Trabajo Fin de Grado

# Statistical Models in Pharmacokinetics and Pharmacodynamics

*Presented by:*
**Sandra Benítez Peña**

*Supervisors:*
Dr. Rafael Blanquero Bravo, Universidad de Sevilla
Dr. Emilio Carrizosa Priego, Universidad de Sevilla


UNIVERSIDAD D SEVILLA

June 21, 2015

# Contents


**1 Introduction** **4**

**2 Nonlinear Mixed Effects Models** **5**
2.1 Nonlinear Models . . . . . 5
2.2 Nonlinear Fixed Effects Models . . . . . 7
2.3 Nonlinear Random and Mixed Effects Models . . . . . 8
2.3.1 The Two-Stage Hierarchy . . . . . 9
2.3.2 Summary . . . . . 15
2.4 Inference in Nonlinear Mixed Effects Models . . . . . 16
2.4.1 Inferential Approaches . . . . . 17
2.4.2 Inference . . . . . 19

**3 Applications in Pharmacokinetics and Pharmacodynamics** **25**
3.1 Pharmacokinetics and Pharmacodynamics . . . . . 25
3.1.1 Pharmacokinetics . . . . . 27
3.2 Some Pharmacokinetic Models . . . . . 33
3.2.1 One compartment models . . . . . 34
3.2.2 Two compartments models . . . . . 41
3.3 Inference . . . . . 45
3.4 Some Pharmacodynamic Models . . . . . 48
3.4.1 Inmediate response models . . . . . 48
3.4.2 Turnover response models . . . . . 50

# Chapter 1

# Introduction

Population Pharmacokinetic / Pharmacodymanic analysis using Nonlinear Mixed Effects Modeling is an approach that has gained importance in pharmacometrics. One of the principal purposes of this analysis, as we will see later, is to keep the concentration of a drug high enough to produce a desirable response, but low enough to avoid toxicity. For this objective, studies in population are made.

Nonlinear Mixed Effects Modeling is the primary tool employed on this analysis, so the first objects that are introduced in this work are Nonlinear Models.

This work is structured in two chapters apart from this introduction. As mentioned before, Nonlinear Models are introduced in Chapter 2. Then, with this element defined, and after we define Nonlinear Fixed and Random Effects Models, Nonlinear Mixed Effects Models are presented.

After that, and also in Chapter 2, inference on those models is made. We expose some methods of inference that can be made, as Methods Based on Individuals Estimates or Methods Based on Approximations of the Likelihood.

Then, in Chapter 3 we begin explaining what Pharmacokinetics and Pharmacodynamics are. Once we have made that, some Pharmacokinetic and Pharmacodynamic models are presented. Also, we make inference using [10] in some of those models.

# Chapter 2

# Nonlinear Mixed Effects Models

## 2.1 Nonlinear Models

In order to explain what a nonlinear model is, we must remember the definition of linear model.

Let

- $Y \in \mathbb{R}$ be a variable of responses (called dependent variable),
- $\boldsymbol{X} \in \mathbb{R}^p$ a vector of predictor variables (called independent variables or covariates),
- $\boldsymbol{\theta} \in \mathbb{R}^r$ a vector of regression parameters,
- $\varepsilon \in \mathbb{R}$ a variable of random intra-individual errors.

A model

$$Y = f(\boldsymbol{X}; \boldsymbol{\theta}) + \varepsilon,$$

is said to be linear if for any $\boldsymbol{X}$, the function

$$\boldsymbol{\theta} \mapsto f(\boldsymbol{X}; \boldsymbol{\theta})$$

is affine. In particular, such $f$ is differentiable and its partial derivatives with respect to any of the model parameters

$$\left\{ \frac{\partial f}{\partial \theta_k} \right\}_{k=1,\cdots,p}$$

are independent of the other parameters.

Otherwise, a model

$$Y = f(\boldsymbol{X}; \boldsymbol{\theta}) + \varepsilon$$

is **nonlinear** if any of the partial derivatives with respect to any of the model parameters

$$\frac{\partial f}{\partial \theta_i}$$

are dependent on any other model parameter $\theta_j \quad (j \neq i)$ or if any of the derivatives do not exist or are discontinuous.

**Example 2.1.1** *An example of nonlinear model is the $E_{max}$ pharmacodynamic model, see [2], [8] or [1].*

*It has a two dimensional parameter $\boldsymbol{\theta} = (\theta_1, \theta_2) = (E_{max}, EC_{50})$ , where*

- *$E_{max}$ is the maximal effect,*
- *$EC_{50}$ is the concentration that produces 50% of the maximal effect.*

*Define:*

- *$E$ as the observed effect,*
- *$C$ as the concentration.*

*The $E_{max}$ model is given by:*

$$Y = E = \frac{E_{max} C}{EC_{50} + C} = f(\boldsymbol{X}; \boldsymbol{\theta}) + \varepsilon,$$

*where $C \equiv \boldsymbol{X}$.*

*The partial derivatives are given by*

$$\frac{\partial Y}{\partial \theta_1} = \frac{\partial E}{\partial E_{max}} = \frac{C}{EC_{50} + C}$$

*and*

$$\frac{\partial Y}{\partial \theta_2} = \frac{\partial E}{\partial EC_{50}} = \frac{-E_{max} C}{(EC_{50} + C)^2}.$$

*We note that both partial derivatives are dependent on the another parameter, so this model is effectively a nonlinear model.*

**Example 2.1.2** *Another example of nonlinear model is*

$$f(\boldsymbol{X};\boldsymbol{\theta}) = \begin{cases} \theta_0 + \theta_1 X, & X \leq X_0, \\ \theta_2 + \theta_3 X, & X > X_0, \end{cases}$$

*see [2].*

*We must note that in those models the regression parameters themselves are not dependent on any other model parameter, but the estimates may be dependent on the value of* $X_0$*.*

*For those models, the derivative is not generally continuous at* $X = X_0$*, so they are nonlinear models.*

Models that are transformed in predictor variables are not, however, nonlinear. To illustrate that, we use the following example.

**Example 2.1.3** *Let*

$$f(\boldsymbol{X};\boldsymbol{\theta}) = \theta_0 + \theta_1\sqrt{X_1} + \theta_2 sin(X_2).$$

*be a regression model. This model is not nonlinear because the partial derivatives are not dependent on any other model parameter and the derivative is not discontinuous (in fact, this is a linear model), see [2].*

Now, we know how nonlinear models are, so we can advance and study the next section, the **Nonlinear Fixed Effects Models**.

## 2.2 Nonlinear Fixed Effects Models

Before we explain what nonlinear fixed effects models are, we are going to define what a fixed effect variable is.

A **fixed effect variable** is one where the researcher can choose the level(s) of the variable to represent the precise contrast of interest. Alternatively, **fixed effect variables** are those variables whose levels in a study represent an exhaustive set of all possible levels. Later, we will define those **fixed effects variables** more mathematically, and they will be represented as the $(q \times 1)$ vector $\boldsymbol{\beta}$.

**Example 2.2.1** *The doses of drug used in a study or the time points that blood samples are measured represent fixed effects variables, see [2].*

**Example 2.2.2** *Using both males and females in a study exhaust the possible levels of this variable, thus making sex a fixed effect, see [2].*

With fixed variables defined, we can now define **nonlinear fixed effects models** as those that only contain fixed effects variables.

So, as we have already defined nonlinear fixed effects models, we can continue to define **random** and **mixed effects models**.

## 2.3 Nonlinear Random and Mixed Effects Models

As we made before with fixed effects models, we will explain what random effects variables are, in order to define nonlinear random effects models, and then we will define what mixed effects models are.

**Random effects variables** are variables whose levels do not exhaust the set of possible levels and each level is equally representative of other levels. As we have made with fixed effects variables, we also explain later and more mathematically **random effects variables**, which will be denoted as the $(k \times 1)$ vector $\mathbf{b}_i$.

**Example 2.3.1** *The most commonly seen random effect variables in clinical research are the subjects used in an experiment, since in most cases researchers are not specifically interested in the particular set of subjects that were used in a study.*

Now we can define what nonlinear random and nonlinear mixed effects models are.

Any nonlinear model that only contains random effects variables is a **nonlinear random effects model**.

Any nonlinear model that contains both fixed and random effects variables is a **nonlinear mixed effects model**.

In the remainder of this section, we describe the nonlinear mixed effects model and situations for which it is an appropriate framework, see [4].

**Nonlinear mixed effects models** are used to embed the model describing individual trajectories in a statistical model. It formalizes knowledge and assumptions about variation in outcomes and mechanisms within and among individuals and provides a framework for inference based on repeated measurement data from $m$ individuals.

We must keep in mind that we are not interested in a descriptive model which fits the data, but rather in a mechanistic model which has some biological meaning and which is a function of some physiological parameters.

In order to develop nonlinear mixed effects models, we can use a hierarchical approach, as we will explain below. Before that, some notation is introduced.

Let $Y_{i,j}$ denote the $j^{th}$ measurement of the response of the $i^{th}$ individual under condition $t_{i,j}, i \in \{1, \cdots, m\},\ j \in \{1, \cdots, n_i\}$ (usually, the set $(t_{i,j})_{i=1,\cdots,m;j=1,\cdots,n_i}$ is the set of time points), and possible additional conditions $\boldsymbol{u}_i$ (for example, $\boldsymbol{u}_i = D_i$, where $D_i$ denotes the dose for individual $i^{th}$ at time zero). We write for brevity $\boldsymbol{X}_{i,j} = (t_{i,j}, \boldsymbol{u}_i)$, but note dependence on $t_{i,j}$ where appropriate. Assume that there may be a vector of characteristics $\boldsymbol{a}_i$ for each individual that do not change with time (for example: age, weight,...). Letting $\boldsymbol{Y}_i = (Y_{i,1}, \cdots, Y_{i,n_i})'$ it is ordinarily assumed that the triplets $(\boldsymbol{Y}_i, \boldsymbol{u}_i, \boldsymbol{a}_i)$ are independent across $i$, reflecting the belief that individuals are unrelated. Finally, let $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \cdots, \varepsilon_{i,n_i})$ be the vector of intra-individuals errors for individual $i$.

In this work, as we have mentioned above, nonlinear mixed effects models will be written as a two-stage hierarchy, as follows.

### 2.3.1 The Two-Stage Hierarchy

#### Stage 1 - Individual-Level Model

*Stage* 1 consists of writing the model

$$Y_{i,j} = f(\boldsymbol{X}_{i,j}; \boldsymbol{\theta}_i) + \varepsilon_{i,j}, \qquad j \in \{1, 2, \cdots, n_i\}, \tag{2.1}$$

where $\boldsymbol{\theta}_i$ is the $\mathbb{R}^r$ vector of regression parameters.

In (2.1), $f$ is a function governing within-individual behavior, depending on an $(r \times 1)$ vector of parameters $\boldsymbol{\theta}_i$ specific to individual $i$.

**Example 2.3.2** *For example, in Example 2.1.1, we had that* $(E_{max_i}, EC_{50_i})' = (\theta_{1i}, \theta_{2i})' = \boldsymbol{\theta_i}$.

The intra-individuals errors $\varepsilon_{i,j}$,

$$\varepsilon_{i,j} = Y_{i,j} - f(\boldsymbol{X}_{i,j}; \boldsymbol{\theta}_i)$$

are assumed to satisfy

$$E[\varepsilon_{i,j} | \boldsymbol{u}_i, \boldsymbol{\theta}_i] = 0 \ \forall j.$$

We will say more about other properties of the $\varepsilon_{i,j}$ shortly.

**Stage 2 - Population Model**

*Stage* 2 consists of writing

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\mathbf{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \qquad i \in \{1, 2, \cdots, m\}; \tag{2.2}$$

where $\boldsymbol{d}$ is an $r$-dimensional function describing relationship between $\boldsymbol{\theta}_i$ and $\boldsymbol{a}_i$ in terms of

- $\boldsymbol{\beta}$, which is an $\mathbb{R}^q$ fixed parameter (called "fixed effects").
- $\boldsymbol{b}_i$, which is a $\mathbb{R}^k$ random parameter (called "random effects"), associated with individual $i$.

that characterize how elements of $\boldsymbol{\theta}_i$ vary across individual due both to

- Systematic association with individual attributes $\boldsymbol{a}_i$.
- Unexplained variation in the population of individuals, represented by $\boldsymbol{b}_i$.

