1

Overview

1.1 The population approach

The desire to model a biological or physical phenomenon often arises when we are able to record some observations issued from it. Nothing would therefore be more natural than to begin this introduction by looking at some observed data. We will not go into the details of the data here; we use it only to illustrate in an intuitive and introductory way what we call the population approach.

The first example involves weight data for rats measured over 14 weeks for a subchronic toxicity study related to the question of genetically modified (GM) corn. To evaluate health risks associated with GM organisms, individual rats were put on diets containing different quantities of GM corn. To discover whether there is an effect due to GM food on weight, for instance, we look to find whether part of the variability observed among the growth curves can be explained by diet. For each rat, the sequence of observations is longitudinal data (Fitzmaurice et al., 2008; Molenberghs and Verbeke, 2005); the same information (weight) is measured at multiple time points (Figure 1.1).

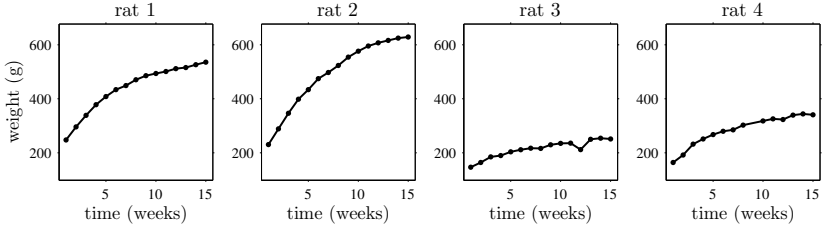

FIGURE 1.1: Weight curves for four rats.

The next set of plots displays the viral load of four patients with hepatitis C who started a treatment at time $t = 0$. Viral load data for hepatitis C carriers exhibit great clinical response variability under the

same treatment (Snoeck et al., 2010). Some patients respond to treatment and their viral load decreases, while others have no response or see their viral load rise again at the treatment's end (Figure 1.2).

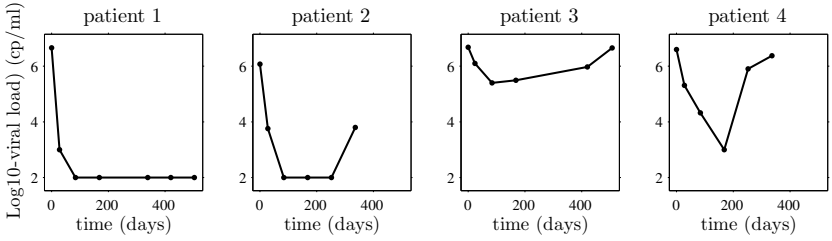

FIGURE 1.2: Viral load of four patients with hepatitis C.

In the next example (Figure 1.3), the data are fluorescence intensities measured over time in a cellular biology experiment which aims to describe the cell cycle and the creation of free radicals (Faure et al., 2013). Several repeats of the same experiment are necessary because the same experimental conditions can lead to great variability in fluorescence activity.

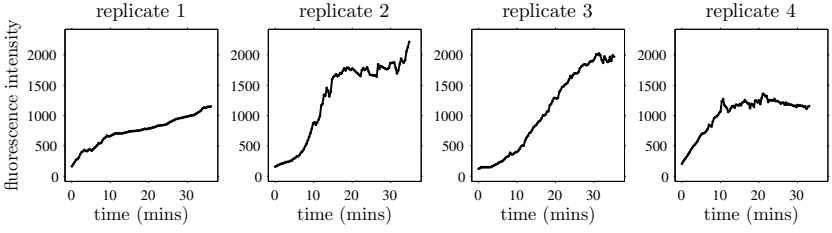

FIGURE 1.3: Fluorescence intensities from four replicates of the same experiment.

Note that repeated measurements are not necessarily always functions of time. For example, we may be interested in corn production as a function of fertilizer quantity (Makowski and Lavielle, 2006). Such studies are important in agriculture in order to provide recommendations for farmers as to optimal fertilizer use. In nature there exists variability in soil conditions, and studies need to be performed over several parcels of land (Figure 1.4).

