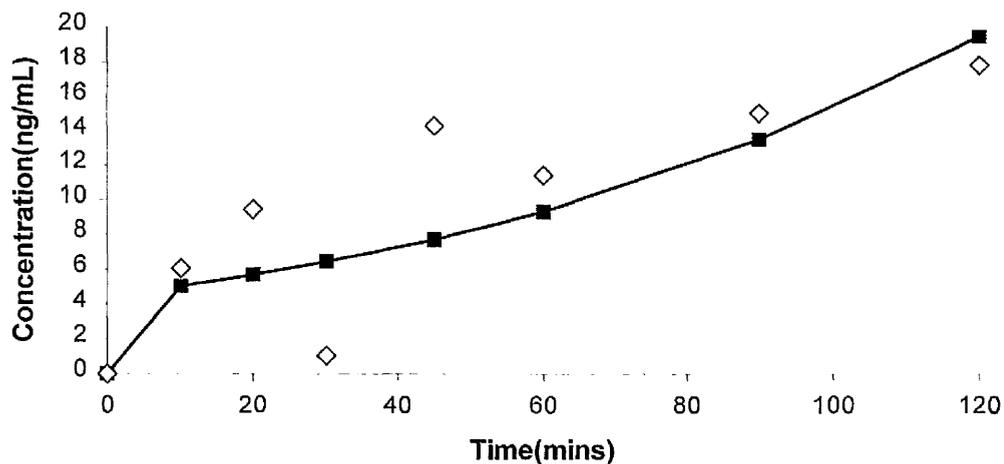
A plot of two individuals from the 20 mg dose group are given in figures 5.5 and 5.6. Figure 5.5 shows a good fit of the model but figure 5.6 shows the amount of variability in the data. As can be seen, the empirical model chosen is flexible enough to model the very different profiles. The two peaks in the observed data can be seen easily in figure 5.6 and in figure 5.5 to a lesser extent. Residual plots are given in figures 5.7-5.9.

Figure 5.7. *Residual versus time plot for population model defined in equation (5.5).*
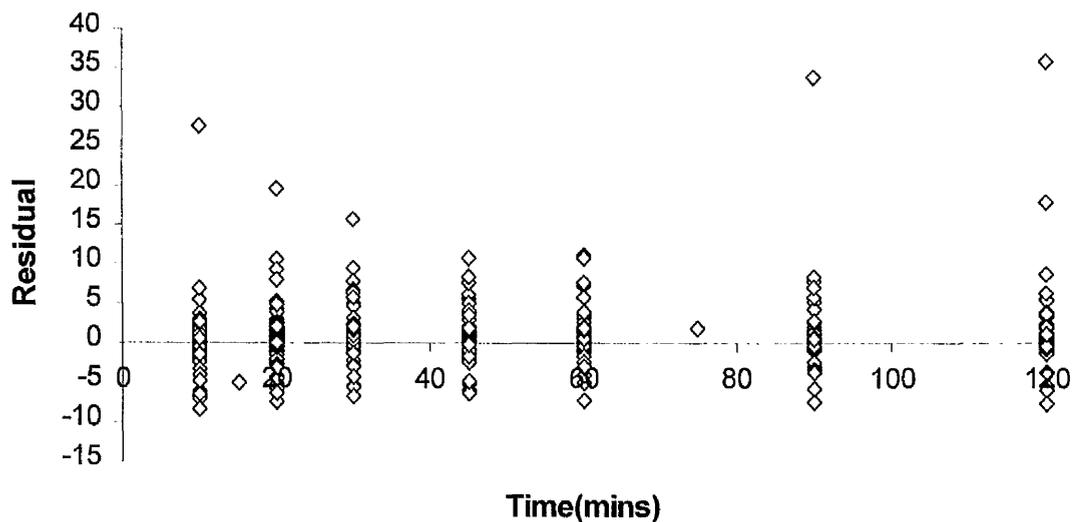
Figure 5.8. *Residual versus population predicted plot for model defined in equation*
(5.5).
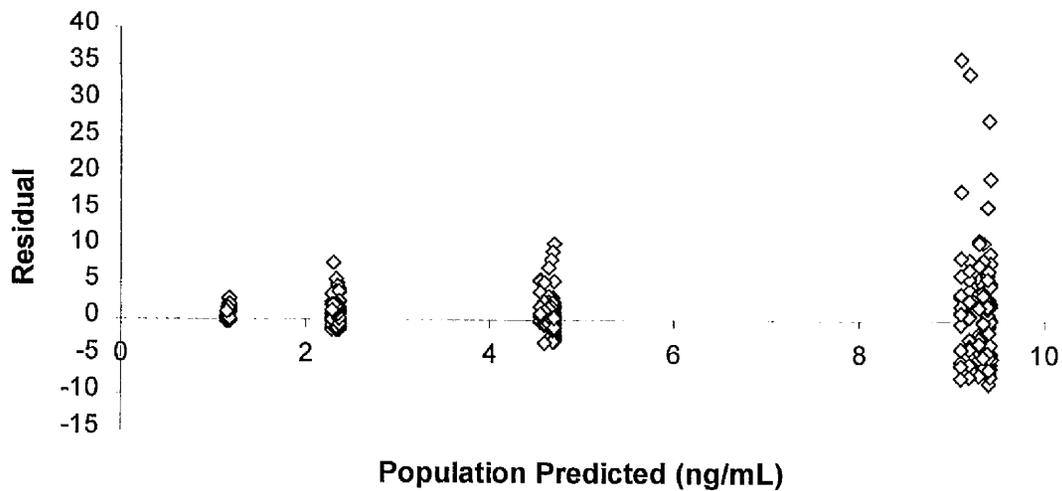
Figure 5.9. *Weighted residual versus population predicted plot for model defined in*
equation (5.5).
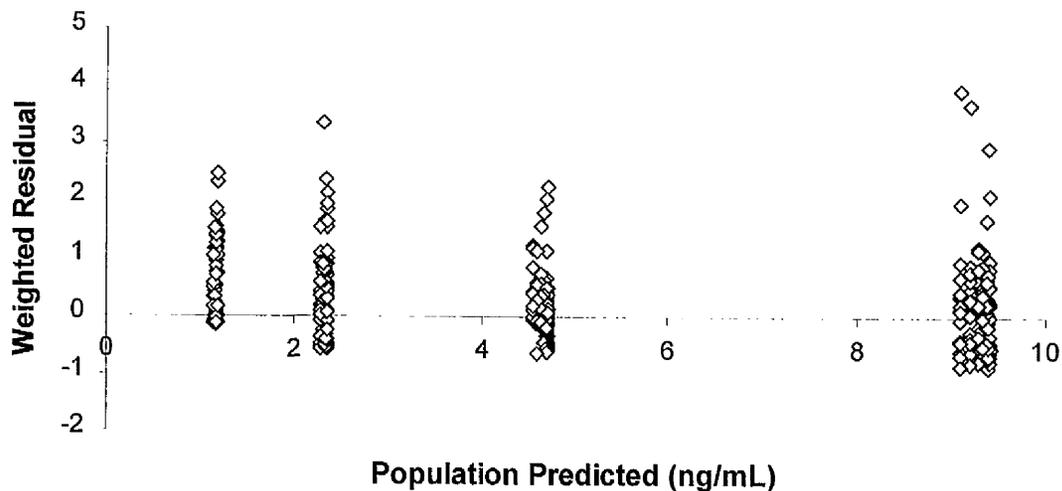
The weighted residual is defined in equation (5.6).

$$wr_{ij} = \frac{\sqrt{\tau}(Y_{ij} - E(Y_{ij}))}{\sqrt{E(Y_{ij})}}$$

(5.6)

The residual versus time plot in figure 5.7 shows a good fit of the model to the data. The residual versus population predicted shows a reasonable amount of evenness but the weighted residual versus population predicted appears to show that possibly the wrong structural model was used.

5.4 Pharmacodynamic Analysis

The variable measured for the pharmacodynamics of intranasal sumatriptan was migraine severity. This was measured on a categorical scale as described in section 5.2. The number of scores in each category for a particular time and dose are given in tables 5.4a-e.

Table 5.4a. *Number of scores in each category for 0 mg dose group.*

Category	Time(minutes)						Total
	0	15	30	60	90	120	
0	0	0	0	1	1	1	3
1	0	1	2	2	1	1	7
2	6	5	4	3	4	4	26
3	5	5	5	5	5	5	25
Total	11	11	11	11	11	11	66

Table 5.4b. *Number of scores in each category for 2.5 mg dose group.*

Category	Time(minutes)						Total
	0	15	30	60	90	120	
0	0	0	1	2	3	3	9
1	0	1	1	1	0	3	6
2	6	8	7	7	6	3	37
3	9	6	6	5	5	3	36
Total	15	15	15	15	14	14	88

Table 5.4c. *Number of scores in each category for 5 mg dose group.*

Category	Time(minutes)						Total
	0	15	30	60	90	120	
0	0	2	3	3	3	3	14
1	0	2	1	3	4	7	17
2	10	8	10	7	5	2	42
3	9	7	5	5	5	5	36
Total	19	19	19	18	17	17	109

Table 5.4d. *Number of scores in each category for 10 mg dose group.*

Category	Time(minutes)						Total
	0	15	30	60	90	120	
0	0	0	1	2	3	4	10
1	0	2	4	5	4	2	17
2	11	11	9	6	5	5	47
3	7	5	4	4	5	6	31
Total	18	18	18	17	17	17	105

Table 5.4e. *Number of scores in each category for 20 mg dose group.*

Category	Time(minutes)						Total
	0	15	30	60	90	120	
0	0	1	4	7	8	9	39
1	0	6	5	5	6	3	25
2	12	5	6	5	4	4	36
3	8	8	5	3	2	2	28
Total	20	20	20	20	20	18	118

There were 486 observations out of a possible 498 observations and so no attempt was made to account for missing data. As can be seen in tables 5.4a-e, the number of scores observed in the no and mild pain categories (0 and 1) increases as dose and time increases. To see if there was a placebo effect (time effect for placebo dose), the placebo data was modelled alone. A look at the data for the placebo group seems to suggest that there is no time trend. Firstly, a constant baseline proportional odds model was fitted to the placebo data which is defined in equation (5.7) with likelihood specification and priors. This model was parameterised so that the cumulative categories were grouped in terms of the highest value categories.

$$\text{logit}(\Pr(Y_{ij} \geq k \mid b_i)) = \text{logit}(\gamma_{kij}) = \theta_k + b_i$$

$$L_{ij} = \left(\left(\frac{r_{1ij}}{r_{0ij}} \right)^{r_{1ij}} \left(\frac{r_{0ij} - r_{1ij}}{r_{0ij}} \right)^{r_{0ij} - r_{1ij}} \right) \left(\left(\frac{r_{2ij}}{r_{1ij}} \right)^{r_{2ij}} \left(\frac{r_{1ij} - r_{2ij}}{r_{1ij}} \right)^{r_{1ij} - r_{2ij}} \right) \left(\left(\frac{r_{3ij}}{r_{2ij}} \right)^{r_{3ij}} \left(\frac{r_{2ij} - r_{3ij}}{r_{2ij}} \right)^{r_{2ij} - r_{3ij}} \right)$$

$$\gamma_m = \pi_m + \ldots + \pi_3, m = 0,1,2,3$$
$$r_m = y_m + \ldots + y_3, m = 0,1,2,3$$

$$\theta_1 \sim N(5,1 \times 10^4), \theta_2 \sim N(3,1 \times 10^4), \theta_3 \sim N(1,1 \times 10^4)$$

$$b_i \sim N(0, \tau^{-1})$$
$$\tau \sim G(0.001, 0.001)$$

$$(5.7)$$

Table 5.5. *Results of constant baseline proportional odds model.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	12.84	12.65	4.396	(5.865,22.88)
θ_2	10.2	10.04	4.323	(3.436,20.23)
θ_3	-2.968	-2.968	4.063	(-12.05,4.563)
σ	13.21	12.22	6.041	(5.24,28.27)

The results of this model are given in table 5.5. The cut points correspond to individual category probabilities of $\Pr(Y=0) = 0$, $\Pr(Y=1) = 0$, $\Pr(Y=2) = 0.949$ and $\Pr(Y=3) = 0.051$. This is not a particularly good representation of the observed data. A new model was run with a time effect as defined in equation (5.8).

$$\text{logit}(\Pr(Y_{ij} \geq k \mid b_i)) = \log it(\gamma_{kij}) = \theta_k + \beta time_{ij} + b_i$$
$$\beta \sim N(0,10)$$

$$(5.8)$$

A relatively informative prior (compared to previous priors on gradient parameters) was assigned to the gradient term as this parameter was not expected to be estimated well due to only 11 patients being in the placebo group. It was expected *a priori* that there would be no time trend over 2 hours. The results are given in table 5.6. Table 5.6 shows that the 95% credible interval for the time effect parameter includes zero. Based on this, it was assumed that there was no placebo effect in the pharmacodynamic data for intranasal sumatriptan.

Table 5.6. *Results of baseline proportional odds model with time effect.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	14.84	14.55	4.073	(7.913,23.93)
θ_2	11.59	11.3	3.831	(5.311,20.05)
θ_3	-3.012	-2.493	4.457	(-13.3,4.332)
β	-0.0246	-0.02363	0.0136	$(-0.0548,8.595\times10^{-4})$
σ	15.39	13.4	8.862	(5.813,36.4)

The placebo data was also analysed in terms of a Poisson regression model where the counts in table 5.4a became the response variable rather than the probability of being in a particular set of cumulative probabilities. The model is defined in equation (5.9) and uses the log-linear transformation on the counts. This model was estimated using Splus version 4.5 as only fixed effects were to be estimated. This program uses a weighted least squares method to estimate the parameters. The results are given in table 5.7.

$$\log(E(count)) = \beta_0 + \beta_1 time + \beta_2 cat2 + \beta_3 cat1 + \beta_4 cat0 \qquad (5.9)$$

The covariates *cat2*, *cat1* and *cat0* are binary variables defining whether the count was in category 3, 2, 1 or 0.

Table 5.7. *Results of Poisson model.*

	Mean	Standard Deviation	t value
β_0	1.609	0.447	3.599
β_1	2×10^{-4}	0.0031	0.0596
β_2	-0.0861	0.502	-0.172
β_3	-1.379	0.401	-3.440
β_4	-2.225	0.582	-3.822

It can be seen from table 5.7 that the time parameter was not significantly different from zero based on the t value. This confirmed the results of the BUGS analysis for the placebo model and no time trend need be taken into account.

Once the placebo model was taken into account, then the whole data set was analysed. The first model to be examined was a simple proportional odds model with linear dose as a predictor. This dose-effect model is defined in equation (5.10) where the likelihood is as in equation (5.7). The term ε_{ij} acts like a residual term and τ_{ε}^{-1} is the estimate of the residual variance on the logit transformed scale. The results of this model are given in table 5.8 where the parameters are based on an initial burn-in of 2000 iterations then a sample of 5000 iterations.

$$\text{logit}(\Pr(Y_{ij} \geq k \mid b_i)) = \theta_k + \beta dose_i + b_i + \varepsilon_{ij}$$
$$\theta_1 \sim N(5,1000), \theta_2 \sim N(3,1000), \theta_3 \sim N(1,1000)$$
$$\beta \sim N(-0.1,1000) \tag{5.10}$$
$$b_i \sim N(0,\tau_b^{-1}), \tau_b \sim G(0.001,0.001)$$
$$\varepsilon_{ij} \sim N(0,\tau_{\varepsilon}^{-1}), \tau_{\varepsilon} \sim G(0.001,0.001)$$

The dose effect is significant in terms of dose with the coefficient being negative implying for example that as dose increases, the probability of being in the category 3 (severe pain) decreases. A plot of the mode is give in figure 5.10.

Table 5.8. *Results of proportional odds model with linear dose.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	4.679	4.647	0.587	(3.681,5.829)
θ_2	2.972	2.953	0.508	(2.023,3.992)
θ_3	-0.323	-0.306	0.452	(-1.313,0.532)
β	-0.119	-0.119	0.0349	(-0.190,-0.0472)
σ_b	2.746	2.705	0.374	(2.148,3.617)
σ_{ε}	0.475	0.397	0.382	(0.0523,1.571)

Figure 5.10. *Plot of linear dose proportional odds model with observed data averaged over time.*
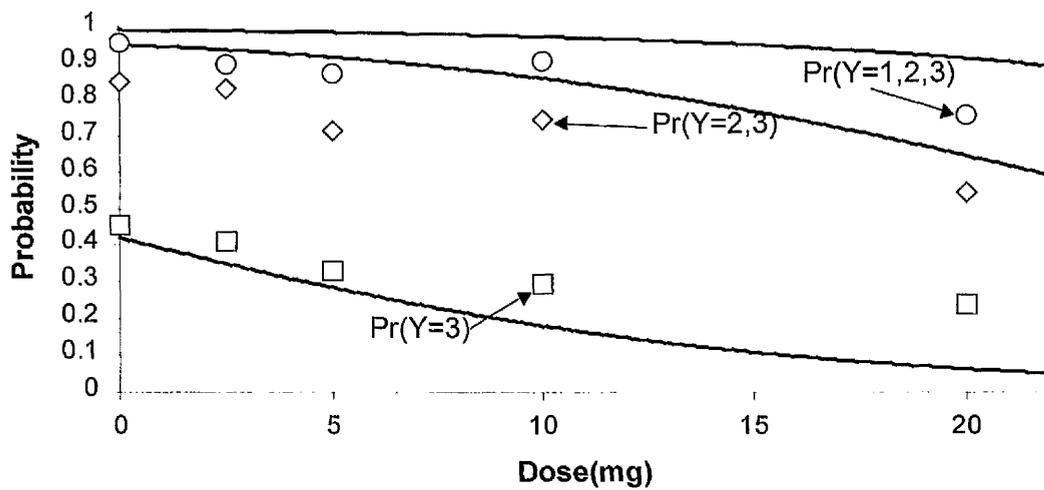
The plot of the model in figure 5.10 shows observed probabilities averaged over time. This might not be appropriate, as there might be a time effect in the active treatment groups. The patients were administered a single dose of intranasal sumatriptan once the migraine had begun and hence for there to be a difference in the number of patients with lower scores in the active treatment groups at later times as compared to earlier times would require a model with time. To check this, the dose effect model can be expanded to test for a linear trend in time. More complicated models in time could have been tested but these time trends can be included in the measurements of the concentrations or truncated AUC values. If there were any time effect in the data then a linear model would hopefully indicate such a correlation. The model is given in equation (5.11). No residual term, ε_{ij}, was included in the model as initial attempts at estimating this parameter were unsuccessful so this parameter was dropped from the model temporarily.

$$\text{logit}(\Pr(Y_{ij} \geq k \mid b_i)) = \theta_k + \beta_1 dose_i + \beta_2 time_{ij} + b_i$$
$$\theta_1 \sim N(4,1000), \theta_2 \sim N(2,1000), \theta_3 \sim N(0,1000)$$
$$\beta_1 \sim N(0,1000), \beta_2 \sim N(0,1000)$$
$$b_i \sim N(0,\tau_b^{-1}), \tau_b \sim G(0.001,0.001)$$

(5.11)

The results from this BUGS run are given in table 5.9.

Table 5.9. *Results of dose+time proportional odds model.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	6.683	6.681	0.723	(5.28,8.182)
θ_2	4.661	4.655	0.673	(3.337,6.027)
θ_3	0.889	0.883	0.616	(-0.291,2.151)
β_1	-0.126	-0.126	0.0515	(-0.229,-0.0223)
β_2	-0.0259	-0.0258	0.0029	(-0.0318,-0.0202)
σ_b	3.208	3.186	0.379	(2.534,4.004)

The 95% credible interval for both the dose and time parameters do not include zero so there was a time effect in the data. This model implies that when the dose is zero (placebo), there is a change in probability over time given by the time parameter. This contradicts the placebo model but was required for the active treatment groups. As the time parameter estimate was negative, this infers that as time increases, the probability of being in a higher category decreases.

As both the dose and time parameter estimates were important to explaining the data, an interaction term was included in a new model but unfortunately, this model was unable to be estimated in BUGS. The problem was the Markov chains corresponding to the conditional distributions would not settle down to a stationary distribution.

A simple generalised linear model in terms of dose and time was significant in terms of the parameter estimates being different from zero. The linear assumption was an

212

obvious and simple first step to make in fitting such a proportional odds model. Linear time and dose is just one function of time and dose to be used to predict the pharmacodynamics. Another function of time and dose (as well as other pharmacokinetic parameters) to predict the pharmacodynamics is the concentration time profile. As stated in chapter 3, it is generally believed that using the plasma concentration data should be better at correlating with the pharmacodynamics of a drug than dose alone (although this is not exclusively the case).

There are different ways in which the concentrations can be modelled to the migraine pain severity scores. One way is to determine the hypothetical effect site concentrations from the plasma concentration and pharmacodynamic data and use these measurements as regressors for the pain severity scores. The concentrations at the effect site are believed to correlate better with the pharmacodynamic outcome than say plasma concentration although this is not always the case. Using the observed plasma concentrations as a predictor for the effect measurements is another way of using the pharmacokinetics to predict the pharmacodynamics. This can be problematic if the times at which the concentrations and scores are taken are different and interpolation is required to get values for the concentrations at the same times as the score measurements. Also, assay error and other error components may not be taken into consideration which will cause a possible underestimate of the variability in the pharmacodynamic model. Instead of using observed concentrations, predicted concentrations could be used. This method allows interindividual variability in the pharmacokinetics to be included in the analysis but residual variability is not accounted for. Again, this would cause an underestimate of the variability, however it would circumvent the problem of different measurement times. A simultaneous method of

analysing both the pharmacokinetics and pharmacodynamics is another, more general way of analysing such data. This allows the possibility of finding correlations between pharmacokinetic and pharmacodynamic parameters that would not be possible with sequential methods discussed above. This method allows all levels of variability associated with the pharmacokinetics to be included in the analysis of the pharmacodynamics. This supposedly gives better estimates of variability in the pharmacodynamics conditional on the variability in the pharmacokinetics. This method is usually difficult to implement but with MCMC methods, such issues as analytical intractability should be overcome. An example of a simultaneous analysis for simple linear regression models is given at the end of this chapter as an appendix. A simultaneous analysis within BUGS 0.6 was attempted but failed to run. The model compiled but iterations were unable to be produced. The reason for this was not resolved.

It was decided initially to use the predicted concentrations from the BUGS run of the pharmacokinetic model. The model used was a proportional odds model with linear concentration as a predictor. This model could also be considered as a nonlinear model in terms of dose, time and predicted pharmacokinetic parameters but this is not the case as the pharmacodynamic parameters were linear in the model. The model is defined in equation (5.12). The results of this model are given in table 5.10.

$$
\begin{aligned}
\log it(\Pr(Y_{ij} \geq k \mid b_i)) &= \theta_k + \beta conc_{ij} + b_i + \varepsilon_{ij} \\
\theta_1 &\sim N(4,1000), \theta_2 \sim N(2,1000), \theta_3 \sim N(-1,1000) \\
\beta &\sim N(-0.005,1000) \\
b_i &\sim N(0,\tau_b^{-1}), \tau_b \sim G(0.001,0.001) \\
\varepsilon_{ij} &\sim N(0,\tau_\varepsilon^{-1}), \tau_\varepsilon \sim G(0.001,0.001)
\end{aligned}
\tag{5.12}
$$

The results in table 5.10 are based on an initial burn-in of 2000 iterations and a sample of 8000 iterations.

Table 5.10. *Results of proportional odds model with linear plasma concentration predictor.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	5.165	5.161	0.469	(4.278,6.102)
θ_2	3.221	3.207	0.402	(2.459,4.039)
θ_3	-0.284	-0.283	0.364	(-0.990,0.434)
β	-0.327	-0.326	0.0443	(-0.418,-0.243)
σ_b	2.782	2.768	0.316	(2.222,3.465)
σ_ε	0.232	0.196	0.144	(0.0492,0.605)

Figure 5.11. *Plot of proportional odds model versus predicted concentration.*
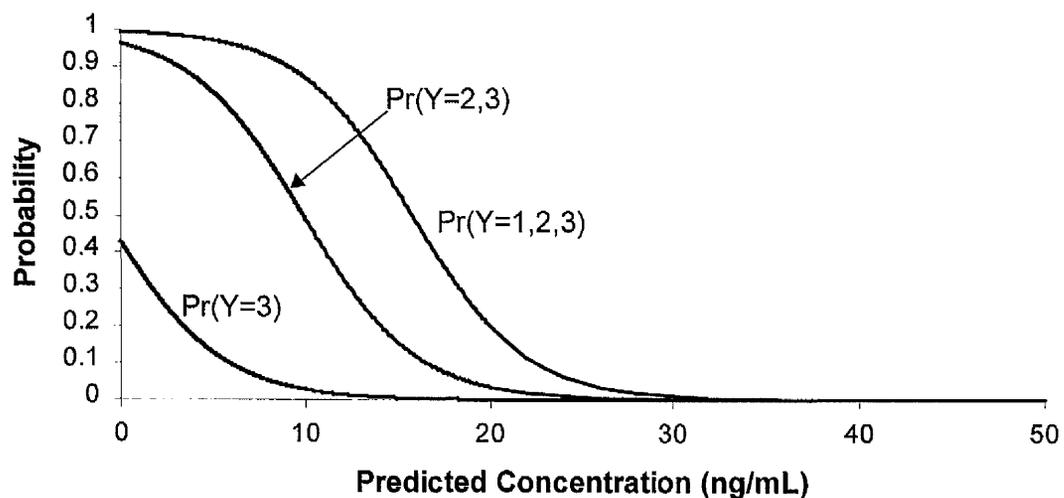
Figure 5.12. *Plot of individual category probabilities versus predicted concentration.*
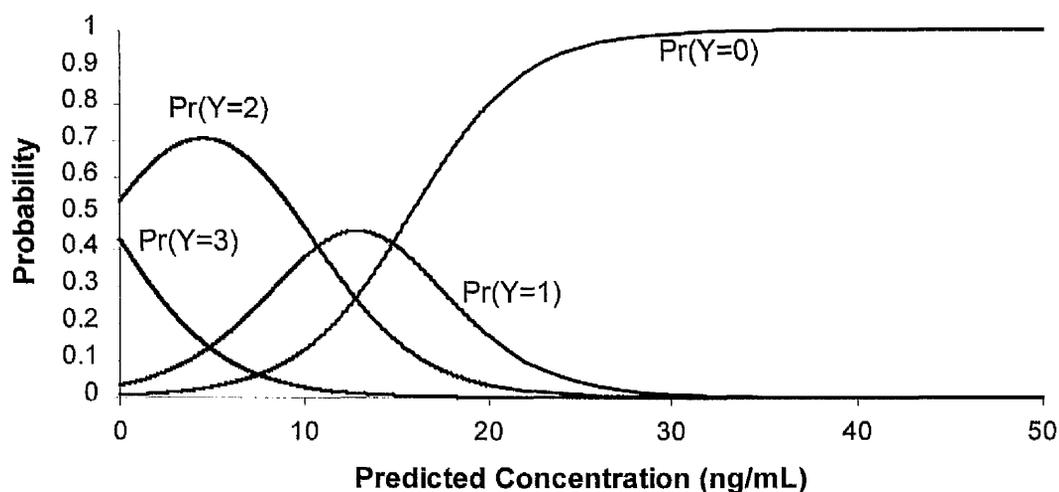
One of the problems of modelling the effect as a linear function of concentration is that the model implies that the effect at the present time only depends on the current concentration. This could be misleading if, for example, the current concentration was high and the previous concentration was low, there could be a time lag before the effect of the drug sets in which a model with an independence assumption from time to time could not account for. A model that includes an autoregressive type correlation could be introduced into the model as a way of getting round the possible time lag. This type of model is difficult to implement when the time spacing between observations are unequal. Another possible solution is to look at the exposure to the drug by using an individual truncated AUC value as a predictor for the pharmacodynamics. The truncated AUC was the integral of the concentration-time profile up to time t as given in equation (5.13).

$$AUC_i(t) = Dose_i \times e^{\beta_1} \left[\frac{1}{\beta_{2i}} (e^{\beta_{2i}t_i} - 1) - \frac{1}{(\beta_{2i} - \beta_3)} (e^{(\beta_{2i} - \beta_3)t_i} - 1) \right] \qquad (5.13)$$

The model involving AUC is given in equation (5.14) and the results are in table 5.11.

$$\mathrm{log}\,it(E(\mathrm{Pr}(Y_{ij} \geq k \mid b_i))) = \theta_k + \beta AUC_{ij} + b_i$$

$$\theta_1 \sim N(4,1000), \theta_2 \sim N(2,1000), \theta_3 \sim N(0,1000)$$

$$\beta \sim N(0,1000) \tag{5.14}$$

$$b_i \sim N(0, \tau_b^{-1}), \tau_b \sim G(0.001, 0.001)$$

Table 5.11. *Results of proportional odds model with linear AUC as predictor.*

	Mean	Median	Standard Deviation	95% Credible Interval
θ_1	5.182	5.167	0.494	(4.269,6.179)
θ_2	3.079	3.068	0.424	(2.288,3.929)
θ_3	-0.602	-0.604	0.388	(-1.361,0.178)
β	-0.0048	-0.0048	5.636×10^{-4}	(-0.0059,-0.0038)
σ_b	3.031	3.007	0.356	(2.409,3.789)

The parameter estimate corresponding to the AUC term was significant and negative.

5.5 Model Checking

It was important to compare the models derived for the migraine pain severity scores so it could be seen which model fitted the data best in the set of models considered. Before the models were checked, it was considered that the model in AUC would fit the data best, whereas the model with just dose would fit the worst. This was due to the preconceived idea that the pharmacokinetics would correlate better with the pharmacodynamics than dose alone.

The method used in this chapter to determine goodness-of-fit was to define an objective function and use the value obtained for a particular model as a measure of how well the model fit the data. The objective function used is defined in equation (5.15) and is a possible generalisation of the deviance statistic for dichotomous data. The categories

{0,1,2,3} have been renamed {1,2,3,4} for convenience. The objective function defined in equation (5.15) is in terms of individual category probabilities rather than cumulative category probabilities.

$$llike_{ij} = \sum_{k=1}^{4} \delta\left(\frac{Y_{ij}}{k}\right)\left(\frac{Y_{ij}}{k}\log(\pi_{ij}) + \left(1 - \frac{Y_{ij}}{k}\right)\log(1 - \pi_{ij})\right)$$

$$llike.sat_{ij} = \sum_{k=1}^{4} \delta\left(\frac{Y_{ij}}{k}\right)\left(\frac{Y_{ij}}{k}\log(Y_{ij}) + \left(1 - \frac{Y_{ij}}{k}\right)\log(1 - Y_{ij})\right)$$

$$\delta(B) = \begin{cases} 1, B = 1 \\ 0, otherwise \end{cases}, B \in \Re$$

$$D = 2\left(\sum_{i,j} llike.sat_{ij} - \sum_{i,j} llike_{ij}\right)$$

(5.15)

Apart from the dose and dose+time models, these models are non-nested. The standard distributional assumptions about deviance residuals could not be used here to compare the models partly due to the non-nestedness of the models but also since these models included random effects and such approximations (χ^2 distribution) are known not to be appropriate. As such, comparisons between models were made on an informal basis where the lower the objective function value, the better the model was at describing the data. The results are given in table 5.12.

Table 5.12. *Objective function values for the proportional odds models.*

Model	Objective Function Value
Dose	843
Dose+Time	733
Predicted Concentration	779
Predicted AUC	727

The model with the lowest objective function value was the predicted AUC model with an objective function value of 727 and the worst was the dose model with a value of 843 as expected. For the model with linear time and dose, the objective function value was 733, only a difference of 6 from the AUC model. This does not seem to be a big

difference when the values of the objective function values are of the order of magnitude of hundreds. Based on this, it was decided that the dose+time model was best as it did not require complicated calculations of the truncated AUC values.

5.6 Discussion

The purpose of this work was to firstly carry out a general population pharmacokinetic/pharmacodynamic analysis of intranasal sumatriptan. The population pharmacokinetic analysis was not in the usual form of a compartmental model (other published analyses for sumatriptan have used a two compartment disposition model). As the data were only up to two hours, very little information if any was available on the disposition of intranasal sumatriptan. For this reason, an empirical parametric model was used to model the pharmacokinetic data. The obvious problem of doing this sort of analysis was that no physiologically interpretable pharmacokinetic parameters were estimated. It would not have been straight forward to fit such a model to the data available, as many of the parameters would have to have been fixed for the disposition stage of the profile. The empirical model used was adequate for describing the population and the individual.