In this work we will assume for each $i \in \{1, 2, \cdots, m\}$ that the distribution of the $\boldsymbol{b}_i$ conditional on $\boldsymbol{a}_i$ does not depend on $\boldsymbol{a}_i$, so:

- $E[\boldsymbol{b}_i | \boldsymbol{a}_i] = E[\boldsymbol{b}_i] = 0$.
- $Var[\boldsymbol{b}_i | \boldsymbol{a}_i] = Var[\boldsymbol{b}_i] = \boldsymbol{G}$, where $\boldsymbol{G}$ is a covariance matrix that is the same $\forall i$ which characterizes the magnitude of unexplained variation in the elements of $\boldsymbol{\theta}_i$ and associations among them.

- $\boldsymbol{b}_i \sim (\mathbf{0}, \boldsymbol{G})$. A popular assumption is $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{G})$.

(2.2) allows nonlinear specifications (in $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$) for elements of $\boldsymbol{\theta}_i$.

However, a common special case of (2.2) is that of a linear relationship between $\boldsymbol{\theta}_i$ and fixed and random effects as in usual, empirical statistical linear modeling, i.e.,

$$\boldsymbol{\theta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i \tag{2.3}$$

where $\boldsymbol{A}_i$ is a design matrix depending on elements of $\boldsymbol{a}_i$ and $\boldsymbol{B}_i$ is a design matrix typically involving only zeroes and ones allowing some elements of $\boldsymbol{\theta}_i$ to have no associated random effect.

**Completing the Nonlinear Mixed Model. Within-individual variation**

To complete the full nonlinear mixed effects model, a specification for variation within individuals is required. In this work we discuss this feature in some detail, focusing on phenomena taking place within the $i^{th}$ single individual. Our discussion focuses on model (2.1), but the same considerations are relevant for linear modeling.

According to the individual model (2.1), one has that

$$E[Y_{i,j}|\mathbf{u}_i, \boldsymbol{\theta_i}] = f(t_{i,j}, \mathbf{u}_i; \boldsymbol{\theta_i}),$$

so that $f$ represents what happens on average for the $i^{th}$ subject.

Each $Y_{i,j}$ observed is the sum of one realized profile and one set of measurement errors at intermittent time points $t_{i,j}$, formalized by writing (2.1) as

$$Y_{i,j} = f(t_{i,j}, \mathbf{u}_i; \boldsymbol{\theta_i}) + \varepsilon_{R;i,j} + \varepsilon_{M;i,j}, \tag{2.4}$$

where $\varepsilon_{i,j}$ has been partitioned into

- $\varepsilon_{R;i,j}$, the commonly named Realization Deviation Process (this deviation is due to the particular realization observed)
- $\varepsilon_{M;i,j}$, the Measurement Error Deviation Process (this deviation is due to a possible measurement error)

at each time point $t_{i,j}$.

In (2.4), the actual realized response at time point $t_{i,j}$, if it could be observed without error, is thus

$$f(t_{i,j}, \mathbf{u}_i; \boldsymbol{\theta_i}) + \varepsilon_{R;i,j}.$$

We may think of (2.4) as following from a within-subject stochastic process of the form

$$Y_i(t, \mathbf{u}_i) = f(t, \mathbf{u}_i; \boldsymbol{\theta_i}) + \varepsilon_{R;i}(t, \mathbf{u}_i) + \varepsilon_{M;i}(t, \mathbf{u}_i) \tag{2.5}$$

with

$$E[\varepsilon_{R;i}(t, \mathbf{u}_i)|\mathbf{u}_i, \boldsymbol{\theta_i}] = E[\varepsilon_{M;i}(t, \mathbf{u}_i)|\mathbf{u}_i, \boldsymbol{\theta_i}] = 0$$

where

- $\varepsilon_{R;i}(t_{i,j}, \mathbf{u}_i) = \varepsilon_{R;i,j}$
- $\varepsilon_{M;i}(t_{i,j}, \mathbf{u}_i) = \varepsilon_{M;i,j}$

and hence

$$E[\varepsilon_{R;i,j}|\mathbf{u}_i, \boldsymbol{\theta_i}] = E[\varepsilon_{M;i,j}|\mathbf{u}_i, \boldsymbol{\theta_i}] = 0.$$

Assumptions on $\varepsilon_{R,i}(t, \mathbf{u}_i)$ and $\varepsilon_{M,i}(t, \mathbf{u}_i)$ lead to a model for $Var[\boldsymbol{\varepsilon_i}|\mathbf{u}_i, \boldsymbol{\theta_i}]$ and hence for $Var[\mathbf{Y}_i|\mathbf{u}_i, \boldsymbol{\theta_i}]$:

- **Realization deviation process**
  It is natural to expect $\varepsilon_{R,i}(t, \mathbf{u}_i)$ and $\varepsilon_{R,i}(s, \mathbf{u}_i)$ at different close times points $t$ and $s$ to be positively correlated, and realizations in time points far apart bear little relation to the other. One possible way of expressing this is as

  $$corr[\varepsilon_{R,i}(t, \boldsymbol{u}_i), \varepsilon_{R,i}(s, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = e^{(-\rho|t-s|)}, \quad \rho \geq 0.$$

  We can also assume that variation of realizations are of similar magnitude over time and individuals, e.g.

  $$Var[\varepsilon_{R,i}(t, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \sigma_R^2 \geq 0 \quad (\sigma_R^2 \text{ constant } \forall t)$$

  or that variation depends on $f(t, \boldsymbol{u}_i, \boldsymbol{\theta_i})$, e.g.

  $$Var[\varepsilon_{R,i}(t, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \sigma_R^2 \left(f(t, \boldsymbol{u}_i, \boldsymbol{\theta_i})\right)^{2\delta}$$

  Letting

  $$\boldsymbol{\varepsilon_{R;i}} = \begin{pmatrix} \varepsilon_{R;i,1} \\ \vdots \\ \varepsilon_{R;i,n_i} \end{pmatrix},$$

under specific variance and autocorrelation functions, define

$$\mathbf{T}_i(\boldsymbol{u}_i, \boldsymbol{\theta_i}, \boldsymbol{\delta})$$

to be the $(n_i \times n_i)$ diagonal matrix with diagonal elements

$$Var[\varepsilon_{R;i,j}|\boldsymbol{u}_i, \boldsymbol{\theta_i}]$$

depending on parameters $\boldsymbol{\delta}$, say, and

$$\boldsymbol{\Gamma}_i(\boldsymbol{\rho})$$

the $(n_i \times n_i)$ matrix with $(j, j')$ elements

$$corr[\varepsilon_{R;i,j}, \varepsilon_{R;i,j'}|\boldsymbol{u}_i, \boldsymbol{\theta_i}]$$

depending on parameters $\boldsymbol{\rho}$, then

$$Var[\boldsymbol{\varepsilon_{R;i}}|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \mathbf{T}_i^{1/2}(\boldsymbol{u}_i, \boldsymbol{\theta_i}, \boldsymbol{\delta})\boldsymbol{\Gamma}_i(\boldsymbol{\rho})\mathbf{T}_i^{1/2}(\boldsymbol{u}_i, \boldsymbol{\theta_i}, \boldsymbol{\delta}) \qquad (2.6)$$

is the covariance matrix $\in \mathbb{R}^{n_i \times n_i}$ for $\boldsymbol{\varepsilon_{R;i}}$.

- **Measurement error deviation process**
  As we know, all measuring devices commit haphazard errors, so we can assume there is no correlation between $\varepsilon_{M,i}(t, \boldsymbol{u}_i)$ and $\varepsilon_{M,i}(s, \boldsymbol{u}_i)$, if $t \neq s$:
  $$corr[\varepsilon_{M,i}(t, \boldsymbol{u}_i), \varepsilon_{M,i}(s, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = 0, \quad t \neq s.$$
  We can assume that the magnitude of errors is similar regardless of the level, e.g.
  $$Var[\varepsilon_{M,i}(t, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \sigma_M^2 \geq 0 \quad (\sigma_M^2 \text{ constant } \forall t)$$
  or assume that the magnitude changes with the level, e.g
  $$Var[\varepsilon_{M,i}(t, \boldsymbol{u}_i)|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \sigma_M^2 \left(f(t, \boldsymbol{u}_i, \boldsymbol{\theta_i})\right)^{2\lambda}$$
  Defining
  $$\boldsymbol{\varepsilon}_{M;i} = \begin{pmatrix} \varepsilon_{M;i,1} \\ \vdots \\ \varepsilon_{M;i,n_i} \end{pmatrix},$$
  the covariance matrix of $\boldsymbol{\varepsilon}_{M;i}$ would thus be diagonal with diagonal elements
  $$Var[\varepsilon_{M;i,j}|\boldsymbol{u}_i, \boldsymbol{\theta_i}];$$

i.e.,

$$Var[\boldsymbol{\varepsilon}_{M;i}|\boldsymbol{u}_i, \boldsymbol{\theta_i}] = \boldsymbol{\Lambda}_i(\boldsymbol{u}_i, \boldsymbol{\theta_i}, \boldsymbol{\lambda}), \tag{2.7}$$

a diagonal matrix $\in R^{n_i \times n_i}$ depending on a parameter $\boldsymbol{\lambda}$. However, (2.7) may also depend on $\boldsymbol{\theta_i}$.

For constant measurement error variance, $\boldsymbol{\Lambda}_i(\boldsymbol{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\lambda}) = \sigma_M^2 \boldsymbol{I}_{n_i}$ and $\boldsymbol{\lambda} = \sigma_M^2$.

A common assumption is that the realization and measurement error processes in (2.5) are conditionally independent, which implies

$$Var[\mathbf{Y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_i] = Var[\boldsymbol{\varepsilon}_{R;i}|\boldsymbol{u}_i, \boldsymbol{\theta}_i] + Var[\boldsymbol{\varepsilon}_{M;i}|\boldsymbol{u}_i, \boldsymbol{\theta}_i]. \tag{2.8}$$

If such independence were not thought to hold, $Var[\mathbf{Y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_i]$ would also involve a conditional covariance term. Thus, combining the foregoing considerations and adopting (2.8) as is customary, a general representation of the components of within-subject variation is

$$\begin{gathered} Var[\mathbf{Y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_i] = \\ = \mathbf{T}_i^{1/2}(\mathbf{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\delta})\boldsymbol{\Gamma}_i(\boldsymbol{\rho})\mathbf{T}_i^{1/2}(\mathbf{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\delta}) + \boldsymbol{\Lambda}_i(\mathbf{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\lambda}) = \\ = \mathbf{R}_i(\mathbf{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}), \end{gathered} \tag{2.9}$$

where $\boldsymbol{\xi} = (\boldsymbol{\delta}', \boldsymbol{\rho}', \boldsymbol{\lambda}')'$, which are common to all individuals, fully describes the overall pattern of within-individual variation.

The representation (2.9) provides a framework for thinking about sources that contribute to the overall pattern of within-individual variation. It is common in practice to adopt models that are simplifications of (2.9).

In some contexts, measurement error may be taken to be the primary source of variation about $f$, and it is a common approximation in pharmacokinetics, see [6].

In fact, if

$$Var[\varepsilon_{R;i,j}|\boldsymbol{u}_i, \boldsymbol{\theta}_i] << Var[\varepsilon_{M;i,j}|\boldsymbol{u}_i, \boldsymbol{\theta}_i],$$

one might regard the first term in (2.9) as negligible. This leads to a standard model in this application, where

$$\boldsymbol{R}_i(\boldsymbol{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) = \boldsymbol{\Lambda}_i(\boldsymbol{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\lambda}),$$

with $\boldsymbol{\Lambda}_i(\boldsymbol{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\lambda})$ a diagonal matrix with diagonal elements $\sigma^2(f(t_{i,j}, \boldsymbol{u}_i, \boldsymbol{\theta}_i))^{2\lambda}$ for some $\boldsymbol{\lambda} = (\sigma^2, \lambda)'$; often $\lambda = 1$.