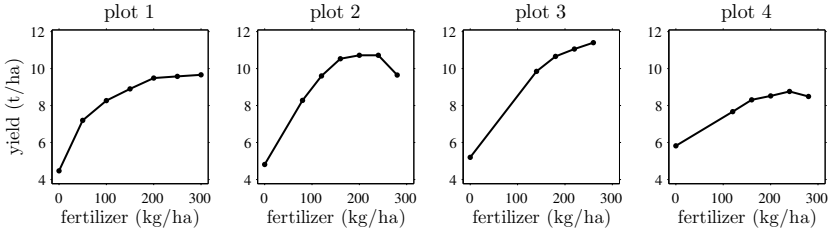

FIGURE 1.4: Corn production as a function of fertilizer quantity.

Even though these examples come from quite different domains, in each case the data is made up of repeated measurements on several individuals from a population. What we will call a "population approach" will be relevant for characterizing and modeling this data. The modeling goal is thus twofold: characterize, first, the biological or physical phenomena observed for each individual and, second, the variability seen between individuals.

In the example with rats, the model needs to integrate a growth model that describes how a rat's weight increases with time and a statistical model that describes why this process can vary from one rat to another. The goal is thus to finish with a "typical" growth curve for the population and be able to describe and explain, as much as possible, the variability in the individual's curves around this population curve.

As these are longitudinal data, a curve with weight as a function of time gives a natural graphical representation of a mathematical growth model. Thus, after looking at the data, the first step is to choose a growth model well adapted to the data. For example, we might consider a very simple growth model of the form $f(t) = a + b(1 - e^{-k\,t})$. We can then check that curves created using this model "look like" the observed growth curves for a suitably chosen set of parameters (a, b, k). In addition to its simplicity, another advantage of this model is that its parameters can be interpreted biologically: a is the rat's weight at time 0, i.e., at birth, b the maximum weight gain (as time t tends to infinity) and k a measure of the speed at which the rat grows. Obviously, other growth models are available of which some will be more adapted than others to modeling certain types of growth. This leads to the need for model selection criteria, kept out of this introductory section for simplicity, but looked at later in the book.

Beyond these technical aspects of model selection, it is important to have in mind that the selection of a model must first and foremost be guided by the usage we intend to make of it. In the above example, the model was chosen quite empirically and we have no misconceptions that it is a true biological model, i.e., the equation form of some physiological

process. It is merely a reasonable approximation of what was empirically observed. Clearly, the model does not allow us to precisely characterize the rat's growth process, its limbs and organs, etc. If it is these types of complex phenomena that need to be studied, we would have to develop much more sophisticated models to describe them. If on the other hand we are only interested in a basic characterization of the phenomenon, our simple model may just do the trick; it could be useful for such things as predicting the weight of a rat at 15 or 16 weeks, or detecting rats with atypical growth patterns.

The use of such types of parametric models will allow us to formalize and quantify the concept that we have so far introduced intuitively yet informally: the *population approach*. First, defining a typical population curve means in essence defining a typical set of population parameters $(a_{\text{pop}}, b_{\text{pop}}, k_{\text{pop}})$. Then, each individual's curve can be seen as depending on its own set of individual parameters (a_i, b_i, k_i). Characterizing the *inter-individual* variability of these curves around the typical population curve thus means characterizing the inter-individual parameter variability around the typical population parameters. Figure 1.5 shows the population curve as well as the four individual's curves.[1]

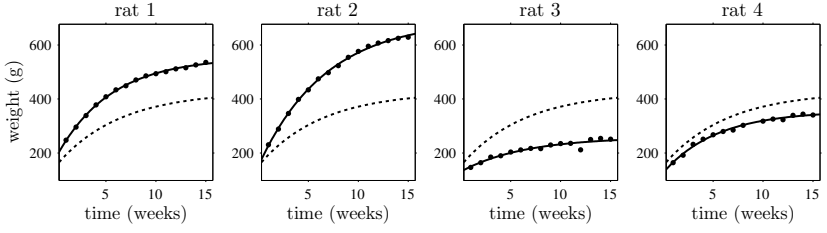

FIGURE 1.5: Weights of four rats. The observed weights are displayed with dots, the individual growth curves with solid lines, the population curves with dotted lines.

This step of modeling the inter-individual variability happens in a probabilistic framework: we consider that each individual in the sample was drawn at random from the population of interest. The individual parameters are then regarded as random variables that possess a certain distribution, precisely representing the way in which the values of these parameters are distributed within the population. Such types of statistical models based on probability distributions allow us to describe the inter-individual variability of the parameters. We would like not only

[1] Population and individual parameters have been estimated using the software Monolix and the methods described in Chapter 7. Modeling of this data is presented in detail in Section 8.1.

to characterize this variability, but to understand it as best as possible. Thus, in our example with rats, we can see that some rats are heavier than others, but can we say why? Do we have, for example, relevant information perhaps to explain why rats 1 and 2 are heavier and put on weight faster that rats 3 and 4? Well yes! We notice that these two rats are male while the other two are female. Gender should therefore help explain some of the variability seen here.