The emphasis of this chapter was on the analysis of repeated measurement pharmacokinetic/pharmacodynamic data when the effect was measured on a categorical scale. The pharmacodynamics in this study was not easy to analyse, as is often the case with modelling categorical data. Techniques for analysing this sort of population pharmacodynamic data have been taken from the statistical literature as well as more

219

recent publications where such techniques have been applied to categorical pharmacodynamic data, for example Sheiner (1994), Sheiner *et al* (1997) and Mandema and Stanski (1996). These methods (for categorical data) are not as well developed as those for nonlinear mixed effects models as for population pharmacokinetic data. The difficulty in obtaining individual estimates (by including random effects components) from population categorical data is due to the nature of categorical data itself. Categorical data contains little information when compared to continuous data and extracting adequate models of the individual and population can be difficult. Standard statistical packages do allow the estimation of proportional odds and logistic regression models but usually only in a fixed effects setting. SAS can be used to fit generalised linear mixed effects models by using PROC MIXED within a macro but this is only one of the few commercially available programs that has the capability. Currently, Splus 4.5 has a nonlinear mixed effects routine but no random effects for generalised linear models. NONMEM is capable of mixed effects modelling for categorical data and is quite flexible as it allows the user to specify their own likelihood. The models considered for intranasal sumatriptan only included one subject specific random effect. This was assigned to the cut points additively so it acted like a shift on the model on the logit scale. Random effects were often not possible on the gradient parameters such as dose and time. This type of random effect is possible in NONMEM but it is not known how good the estimate of the variability is. The residual term used in a couple of the models (dose and predicted concentration) acts like a residual term in linear mixed effects models as it enters the model on the linear logit scale rather than on the probability scale. NONMEM is not capable of estimating such a term as it uses the residual term to indicate a switch between normally distributed data and user defined distributions. Comparing the models with only a random component on the cut points,

(dose and predicted AUC) the standard deviation of the between subject variability for the two models were 3.208 and 3.031 implying that the AUC model could explain more of the variability than dose alone when no residual component was included. Comparing the models with a residual component as well (dose+time and predicted concentration), the between subject variability estimates were 2.746 (0.475) and 2.782 (0.232) where the residual standard deviation is given in parentheses. The between subject variability was greater for the predicted concentration model but the residual variability was approximately half that for the dose+time model. This implies that both models estimate the between subject variability to approximately the same level but dose+time does not explain as much of the residual variability.

The models that were used to model the pharmacodynamics of intranasal sumatriptan were different from the usual model that is commonly associated with pharmacodynamics, as they do not use any variant of the E_{max} model. Instead, they are generalised linear regression models with predictors being combinations of dose, time, predicted concentration or predicted AUC. The models used were linear on the logit scale. An E_{max} model, which takes the modelling into generalised nonlinear mixed effects modelling, is very complicated and possibly only NONMEM is capable of fitting such models although how well is unknown. The models used in the above analysis are adequate in fitting the data and as previously stated, AUC was the best model for fitting the data according to the objective function defined in equation (5.14). Not far behind, in terms of the objective function was the model in terms of linear time and dose. If the model fit was not significantly worse than the AUC model then this implies that including the pharmacokinetic information in the model to predict the pharmacodynamics is not necessarily any better than simple functions of time and dose.

221

Had an interaction term in time and dose been able to be estimated, such a model might have been superior to the AUC model. This could be due to the inherent lack of information in categorical data or due to greater variability in the pharmacodynamics which the pharmacokinetic data could not explain. It is often the case that there is greater variability in the pharmacodynamics between subjects than in the pharmacokinetic data.

The computer program BUGS 0.6 was used instead of NONMEM to analyse the intranasal sumatriptan data set. Apart from the main difference between the two programs of BUGS being a Bayesian package and NONMEM a frequentist package, there were several other differences of importance. The ability to estimate nonlinear models was restricted in this version of BUGS as the Griddy-Gibbs sampling algorithm, although useful in the estimation of nonlinear statistical problems, was not always up to the task of estimating n parameters in complex nonlinear models. Such an empirical parametric model as defined in equation (5.1) would have been estimated without too much difficulty in NONMEM. The Griddy-Gibbs sampler also inhibited the ability to estimate more complicated models in BUGS, although whether these would have been able to be estimated in either BUGS or NONMEM would be debatable. The ability to specify more complicated random effects structures (such as exponential errors) is possible in NONMEM but whether the estimates of these parameters are any good is questionable. More complicated error structures in BUGS usually resulted in the model crashing which was typically an indication of the model being more complicated than the data allowed and was not worth considering further. The random effects distribution can be changed easily in BUGS from a library of distributions but in NONMEM, only normal or log-normal distributions are specified. The ease with which

222

user defined objective functions can be created is another useful feature in BUGS that is not possible in NONMEM when the objective function given is not the one wanted for analysis.

Appendix

This appendix is an example of how a simultaneous analysis could be carried out using Gibbs sampling and BUGS. Code was written in BUGS for a simultaneous analysis of the intranasal sumatriptan data, which compiled but failed to run. As a simple example, two linear models are to be estimated using Bayesian methods. The two linear models can be considered to be the pharmacokinetic and pharmacodynamic models. With Gibbs sampling, the ideas given here can be generalised to any models for the pharmacokinetic and pharmacodynamic data, although the sampling algorithms become more complicated as the models increase in complexity.

Consider the situation where we have data that is of the form given in table A5.1:

Table A5.1. *Data set to be modelled.*

X	Y	Z
x_1	y_1	z_1
x_2	y_2	z_2
...	...	...
x_n	y_n	z_n

It is required to fit two linear regression models, one for y on x and another for z on y so the models are as in equation (A5.1).

$$y_i = \alpha + \beta x_i + \varepsilon_i$$
$$z_i = \gamma + \delta y_i + \eta_i \qquad\qquad (A5.1)$$
$$i = 1,...,n$$

If it were required to fit the first of the above models using Bayesian methods, distributional assumptions would be made for the likelihood and the prior distributions on the parameters. For linear models, these might be of the form in equation (A5.2).

$$Y_i \sim N(\alpha + \beta x_i, \sigma_\varepsilon^2)$$
$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$$
$$\beta \sim N(\mu_\beta, \sigma_\beta^2) \qquad\qquad (A5.2)$$
$$\frac{1}{\sigma_\varepsilon^2} \sim G(c, d)$$

As is the usual case for linear regression models, a normal likelihood is assigned to the data and an additive error model is used. Rather than being assumed known and fixed, the parameters are assumed to be random variables and are given prior distributions. Conjugate priors are used for the parameters α and β, and the residual term is given an inverse gamma distribution which is also conjugate to the normal likelihood. The reason why a linear model is being used in this appendix is because it is tractable and easy to conceptualise. More complicated models could have been used but this would have detracted from the main idea of the appendix.

Suppose it was required to carry out Gibbs sampling to estimate the parameters of the linear model as in equation (A5.1). This is not necessary as the posterior can be determined exactly but it is an easy example to work with. The full conditional distributions need to be defined to sample from to estimate the parameters. These posterior distributions are given in equation (A5.3).

$$p(\alpha \mid \beta, \tau_\varepsilon, y) \sim N\left(\frac{\tau_\varepsilon \sum_i y_i - \beta \sum_i x_i + \tau_\alpha \mu_\alpha}{n\tau_\varepsilon + \tau_\alpha}, \frac{1}{n\tau_\varepsilon + \tau_\alpha}\right)$$

$$p(\beta \mid \alpha, \tau_\varepsilon, y) \sim N\left(\frac{\tau_\varepsilon \sum_i y_i x_i - \alpha \sum_i y_i + \tau_\beta \mu_\beta}{\tau_\varepsilon \sum_i x_i^2 + \tau_\beta}, \frac{1}{\tau_\varepsilon \sum_i x_i^2 + \tau_\beta}\right) \qquad \text{(A5.3)}$$

$$p(\tau_\varepsilon \mid \alpha, \beta, y) \sim G\left(\frac{n}{2} + c - 1, d + \frac{1}{2}\sum_i (y_i - (\alpha + \beta x_i))^2\right)$$

Taking samples from these full conditional distributions is equivalent to taking samples from the joint posterior distribution.

Suppose the next step was to fit the second model in equation (A5.1). A Bayesian approach is to be used again by implementing Gibbs sampling. The full conditional distributions are the same as in equation (A5.3) but the parameters of the second model replace the parameters for the first model in equation (A5.1).

$$p(\gamma \mid \delta, \tau_\eta, z) \sim N\left(\frac{\tau_\eta \sum_i z_i - \delta \sum_i y_i + \tau_\gamma \mu_\gamma}{n\tau_\eta + \tau_\gamma}, \frac{1}{n\tau_\eta + \tau_\gamma}\right)$$

$$p(\delta \mid \gamma, \tau_\eta, z) \sim N\left(\frac{\tau_\eta \sum_i z_i y_i - \gamma \sum_i z_i + \tau_\delta \mu_\delta}{\tau_\eta \sum_i y_i^2 + \tau_\delta}, \frac{1}{\tau_\eta \sum_i y_i^2 + \tau_\delta}\right) \qquad \text{(A5.4)}$$

$$p(\tau_\eta \mid \gamma, \delta, z) \sim G\left(\frac{n}{2} + e - 1, f + \frac{1}{2}\sum_i (z_i - (\gamma + \delta y_i))^2\right)$$

By sampling from these distributions, the parameter estimates for the second model can be obtained. Two models would then have been fitted independently.

Now assume that the following is known: X is measured without error, Y and Z are measured with error. With this in mind, using the observed y values to fit a model to z is not desirable, as this will introduce bias in to the parameter estimates. Instead of using the observed y data, the predicted y values could be used instead. Without taking into account the parameter uncertainty, using predicted values will be as biased as using

the observed data. It is therefore necessary to take into account the error in estimating the parameters. The full conditional distributions are therefore going to be of the form given in equation (A5.5).

$$p(\gamma \mid \delta, \tau_\eta, z) \sim N\left(\frac{\tau_\eta \sum_i z_i - \delta \sum_i (\alpha + \beta x_i) + \tau_\gamma \mu_\gamma}{n \tau_\eta + \tau_\gamma}, \frac{1}{n \tau_\eta + \tau_\gamma}\right)$$

$$p(\delta \mid \gamma, \tau_\eta, z) \sim N\left(\frac{\tau_\eta \sum_i z_i (\alpha + \beta x_i) - \gamma \sum_i z_i + \tau_\delta \mu_\delta}{\tau_\eta \sum_i (\alpha + \beta x_i)^2 + \tau_\delta}, \frac{1}{\tau_\eta \sum_i (\alpha + \beta x_i)^2 + \tau_\delta}\right)$$

$$p(\tau_\eta \mid \gamma, \delta, z) \sim G\left(\frac{n}{2} + e - 1, f + \frac{1}{2} \sum_i z_i - (\gamma + \delta(\alpha + \beta x_i))^2\right)$$

$$p(\alpha \mid \beta, \tau_\varepsilon, y) \sim N\left(\frac{\tau_\varepsilon \sum_i y_i - \beta \sum_i x_i + \tau_\alpha \mu_\alpha}{n \tau_\varepsilon + \tau_\alpha}, \frac{1}{n \tau_\varepsilon + \tau_\alpha}\right)$$

$$p(\beta \mid \alpha, \tau_\varepsilon, y) \sim N\left(\frac{\tau_\varepsilon \sum_i y_i x_i - \alpha \sum_i y_i + \tau_\beta \mu_\beta}{\tau_\varepsilon \sum_i x_i^2 + \tau_\beta}, \frac{1}{\tau_\varepsilon \sum_i x_i^2 + \tau_\beta}\right)$$

$$p(\tau_\varepsilon \mid \alpha, \beta, y) \sim G\left(\frac{n}{2} + c - 1, d + \frac{1}{2} \sum_i (y_i - (\alpha + \beta x_i))^2\right) \tag{A5.5}$$

As can be seen from the full conditionals, the y-x model does not depend on the z-y model but the z-y model does depend on the y-x model through the conditional distributions for α and β.

The following algorithm then gives the sampling procedure necessary to estimate the parameters.

1.   Set initial estimates to the parameters of the two models $(\alpha, \beta, \gamma, \delta)$.

2.   Sample parameters for y-x model.

3.   Using current sampled parameters from y-x model, sample z-y parameters from full conditionals given in equation (A5.5).

4.   Go back to 2.

This method will allow the parameters from both the y-x and z-y model to be estimated and enable the variability from the y-x model to be incorporated into the z-y model. Such a method is easily implemented in BUGS and could be easily generalised to include nonlinear models for the pharmacokinetic data and any model for the pharmacodynamic data.

6 Oxybutynin – Phase III Efficacy and Toxicity Data

6.1 Review of Oxybutynin and Incontinence

Urinary incontinence refers to the involuntary loss of urine in sufficient amounts to be considered a social and health problem. While it is psychologically distressing and socially disruptive, urinary incontinence is believed to be an under reported and under diagnosed medical condition. There are four basic categories of incontinence: stress, urge, functional and overflow. Urge incontinence is associated with detrusor muscle overactivity. Stress incontinence is associated with an incompetent outlet. In geriatric patients a major cause of incontinence is bladder instability due to detrusor instability and detrusor hyperreflexia. Initial management involves identification and treatment of reversible causes. In patients with urinary incontinence and detrusor instability, bladder relaxant medications, such as oxybutynin, have shown some effectiveness.

Oxybutynin is a tertiary amine compound with direct antispasmodic effect on smooth muscle, an anticholinergic effect and local anaesthetic effect. Studies in patients with uninhibited neurogenic and reflex neurogenic bladders demonstrate that oxybutynin increases urinary bladder capacity, decreases the frequency of uninhibited contractions of the detrusor muscle and delays the initial desire to void. Oxybutynin has been officially approved by the FDA (Food and Drug Administration) for the pharmacological treatment of incontinence.

Oxybutynin is a good candidate for controlled drug delivery because it is an effective therapeutic agent, it has a short plasma half life and it undergoes extensive first pass metabolism. Controlled drug delivery by OROS oxybutynin chloride (sustained release formulation) or TTS (Transdermal Therapeutic System formulation) would eliminate the need for dosing three times a day, enhance patient compliance, and possibly eliminate or reduce troublesome side effects. The most common side effect experienced with oxybutynin is dry mouth; other adverse effects included constipation, drowsiness, blurred vision, nausea, abdominal discomfort, facial flushing and difficulty in micturition. In general, clinical studies with oxybutynin show that it is significantly better than placebo in improving urinary incontinence, but many patients drop out because of side effects (Kirkali and Whitaker (1987)).

Oxybutynin is rapidly and well absorbed. The recommended daily dose in adults is 5 mg, two or three times a day. Following oral administration of 5 mg oxybutynin in normal volunteers, the mean maximum plasma concentration (C_{max}) was 8.2 ng/mL at 0.8 hours. Mean half life value reported for 5 mg tablet (instant release) was 2.44 hours, and 2.31 hours following a intravenous infusion of 5 mg over 10-20 minutes.

6.2 Study Design

OROS oxybutynin chloride and TTS oxybutynin are being developed for the treatment of urge incontinence. Potential benefits of these dosage forms are improved efficacy and reduced frequency of dosing and reduced side effects. The study carried out was a randomised, double-blind, placebo controlled, escalating dose, parallel group study with

one week of single-blind placebo run in followed by randomisation to one of the 5 treatment groups. After randomisation, the study was double-blind with respect to active versus placebo treatment. The study was to be performed in 120 evaluable middle aged and elderly women (40 years and older) with urge incontinence to compare the safety and efficacy of OROS oxybutynin chloride and TTS oxybutynin with IR (immediate release formulation) oral oxybutynin and with placebo. 176 patients were enrolled to complete 120 patients as follows: 30 patients for OROS oxybutynin treatment, 30 patients for TTS oxybutynin treatment, 30 patients for the IR oxybutynin treatment and 15 for each of the two placebo (oral and trans-dermal) groups. The actual number of patients available for analysis is given in table 6.1. Out of a possible 134 evaluable patients, 118 completed the study. Patients dropped out for reasons unrelated to the study.

Table 6.1. *Number of patients in each treatment group.*

	OROS	IR	TTS	Placebo (TTS)	Placebo (oral)	Total
Number at start	34	32	35	17	16	134
Number completed	32	30	25	16	15	118

The study objectives were:

(1) To compare the efficacy of OROS oxybutynin chloride and TTS oxybutynin to that of active control, IR oral oxybutynin chloride, and to placebo in middle aged and elderly patients with urinary urge incontinence. The primary efficacy measure was number of urge incontinence episodes per week (as recorded in the patient urinary diary).

(2) To evaluate the side effect profile of OROS, TTS and IR before and after the first dose of each week through subjective assessment of anticholinergic effects: dry mouth, constipation, blurred vision, nausea and drowsiness.

(3) To determine steady state plasma concentrations of oxybutynin and its metabolite, at weekly time points before the first dose of each week and one hour after the last dose of the study.

(4) To determine the urine concentration of oxybutynin and desethyloxybutynin (metabolite) at the end of each two week interval.

Only study objectives (1) and (2) were considered here as concentration of the parent and metabolite were not available.

These data were analysed also by Gupta *et al* (1999) using a NONMEM approach and their results were contrasted to those obtained in the following work.

6.3 Efficacy Data Analysis

The pharmacodynamic measure used for the efficacy of oxybutynin was the number of urge incontinence episodes in a seven day period. This measure is a count and in this work, Poisson log-linear regression models were used to analyse the data. The covariates available for the analysis of these data were dose, time and baseline count (the number of urge incontinence episodes in the first week). The data analysed corresponds to the OROS, IR and oral placebo groups. The analysis for TTS and patch placebo is not reported here. BUGS 0.6 was used for the analysis of the data throughout this chapter.

231

6.3.1 Placebo Model

As the design of this study was a forced dose escalation study then dose was confounded with time. With the inclusion of a placebo group, it was then possible to model the time trend of urge incontinence independent of any dose effect.

As the data were in terms of counts of urge incontinence episodes per week, it was assumed that the data were from a Poisson distribution. The Poisson distribution is a discrete distribution (equation (6.1)) based on a non-negative integer number of events per unit of time or space.

$$\Pr(Y = y) = e^{-\mu}\frac{\mu^{y}}{y!}, y = 0,1,2,...,\mu > 0. \tag{6.1}$$

The link function associated with the count data is the log link function. This can be derived based on distributional assumptions that the Poisson distribution is a member of the exponential family and then the link function can be derived from the cumulant function. Details are given in McCullagh and Nelder (1989).

The first model to be fitted to the placebo data was a mixed effects model linear in time as defined in equation (6.2). Normal priors were assigned to the fixed effect parameters and a normal distribution was assigned to the random effect with the inverse of the variance being assigned a gamma distribution.

$$\begin{aligned}
\log(E(Y_{ij} \mid b_i)) &= \beta_o + \beta_1 time_{ij} + b_i \\
Y_{ij} \mid \underline{\beta}, b_i &\sim Pois(E(Y_{ij} \mid b_i)) \\
\beta_0 &\sim N(0,10000), \beta_1 \sim N(0,1\times10^4) \\
b_i &\sim N(0,\sigma_b^2), \frac{1}{\sigma_b^2} \sim G(0.001,0.001)
\end{aligned} \tag{6.2}$$

The random effect b_i, is additive on the intercept term. The parameter values were determined from taking a sample of 5000 iterations after a burn-in of 1500 iterations. The results are given in table 6.2.

Table 6.2. *Results of log-linear model with linear time.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_0	2.87	0.236	2.877	(2.303,3.346)
β_1	-0.0168	0.00186	-0.0168	(-0.0205,-0.0133)
σ_b	0.889	0.188	0.860	(0.607,1.329)

From these placebo data, there was a significant time effect based on the 95% credible interval. This implies that without giving oxybutynin or any other drug, the number of urge incontinence episodes will converge to zero. If this were truly the case, then there would be no need to administer any drug to alleviate incontinence. To see if more complex models might describe the data better, a cubic and quadratic polynomial model was fitted to the data and the results are given in tables 6.3 and 6.4.

Table 6.3. *Results of cubic polynomial for efficacy placebo data.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_0	2.875	0.280	2.845	(2.375,3.451)
β_1	-0.0264	0.0231	-0.0256	(-0.0763,0.0138)
β_2	4.584×10^{-4}	9.65×10^{-4}	5.647×10^{-4}	(-0.00124,0.00246)
β_3	-6×10^{-6}	1.182×10^{-5}	-6.819×10^{-6}	(-2.97×10^{-5},1.57×10^{-5})
σ_b	0.887	0.180	0.860	(0.614,1.306)

Table 6.4. *Results of quadratic polynomial for efficacy placebo data.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_0	2.946	0.279	2.907	(2.51,3.67)
β_1	-0.0165	0.00796	-0.0171	(-0.0312,2.551×10^{-6})
β_2	-8.071×10^{-6}	1.456×10^{-4}	4.987×10^{-6}	(-3.131×10^{-4},2.62×10^{-4})
σ_b	0.904	0.198	0.875	(0.61,1.386)

The cubic term in table 6.3 was small in magnitude and the 95% credible interval included zero. The quadratic term in table 6.4 was also non-significant so it appeared that the model linear in time was as good as the quadratic and cubic models. Another possible model for the efficacy placebo data was to use logarithm of time rather than linear time. The results of such a model are given in table 6.5.

Table 6.5. *Results of log(time) model efficacy placebo data.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_0	3.364	0.229	3.371	(2.867,3.797)
β_1	-0.334	-0.0351	-0.333	(-0.407,-0.268)
σ_b	0.895	0.188	0.869	(0.608,1.338)

A plot of the log(time) and linear time models are given in figure 6.1. Pearson residuals are given in figure 6.2 and figure 6.3 for the linear time and log(time) models respectively.

Figure 6.1. Comparison of log(time) and linear time models.
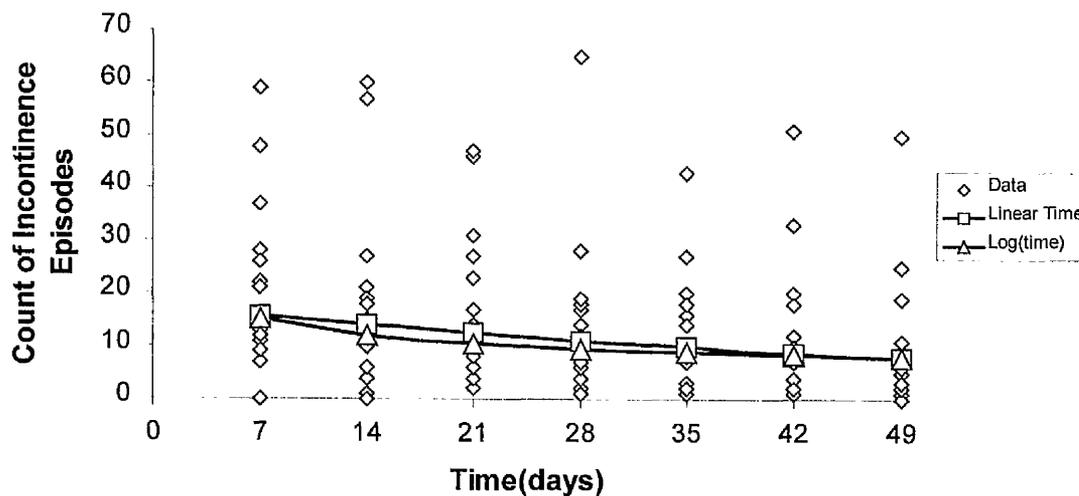
Figure 6.2. *Pearson Residuals versus model predicted linear time model.*
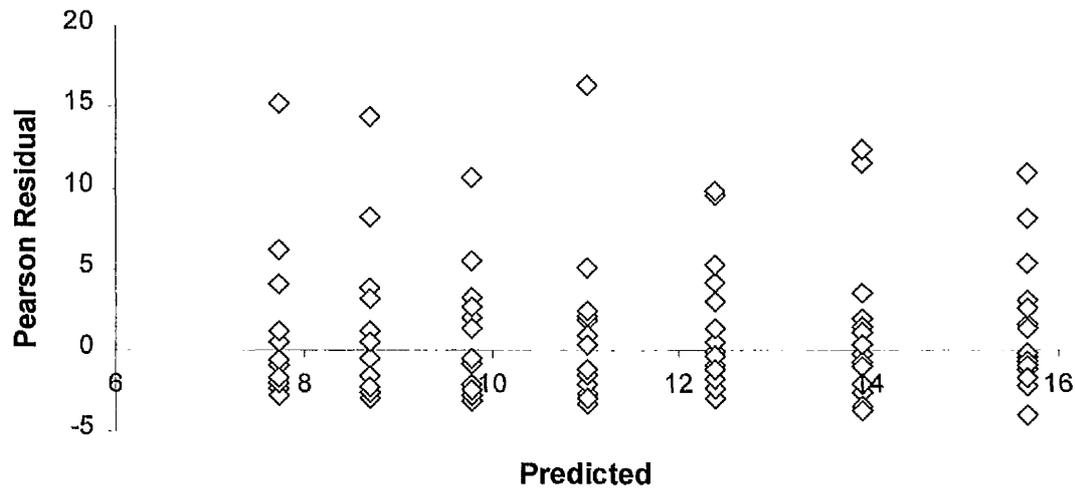
Figure 6.3. *Pearson Residuals versus model predicted log(time) model.*
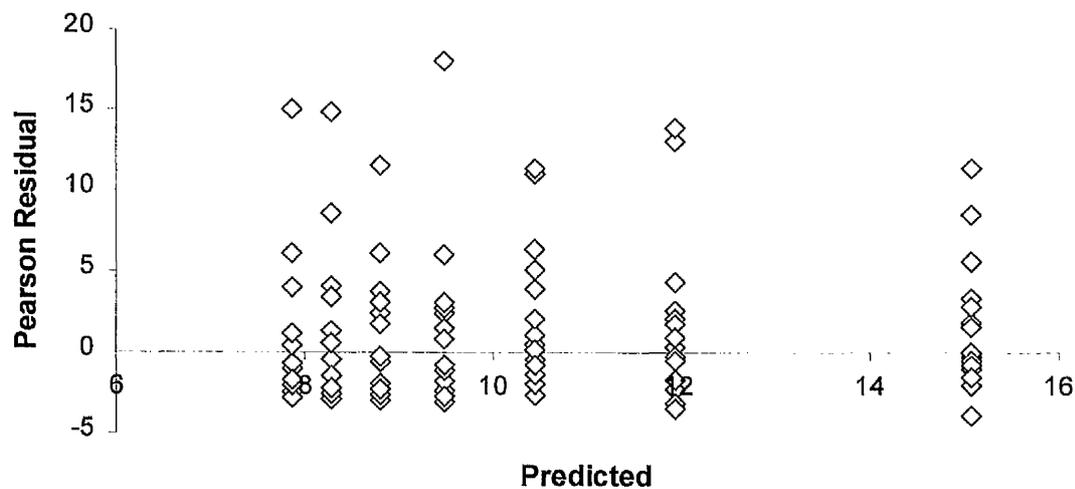
The Pearson residuals for the Poisson distribution are defined in equation (6.3).

$$r_{P_i} = \frac{Y_{ij} - E(Y_{ij})}{\sqrt{E(Y_{ij})}} \tag{6.3}$$

There is little difference between the fit of the model in figure 6.1, and the Pearson residual plots also show little difference in the quality of the fit. Based on this, it was decided to choose the linear time model.

It can be seen that for the population model, the Pearson residuals are large in magnitude and unevenly distributed around the axis implying that the model is not satisfactory in describing the data.

As already mentioned previously, the linear time model predicts that the number of incontinence episodes will converge to zero as time increases implying no need for drug administration. The arithmetic mean of the number of incontinence episodes per week at 49 days in the placebo group was 10.2. A model that does not converge to zero by default could be fitted to see if the number of incontinence episodes as time increased asymptoted to a non-zero value. Such a model is given in equation (6.4).

$$E(Y_{ij}) = y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$
(6.4)

In this model, y_{0i} is the baseline count of urge incontinence episodes (first 7 days). For future collected data, y_0 could be treated as the average number of the incontinence episodes in a week. This covariate acts like a normalising constant so that at the end of the first week (day 7) then the model predicts the baseline incontinence count. As time increases, this model converges to $(\beta_1 - \beta_2)$ which is the number of incontinence episodes that will be reached without administering any drug. To ensure non-negativity, $\beta_1 > \beta_2$. The parameter β_3 is the rate at which the incontinence episodes decrease. The desired Bayesian model specification is given in equation (6.5). $I(a,b)$ denotes the distribution is truncated between a and b.

$$E(Y_{ij}) = y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$
$$Y_{ij} \sim Pois(E(Y_{ij}))$$
(6.5)
$$\beta_1 \sim N(\mu_{\beta_1}, 1 \times 10^4), \beta_2 \sim N(\mu_{\beta_2}, 1 \times 10^4)I(\beta_1, \infty), \log(\beta_3) \sim N(\mu_{\beta_3}, 1 \times 10^4)$$

There are several problems with this specification of the model when using BUGS for estimation. Firstly, this model was not a generalised linear model as the model is

nonlinear in the three parameters for a log-linear Poisson model. BUGS 0.6 uses an adaptive rejection sampling algorithm for generalised linear models (other than normally distributed models) and the Griddy-Gibbs algorithm for normal distributed nonlinear models (implying a non-log-concave full conditional distribution). BUGS requires an adaptive Metropolis algorithm for sampling from generalised nonlinear models. As BUGS 0.6 does not have such an algorithm, then this model specification could not be estimated. Instead, the following model specification was used.

$$E(Y_{ij}) = y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$
$$Y_{ij} \sim N(E(Y_{ij}), \sigma^2)I(0, \infty)$$
$$\frac{1}{\sigma^2} \sim G(0.001, 0.001)$$
$$\beta_1 \sim U(0,5), \beta_2 \sim U(0, \beta_1), \beta_3 \sim U(0,1)$$

(6.6)

The Poisson likelihood was approximated by a normal distribution which was truncated to be positive. Whether a log normal distribution would have been more appropriate was not considered. Instead of using normal priors on the parameters, uniform priors were used so that $\beta_1 > \beta_2$ and β_3 was positive. The parameter estimates are given in table 6.6 and are based on a run of 10,000 iterations where every second sample was saved to estimate the parameter values.

Table 6.6. *Results of nonlinear placebo model.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_1	1.337	0.2135	1.293	(1.067,1.805)
β_2	1	0.173	0.990	(0.685,1.351)
β_3	0.054	0.0263	0.47	(0.0233,0.114)
σ	10.41	0.743	10.36	(9.121,11.95)

The number of counts converges to $0.337 \times y_0$, which for an initial count of 25 incontinence episodes is 8.5 episodes per week as time increases. The residual term is on a different scale in the normal model compared to that in the Poisson distribution.

The residual is additive in the normal distribution and multiplicative in the Poisson distribution. Individual population predicted plots (no subject specific random effects) are given in figures 6.4-6.7. Residual and weighted residual plots are given in figures 6.8 and 6.9 respectively.