### 2.3.2 Summary

We are now in a position to summarize the basic nonlinear mixed effects model. Let

$$\boldsymbol{f}_i(\boldsymbol{u}_i; \boldsymbol{\theta}_i) = \begin{pmatrix} f(\boldsymbol{X}_{i,1}; \boldsymbol{\theta}_i) \\ \vdots \\ f(\boldsymbol{X}_{i,n_i}; \boldsymbol{\theta}_i) \end{pmatrix},$$

and let $\boldsymbol{z}_i = (\boldsymbol{u}_i', \boldsymbol{a}_i')'$ summarize all covariate information on subject $i$. Then, we may write the model in (2.1) and (2.2) succinctly as

Stage 1:

$$\begin{aligned} E[\boldsymbol{Y}_i|\boldsymbol{z}_i, \boldsymbol{b}_i] &= \boldsymbol{f}_i(\boldsymbol{u}_i; \boldsymbol{\theta}_i) = \boldsymbol{f}_i(\boldsymbol{z}_i; \boldsymbol{\beta}, \boldsymbol{b}_i) \\ Var[\boldsymbol{Y}_i|\boldsymbol{z}_i, \boldsymbol{b}_i] &= \boldsymbol{R}_i(\boldsymbol{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) = \boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi}) \end{aligned} \tag{2.10}$$

Stage 2:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \qquad \boldsymbol{b}_i \sim (\boldsymbol{0}, \boldsymbol{G}) \tag{2.11}$$

In (2.10), dependence of $\boldsymbol{f}_i$ and $\boldsymbol{R}_i$ on the covariates $\boldsymbol{a}_i$ and fixed and random effects through $\boldsymbol{\theta}_i$ is emphasized. This model represents individual behavior conditional on $\boldsymbol{\theta}_i$ and hence on $\boldsymbol{b}_i$, the random component in (2.11). In (2.11), we assume that the distribution of $\boldsymbol{b}_i|\boldsymbol{a}_i$ does not depend on $\boldsymbol{a}_i$, so that all $\boldsymbol{b}_i$ have common distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{G}$.

*Within-individual correlation:*

The nonlinear model (2.10)-(2.11) implies a model for the marginal mean and covariance matrix of $\boldsymbol{Y}_i$ given all covariates $\boldsymbol{z}_i$; i.e. averaged across population.

Letting $F_b(\boldsymbol{b}_i)$ denote the cumulative distribution function of $\boldsymbol{b}_i$, we have

$$E[\boldsymbol{Y}_i|\boldsymbol{z}_i] = \int \boldsymbol{f}_i(\boldsymbol{z}_i;\boldsymbol{\beta},\boldsymbol{b}_i)dF_b(\boldsymbol{b}_i)$$

$$Var[\boldsymbol{Y}_i|\boldsymbol{z}_i] = E[\boldsymbol{R}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i,\boldsymbol{\xi})|\boldsymbol{z}_i] + Var[\boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i)|\boldsymbol{z}_i], \quad (2.12)$$

where expectation and variance are with respect to the distribution of $\boldsymbol{b}_i$. In (2.12), $E[\boldsymbol{Y}_i|\boldsymbol{z}_i]$ characterizes the typical response profile among individuals with covariates $\boldsymbol{z}_i$.

$Var[\boldsymbol{Y}_i|\boldsymbol{z}_i]$ involves two terms as we have seen in (2.12):

- $E[\boldsymbol{R}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i,\boldsymbol{\xi})|\boldsymbol{z}_i]$, which averages realization and measurement variation that occur within individuals across individuals having covariates $\boldsymbol{z}_i$
- $Var[\boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i)|\boldsymbol{z}_i]$, which describes how inherent trajectories vary among individuals which have the same $\boldsymbol{z}_i$.

Note that $E[\boldsymbol{R}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i,\boldsymbol{\xi})|\boldsymbol{z}_i]$, is a diagonal matrix only if $\boldsymbol{\Gamma}_i(\boldsymbol{\rho})$ in (2.9), reflecting correlation due to within-individual realizations, is an identity matrix. However, $Var[\boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i)|\boldsymbol{z}_i]$, has non-zero off-diagonal elements in general due to common dependence of all elements of $\boldsymbol{f}_i$ on $\boldsymbol{b}_i$. Thus, correlation at the marginal level is always expected due to variation among individuals, while there is correlation from within-individual sources only if serial associations among intra-individual realizations are nonnegligible. In general, then, both terms contribute to the overall pattern of correlation among responses on the same individual represented in $Var[\boldsymbol{Y}_i|\boldsymbol{z}_i]$.

In many applications, the effect of within-individual serial correlation reflected in the first term of $Var[\boldsymbol{Y}_i|\boldsymbol{z}_i]$ is dominated by that from among individual variation in $Var[\boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\boldsymbol{b}_i)|\boldsymbol{z}_i]$. This explains why many published applications of nonlinear mixed models adopt simple, diagonal models for $\boldsymbol{R}_i(\boldsymbol{u}_i,\boldsymbol{\theta}_i,\boldsymbol{\xi})$ that emphasize measurement error.

## 2.4 Inference in Nonlinear Mixed Effects Models

In this section we will introduce the theory behind nonlinear mixed effects models. Modeling and estimation of model parameters will be discussed, as

the relationship between the covariate and the dependent variable is modeled.

### 2.4.1 Inferential Approaches

The main inferential objectives are:

- Determine an appropriate population model $\boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ and inference on elements of $\boldsymbol{\beta}$ as central interests.
- Inference on $\boldsymbol{G}$ and, in particular, on diagonal elements.
- Inference on $\boldsymbol{\theta}_i$ and $f(t_0, \boldsymbol{u}_i, \boldsymbol{\theta}_i)$ at some specific time point $t_0$, for $i \in \{1, 2, \cdots, m\}$.

The nonlinear mixed effects model (2.10)-(2.11) is a so called *subject-specific* model. The distinction between *subject-specific* and *population-averaged* (or marginal) models may not be important for linear mixed effects models, but it is critical under nonlinearity. A *population-averaged* model assumes that interest focuses on parameters that describe the marginal distribution of $\mathbf{Y}_i$ given covariates $\mathbf{z}_i$. From the discussion following (2.12), if $E(\mathbf{Y}_i|\mathbf{z}_i)$ were modeled directly as a function of $\mathbf{z}_i$ and a parameter $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ would represent the parameter corresponding to the *typical response profile* among individuals with covariates $\mathbf{z}_i$. This is to be contrasted with the meaning of $\boldsymbol{\beta}$ in (2.11) as the *typical value* of individual-specific parameters $\boldsymbol{\theta}_i$ in the population.

Consider first linear models. A linear subject-specific model with second stage

$$\boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i$$

as in (2.3), for design matrix $\mathbf{A}_i$ depending on $\mathbf{a}_i$, and first stage

$$E[Y_{i,j}|\mathbf{u}_i, \boldsymbol{\theta}_i] = \mathbf{U}_i\boldsymbol{\theta}_i,$$

where $\mathbf{U}_i$ is a design matrix depending on the $t_{i,j}$ and $\mathbf{u}_i$, leads to the linear mixed effects model

$$E[\mathbf{Y}_i|\mathbf{z}_i, \mathbf{b}_i] = \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i) = \widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

for

$$\widetilde{\mathbf{X}}_i = \mathbf{U}_i\mathbf{A}_i$$

and

$$\mathbf{Z}_i = \mathbf{U}_i\mathbf{B}_i$$

where $\widetilde{\mathbf{X}}_i$ thus depends on $\mathbf{z}_i$. From (2.12), this model implies that

$$E[\mathbf{Y}_i|\mathbf{z}_i] = \int (\widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)dF_b(\mathbf{b}_i) = \widetilde{\mathbf{X}}_i\boldsymbol{\beta},$$

as $E[\mathbf{b}_i] = 0$. Thus, in a linear subject-specific model, $\boldsymbol{\beta}$ fully characterizes both the typical value os $\boldsymbol{\theta}_i$ and the typical response profile, so that the interpretation is valid. Here, then, postulating the linear subject-specific model is equivalent to postulating a population-averaged model of the form $E[\mathbf{Y}_i|\mathbf{z}_i] = \widetilde{\mathbf{X}}_i\boldsymbol{\beta}$ directly, in that both approaches yield the same representation of the marginal mean and hence allow the same interpretation of $\boldsymbol{\beta}$. Consequently, the distinctions between subject-specific and population-averaged approaches have not generally been of concern in the literature on linear modeling.

For nonlinear models, however, this is not longer the case. For instance, suppose that $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$, and consider a subject-specific model of the form in (2.10) and (2.11) for some function $f$ nonlinear in $\boldsymbol{\theta}_i$ and hence $\mathbf{b}_i$. Then, from (2.12), the implied marginal mean is

$$E[\mathbf{Y}_i|\mathbf{z}_i] = \int \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)p(\mathbf{b}_i; \mathbf{G})d\mathbf{b}_i, \tag{2.13}$$

where $p(\mathbf{b}_i; \mathbf{G})$ is the $\mathcal{N}(\mathbf{0}, \mathbf{G})$ density. For nonlinear $f$, this integral is clearly intractable and $E[\mathbf{Y}_i|\mathbf{z}_i]$ is an expression that may not be available in a closed form and depends on both $\boldsymbol{\beta}$ and $\mathbf{G}$ in general. As a result, if we start with a nonlinear subject-specific model, the implied population-averaged marginal mean model involves both the typical value of $\boldsymbol{\theta}_i$ ($\boldsymbol{\beta}$) and $\mathbf{G}$. Accordingly, $\boldsymbol{\beta}$ does not fully characterize the typical response profile and thus cannot enjoy both interpretations. Conversely, if we were to take a population-averaged approach and model the marginal mean directly as a function of $\mathbf{z_i}$ and a parameter $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ would indeed have the interpretation of describing the typical response profile. But it seems unlikely that it could also have the interpretation as the typical value of specific parameters $\boldsymbol{\theta}_i$ in a subject-specific model. Thus, for nonlinear models, the interpretation of $\boldsymbol{\beta}$ in subject-specific and population-averaged models cannot be the same in general. The implication is that the modeling approach must be carefully considered to ensure that the interpretation of $\boldsymbol{\beta}$ coincides with the questions of scientific interest.

### 2.4.2 Inference

A number of inferential methods for the nonlinear mixed effects model are now in common use, see [4] and [7]. We provide a brief overview that can be seen below, beginning with *Maximum Likelihood*.

Maximum Likelihood

A natural starting point for inference is *Maximum Likelihood*. This is only a starting point here because the analytical intractability of likelihood inference has motivated many approaches on approximations.

The individual model (2.10) along with an assumption on the distribution of $\boldsymbol{Y}_i$ given $(\boldsymbol{z}_i, \boldsymbol{b}_i)$ yields a conditional density $p(\mathbf{Y}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi})$, say; the ubiquitous choice is the normal. Under the popular assumption that $\mathbf{R}_i(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i, \boldsymbol{\xi})$ is diagonal, the density may be written as the product of $m$ contributions $p(Y_{i,j}|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\xi})$. Under this condition, the lognormal has also been used. At Stage 2, (2.11), adopting independence of $\mathbf{b}_i$ and $\mathbf{a}_i$, one assumes a $k$-variate density $p(\mathbf{b}_i, \mathbf{G})$ for $\mathbf{b}_i$. As with other mixed models, normality is standard. With these specifications, the joint density of $(\boldsymbol{Y}_i, \boldsymbol{b}_i)$ given $\mathbf{z}_i$ is

$$p(\boldsymbol{Y}_i, \mathbf{b}_i|\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G}) = p(\boldsymbol{Y}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi})p(\mathbf{b}_i; \mathbf{G}) \tag{2.14}$$

A likelihood for $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{G}$ may be based on the joint density of the observed data $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m$ given $\mathbf{z}_i$,

$$\prod_{i=1}^{m} \int p(\boldsymbol{Y}_i, \mathbf{b}_i|\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G}) d\mathbf{b}_i = \prod_{i=1}^{m} \int p(\boldsymbol{Y}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi}) p(\mathbf{b}_i; \mathbf{G}) d\mathbf{b}_i \tag{2.15}$$

by independence across $i$. Nonlinearity implies that the $m$ $k$-dimensional integrations in (2.15) generally cannot be done in a closed form; thus, iterative algorithms to maximize (2.15) in $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{G}$ require a way to handle these integrals. Although numerical techniques for evaluation of an integral are available, these can be computationally expensive when performed at each internal iteration of the algorithm. Hence, many approaches to fitting (2.10)-(2.11) are instead predicated on analytical approximations.