In addition to the population curve and the four individual growth curves, Figure 1.6 shows the typical growth curves for males and females.

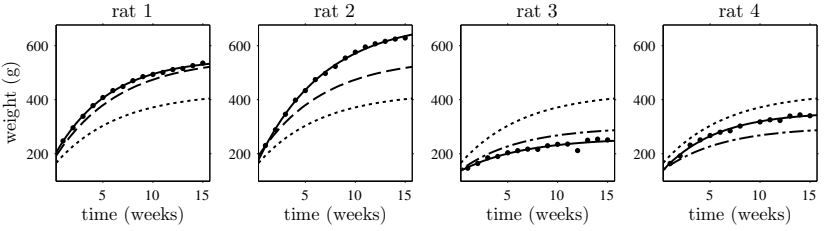

FIGURE 1.6: The typical male curve is displayed with dashed line and female curve with dash-dot line. Rats 1 and 2 are male, rats 3 and 4 are female.

This figure thus shows a decomposition of the variability of the growth curves around the typical population curve: first we decompose the population into two subpopulations, male and female, each of which possesses a typical curve. The differences between these curves represent the *inter-group* variability and already explain much of the variability observed between rats. Of course, all males (nor the females) are not the same weight, so there exists intra-group variability represented by the variability of individual curves around the group ones.

The question then naturally arises as to whether this *intra-group* variability can itself be explained by other *covariates*. In the case of a toxicological study for GM crops, for example, we can wonder if these weight differences are related to diet.

But whether we are in the context of toxicological studies, clinical trials or agronomic trials, modeling is not only for characterizing the collected data. The main goal is the building of a *population model* which applies by definition to the entire population. We therefore move from a purely descriptive treatment of the data to an inferential one. The question is then whether some characteristic observed in the sample can be generalized to the whole population. For example, can the weight difference observed between males and females be truly "explained" by a "gender effect," or can it be attributed to random sampling? In other

words, would we have a good chance of reaching the same conclusions if the experiment were repeated with new individuals?

1.2 About models

1.2.1 Mixed effects models and hierarchical models

When observations (rat weights, for example) are available, the observed data have no reason to be absolutely identical to the predictions provided by the model (the growth model in our example). Indeed, the model is chosen to try to mimic the studied phenomenon; thus, the prediction it provides is not necessarily the rat's actual weight. Furthermore, the observations are subject to measurement errors; whatever the quality of the measuring device, it is not possible to measure "exactly" the weight of a rat. The model must therefore include a statistical model that can describe how observations are distributed around the values predicted by the model.

Yet another component of the model must describe the inter-individual variability of the model's parameters. The model will explain some of this variability using individual covariates such as sex, but some variability will remain unexplained and have to be considered as random. Integrating both fixed and random effects into the same model leads naturally to the use of mixed effects models.

Linear mixed effects models (Verbeke and Molenberghs, 2000; West et al., 2006), generalized linear mixed models (Dobson, 2002; McCulloch et al., 2008) and nonlinear mixed effects models (Lindstrom and Bates, 1990; Davidian and Giltinan, 1995; Pinheiro and Bates, 2009) are widely used in applications such as sociology and demography (O'Brien et al., 2008), psychophysiology (Bagiella et al., 2000), ecology and evolution (Bolker et al., 2009), genomics (Haldermans et al., 2007), portfolio risk (McNeil and Wendin, 2007), agronomy (Bürger et al., 2012), neuroimaging (Friston et al., 2005), medicine (Brown and Prescott, 2006), genetics (Ott, 1979), systems biology (Gonzalez et al., 2013) and clinical trials (Andersen and Millen, 2013). Nonlinear mixed effects models have also been an essential tool for pharmacokinetic-pharmacodynamic (PKPD) modeling for several decades (Bonate, 2011). The adoption of population approaches in pharmacometrics owes a lot to the development of the NONMEM software in the 1980's (Beal and Sheiner, 1980) as well as the pioneering work of Lewis Sheiner (Rowland, 2005). We will see in this book that mixed effects models are particularly well-suited sta-

tistical tools for population modeling of broad types of data including continuous, categorical, count and time-to-event data.

