# *Chapter 1*

# *Introduction*

## 1.1 Goals

In follow-up studies different types of outcomes are typically collected for each sample unit, which may include several longitudinally measured responses, and the time until an event of particular interest occurs. The research questions of interest in such studies often require the separate analysis of the recorded outcomes, but in many occasions interest may also lie in studying their association structure. A frequently encountered example of the latter case can be found in biomarker research, where many clinical studies are designed to identify biomarkers with strong prognostic capabilities for event time outcomes. Standard examples include among others, HIV research in which interest lies in the association between CD4 cell counts or viral load and the time to AIDS, liver cirrhosis studies which investigate the association between serum bilirubin and the time to death, and prostate cancer studies in which interest lies in the association between PSA levels and the time to the development of prostate cancer. An important inherent characteristic of these medical conditions is their dynamic nature. That is, the rate of progression is not only different from patient to patient but also dynamically changes in time for the same patient. Thus, the true potential of a biomarker in describing disease progression and its association with survival can only be revealed when repeated evaluations of the marker are considered in the analysis.

To address research questions involving the association structure between repeated measures and event times, a class of statistical models has been devel-

oped known as joint models for longitudinal and time-to-event data. Currently, the study of these models constitutes an active area of statistics research that has received a lot of attention in the recent years. In particular, after the early work on joint modeling approaches with application in AIDS research by Self and Pawitan (1992) and DeGruttola and Tu (1994), and the seminal papers by Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) who introduced what it could be nowadays called the standard joint model, there has been an explosion of developments in this field. Numerous papers have appeared proposing several extensions of the standard joint model, including among others, the flexible modeling of longitudinal trajectories, the incorporation of latent classes to account for population heterogeneity, the consideration of multiple longitudinal markers, modeling multiple failure times, and the calculation of dynamic predictions and accuracy measures.

The primary goal of this monograph is to offer a comprehensive introduction to this joint modeling framework. In particular, we will focus on the type of research questions joint models attempt to answer and the circumstances under which these models are appropriate to answer these questions. We will explain which are the key assumptions behind them, and how they can be optimally utilized to extract relevant information from the data. An additional aim of this book is to promote the use of these models in everyday statistical practice. To this end, (almost) all the theoretical material covered in the text is illustrated in real data examples using package **JM** (Rizopoulos, 2012b, 2010) developed for the R software environment for statistical computing and graphics (R Development Core Team, 2012).

## 1.2 Motivating Studies

### *1.2.1 Primary Biliary Cirrhosis Data*

Primary biliary cirrhosis (PBC) is a chronic, fatal, but rare liver disease characterized by inflammatory destruction of the small bile ducts within the liver, which eventually leads to cirrhosis of the liver. The dataset we consider here comes from a study conducted by the Mayo Clinic from 1974 to 1984 (Murtaugh et al., 1994) that includes 312 patients, 158 randomized to D-penicillamine and 154 to placebo. The outcome of primary interest was patient survival and whether this could be prolonged by D-penicillamine. In addition, we have information on baseline covariates (e.g., age at baseline, gender, etc.), and follow-up measurements for several biomarkers. These included among others, serum bilirubin, the presence of spiders (blood vessel malformations in the skin) and hepatomegaly (enlarged liver). Here we will focus on the serum bilirubin level which is considered a strong indicator of disease progression, and in particular, we are interested in the association of this marker with survival. The original clinical protocol for these patients specified visits

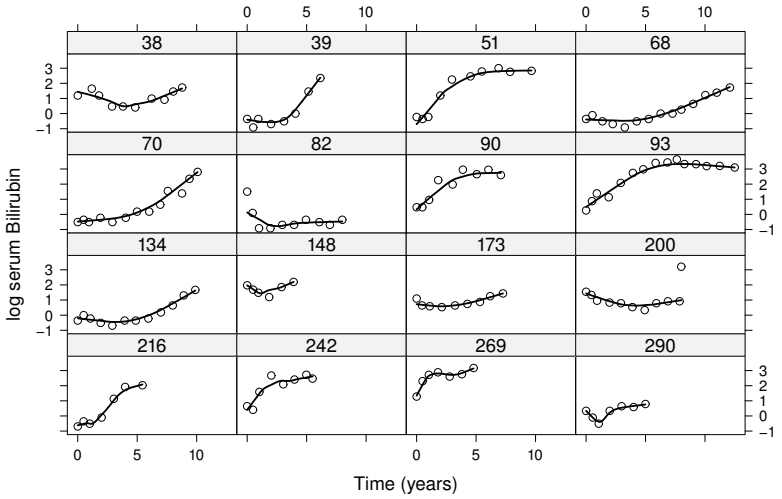


**FIGURE 1.1**: Smooth longitudinal profiles of 16 subjects from the PBC dataset. The solid line represents the fit of the loess smoother.