Methods Based On Individual Estimates

If the $n_i$ are sufficiently large, an intuitive approach is to summarize the responses $\boldsymbol{Y}_i$ for each $i$ through individual-specific estimates $\hat{\boldsymbol{\theta}}_i$ and then

use these as the basis for inference on $\boldsymbol{\beta}$ and $\boldsymbol{G}$. In particular, viewing the conditional moments in (2.10) as functions of $\boldsymbol{\theta}_i$, i.e.,

$$E[\mathbf{Y}_i|\mathbf{u}_i, \boldsymbol{\theta}_i] = f(\mathbf{X}_{i,j}, \boldsymbol{\theta}_i),$$

$$Var[\mathbf{Y}_i|\mathbf{u}_i, \boldsymbol{\theta}_i] = \mathbf{R}_i(\mathbf{u}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}),$$

fit the model specified by these moments for each individual.

Usual large-sample theory implies that the individual estimators $\hat{\boldsymbol{\theta}}_i$ are asymptotically normal. Each individual is treated separately, so the theory may be viewed as applying conditionally on $\boldsymbol{\theta}_i$ for each $i$, yielding

$$\hat{\boldsymbol{\theta}}_i|\mathbf{u}_i, \boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\theta}_i, \boldsymbol{C}_i).$$

Because of the nonlinearity of $f$ in $\boldsymbol{\theta}_i$, $\mathbf{C}_i$ depends on $\boldsymbol{\theta}_i$ in general, so $\mathbf{C}_i$ is replaced by $\hat{\mathbf{C}}_i$ in practice, where $\hat{\boldsymbol{\theta}}_i$ is substituted.

To see how this is exploited for inference on $\boldsymbol{\beta}$ and $\mathbf{G}$, consider the linear second-stage model (2.3); the same developments apply to any general $\boldsymbol{d}$ in (2.11). The asymptotic result may be expressed alternatively as

$$\boldsymbol{\theta}_i^* \approx \boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i^* = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i^*,$$

$$\boldsymbol{\varepsilon}_i^*|\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{C}}_i), \quad \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{G}) \tag{2.16}$$

where $\hat{\mathbf{C}}_i$ is treated as known for each $i$, so that the $\boldsymbol{\varepsilon}_i^*$ do not depend on $\mathbf{b}_i$. This is the form of a linear mixed effects model with known error covariance matrix $\hat{\mathbf{C}}_i$, which suggest using standard techniques for fitting such models to estimate $\boldsymbol{\beta}$ and $\mathbf{G}$. Some authors propose use of the *EM algorithm* in the case $k = p$ and $\mathbf{B}_i = \mathbf{I}_p$. Alternatively, it is possible to use linear mixed model software to fit (2.16).

**EM Algorithm**

The EM algorithm is a procedure to compute the maximum likelihood estimates in the presence of missing or hidden data.

Each iteration of the EM algorithm consists of two processes:

1. the E-step, in which the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation,

2. the M-step, in which the likelihood function is maximized under the assumption that the missing data are known. The estimate of the missing data from the E-step are used instead of the actual missing data.

Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration. For more details, see [3].

In the case treated in this work, the EM-algorithm can be specified as follows, see [9]. At $(k+1)^{th}$ iteration,

1. E-step: Update the estimates of $\boldsymbol{\theta}_i, \quad i = 1, 2, \cdots, m$ as

$$\hat{\boldsymbol{\theta}}_{i,(k+1)} = (\hat{\mathbf{C}}_i^{-1} + \hat{\mathbf{G}}_{(k)}^{-1})^{-1}(\hat{\mathbf{C}}_i^{-1}\boldsymbol{\theta}_i^* + \hat{\mathbf{G}}_{(k)}^{-1}\mathbf{A}_i\hat{\boldsymbol{\beta}}_{(k)}), \quad i = 1, 2, \cdots, m.$$

2. M-step: Update the estimates of the population parameters as

$$\hat{\boldsymbol{\beta}}_{(k+1)} = \left(\sum_{i=1}^{m} \mathbf{A}_i'\hat{\mathbf{G}}_{(k)}^{-1}\mathbf{A}_i\right)^{-1} \sum_{i=1}^{m} \mathbf{A}_i'\hat{\mathbf{G}}_{(k)}^{-1}\hat{\boldsymbol{\theta}}_{i,(k+1)},$$

$$\hat{\mathbf{G}}_{(k+1)} = m^{-1}\sum_{i=1}^{m}(\hat{\boldsymbol{\theta}}_{i,(k+1)} - \mathbf{A}_i\hat{\boldsymbol{\beta}}_{(k+1)})(\hat{\boldsymbol{\theta}}_{i,(k+1)} - \mathbf{A}_i\hat{\boldsymbol{\beta}}_{(k+1)})' +$$

$$+m^{-1}\sum_{i=1}^{m}(\hat{\mathbf{C}}_i^{-1} + \hat{\mathbf{G}}_{(k)}^{-1})^{-1}$$

Note that as starting values for the EM algorithm one can use the following estimates:

- $$\hat{\boldsymbol{\beta}}_{(0)} = \left(\sum_{i=1}^{m} \mathbf{A}_i'\mathbf{A}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{A}_i'\boldsymbol{\theta}_i^*\right),$$

- $$\hat{\mathbf{G}}_{(0)} = (m-1)^{-1}\sum_{i=1}^{m}(\boldsymbol{\theta}_i^* - \mathbf{A}_i\hat{\boldsymbol{\beta}}_{(0)})(\boldsymbol{\theta}_i^* - A_i\hat{\boldsymbol{\beta}}_{(0)})'$$

This algorithm is iterated until the M-step converges.

This last point is critical when the $n_i$ are not large. Although this provides rich information for building population models $\boldsymbol{d}$, there are insufficient data to fit the pharmacokinetic model $f$ to any one subject. Implementation of (2.15) in principle imposes no requirements on the magnitude of the $n_i$. Thus, an attractive strategy is instead to approximate (2.15) in a way that intractable integration is avoided. In particular, for each $i$, an approximation to

$$p(\mathbf{Y}_i|\mathbf{z}_i;\boldsymbol{\beta},\boldsymbol{\xi},\mathbf{G}) = \int p(\mathbf{Y}_i|\mathbf{z}_i,\mathbf{b}_i;\boldsymbol{\beta},\boldsymbol{\xi})p(\mathbf{b}_i;\mathbf{G})d\mathbf{b}_i$$

is obtained.

First Order Methods

An approach is motivated by letting $\mathbf{R}_i^{1/2}$ be the Cholesky decomposition of $\mathbf{R}_i$ and writing (2.10)-(2.11) as

$$\mathbf{Y}_i = \boldsymbol{f}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{b}_i) + \mathbf{R}_i^{1/2}(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{b}_i,\boldsymbol{\xi})\boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i|\mathbf{z}_i,\mathbf{b}_i \sim (\mathbf{0},\mathbf{I}_{n_i}). \tag{2.17}$$

As nonlinearity in $\mathbf{b}_i$ causes the difficulty for integration in (2.15), it is natural to consider a linear approximation. A Taylor series of (2.17) about $\mathbf{b}_i = \mathbf{0}$ to linear terms, disregarding the term involving $\mathbf{b}_i\boldsymbol{\epsilon}_i$ as small and letting

$$\mathbf{Z}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{b}^*) = \frac{\partial}{\partial \mathbf{b}_i}\left(\boldsymbol{f}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{b}_i)\right)|_{b_i=b^*}$$

leads to

$$\mathbf{Y}_i \approx \boldsymbol{f}(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0}) + \mathbf{Z}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0})\mathbf{b}_i + \mathbf{R}_i^{1/2}(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0},\boldsymbol{\xi})\boldsymbol{\epsilon}_i \tag{2.18}$$

$$E[\mathbf{Y}_i|\mathbf{z}_i] \approx \boldsymbol{f}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0}),$$
$$Var[\mathbf{Y}_i|\mathbf{z}_i] \approx \mathbf{Z}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0})\mathbf{G}\mathbf{Z}_i'(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0}) + \mathbf{R}_i(\mathbf{z}_i,\boldsymbol{\beta},\mathbf{0},\boldsymbol{\xi}). \tag{2.19}$$

When $p(\mathbf{Y}_i|\mathbf{z}_i,\mathbf{b}_i;\boldsymbol{\beta},\boldsymbol{\xi})$ in (2.15) is a normal density, (2.18) amounts to approximating it by another normal density whose mean and covariance matrix are linear in and free of $\mathbf{b}_i$, respectively. If $p(\mathbf{b}_i;\mathbf{G})$ is also normal, the integral is analytically calculable analogous to a linear mixed model and yields a $n_i$-variate normal density $p(\mathbf{Y}_i|\mathbf{z}_i;\boldsymbol{\beta},\boldsymbol{\xi},\mathbf{G})$ for each $i$ with mean and covariance matrix (2.19).

This suggests the proposal to estimate $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\mathbf{G}$ by jointly maximizing

$$\prod_{i=1}^{m} p(\mathbf{Y}_i|\mathbf{z}_i;\boldsymbol{\beta},\boldsymbol{\xi},\mathbf{G}),$$

which is equivalent to maximum likelihood under the assumption the marginal distribution $\mathbf{Y}_i|\mathbf{z}_i$ is normal with moments (2.19). The advantage is that this approximate likelihood is available in a closed form. Standard errors are obtained from the information matrix assuming the approximation is exact. As (2.19) defines an approximate marginal mean and covariance matrix for $\mathbf{Y}_i$ given $\mathbf{z}_i$, an alternative approach is to estimate $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\mathbf{G}$ by solving generalized estimating equations.

An obvious drawback of all first order methods is that the approximation may be poor, as they essentially replace

$$E[\mathbf{Y}_i|\mathbf{z}_i] = \int f(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i) p(\mathbf{b}_i; \mathbf{G})$$

by

$$f(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{0}).$$

This suggest that more refined approximation would be desirable, like *First Order Conditional Methods*.

First Order Conditional Methods

As $p(\mathbf{Y}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi})$ and $p(\mathbf{b}_i; \mathbf{G})$ are ordinarily taken as normal densities, a natural way to approximate integrals like those in (2.15) is to exploit Laplace's method, a standard technique to approximate an integral of the form $\int e^{-l(\mathbf{b})} d\mathbf{b}$ that follows from Taylor series expansion of $-l(\mathbf{b})$ about the value $\hat{\mathbf{b}}$, say, maximizing $l(\mathbf{b})$. The result is that $p(\mathbf{Y}_i|\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G})$ may be approximated by a normal density with

$$\begin{aligned} E(\mathbf{Y}_i|\mathbf{z}_i) &\approx \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i) - \mathbf{Z}_i(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i)\hat{\mathbf{b}}_i \\ Var[\mathbf{Y}_i|\mathbf{z}_i] &\approx \mathbf{Z}_i(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i)\mathbf{G}\mathbf{Z}_i'(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i) + \mathbf{R}_i(\mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\xi}) \end{aligned} \tag{2.20}$$

$$\hat{\mathbf{b}}_i = \mathbf{G}\mathbf{Z}_i'(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i)\mathbf{R}_i(\mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\xi})[\mathbf{Y}_i - \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \hat{\mathbf{b}}_i)], \tag{2.21}$$

where $\mathbf{Z}_i$ is defined as before, and $\hat{\mathbf{b}}_i$ maximizes

$$l(\mathbf{b}_i) = [\mathbf{Y}_i - \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)]'\mathbf{R}_i^{-1}(\mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\xi})[\mathbf{Y}_i - \boldsymbol{f}_i(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)] + \mathbf{b}_i'\mathbf{G}\mathbf{b}_i$$

in $\mathbf{b}_i$. In fact, $\hat{\mathbf{b}}_i$ maximizes in $\mathbf{b}_i$ the posterior density for $\mathbf{b}_i$

$$p(\mathbf{b}_i|\mathbf{Y}_i, \mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G}) = \frac{p(\mathbf{Y}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi}) p(\mathbf{b}_i; \mathbf{G})}{p(\mathbf{Y}_i|\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G})}. \tag{2.22}$$

Equations (2.20) and (2.21) suggest an iterative scheme whose essential steps are

1. Given current estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\xi}}$, $\hat{\mathbf{G}}$ and $\hat{\mathbf{b}}_i$, say, update $\hat{\mathbf{b}}_i$ by substituting these in the right hand side of (2.21)

2. Holding $\hat{\mathbf{b}}_i$ fixed, update estimation of $\boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G}$ based on the moments in (2.20)

It is well-documented by numerous authors that these first order conditional approximations work extremely well in general, even when $n_i$ are not large or the assumptions of normality that dictate the form of (2.22) on which $\hat{\mathbf{b}}_i$ is based are violated. These features and the availability of supported software have made this approach probably the most popular way to implement nonlinear mixed models in practice.