An alternative yet equivalent approach considers the model to be hierarchical; each individual data series is described by a single model, and variability between individual models is described by a population model. In the case of parametric models, this means that the observations for a given individual are described by a probability distribution that depends on a vector of individual parameters; this is the classic individual approach. The population approach is then a direct extension of this; we add a new component to the model, a probability distribution that describes variability of the individual parameters within the population.

A mixed effects model can thus be seen as a hierarchical model, i.e., as a joint probability distribution for the observations and individual parameters. This joint distribution can easily be extended to the case where other variables in the model are considered as random variables, such as covariates, population parameters or the design.

The use of hierarchical models has grown enormously since the development of Bayesian inference methods (Congdon, 2006; Gelman et al., 2003; Robert, 2007). Markov chain Monte Carlo (MCMC) methods are particularly well-suited to this type of model (Gilks et al., 1996; Robert and Casella, 2004). Hierarchical models are thus used today in a great number of domains including spatial data modeling (Banerjee et al., 2003), marketing (Allenby and Rossi, 1998), political science (Hurwitz and Peffley, 1987), climate studies (Cooley et al., 2007; New and Hulme, 2000), plant growth (Baey et al., 2013), leisure sciences (Crawford et al., 1991), sociology (Raftery, 1995), PKPD modeling (Wakefield et al., 1998) and epidemiology (Lawson, 2013).

Mixed effects models and hierarchical models turn out to be complementary approaches which lead to different representations of the same model.

1.2.2 Description, representation and implementation of models

A model can be used in the real world if it can be implemented in a software program. To do this, we need a language that can be understood by the software. Before even arriving at this point, it is important to be very clear and systematic about what a model is and how we want to use it.

It is of fundamental importance to distinguish between the description, representation and implementation of a model. Each of these three concepts uses a specific language. In the context of mixed effects models,

models that we want to implement can be decomposed into two components: the structural model and the statistical model. Both have to be described, represented and implemented with precision.

Consider first our rat growth model:

1. We start describing a model with words, i.e., a human language:

   "The weight of a rat increases over time with a growth rate that decreases exponentially"

2. Then we represent the model using a mathematical language:

   $$w(t) = a + b(1 - e^{-k\,t})$$

3. Lastly, we implement the model via a language understood by the software:

   ```
   WEIGHT = a + b*(1-exp(-k*TIME))
   ```

We can follow the same process for the statistical model. A description of the model for the initial weight a might be, "The weight at birth is log-normally distributed in the population." This model can then be mathematically represented:

$$\log(a) \sim \mathcal{N}(\log(a_{\text{pop}}), \omega_a^2), \tag{1.1}$$

and implemented, for instance, with MLXTRAN (see Section 5.2.3), a powerful declarative language for implementing models:

```
a={distribution=lognormal, reference=a_pop, sd=omega_a}
```

This model representation is not unique. Indeed, model (1.1) could also be represented (and implemented) using an equation rather than a definition:

$$a = a_{\text{pop}} e^{\eta}, \tag{1.2}$$

where $\eta \sim \mathcal{N}(0, \omega_a^2)$. This gives us two representations of the same statistical model where the distribution of a is given explicitly in (1.1) and implicitly in (1.2). The choice of representation should be driven by the tasks we want to execute. If the model is only used to perform simulation, both representations contain all the information required. On the other hand, if the log-normal assumption needs to be tested or the probability distribution function (pdf) of a needs to be computed without using approximation, the distribution needs to be represented by an explicit definition as in (1.1).

1.3 Tasks, methods and tools

We will talk about models all the way through the book, but remember at all times that the main purpose of a model is to be *used*. We will essentially focus on two types of task *i)* modeling, in which we have data and wish to construct a model that could have generated this data or is capable of explaining its variability and *ii)* simulation, where we want to generate virtual data using a model.

Remark: This book does not pretend to cover all possible tasks one can perform with a model. Optimization of the design, for instance, is of real importance in practice but will not be considered here. There exists an extensive bibliography on the topic (see, for example, Fedorov and Leonov, 2013, and the many references therein). Several software packages for optimal design also exist, in particular for applications in pharmacometrics (see Mentré et al., 2007, for a comparison of existing tools).