at six months, one year, and annually thereafter. However due to death and censoring, patients made on average 6.2 visits (st.dev. 3.8 visits), resulting in a total of 1945 observations of serum bilirubin. By the end of the study 140 patients had died, 29 received a transplant, and 143 were still alive. Figure 1.1 shows smoothed longitudinal profiles of the log serum bilirubn for a sample of patients, from which it can be seen that many of these profiles are nonlinear in time.

### *1.2.2 AIDS Data*

In the AIDS dataset we consider 467 patients with advanced human immunodeficiency virus infection during antiretroviral treatment who had failed or were intolerant to zidovudine therapy. The main aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC), in the time-to-death. Patients were randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry, where randomization took place, as well as at 2, 6, 12, and 18 months thereafter. More details regarding the design of this study can be found in Abrams et al. (1994).

By the end of the study 188 patients had died, resulting in about 59.7% censoring, and out of the 2335 planned measurements, 1405 were actually recorded, leading to 39.8% of missing responses. Figure 1.2 presents the Kaplan-Meier estimate of the survival function for the time to death as well

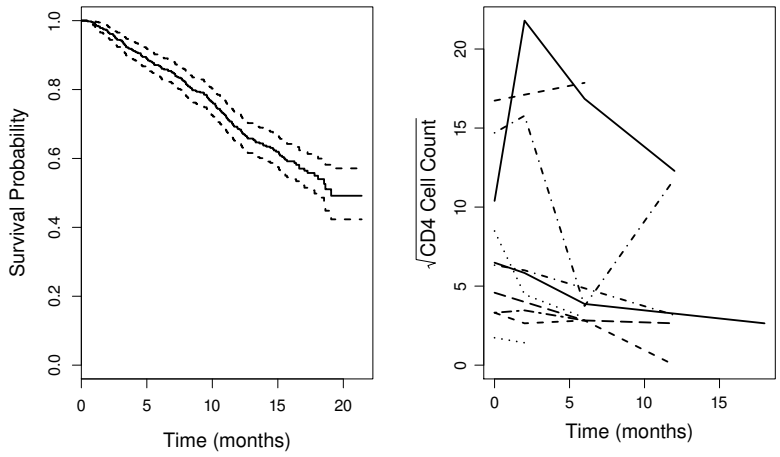


**FIGURE 1.2**: Left panel: Kaplan-Meier estimate of the survival function for the time-to-death in the AIDS dataset. The dashed lines correspond to 95% pointwise confidence intervals. Right panel: Longitudinal trajectories for square root CD4 cell counts for six randomly selected subjects in the AIDS dataset.

as longitudinal trajectories of the square root of the CD4 cell count for a random sample of ten patients (for more details on the Kaplan-Meier estimator the reader is referred to Section 3.2). For our illustrations we focus in one of the secondary aims of this study, which was to study the association structure between the CD4 count and the risk for death for these advanced HIV infected patients. In particular, the CD4 cells are a type of white blood cells made in the spleen, lymph nodes, and thymus gland and are part of the infection-fighting system. The CD4 count measures the number of CD4 cells in a sample of blood and constitutes an important marker of the strength of the immune system. Therefore, a decrease in the CD4 cell count over time is indicative of a worsening of the condition of the immune system of the patient, and thus to higher susceptibility to infection.