Some final remarks follow.

The methods in this section may be implemented for any $n_i$. Although they involve closed-form expressions for $p(\mathbf{Y}_i|\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{G})$ and moments (2.19) and (2.20), maximization or solution of likelihoods or estimating equations can still be computationally challenging, and selection of suitable starting values for the algorithms is essential. Results from first order methods may also be used as starting values for a more refined "conditional" fit. A common practical strategy is to first fit a simplified version of the model and use the results to suggest starting values for the intended analysis. For instance, one might take $\mathbf{G}$ to be a diagonal matrix, which can often speed convergence of the algorithms; this implies the elements of $\boldsymbol{\theta}_i$ are uncorrelated, which is usually highly unrealistic.

# Chapter 3

# Applications in Pharmacokinetics and Pharmacodynamics

The theory we have seen before is the basis of pharmacokinetics (PK) and pharmacodynamics (PD) models, since they need of nonlinear mixed effects models to be fitted.

In the following section, a brief introduction to pharmacokinetics and pharmacodynamics is presented.

## 3.1 Pharmacokinetics and Pharmacodynamics

When someone takes a drug, the body and the drug interact.

Intuitively speaking,

**Pharmacokinetics (PK)** is what the body does to the drug. Pharmacokinetics studies how the drug moves through the body, the so-called *ADME processes* (absorption, distribution and elimination (metabolism and excretion) that govern this movement and how these processes vary across subjects.

**Pharmacodynamics (PD):** is what the drug does to the body.

Pharmacokinetics and pharmacodynamics studies are necessary to answer the following questions:

- What is a good drug concentration?
- What is the therapeutic window? Is it wide or narrow? Is it the same for everyone?
- What is the relationship between response and drug concentration?

A PK/PD study is based on collecting both concentrations and response from each subject.

One of the purposes of why all this theory has been developed is that it must be kept in mind that in the body in which the drug is applied, concentrations must be kept high enough to produce a desirable response, but low enough to avoid toxicity, in a region called **therapeutic window**; so it must be found the optimal dose to get that, by studying those models.

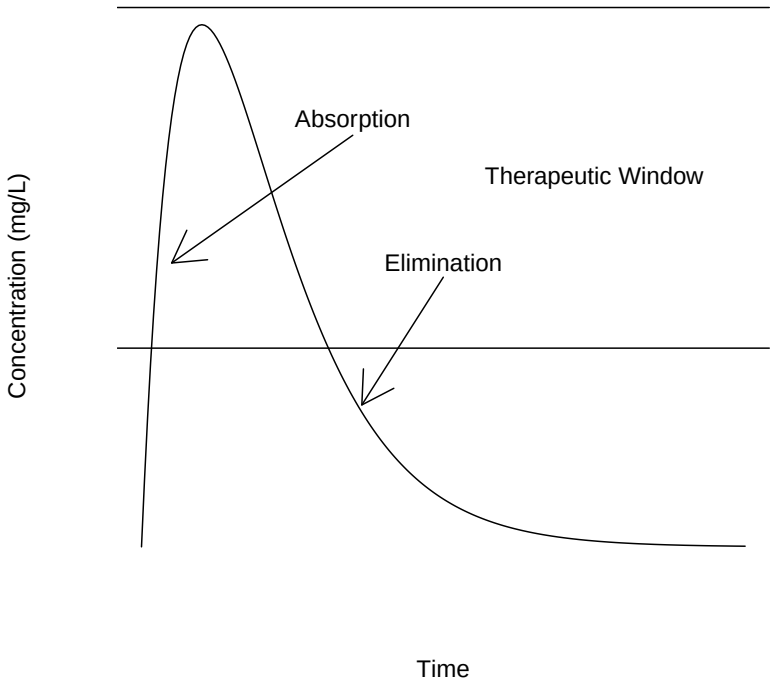


Figure 3.1: Therapeutic Window

Some basic assumptions and principles we must keep in mind are:

- There is an effect site where the drug will have its effect.
- Magnitudes of response and toxicity depend on drug concentration at the effect site.
- The drug can not be placed directly at the effect site, since it must move there.
- Concentrations at the effect site are determined by ADME processes.
- Concentrations must be kept inside the therapeutic window.
- Usually, concentrations cannot be measured at effect site directly, but in blood, plasma, etc.

### 3.1.1 Pharmacokinetics

The broad goal of pharmacokinetics analysis is to understand and characterize intra-subject ADME processes of drug absorption, distribution, metabolism and excretion governing achieved drug concentrations and how these processes vary across subjects (inter-subjects variation).

In practice, two kinds of pharmacokinetics studies are made in humans: *Intensive* and *Population* studies.

**Pharmacokinetics studies in humans (Intensive studies).**

They are characterized by the following features:

- Small number of subjects (usually healthy volunteers).
- Frequent samples over time, often following single dose.
- Usually in early stages of drug development.
- Useful for gaining initial information on typical pharmacokinetics behavior in humans and for identifying an appropriate pharmacokinetics model.
- Preclinical pharmacokinetics studies in animals are generally intensive studies.

**Pharmacokinetics studies in humans (Population studies).**

They have the following features:

- Large number of subjects (both healthy and unhealthy volunteers).
- Often in later stages of drug development or after a drug is in routine use.
- Sparse sampling over time and multiple dosing intervals.
- Extensive demographic and physiologic characteristics.
- Useful for understanding associations between patient characteristics and pharmacokinetics behavior.

**Example 3.1.1** *An example of intensive pharmacokinetics studies is the Theophylline study outlined below. Theophylline is a drug used in therapy for respiratory diseases such a COPD and asthma, see [4].*

*In this study there were 12 subjects in which the same oral dose (mg/Kg) was applied. Blood samples were collected to measure the Theophylline concentration (mg/L). These concentrations are presented in Figure 3.2.*

*In Figure 3.2, we can observe the following features:*

- *It is an intensive study, as we mentioned before: the number of subjects is quite small (only twelve volunteers), and there are many samples in a short period of time.*
- *All concentration-time profiles have similar shapes.*
- *Peak, rise and decay vary across subjects, attributable to inter-subject variation underlying pharmacokinetics behavior.*

In PK/PD many kinds of models can be obtained. To study those models, we represent the body by a simple system of compartments, which is very useful. Later in this work we will focus our study on one-compartment and two-compartment models; whose description is given below.

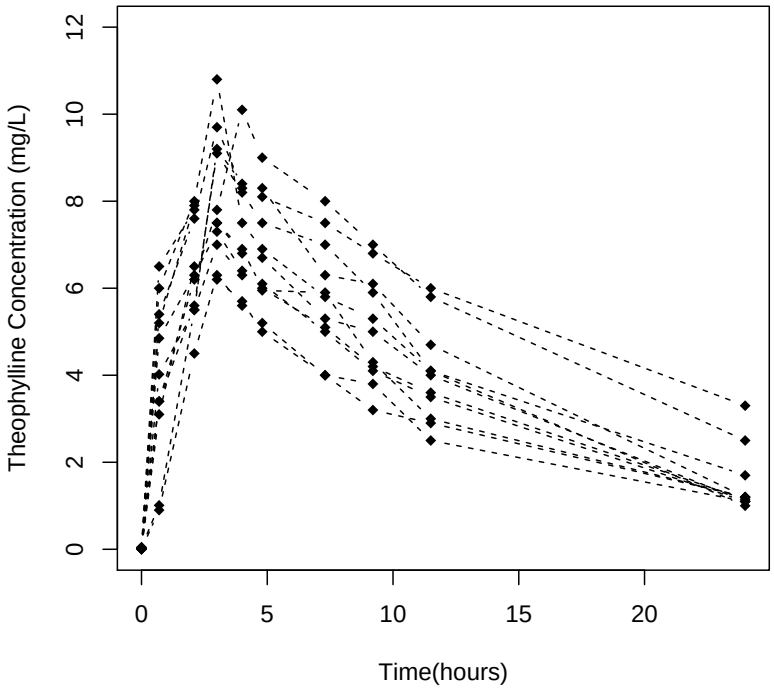


Figure 3.2: Theophylline Study

- **One compartment models**: The drug is considered to be distributed instantaneously into a unique compartment in the body. This compartment is characterized by a distribution volume. The drug input into this volume depends on the dosage regimen. The drug output from this volume is characterized by an elimination constant rate $k_e$. Several dosage regimens can be considered here:
  - Intravenous bolus injection.
  - Intravenous infusion.
  - Extravascular dose.
  - ...
- **Two compartment models**: The drug is not considered to be distributed instantaneously into a unique compartment in the body. Instead, drug first goes to a central compartment and then it goes to a peripheral compartment. After a time, there is an equilibrium of distribution of the drug in the body.

# One compartment model with first-order absorption and elimination.

Let $A(t)$ and $A_a(t)$ denote, respectively, the amount at time $t$ in the compartment and at the absorption site.

In this model it is assumed that the absorption rate $k_a$ and elimination rate $k_e$ are constants ($k_a > k_e$), with

$$k_e = \frac{Cl}{V},$$

where $V$ is the volume of the compartment and $Cl$ is the clearance.

Moreover, it is assumed that $A$, $A_a$ satisfy the linear differential equations

$$\frac{dA(t)}{dt} = k_a A_a(t) - k_e A(t), A(0) = 0 \tag{3.1}$$

$$\frac{dA_a(t)}{dt} = -k_a A_a(t), A_a(0) = D$$

Under the initial value conditions

$$A(0) = 0 \tag{3.2}$$
$$A_a(0) = D,$$

where D is the dose.

It is easily seen that the Cauchy problem given by (3.1) under the initial value (3.2) has as solution

$$A(t) = \frac{k_a D}{k_a - k_e}\left(e^{-k_e t} - e^{-k_a t}\right) \tag{3.3}$$
$$A_a(t) = De^{-k_a t}$$

So concentration at time point t in the compartment is finally given by

$$C(t) = \frac{A(t)}{V} = \frac{k_a D}{V(k_a - k_e)}\left(e^{-k_e t} - e^{-k_a t}\right)$$

Individual pharmacokinetics behavior is characterized by the three-dimensional pharmacokinetics parameter

$$\boldsymbol{\theta} = \begin{pmatrix} k_a \\ V \\ Cl \end{pmatrix}$$

Knowing the value of $\boldsymbol{\theta}$ allows one to determine concentrations achieved at any time point $t$ under different doses, because the pharmacokinetics model assumes that pharmacokinetics processes are dose-independent. So it can be used to develop dosing regimens.

Now it can be studied its therapeutic window. It must be calculated the dose ($D$) which allows concentration to be bigger than a certain concentration, say $C_m$ but lower than another concentration, $C_M$.