To use a model in practice, we clearly need modeling and simulation tools, but also a language that allows us to implement it, which is "understood" and can be used by these tools.

1.3.1 Model coding

In this book, we will be working with MLXTRAN, a flexible and powerful declarative language[2] designed for implementing complex hierarchical models.

MLXTRAN allows users to write ordinary differential equation (ODE) based models, implement pharmacokinetic models with complex administration schedules, include inter-individual variability in parameters and define statistical models for different types of data (including continuous, categorical, count and time-to-event data). Another crucial property of MLXTRAN is that it rigorously adopts the model representation formalism proposed in this book. In other words, the model implementation is fully consistent with its mathematical representation.

Furthermore, it is important to note that all of the tools we are going to use, for both modeling and simulation, are based on MLXTRAN. This will allow us to define a complete workflow using the same model implementation, i.e., run several tasks based on the same model.

[2] A declarative language allows variables to be defined using equations and definitions, rather than simply the calculation of their values.

1.3.2 Model exploration

Before using a model to perform modeling and simulation, it may be useful to visualize, i.e., represent graphically, the various functions of time given in the model for several parameter values and run a sensitivity analysis, i.e., "look at" what happens when we change the value of one or several parameters.

Coming back to our usual example, we may want to visualize the growth model $f(t) = a+b(1-e^{-k\,t})$ and see how a rat's weight converges to its limiting weight as a function of the parameter k. We may also want to visualize the probability distribution of f when a, b and k are random variables with distributions defined in the model.

Whenever we want to visualize a model, we will use the MLXPLORE software. MLXPLORE allows us to visualize not only the structural model but also the statistical one, for example, visualizing the impact of covariates and inter-individual variability of model parameters on predictions.

1.3.3 Modeling

Modeling is clearly the most difficult and challenging task to perform. It can only be partially automated and requires experience on the modeler's part. Tools for modeling are not designed to build a model automatically, but rather to help a modeler construct a model with the aid of a set of observations.

It is important to differentiate clearly between these distinct concepts: the task to perform, the method to use for this task, the algorithm to implement the method, and the implementation of the algorithm itself. For example, imagine that we want to estimate the population parameters. In this case, the task to perform is estimation and a possible method is maximum likelihood estimation. SAEM (Stochastic Approximation of Expectation-Maximization, Delyon et al., 1999) is an algorithm for computing this estimate and MONOLIX a software in which the algorithm is implemented.

Alternatively, we could choose to perform a Bayesian estimate of the population parameters using an MCMC algorithm implemented, for example, in WinBUGS (Ntzoufras, 2011; Spiegelhalter et al., 2003).

As well as estimating the population parameters, it is important to look at the precision of these estimates. Here again, several methods are available, such as the Fisher information matrix, bootstrapping to obtain standard errors, or posterior distributions.

The estimation of population parameters and their uncertainty supposes that the model is given; we then look for parameter values for this model. Of course, the "best" model is rarely known, so a modeler

must progress by trial and error, trying different models, improving and comparing them. We therefore need diagnostic and model selection tools.

Diagnostic tools are used to decide whether a model is acceptable, and if not, to identify which model hypotheses are not valid. We will therefore be able to work in the context of statistical hypothesis testing:

$$H_0: \quad \text{“}\mathcal{M} = \mathcal{M}_0\text{”} \qquad \text{vs} \qquad H_1: \quad \text{“}\mathcal{M} \neq \mathcal{M}_0\text{”}$$

where $\mathcal{M}_0$ is the model the modeler wishes to evaluate. A model remains acceptable if the null hypothesis H_0 is not rejected. The more the diagnostic tools are able to detect model misspecification, the more powerful the test will be. Using this methodology does not imply that we know explicitly the rejection zones of the tests; the approach remains largely empirical, based on graphical outputs and simulation methods.

Methods for model selection that are widely used in our context and that are presented in this book are standard, based on statistical tests such as the likelihood ratio test and Wald test (see Engle 1984; Freedman et al. 2007) and on information criteria such as the Akaike information criterion (Greven and Kneib, 2010) and Bayesian information criterion (Schwarz, 1978). The art of modeling then consists of correctly interpreting and adequately combining all the graphical outputs and statistical criteria in order to improve the model.