### *1.2.3 Liver Cirrhosis Data*

The Liver Cirrhosis dataset includes 488 patients with histologically verified liver cirrhosis, with 251 patients randomized to a treatment with prednisone and the remaining received placebo. Liver cirrhosis is a generic term that includes all forms of chronic diffuse liver disease characterized by extensive loss of liver cells and extensive disorganization of the hepatic lobular architec-

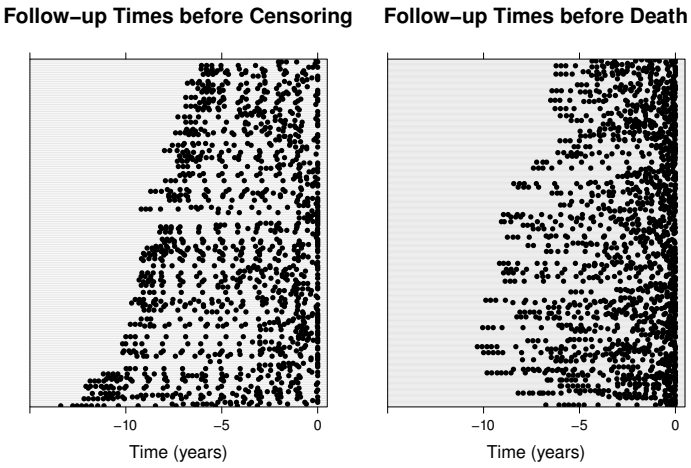


**FIGURE 1.3**: Distribution of follow-up times before censoring (left panel) and death (right panel) for the Liver Cirrhosis data.

ture. The study took place from 1962 to 1974 in Copenhagen, and its main purpose was to evaluate whether prednisone prolongs survival for patients with cirrhosis (Andersen et al., 1993). By the end of follow-up, 142 (56.6%) prednisone-treated, and 150 (63.3%) placebo-treated patients died.

Patients were scheduled to return at 3, 6, and 12 months, and yearly thereafter, and provide records for several clinical and biochemical variables. The clinical variables included information on alcohol consumption, nutritional status, bleeding and degree of ascites, whereas the most important biochemical variables are albumin, bilirubin, alkaline phosphatase and prothrombin. Even though patients were supposed to provide measurements on the aforementioned predetermined visit times, the actual follow-up times varied considerably around the scheduled visits. Moreover, as it can be seen from Figure 1.3, patients who died had more visits taking place shortly prior to death.

For our illustrations we will concentrate on the association between the prothrombin index and the risk for death. This index is a measurement based on a blood test of coagulation factors II, VII, and X produced by the liver. Figure 1.4 depicts the subject-specific longitudinal trajectories per treatment group. In addition, we are interested in investigating the capability of the prothrombin index in discriminating between subjects who died within a medically relevant time interval after their last assessment and subjects who lived longer than that. That is, for a future patient from the same population, we would like to inform the treating physicians about her survival probability that is calculated based on her baseline covariates and her available prothrombin measurements, and assist them in further deciding upon their actions.

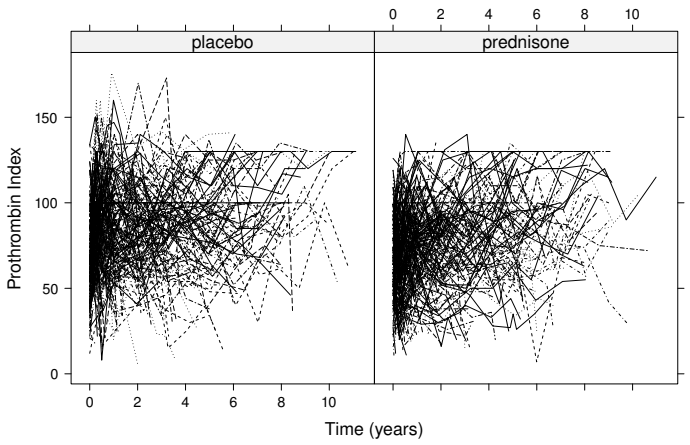


**FIGURE 1.4**: Subject-specific longitudinal trajectories of the prothrombin index for the Liver Cirrhosis data per treatment group.

### *1.2.4 Aortic Valve Data*

The Aortic Valve dataset includes 289 patients with Aortic valve Stenosis (AS) who underwent allograft aortic root replacement (RR) or subcoronary implantation (SI) procedures at Erasmus University Medical Center between 1992 and 2005 (Takkenberg et al., 2002, 2006). Aortic stenosis occurs when the opening of the aortic valve located between the left ventricle of the heart and the aorta is narrowed, and is one of the most common valvular heart diseases. All patients in this dataset have been followed-up prospectively by annual telephone interviews and through visits to their cardiologist. Echocardiographic follow-up at Erasmus MC were obtained at six months postoperative, one year postoperative, and thereafter, biennially by means of serial standardized echocardiography. By the end of the study 61 (21.1%) patients have died and 78 (27%) had a re-operation.