First of all it must be calculated the global maxima of $C(t)$. Then, at the time point $t^*$ in which $C(t)$ reaches its maximum is imposed to be $C_m \leq C(t^*) \leq C_M$.

After some simple calculations, it can be observed that such time point $t^*$ is given by the expression

$$t^* = \frac{ln\left(\frac{k_a}{k_e}\right)}{k_a - k_e}.$$

If it is desired to be $C(t^*) \leq C_M$, after some operations it is easily obtained that it must be

$$D \leq \frac{C_M V(k_a - k_e)}{k_a(e^{-k_e t^*} - e^{-k_a t^*})}$$

Equivalently, if it is wanted to be $C(t) \geq C_m$, it must be

$$D \geq \frac{C_m V(k_a - k_e)}{k_a(e^{-k_e t^*} - e^{-k_a t^*})}$$

So, in summary, in order to be concentration inside the therapeutic window, dose must be

$$D \in \left[ \frac{C_m V(k_a - k_e)}{k_a(e^{-k_e t^*} - e^{-k_a t^*})}, \frac{C_M V(k_a - k_e)}{k_a(e^{-k_e t^*} - e^{-k_a t^*})} \right].$$

**Example 3.1.2** ***Argatroban PK/PD study:***

*Note: Argatroban is an anticoagulant that is a small molecule direct thrombin inhibitor.*

*For the study described in [5], a sample of $N = 37$ subjects is used. Different constant infusion rates (doses) of 1 to 5 $\mu/Kg/min$ of argatroban are assigned. This infusion was administered by intravenous infusion for 4 hours (240 minutes). Blood samples for pharmacokinetics study were collected at $\{30, 60, 90, 115, 160,\ 200, 240, 250, 260, 275, 295, 320\}$ minutes. For pharmacodynamics study additional samples at 5-9 points, measuring activated partial thromboplastin time (aPTT, the response), were collected.*

*The effect site of this example is blood.*

***Argatroban pharmacokinetic model:***

*Let the function $(\cdot)_+$ be defined by $x_+ = 0$ if $x \leq 0$ and $x_+ = x$ if $x > 0$. Denoting by $D$, $Cl$, $T_{inf}$ respectively the dose, clearance and duration of the infusion, the Argatroban concentration $(C(t))$ over time at the effect compartment is given by the equation*

$$C(t) = \frac{D}{Cl}\left[e^{-\frac{Cl}{V}(t-t_{inf})_+} - e^{-\frac{Cl}{V}t}\right],$$

*so in this model we have a two-dimensional parameter $\boldsymbol{\theta}$, namely,*

$$\boldsymbol{\theta} = \begin{pmatrix} Cl \\ V \end{pmatrix}$$

*We must estimate typical values of $\boldsymbol{\theta}$ and how they vary in the population of subjects and understand the relationship between concentrations and response.*

***Argatroban pharmacodynamic model:***

*Argatroban pharmacodynamics model follows the $E_{max}$ model presented in Example 2.1.1.*

$$E(t) = \frac{E_{max}C(t)}{EC_{50} + C(t)}$$

The ultimate objective is to put pharmacokinetics and pharmacodynamics together to characterize the therapeutic window and how it varies across subjects and for develop dosing regimens targeting achieved concentrations leading to therapeutic response.

## 3.2 Some Pharmacokinetic Models

In this section, we introduce some examples of pharmacokinetic models, see [1]. Some of them are models that we have already presented, but we also show other new models. We use the package PKfit in R, see [10], to simulate data of all this set of models, and then we are going to do the reverse: given some data, we try to fit the data into an specific model.

Later in this work, some pharmacodynamic models are shown, but not in the same extension we present pharmacokinetic models.

The presentation of the pharmacokinetic models is organised as follows:

- First level: number of compartment
  - One compartment
  - Two compartments
- Second level: route of administration
  - Intravenous (IV) bolus
  - Infusion
  - First order absorption
  - Zero order absorption

- Third level: elimination process
  - Linear
  - Michaelis-Menten
- Fourth level: existence of a lag time for first and zero order absorption only
- Last level: administration profile

The equations presented here express the concentration $C(t)$ in the central compartment at a time $t$ after a dose $D$ given at time $t_D$ ($t \geq t_D$).

**Note:** For infusion, the duration of infusion is $T_{inf}$ and $D$ is the total dose administrated.

**Note:** For models with 1 and 2 compartments (as the models presented in this work), equations C(t) express concentration in the central compartment at a time $t$ after drug administration. PK/PD analysis, with intermediate response models, can use concentration $C(t)$ in the central compartment but alternatively concentration $C_e(t)$ in the effect compartment.
There is an additional parameter to estimate, $k_{e0}$, the equilibrium rate constant between central and effect compartment.
For each model, the equation for $C_e(t)$ is given after the corresponding one for $C(t)$.

### 3.2.1 One compartment models

**Parameters**

- $V$ = volume of distribution
- $k_e$ = elimination rate constant
- $Cl$ = clearance of elimination
- $V_m$ = maximum elimination rate (in amount per time unit)
- $K_m$ = Michaelis-Menten constant (in concentration unit)
- $k_a$ = absorption rate constant
- $Tlag$ = lag time

- $Tk_0$ = absorption duration for zero order absorption

**Parameterisation.**

There are two parameterisations for one compartment models, ($V$ and $k_e$) or ($V$ and $Cl$). The equations are given for the first parameterisation ($V$,$k_e$). The equations for the second parameterisation ($V$,$Cl$) are derived using $k_e = \frac{Cl}{V}$.

**Example 3.2.1** ***IV bolus. Linear elimination.***

$$C(t) = \frac{D}{V} e^{-k_e(t-t_D)}$$

$$C_e(t) = \frac{D}{V} \frac{k_{e0}}{(k_{e0} - k_e)} (e^{-k_e(t-t_D)} - e^{-k_{e0}(t-t_D)})$$

*R simulation data.*
`Total subjects=6`
`D=300`
$k_e$`=0.21`
`V=11.7`
`Time points: 0,0.1,0.2,0.3,0.4,0.6,0.8,1,2,4,6,8,12,14,16,18,24`

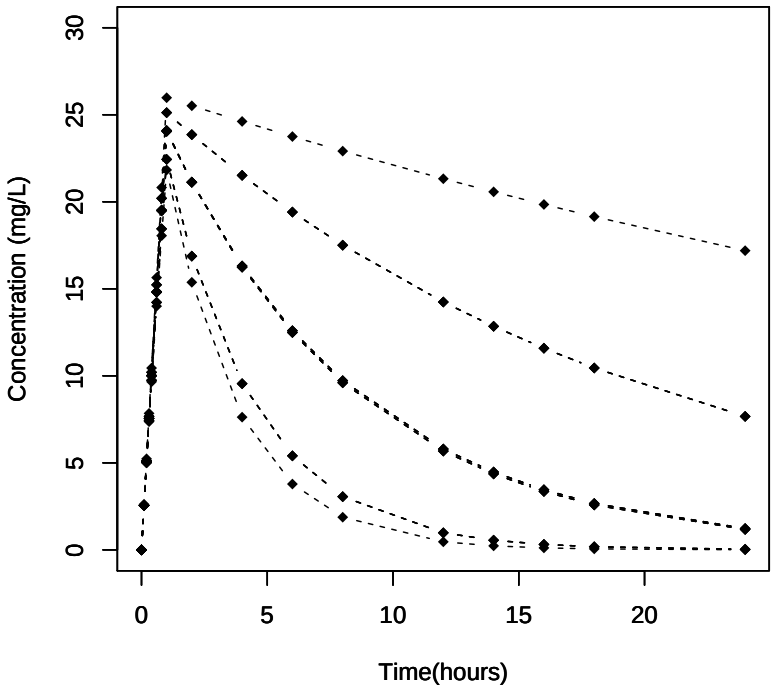


Figure 3.3: Simulation IV Bolus. Linear elimination.

**Example 3.2.2** ***IV bolus. Michaelis-Menten elimination.***

$$\textit{Initial conditions:} \begin{cases} C(t) &= 0 \ \forall t < t_D \\ C_e(t) &= 0 \ \forall t \leq t_D \\ C(t_D) &= \dfrac{D}{V} \end{cases}$$

$$\frac{dC}{dt} = \frac{\frac{V_m}{V}C}{K_m + C}$$

$$\frac{dC_e}{dt} = k_{e0}(C - C_e)$$

*R simulation data.*
*Total subjects=6*
*D=300*
$K_m$*=4.84*
*V=11.7*
$V_m$*=2.17*
*Time points: 0,0.1,0.2,0.3,0.4,0.6,0.8,1,2,4,6,8,12,14,16,18,24*

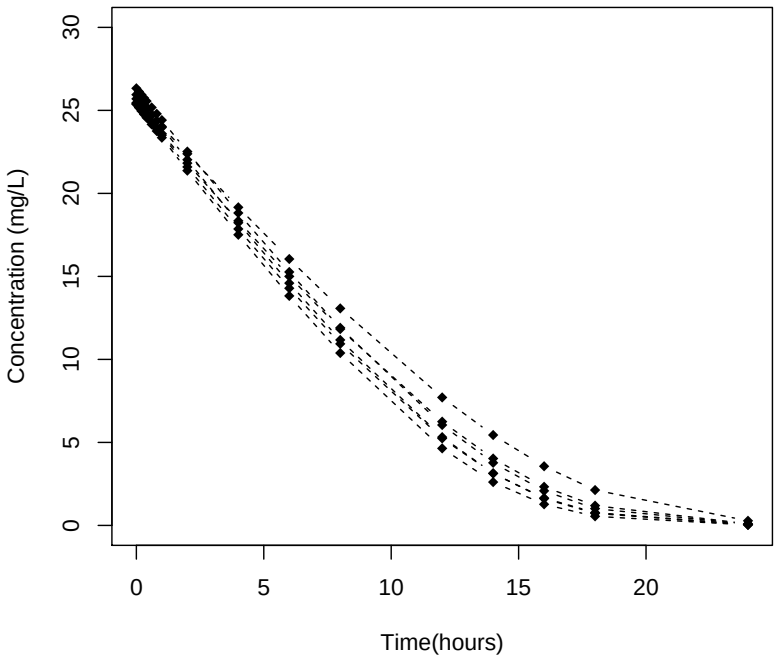


Figure 3.4: Simulation IV Bolus. Michaelis-Menten elimination.

**Example 3.2.3** ***IV infusion. Linear elimination.***

$$C(t) = \begin{cases} \dfrac{D}{T_{inf}} \dfrac{1}{k_e V}(1 - e^{-k_e(t-t_D)}) & if\ t - t_D \geq T_{inf} \\ \dfrac{D}{T_{inf}} \dfrac{1}{k_e V}(1 - e^{-k_e T_{inf}}) e^{-k_e(t-t_D-T_{inf})} & else \end{cases}$$

$$C_e(t) = \begin{cases} \dfrac{D}{T_{inf}} \dfrac{1}{k_e V(k_{e0} - k_e)} [k_{e0}(1 - e^{-k_e(t-t_D)}) - k_e(1 - e^{-k_{e0}(t-t_D)})] & if\ t - t_D \leq T_{inf} \\ \dfrac{D}{T_{inf}} \dfrac{1}{k_e V(k_{e0} - k_e)} [k_{e0}(1 - e^{-k_e T_{inf}}) e^{-k_e(t-t_D-T_{inf})} - \\ \qquad -k_e(1 - e^{-k_{e0} T_{inf}}) e^{-k_{e0}(t-t_D-T_{inf})}] & else \end{cases}$$

*R simulation data.*

```
Total subjects:  6
Dose:  300
T_inf:  1
K_e:  0.12
V: 11.7
Time dose:  0,0.1,0.2,0.3,0.4,0.6,0.8,1,2,4,6,8,12,14,16,18,24,48,72
```

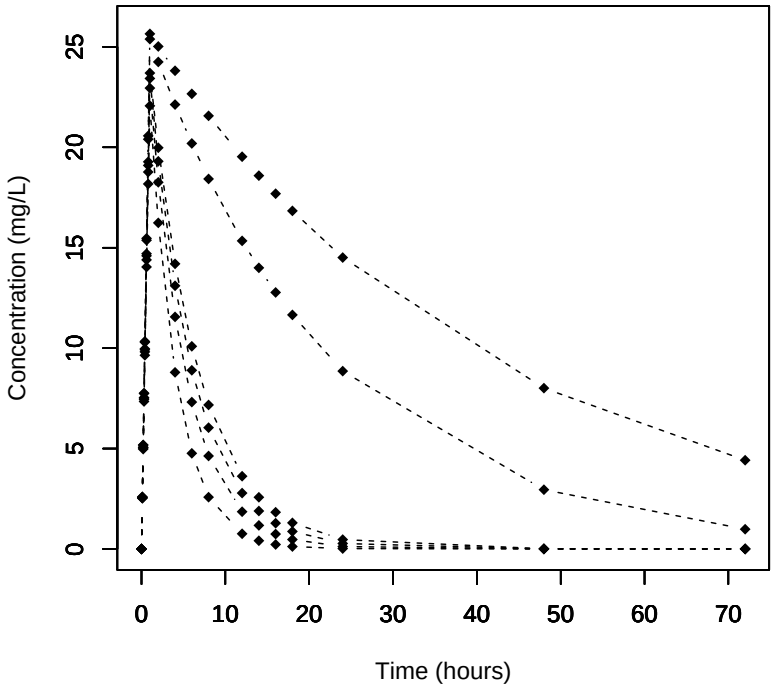


Figure 3.5: Simulation IV Infusion. Linear elimination.