Figure 6.4. *Plot of nonlinear model for subject 112.*
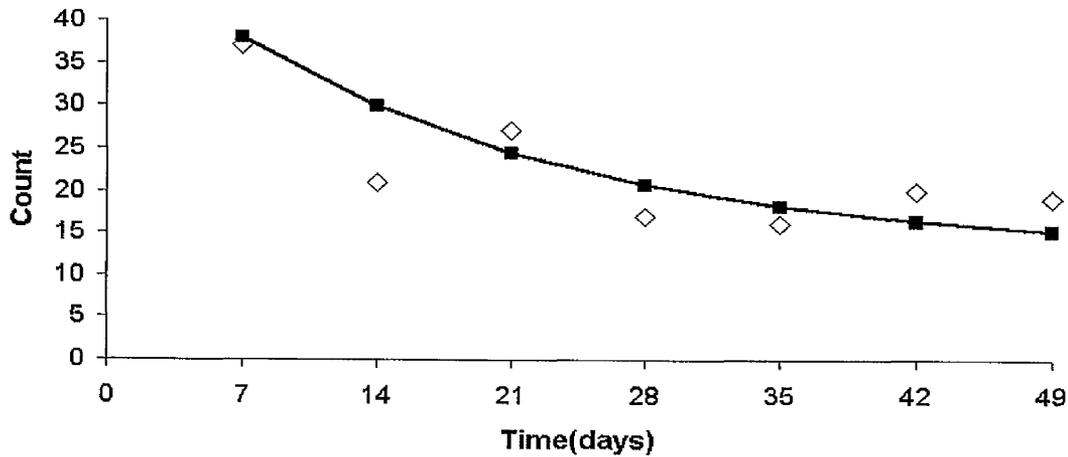
Figure 6.5. *Plot of nonlinear model for subject 301.*
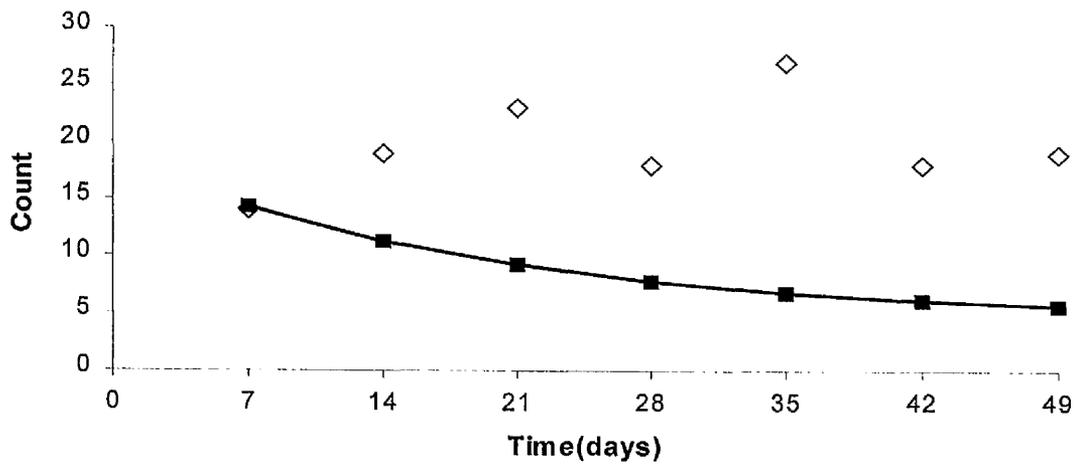
Figure 6.6. *Plot of nonlinear model for subject 516.*
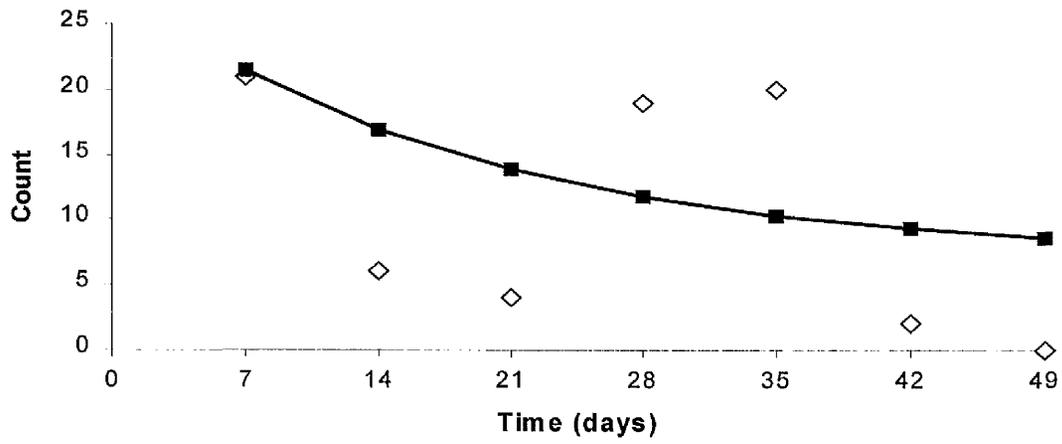
Figure 6.7. *Plot of nonlinear model for subject 901.*
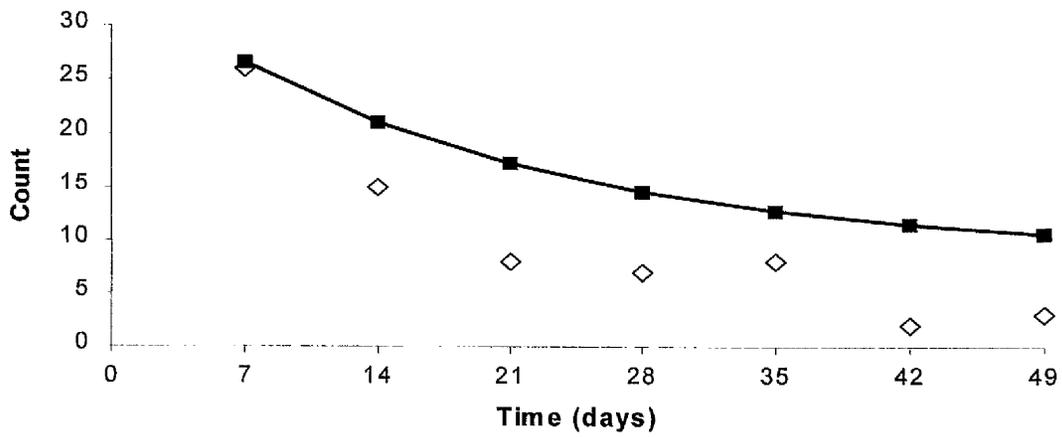
Figure 6.8. *Residual versus time plot for nonlinear placebo model.*
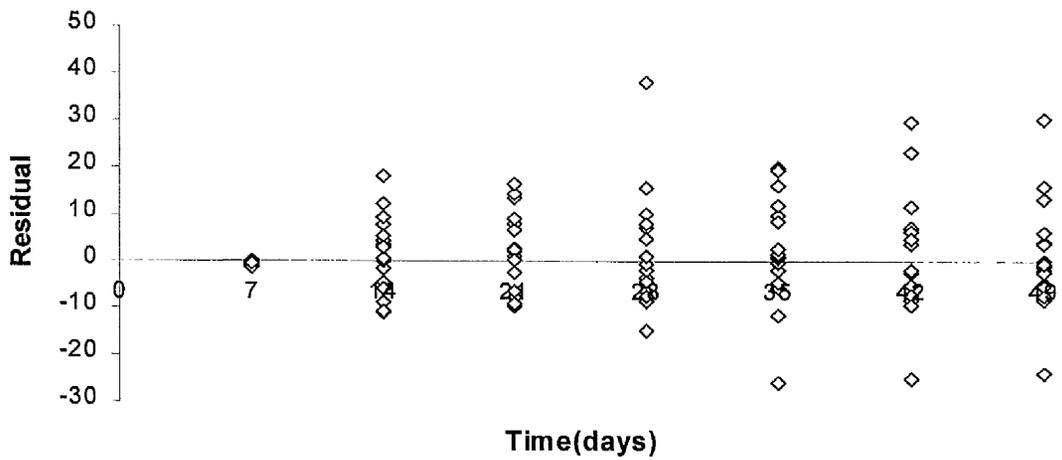
Figure 6.9. *Weighted residual versus population predicted plot for nonlinear model.*
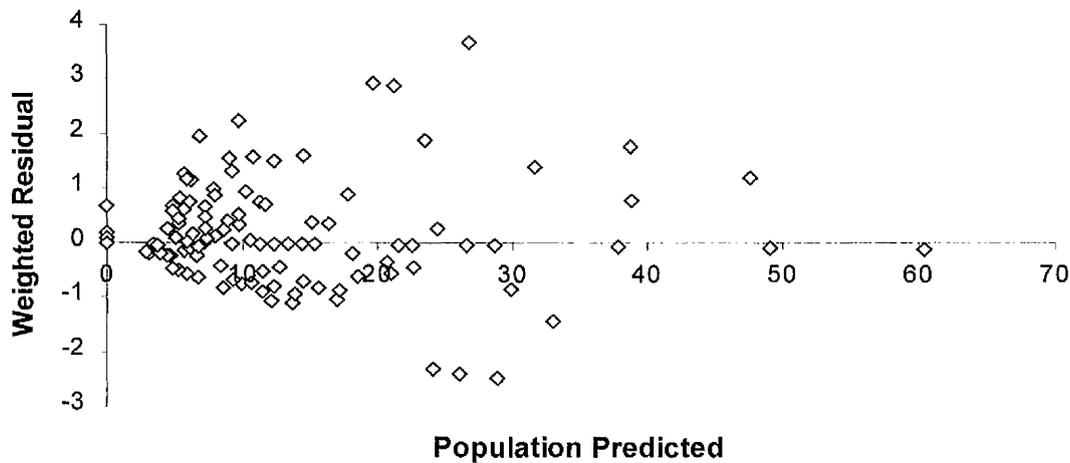
The individual fits in figures 6.4-6.7 show a varying degree of quality of fit. Individual 112 was a reasonable individual fit, individual 301 was under predicted, 901 was over predicted and individual 516 had considerable inter-week variability which made the data difficult to model. The residual plot (figure 6.8) shows a good fit to the population data and the weighted residual plot (figure 6.9) appears satisfactory.

Gupta *et al* (1999) reported a similar type of placebo model where the equivalent parameter of β_3 was 0.029 compared to 0.054 estimated here. The result is not directly comparable but gives an indication of the different estimates. A pseudo half-life can be estimated by $\ln(2)/\beta_3$ and are 24 and 13 days for the number of incontinence episodes to drop by half for Gupta *et al* and this work respectively.

6.3.2 Active Treatment Groups

The purpose of the placebo model was to be able to correct for any time effects in the active treatment groups that were not due to the drug. As there was a time effect in the placebo data, this needed to be included in the active treatment group model. The placebo model can be included in the active treatment model easily by scaling the population placebo model by the initial seven day run-in urge incontinence count. To obtain a correct dose effect model, the placebo model must be included but the form of the dose model was probably dependent on the placebo model chosen.

As described earlier, there are two active treatment groups, the IR (immediate release) group and the OROS (sustained release) treatment group. One of the primary study objectives was to compare the efficacy and adverse effect profiles of the two formulations. The purpose of this section was to define models for the two treatment groups and compare the dose effects of the two formulations in terms of the efficacy.

The number of women in each treatment group at the end of each week are given in table 6.7.

Table 6.7. *Sample sizes in each group at the end of each week.*

Time(days)	7	14	21	28	35	42	49
IR	32	32	32	32	31	30	30
OROS	34	34	34	34	34	34	32
Placebo	16	16	16	15	15	15	15

As can be seen from table 6.7, by the end of day 49, 2 women had dropped out of both the IR and OROS treatment groups and 1 woman dropped out of the placebo treatment

group. As this corresponds to only 6% of the total number of women in these three groups, it was decided not to incorporate any method to adjust for the missing data. It was reported in the data summary supplied by the company that missing data was not associated with the drug and can be treated as missing completely at random. It should be noted that when BUGS encounters a missing variable, it treats it as a random variable that needs to be determined and estimates the missing data.

The general form in which the model can be written is given in equation (6.7).

$$\log(E(Y_{ij})) = f_P + f_D + f_{PD} \tag{6.7}$$

f_P, f_D and f_{PD} are the placebo, dose and interaction models respectively. The placebo model is as described in section 6.3.1, the dose model is a function in dose alone and the interaction model is a model that takes into account any combination effects between dose and time. The f_{PD} model was not considered here. The log link function was used as the data were urge incontinence episodes.

The first model fitted to all the data was that defined in equation (6.8).

Placebo data

$$Y_{ijplac} \sim N(E(Y_{ijplac}), \sigma_{plac}^2)I(0,\infty), E(Y_{ijplac}) = y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$

$$\frac{1}{\sigma_{plac}^2} \sim G(0.001, 0.001), \beta_1 \sim U(0,5), \beta_2 \sim U(0, \beta_1), \beta_3 \sim U(0,1)$$

Active treatment data

$$Y_{ijactive} \sim Pois(E(Y_{ijactive})) \tag{6.8}$$

$$E(Y_{ijactive}) = trt_i(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + \beta_{OROS} dose_i)$$

$$+ (1 - trt_i)(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + \beta_{IR} dose_i) + \varepsilon_{ij}$$

$$\beta_{OROS} \sim N(0,1000), \beta_{IR} \sim N(0,1000),$$

$$\varepsilon_{ij} \sim N(0, \sigma_{active}^2), \frac{1}{\sigma_{active}^2} \sim G(0.001, 0.001)$$

In this model specification, the placebo model is refitted using the same Bayesian specification so as to allow the parameter uncertainty to be included in the time model for the active treatment groups. The population placebo model was still used but normalised to the active treatment week 1 urge incontinence count. The results of this model are given in table 6.9. The placebo parameters are the same as those given in table 6.6.

Table 6.9. *Results of linear dose model for active treatment groups.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS}	-0.145	0.0132	-0.145	(-0.168,-0.121)
β_{IR}	-0.105	0.0122	-0.105	(-0.126,-0.0835)
σ_{active}	0.889	0.0486	0.887	(0.797,0.986)

The estimates of the dose parameters are both negative with the OROS data having a steeper gradient. This might imply that OROS has better efficacy than IR as the gradient is steeper implying that a lower dose is required to attain the same effect as IR. Plots of the linear dose OROS and IR models are given in figure 6.10 and 6.11 respectively.

Figure 6.10. *Plot of linear dose OROS model.*
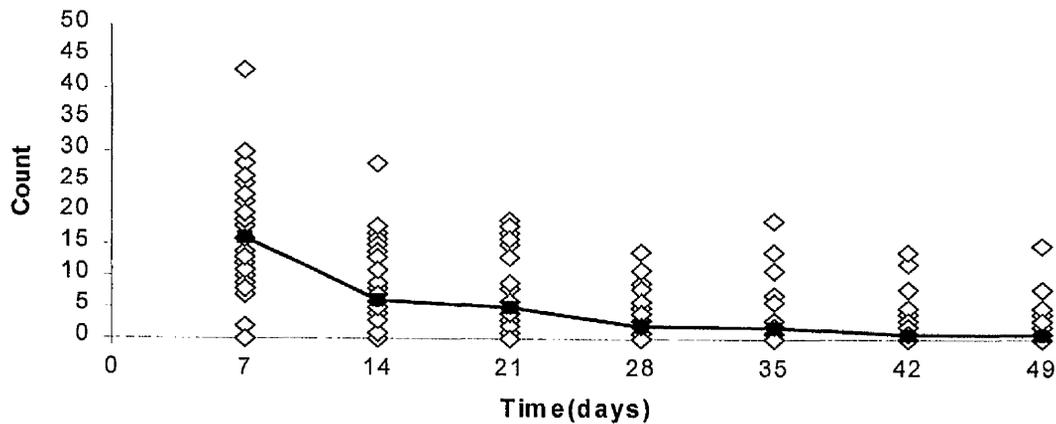
Figure 6.11. *Plot of linear dose IR model.*
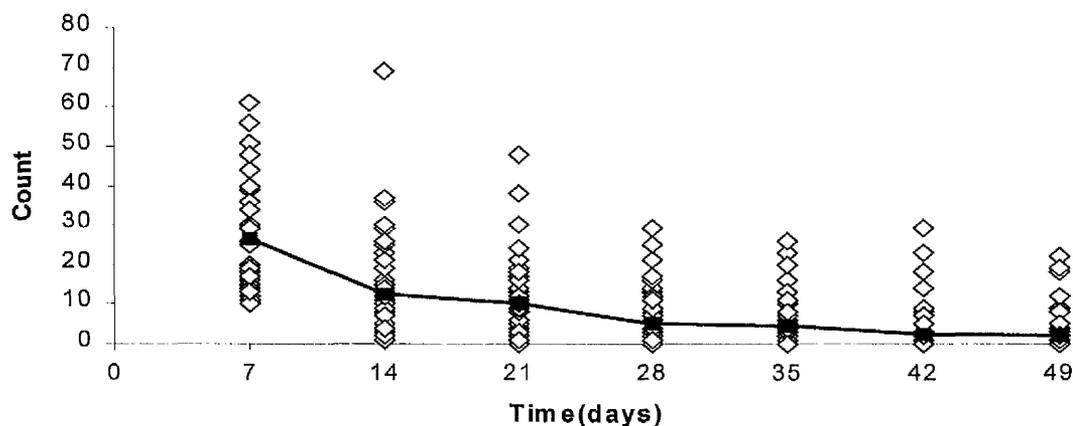
To see if there was a statistically significant difference between the gradient terms for the two treatment groups, the model was redefined in terms of equation (6.9). This model parameterisation allowed the checking of the difference of the dose estimates by considering the 95% credible intervals.

$$E(Y_{ijactive}) = trt_i(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + \beta_{OROS}dose_i)$$
$$+ (1 - trt_i)(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + (\beta_{OROS} + \beta_{IR-OROS})dose_i) + \varepsilon_{ij} \qquad (6.9)$$
$$\beta_{OROS} \sim N(0,1000), \beta_{IR-OROS} \sim N(0,1000)$$

The results of this model are given in table 6.10.

Table 6.10. *Results of reparameterised linear dose model for active treatment groups.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS}	-0.145	0.0132	-0.146	(-0.168,-0.122)
$\beta_{IR-OROS}$	0.0402	0.0116	0.0403	(0.0172,0.0622)
σ_{active}	0.8879	0.0497	0.887	(0.798,0.991)

The 95% credible interval on the parameter, $\beta_{IR-OROS}$ does not include zero so there was a statistically significant difference between the OROS and IR dose effect. The term σ_{active} is an overall residual error term and no random effect was included in the model.

Another linear dose model was run with a subject specific additive random effect on the intercept and a residual error term for each active dose group. The model is given in equation (6.10) and the results are given in table 6.11.

$$E(Y_{ij active}) = trt_i(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + \beta_{OROS} dose_i + b_{OROSi} + \varepsilon_{OROSij})$$
$$+ (1 - trt_i)(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}})) + (\beta_{OROS} + \beta_{IR-OROS}) dose_i + b_{IRi} + \varepsilon_{IRij})$$
$$b_{OROSi} \sim N(0, \sigma^2_{OROSb}), \varepsilon_{OROSij} \sim N(0, \sigma^2_{OROS\varepsilon})$$
$$b_{IRi} \sim N(0, \sigma^2_{IRb}), \varepsilon_{IRij} \sim N(0, \sigma^2_{IR\varepsilon})$$

(6.10)

Table 6.11. *Results of linear dose model with additive random effects and residual terms.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS}	-0.132	0.0127	-0.132	(-0.157,-0.107)
β_{IR}	-0.101	0.0132	-0.102	(-0.126,-0.0741)
σ_{OROSb}	0.519	0.0854	0.511	(0.368,0.704)
σ_{IRb}	0.772	0.119	0.760	(0.574,1.032)
$\sigma_{OROS\varepsilon}$	0.507	0.0616	0.504	(0.403,0.623)
$\sigma_{IR\varepsilon}$	0.633	0.0617	0.631	(0.532,0.747)

The interindividual and residual variability appears to be larger in the IR treatment group than the OROS treatment group. The dose parameter estimates were smaller in magnitude compared to the model with only an overall residual error term. The model was rerun without the treatment group specific residual term to see if the parameters changed substantially. The parameter values are given in table 6.12.

Table 6.12. *Results of linear dose model with additive random effects.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS}	-0.0944	0.00987	-0.0947	(-0.113,-0.0744)
β_{IR}	-0.0541	0.00906	-0.0544	(-0.0707,-0.0353)
σ_{OROSb}	0.388	0.0626	0.383	(0.283,0.527)
σ_{IRb}	0.524	0.0753	0.517	(0.399,0.693)

Again, the dose parameter estimates were down on the previous estimates and so were the between subject variability estimates. This could be due to the different model specification or just inter-BUGS run variability where different seeds are used to start the sampling procedure at each run.

Another model with linear dose as a predictor was run in BUGS where there was no additive random effect but with an additive residual component. The parameter estimates are given in table 6.13.

Table 6.13. *Results of linear dose model with residual error terms.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS}	-0.137	0.0115	-0.137	(-0.16,-0.115)
β_{IR}	-0.109	0.0121	-0.109	(-0.132,-0.085)
$\sigma_{OROS\varepsilon}$	0.744	0.0661	0.742	(0.622,0.880)
$\sigma_{IR\varepsilon}$	1.005	0.0732	1.002	(0.871,1.16)

The parameter estimates for the dose effect are closer to those in tables 6.9 and 6.11.The residual components also show that there was greater variability in the IR treatment group than the OROS treatment group.

The next step was to check whether more complicated dose models would be able to fit the data any better. A cubic polynomial in dose was considered as defined in equation (6.11).

$$E(Y_{ijactive}) = trt_i(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$
$$+ \beta_{OROS1}dose_i + \beta_{OROS2}dose_i^2 + \beta_{OROS3}dose_i^3 + b_{OROSi} + \varepsilon_{OROSij})$$
$$+ (1 - trt_i)(y_{0i}(\beta_1 - \beta_2(1 - e^{-\beta_3 t_{ij}}))$$
$$+ \beta_{IR1}dose_i + \beta_{IR2}dose_i^2 + \beta_{IR3}dose_i^3 + b_{IRi} + \varepsilon_{IRij})$$

(6.11)

The results of this model are given in table 6.14.

Table 6.14. *Results of cubic dose model.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS1}	-0.0541	0.0607	-0.0573	(-0.169,0.0706)
β_{OROS2}	-0.0191	0.0115	-0.0192	(-0.0424,0.00132)
β_{OROS3}	9.836×10^{-4}	5.164×10^{-4}	9.963×10^{-4}	$(5.66\times10^{-5},1.968\times10^{-3})$
σ_{OROSb}	0520	0.0882	0.511	(0.369,0.713)
$\sigma_{OROS\varepsilon}$	0.499	0.0621	0.497	(0.393,0.619)
β_{IR1}	-0.0649	0.0440	-0.0663	(-0.158,00237)
β_{IR2}	-0.00951	0.00756	-0.00955	(-0.0225,0.00825)
β_{IR3}	5.125×10^{-4}	3.662×10^{-4}	5.402×10^{-4}	$(-2.953\times10^{-4},1.082\times10^{-3})$
σ_{IRb}	0.772	0.120	0.760	(0.571,1.046)
$\sigma_{IR\varepsilon}$	0.628	0.0619	0.624	(0.529,0.742)

The cubic parameter estimate for the OROS treatment group was significantly different from zero but the magnitude of the parameter was very small. The cubic term for the IR treatment group was not significant. The model was re-analysed without the cubic terms and the results are given in table 6.15.

Table 6.15. *Results of quadratic dose model.*

	Mean	Standard Deviation	Median	95% Credible Interval
β_{OROS1}	-0.166	0.0309	-0.168	(-0.223,-0.107)
β_{OROS2}	0.00274	0.00209	0.00278	(-0.00121,0.00651)
σ_{OROSb}	0.519	0.0882	0.511	(0.373,0.719)
$\sigma_{OROS\varepsilon}$	0.508	0.0616	0.505	(0.402,0.630)
β_{IR1}	-0.131	0.0310	-0.131	(-0.193,-0.0698)
β_{IR2}	0.00225	0.00201	0.00221	(-0.00173,0.00617)
σ_{IRb}	0.776	0.120	0.763	(0.577,1.043)
$\sigma_{IR\varepsilon}$	0.634	0.0627	0.630	(0.530,0.753)

It can be seen from table 6.15 that both quadratic terms are not significant and hence are not any better at predicting the active efficacy treatment data than the linear dose model. As the IR data was best described by a linear model, these data were not considered further. To see whether there was a practically significant difference between the cubic and linear models for the OROS treatment group, a simple comparison of the two

models was made. In figure 6.12, a plot of the cubic and linear models are shown. There does not appear to be much of a difference between the two models in terms of model fit. In figures 6.13 and 6.14 are Pearson residual plots to see if this shows any difference between the models. Again, there was not much difference but the cubic model appears to be more evenly distributed. One problem with the cubic dose model is that the cubic term is positive implying that as dose increases, then the model will predict that the number of urinary urge incontinence episodes will eventually increase with dose whereas the linear dose model implies that the count converges to zero as dose increases.

Figure 6.12. *Cubic and linear dose model for OROS data.*
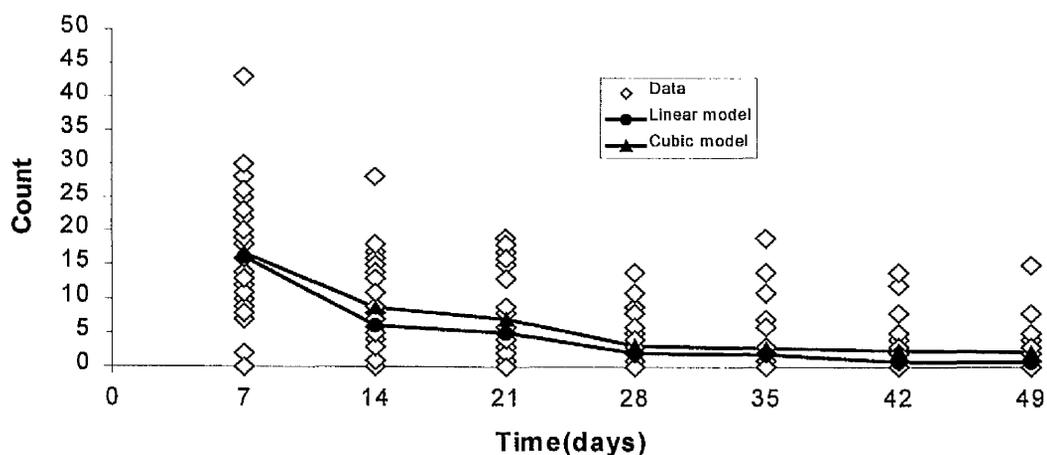
Figure 6.13. *Pearson residual versus predicted for linear dose model for OROS data.*

Population Predicted
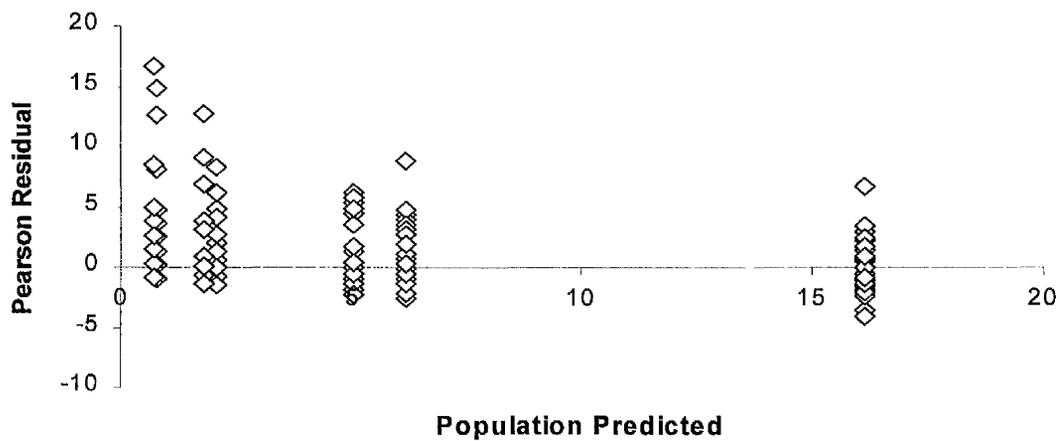
Figure 6.14. *Pearson residual versus predicted for cubic dose model for OROS data.*

Population Predicted
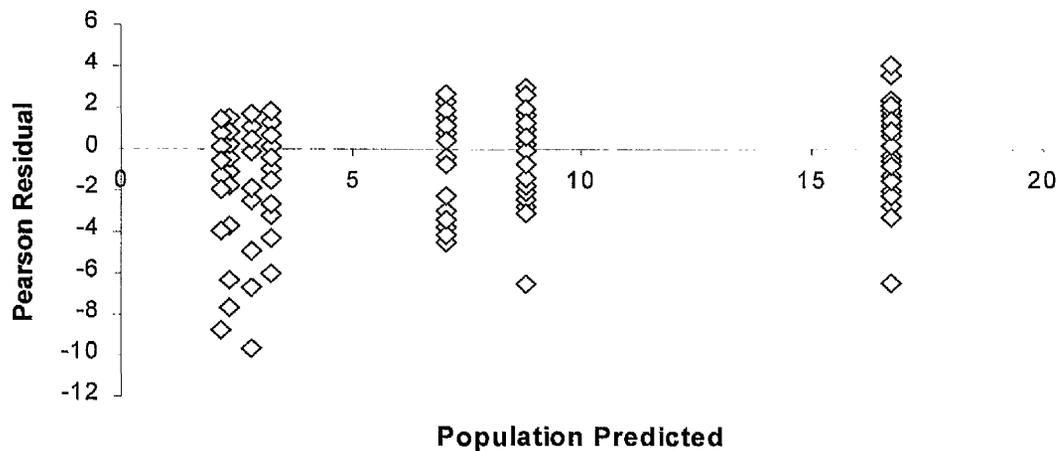
The results of Gupta *et al* (1999) were based on assuming a linear dose effect on the log scale. They showed that there was a significant difference between the active treatment intercepts but no difference between the rate of the dose effect. This was probably due to the use of a different model to that presented here and will be discussed later.

6.4 Adverse Effects Data Analysis

The pharmacodynamic measure used for the adverse effects of oxybutynin was dry mouth. This was measured on a 4 point ordinal categorical scale where 0 = no dry mouth, 1 = mild dry mouth, 2 = moderate dry mouth and 3 = severe dry mouth. As with other categorical data previously analysed, the proportional odds model was used.

6.4.1 Placebo Model

As with the efficacy data of oxybutynin, a placebo model needed to be considered first as the data between time and dose were confounded. Since categorical data has less information in the data compared to count data, models that were tried were necessarily less sophisticated. The number of counts in each category at each time is given in table 6.16.

Table 6.16. *Counts of observations in each category at each time.*

Time(days)	Category			
	0	1	2	3
7	12	4	0	0
14	11	4	1	0
21	11	4	1	0
28	10	4	1	0
35	11	3	1	0
42	12	2	1	0
49	11	3	1	0
Total	78	24	6	0

As can be seen in table 6.14, the counts stay reasonably constant over time. The first model to be considered for the adverse effect data was a simple linear model in time.

250

The full model specification is given in equation (6.12) where π is the individual category probability and 1 is the number of samples for the particular (i,j)th response.

$$W_{ij} \sim Multi(\pi_{0ij}, \pi_{1ij}, \pi_{2ij}, \pi_{3ij}, 1)$$
$$\log it(\Pr(W_{ij} \leq k \mid b_i)) = \theta_k + \beta_1 time_{ij} + b_i, k = 0,1,2$$
$$\theta_0 \sim N(3, 1 \times 10^5), \theta_1 \sim N(11, 1 \times 10^5), \theta_2 \sim N(800, 1 \times 10^5) \qquad (6.12)$$
$$\beta_1 \sim N(0, 1 \times 10^5), b_i \sim N(0, \sigma^2), \frac{1}{\sigma^2} \sim G(0.001, 0.001)$$

The results of this model are given in table 6.17.

Table 6.17. *Results of linear time adverse effects placebo model.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_0	2.823	3.303	3.075	(-5.735,8.190)
θ_1	13.6	5.71	12.63	(5.158,26.94)
θ_2	1153	737.5	1056	(75.23,2771)
β_1	0.0036	0.0247	0.00463	(-0.0447,0.0520)
σ	8.435	4.113	7.508	(3.244,19.14)

It can be seen from these results that the proportional odds model with linear time did not show any time effect (95% credible interval included zero). It was not expected therefore that any other placebo model would do any better. This was expected because if the linear time effect model did not show a significant effect then more complicated models which would require more parameters to be estimated were unlikely to show any effect. Based on the results in table 6.17, no placebo model was included when modelling the active treatment group data.

6.4.2 Active Treatment Groups

The number of counts in each adverse effect category at each time for OROS and IR oxybutynin is given tables 6.18a and b respectively. The number of counts in each

category at each dose for OROS and IR oxybutynin are given in tables 6.19a and b respectively.