Here we are interested in the association between the aortic jet velocity (aortic gradient) and the risk for death or re-operation. Due to the fact that the aortic gradient levels exhibit right skewed shapes of distribution, we typically work with their square root transform. Figure 1.5 shows the scatterplots of the subject-specific intercepts and slopes, from a simple linear regression for the square root of the aortic gradient, ordered according to the event time. We observe much greater variability in the subject-specific intercepts than in the slopes, and for the latter we see that the variability decreases for later event times. This is partially explained by the fact that as time progresses more aortic gradient measurements are recorded for each patient, which in turn results in a more accurate estimate of the slope.

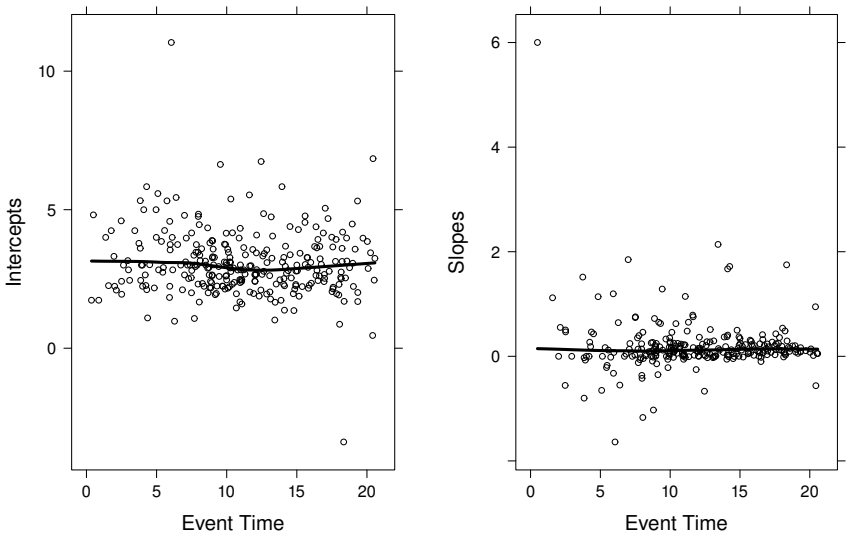


**FIGURE 1.5**: Subject-specific intercepts and slopes per type of operation for the evolution of the square root of the aortic gradient in time for the Aortic Valve Dataset. Solid lines denote the fit of the loess smoother.

### *1.2.5 Other Applications*

The previous sections have focused in datasets from human studies focusing in the association between longitudinal biomarker measurements and patient survival. Nevertheless, the interrelationships between longitudinal responses and event times are of interest in many other disciplines as well. Below we give three such examples from other areas than biomarker research.

Ex1: In gregarious animal studies longitudinal measurements of the sociodynamics of the herd may be associated with the time to relocation to an other area.

Ex2: In sociology and educational testing but also in some epidemiological studies it is often of interest to relate the performance of respondents to questionnaires to event time outcomes. For example, in dementia research questionnaires are used to measure the status of a patient's mood, and her memory and communication capabilities. Since dementia is progressive, patients take these questionnaires at frequent intervals and interest lies in the relation between the evolution of the performance of a patient in these psychometric tests and the clinical onset of the disease.

Ex3: In civil engineering it is often of interest to study the time until a building is no longer useable. To achieve this several indicators of structural integrity are recorded at regular time intervals, with aim to evaluate if these indicators are strong predictors for the risk of failure of the structure in question.