**Example 3.2.4** ***IV infusion. Michaelis-Menten elimination.***

$$\begin{aligned} \textit{Initial condition:} \quad & C(t) = 0 \ \forall t < t_D \\ & C_e(t) = 0 \ \forall t < t_D \end{aligned}$$

$$\frac{dC}{dt} = \frac{\frac{V_m}{V}C}{K_m + C} + input$$

$$\frac{dC_e}{dt} = k_{e0}(C - C_e)$$

$$input(t) \ = \begin{cases} \frac{D}{T_{inf}}\frac{1}{V} & if\, 0 \le t - t_D \le T_{inf} \\ 0 & else \end{cases}$$

**Example 3.2.5** ***First order absorption. Linear elimination. In absence of a lag time.***

$$C(t) = \frac{D}{V}\frac{k_a}{k_a - k_e}(e^{-k_e(t-t_D)} - e^{-k_a(t-t_D)})$$

$$C_e(t) = \frac{Dk_ak_{e0}}{V}(\frac{e^{-k_a(t-t_D)}}{(k_e - k_a)(k_{e0} - k_a)} + \frac{e^{-k_e(t-t_D)}}{(k_a - k_e)(k_{e0} - k_e)} + \frac{e^{-k_{e0}(t-t_D)}}{(k_a - k_{e0})(k_e - k_{e0})})$$

**Example 3.2.6** ***First order absorption. Linear elimination. In presence of a lag time.***

$$C(t) = \begin{cases} 0 \quad if \quad t - t_D \le Tlag \\ \\ \frac{D}{V}\frac{k_a}{k_a - k_e}(e^{-k_e(t-t_D-Tlag)} - e^{-k_a(t-t_D-Tlag)}) \quad else \end{cases}$$

$$C_e(t) = \begin{cases} 0 \quad if \quad t - t_D \le Tlag \\ \\ \frac{Dk_ak_{e0}}{V}\left[\frac{e^{-k_a(t-t_D-Tlag)}}{(k_e - k_a)(k_{e0} - k_a)} + \frac{e^{-k_e(t-t_D-Tlag)}}{(k_a - k_e)(k_{e0} - k_e)} + \right. \\ \\ \left. + \frac{e^{-k_{e0}(t-t_D-Tlag)}}{(k_a - k_{e0})(k_e - k_{e0})}\right] \quad else \end{cases}$$

**Example 3.2.7** ***First order absorption. Michaelis-Menten elimination. In absence of a lag time.***

$$\begin{aligned} \textit{Initial condition: } & C(t) = 0 \textit{ for } t < t_D \\ & C_e(t) = 0 \textit{ for } t < t_D \end{aligned}$$

$$\begin{aligned} \frac{dC}{dt} &= -\frac{\frac{V_m}{V}C}{K_m + C} + input \\ \frac{dC_e}{dt} &= k_{e0}(C - C_e) \\ input(t) &= \frac{D}{V}k_a e^{-k_a(t-t_D)} \end{aligned}$$

**Example 3.2.8** ***First order absorption. Michaelis-Menten elimination. In presence of a lag time.***

$$\begin{aligned} \textit{Initial condition: } & C(t) = 0 \textit{ for } t < t_D \\ & C_e(t) = 0 \textit{ for } t < t_D \end{aligned}$$

$$\begin{aligned} \frac{dC}{dt} &= -\frac{\frac{V_m}{V}C}{K_m + C} + input \\ \frac{dC_e}{dt} &= k_{e0}(C - C_e) \\ input(t) &= \begin{cases} 0 \quad if \quad t - t_D < Tlag \\ \frac{D}{V}k_a e^{-k_a(t-t_D-Tlag)} \quad else \end{cases} \end{aligned}$$

**Example 3.2.9** ***Zero order absorption. Linear elimination. In absence of a lag time.***

$$C(t) = \begin{cases} \frac{D}{Tk_0}\frac{1}{k_e V}[1 - e^{-k_e(t-t_D)}] \quad if \quad t - t_D \leq Tk_0 \\ \\ \frac{D}{Tk_0}\frac{1}{k_e V}[1 - e^{-k_e Tk_0}]e^{-k_e(t-t_D-Tk_0)} \quad else \end{cases}$$

$$C_e(t) = \begin{cases} \frac{D}{Tk_0}\frac{1}{k_e V(k_{e0} - k_e)}[k_{e0}(1 - e^{-k_e(t-t_D)}) - k_e(1 - e^{-k_{e0}(t-t_D)})] \quad if \quad t - t_D \leq Tk_0 \\ \\ \frac{D}{Tk_0}\frac{1}{k_e V(k_{e0} - k_e)}[k_{e0}(1 - e^{-k_e Tk_0})e^{-k_e(t-t_D-Tk_0)} - \\ \qquad\qquad -k_e(1 - e^{-k_{e0}Tk_0})e^{-k_{e0}(t-t_D-Tk_0)})] \qquad\qquad else \end{cases}$$

**Example 3.2.10** ***Zero order absorption. Linear elimination. In presence of a lag time.***

$$C(t) = \begin{cases} 0 \quad if \quad t - t_D \leq Tlag \\ \\ \dfrac{D}{Tk_0}\dfrac{1}{k_e V}(1 - e^{-k_e(t-t_D-Tlag)}) \quad if \quad Tlag < t - t_D \leq Tlag + Tk_0 \\ \\ \dfrac{D}{Tk_0}\dfrac{1}{k_e V}(1 - e^{-k_e Tk_0})e^{-k_e(t-t_D-Tlag-Tk_0)} \quad else \end{cases}$$

$$C_e(t) = \begin{cases} 0 \quad if \quad t - t_D \leq Tlag \\ \\ \dfrac{D}{Tk_0}\dfrac{1}{k_e V(k_{e0} - k_e)}[k_{e0}(1 - e^{-k_e(t-t_D-Tlag)}) - \\ \\ \qquad -k_e(1 - e^{-k_{e0}(t-t_D-Tlag)})] \quad if \quad Tlag < t - t_D \leq Tlag + Tk_0 \\ \\ \dfrac{D}{Tk_0}\dfrac{1}{k_e V(k_{e0} - k_e)}[k_{e0}(1 - e^{-k_e Tk_0})e^{-k_e(t-t_D-Tlag-Tk_0)} - \\ \\ \qquad -k_e(1 - e^{-k_{e0}Tk_0})e^{-k_{e0}(t-t_D-Tlag-Tk_0)}] \quad else \end{cases}$$

### 3.2.2 Two compartments models

**Parameters**

- $V = V_1$ = volume of distribution of first compartment
- $k_e$ = elimination rate constant
- $Cl$ = clearance of elimination
- $V_m$ = maximum elimination rate (in amount per time unit)
- $K_m$ = Michaelis-Menten constant (in concentration unit)
- $k_{12}$ = distribution rate constant from compartment 1 to comp. 2
- $k_{21}$ = distribution rate constant from compartment 2 to comp. 1
- $Q$ = inter-compartmental clearance
- $V_2$ = volume of distribution of second compartment

- $k_a$ = absorption rate constant
- $\alpha$ = first rate constant
- $\beta$ = second rate constant
- $A$ = first macro-constant
- $B$ = second macro-constant

**Parameterisation.**

There are three parameterisations for two compartment models: $(V, k_e, k_{12}$ and $k_{21})$, $(Cl, V_1, Q$ and $V_2)$ or $(\alpha, \beta, A$ and $B)$ except for Michaelis-Menten elimintion where the last parameterisation is not used. The second parameterisation terms are derived using:

- $V_1 = V$
- $Cl = k_e V_1$
- $Q = k_{12} V_1$
- $V_2 = \dfrac{k_{12}}{k_{21}} V_1$
- $\dfrac{V_1}{V_2} = \dfrac{k_{21}}{k_{12}}$

The equations are given for the third parameterisation with:

- $$\alpha = \frac{k_{21} k_e}{\beta} = \frac{\dfrac{Q}{V_2}\dfrac{Cl}{V_1}}{\beta}$$
- $$\beta = \begin{cases} \dfrac{1}{2}\left[k_{12} + k_{21} + k_e - \sqrt{\left(k_{12} + k_{21} + k_e\right)^2 - 4k_{21}k_e}\right] \\ \dfrac{1}{2}\left[\dfrac{Q}{V_1} + \dfrac{Q}{V_2} + \dfrac{Cl}{V_1} - \sqrt{\left(\dfrac{Q}{V_1} + \dfrac{Q}{V_2} + \dfrac{Cl}{V_1}\right)^2 - 4\dfrac{Q}{V_2}\dfrac{Cl}{V_1}}\right] \end{cases}$$

In the following, $C(t) = C_1$ represents the drug concentration in the first compartment and $C_2$ represents the drug concentration in the second compartment.

**Example 3.2.11** ***IV bolus. Linear elimination.***

- $A = \dfrac{1}{V}\dfrac{\alpha - k_{21}}{\alpha - \beta} = \dfrac{1}{V_1}\dfrac{\alpha - \dfrac{Q}{V_2}}{\alpha - \beta}$
- $B = \dfrac{1}{V}\dfrac{\beta - k_{21}}{\beta - \alpha} = \dfrac{1}{V_1}\dfrac{\beta - \dfrac{Q}{V_2}}{\beta - \alpha}$
- $A^e = \dfrac{k_{e0}A}{k_{e0} - \alpha}$
- $B^e = \dfrac{k_{e0}B}{k_{e0} - \beta}$

$$C(t) = D(Ae^{-\alpha(t-t_D)} + Be^{-\beta(t-t_D)})$$
$$C_e(t) = D(A^e e^{-\alpha(t-t_D)} + B^e c^{-\beta(t-t_D)} - (A^e + B^e)e^{-k_{e0}(t-t_D)})$$