We mainly use the software Monolix in this book to illustrate the various modeling tasks with which the modeler is faced. Monolix has been developed with this in mind: it serves modelers well while fully complying with a coherent mathematical formalism that uses well-known and theoretically justified methods. The algorithms implemented in Monolix (SAEM, MCMC, simulated annealing, importance sampling, etc.) are extremely efficient for a wide variety of complex models including count data models (Savic and Lavielle, 2009), categorical data models (Savic et al., 2011), censored data models (Samson et al., 2006), time-to-event data models (Mbogning et al., 2014), mixture models (Lavielle and Mbogning, 2013), hidden Markov models (Delattre and Lavielle, 2012) and SDE-based models (Delattre and Lavielle, 2013; Donnet and Samson, 2008). We also emphasize that the convergence of SAEM and its extensions has been established under quite general hypotheses (Allassonnière et al., 2010; Delyon et al., 1999; Kuhn and Lavielle, 2004).

1.3.4 Simulation

Once a model has been constructed, we may wish to simulate the phenomenon we have been studying by generating data from the model. Here, simulation means simulating both virtual individuals and longitu-

dinal data for them, or simply calculating the values of quantities defined in the model (e.g., the components of a system of ODEs).

We have available to us Simulx, which is both an R and MATLAB function that enables us to compute predictions and sample data from any MLXTRAN model. Simulx combines the flexibility of R and MATLAB scripts with the power of MLXTRAN to easily encode complex models.

1.4 Contents of the book

Part I presents several preliminary concepts that will help the reader to be more comfortable with the methodology that follows. We start in Chapter 2 with linear and nonlinear models, used for modeling a single sequence of longitudinal data. In the population approach, linear and nonlinear mixed models can then be naturally introduced to model continuous longitudinal data and take into account variability in a model's individual parameters. Generalized mixed models allow us to model other data such as categorical data. We will see how these different mixed models can all be also seen as hierarchical ones.

Following this, in Chapter 3 we show how the hierarchical structure of a model leads to a natural decomposition of the joint distribution into a product of conditional and marginal distributions.

Part II focuses on model description, representation and implementation. Models for longitudinal data are described in Chapter 4. In particular, models for continuous, count, categorical and time-to-event data – including joint models which combine several types of outcome – are presented and illustrated in various examples.

Models for individual parameters are described in Chapter 5. We will be particularly interested in Gaussian models, i.e., ones in which there exists some transform (log, logit, probit, etc.) for which the transformed parameters have a Gaussian distribution. Individual covariates may also be introduced into a model in order to explain some of the variability in the individual parameters. This chapter also shows how several levels of variability can be included in a model with the aim of modeling things like *inter-occasion* and *inter-group* variability.

Extensions for mixture models, hidden Markov models and stochastic differential equation-based models are presented in Chapter 6.

Part III is concerned with ways to use models to perform tasks. Chapter 7 presents several methods involved in the practical use of these models for tasks such as parameter estimation, model diagnostics and

model selection. We use a classical population pharmacokinetics (PK) example to illustrate these methods.

Chapter 8 illustrates many of the proposed approaches and methods using several examples. The first example describes how to model the growth curves of rats in a toxicity study related to GM corn. The second example involves a classic pharmacokinetics-pharmacodynamics application. We show how to extend a proposed joint model for continuous outcomes to a joint model for continuous and categorical data. The last example then shows that mixed effects models can also be successfully used in quantitative biology for modeling the dynamics of biological networks in cell populations.

Chapter 9 describes several algorithms useful for running tasks that interest us. In particular, the SAEM algorithm, used for estimating population parameters by maximum likelihood, is presented and its implementation described. A simple example then allows us to better understand its excellent theoretical and practical properties.

Appendix A describes how to deal with longitudinal data in the classical context of a unique individual. A PK example accompanied by R code illustrates this approach.

Some prerequisites are required for a good understanding of this book. To help with this, Appendix B recalls some basic probability and statistics results which play fundamental roles in model construction.

This book does not treat only pharmacometric modeling and is not intended for pharmacometricians only. However, many of the illustrative examples do come from pharmacokinetics because it is a domain that frequently uses the population approach and mixed models. For this reason, an overview of PK modeling is proposed in Appendix C to help nonspecialists in this domain better understand the book's examples.

Lastly, Appendix D gives a brief introduction to various modeling tools such as Monolix, Datxplore, Mlxplore and Simulx. All of these tools use the same modeling language Mlxtran. This appendix is not a user guide as such, but its content should be enough to provide the minimal required background knowledge to enable any reader to understand how models are implemented in the book's examples and how to run basic examples.