Table 6.18a. *Counts of observations in each category at each time for OROS oxybutynin.*

Time(days)	Category			
	0	1	2	3
7	24	9	1	0
14	20	11	3	0
21	17	14	3	0
28	11	14	6	3
35	12	13	7	2
42	8	10	10	5
49	6	13	9	4
Total	98	84	39	14

Table 6.18b. *Counts of observations in each category at each time for IR oxybutynin.*

Time(days)	Category			
	0	1	2	3
7	27	4	0	1
14	13	16	1	1
21	13	16	1	1
28	6	16	9	1
35	7	17	4	3
42	2	15	8	6
49	4	15	7	4
Total	72	99	30	17

Table 6.19a. *Counts of observations in each category at each dose for OROS oxybutynin.*

Dose(mg)	Category				
	0	1	2	3	Total
0	24	9	1	0	34
5	37	25	6	0	68
10	23	27	13	5	68
15	14	23	19	9	65
Total	98	84	39	14	235

Table 6.19b. *Counts of observations in each category at each dose for IR oxybutynin.*

Dose(mg)	Category				
	0	1	2	3	Total
0	27	4	0	2	33
5	26	32	2	1	61
10	13	33	13	4	63
15	6	30	15	10	61
Total	72	99	30	17	218

The first model considered for the active treatment group was a proportional odds model in terms of linear time, dose and dose×time interaction. The full model specification is given in equation (6.13).

$$W_{ij} \sim Multi(\pi_{0ij}, \pi_{1ij}, \pi_{2ij}, \pi_{3ij}, 1)$$

$$\log it(\Pr(W_{ij} \leq k \mid b_i)) =$$

$$trt_i(\theta_{OROSk} + \beta_{OROS1}dose_i + \beta_{OROS2}time_{ij} + \beta_{OROS3}dose_i time_{ij} + b_{OROSi})$$

$$+ (1 - trt_i)(\theta_{IRk} + \beta_{IR1}dose_i + \beta_{IR2}time_{ij} + \beta_{IR3}dose_i time_{ij} + b_{IRi}), k = 0,1,2$$

$$\theta_{OROS0} \sim N(1.7, 1 \times 10^5), \theta_{OROS1} \sim N(4.8, 1 \times 10^5), \theta_{OROS2} \sim N(7.3, 1 \times 10^5)$$

$$\beta_{OROS1} \sim N(-0.3, 1 \times 10^5), \beta_{OROS2} \sim N(0, 1 \times 10^5), \beta_{OROS3} \sim N(0, 1 \times 10^5)$$

$$b_{OROSi} \sim N(0, \sigma^2_{OROS}), \frac{1}{\sigma^2_{OROS}} \sim G(0.001, 0.001)$$

$$\theta_{IR0} \sim N(1.2, 1 \times 10^5), \theta_{IR1} \sim N(4.3, 1 \times 10^5), \theta_{IR2} \sim N(6.2, 1 \times 10^5)$$

$$\beta_{IRS1} \sim N(-0.4, 1 \times 10^5), \beta_{IR2} \sim N(0, 1 \times 10^5), \beta_{IR3} \sim N(0, 1 \times 10^5)$$

$$b_{IRi} \sim N(0, \sigma^2_{IR}), \frac{1}{\sigma^2_{IR}} \sim G(0.001, 0.001)$$

(6.13)

The results obtained from this model are given in table 6.20.

253

Table 6.20. *Results of proportional odds model in terms of linear time, dose and dose×time interaction.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_{OROS0}	2.585	0.913	2.632	(0.949,4.314)
θ_{OROS1}	5.736	1.015	5.759	(3.888,7.662)
θ_{OROS2}	8.204	1.101	8.188	(6.153,10.42)
β_{OROS1}	-0.335	0.147	-0.347	(-0.592,-0.0284)
β_{OROS2}	-0.0456	0.0543	-0.0504	(-0.141,0.0524)
β_{OROS3}	0.00340	0.00290	0.00376	(-0.00272,0.00793)
σ_{OROS}	2.459	0.450	2.407	(1.693,3.437)
θ_{IR0}	2.215	0.629	2.163	(1.072,3.552)
θ_{IR1}	5.410	0.757	5.361	(4.064,6.989)
θ_{IR2}	7.265	0.848	7.186	(5.799,8.992)
β_{IR1}	-0.510	0.138	-0.517	(-0.779,-0.256)
β_{IR2}	-0.0129	0.0523	-0.00288	(-0.124,0.0807)
β_{IR3}	0.00513	0.00221	0.00498	(0.00103,0.00939)
σ_{IR}	0.973	0.250	0.954	(0.549,1.518)

The interaction term for the OROS data was not significant whereas it was for the IR data. The model was rerun without the interaction term in the OROS group. The results for the OROS model are given in table 6.21.

Table 6.21. *Results of linear dose+time model for OROS treatment group.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_{OROS0}	1.807	0.576	1.753	(0.745,0.2825)
θ_{OROS1}	4.888	0.726	4.852	(3.628,6.269)
θ_{OROS2}	7.337	0.873	7.28	(5.83,9.023)
β_{OROS1}	-0.245	0.0834	-0.253	(-0.400,-0.0859)
β_{OROS2}	-0.0111	0.0291	-0.0111	(-0.0655,0.0474)
σ_{OROS}	2.385	0.473	2.343	(1.609,3.412)

The time parameter estimate 95% credible interval includes zero so the model was re-analysed without the time effect. The results are given in table 6.22.

Table 6.22. *Results of linear dose model for OROS treatment group.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_{OROS0}	1.757	0.524	1.728	(0.804,2.866)
θ_{OROS1}	4.877	0.626	4.865	(3.686,6.118)
θ_{OROS2}	7.355	0.745	7.343	(5.963,8.849)
β_{OROS1}	-0.278	0.0371	-0.278	(-0.353,-0.207)
σ_{OROS}	2.428	0.453	2.378	(1.691,3.516)

The dose parameter estimate was significant so the dose parameter remained in the model. The IR data was also run with just the dose term included and the results are given in table 6.23.

Table 6.23. *Results of linear dose proportional odds model for IR dose group.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_{IR0}	1.361	0.342	1.357	(0.691,2.038)
θ_{IR1}	4.48	0.501	4.470	(3.539,5.523)
θ_{IR2}	6.337	0627	6.320	(5.176,7.628)
β_{IR1}	-0.273	0.0349	-0.272	(-0.343,-0.207)
σ_{IR}	0.902	0.469	0.900	(0.381,1.426)

Figures 6.15 and 6.16 are plots of the linear dose proportional odds models for the OROS and IR treatment groups respectively.

Figure 6.15. *Plot of linear dose proportional odds model for OROS dose group (results in table 6.22).*
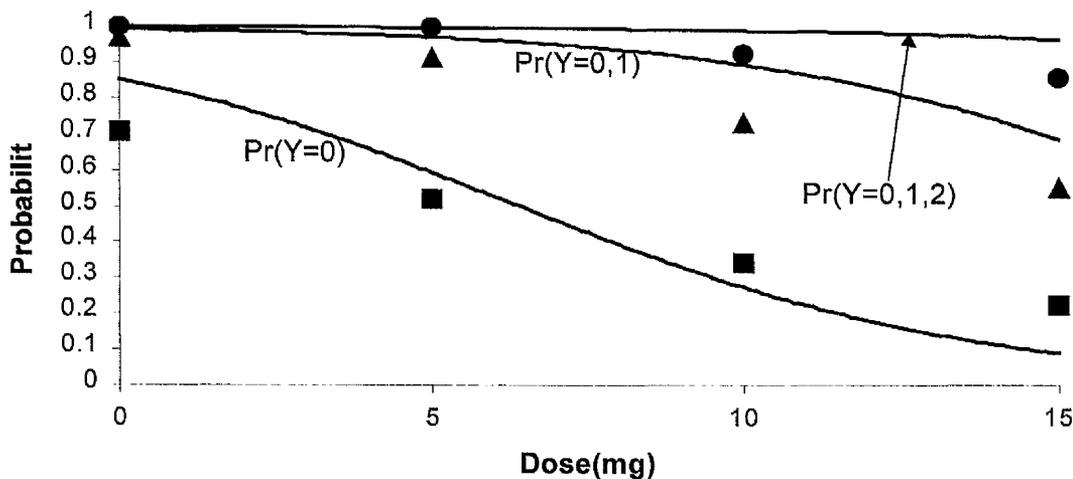
Figure 6.16. *Plot of linear dose proportional odds model for IR dose group (results in table 6.23).*
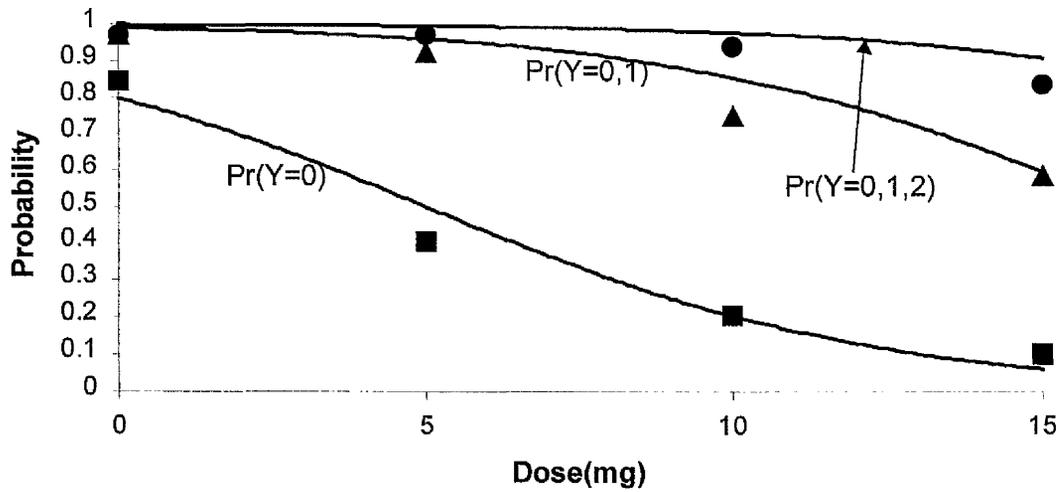
To see if there was a significant difference between the dose effects in the OROS and IR dose groups, a model of the form specified in equation (6.14) was examined.

$$\log it(E(\Pr(W_{ij} \le k \mid b_i))) = trt_i(\theta_{OROSk} + \beta_{OROS1}dose_i + b_{OROSi})$$
$$+ (1 - trt_i)(\theta_{IRk} + (\beta_{OROS1} + \beta_{IR1-OROS1})dose_i + b_{IRi}), k = 0,1,2 \tag{6.14}$$

The results for this model are given in table 6.24.

Table 6.24. *Results of linear dose comparison model.*

	Mean	Standard Deviation	Median	95% Credible Interval
θ_{OROS0}	1.735	0.544	1.732	(0.679,2.841)
θ_{OROS1}	4.854	0.647	4.842	(3.602,6.149)
θ_{OROS2}	7.405	0.780	7.393	(5.918,8.955)
β_{OROS1}	-0.273	0.0350	-0.273	(-0.342,-0.202)
σ_{OROS}	2.487	0.466	2.437	(1.735,3.528)
θ_{IR0}	1.347	0.338	1.338	(0.695,2.043)
θ_{IR1}	4.458	0.490	4.446	(3.540,5.474)
θ_{IR2}	6.303	0.613	6.271	(5.162,7.551)
$\beta_{IR1-OROS1}$	7.431×10^{-4}	0.0491	0.00165	(-0.0969,0.0968)
σ_{IR}	0.898	0.0249	0.890	(0.430,1.404)

The 95% credible interval corresponding to the parameter for the difference in the dose effects included zero so there was not a significant difference between the toxicity of the two formulations.

Gupta *et al* (1999) showed similar results as the work presented here. They showed that a placebo model was not needed to be accounted for and there was not a significant difference between the dose effects in the adverse effects data.

6.5 Decision Analysis

Decision analysis is concerned with the problem of making decisions. These decisions are made in the presence of information through data, knowledge through previous analyses and *a priori* beliefs and preferences to different decisions that could be made. The statistical analysis in work carried out in this thesis was usually geared towards describing and comparing data in the hope of defining a model for future use in prediction of new patients' outcomes or comparing new dosage regimens for example. The modelling process was ended when it was considered that a suitable model had been found. Being able then to go on and use the model for some practical purpose is of prime importance in clinical studies. A general framework such as decision analysis is described in detail by Berger (1985) and Bernardo and Smith (1994) and is well placed for the types of decisions that are made before, during and after the implementation of clinical trials. Decision problems can be framed in terms of frequentist or Bayesian ideas, but here the problem will be in terms of Bayesian decision theory.

The purpose of this section is to define an optimal dose for the use of once daily oxybutynin chloride based on the efficacy and toxicity data studied previously. The optimal dose was to be defined based on the Bayesian analysis of the data and clinicians' beliefs on what would be suitable levels of efficacy and toxicity for oxybutynin. These beliefs would then be described in terms of a mathematical function which could then be included into the Bayesian analysis to produce an answer to the question of an optimal dose.

6.5.1 Decision Analysis Framework

The basic elements of a decision problem were outlined briefly above. This section is described in more detail in Berger (1985) and is a summary of the elements involved in making decisions.

The unknown quantity which affects the decision (or action) that might be taken is commonly called the state of nature and is denoted by ϕ and the set of all possible states of nature is given by Φ. In making decisions it is clearly important to consider what the possible states of nature are. Of course, ϕ can be multivariate and in the work undertaken here, the possible states of nature are the parameters of the models for the efficacy and toxicity data.

Decisions are commonly called actions in the literature, Whitehead and Brunier (1995) for example. Particular actions are denoted by a while the set of all possible actions are denoted by A. In the oxybutynin study, the set of actions could be the decision regarding which formulation of oxybutynin to give a patient.

A key element of decision theory is the utility (gain) function or sometimes referred to in terms of loss (negative gain) (Wakefield (1994)). If a particular action a_l is taken and ϕ_l turns out to be the true state of nature then a utility of $U_l(\phi_l, a_l)$ will be observed. Thus, a utility function is defined for all $(\phi, a) \in \Phi \times A$. The loss function in the oxybutynin data set is defining the relative desirability of the efficacy and toxicity responses.

When a statistical investigation is performed to obtain information about ϕ, the outcome (a random variable) will be denoted by Y. Often, Y will be a vector, as when $Y = (Y_1, Y_2, \ldots, Y_n)$, the Y_i's being independent observations from a common distribution. The probability distribution of Y will depend on the unknown parameter ϕ. The distribution of Y conditional on ϕ is denoted by $P(Y|\phi)$ (which is the likelihood).

The last piece of information to be included in the decision problem is the prior information concerning ϕ. The symbol $\xi(\phi)$ will be used to represent a prior density on the parameters ϕ.

As the decision problem is usually formulated in terms of uncertainties, it is usual to work with expected utilities. Hence the actual incurred utility will never be known with certainty (at the time of decision making). The Bayesian expected utility is defined in equation (6.15).

$$\rho(\xi, a) = E(U_l(\phi, a)) \propto \int_\Phi U_l(\phi, a) P(y|\phi) \xi(\phi) d\phi \tag{6.15}$$

This expectation is being evaluated with respect to the posterior distribution and therefore equation (6.15) can be re-expressed as equation (6.16).

$$\rho(\xi, a) = E(U_l(\phi, a)) = \int_\Phi U_l(\phi, a) P(\phi|y) d\phi \tag{6.16}$$

The expected utility is maximised over the posterior distribution so that both uncertainty in the random variable Y and the parameters, ϕ can be taken into account. If there are n possible actions that can be taken, then the action that maximises the expected utility (ρ^{max}) is chosen as defined in equation (6.17).

$$\rho^{max} = \max_i(\rho(\phi, a_i)), i = 1,...,n \qquad (6.17)$$

6.5.2 Utility Function for Oxybutynin

The utility function for the oxybutynin data was defined in two parts: one part for the efficacy data and one part for the adverse effects measure. The utility functions had to be defined in a certain way so as to bring both variables onto the same scale. The efficacy data was measured on a non-negative integer scale and the adverse effects were measured on a categorical scale. The expected response was a non-negative continuous variable for the efficacy data and a probability for the toxicity data. The way in which the clinicians defined the desired levels of efficacy and toxicity was in terms of a population mean response. The required level of efficacy and toxicity defined by the clinicians was the following:

Efficacy

"The number of urinary urge incontinence episodes in a week must be less than or equal to 6."

Toxicity

"The probability of having severe dry mouth must be less than 0.1 and the probability of moderate or severe dry mouth be less than 0.7."

Another aspect of the utility function that could have been specified was the 'weighting' of the toxicity and efficacy measures. This weighting would determine how much

importance was assigned to the different aspects of the pharmacodynamics of oxybutynin. If the toxicity were given more weight then this would imply that toxicity is more important in determining the optimal dose, whereas if efficacy were given a greater weight then the converse would be true. As the clinicians did not make any specification about whether efficacy or toxicity was more important, they were both given equal weighting.

The defined level of desired efficacy was such that there be less than 7 urinary urge incontinence episodes a week. It was assumed for the efficacy utility function, there were on average, 20 incontinence episodes a week. This value, if changed would have an impact on the final optimal dose but the sensitivity on this value was not carried out. The range of the efficacy utility function was between 0 and 1. When the number of incontinence episodes is greater than 6, the efficacy utility function returns a value of 0. When the number of urge incontinence episodes was 0 then the utility function returned 1. When the number of incontinence episodes are between 0 and 6 then the utility function increases from 0 to 1 as the number of episodes decreases from 6 to 0. The utility function is defined in equation (6.18) and exemplified in figure 6.17.

$$S_i(\underline{\theta}) = (1 - I(20(\theta_{1i} - \theta_{2i}(1 - e^{-\theta_{3i}t}))e^{-\theta_{4i}dose} - 6))\left(6 - \frac{20(\theta_{1i} - \theta_{2i}(1 - e^{-\theta_{3i}t}))e^{-\theta_{4i}dose}}{6}\right)$$

$$I(A) = \begin{cases} 1, A \geq 0 \\ 0, A < 0 \end{cases}$$

(6.18)

**Figure 6.17.** *Example plot of efficacy utility function.*
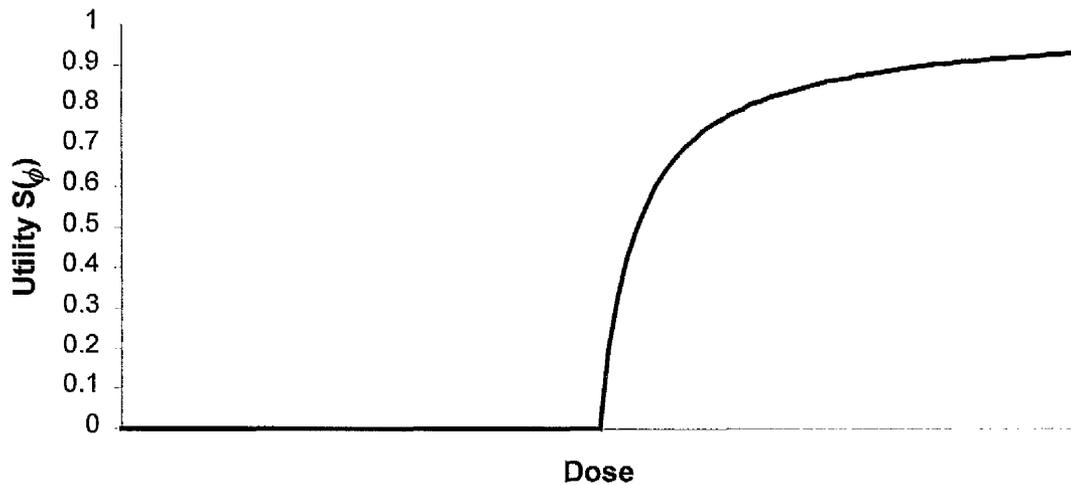
A step function could have been defined but it was thought more appropriate to use a smooth function as the number of incontinence episodes decrease from 6.

The defined level of desired toxicity was based on two probabilities: the probability of having severe dry mouth being less than 0.1 and the probability of having severe or moderate dry mouth being less than 0.7. The range of the toxicity utility function was from −1 to 0. The utility function was 0 when the probability of being in category 3 was less than 0.1 and the probability of being in category 2 and 3 was less than 0.7. If the probability of being in categories 2 and 3 is less than 0.7 and the probability of being category 3 is greater than 0.1 then the utility is defined by the logistic curve $\Pr(Y \leq 1)$. When the probability of being in categories 2 and 3 is greater than 0.7 and the probability of being in category 3 is greater than 0.1, then the utility function returns −1. The utility function is defined in equation (6.19) and is graphically exemplified in figure 6.18.

$$T_i(\gamma) = I(\Pr(Y_i \leq 2) - 0.9) + (1 - I(\Pr(Y_i \leq 2) - 0.9))) \Pr(Y_i \leq 1) I(\Pr(Y_i \leq 1) - 0.3) - 1$$

(6.19)

Figure 6.18. *Example plot of efficacy utility function.*
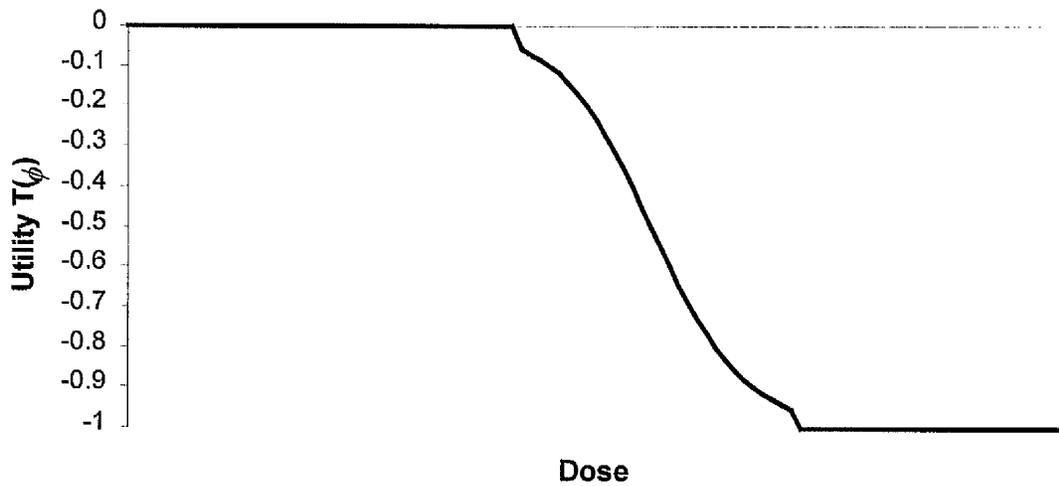
The combined utility function was simply set to the sum of the efficacy and toxicity utility functions. A small arbitrary amount was subtracted from the combined utility function to account for hypersensitivity (Rowland and Tozer (1995)). The combined utility function is defined in equation (6.20).

$$U_i = S_i + T_i - 0.05 \qquad\qquad (6.20)$$

To determine the expected posterior utility, the efficacy and toxicity models needed to be rerun in the same BUGS code so that parameter uncertainty in the population parameter estimates could be accounted for in the utility function. Linear dose effect models were used for both the efficacy and toxicity data.

The results of the efficacy, toxicity and combined utility functions are shown in figures 6.19-6.21 respectively.

Figure 6.19. *Results of efficacy utility function.*
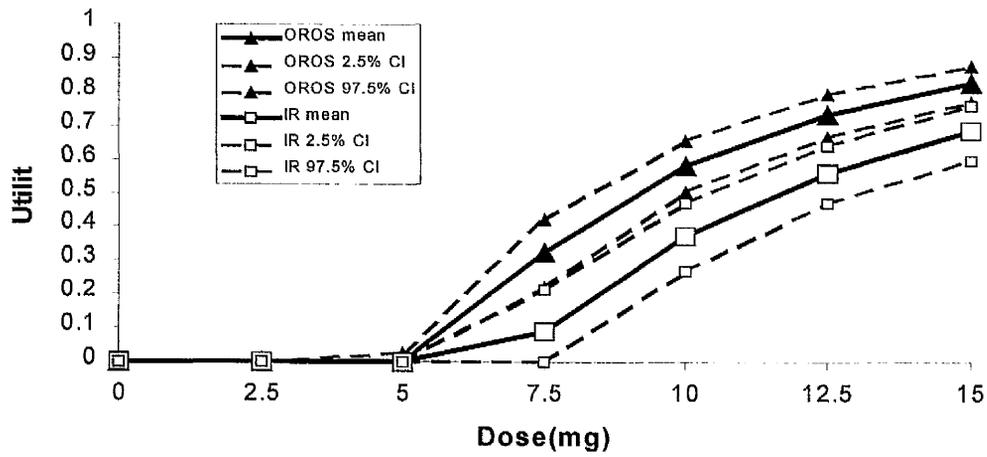
**Figure 6.20**. *Results of toxicity utility function.*
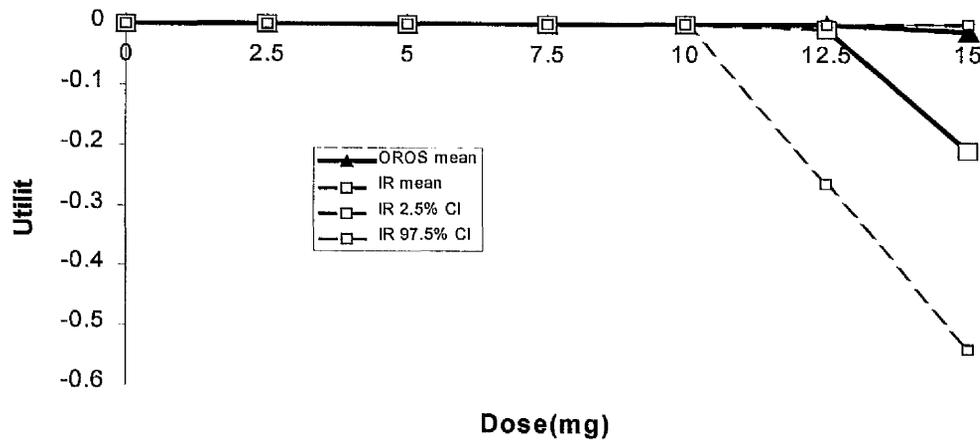
Figure 6.21. *Results of combined utility function.*

264
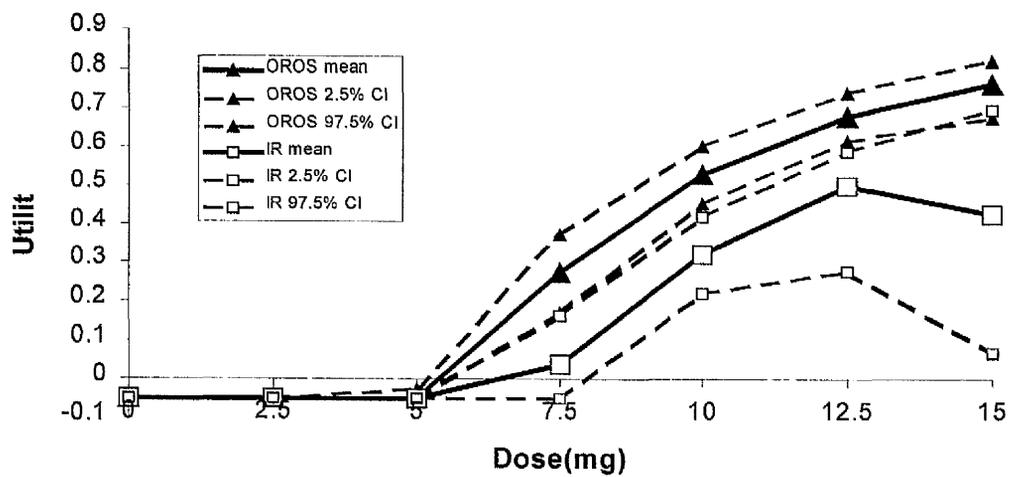
From figure 6.19, the utility for OROS was greater than the utility for the IR formulation of oxybutynin, once dose levels had exceeded 5 mg. There was very little difference between the toxicity utility of the two formulations except after approximately 12.5 mg when the utility of IR was lower but the 95% credible interval was very wide and included that of the OROS formulation. The combined utility of efficacy and toxicity shows that most of the combined result comes from the efficacy outcome and toxicity does not really start to have any impact until the top range of the doses were considered. Although there was not a significant difference between dose effects in the toxicity proportional odds models, there appears to be some down turn in the IR combined utility function before that of OROS. This was certainly the case but the 95% credible interval gets wider at this point and shows a lot of variability. The optimal dose of IR oxybutynin appears to be at approximately 12.5 mg and for the OROS oxybutynin, somewhere just above 15 mg.

6.6 Discussion

The study design was a forced dose escalation study. This type of design is not as frequently met in the analysis of clinical trials as the parallel dose group design but is useful for the study of efficacy and toxicity of a drug. Although dose and time as predictors for the pharmacodynamics of the drug are restricted to a certain degree by the study design, it was still possible to determine dose and time effects by the active treatment group data alone. The placebo data allowed the estimation of a more complicated placebo model but the active treatment data allowed the estimation of linear time trends only. This was because while dose remained the same for 2 weeks at

a time, the time component changed and there were two longitudinal measurements within a particular dose for a particular individual. This was the same for both the efficacy and toxicity data. A linear time component could have been tested in the efficacy data but because a more descriptive model was derived from the placebo data, it was not considered necessary. Other models for the placebo data could have been considered such as the E_{max} model but only the model in equation (6.4) was considered.

The efficacy placebo model was able to describe the data well with the inclusion of the placebo run-in episode count as a covariate in the model. This was a good choice of model because it allowed easy application to the active treatment groups as all that was required was a weekly count of the number of urinary urge incontinence episodes. The model chosen by Gupta *et al* (1999) is defined in equation (6.21) where λ is the mean number of incontinence episodes.

$$\log(\lambda) = \alpha + e^{-\beta time} \tag{6.21}$$

This model does not include a subject specific covariate for the baseline count and so any individual estimates need to be estimated with the use of subject specific random effects. This model also only requires two population parameters which was an improvement on the model used in this work. The model used previously could be simplified by assuming one of the parameters to be fixed (β_2=1) so as to make the placebo model go through the baseline count and the other parameter (β_2) being an offset to estimate the asymptote as time increases. The model used here could easily be implemented in the NONMEM analysis to obtain a reasonable estimate of the individual data. In the NONMEM analysis, a random effect was assigned to the time effect parameter which allowed for different rates of decrease (or increase) in incontinence counts between individuals to be estimated but in the BUGS analysis, only

the population parameter was estimated. This was due to the use of the Griddy Gibbs algorithm in BUGS which requires the use of a truncated distribution to be discretised into a histogram and so a uniform distribution was selected. Individual estimates could have been obtained but this would have required a slightly different Bayesian model specification. Such a model would be estimated easily in WinBUGS as there is a Metropolis algorithm step implemented. This would allow both individual and population parameter estimates on this parameter to be estimated.

The above analysis of the active treatment group data showed different results to that of the analysis by Gupta *et al* (1999). The IR dose group had the same linear dose trend on the log scale as the NONMEM analysis but the OROS data showed a cubic relationship between incontinence counts and dose in the BUGS analysis. It was not reported whether other dose models were considered for the active treatment group efficacy data in the NONMEM analysis. Although a cubic dose model was found for OROS, the cubic term was positive which when extrapolated to higher doses would lead to an incorrect dose-effect relationship. In the NONMEM final model, the dose parameter estimate for OROS and IR was -0.144 as it was assumed that both treatment groups had the same dose effect but different intercepts compared to -0.145 and -0.105 in the BUGS analysis. Whereas the BUGS analysis showed there was a difference between the dose effect of the two formulations, the NONMEM analysis did not but did show a significant difference between the intercept terms. The difference in results could also be due to the different placebo models used having an effect on the estimation of the dose parameters. In both analyses, it was shown that there was a higher number of incontinence episodes in the IR treatment group as compared to the OROS treatment group.

Other efficacy models could have been considered for these data. One important class of models that could have been studied were those that included autoregressive components. Zeger (1988) reported an approach to modelling a time series of counts where the correlation was assumed to arise from an unobservable process added to the linear predictor in a log linear model. Other autoregressive type models have been reported by Azzalini (1994), Cessie and van Houwelingen (1994) and Fahrmeir and Kaufman (1987) which have been applied to binary and categorical data. Stiratelli *et al* (1984) took a random effect approach to accounting for autoregression between the serially correlated data. Such models can be expressed as in equation (6.22).

$$g(E(Y_{ij})) = \underline{\beta}\underline{x}_{ij} + \gamma_1 Y_{ij-1} + \gamma_2 Y_{ij-2} + ... + \gamma_{j-1} Y_{i1} \qquad (6.22)$$

The first part of equation (6.22) is the standard covariate part of the model with regression parameter $\underline{\beta}$ and the second part of the model is made of previously observed response data giving the model a time series flavour. These models are usually only applicable when the spacing between serially observed data are equal as was the case for the oxybutynin data. It is possible to take into account unequal spacing as described by Jones and Boadi-Boateng (1991) but this was not necessary. Random effects models as described by Gupta *et al* (1999) could have been applied and even more complicated random effects structures but were not due to the difficulty of implementation in BUGS.