## 1.3 Inferential Objectives in Longitudinal Studies

It is evident from the previous section that in longitudinal studies typically a wealth of information is recorded for each patient, and interest often lies in complex associations between these different pieces of information. Before discussing in more detail possible interrelationships that could be of interest in a longitudinal setting, we will first make a distinction between two types of outcomes recorded in these studies. First, we will call *explicit* outcomes the outcomes that are explicitly specified in the study protocol to be recorded during follow-up. For example, in the PBC study these include the survival information on the patients, and the longitudinal measurements on biomarkers. The second type of outcomes, which we call *implicit* outcomes, are the outcomes that are not of direct interest but nonetheless may complicate the analysis of the explicit ones. A characteristic example is missing data. In particular, even though according to the protocol patients are typically required to appear at the study center(s) at prespecified occasions to provide information, this is rarely done in practice. Patients often miss some of their visits or they may completely drop out from the study for a variety of reasons. For instance, as we saw in the AIDS dataset, out of the 2335 planned measurements, 1405 were actually recorded, leading to 39.8% of missing responses. Another example of an implicit outcome, closely related to the missing data, is the visiting process, which is defined as the mechanism (stochastic or deterministic) that generates the time points at which longitudinal measurements are collected (Lipsitz et al., 2002). Random visit times are more often encountered in observational studies where the time points at which the longitudinal measurements are taken are not fixed by design but rather determined by the physician or even the patients themselves. Nonetheless, random visit times may even occur in randomized studies that have pre-specified by the protocol visit times. For example, for the PBC dataset and during the first two years of follow-up, measurements of serum bilirubin were taken at baseline, half, one, and two years, with little variability, whereas, in later years the variability in the visit times increased considerably. In the following we present a categorization of the possible research questions we could formulate in a longitudinal study.

### *1.3.1 Effect of Covariates on a Single Outcome*

The most common type of research questions in medical studies, in general, is to estimate or test for the effect of a set of covariates in some outcomes of interest. For example, for the AIDS patients we would like to know whether ddI improves the survival rates, or whether there is a difference in the average longitudinal profiles of the CD4 cell counts between males and females. The answer to such questions requires to postulate a suitable statistical model that relates the covariate(s) to the outcome of interest. Depending on the nature of the outcome several types of statistical models are available. A review of the basic modeling frameworks for longitudinal and event time data is given in Chapters 2 and 3, respectively.

### *1.3.2 Association Between Outcomes*

Often it is also of interest to investigate the association structure between outcomes. For instance, in the PBC dataset physicians are interested in measuring how strongly associated is the current level of serum bilirubin with the risk for death. A similar example occurs in asthma studies, where the risk for an asthma attack may be correlated with the levels of air pollution. At first glance, these research questions seem similar in spirit to the ones posed in Section 1.3.1, with the only difference being that the covariate process is now time-dependent. Thus, one could simply proceed by postulating suitable models that relate the two outcomes of interest. For example, we could simply formulate a time-dependent Cox model for the hazard for death and include the longitudinal CD4 cell count measurements as a time-dependent covariate (Andersen and Gill, 1982). Nevertheless, an important feature that we need to carefully consider is the fact that in such models the outcome variables play the role of both the response and the covariate. To proceed in this setting we first need to discern the type of the covariate-outcome process, and in particular whether the covariate-outcome is internal (also known as endogenous) or external (also known as exogenous) to the response-outcome. Formal definitions of endogeneity are given later in Chapter 3. However, a more intuitive manner to distinguish between internal and external covariates is by understanding the nature of a time-dependent covariate process. To put it loosely, internal covariates are generated from the patient herself and therefore require the existence of the patient. Revisiting the previous two examples, we note that the CD4 cell count and the hazard for death are stochastic processes generated by the patient herself, and therefore the CD4 cell count constitutes an internal covariate process. On the other hand, air pollution is an external covariate to asthma attacks, since the patient has no influence on air pollution. When the covariate-outcome is external to the response-outcome, we can use the majority of the standard models mentioned in Section 1.3.1, with relatively small modifications. However, as we will see later, statistical analysis with internal covariates poses several additional difficulties.

### 1.3.3 Complex Hypothesis Testing

Combinations of the previous two types of research questions are also often of interest. A typical example of this setting constitutes the evaluation of surrogate markers. In particular, for chronic conditions, such as PBC, we could be interested to assess treatment efficacy using the short term longitudinal serum bilirubin measurements instead of the survival endpoint, which is lengthy to ascertain. Prentice (1989) set three conditions for surrogacy: (I) treatment must have an effect on patient survival; (II) treatment must have an effect on the marker, i.e., serum bilirubin; and (III) the effect of treatment should manifest through the marker, i.e., the risk for death given a specific marker trajectory should be independent of treatment. It is evident that to assess conditions (I) and (II) we need to posit separate models for the survival and longitudinal outcomes each one containing treatment as a predictor. However, to check condition (III) a model for the survival outcome that conditions on both treatment and serum bilirubin is required instead. Given the special characteristics of serum bilirubin as an endogenous time-dependent covariate explained above, joint models provide a flexible modeling framework to determine whether treatment has an effect on survival after accounting for serum bilirubin.