**Example 3.2.12** ***IV infusion. Linear elimination.***

- $A = \dfrac{1}{V}\dfrac{\alpha - k_{21}}{\alpha - \beta} = \dfrac{1}{V_1}\dfrac{\alpha - \dfrac{Q}{V_2}}{\alpha - \beta}$
- $B = \dfrac{1}{V}\dfrac{\beta - k_{21}}{\beta - \alpha} = \dfrac{1}{V_1}\dfrac{\beta - \dfrac{Q}{V_2}}{\beta - \alpha}$
- $A^e = \dfrac{k_{e0}A}{k_{e0} - \alpha}$
- $B^e = \dfrac{k_{e0}B}{k_{e0} - \beta}$

$$C(t) = \begin{cases} \dfrac{D}{T_{inf}}\left[\dfrac{A}{\alpha}(1-e^{-\alpha(t-t_D)}) + \dfrac{B}{\beta}(1-e^{-\beta(t-t_D)})\right] & if \quad t-t_D \leq T_{inf} \\ \dfrac{D}{T_{inf}}\left[\dfrac{A}{\alpha}(1-e^{-\alpha T_{inf}})e^{-\alpha(t-t_D-T_{inf})} + \dfrac{B}{\beta}(1-e^{-\beta Tinf})e^{-\beta(t-t_D-T_{inf})}\right] & else \end{cases}$$

$$C_e(t) = \begin{cases} \dfrac{D}{T_{inf}}\left[\dfrac{A^e}{\alpha}(1-e^{-\alpha(t-t_D)}) + \dfrac{B^e}{\beta}(1-e^{-\beta(t-t_D)}) - \right. \\ \qquad \left. - \dfrac{A^e+B^e}{k_{e0}}(1-e^{-k_{e0}(t-t_D)})\right] & if \quad t-t_D \leq T_{inf} \\ \dfrac{D}{T_{inf}}\left[\dfrac{A^e}{\alpha}(1-e^{-\alpha T_{inf}})e^{-\alpha(t-t_D-T_{inf})} + \dfrac{B^e}{\beta}(1-e^{-\beta T_{inf}})e^{-\beta(t-t_D-T_{inf})} - \right. \\ \qquad \left. - \dfrac{A^e+B^e}{k_{e0}}(1-e^{-k_{e0}T_{inf}})e^{-k_{e0}(t-t_D-T_{inf})}\right] & else \end{cases}$$

**Example 3.2.13** ***First Order Absorption. Linear elimination. In absence of a lag time.***

- $A = \dfrac{k_a}{V}\dfrac{k_{21}-\alpha}{(k_a-\alpha)(\beta-\alpha)} = \dfrac{k_a}{V_1}\dfrac{\dfrac{Q}{V_2}-\alpha}{(k_a-\alpha)(\beta-\alpha)}$
- $B = \dfrac{k_a}{V}\dfrac{k_{21}-\beta}{(k_a-\beta)(\alpha-\beta)} = \dfrac{k_a}{V_1}\dfrac{\dfrac{Q}{V_2}-\beta}{(k_a-\beta)(\alpha-\beta)}$
- $A^e = \dfrac{k_{e0}A}{k_{e0}-\alpha}$
- $B^e = \dfrac{k_{e0}B}{k_{e0}-\beta}$
- $C^e = -\dfrac{A^e(k_a-\alpha)+B^e(k_a-\beta)}{k_a-k_{e0}}$

$$C(t) = D(Ae^{-\alpha(t-t_D)} + Be^{-\beta(t-t_D)} - (A+B)e^{-k_a(t-t_D)})$$

$$C_e(t) = D(A^e e^{-\alpha(t-t_D)} + B^e e^{-\beta(t-t_D)} + C^e e^{-k_{e0}(t-t_D)} - (A^e+B^e+C^e)e^{-k_a(t-t_D)})$$

## 3.3 Inference

Inference will focus on *one compartment models*. In particular, we will make inference on IV bolus and IV infusion, both with Michaelis-Menten elimination. Inference on other kind of models can be made too with [10], even in more-than-one compartment models.

Some models described before has been done using the package [10]. Now, for inference, we use as initial values some values close to those given by the package, and we will see if, after some iterations of the algorithm, R is capable of obtaining the true parameters that have been used.

**IV bolus. Michaelis-Menten elimination**

If we simulate with [10] 24 subject data with parameters

- $D = 300.00$
- $V_m = 2.17$
- $K_m = 4.84$
- $V = 11.70$,

we obtain a table that begins like that:

| *Subject* | *Time* | *Concentration* |
|---|---|---|
| 1 | 0 | 25.421 |
| 1 | 0.1 | 25.248 |
| 1 | 0.2 | 25.074 |
| 1 | 0.3 | 24.091 |
| 1 | 0.4 | 24.728 |
| ⋮ | ⋮ | ⋮ |

So we can consider it as data in which make inference in order to determine $V_m$, $K_m$ and $V$.

We can choose, for example, the initial values

- $D = 300$ (We consider the dose as known)
- $V_m = 2$

- $K_m = 4$
- $V = 11$,

and R give us as a result the following:

```
--- initial values for parameters ----

   Parameter  Initial
 1      D        300
 2     Vm         2
 3     Km         5
 4      V        11
--- wighting scheme:  equal weight

--- model selection:  a one-compartment, iv bolus pk model with
M-M elim.

<< PK parameter obtained from Nelder-Mead Simplex algorithm >>

   Parameter  Value
 1     Vm      2.024
 2     Km      4.834
 3      V     11.779
```

As we see, we have get similar results as the original true values.

### IV infusion. Michaelis-Menten elimination.

As before, if we simulate with [10] 24 subject data with parameters

- $D = 300.00$
- $T_{inf} = 0.5$
- $V_m = 2.31$
- $K_m = 4.74$
- $V = 11.70$,

we obtain again a table that begins like that:

| *Subject* | *Time* | *Concentration* |
|---|---|---|
| 1 | 0 | 0 |
| 1 | 0.1 | 5.016 |
| 1 | 0.2 | 9.964 |
| 1 | 0.3 | 14.885 |
| 1 | 0.4 | 19.791 |
| ⋮ | ⋮ | ⋮ |

So we can consider again the data above as data in which we can make inference for determine $V_m$, $K_m$ and $V$.

We can choose as initial values

- $D = 300$ (We consider the dose as known)
- $T_{inf} = 0.5$ (We consider the infusion time as known too)
- $V_m = 2$
- $K_m = 4$
- $V = 11$,

and R give us as a result the following:

```
--- initial values for parameters ----

   Parameter  Initial
 1      D        300
 2     T_inf       0.5
 3     V_m         2
 4     K_m         4
 5      V         11
--- wighting scheme:  equal weight

--- model selection:  one-compartment, iv infusion PK model
with M-M elim.

<< PK parameter obtained from Nelder-Mead Simplex algorithm >>

   Parameter  Value
 1     V_m      2.183
 2     K_m      4.792
 3      V      11.670
```

As we see, we have get once more similar results as the original true values.

## 3.4 Some Pharmacodynamic Models

Two different type of models are presented here:

- Inmediate response models
- Turnover models

### 3.4.1 Inmediate response models

For these type of models, the effect (denoted as $E(t)$) is expressed as:

$$E(t) = A(t) + S(t) \tag{3.4}$$

where $A(t)$ represents the model of drug action and $S(t)$ corresponds to the baseline/disease model. A(t) is a function of the concentration $C(t)$ in the central compartment or of the concentration $C_e(t)$ in the effect compartment.

**Parameters**

- $A_{lin}$ = constant associated to $C(t)$
- $A_{quad}$ = constant associated to the square of $C(t)$
- $A_{log}$ = constant associated to the logarithm of $C(t)$
- $E_{max}$ = maximal agonistic response
- $I_{max}$ = maximal antagonistic response
- $EC_{50}$ = concentration to get half of the maximal response
- $\gamma$ = sigmoidicity factor
- $S_0$ = baseline value of the studied effect
- $k_{prog}$ = rate constant of disease progresion

**Drug action models**

- linear model
$$A(t) = A_{lin}C(t)$$
- quadratic model
$$A(t) = A_{lin}C(t) + A_{quad}C(t)^2$$
- logarithmic model
$$A(t) = A_{log}log(C(t))$$
- $E_{max}$ model
$$A(t) = \frac{E_{max}C(t)}{C(t) + EC_{50}}$$
- sigmoïd $E_{max}$ model
$$A(t) = \frac{E_{max}C(t)^\gamma}{C(t)^\gamma + EC_{50}^\gamma}$$
- $I_{max}$ model
$$A(t) = 1 - \frac{I_{max}C(t)}{C(t) + EC_{50}}$$
- sigmoïd $I_{max}$ model
$$A(t) = 1 - \frac{I_{max}C(t)^\gamma}{C(t)^\gamma + EC_{50}^\gamma}$$

**Baseline/disease models**

- null baseline

$$S(t) = 0$$

- constant baseline with no disease progression

$$S(t) = S_0$$

- linear disease progression

$$S(t) = S_0 + k_{prog}t$$

- exponential disease increase

$$S(t) = S_0 e^{-k_{prog}t}$$

- exponential disease decrease

$$S(t) = S_0(1 - e^{-k_{prog}t})$$

**Note:** Only, for the $I_{max}$ and sigmoïd $I_{max}$ models $A(t)$ is not added to $S(t)$ but $S_0$ is multiplied by $A(t)$ in the expression of $S(t)$. For instance, for $I_{max}$ model with linear baseline we have

$$E(t) = S_0 * A(t) + k_{prog}t.$$

### 3.4.2 Turnover response models

In these models, the drug is not acting on the effect $E$ directly but rather on $R_{in}$ (input (synthesis) rate) or $k_{out}$ (output (elimination) rate constant).

Thus the system is described with differential equations, given $\frac{dE}{dt}$ as a function of $R_{in}$, $k_{out}$ and $C(t)$, the drug concentration at time $t$.

The initial condition is: while $C(t) = 0$, $E(t) = \frac{R_{in}}{k_{out}}$.

**Parameters:**

- $E_{max}$ = maximal agonistic response
- $I_{max}$ = maximal antagonistic response

- $EC_{50}$ = concentration to get half of the maximal response
- $\gamma$ = sigmoidicity factor
- $R_{in}$ = input (synthesis) rate
- $k_{out}$ = output (elimination) rate constant

**Models with impact on the input** ($R_{in}$)

- $E_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 + \frac{E_{max}C}{C + EC_{50}}\right) - k_{out}E$$
- sigmoïd $E_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 + \frac{E_{max}C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right) - k_{out}E$$
- $I_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 - \frac{I_{max}C}{C + EC_{50}}\right) - k_{out}E$$
- sigmoïd $I_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 - \frac{I_{max}C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right) - k_{out}E$$
- full $I_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 - \frac{C}{C + EC_{50}}\right) - k_{out}E$$
- sigmoïd full $I_{max}$ model
$$\frac{dE}{dt} = R_{in}\left(1 - \frac{C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right) - k_{out}E$$

**Models with impact on the output** ($k_{out}$)

- $E_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 + \frac{E_{max}C}{C + EC_{50}}\right)E$$

- sigmoïd $E_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 + \frac{E_{max}C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right)E$$

- $I_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 - \frac{I_{max}C}{C + EC_{50}}\right)E$$

- sigmoïd $I_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 - \frac{I_{max}C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right)E$$

- full $I_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 - \frac{C}{C + EC_{50}}\right)E$$

- sigmoïd full $I_{max}$ model

$$\frac{dE}{dt} = R_{in} - k_{out}\left(1 - \frac{C^{\gamma}}{C^{\gamma} + EC_{50}^{\gamma}}\right)E$$

# Bibliography


[1] Julie Bertrand and France Mentré. Mathematical expressions of the pharmacokinetic and pharmacodynamic models implemented in the monolix software. Technical report, Technical 10 report, INSERM, 2008.

[2] Peter L. Bonate. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer, 2nd edition, 2011.

[3] Sean Borman. The expectation maximization algorithm-a short tutorial. 2004.

[4] Marie Davidian. *Nonlinear Mixed Effects Models: An Overview and Update*. Journal of Agricultural, Biological, and Environmental Statistics, 2004.

[5] Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. *Longitudinal data analysis*. CRC Press, 2008.

[6] Mats O Karlsson, Stuart L Beal, and Lewis B Sheiner. Three new residual error models for population pk/pd analyses. *Journal of pharmacokinetics and biopharmaceutics*, 23(6):651–672, 1995.

[7] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(0):1020–1038, 2005.

[8] B. Meibohm and H. Derendorf. Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modeling. *International Journal of Clinical Pharmacology and Therapeutics*, 35(0):401–413, 1997.

[9] Hulin Wu and Jin-Ting Zhang. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*, volume 515. John Wiley & Sons, 2006.

[10] Chun ying Lee and Yung jin Lee. *PKfit: A Data Analysis Tool for Pharmacokinetics*, 2014. R package version 1.2.4.