The placebo adverse effects data in both the BUGS and NOMEM analysis included no time effect. This does not imply that it is not possible to estimate a time effect in the active treatment groups as mentioned previously.

In the active treatment groups, again there was a difference in results between that from BUGS and NONMEM although this time, there was greater similarity in the types of

models used. For the IR data, the BUGS analysis showed that there was a statistically significant interaction between dose and time which was not reported in the NONMEM analysis. The OROS models were the same for the two different analyses. When a linear dose model was run in BUGS, there was no statistically significant difference between the IR and OROS formulations.

Like the efficacy data, the adverse data could have been modelled using a model with an autoregressive component. For categorical data, such models have a dynamic Markov chain feel as the probability of going from category to category can be modelled but the probabilities change due to the influence of covariates. The baseline category could have also been used as a covariate in the model although it was not tried and therefore not known whether such a model would improve the fit.

The analysis of the efficacy and toxicity data was carried out independently of each other. A more suitable analysis could have been to use a multivariate technique to take into account the correlation between the toxicity and efficacy. Work has been carried out on the analysis of bivariate toxicity and efficacy data by Tubert-Bitter *et al* (1995), Murtaugh and Fisher (1993), Thall and Russell (1998), Jennison and Turnball (1993) and Cook and Farewell (1994). In these studies, the theme was to look at determining an optimal dose and monitoring of efficacy and toxicity for a range of different response variables. A multivariate analysis of oxybutynin might allow the determination of more information from the data on how the efficacy and toxicity are related to pharmacokinetics and other covariate information. Often in clinical trials, there is more than one variable associated with the efficacy and toxicity of the drug so such problems are nearly always going to be seen in a multivariate form.

The decision analysis approach to determining the optimal dose for oxybutynin is a general way of setting up decision procedures in clinical trials. Such a method does not have to be performed in a Bayesian setting and such problems can be defined in a frequentist framework. Defining utility functions for outcomes such as efficacy and toxicity measures is a useful way of looking at clinical end points and determining safe and effective doses. It allows the different variables to be defined on a univariate scale without loss of information from the original data and include prior information through the posterior distribution. From these utility functions, it is possible to make decisions about certain aspects of the drug, such as dosing schedules, therapeutic windows and optimal doses. In the case of oxybutynin, the utility of the OROS formulation was consistently higher than that of IR over the dose range of 5-15 mg. Whereas the IR combined utility function had begun to turn down due to the toxicity of the drug (even though there was no significant difference in the toxicity of the two formulations), the OROS combined utility function was still rising as the toxicity had not reached a previously defined unacceptable level. The utility function was not tested for any sensitivity to any aspect of the model. The actual modelling stage of the analysis was considered to be reasonably stable as it gave similar answers as the NONMEM approach. It can be shown that changing the specification of the utility function would have altered the optimal dose but not the choice of formulation.

7. D-Optimal Design for Ordinal Categorical Data

7.1 Introduction to Optimal Design

Optimal design for categorical response data has received little attention in the statistical and pharmaceutical literature. What work that has been published has been associated with logistic regression for binary or binomial data. The ways in which these problems have been approached has been to use a range of optimality criteria. These criteria have included minimising fiducial intervals, D-optimality, sequential designs, Bayesian designs and constant information measures. Quite often the studies were results of simulations and applications to toxicology examples. With this body of work, the results for optimal design in logistic regression have become well established. Although this is a start to work on optimal designs for categorical response data, there are still many outstanding areas of research for categorical data optimal design, such as categorical mixed effects optimal design but areas such as logistic mixed effects models at a more basic level have also not been tackled.

The purpose of this chapter is to extend the ideas for deriving optimal designs for binary data to where we have more than two categories in the response variable. The work will be restricted to the three category model and to the one independent variable case for ease of understanding and computation. The model investigated was the one used throughout the thesis: the proportional odds model. This model is a generalisation of logistic regression for dichotomous data. This work was originally motivated from the need for more efficient clinical trial designs, especially in Phase II/III clinical trials and

even in pre-clinical studies where categorical data are common. However, initially the fixed effects proportional odds model case was investigated.

7.2 Optimal Design Theory

Optimal design and experimental design dates back to the work of Fisher at the Rothamsted field station. The general idea of optimal design is to design an experiment/data collection procedure where there is optimal efficiency in collecting the data for extracting specific required information. The required information could be to estimate a parameter of interest, prediction at a different predictor level, to test the difference between two means or determine whether a model is linear or quadratic for example. The way in which these requirements are met is by the specification of different optimality criteria. There are many optimality criteria such as A, C, D and G optimality, with D optimality being the most frequently encountered. A optimality minimises the sum of the variances of the parameter estimates, C-optimality minimises the variance of a linear combination of parameters, D optimality minimises the variability associated with parameter estimation and G-optimality minimises the maximum over the design region of the standardised variance (Atkinson and Donev (1992)). These are only a few examples of optimality criteria which require the model to be known before the optimal design is derived for further data collection. For linear models with normally distributed errors, most of the different design criteria produce the same design (which corresponds to taking observations at the extremes of the domain region). Therefore parameter values do not need to be known for the linear model. When the model is nonlinear, the optimal design (for example D-optimal design)

is dependent on parameter values. This is because the variance matrix which is often approximated by the negative inverse of the Fisher information matrix is a function of the model parameters. The paradox of having to know what the parameter values are before they are estimated from the optimal design means that D-optimal designs are impossible to find in practice. As parameter estimation is the main interest in D-optimality, assuming that the parameters are known before the experiment is designed might not seem the most appropriate thing to do. To get round the problem of not knowing the parameter values *a priori*, as is usually the case, a measure of uncertainty can be assigned to the parameters. This can be made possible by the specification of a distribution on the parameters. This would give a measure of location and dispersion of the parameter and allow parameters uncertainty to be incorporated into the design problem. This type of approach to designing experiments is usually referred to as Bayesian optimal design (Chaloner (1984), Firth and Hinde (1997), Chaloner and Larntz (1989) and Merle and Mentre (1995)). Another approach to initial parameter specification is to use a sequential design where the design of the data collecting procedure is split into sections (Spears, Brown and Atkinson (1997)). In this method, data collected at the first stage is used to design the next stage so as to try and make the data collecting procedure as efficient as possible. In this chapter, the work will concentrate on the application of D-optimal designs to ordinal categorical data as commonly found in pharmacodynamic studies.

The definition of D-optimality is to minimise (maximise) the determinant of the variance-covariance (information) matrix of the parameter estimates.

$$\Psi(M(\xi)) = \min_{\xi}\{\det(M^{-1}(\xi))\} \qquad (7.1)$$

273

where Ψ is the optimality criterium, ξ is the design space, M is the Fisher information matrix and det is the determinant of the matrix. The variance-covariance matrix is determined by the (expected) negative inverse of the Fisher information matrix. The Fisher information matrix is defined as the Hessian matrix of the likelihood with respect to the parameters to be estimated.

$$M(\xi) = E_{\underline{Y}}\left(\frac{\partial^2 L(\underline{y} \mid \underline{\theta},\xi)}{\partial \theta_i \, \partial \theta_j}\right) \qquad i,j = 1,...,p \qquad (7.2)$$

where $E_{\underline{Y}}$ is the expectation over the random variable $\underline{Y}$, $\underline{y}$ is the data, L is the likelihood function and $\underline{\theta}$ is the p×1 parameter vector.

7.3 Application to Pharmacokinetic/Pharmacodynamic Studies

Optimal design has not been a widely used method for designing pharmacokinetic and pharmacodynamic studies. The theoretical basis has been applied in the literature but not carried out in practice to any extent. One of the first examples of the design of a pharmacokinetic study is D'Argenio (1981) who used D-optimal design and approximated the variance-covariance matrix by a first-order Taylor series approximation to the normal likelihood. In the pharmacokinetic setting, the goal is to determine the times when drug concentration measurements should be made. Other works have looked at the design of pharmacokinetic experiments. D'Argenio (1990) and Merle and Mentre (1995) examined D-optimal design with parameter uncertainty in fixed effects regression. Tod and Rocchisani (1997) examined D-optimality and variants of D-optimality for fixed effects nonlinear pharmacokinetic models. Mentre *et al* (1997) considered optimal design in a mixed effects setting. Jonsson *et al* (1996)

looked at sparse sampling under varying conditions in practical settings. These approaches have greatly improved knowledge of efficient sampling schedules but more work needs to be carried in a population setting. Some work has been carried out on continuous measure pharmacodynamic models, for example Merle and Mentre (1995). The same ideas used for designing pharmacokinetic experiments can be applied to continuous measure pharmacodynamic studies. Very little has been published specifically for the optimal design of pharmacodynamic studies. As well as continuous measure pharmacodynamic studies, there are of course non-continuous and non-normally distributed data which need to be considered.

7.4 Optimal Design for Binary Data

Finney (1971) applied the ideas of optimal design to binary and binomial data in his work on probit analysis. Other statisticians took up the idea throughout the 1970s and 80s during which the basic methods were developed. The first paper of any real substance was that by Abdelbasit and Plackett (1983). Abdelbasit and Plackett looked at a range of optimality criteria for dichotomous data, such as fiducial intervals, D-optimality, sequential designs and constant information designs. Minkin (1987) then extended and made more rigorous, the work on D-optimal design and made a distinction between local and global D-optimal designs. Chaloner and Larntz (1989) then extended the work to include the situation where the parameter values are not known *a priori* and a prior distribution is assigned to the parameters to formalise the information that is available on the parameters. Then Spears *et al* (1997) generalised the Bayesian design to include multi-stage designs and made available a FORTRAN program.

The context in which this research was undertaken was for the efficient estimation of the parameters of a logistic regression model. Other forms of optimal design were not considered, such as prediction to new ranges of the independent variable or model specification. In many situations where an experiment is designed to gain some important piece of information, the goal is estimating an important parameter. The research described above usually concentrated on models where the dichotomous response was a linear function of one independent variable (for example, dose in the context of a toxicology problem). This is the simplest situation as the functional form is simple and there is only one independent variable.

Consider the situation where we have data (x_i, y_i) for $i = 1, \ldots, n$. The x_i's are the individual values of the independent variable x. It is assumed that this variable is standardised to make it symmetrically distributed around zero. The y_i's are the individual values of the independent variable. It is assumed in this analysis that y is dichotomous and has been assigned values 0 or 1. Binary data of this kind are assumed to be from a Bernoulli distribution. The likelihood is given in equation (7.3).

$$L(\underline{y} \mid \underline{\pi}) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{(1 - y_i)} \tag{7.3}$$

The parameter π_i is the expected probability of 'success' where this is taken to be probability of observing $y_i = 1$. The log-likelihood is given in equation (7.4).

$$l(\underline{y} \mid \underline{\pi}) = \log(L(\underline{y} \mid \underline{\pi})) = \sum_{i=1}^{n} (y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i)) \tag{7.4}$$

Logistic regression relates the probability of observing a response in a particular category given that a particular level of the independent variable x was observed. The model can be written in the form of equation (7.5).

$$\pi_i = \Pr(Y_i = 1 \mid x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \tag{7.5}$$

With binary data there is a variety of link functions such as the logit, probit and complementary log-log link functions. In the following work, the logit link function will be used as it is computationally easier to work with and because it is also the natural link function for binary data. The model can be written in a linear form by transforming equation (7.5) with the logit link function.

$$\log it(\theta_i) = \log it(\Pr(Y_i = 1 \mid x_i)) = \beta_0 + \beta_1 x_i \tag{7.6}$$

An optimal design for this type of model is achieved by minimising the variability in estimating the parameters β_0 and β_1. This allows the parameters to be estimated as efficiently as possible. This minimisation is carried out on the determinant of the asymptotic variance-covariance matrix, which is proportional to the area contained within the confidence ellipsoid for (β_0, β_1). For the model shown in equation (7.6), the asymptotic variance-covariance matrix (expectation of the Hessian matrix of the likelihood with respect to the model parameters) is defined in equation (7.7).

$$Var(\beta_j, \beta_k) = \begin{bmatrix} \sum_{i=1}^{n} \pi_i(1-\pi_i) & \sum_{i=1}^{n} x_i \pi_i(1-\pi_i) \\ \sum_{i=1}^{n} x_i \pi_i(1-\pi_i) & \sum_{i=1}^{n} x_i^2 \pi_i(1-\pi_i) \end{bmatrix}^{-1} \quad j,k = 0,1 \tag{7.7}$$

The difficulty in finding the optimal design for this problem (which is the problem with all nonlinear models) is that the matrix is a function of the parameters for which the optimal design is being sought. In this situation, an initial estimate is used in place of the parameter as though it were known. This could be derived from a pilot study for the purpose of determining the initial values. Another way of determining the initial values is using published results or using some other estimate. The problem of using an exact initial value is that the optimal design might be sensitive to departures from the initial

277

values and the optimality of the design might decay very rapidly as the initial guess moves away from the true value. Another possibility is to specify a distribution on the parameter values, as this is more appropriate when the parameter values are not known exactly *a priori*. In the present context, exact starting values will be used although the problems associated with this are recognised. The standard method of D-optimality is to take the determinant of the asymptotic variance-covariance matrix as given in equation (7.7). Once this is done, then the problem becomes a matter of minimising the variance-covariance matrix (i.e. minimising the variability associated with estimating the parameters). How this is minimised is a matter of choice. A variety of minimisation algorithms could be used but for simplicity, the simplex algorithm has been used (Nelder and Mead (1965)), as it does not require the use of derivatives which are needed for directional search algorithms. The simplex works on the basis of a simplex/polytope searching over the surface of the function in the form of contractions, expansions and reflections until a minimum is obtained.

The results that were obtained were for the simple case of equation (7.5), the case where there is a binary outcome with one independent variable. There are two optimal design points that are symmetric around the point corresponding to probability 0.5. If the initial values are taken to be $(\beta_0, \beta_1) = (0,1)$ then the optimal design points are (-1.543,1.543). These two points correspond to the probabilities, (0.178,0.824) which can be seen in figure 7.1. The design points seem to be intuitively sensible although the exact points would not be easy to guess. The points lie on the logistic curve where there is the most 'informative' change in the direction of the curve. If the design points were in the tails then there would be no information concerning where the curve starts to have a steeper gradient. If the design points were closer to the centre of the curve then again

it might be too difficult to determine how steep the curve is because there is more uncertainty to what the response is in this region.

Figure 7.1. *Plot of design points for logistic regression model with $(\beta_0,\beta_1)=(0,1)$.*
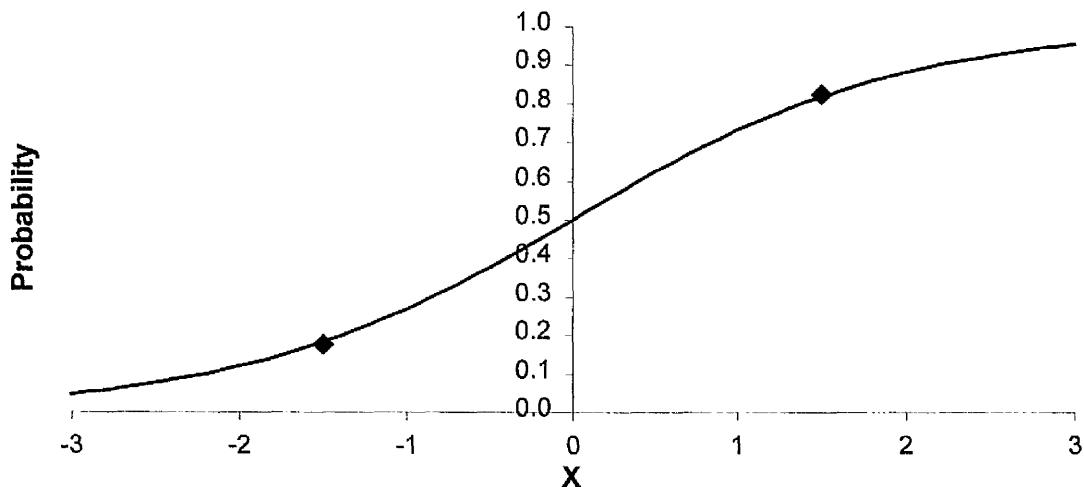
A general formula can be derived for determining where the design points are given the initial starting values. This is given in equation (7.8).

$$(x_1, x_2) = ((1.5434 - \beta_0)/\beta_1, (-1.5434 - \beta_0)/\beta_1) \qquad (7.8)$$

Another important aspect of optimal design is to determine the number of observations at each design point. Assume that it is required to take N observations, how many of these should be assigned to the design point x_1 and how many to x_2? According to Minkin (1987), this is simply a case of dividing equally the observations between x_1 and x_2 for even N whereas for odd N, an extra observation should taken from either design point. This result contradicted that of Abdelbasit and Plackett (1983) that there is a three point optimal design for odd N. Current work using the simplex method of minimising the determinant has shown Minkin to be correct.

An important aspect of optimal design, in particular D-optimal design is to study how well the design performs when the initial parameter values are a distance from the true values. This is usually measured in terms of the ratio of the determinant of the incorrect design to the correct design raised to the inverse power of the number of parameters in the model as defined in equation (7.9) where p is the number of parameters.

$$D_{eff} = \left\{ \frac{M(\xi_{estimate})}{M(\xi_{true})} \right\}^{\frac{1}{p}}$$

(7.9)

D_{eff} is called the relative efficiency. The relative efficiency for the logistic model is given in table 7.1 and is defined in terms of the ED50 (α/β) and scale parameter β as reported by Abdelbasit and Plackett (1983). The initial estimates are given by (α,β) and the true values by (α_0,β_0). Table 7.1 is specified in terms of the ratio of the true gradient to the initial estimate and the difference between the true and initial ED50 scaled by the initial gradient parameter. The efficiencies are given in this form because it gives a symmetric approach to looking at the divergence from the true values. From table 7.1, there is greater departure from optimality when the gradient is overestimated than when it is underestimated. As the difference in the ED50 increases, then there is a greater loss in the relative efficiency.

Table 7.1. *Relative efficiencies for logistic regression.*

$\beta(\alpha_0/\beta_0-\alpha/\beta)$	β_0/β						
	0.8	0.9	0.95	1.0	1.05	1.1	1.2
2.5	20.0	18.1	17.1	16.1	15.2	14.3	12.6
2.0	35.1	33.4	32.2	30.8	29.4	28.0	25.3
1.5	53.9	53.8	52.9	51.6	50.0	48.3	44.6
1.0	72.6	75.4	75.4	74.6	73.4	71.7	67.7
0.5	86.1	91.9	92.9	93.0	92.3	91.0	87.3
0.0	91.1	98.0	99.5	100.0	99.6	98.5	95.0

7.5 Optimal Design of the Proportional Odds Model - 3 Categories

Optimal designs for the three category proportional odds model have not been reported in the statistical or pharmaceutical literature. This could firstly be due to the relatively recent introduction of the proportional odds model for the analysis of categorical response data although it is commonly encountered now in many areas. The model was originally described by McCullagh (1980). The name of the proportional odds model is from the property of the model that the log odds ratio difference at x_j and x_k is independent of the category that has been observed. Changing notation slightly from the previous section, the model is defined in equation (7.10).

$$\text{logit}(\Pr(Y \le m \mid x)) = \alpha_m + \beta_1 x \qquad (7.10)$$

Y is the categorical response variable defined on the K categories $(1,2,\ldots,K)$. The index m can take a value from the K-1 categories $(1,\ldots,K$-1$)$. The predictor variable is x and β_1 is the gradient parameter. The parameters α_m are the so called 'cut points' or intercepts. These represent the baseline logit values when x=0. Equation (7.10) can be written in the untransformed form.

$$\Pr(Y \le m \mid x) = \frac{\exp(\alpha_m + \beta_1 x)}{1 + \exp(\alpha_m + \beta_1 x)} \qquad (7.11)$$

It can be seen that the parameterisation of the equation gives a cumulative probability of the form $\Pr(Y \le m) = \Pr(Y = \bigcup_{k=1}^{m} y_k)$. It is easy to revert back to individual category probabilities by subtracting the relevant cumulative probabilities.

$$\Pr(Y = j) = \Pr(Y \le j) - \Pr(Y \le j - 1) \qquad (7.12)$$

Another part of the problem that needs to be generalised is the distribution of the data. The standard distribution for categorical data is the multinomial distribution and this is usually of the form given in equation (7.13).

$$L(Y \mid \Pi) = \frac{n!}{y_{1i}! y_{2i}! \ldots y_{mi}!} \pi_{1i}^{y_{1i}} \pi_{2i}^{y_{2i}} \ldots \pi_{mi}^{y_{mi}} \qquad (7.13)$$

n is the total number of observations such that $n = y_1 + y_2 + \ldots + y_m$. The π_j's are the

individual category probabilities such that $\pi_1 + \pi_2 + \ldots \pi_m = 1$. This parameterisation is in

the form of individual category probabilities but the proportional odds model is in the

form of cumulative probabilities so the likelihood can be reparameterised into the form

of cumulative probabilities. Using McCullagh's parameterisation, we have

$$L(R \mid \Gamma) = \frac{n!}{r_{1i}! r_{2i}! \ldots r_{mi}!} \left\{ \frac{\gamma_{1i}}{\gamma_{2i}} \right\}^{r_{1i}} \left\{ \frac{\gamma_{2i} - \gamma_{1i}}{\gamma_{2i}} \right\}^{r_{2i} - r_{1i}} \ldots \left\{ \frac{\gamma_{(m-1)i}}{\gamma_{mi}} \right\}^{r_{(m-1)i}} \left\{ \frac{\gamma_{mi} - \gamma_{(m-1)i}}{\gamma_{mi}} \right\}^{r_{mi} - r_{(m-1)i}}$$

$$(7.14)$$

where $r_j = y_1 + \ldots + y_j$ and $\gamma_j = \pi_1 + \ldots + \pi_j$, R is the random variable of cumulative scores

and Γ is the set of cumulative probabilities. As the proportional odds model

corresponds to this parameterisation, then in the likelihood, the cumulative probability

can be replaced by the probability model in equation (7.11). It is with this information

that the asymptotic variance-covariance matrix can be determined by differentiating

with respect to the model parameters twice and inverting the negative of the matrix.

The general formulae for these derivatives for any number of categories are given in an

appendix at the end of this chapter.

The optimality criterion to be used is D-optimality. The determinant is proportional to

the area enclosed by the parameter ellipsoid and the smaller this area, the more accurate

are the estimates. D-optimality is only one way of finding an optimal design but it is

probably one of the easier ones to implement and so was used for the current problem.

D-optimality has been used for logistic regression problems published by several

different authors (Khan and Yazdi (1988), Sitter (1992), Heise and Myers (1996), Minkin (1987) and Abdelbasit and Plackett (1983)).

7.6 Proportional Odds Model Results

The results for the proportional odds model where there are only two categories to choose from were presented in an earlier section and so will not be included here. The results for the three category case are presented in this section and conform to standard results on D-optimality, that is, the minimum number of distinct optimal design points in the optimal design is the same as the number parameters in the model. To check that this is indeed the case, simulations were carried out. These simulations were performed using a double precision program written in Fortran 77 compiled on a Salford Fortran 90 16 bit compiler.

The first simulations were designed to investigate how the design points would change as the initial estimates of the parameters were changed. To make the scenario easier to understand, only symmetric designs were considered. The symmetrical initial parameter assumption implied that the design itself would be symmetric due the symmetry of the model. Also minimisation was limited to return three design points only. As a starting reference point for the initial values, the following initial values will be referred to throughout: $(\alpha_1,\alpha_2,\beta)=(-1,1,1)$. The cut points are symmetric around zero and the gradient is unity. Figure 7.2 shows a plot of the proportional odds model with design points plotted on the curves. The optimal design points are symmetrically distributed around the probability (y) axis as expected, at (-1.866, 0,1.866). It can be

argued from the graph, that these design points are not unreasonable. It must be noted that if the data had been dichotomised as though it were intended to fit two simple linear logistic regression models, then this design would not be the most efficient and would require four optimal design points corresponding to the two intercept and gradient terms given by the formula for the binary case in equation (7.8). The current optimal design is in some way a trade-off between having an optimal design for the non-proportional odds model case and having the constraint on the gradient parameter so as to make both gradients equal. The optimal design can be presented in a slightly different way by plotting the optimal design points as individual category probabilities as in figure 7.3. Again the optimal design points look reasonable even in a slightly different setting, and of course the design points were not determined based on individual category probabilities.

Figure 7.2. *Plot of model and D-optimal design points for 3 category model with* $(\alpha_1,\alpha_2,\beta)=(-1,1,1)$.
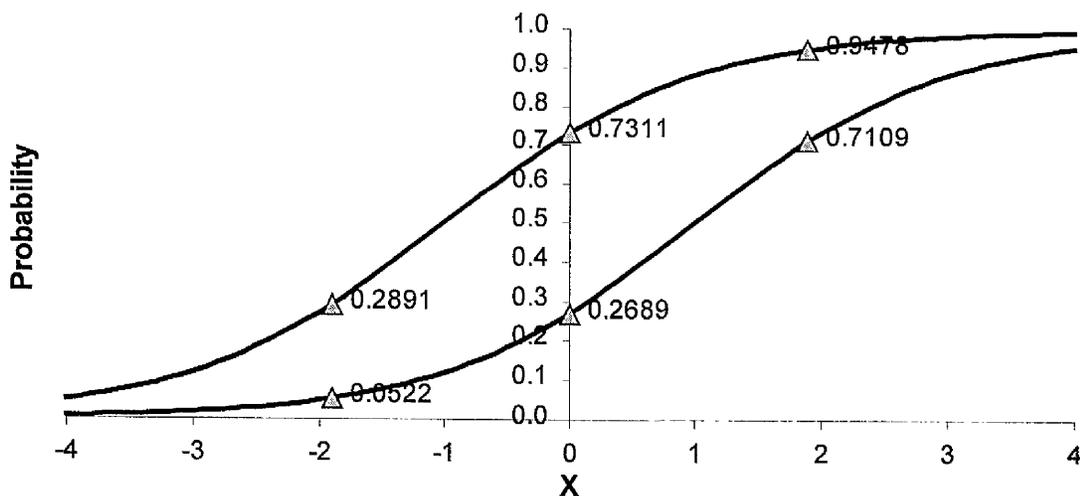
Figure 7.3. *Plot of model in terms of individual category probabilities and D-optimal design points for 3 category model with $(\alpha_1, \alpha_2, \beta) = (-1, 1, 1)$.*
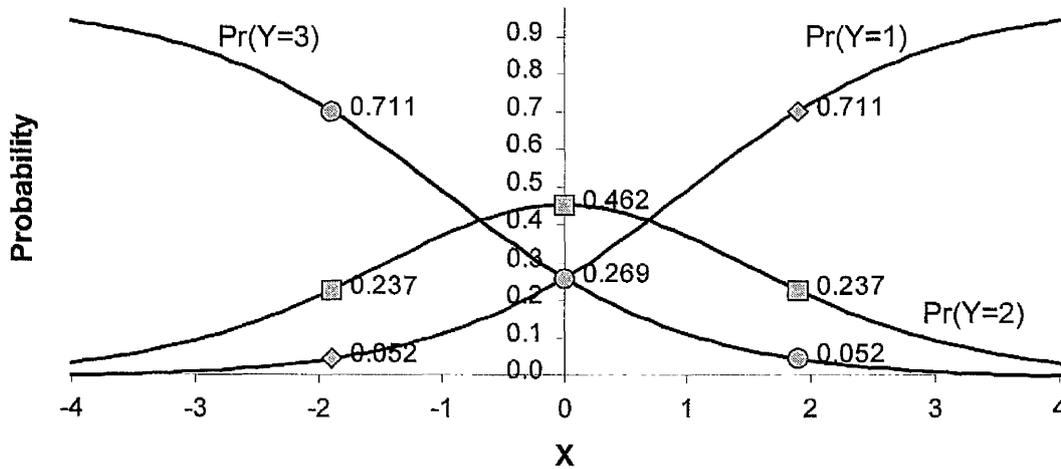
The next step was to systematically change the cut points and the gradient parameter and to investigate what effect this had on the optimal design points. Firstly, the gradient term was held fixed and the cut points were allowed to vary. To keep symmetry, α_2 was made equal to $-\alpha_1$. With this scenario, it is known that the design points are symmetric and any plot requires only that the symmetric optimal design points be plotted against α_1. This is shown in figure 7.4. A relationship is shown between the optimum design point and the cut point. As the cut point converges to zero, i.e. as the two proportional odds lines converge towards each other, so the design point converges to the situation where there are only two categories in the response variable. This is inferred by the fact that as the lines converge together, then the probability of being in the middle category tends to zero leaving only two categories with non-zero probabilities. As the cut points increase in magnitude, the design points also increase in magnitude and asymptote to the line of identity.

Figure 7.4. *Plot of design point versus α (with β held fixed).*
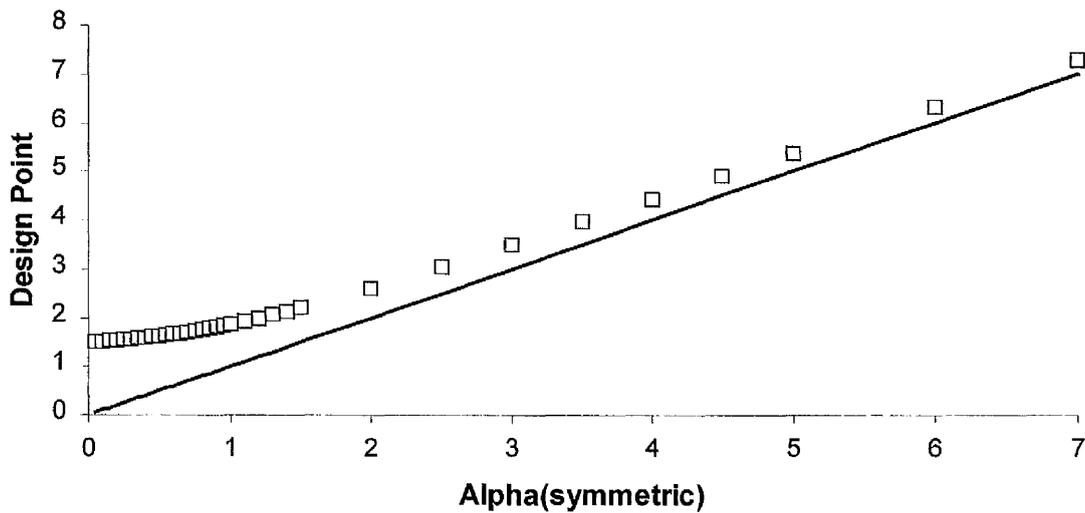
The next step was to see how the optimum design points altered as the gradient was changed and the cut points remained fixed. The results of these simulations are shown in figure 7.5. When the gradient was small in magnitude, the design points become further apart which was as expected as the logistic curve becomes flatter then there is less information in the middle section of the design space. As the gradient increases in magnitude, then according to the simulations, the optimal design points converged to zero. This was also expected since as the gradient becomes steeper, the steepest part of the curve becomes closer to the probability (y) axis.

Figure 7.5. *Plot of design point versus β (with α help fixed).*
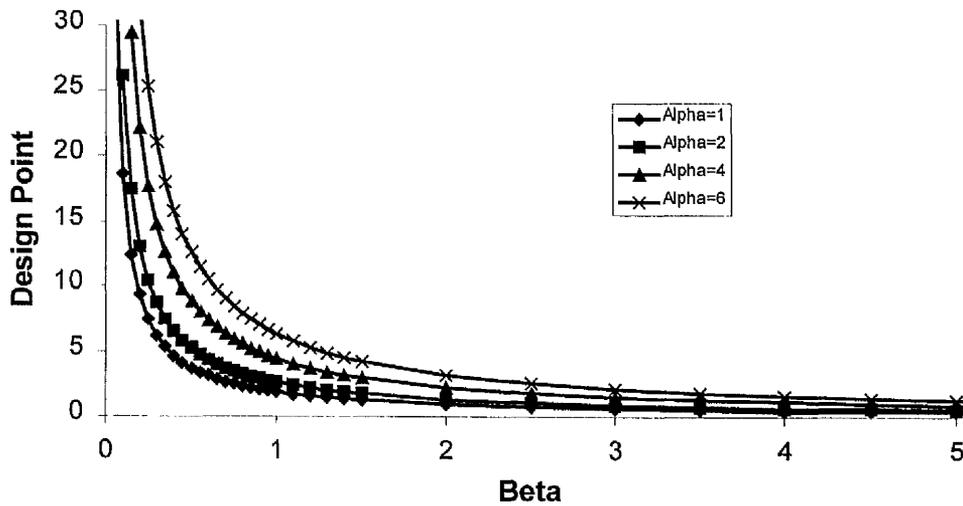
Another important aspect of the optimal design apart from the position of the design points is the allocation of the number of observations to the distinct design points, i.e. assuming it is required to take *n* observations, how do these observations get allocated to the independent variables? To determine the number of allocated observations at each design point, further simulations were carried where the number of design points were varied. The results of the simulations are given in Table 7.2.

Table 7.2. *Results of number of observations and design points.*

Number of observations (*n*)	Design Points(replications)			
	DP1	DP2	DP3	DP4
3	-1.866	0	1.866	-
4	-1.718	-1.185	1.185	1.718
5	-1.685(2)	0	1.685(2)	-
6	-1.866(2)	0(2)	1.866(2)	-
7	-1.615(3)	0	1.615(3)	-
8	-1.750(3)	0(2)	1.750(3)	-
9	-1.866(3)	0(3)	1.866(3)	-
10	-1.687(4)	0(2)	1.687(4)	-

The first row corresponds to the same number of design points as parameters required by the model (standard result of D-optimality). Ignoring the case where $n=4$ for the moment, all the cases have three distinct design points. When n is a multiple of three, then the design points are the same as those for the case when $n=3$ but there is an equal number of replicates at each distinct design point. When the number of points is not a multiple of three then the number of replicates and the position of the design points are not the same as when n is a multiple of 3. In this case, the replicates are arranged in a symmetrical form so that there are more replicates on the outside design points. For example, when $n=5$, there are 2 replicates on the outside design points and 1 replicate of the middle point and when $n=7$, there are 3 replicates on the outside points and 1 in the middle. It seems that there will not be a difference of more than two replicates between the outside points and the middle point. Along with this feature are the positions of the distinct design points. When there are more replicates on the outside points than when there is an equal number of replicates at each design point, the position of the points are pulled towards the middle, for example, when $n=7$, the distinct design points are (-1.615, 0,1.615) and for $n=5$, the design points are (-1.685,0,1.685). It appears that as the ratio of the middle design point number of replicates compared to the number of replicates on an outside design point diverges from 1:1, then the positioning of the points tends towards the middle. This can be seen in the table. 7.3.

Table 7.3. *Position of outside design point compared to ratio of number of replicates of middle design point to outside design point.*

	Ratio of middle point to outer point			
	1:1	2:3	1:2	1:3
Design Point	1.866	1.75	1.685	1.615

This leads to the description of why there are 4 distinct optimum design points when the minimisation was searching on only 4 observations. It seems as though there is a requirement that there be at least three distinct design points. If the problem can not be solved with three distinct design points and symmetric weighting of the replicates, then a solution is to have 4 distinct design points symmetric around 0. The reason for there not being a similar situation when 8 design points are required is because a 3 distinct point design can be derived with symmetric weighting of the replicates in this case and in any other case with a multiple of 4 design points.

As with binary data, the efficiency of the design as the chosen design points vary from the true design points for the parameter set needs to be considered. The parameter values chosen as the reference parameter set were $(\alpha_1,\alpha_2,\beta)=(-1,1,1)$ which has optimal design points $(-1.866,0,1.866)$. The efficiencies are determined by equation (7.9). The results correspond to a combination of design points of DP1= (-4.866,-3.866,-2.866, -1.866,-0.866), DP2=(-0.5,0,0.5) and DP3=(0.866,1.866,2.866,3.866,4.866). The efficiencies are given in tables 7.4-7.6 corresponding to DP2=0, DP2=0.5 and DP2=-0.5.

Table 7.4. *Efficiencies(%) of proportional odds model with middle design point DP2=0.*

DP3		0.866	1.866	2.866	3.866	4.866
DP1	-4.866	72.1	74.7	70.7	63.4	56.1
	-3.866	84.3	85.2	79.5	71.1	63.4
	-2.866	93.8	95.3	88.9	79.5	70.7
	-1.866	94.1	100	95.3	85.2	74.7
	-0.866	78.7	94.1	93.8	84.3	72.1

Table 7.5. *Efficiencies(%) of proportional odds model with middle design point DP2=0.5.*

DP3		0.866	1.866	2.866	3.866	4.866
DP1	-4.866	69.4	69.5	66.2	60.4	54.7
	-3.866	83.5	81.7	76.4	69.5	60.4
	-2.866	94.2	94.4	87.4	76.4	66.2
	-1.866	96.0	99.2	94.4	81.7	69.5
	-0.866	80.5	96.0	94.2	83.5	69.4

Table 7.6. *Efficiencies(%) of proportional odds model with middle design point DP2= -0.5.*

DP3		0.866	1.866	2.866	3.866	4.866
DP1	-4.866	75.0	77.8	72.2	63.3	54.7
	-3.866	85.0	86.4	79.4	69.5	63.3
	-2.866	92.7	94.9	87.4	79.4	72.2
	-1.866	92.4	99.2	94.9	86.4	77.8
	-0.866	80.5	92.4	92.7	85.0	75.0

From these tables, the most striking feature is that it is worse to over estimate the middle design point than to underestimate it. As the design points become wider (increase in magnitude) then the efficiency falls away.

7.7 Discussion

The results obtained here are for the three category proportional odds model and show that the results of D-optimality hold. For the three category proportional odds model, there are three distinct design points, as there are three parameters in the model. The design points are a compromise for the case where two logistic curves are estimated independently of each other giving four parameters and design points.

The number of distinct design points remains at three except for the case where four design points are required. Unlike binary data, where the position of the design points does not depend on the number of observations required, for the three category model, there is a dependence between the two variables. It appears that the position of the design points depends on the ratio of the number of replicates on the outside design point to the number of replicates on the middle design point which was not the case for binary data. Based on this, it might be possible to find a relationship between the position of the design points based on the number of replicates at the middle and end points. For the case of four observations, this was unique because the number of replicates could not be divided equally among three distinct design points with there being at least as many replicates on the outside design point as there is on the middle design point.

These results need to be expanded by considering models with more than three categories. There is no need to consider too many categories (certainly no more than 10) as continuous approximations come into play. To expand the results, 4 and 5 category models could be considered and see whether the results of D-optimality still hold, such as four and five distinct design points for four and five parameter models. Also, whether there is any relation between the number of observations and the position and the number of distinct design points.

There are many potential directions for future work in this area. At the moment the current work needs to be completed and clarified with mathematical and statistical rigour.

As mentioned above the studies need to be extended to problems where there are more than three categories. This would lead to a body of work that was appropriate for any number of categories including the binary case where there is one independent variable. Technically, this would be no more difficult than the current work except it would become more complicated due to the higher dimensionality of the problem. Another direction that could be considered is to include more independent variables in the proportional odds model. Again this would not be any more technically challenging but would be more complex due to the increasing size of the problem.

To bring these two parts together, would require the derivation of an explicit formula for the calculation of the design points from the initial parameter estimates as is the case for linear logistic regression. This would give a definitive answer to the position of the optimum design points for any number of categories and independent variables.

The main push of this work is in the area of population pharmacokinetic/ pharmacodynamic modelling. As already stated earlier, categorical data is commonly encountered for pharmacodynamic responses. Designing pharmacokinetic/ pharmacodynamic studies is an important aspect of clinical trial design and the ability to obtain relevant and useful information is not left to chance. When data is being collected from many individuals and in a longitudinal manner, then it is important that the optimal number of data points in each individual and over all individuals is collected, in order to make the data analysis/population modelling as efficient as possible. This leads to the efficient design of studies for mixed effects modelling. This is generally where work is leading. Corresponding to this is the need to incorporate parameter uncertainty in the design stage. This comes under the area of Bayesian

optimal design and has been widely studied for binary data. This would lead to new difficulties of minimising functions derived from the design criteria as integrating over the prior distribution will generally be analytically intractable.

Finally, there is the need for optimal design of simultaneous pharmacokinetic/ pharmacodynamic studies. This would be a procedure where the data collection in these studies were designed for the purpose of modelling the pharmacokinetic and pharmacodynamic data together rather than sequentially as is usually the case

Appendix A7.1

This appendix gives the Hessian matrix and the Fisher Information matrix (negative expectation of the Hessian) for the three category proportional odds models.

Observed second derivatives

$$\frac{\partial^2 l}{\partial \alpha_1^2} = -\gamma_1(1-\gamma_1)(R_1 - \frac{(R_2-R_1)}{(\gamma_2-\gamma_1)}\gamma_1 + \gamma_2(1-\gamma_1)\frac{(R_2-R_1)}{(\gamma_2-\gamma_1)^2})$$

$$\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_2} = \frac{\partial^2 l}{\partial \alpha_2 \partial \alpha_1} = \gamma_1(1-\gamma_1)\gamma_2(1-\gamma_2)\frac{(R_2-R_1)}{(\gamma_2-\gamma_1)^2}$$

$$\frac{\partial^2 l}{\partial \alpha_1 \partial \beta} = \frac{\partial^2 l}{\partial \beta \partial \alpha_1} = -\gamma_1(1-\gamma_1)R_2 x$$

$$\frac{\partial^2 l}{\partial \alpha_2^2} = \gamma_2(1-\gamma_2)\{[\frac{(R_2-R_1)}{(\gamma_2-\gamma_1)} - \frac{(R_3-R_2)}{(\gamma_3-\gamma_2)}](1-2\gamma_2) - \gamma_2(1-\gamma_2)[\frac{(R_2-R_1)}{(\gamma_2-\gamma_1)^2} + \frac{(R_3-R_2)}{(\gamma_3-\gamma_2)^2}]\}$$

$$\frac{\partial^2 l}{\partial \alpha_2 \partial \beta} = \frac{\partial^2 l}{\partial \beta \partial \alpha_2} = -\gamma_2(1-\gamma_2)(1-R_1)x$$

$$\frac{\partial^2 l}{\partial \beta^2} = -\{\gamma_1(1-\gamma_1)R_2 - \gamma_2(1-\gamma_2)(R_3 - R_1)\}x^2$$

Negative expected second derivatives

$$-E(\frac{\partial^2 l}{\partial \alpha_1^2}) = \frac{\gamma_1(1-\gamma_2)(1-\gamma_1)^2}{(\gamma_2 - \gamma_1)}$$

$$-E(\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_2}) = -E(\frac{\partial^2 l}{\partial \alpha_2 \partial \alpha_1}) = -\frac{\gamma_1(1-\gamma_1)\gamma_2(1-\gamma_2)}{(\gamma_2 - \gamma_1)}$$

$$-E(\frac{\partial^2 l}{\partial \alpha_1 \partial \beta}) = -E(\frac{\partial^2 l}{\partial \beta \partial \alpha_1}) = \gamma_1(1-\gamma_1)\gamma_2 x$$

$$-E(\frac{\partial^2 l}{\partial \alpha_2^2}) = \gamma_2^2(1-\gamma_2)^2(\frac{1}{(\gamma_2 - \gamma_1)} + \frac{1}{(\gamma_3 - \gamma_2)})$$

$$-E(\frac{\partial^2 l}{\partial \alpha_2 \partial \beta}) = -E(\frac{\partial^2 l}{\partial \beta \partial \alpha_2}) = \gamma_2(1-\gamma_2)(1-\gamma_1)x$$

$$-E(\frac{\partial^2 l}{\partial \beta^2}) = \{\gamma_1(1-\gamma_1)\gamma_2 - \gamma_2(1-\gamma_2)(\gamma_3 - \gamma_1)\}x^2$$

Appendix A7.2

Fisher Information matrix for k category proportional odds model with 1 linearly modelled covariate.

Likelihood given by:

$$L(R \mid \Gamma) = \frac{n!}{R_{1i}! R_{2i}! ... R_{ki}!} \left\{\frac{\gamma_{1i}}{\gamma_{2i}}\right\}^{R_{1i}} \left\{\frac{\gamma_{2i} - \gamma_{1i}}{\gamma_{2i}}\right\}^{R_{2i} - R_{1i}} ... \left\{\frac{\gamma_{(k-1)i}}{\gamma_{ki}}\right\}^{R_{(k-1)i}} \left\{\frac{\gamma_{ki} - \gamma_{(k-1)1i}}{\gamma_{ki}}\right\}^{R_{ki} - R_{(k-1)i}}$$

where the notation is given as before.

The model is given by:

$$\log it(\Pr(Y \le m)) = \gamma_m = \theta_m + \beta_1 x_1$$

Fisher information matrix

$$-E(\frac{\partial^2 l}{\partial \alpha_m^2}) = \gamma_m^2 (1-\gamma_m)^2 (\frac{1}{(\gamma_m - \gamma_{m-1})} + \frac{1}{(\gamma_{m+1} - \gamma_m)})\ \ m=1,\ldots,k\text{-}1$$

$$-E(\frac{\partial^2 l}{\partial \alpha_m \partial \alpha_{m+1}}) = -\frac{\gamma_m (1-\gamma_m)\gamma_{m+1}(1-\gamma_{m+1})}{(\gamma_{m+1} - \gamma_m)}$$

$$-E(\frac{\partial^2 l}{\partial \alpha_m \partial \alpha_{m+n}}) = 0, n \ne 0,1\ ;\ m=1,\ldots,k\text{-}3$$

$$-E(\frac{\partial^2 l}{\partial \beta_1^2}) = x_1^2 (\sum_{m=1}^{k-1}\gamma_m (1-\gamma_m)(\gamma_{m+1} - \gamma_{m-1}))$$

$$-E(\frac{\partial^2 l}{\partial \alpha_m \partial \beta_1}) = x_1\gamma_m (1-\gamma_m)(\gamma_{m+1} - \gamma_{m-1})$$

8   Conclusions

The proportional odds model was used consistently throughout this thesis whenever it was required to model categorical pharmacodynamic data. It was shown by the analysis of several different data sets that the proportional odds model can be applied to a variety of different types of analysis. In most of the analyses, such as the pharmacokinetic/pharmacodynamic model in the sumatriptan analysis and the adverse effects model in the oxybutynin data set, the proportional odds model was applied in a longitudinal setting. This is going to be one of the most common applications of the proportional odds model in pharmacodynamics as it is often a requirement to study the time course of the pharmacological effect of the drug. The proportional odds model can accommodate time effects in a similar fashion to any other covariate. Toxicokinetic data set I did not include a longitudinal component (in the pharmacodynamics) as the data were collected at steady state which meant that the proportional odds model was implemented independent of time. This is a common situation in which the proportional odds model can be used to show differences between factors, such as gender differences or between dose groups. Nearly all the models considered were linear in terms of the covariate effects except for one case in which an E_{max} model was used (in toxicokinetic data set II) within the proportional odds model.

Although the proportional odds model can be applied to a variety of different study situations, there are aspects of the model that were restrictive. The main restriction encountered with the proportional odds model was the essential requirement of proportional odds. This requirement forces the model to have the same gradient

parameters for each cumulative category level while the cut points are the only category specific parameters. By allowing the gradient parameters to vary across categories permits more flexible models. One problem with this is that the logistic curves should not cross in the region of interest, as this would lead to illogical results such as negative probabilities. It can be seen in the plots of empirical cumulative probabilities, e.g. figure 6.15, that the odds are not always constant over the categories, but anecdotal evidence says that all that is required for proportional odds is that all the odds be less than or all greater than 1. Another problem with the use of the proportional odds model is the combination of covariates selected for the model. A particular model where this was a problem was the model involving linear time and dose. The critical problem was in interpreting the model, for example in the sumatriptan case where the drug was administered once. Assume that there is a significant dose and time effect. When dose is zero, then the model is dependent only on time and varies accordingly. When the dose is non-zero then the dose of the drug has an effect. The effect of the drug is a constant dose effect according to the model but the response varies over time and this is taken into account by the time parameter. The effect of the dose is instantaneous according to the model but this is not observed in the data where usually the response is changing with respect to time. On average, the same baseline categorical scores are observed at the time the drug is taken (time=0) regardless of the dose given but if the dose is non-zero then the proportional odds model predicts that an instantaneous effect is obtained. To get round this problem, a dose specific model could be implemented where the time effect parameter is estimated separately for each dose. This would probably result in the time parameter values increasing in magnitude as the dose increased, assuming that the effect of the drug increased with dose. Although this circumvents the problem of an instantaneous effect due to dose, it introduces the

problem of extrapolating to doses outside the range of those considered. This problem does not occur with concentration or AUC as these pharmacokinetic predictors of the pharmacodynamics are functions of time and dose, as at time zero, the AUC and concentration is predicted to be zero and increases accordingly with time and dose.

A major criticism of the models employed throughout the different data analyses in this thesis is that they are not mechanistic. It is often the case that the mechanism for how the drug attains an effect is not well understood, such as in the case of sumatriptan (Plosker and McTavish (1994)). Trying to develop a model that incorporates a physiological interpretation is difficult and when it is possible, often results in a complex model. This leads to difficulties in parameter estimation and for categorical pharmacodynamic data, experience shows that this would be difficult to accomplish. The models used have all been empirical and have been devised to describe the data as best as possible given the information available. Modelling categorical pharmacodynamic data as described in this thesis has been of interest over the last 5 years approximately. In all the studies reported, the models have been empirical, even if they have been given a pseudo physiological interpretation. The problem of including mechanistic models into categorical data analysis has not been considered seriously to any great extent. This needs to be studied in depth if categorical data is to be given the same level of treatment as other forms of pharmacodynamic data. As the models are empirical, it is difficult to know how well they are likely to extrapolate to different studies for the same drug and whether the models have any generality.

It is always of importance in any clinical trial to quantify and assess the variability. Interindividual variability is probably of greatest importance in pharmacokinetic/

pharmacodynamic data and so describing the interindividual variability in categorical data should also be one of the main investigations of the modelling process. All the models considered in this thesis were random effects models, but usually the random effect model had a simple structure. The most common way of describing the interindividual variability was in the form of an additive random effect on the cut points. This acted as a shift on the logit scale for the individual response to shift from the population mean response. This was often adequate at describing the variability between individuals but it is common to include random effects on the gradient parameters as well. In BUGS this often proved to be difficult to estimate for the categorical models, but in NONMEM, it seemed possible to estimate any interindividual variance structure required. However the standard errors often could not be estimated. How much confidence can be placed in such estimates is not known but it would probably be unwise to consider them seriously. The interpretation of the random effect estimates needs to be considered carefully as it is different to that of continuous models. The simplest way of interpreting the interindividual variability is to report it as it is estimated on the logit scale. The standard distributional assumption to make for the random effects structure is to use a normal distribution (which can be readily changed in BUGS) so to report the variability on the logit scale gives it the same flavour as that for normally distributed data. It might also be required to consider the variability on the probability scale. In the case of a dose-effect model with an additive random effect, the amount of variability between individuals on the probability scale will depend on the mean response. For example, table 6.23 gives the results of a dose-effect model for oxybutynin with the interindividual standard deviation estimated as 2.428. The probability of being in category 0 given a placebo dose $\pm$ 1 standard deviation is (-0.671,4.185) on the logit scale and (0.338,0.992) on the probability scale and for a

dose of 15 mg is (-4.841,0.015) on the logit scale and (0.008,0.504) on the probability scale. This gives a difference in probabilities of 0.654 and 0.496 for the 0 and 15 mg doses respectively. Interpretation of the variability is therefore of importance as the way in which the variability changes with mean response can have an effect on inferences made from the data.

As well as considering the interindividual variability, it is also of interest to consider the residual variability. The residual variability is not estimated in NONMEM and only possible sometimes in BUGS. It is not often considered in the analysis of categorical data to quantify the residual variability. Why this is the case is not known but should be included in the model if possible to account for the remaining variability. The interpretation of the residual variance component can be made in the same way as that for the interindividual variance component. It is known that including a subject specific random effect can cause a change in the population parameter estimates when compared to a fixed effects model. It has not been described whether including a residual component will do the same thing also. It is not likely to have a great effect if the subject specific components account for most of the variability in the data but may have an effect if the residual component is larger than the interindividual variability.

Checking the ability of a model to describe the data is an important step in the modelling process. Whether it is a case of choosing a selection of different models and then deciding on the best model via some selection criteria or starting with an initial model and building in or taking out different components, there are a number of ways to check the ability of a model to describe the data. A range of techniques were investigated such as the examination of residuals, comparing the deviance statistic,

Bayes factors and the NONMEM objective function. Residuals plots were informative in showing how well the model fitted the data but there were differences in the Pearson and deviance residuals for the same model. Within each type of residual, there was consistency in deciding which model was superior, for example in toxicokinetic data set I. Other residuals were not considered because of the difficulty in interpretation whereas the residuals that were used are more similar to the residuals used in continuous data analysis in terms of their interpretation. Summary statistics such as the deviance statistic are useful when comparing different models but do not give a picture of how each separate model fits the data. The same problem is encountered with the NONMEM objective function but incorrect residuals are given in the NONMEM output file as it is reported as the observed score minus the predicted probability of observing that score. It was considered that the best method of comparing models was by the use of Bayes factors. Although this is a method used in a Bayesian setting, it is a predictive method of comparing models and can therefore be used in a frequentist framework but requires a different interpretation due to the exclusion of prior information. Although the method was used in a pairwise manner in toxicokinetic data set I, it can be formulated to compare multiple models simultaneously. As it is a predictive method, models do not need to be nested and random effects models can be compared to fixed effects models as well as nonlinear models to linear models without having to change the way in which the comparison is made. As well as being applicable to categorical pharmacodynamic models, it could also be applied to pharmacokinetic models.

Two separate analyses in this thesis were from toxicokinetic studies, data set I and II. Each study had varying quality of results but in both cases, sophisticated models were out of the question. In data set I, there were only two dose levels that could be

modelled which meant that essentially, only a comparison of the dosing levels could be made. The modelling showed that the 200 mg/kg dose had a more toxic effect than the 30 mg/kg dose but it would have been appropriate to have the placebo dose group data to compare the treatment groups to the placebo group. The data were rich as there were between 19 and 25 observations per rat and so modelling the data was not a problem, although for the 30 mg/kg dose group there was only one rat with scores in any of the upper two categories. The main restriction in analysing the toxicokinetic studies was in the data that was available. In data set II, the preclinical trials were usually designed with good intentions, i.e. selecting doses that would (hopefully) give a no effect dose, a high toxicity dose and a median effective dose. The problem occurred when the data were collected and there were very few convulsions that could be used to develop a model. This meant that modelling most of the animal studies was difficult if not impossible. Pharmacokinetic modelling for data set II was more straight forward than modelling the pharmacodynamics but even this was not simple. There was considerable variability in the pharmacokinetic data which made obtaining a population model difficult as well as there being small amounts of data in some of the studies. Trying to model the pharmacokinetic and pharmacodynamic data in the same model from such studies did not prove informative in any way as usually the dose-effect models were as good if not better than the models including pharmacokinetic information through the AUC values. It was possible to use the pharmacokinetic information in data set I as there was sufficient data but data set II showed that by doing this was no more informative than using dose. Trying to model the pharmacodynamics when only a few pharmacodynamic events occur based on highly variable pharmacokinetic data was not a good combination for modelling. This brings into question the design of the studies and what data should be collected from animal studies. Certainly, in data set II, a whole

range of measurements were taken representing different parts of the animals and so any variable could have been chosen for the pharmacodynamic model. As this was a retrospective analysis, in that the study was not designed for the analysis carried out here, it begs the question, what was the data to be used for? If one of the goals was to gain information on the compound to help design phase I clinical trials, then some modelling or scaling must take place to help choose dose groups, sampling times, patient numbers and so on. If the preclinical data is not used in this way then it must be questioned whether the collection of so much animal data is needed for the assessment of toxicity only. The scaling of the pharmacokinetics proved to be effective between the three species but the scaling of the pharmacodynamics was non-existent. Most scaling studies are performed with pharmacokinetic variables but correlating the effect of the drug must be of interest as studying the toxicity of the drug in healthy humans is one of the aims of phase I clinical trials. The study of toxicokinetic data given in this thesis shows that if such studies are to be of use then more consideration needs to be given to what is required from the data, how the studies are to be designed and subsequently analysed and how the information is to be used.

Categorical data has already been considered in the literature in terms of population pharmacokinetic/pharmacodynamic analyses (Sheiner (1994), Mandema and Stanski (1996) and Sheiner *et al* (1997)). The approach taken in the literature was to assume that the models were known and had some form of physiological interpretation by including the E_{max} model as a component of the pharmacodynamics. For the analysis of sumatriptan, the models for both the pharmacokinetics and pharmacodynamics were empirical. The primary reason for this was that data was available for the first two hours only. It was known from previously reported clinical trials, that this time frame

corresponded to being before or at C_{max} and so no information was available on the disposition of sumatriptan. For this data set, it was possible to try several different models with different levels of random effects included in the proportional odds models. Interindividual variability was included in all the models considered but residual variability was unable to be estimated for the AUC and dose models. Estimates of the interindividual variability was approximately the same for each of the models but the residual variability when estimated for the dose and time and concentration models were different by a factor of two with the dose and time model having greater residual variability. Model checking with the deviance statistic showed that AUC was the best model in describing the data but the dose and time model was almost as good on this basis. This showed again that including the pharmacokinetic information does not necessarily produce a better fit to the categorical pharmacodynamic data. AUC was superior to concentration and so some form of exposure to the drug may be a better predictor of the drug effect than plasma concentration. For the sumatriptan data set, modelling complex relationships between the pharmacokinetics and categorical pharmacodynamic response measure was possible and some nice relationships were obtained but these relationships remained linear on the logit scale.

There were other interesting features that could have been considered in some of the data sets but were not, such as serial correlation. The models implemented in the data analyses of this thesis all used the assumption of independence from time to time within an individual. This assumption is probably not true in such situations. Serial correlation could have been taken into account for the toxicokinetic data set II, sumatriptan and oxybutynin data sets. It is reasonable to think that for a particular individual that the categories observed over time are longitudinally dependent and this

needs to be checked by the inclusion of an autoregressive component. Random effects models take into account the heterogeneity of a studied population but another way of correcting for unexplained variability is the inclusion of an overdispersion parameter. This takes into account the difference in the variability between the observed and model predicted variability and used to ensure that variance components are not underestimated.

NONMEM and BUGS were the two computer programs used for parameter estimation. NONMEM has been in development for approximately 15 years more than BUGS and so would be expected to be at a more advanced level of development. Although this is the case, BUGS has been extremely useful in the estimation of a wide range of models. Both NONMEM and (Win)BUGS can estimate nonlinear mixed effects models, but NONMEM has built in models for pharmacokinetic modelling as it is specialised for this area. BUGS does not have this ability although there is a proposed add on to BUGS called PKBUGS which has a library of pharmacokinetic models which can be selected, but this is still to be released. NONMEM is flexible in that it allows the specification of a user defined likelihood but there are restrictions on the number of levels of random effects that can be estimated. BUGS has a library of distributions that can be selected for likelihood and prior distributions and any level of random effects can be incorporated. BUGS is more flexible in this respect but NONMEM is generally more stable at estimating nonlinear models, in the sense that it does not crash as much. Both packages are suitable for estimating proportional odds models but BUGS is more flexible in the type of random effects distributions that can be selected whereas NONMEM allows the estimation of nonlinear proportional odds models more freely.

Optimal design for categorical data has not been discussed in the literature before and the results given here are preliminary results for further work in categorical data optimal design. If the analysis of population categorical pharmacodynamic data is to progress then the design of studies for the analysis of such data must also be considered when designing clinical trials. Optimal design has been considered for binary and binomial data but only in terms of fixed effects models. Before population designs are considered for categorical models, it would be worth considering optimal designs for logistic mixed effects regression models. This could then be used to expand to mixed effects proportional odds models. The results obtained for the three category proportional odds model showed that the optimal design for such models are not trivial and require further investigation.

There is considerable scope for further research into the analysis of categorical pharmacodynamic data. One of the most important inadequacies of categorical data is the lack of information when compared to continuous data. This needs to be investigated to see how complex models can become for categorical data. This could be carried out by simulation studies where continuous models are simulated with different levels of variability and then categorised and reanalysed to study the models that can be estimated. The same problem could be assessed in terms of considering when the number of categories become substantial enough to use continuous approximations? As well as studying the estimability of fixed effects parameters, the random effects could be considered to check how well the variability is estimated in terms of interindividual and residual variability. NONMEM and BUGS were used to estimate the models in this thesis but other programs could be used and other techniques such as nonparametric techniques and possibly the implementation of the EM algorithm to maximise

population likelihood functions. Model validation is another important area of research as particularly the asymptotic distribution of binary-normal mixed distributions is not known and this would be useful for model checking in terms of residuals and summary statistics. As well as the analysis of categorical data, it would be appropriate that the trials be designed appropriately for the modelling of such data. Optimal design is one method of designing preclinical and clinical trials but in terms of population analysis, optimal design is still in the early stages for mixed effects models, even for continuous data. Clinical trial simulation has become a popular method for designing trials and so this could also be considered.

Bibliography

Aarons L: Software for population pharmacokinetics and pharmacodynamics. Clinical Pharmacokinetics, 1999, Vol. 36, No. 4, 255-264.

Aarons L, Vozeh S, Wenk M, Weiss Ph, Follath F: Population pharmacokinetics of tobramycin. British Journal of Clinical Pharmacology, 1989, Vol. 28, 305-314.

Agresti A: Categorical Data Analysis. Wiley, New York, 1990.

Abdelbasit KM, Plackett RL: Experimental design for binary data. Journal of the Amercian Statistical Association, 1983, Vol. 78, No. 381, 90-98.

Aitken M: Posterior Bayes Factors. Journal of the Royal Statistical Society Series B, Vol. 53, No. 1, 111-142.

Albert J, Chib S: Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 1993, Vol. 88, No. 422, 669-679.

Albert J, Chib S: Bayesian residual analysis for binary response regression models. Biometrika, 1995, Vol. 82, 747-759.

Albert PS, Hunsberger SA, Biro FM: Modelling repeated measures with monotonic ordinal responses and misclassification, with applications to studying maturation. Journal of the American Statistical Association, 1997, Vol. 92, No. 440, 1304-1311.

Anderson JA: Regression and ordered categorical variables. Journal of the Royal Statistical Society Series B, 1984, Vol. 46, No. 1, 1-30.

Anscombe FJ: Contribution to the discussion of a paper by H Hotelling. Journal of the Royal Statistical Society Series B, 1953, Vol. 15, 229-230.

D'Argenio DZ: Optimal sampling times for pharmacokinetic experiments. Journal of Pharmacokinetics and Biopharmaceutics, 1981, Vol. 8, 739-756.

D'Argenio DZ: Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. Mathematical Biosciences, 1990, Vol. 99, 105-118.

Armstrong BG, Sloan M: Ordinal regression for epidemiologic data. American Journal of Epidemiology, 1989, Vol. 129, No. 1, 191-204.

Atkinson AC, Donev AN: Optimum Experimental Designs. Clarendon Press, Oxford, 1992.

Azzalini A: Logistic regression for autocorrelated data with application to repeated measures. Biometrika, 1994, Vol. 81, No. 4, 767-775.

Beal SL, Sheiner LB: NONMEM users guides, parts I-VII. NONMEM Project Group, 1979-1992, University of California, San Francisco.

Bennett JE, Wakefield JC: A comparison of a Bayesian population method with two methods as implemented in commercially available software. Journal of Pharmacokinetics and Biopharmaceutics, 1996, Vol. 24, No. 4, 403-432.

Berger JO: Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York, 1985.

Bernardo JM, Smith AFM: Bayesian Theory. Wiley, Chichester, 1994.

Best NG, Tan KKC, Gilks WR Spiegelhalter DJ: Estimation of population pharmacokinetics using the Gibbs sampler. Journal of Pharmacokinetics and Biopharmaceutics, 1995, 407-435.

Bolland K, Sooriyarachchi MR, Whitehead J: Sample size review in a head injury trial with ordered categorical data. Statistics in Medicine, 1998, Vol. 17, 2835-2847.

Booth JG, Hobert JP: Maximising generalised linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society Series B, 1999, Vol. 61, Part 1, 265-285.

Box GEP: Sampling and Bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society Series B, 1980, Vol 143, Part 4, 383-430.

Boxtel CJ van, Holford NHG, Danhof M: The In Vivo Study of Drug Action. Elsevier, Amsterdam, 1992.

Brunier HC, Whitehead J: Samples sizes for phase II clinical trials derived from Bayesian decision theory. Statistics in Medicine, 1994, Vol. 13, 2493-2502.

Carlin BP, Chib S: Bayesian model choice via Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B, 1995, No. 3, 473-484.

le Cessie S, van Houwelingen JC: Logistic regression for correlated data. Applied Statistics, 1994, Vol. 43, No. 1, 95-108.

Chaloner K: Optimal Bayesian experimental design for linear models. Annals of Statistics, 1984, Vol. 12, No. 1, 283-300.

Chaloner K, Larntz K: Optimal Bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference, 1989, Vol. 19, 191-208.

Chib S, Greenberg E: Understanding the Metropolis-Hastings algorithm. The American Statistician, 1995, Vol. 49, No. 4, 327-335.

Chuang C, Agresti A: A new model for ordinal pain data from a pharmaceutical study. Statistics in Medicine, 1986, Vol. 5, 15-20.

Collett D: Modelling Binary Data. Chapman and Hall, London, 1991.

309

Conaway MR, Waternaux C: Pre-natal blood lead levels and learning difficulties in children: an analysis of non-randomly missing categorical data. Statistics in Medicine, 1992, Vol. 11, 799-811.

Conaway MR: Analysis of repeated categorical measurements with conditional likelihood methods. Journal of the American Statistical Association, 1989, Vol. 84, No. 405, 53-62.

Cook RJ, Farewell VT: Guidelines for monitoring efficacy and toxicity responses in clinical trials. Biometrics, 1994, Vol. 50, 1146-1152.

Cosson VF, Fuseau E, Efthymiopoulos C: Mixed effect modelling of sumatriptan pharmacokinetics during drug development. I: Interspecies allometric scaling. Journal of Pharmacokinetics and Biopharmaceutics, 1997, Vol. 25, No. 2, 149-167.

Cosson VF, Fuseau E: Mixed effect modeling of sumatriptan pharmacokinetics during drug development: II. From healthy subjects to phase 2 dose ranging studies in patients. Journal of Pharmacokinetics and Biopharmaceutics, 1998, Vol. 26, 149-171.

Cox C: Multinomial regression models on continuation ratios. Statistics in Medicine, 1988, Vol. 7, 435-441.

Danhof M, Mandema JW, Stijnen AM: Pharmacokinetic complexities in pharmacodynamic studies *in vivo*. The In Vivo Study of Drug Action, Elsevier, Amsterdam, 1992.

Davidian M, Gallant AR: Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. Journal of Pharmacokinetics and Biopharmaceutics, 1992, Vol. 20, 529-556.

Davidian M, Giltinan DM: Nonlinear models for repeated measurement data. Chapman and Hall, London, 1995.

Dean CB, Balshaw R: Efficiency by analysing counts rather than event times in Poisson and overdispersed Poisson regression models. Journal of the American Statistical Association, 1997, Vol. 92, No. 440, 1387-1398.

Diggle P, Kenward MG: Informative dropout in longitudinal data analysis. Applied Statistics, 1994, Vol. 43, No.1 49-93.

Diggle PJ, Liang KY, Zeger SL: Analysis of Longitudinal Data. 1994, Oxford Science Publications, Oxford.

Duquesnoy C, Mamet JP, Sumner D, Fuseau E: Comparative clinical pharmacokinetics of single doses of sumatriptan following subcutaneous, oral, rectal and intranasal administration. European Journal of Pharmaceutical Sciences, 1998, Vol. 6, 99-104.

Efron B: Comparing non-nested linear models. Journal of the American Statistical Association, 1984, Vol. 79, No. 388, 791-802.

Ette EI, Kelman AW, Howie CA, Whiting B: Analysis of animal pharmacokinetic data: performance of the one point per animal design. Journal of Pharmacokinetics and Biopharmaceutics, 1995, Vol. 23, No. 6, 551-566.

Farhmeir L, Kaufmann H: Regression models for non-stationary categorical time series. Journal of Time Series, 1987, Vol. 8, No. 2, 147-160.

Farrington, CP: On assessing goodness of fit of generalised linear models to sparse data. Journal of the Royal Statistical Society Series B, 1996, Vol. 58, No. 2, 349-360.

Fattinger K, Vozeh S, Borner HRHAM, Follath: Population pharmacokinetics of quinidine. British Journal of Clinical Pharmacokinetics, 1991, Vol. 31, 279-286.

Faucett CL, Schenker N, Elashoff RM: Analysis of censored survival data with intermittently observed time-dependent binary covariates. Journal of the American Statistical Association, 1998, Vol. 93, No. 442, 427-437.

Fedorov VV: Theory of Optimal Experiments. Academic Press, New York, 1972.

Firth D, Hinde JP: On Bayesian D-optimum design criteria and the equivalence theorem in non-linear models. Journal of the Royal Statistical Society Series B, 1997, Vol. 59, No.4, 793-797.

Fitzmaurice GM, Molenberghs G, Lipsitz SR: Regression models for longitudinal binary responses with informative drop-outs. Journal of the Royal Statistical Society Series B, 1995, Vol. 57, No. 4, 691-704.

Follman D, Wu M: An approximate generalised linear model with random effects for informative missing data. Biometrics, 1995, Vol. 51, 151-168.

Fuseau E, Sheiner LB: Simultaneos modelling of pharmacokinetics and pharmacodynamics with a nonparametric pharmacodynamic model. Clinical Pharmacology and Therapeutics, 1984, Vol. 35, No. 6, 733-741.

Gelfand AE, Smith AFM: Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 1990, Vol. 85, No. 410, 398-409.

Geman S, Geman D: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, Vol. PAMI-6, No. 6, 721-741.

Gilks WR, Wild P: Adaptive Rejection Sampling. Applied Statistics, 1992, Vol. 41, No. 2, 337-348.

Gilks WR, Thomas A, Spiegelhalter DJ: A language and Program for complex Bayesian Modelling. The Statistician, 1994, Vol. 43, No. 1, 169-177.

Gilks WR, Best NG, Tan KKC: Adaptive Rejection Metropolis sampling within Gibbs Sampling. Applied Statistics, 1995, Vol. 44, No. 4, 455-472.

Gordon NH, Wilson JKV: Using toxicity grades in the design and analysis of cancer phase I clinical trials. Statistics in Medicine, 1992, Vol. 11, 2063-2075.

Gupta, SK, Sathyan G, Lindemulder EA, Ho PL, Sheiner LB, Aarons L: Quantitative characterisation of therapeutic index: Application of mixed-effects modelling to evaluate oxybutynin dose-efficacy and dose-side effect relationships. Clinical Pharmacology and Therapeutics, 1999, Vol. 65, 672-684.

Hastings WK: Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 1970, Vol. 57, No. 1, 97-109.

Heise MA, Myers RH: Optimal designs for bivariate logistic regression. Biometrics, 1996, Vol. 52, 613-624.

Heyting A, Tolboom JTBM, Essers JGA: Statistical handling of drop-outs in longitudunal clinical trials. Statistics in Medicine, 1992, Vol. 11, 2043-2061.

Houwelingen JC van: Use and abuse of variance models in regression. Biometrics, 1988, Vol. 44, 1073-1081.

Jennings DE: Outliers and residual distributions in logistic regression. Journal of the American Statistical Association, 1986, Vol. 81, No. 396, 987-990.

Jennison C, Turnbull BW: Group sequential tests for bivariate response: Interim analyses of clinical trials with both efficacy and safety endpoints. Biometrics, 1993, Vol.49, 741-752.

Jones RH, Boadi-Boateng F: Unequally spaced longitudinal data with AR(1) serial correlation. Biometrics, 1991, Vol. 47, 161-175.

Jonsson EN, Wade JR, Karlsson MO: Comparison of some practical sampling strategies for population pharmacokinetic studies. Journal of Pharmacokinetics and Biopharmaceutics, 1996, Vol. 24, No. 2, 245-263.

Kahn MJ, Raftery AE: Discharge Rates of Medicare stroke patients to skilled facilities: Bayesian Logistic regression with unobserved heterogeneity. Journal of the American Statistical Association, 1996, Vol. 91, No. 433, 29-41.

Karlsson MO, Molnar V, Bergh J, Freijs A, Larsson R: A general model for time-dissociated pharmacokinetic-pharmacodynamic relationships exemplified by paclitaxel myelosuppression. Clinical Pharmacology and Therapeutics, 1998, Vol. 63, 11-25.

Karlsson MO, Beal SL, Sheiner LB: Three new residual error models for population PK/PD analyses. Journal of Pharmacokinetics and Biopharmaceutics, 1995, Vol. 23, No. 6, 651-672.

Karlsson MO, Sheiner LB: The importance of modelling interoccasion variability in population pharmacokinetic analyses. Journal of Pharmacokinetics and Biopharmaceutics, 1993, Vol. 21, No. 6, 735-750.

312

Kass RE, Carlin BP, Gelman A, Neal RM: Markov chain Monte Carlo in practice: A roundtable discussion. The American Statistician, 1998, Vol. 52, No. 2, 92-100.

Kass RE, Raftery AE: Bayes factors. Journal of the American Statistical Association, 1995, Vol. 90, No. 430, 773-795.

Kenward MG. Lesaffre E, Molenberghs G: An application of maximum likelihood and generalised estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. Biometrics, 1994, Vol. 50, 945-953.

Khan MK, Yazdi AA: On D-optimal designs for binary data. Journal of Statistical Planning and Inference 1988, Vol. 18, 83-91.

Kirkali Z, Whitaker RH: The Use of Oxybutynin in Practice. International Urology and Nephrology, 1987, Vol. 19, No. 4, 385-391.

Koch GG, Landis JR, Freeman JL, Freeman DH, Lehnen RG: A general methodology for the analysis of experiments with repeated measurements of categorical data. Biometrics, 1977, Vol. 33, 133-158.

Korn EL, Whittmore AS: Methods for analysing panel studies of acute health effects of air pollution. Biometrics, 1979, Vol. 35, 795-802.

Landwehr JM, Pregibon D, Shoemaker AC: Graphical methods for assessing logistic regression models. Journal of the American Statistical Association, 1984, Vol. 79, No. 385, 61-83.

Laird NM: Missing data in longitudinal studies. Statistics in Medicine, 1988, Vol. 7, 305-315.

Lee Y, Nelder JA: Hierarchical Generalised Linear Models. Journal of the Royal Statistical Society Series B, 1996, Vol. 58, No. 4, 619-678.

Lemeshow S, Hosmer DW: A review of goodness of fit statistics for use in the development of logistic regression models. American Journal of Epidemiology, 1982, Vol. 115, No. 1, 92-106.

Levy G: Variability in animal and human pharmacodynamic studies. Variability in Drug Therapy, 1985, Raven Press, 125-138.

Lindstrom MJ, Bates DM: Newton-Raphson and EM algorithms for linear mixed-effects for repeated-measures data. Journal of the American Statistical Association, 1988, Vol. 83, No. 404, 1014-1022.

Lindstrom MJ, Bates DM: Nonlinear Mixed Effects models for repeated measures data. Biometrics, 1990, Vol. 46, 673-687.

Lipsitz, SR, Fitzmaurice GM: Sample size for repeated measures studies with binary responses. Statistics in Medicine, 1994, Vol. 13, 1233-1239.

Longford NT: Logistic regression with random coefficients. Computational Statistics and Data Analysis, 1994, Vol. 17, 1-15.

Lunn DJ, Aarons LJ: Markov chain Monte Carlo techniques for studying interoccasion and intersubject variability: Application to pharmacokinetic data. Applied Statistics, 1997, Vol. 47, No. 1, 73-91.

McArthur RD: Parameter estimation in a two-compartment population pharmacokinetic model with destructive sampling. Mathematical Biosciences, 1988, Vol. 91, 157-173.

McCullagh P: Regression models for ordinal data. Journal of the Royal Statistical Society Series B, 1980, Vol. 42, No. 2, 109-142.

McCullagh P, Nelder JA: Generalised Linear Models. Chapman and Hall, London, 1989.

Mallet A, A maximum likelihood estimation method for random coefficient regression models. Biometrika, 1986, Vol. 73, No. 3, 645-656.

Mallet A, Mentre F, Steimer JL, Lokiec F: Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to Cyclosporine. Journal of Pharmacokinetics and Biopharmaceutics, 1998, Vol. 16, No. 3, 311-327.

Mandema JW, Verotta D, Sheiner LB: Building population pharmacokinetic-pharmacodynamic models. I. Models for covariate effects. Journal of Pharmacokinetic and Biopharmaceutics, 1992, Vol. 20, No. 5, 511-528.

Mandema JW, Stanski DR: Population pharmacodynamic model for the ketorolac analgesia. Clinical Pharmacology and Therapeutics, 1996, Vol. 60, 619-635.

Mentre F, Gomeni R, A two-step iterative algorithm for estimation in nonlinear mixed-effect models with an evaluation population pharmacokinetics. Journal of Biopharmaceutical Sciences, 1995, Vol. 5, No. 2, 141-158.

Mentre F, Mallet A, Baccar D: Optimal design in random-effects regression models. Biometrika, 1997, Vol. 84, No. 2, 429-442.

Merle Y, Mentre F: Bayesian design criteria: computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. Journal of Pharmacokinetics and Biopharmaceutics, 1995, Vol. 23, No. 1, 101-125.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E: Equations of state calculations by fast computing machines. Journal of Chemical Physics, 1953, Vol. 21, 1087-1091.

Minkin S: Optimal designs for binary data. Journal of the American Statistical Association, 1987, Vol. 82, No. 400, 1098-1103.

Moore KHP, Hussey EK, Fuseau E, Duquesnoy C, Pakes GE: Safety, tolerability, and pharmacokinetics of sumatriptan in healthy subjects following ascending single intranasal doses and multiple intranasal doses. Cephalalgia, 1997, Vol. 17, 541-550. Moore MJ, Theisson JJ: Cytotoxics and irreversible effects. The In Vivo Study of Drug Action, Elsevier, Amsterdam, 1992.

Murtaugh PA, Fisher LD: Random-effects model of dependence between efficacy and toxicity in dose-ranging studies. Communications in Statistics and Simulations, 1993, Vol. 22, No. 2, 507-522.

Nelder JA, Mead R: A simplex method for function minimisation. Computer Journal, 1965, Vol. 7, 308-313.

Peck CC, Beal SL, Sheiner LB, Nichols AI: Extended least squares nonlinear regression: a possible solution to the "choice of weights" problem in analysis of individual pharmacokinetic data. Journal of Pharmacokinetics and Biopharmaceutics, 1984, Vol. 12, No. 5, 545-558.

Plosker GL McTavish D: Sumatriptan: a reappraisal of its pharmacology and therapeutic efficacy in the acute treatment of migraine and cluster headache. Drugs, 1994, Vol. 47, No. 4, 622-651.

Pregibon D: Goodness of link tests for generalised linear models. Applied Statistics, 1980, Vol. 29, No. 1, 15-24.

Pregibon D: Logistic regression diagnostics. The Annals of Statistics, 1981, Vol. 9, No. 4, 705-724.

Pulkstenis EP, Ten Have TR, Landis JR: Model for the analysis of binary longitudinal pain data subject to informative dropout through remediation. Journal of the American Statistical Association, 1998, Vol. 93, No. 442, 438-450.

Racine-Poon A: A Bayesian approach to nonlinear random effects models. Biometrics, 1985, Vol. 41, 1015-1023.

Racine-Poon A, Wakefield, J: Statistical methods for population pharmacokinetic modelling. Statistical Methods in Medical Research, 1998, Vol. 7, 63-84.

Ragueneau I, Lavelille C, Jochemsen R, Resplandy G, Funck-Brentano C, Jaillon P: Pharmacokinetic-pharmacdynamic modelling of the effects of ivabradine, a direct sinus node inhibitor, on heart rate in healthy volunteers. Clinical Pharmacology and Therapeutics, 1998, Vol. 64, 192-203.

Ripley BD: Stochastic Simulation. Wiley, New York, 1987.

Ritter C, Tanner MA: Facilitating the Gibbs Sampler: The Gibbs stopper and the griddy-Gibbs sampler. Journal of the American Statistical Association, 1992, Vol. 87, No. 419, 861-868.

Rowland M, Tozer TN: Clinical Pharmacokinetics Concepts and Applications. Third Edition. Williams and Wilkins, Baltimore, 1995.

Rydberg T, Jonsson A, Karlsson MO, Melander A: Concentration-effect relations of glibenclamide and its active metabolites in man: modelling of pharmacokinetics and pharmacodynamics. British Journal of Clinical Pharmacology, 1997, Vol. 43, 373-381.

Schluchter MD: Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine, 1992, Vol. 11, 1861-1870.

Schumitzky A: Nonparametric EM algorithm for estimating prior distributions. Applied Mathematics and Computation, 1991, Vol. 45, 143-157.

Searle SR, Casella G, McCulloch CE: Variance Components. Wiley, New York, 1992.

Senn S: A random effects model for ordinal responses from a crossover trial. Statistics in Medicine, 1993, Vol. 12, 2147-2151.

Sheiner LB: A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. Clinical Pharmacology and Therapeutics, 1994, Vol. 56, 309-322.

Sheiner LB, RosenBerg B, Marathe VV: Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. Journal of Pharmacokinetics and Biopharmaceutics, 1977, Vol. 5, No. 5, 445-479.

Sheiner LB, Beal SL, Dunne A: Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. Journal of the American Statistical Association, 1997, Vol. 92, No. 440, 1235-1255.

Smith AFM, Roberts GO: Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B, 1993, Vol. 55, No. 1, 3-23.

Smith AFM Gelfand AE: Bayesian Statistics without tears: a sampling resampling perspective. The American Statistician, 1992, Vol. 46, No. 2, 84-88.

Smith AFM, Spiegelhalter: Bayes Factors and choice for linear models. Journal of the Royal Statistical Society Series B, 1980, Vol. 42, No. 2, 213-220.

Spears FM, Brown BW, Atkinson EN: The effect of incomplete knowledge of parameter values on single and multiple stage designs for logistic regression. Biometrics, 1997, Vol. 53, 1-10.

Spiegelhalter DJ, Thomas A, Best N, Gilks WR: BUGS: Bayesian Inference using Gibbs Sampling, Version 0.50. 1995, Cambridge University, Medical Research Council Biostatistics Unit, Institute of Public Health. http://www.mrc-bsu.cam.ac.uk/bugs

Spiegelhalter DJ: Bayesian graphical modelling: a case-study in monitoring health outcomes. Applied Statistics, 1998, Vol. 47, Part 1, 115-133.

Stemier JL, Mallet A, Mentre F: Estimating interindividual pharmacokinetic variability. Variability in Drug Therapy, 1985, Raven Press, 65-112.

Stiratelli R, Laird N, Ware JH: Random effects models for serial observations with binary response. Biometrics, 1984, Vol. 40, 961-971.

Sun D, Tsutakawa RK: Bayesian design for dose-response curves with penalised risk. Biometrics, 1997, Vol. 53, 1262-1273.

Ten Have TR, Kunselman AR, Pulkstenis EP, Landis JR: Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. Biometrics. 1998, Vol. 54, 367-383.

Thall PF, Russell KE: A strategy for dose finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Biometrics, 1998, Vol. 54, 251-264.

Tierney L, Kadane JB: Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association, 1986, Vol. 81, 82-86.

Tierney L: Markov chains for exploring posterior distributions. The Annals of Statistics, 1994, Vol. 24, No. 4, 1701-1762.

Tod M, Rocchisani JM: Comparison of ED, EID, and API criteria for the robust optimisation of sampling times in pharmacokinetics. Journal of Pharmacokinetics and Biopharmaceutics, 1997, 515-537.

Tsutakawa RK: Selection of dose levels for estimating a percentage point of a logistic quantal response curve. Applied Statistics, 1980, Vol. 29, No. 1, 25-33.

Tubert-Bitter P, Bloch DA, Raynauld JP: Comparing the bivariate effects of toxicity and efficacy of treatments. Statistics in Medicine, 1995, Vol. 14, 1129-1141.

Turi M, Stein G: The determination and practically useful doses of new drugs: some methodological considerations. Statistics in Medicine, 1986, Vol. 5, 449-457.

Vozeh S: The population approach: 1991-1997. COST B1 – The Population Approach: Measuring and Managing Variability in Response, Concentration and Dose. 1997, European Commission, 3-8.

Wakefield J: The Bayesian analysis of population pharmacokinetic models. Journal of the American Statistical Society, 1996, Vol. 91, No. 433, 62-75.

Wakefield J: Bayesian individualisation via sampling based methods. Journal of Pharmacokinetics and Biopharmaceutics, 1995, Vol. 25, 103-131.

Wakefield J: An expected loss approach to the design of dosage regimens via sampling based methods. The Statistician, 1994, Vol. 43, N0. 1, 13-29.

Wakefield J, S Walker S: Bayesian Nonparametric population models: formulation and comparison with likelihood approaches. Journal of Pharmacokinetics and Biopharmaceutics, 1997, Vol. 25, No. 2, 235-253.

Wakefield J, Benett J: The Bayesian modelling of covariates for population pharmacokinetic models. Journal of the American Statistical Society, 1996, Vol. 91, No. 435, 917-926.

Wakefield J, Racine-Poon A: An Application of Bayesian population pharmacokinetic/pharmacodynamic models to dose recommendations. Statistics in Medicine, 1995, Vol. 14. 971-986.

Whitehead J: Sample size calculations for ordered categorical data. Statistics in Medicine, 1993, Vol. 12, 2257-2271.

Whitehead J, Brunier H: Bayesian Decision procedures for dose determining experiments. Statistics in Medicine, 1995, Vol. 14, 885-893.

Wu MC, Carroll RJ: Estimation and comparison of changes of the presence of informative right censoring by modelling the censoring process. Biometrics, 1988, Vol. 44, 175-188.

Zeger SL, Liang SL: An overview of methods for the analysis of longitudinal data. Statistics in Medicine, 1992, Vol. 11, 1825-1839.

Zeger SL, Karim MR: Generalised linear models with random effects; a Gibbs sampling approach. Journal of the American Statistical Society, 1991, Vol. 86, No. 413, 79-86.

Zeger SL, Liang KY, Albert PS: Models for longitudinal data: A generalised estimating equation approach. Biometrics, 1988, Vol. 44, 1049-1060.

Appendix. Programs for Parameter Estimation

A1 BUGS Code

A1.1 BUGS Code For Proportional Odds Model (From Toxicokinetic Data Set I)

```
model auc1b.bug;

const
  I = 12,  # number of rats
  J = 5;   # number of categories

var
  pi[I,J], gamma[I,J-1], theta[J-1], beta1, eta[I], tau, sigma, Y[I,J],
  E[I,J], beta2, beta12, auc[I], sex[I], dose[I], ncount[I];

data auc, sex in "auc1.dat", Y, ncount in "auc1a.dat";

inits in "auc1.in";

{
  for (i in 1:I) {
    for (j in 1:J-1) {
# logit function
      cloglog(gamma[i,j]) <- theta[j] + beta1*auc[i] + beta2*sex[i]
                             + beta12*auc[i]*sex[i] + eta[i];
    }
# Probability of being in category j for rat i
    pi[i,1] <- gamma[i,1];
    for (j in 2:J-1) {
      pi[i,j] <- gamma[i,j] - gamma[i,j-1];
    }
    pi[i,J] <- 1 - gamma[i,J-1];
# Distribution of Y
    Y[i,] ~ dmulti(pi[i,],ncount[i]);
    eta[i] ~ dnorm(0,tau);
  }
  for (i in 1:I) {
    for (j in 1:J) {
      E[i,j] <- ncount[i]*pi[i,j];
    }
  }
  tau ~ dgamma(0.0001,0.0001);
  sigma <- 1/sqrt(tau);
```

```
# Ordered cut points
   theta[1] ~ dnorm(1,1.0E-5)I(,theta[2]);
   theta[2] ~ dnorm(2,1.0E-5)I(theta[1],theta[3]);
   theta[3] ~ dnorm(3,1.0E-5)I(theta[2],theta[4]);
   theta[4] ~ dnorm(4,1.0E-5)I(theta[3],);

# Gradient paramters
   beta1 ~ dnorm(0,1.0E-5);
   beta2 ~ dnorm(0,1.0E-5);
   beta12 ~ dnorm(0,1.0E-5);
}
```

A1.2 BUGS Code For Bayes Factors Estimation

```
model comaucd3.bug;

const
   I = 12,     # Number of rats
   J = 5,      # Number of categories (lesion scores)
   M = 2;      # Number of models

var
   Y[I,J],          # response for rat i, category j
   ncount[I],       # number of observations per rat
   p[M],            # prior for model m
   pM2,             # probability for model 2
   m,               # true model
   pi[M,I,J],       # probability for rat i, having lesion score j and being in model m
   gamma[M,I,J-1],  # cumulative probability for rat i, lesion score j and model m
   theta[J-1],      # cut points for model 1
   phi[J-1],        # cut points for model 2
   dose[I],         # dose covariate
   AUC[I],          # AUC covariate
   sex[I],          # gender covariate
   beta11,          # coefficient for dose
   beta12,          # coefficient for AUC
   beta21,          # coefficient for sex model 1 (dose model)
   beta22,          # coefficient for sex model 2 (AUC model)
   beta31,          # coefficient for interaction model 1 (dose model)
   beta32,          # coefficient for interaction model 2 (AUC model)
   mu.theta[M,J],   # mean of theta for category j, model m
   tau.theta[M,J],  # precision of theta for category j, model m
   mu.phi[M,J],     # mean of phi for category j, model m
   tau.phi[M,J],    # precision of phi for category j, model m
```

```
mu.beta11[M],   # mean of beta11 for model m
tau.beta11[M],  # precision of beta11 for model m
mu.beta12[M],   # mean of beta12 for model m
tau.beta12[M],  # precision of beta12 for model m
mu.beta21[M],   # mean of beta21 for model m
tau.beta21[M],  # precision of beta21 for model m
mu.beta22[M],   # mean of beta22 for model m
tau.beta22[M],  # precision of beta22 for model m
mu.beta31[M],   # mean of beta31 for model m
tau.beta31[M],  # precision of beta32 for model m
mu.beta32[M],   # mean of beta32 for model m
tau.beta32[M];  # precision of beta32 for model m


data auc, dose, sex, ncount in "comaucd.dat", Y in "comaucde.dat";
inits in "comaucd.in";
{
# model (node) prior distribution.  Categorical distribution
   m ~ dcat(p[]);
   p[1] <- 0.999; p[2] <- 0.001; # use for joint modelling
   pM2 <- step(m - 1.5);
# model structure
   for (i in 1:I) {
      for (j in 1:J-1) {
         logit(gamma[1,i,j]) <- theta[j]+beta11*dose[i];
         logit(gamma[2,i,j]) <- phi[j]+beta12*auc[i];
      }
# Probability of being in category j and covariate pattern i and model m
      pi[1,i,1] <- gamma[1,i,1];
      pi[2,i,1] <- gamma[2,i,1];
      for (j in 2:J-1) {
         pi[1,i,j] <- gamma[1,i,j]-gamma[1,i,j-1];
         pi[2,i,j] <- gamma[2,i,j]-gamma[2,i,j-1];
      }
      pi[1,i,J] <- 1-gamma[1,i,J-1];
      pi[2,i,J] <- 1-gamma[2,i,J-1];
   }
# Distribution of Y
   for (i in 1:I) {
      Y[i,] ~ dmulti(pi[m,i,],ncount[i]);
   }
# Model 1
   theta[1] ~ dnorm(mu.theta[m,1],tau.theta[m,1]);
   theta[2] ~ dnorm(mu.theta[m,2],tau.theta[m,2]);
   theta[3] ~ dnorm(mu.theta[m,3],tau.theta[m,3]);
   theta[4] ~ dnorm(mu.theta[m,4],tau.theta[m,4]);
   beta11   ~ dnorm(mu.beta11[m],tau.beta11[m]);
#  beta21   ~ dnorm(mu.beta21[m],tau.beta21[m]);
#  beta31   ~ dnorm(mu.beta31[m],tau.beta31[m]);
# Estimation priors
   mu.theta[1,1] <- 0; tau.theta[1,1] <- 0.001;
```

```
    mu.theta[1,2] <- 0; tau.theta[1,2] <- 0.001;
    mu.theta[1,3] <- 0; tau.theta[1,3] <- 0.001;
    mu.theta[1,4] <- 0; tau.theta[1,4] <- 0.001;
    mu.beta11[1]  <- 0; tau.beta11[1]  <- 0.001;
#   mu.beta21[1]  <- 0; tau.beta21[1]  <- 0.001;
#   mu.beta31[1]  <- 0; tau.beta31[1]  <- 0.001;
# Pseudo-priors
    mu.phi[1,1] <- 1.12; tau.phi[1,1] <- 13;
    mu.phi[1,2] <- 2.368; tau.phi[1,2] <- 9.6;
    mu.phi[1,3] <- 4.124; tau.phi[1,3] <- 7;
    mu.phi[1,4] <- 5.836; tau.phi[1,4] <- 3.8;
    mu.beta12[1] <- -0.037928; tau.beta12[1] <- 48000;
#   mu.beta22[1] <- 0.5669; tau.beta22[1] <- 17;
#   mu.beta32[1] <- 0; tau.beta32[1] <- 30000;
# Model 2
    phi[1] ~ dnorm(mu.phi[m,1],tau.phi[m,1]);
    phi[2] ~ dnorm(mu.phi[m,2],tau.phi[m,2]);
    phi[3] ~ dnorm(mu.phi[m,3],tau.phi[m,3]);
    phi[4] ~ dnorm(mu.phi[m,4],tau.phi[m,4]);
    beta12 ~ dnorm(mu.beta12[m],tau.beta12[m]);
#   beta22 ~ dnorm(mu.beta22[m],tau.beta22[m]);
#   beta32 ~ dnorm(mu.beta32[m],tau.beta32[m]);
# Estimation priors
    mu.phi[2,1]  <- 0; tau.phi[2,1]  <- 0.001;
    mu.phi[2,2]  <- 0; tau.phi[2,2]  <- 0.001;
    mu.phi[2,3]  <- 0; tau.phi[2,3]  <- 0.001;
    mu.phi[2,4]  <- 0; tau.phi[2,4]  <- 0.001;
    mu.beta12[2] <- 0; tau.beta12[2] <- 0.001;
#   mu.beta22[2] <- 0; tau.beta22[2] <- 0.001;
#   mu.beta32[2] <- 0; tau.beta32[2] <- 0.001;
# Pseudo-priors
    mu.theta[2,1] <- 0.87; tau.theta[2,1] <- 17;
    mu.theta[2,2] <- 2.08; tau.theta[2,2] <- 14;
    mu.theta[2,3] <- 3.78; tau.theta[2,3] <- 10;
    mu.theta[2,4] <- 5.47; tau.theta[2,4] <- 4.3;
    mu.beta11[2]  <- -0.0128; tau.beta11[2]  <- 450000;
#   mu.beta21[2]  <- 0.784; tau.beta21[2]  <- 16;
#   mu.beta31[2]  <- 0; tau.beta31[2] <- 30000;
}
```

A1.3 BUGS Code For Simultaneous Analysis of Sumatriptan Data

```
model pkpd4.bug;

const
```

```
N = 83,    # number of individuals in PD data
T = 6,     # number of time points in PD data
T2= 7,     # number of time points in PK data
K = 4;     # number of cut points
var
    alpha[N], lambda[N], time[N,T], W[N,T], phi[N,T,K-1], beta, theta[K-1],
    a[N], tau.a, AUC[N,T], sigma.a, pi[N,T,K], dose[N], id[N], conc[N,T],
    lnmu[N,T2], epsilon[N,T2], abs, tau, sigma, Y[N,T2], mu.alpha, mu.lambda,
    tau.alpha, tau.lambda, sigma.alpha, sigma.lambda, nobs[N], idpk[N],
    timepk[N,T2], b[N,T], tau.b, sigma.b;
data idpk, timepk, nobs, Y in "logpk.dat", id, dose, time, W in "effic.txt";
inits in "pkpd4a.in";
{
  for (i in 1:N) {
    for (j in 1:nobs[i]) {
      lnmu[i,j] <- log(dose[i])+alpha[i]+lambda[i]*timepk[i,j]+log(1-exp(-abs*timepk[i,j]));
      Y[i,j] ~ dnorm(lnmu[i,j],tau);
    }
    alpha[i] ~ dnorm(mu.alpha,tau.alpha);
    lambda[i] ~ dnorm(mu.lambda,tau.lambda);
  }
  mu.alpha ~ dnorm(-0.5,0.0001);
  mu.lambda ~ dnorm(0,0.0001);
  tau.alpha ~ dgamma(0.001,0.001);
  tau.lambda ~ dgamma(0.001,0.001);
  sigma.alpha <- 1/sqrt(tau.alpha);
  sigma.lambda <- 1/sqrt(tau.lambda);
  abs ~ dunif(0,3);
  tau ~ dgamma(0.001,0.001);
  sigma <- 1/sqrt(tau);
  for (i in 1:N) {
    for (j in 1:T) {
#       AUC[i,j] <- dose[i]*exp(alpha[i])*((1/lambda[i])*exp(lambda[i]*
#             time[i,j]) - (1/(lambda[i]-abs))*exp((lambda[i]-abs)*time[i,j])
#             + 1/(lambda[i]*(lambda[i]-abs)));
      conc[i,j] <- dose[i]*exp(alpha[i]+lambda[i]*time[i,j])*(1-exp(-abs*time[i,j]))
             *(1-equals(dose[i],0));
      for (k in 1:K-1) {
        logit(phi[i,j,k]) <- theta[k]+beta*conc[i,j]+a[i]+b[i,j];
      }
      W[i,j] ~ dcat(pi[i,j,1:K]);
      pi[i,j,1] <- 1 - phi[i,j,1];
      pi[i,j,2] <- phi[i,j,1] - phi[i,j,2];
      pi[i,j,3] <- phi[i,j,2] - phi[i,j,3];
      pi[i,j,K] <- phi[i,j,K-1];
      b[i,j] ~ dnorm(0,tau.b)
    }
    a[i] ~ dnorm(0,tau.a);
  }
  tau.a ~ dgamma(0.001,0.001);
```

```
    tau.b ~ dgamma(0.001,0.001);
    sigma.a <- 1/sqrt(tau.a);
    sigma.b <- 1/sqrt(tau.b);
    theta[1] ~ dnorm(10,0.0001);#I(,theta[2]);
    theta[2] ~ dnorm(5,0.0001);#I(theta[1],theta[3]);
    theta[3] ~ dnorm(-1,0.0001);#I(theta[2],);
    beta ~ dnorm(0,0.0001);
}
```

A1.4 BUGS Code For Deviance Statistic in Sumatriptan Data Analysis

```
model pkpd6.bug;

const
   N = 83,    # number of individuals in PD data
   T = 6,     # number of time points in PD data
   T2= 7,     # number of time poinys in PK data
   abs = 0.5364,
   K = 4;     # number of cut points
var
   alpha[N], lambda[N], time[N,T], W[N,T], phi[N,T,K-1], beta, theta[K-1],
   AUC[N,T], pi[N,T,K], dose[N], id[N], conc[N,T], lnmu[N,T2],
   tau, sigma, Y[N,T2], nobs[N], idpk[N], timepk[N,T2],
   a[N], tau.a, sigma.a, dosepk[N], beta1, beta2, beta3,
   llike[N,T], llike.sat[N,T], deviance,b[N,T],tau.b,sigma.b,
   Wlike1[N,T], Wlike2[N,T], Wlike3[N,T], Wlike4[N,T];
data alpha, lambda in "pkparam.txt", id, dose, time, W in "effic.txt";
inits in "pkpd4a.in";
{
  for (i in 1:N) {
     for (j in 1:T) {
#      AUC[i,j] <- dose[i]*exp(alpha[i])*(1/lambda[i]*(exp(lambda[i]*time[i,j])-1)
#              -1/(lambda[i]-abs)*(exp((lambda[i]-abs)*time[i,j])-1));
#      conc[i,j] <- dose[i]*exp(alpha[i]+lambda[i]*time[i,j])*(1-exp(-abs*time[i,j]))
       for (k in 1:K-1) {
          logit(phi[i,j,k]) <- -(theta[k]+beta1*dose[i]+beta2*time[i,j]
                     +a[i]+b[i,j]);
       }
       b[i,j] ~ dnorm(0,tau.b);
       W[i,j] ~ dcat(pi[i,j,1:K]);
       pi[i,j,1] <- phi[i,j,1];
       pi[i,j,2] <- phi[i,j,2] - phi[i,j,1];
       pi[i,j,3] <- phi[i,j,3] - phi[i,j,2];
       pi[i,j,K] <- 1 - phi[i,j,K-1];
       Wlike1[i,j] <- equals(W[i,j],1);
```

```
        Wlike2[i,j] <- equals(W[i,j],2);
        Wlike3[i,j] <- equals(W[i,j],3);
        Wlike4[i,j] <- equals(W[i,j],4);
        llike[i,j] <- Wlike1[i,j]*(W[i,j]*log(pi[i,j,1])+(1-W[i,j])*log(1-pi[i,j,1]))
                  + Wlike2[i,j]*((W[i,j]/2)*log(pi[i,j,2])+(1-W[i,j]/2)*log(1-pi[i,j,2]))
                  + Wlike3[i,j]*((W[i,j]/3)*log(pi[i,j,3])+(1-W[i,j]/3)*log(1-pi[i,j,3]))
                  + Wlike4[i,j]*((W[i,j]/4)*log(pi[i,j,4])+(1-W[i,j]/4)*log(1-pi[i,j,4]));

        llike.sat[i,j] <- Wlike1[i,j]*(W[i,j]*log(W[i,j])+(1-W[i,j])*log(1-W[i,j]))
                  + Wlike2[i,j]*((W[i,j]/2)*log(W[i,j]/2)+(1-W[i,j]/2)*log(1-W[i,j]/2))
                  + Wlike3[i,j]*((W[i,j]/3)*log(W[i,j]/3)+(1-W[i,j]/3)*log(1-W[i,j]/3))
                  + Wlike4[i,j]*((W[i,j]/4)*log(W[i,j]/4)+(1-W[i,j]/4)*log(1-W[i,j]/4));

      }
      a[i] ~ dnorm(0,tau.a);
   }
   theta[1] ~ dnorm(4,0.0001);
   theta[2] ~ dnorm(2,0.0001);
   theta[3] ~ dnorm(0,0.0001);
   beta1 ~ dnorm(0,0.0001);
   beta2 ~ dnorm(0,0.0001);
   beta3 ~ dnorm(0,0.0001);
   tau.a ~ dgamma(0.001,0.001);
   sigma.a <- 1/sqrt(tau.a);
   tau.b ~ dgamma(0.001,0.001);
   sigma.b <- 1/sqrt(tau.b);
   deviance <- 2*(sum(llike.sat[,]) - sum(llike[,]));
}
```

A1.5 BUGS Code For Utility Functions in Oxybutynin Data Analysis

```
model oxyutil;

const
   N = 65,    # number of individuals
   M = 82,
   K = 4,     # number of categories
   D = 2,     # number of treatments
   T = 7;     # number of time points per individual
var
   trt1[M],trt2[M],trt3[M],Y[M,T],dose[M,T],        # data
   Ytime1[M,T],time[T],W[M,T],id[M],Y1[M],          # data
   plac[M,T],mu[M,T],phi[N,T,K-1],pi[N,T,K],        # mean responses
   plac1,plac2,plac3,tau.plac,sigma.plac,           # efficacy placebo parameters
   betae11,betae12,betae13,betae21,betae22,betae23, # efficacy parameters
```

```
    betaa11,betaa12,betaa13,betaa21,betaa22,betaa23,  # adverse parameters
    theta1[K-1],theta2[K-1],                          # adverse parameters (cut points)
    etaa1[N],etaa2[N],taua1,taua2,sigmaa1,sigmaa2,    # adverse random effects
    etae1[M],etae2[M],taue1,taue2,sigmae1,sigmae2,    # efficacy random effects
    epsilon1[N,T],tau1,sigma1,epsilon2[N,T],tau2,sigma2, # efficacy residual
    epsilona[M,T],taua,sigmaa,                        # adverse residual
    S[2,T],V[2,T],U[2,T],d[T],heure[T];               # loss function variables
data id, trt1, trt2, trt3, dose, Y, W, Y1, Ytime1 in "oxy1.dat", time in "time.dat";
inits in "oxyutil.in";
{
    for (i in 66:M) {
        for (j in 1:T) {
            plac[i,j] <- Y1[i]*(plac1-plac2*(1-exp(-plac3*time[j])));
            Y[i,j] ~ dnorm(plac[i,j],tau.plac)I(0,);
            mu[i,j] <- 0.0;
        }
    }
    for (i in 1:N) {
        etae1[i] ~ dnorm(0,taue1);
        etae2[i] ~ dnorm(0,taue2);
        for (j in 1:T) {
            log(mu[i,j]) <- trt1[i]*(log(Y1[i]*(plac1-plac2*(1-exp(-plac3*time[j]))))
                            +betae11*dose[i,j]+betae12*pow(dose[i,j],2)
                            +etae1[i]+epsilon1[i,j])
                        +trt2[i]*(log(Y1[i]*(plac1-plac2*(1-exp(-plac3*time[j]))))
                            +(betae21)*dose[i,j]
                            +(betae22)*pow(dose[i,j],2)
                            +etae2[i]+epsilon2[i,j]);
            epsilon1[i,j] ~ dnorm(0,tau1);
            epsilon2[i,j] ~ dnorm(0,tau2);
            for (k in 1:K-1) {
                logit(phi[i,j,k]) <- -trt1[i]*(theta1[k]+betaa11*dose[i,j]+betaa12*pow(dose[i,j],2)
                            +etaa1[i])
                        -trt2[i]*(theta2[k]+(betaa21)*dose[i,j]
                            +(betaa22)*pow(dose[i,j],2)+etaa2[i]);
            }
            W[i,j] ~ dcat(pi[i,j,1:K]);
            Y[i,j] ~ dpois(mu[i,j]);
            plac[i,j] <- 0.0;
            pi[i,j,1] <- 1-phi[i,j,1];
            for (k in 2:K-1) {
                pi[i,j,k] <- phi[i,j,k-1]-phi[i,j,k];
            }
            pi[i,j,K] <- phi[i,j,K-1];
        }
        etaa1[i] ~ dnorm(0,taua1);
        etaa2[i] ~ dnorm(0,taua2);
    }
    plac1 ~ dunif(0,5);
    plac2 ~ dunif(0,plac1);
```

```
plac3 ~ dunif(0,1);
tau.plac ~ dgamma(0.001,0.001);
sigma.plac <- 1/sqrt(tau.plac);
betae11 ~ dnorm(0,0.0001);
betae12 ~ dnorm(0,0.0001);
betae21 ~ dnorm(0,0.0001);
betae22 ~ dnorm(0,0.0001);
tau1 ~ dgamma(0.001,0.001);
sigma1 <- 1/sqrt(tau1);
tau2 ~ dgamma(0.001,0.001);
sigma2 <- 1/sqrt(tau2);
theta1[1] ~ dnorm(1.7,0.0001);
theta1[2] ~ dnorm(4.8,0.0001);
theta1[3] ~ dnorm(7.3,0.0001);
theta2[1] ~ dnorm(1.2,0.0001);
theta2[2] ~ dnorm(4.3,0.0001);
theta2[3] ~ dnorm(6.1,0.0001);
betaa11 ~ dnorm(0,0.0001);
betaa12 ~ dnorm(0,0.0001);
betaa21 ~ dnorm(0,0.0001);
betaa22 ~ dnorm(0,0.0001);
taua1 ~ dgamma(0.001,0.001);
taua2 ~ dgamma(0.001,0.001);
sigmaa1 <- 1/sqrt(taua1);
sigmaa2 <- 1/sqrt(taua2);
taue1 ~ dgamma(0.001,0.001);
taue2 ~ dgamma(0.001,0.001);
sigmae1 <- 1/sqrt(taue1);
sigmae2 <- 1/sqrt(taue2);
d[1] <- 0; d[2] <- 2.5; d[3] <- 5; d[4] <- 7.5;
d[5] <- 10; d[6] <- 12.5; d[7] <- 15;
heure[1] <- 7; heure[2] <- 14; heure[3] <- 21;
heure[4] <- 28; heure[5] <- 35; heure[6] <- 42;
heure[7] <- 49;
for (l in 1:T) {
   S[1,l] <- (1-step(20*(plac1-plac2*(1-exp(-plac3*heure[l])))*exp(betae11*d[l])-6))*
       (6-20*(plac1-plac2*(1-exp(-plac3*heure[l])))*exp(betae11*d[l]))/6;
   S[2,l] <- (1-step(20*(plac1-plac2*(1-exp(-plac3*heure[l])))*exp((betae21)*d[l])-6))*
       (6-20*(plac1-plac2*(1-exp(-plac3*heure[l])))*exp((betae21)*d[l]))/6;
   V[1,l] <- step(exp(theta1[3]+betaa11*d[l])/(1+exp(theta1[3]+betaa11*d[l]))-0.9)+
       (1-step(exp(theta1[3]+betaa11*d[l])/(1+exp(theta1[3]+betaa11*d[l]))-0.9))*
       exp(theta1[2]+betaa11*d[l])/(1+exp(theta1[2]+betaa11*d[l]))*
       step(exp(theta1[2]+betaa11*d[l])/(1+exp(theta1[2]+betaa11*d[l]))-0.3)-1;
   V[2,l] <- step(exp(theta2[3]+(betaa21)*d[l])/(1+exp(theta2[3]+(betaa21)*d[l]))-0.9)+
       (1-step(exp(theta2[3]+(betaa21)*d[l])/(1+exp(theta2[3]+(betaa21)*d[l]))-0.9))*
       exp(theta2[2]+(betaa21)*d[l])/(1+exp(theta2[2]+(betaa21)*d[l]))*
       step(exp(theta2[2]+(betaa21)*d[l])/(1+exp(theta2[2]+(betaa21)*d[l]))-0.3)-1;
   U[1,l] <- S[1,l] + V[1,l] - 0.05;
   U[2,l] <- S[2,l] + V[2,l] - 0.05;
} }
```

A2 NONMEM Code

A2.1 NONMEM Code For Two Compartment First Order Absorption Model

```
C   TWO COMPARTMENT FIRST ORDER ABSORPTION MODEL
C
    SUBROUTINE PRED (ICALL,NEWIND,THETA,DATREC,INDXS,F,G,H)
    DIMENSION  THETA(*),DATREC(*),INDXS(*), G(*),H(*)
    DIMENSION  ETA(10), EPS(10)
    REAL KA,K10,K12,K21,L1,L2
    SAVE
    COMMON /NMPRD4/ CO,CL,V1
    IF (ICALL.EQ.4) THEN
      IF (NEWIND.NE.2) CALL SIMETA(ETA)
      CALL SIMEPS(EPS)
            ELSE
      IF (NEWIND.NE.2) THEN
        CALL GETETA(ETA)
        EPS(01)=0.0
      END IF
    END IF
    TAU=24.
    ID=NINT(DATREC(1))
    TIME=DATREC(2)
    DOSE=DATREC(4)*1000.0
    CL=THETA(1)*EXP(ETA(1))
    V1=THETA(2)*EXP(ETA(2))
    V2=THETA(3)
    VSS=V1+V2
    CLD=THETA(4)
    KA=10.0
    R0=DOSE*KA
    K12=CLD/V1
    K21=CLD/V2
    K10=CL/V1
    SUM=K10+K12+K21
    ROOT=SUM*SUM-4.0*K10*K21
    L1=0.50*(SUM+SQRT(ROOT))
    L2=0.50*(SUM-SQRT(ROOT))
    C1=(L1-K21)/((KA-L1)*(L1-L2)*V1)
    C2=(K21-L2)/((KA-L2)*(L1-L2)*V1)
    C3=(K21-KA)/((L1-KA)*(L2-KA)*V1)
    DL1CL=0.5*(1.0/V1+0.5*(2.0*SUM/V1-4.0*K21/V1)/(SQRT(ROOT)))
    DL2CL=0.5*(1.0/V1-0.5*(2.0*SUM/V1-4.0*K21/V1)/(SQRT(ROOT)))
    DC1CL=DL1CL*(KA-K21)/((KA-L1)*(KA-L1)*(L1-L2)*V1)
  1    -(DL1CL-DL2CL)*(L1-K21)/((KA-L1)*(L1-L2)*(L1-L2)*V1)
```

```
      DC2CL=-DL2CL*(KA-K21)/((KA-L2)*(KA-L2)*(L1-L2)*V1)
   1    -(DL1CL-DL2CL)*(K21-L2)/((KA-L2)*(L1-L2)*(L1-L2)*V1)
      DC3CL=-(K21-KA)*(DL1CL/(L1-KA)+DL2CL/(L2-KA))/((L1-KA)*(L2-KA)*V1)
      DC1KA=-C1/(KA-L1)
      DC2KA=-C2/(KA-L2)
      DC3KA=-C3/(K21-KA)+C3/(L1-KA)+C3/(L2-KA)
      DL1V1=0.5*(-K10/V1-K12/V1+0.5*(-2.0*SUM*(K10/V1
   1    +K12/V1)+4.0*K10*K21/V1)/
   2    (SQRT(ROOT)))
      DL2V1=0.5*(-K10/V1-K12/V1-0.5*(-2.0*SUM*(K10/V1
   1    +K12/V1)+4.0*K10*K21/V1)/
   2    (SQRT(ROOT)))
      DC1V1=DL1V1*(KA-K21)/((KA-L1)*(KA-L1)*(L1-L2)*V1)
   1    -(DL1V1-DL2V1)*(L1-K21)/((KA-L1)*(L1-L2)*(L1-L2)*V1)
   2    -C1/V1
      DC2V1=-DL2V1*(KA-K21)/((KA-L2)*(KA-L2)*(L1-L2)*V1)
   1    -(DL1V1-DL2V1)*(K21-L2)/((KA-L2)*(L1-L2)*(L1-L2)*V1)
   2    -C2/V1
      DC3V1=-(K21-KA)*(DL1V1/(L1-KA)+DL2V1/(L2-KA))/((L1-KA)*(L2-KA)*V1)
   1    -C3/V1
      F1=C1*(EXP(-L1*TIME))/(1.-EXP(-L1*TAU))
      F2=C2*(EXP(-L2*TIME))/(1.-EXP(-L2*TAU))
      F3=C3*(EXP(-KA*TIME))/(1.-EXP(-KA*TAU))
      F=(F1+F2+F3)*R0
      CO=F
      G(1)=DC1CL*EXP(-L1*TIME)/(1.-EXP(-L1*TAU))
   1    -C1*TIME*EXP(-L1*TIME)*DL1CL/(1.-EXP(-L1*TAU))
   1    -C1*EXP(-L1*TIME)*DL1CL*TAU*EXP(-L1*TAU)
   1    /((1.-EXP(-L1*TAU))*(1.-EXP(-L1*TAU)))
   2    +DC2CL*EXP(-L2*TIME)/(1.-EXP(-L2*TAU))
   3    -C2*TIME*EXP(-L2*TIME)*DL2CL/(1.-EXP(-L2*TAU))
   3    -C2*EXP(-L2*TIME)*DL2CL*TAU*EXP(-L2*TAU)
   3    /((1.-EXP(-L2*TAU))*(1.-EXP(-L2*TAU)))
   4    +DC3CL*EXP(-KA*TIME)/(1.-EXP(-KA*TAU))
      G(1)=G(1)*R0*CL
      G(2)=DC1V1*EXP(-L1*TIME)/(1.-EXP(-L1*TAU))
   1    -C1*TIME*EXP(-L1*TIME)*DL1V1/(1.-EXP(-L1*TAU))
   1    -C1*EXP(-L1*TIME)*DL1V1*TAU*EXP(-L1*TAU)
   1    /((1.-EXP(-L1*TAU))*(1.-EXP(-L1*TAU)))
   2    +DC2V1*EXP(-L2*TIME)/(1.-EXP(-L2*TAU))
   3    -C2*TIME*EXP(-L2*TIME)*DL2V1/(1.-EXP(-L2*TAU))
   3    -C2*EXP(-L2*TIME)*DL2V1*TAU*EXP(-L2*TAU)
   3    /((1.-EXP(-L2*TAU))*(1.-EXP(-L2*TAU)))
   4    +DC3V1*EXP(-KA*TIME)/(1.-EXP(-KA*TAU))
      G(2)=G(2)*R0*V1
C     G(3)=DC1KA*EXP(-L1*TIME)/(1.-EXP(-L1*TAU))
C  1    +DC2KA*EXP(-L2*TIME)/(1.-EXP(-L2*TAU))
C  2    +DC3KA*EXP(-KA*TIME)/(1.-EXP(-KA*TAU))
C  3    -TIME*C3*EXP(-KA*TIME)/(1.-EXP(-KA*TAU))
C  4    -C3*EXP(-KA*TIME)*TAU*EXP(-KA*TAU)
```

```
C   5   /((1.-EXP(-KA*TAU))*(1.-EXP(-KA*TAU)))
C      G(3)=G(3)*R0*KA
C      G(1)=F
       H(1)=F
C      H(2)=1.
       END
```

A2.2 NONMEM Code For Proportional Odds Model

```
SUBROUTINE PRED(ICALL,NEWIND,THETA, DATREC, INDXS, F, G, H)
      DIMENSION THETA(*), DATREC(*), INDXS(*), G(10,*), H(10,*)
      DIMENSION ETA(10),EPS(10)
      DIMENSION PC(5), PB(5), DPCDE(5),DDPCDE(5)
      COMMON /NMPRD4 / Y
      COMMON /ROCM12/ MSEC
      INTEGER PID
      NCAT = 5
      IF(ICALL.EQ.4)THEN
        IF (ICALL.NE.2)CALL SIMETA(ETA)
          CALL SIMEPS(EPS)
              ELSE
        IF(NEWIND.NE.2)THEN
          CALL GETETA(ETA)
          EPS(1)=0.
        ENDIF
      ENDIF
      D = DATREC(2)
      PLP =  THETA(5)*D+ETA(1)
      DO 10 I = 1,NCAT-1
      PLP = PLP + THETA(I)
      X = EXP(PLP)
      PC(I) = X/(1.0+X)
      DPCDE(I) = PC(I) - PC(I)*PC(I)
      DDPCDE(I) = DPCDE(I)-2.*PC(I)*DPCDE(I)
10    CONTINUE
      PID = NINT(DATREC(3))
      PB(1) = PC(1)
      PC(NCAT) = 1
      DPCDE(NCAT) = 0.
      DO 20 I = 2,NCAT
      PB(I) = PC(I) - PC(I-1)
20    CONTINUE
      F = PB(PID)
      Y=F
      IF (PID.GT.1) THEN
        G(1,1) = DPCDE(PID) - DPCDE(PID-1)
          ELSE
```

```
      G(1,1) = DPCDE(PID)
    ENDIF
    IF(MSEC.EQ.1)THEN
     IF(PID.GT.1)THEN
       G(1,2)=DDPCDE(PID)-DDPCDE(PID-1)
        ELSE
       G(1,2)=DDPCDE(PID)
     ENDIF
    ENDIF
    H(1,1)=G(1,1)
    END
```

A3 FORTRAN Code

A3.1 FORTRAN Code For Categorical Data D-Optimal Design

```
PROGRAM SIMPLEX
    IMPLICIT DOUBLE PRECISION (A-H,O-Z)
    PARAMETER(NP=3,MP=4,FTOL=1.0E-14)
    DOUBLE PRECISION X, P, Y, FTOL, DET
    DIMENSION P(MP,NP), Y(MP), X(NP)
    INTEGER MP,NP,ITER,NDIM
    EXTERNAL FUNK
    P(1,1) = 1.3
    P(1,2) = 0.6
    P(1,3) = -0.8
    P(2,1) = -1.0
    P(2,2) = 0.2
    P(2,3) = 1.1
    P(3,1) = 0.0
    P(3,2) = 0.4
    P(3,3) = -0.6
    P(4,1) = 0.1
    P(4,2) = 0.4
    P(4,3) = -0.2
    NDIM = NP
    DO 12 I = 1,MP
      DO 11 J = 1,NP
        X(J) = P(I,J)
11    CONTINUE
      Y(I) = FUNK(X)
12  CONTINUE
    CALL AMOEBA(P,Y,MP,NP,NDIM,FTOL,FUNK,ITER,RTOL)
    DET=-Y(1)
    PRINT *, RTOL
    PRINT *, "X(1) = ",P(1,1),"X(2) = ",P(1,2),"X(3) = ",P(1,3)
```

```fortran
      PRINT *, DET
      END PROGRAM

      DOUBLE PRECISION FUNCTION FUNK(X)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      DOUBLE PRECISION X,ALPHA1,ALPHA2,BETA1
      DOUBLE PRECISION G11,G12,G13,G21,G22,G23
      DOUBLE PRECISION A11,A12,A13,A21,A22,A23,A31,A32,A33
      DIMENSION X(3)
      ALPHA1=-1.
      ALPHA2=1.
      BETA1=1.
      G11=DEXP(ALPHA1+BETA1*X(1))/(1.+DEXP(ALPHA1+BETA1*X(1)))
      G12=DEXP(ALPHA1+BETA1*X(2))/(1.+DEXP(ALPHA1+BETA1*X(2)))
      G13=DEXP(ALPHA1+BETA1*X(3))/(1.+DEXP(ALPHA1+BETA1*X(3)))
      G21=DEXP(ALPHA2+BETA1*X(1))/(1.+DEXP(ALPHA2+BETA1*X(1)))
      G22=DEXP(ALPHA2+BETA1*X(2))/(1.+DEXP(ALPHA2+BETA1*X(2)))
      G23=DEXP(ALPHA2+BETA1*X(3))/(1.+DEXP(ALPHA2+BETA1*X(3)))
      A11= G11*G21*(1.-G11)*(1.-G11)/(G21-G11)+G12*G22*(1.-G12)*
     1    (1.-G12)/(G22-G12)+G13*G23*(1.-G13)*(1.-G13)/(G23-G13)
      A12= -G11*(1.-G11)*G21*(1.-G21)/(G21-G11)-G12*(1.-G12)*G22*
     1    (1.-G22)/(G22-G12)-G13*(1.-G13)*G23*(1.-G23)/(G23-G13)
      A13= X(1)*G21*(1.-G11)*G11+X(2)*G22*(1.-G12)*G12+X(3)*G23*
     1    (1.-G13)*G13
      A21= -G11*(1.-G11)*G21*(1.-G21)/(G21-G11)-G12*(1.-G12)*G22*
     1    (1.-G22)/(G22-G12)
     2    -G13*(1.-G13)*G23*(1.-G23)/(G23-G13)
      A22= G21*G21*(1.-G21)*(1.-G21)*(1./(G21-G11)+1./(1.-G21))
     1    +G22*G22*(1.-G22)*(1.-G22)*(1./(G22-G12)+1./(1.-G22))
     2    +G23*G23*(1.-G23)*(1.-G23)*(1./(G23-G13)+1./(1.-G23))
      A23= X(1)*G21*(1.-G11)*(1.-G21)+X(2)*G22*(1.-G12)*(1.-G22)
     1    +X(3)*G23*(1.-G13)*(1.-G23)
      A31= X(1)*G11*G21*(1.-G11)+X(2)*G12*G22*(1.-G12)+X(3)*
     1    G13*G23*(1.-G13)
      A32= X(1)*G21*(1.-G11)*(1.-G21)+X(2)*G22*(1.-G12)*(1.-G22)
     1    +X(3)*G23*(1.-G13)*(1.-G23)
      A33= X(1)*X(1)*(G11*G21*(1.-G11)+G21*(1.-G21)*(1.-G11))
     1    +X(2)*X(2)*(G12*G22*(1.-G12)+G22*(1.-G22)*(1.-G12))
     2    +X(3)*X(3)*(G13*G23*(1.-G13)+G23*(1.-G23)*(1.-G13))
      FUNK = -(A11*(A22*A33-A23*A32) -
     1      A12*(A21*A33-A31*A23) +
     2      A13*(A21*A32-A31*A22))
      END FUNCTION

      SUBROUTINE AMOEBA(P,Y,MP,NP,NDIM,FTOL,FUNK,ITER,RTOL)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      PARAMETER (NMAX=20,ALPHA=1.0,BETA=0.5,GAMMA=2.0,ITMAX=1000)
      DIMENSION P(MP,NP),Y(MP),PR(NMAX),PRR(NMAX),PBAR(NMAX)
      DOUBLE PRECISION ALPHA,BETA,GAMMA,P,PR,PRR,PBAR,FTOL,RTOL
      INTEGER MPTS,NDIM,ITER,ILO,IHI,INHI
```

```
C     OPEN(6,FILE="PO3SIMP.RES")
      MPTS=NDIM+1
      ITER=0
1     ILO=1
      IF(Y(1).GT.Y(2))THEN
       IHI=1
       INHI=2
      ELSE
       IHI=2
       INHI=1
      ENDIF
      DO 11 I=1,MPTS
       IF(Y(I).LT.Y(ILO)) ILO=I
       IF(Y(I).GT.Y(IHI))THEN
         INHI=IHI
         IHI=I
       ELSE IF(Y(I).GT.Y(INHI))THEN
         IF(I.NE.IHI) INHI=I
       ENDIF
11    CONTINUE
      IF(Y(IHI).EQ.0.AND.Y(ILO).EQ.0) THEN
        RTOL= 0.
        ELSE
      RTOL=2.*DABS(Y(IHI)-Y(ILO))/(DABS(Y(IHI))+DABS(Y(ILO)))
      PRINT*,"Iteration =",ITER,"RTOL =",RTOL,"Y =",Y(ILO)
      ENDIF
      IF(RTOL.LT.FTOL)THEN
       RETURN
      ENDIF
      IF(ITER.EQ.ITMAX) PAUSE 'Amoeba exceeding maximum iterations.'
      ITER=ITER+1
      DO 12 J=1,NDIM
       PBAR(J)=0.
12    CONTINUE
      DO 14 I=1,MPTS
       IF(I.NE.IHI)THEN
         DO 13 J=1,NDIM
          PBAR(J)=PBAR(J)+P(I,J)
13       CONTINUE
       ENDIF
14    CONTINUE
      DO 15 J=1,NDIM
       PBAR(J)=PBAR(J)/NDIM
       PR(J)=(1.+ALPHA)*PBAR(J)-ALPHA*P(IHI,J)
15    CONTINUE
      YPR=FUNK(PR)
      IF(YPR.LE.Y(ILO))THEN
        DO 16 J=1,NDIM
         PRR(J)=GAMMA*PR(J)+(1.-GAMMA)*PBAR(J)
16      CONTINUE
```

```fortran
      YPRR=FUNK(PRR)
      IF(YPRR.LT.Y(ILO))THEN
        DO 17 J=1,NDIM
         P(IHI,J)=PRR(J)
17      CONTINUE
        Y(IHI)=YPRR
      ELSE
        DO 18 J=1,NDIM
         P(IHI,J)=PR(J)
18      CONTINUE
        Y(IHI)=YPR
      ENDIF
    ELSE IF(YPR.GE.Y(INHI))THEN
      IF(YPR.LT.Y(IHI))THEN
        DO 19 J=1,NDIM
         P(IHI,J)=PR(J)
19      CONTINUE
        Y(IHI)=YPR
      ENDIF
      DO 21 J=1,NDIM
        PRR(J)=BETA*P(IHI,J)+(1.-BETA)*PBAR(J)
21    CONTINUE
      YPRR=FUNK(PRR)
      IF(YPRR.LT.Y(IHI))THEN
        DO 22 J=1,NDIM
         P(IHI,J)=PRR(J)
22      CONTINUE
        Y(IHI)=YPRR
      ELSE
        DO 24 I=1,MPTS
          IF(I.NE.ILO)THEN
            DO 23 J=1,NDIM
              PR(J)=0.5*(P(I,J)+P(ILO,J))
              P(I,J)=PR(J)
23          CONTINUE
            Y(I)=FUNK(PR)
          ENDIF
24      CONTINUE
      ENDIF
    ELSE
      DO 25 J=1,NDIM
       P(IHI,J)=PR(J)
25    CONTINUE
      Y(IHI)=YPR
    ENDIF
    GO TO 1
    END
```