A similar type of analysis is required when we are interested in simultaneously testing for the effect of a baseline covariate in several outcomes. For instance, continuing on the same example mentioned above, serum bilirubin may not be a good biomarker in describing disease progression, and therefore treatment may still have an influence on patient survival, even after conditioning on serum bilirubin. In this situation we could extend our analysis, and include additional biomarkers of disease progression, such as spiders and hepatomegaly. Interest could then be in testing for the effect of treatment in all markers simultaneously or in testing for the association of one specific marker with the risk for death after correcting for the other markers. It is evident that in order to perform such tests we need a modeling approach that can flexibly capture the interrelationships between these outcomes.

### 1.3.4 Prediction

Statistical models are also often built to provide predictions of patient-related outcomes. In particular, due to current trends in medical practice towards personalized medicine, models that can provide subject-specific predictions of high quality can be proven quite valuable. In practice, for a specific patient and at a specific time point during follow-up, physicians would like to utilize all available information they have at hand (including both baseline information and accumulated biomarker levels) to produce predictions of medically relevant outcomes, gain a better understanding of the disease dynamics, and ultimately take the most optimal decision at that time. When new information

is recorded, physicians would be interested in updating these predictions, and therefore proceed in a time dynamic manner.

When good quality predictions are of interest, it would be useful to combine all available information we have for a patient in order to account for the biological interrelationships between the outcomes. In the PBC dataset for example, it is clear from the definition of the biomarkers that they measure different aspects of liver functioning. Thus, if we were to base predictions on one of those markers and ignore the others, we would discard valuable information. This would unavoidably imply that we would not reach the maximum of the predictive capability that we could have achieved had all biomarkers been simultaneously combined. It is evident therefore that a modeling approach that combines all markers in a single model is advantageous because it utilizes all available information. The added value of combining markers for prediction has been empirically illustrated by Fieuws et al. (2008) who noted that predictions of graft failure in a kidney transplant study based on a joint model using all recorded biomarkers of kidney functioning substantially outperformed the separate analyses per marker.

### *1.3.5 Statistical Analysis with Implicit Outcomes*

In all the above types of research questions we have focused on explicit outcomes. However, as mentioned earlier, in longitudinal studies more often than not implicit outcomes are also generated and their appropriate handling is required even though they are not the outcomes of primary interest. In particular, in the presence of implicit outcomes, and before proceeding in the analysis of interest one must carefully consider the nature of the probabilistic mechanism describing the process generating the implicit outcome(s) (missing data and/or visit times) because it can greatly determine how the analysis should be adjusted in order to obtain valid inferences.

## 1.4 Overview

Chapters 2 and 3 aim at introducing the building blocks of joint models, namely linear mixed-effects models for longitudinal data and relative risk models for survival data. In particular, in Chapter 2 we discuss the complications arising in the analysis of longitudinal responses, and we introduce the linear mixed-effects model as a flexible modeling framework to handle correlated data. We refer to estimation and inference, and then focus on the problem of missing data that is frequently encountered in longitudinal studies. We define the different missing data mechanisms and explain under which circumstances the linear mixed model provides valid inferences.

Chapter 3 starts by explaining the special features of event time data, such as censoring and truncation, and how they complicate the analysis of

such data. Following, we introduce relative risk models and in particular the Cox model. As in Chapter 2, we briefly refer to estimation, under partial and full likelihood, and inference. For the last part of this chapter we focus on time-dependent covariates. More specifically, we provide the definitions of endogenous and exogenous time-dependent covariates, and we discuss under which settings the extended (time-dependent) Cox model provides valid inferences.

Chapter 4 introduces the basics of the joint modeling framework. In particular, continuing from the end of Chapter 3, we motivate joint models first from the survival point of view as a modeling framework to handle endogenous time-dependent covariates. We introduce the standard joint model, discuss the assumptions behind it, and present maximum likelihood estimation. Following, we make the connection with the missing data framework presented in Chapter 2, and additionally motivate joint models as models that can handle nonrandom dropout.

In Chapter 5 we explore several extensions of the standard joint model. Extensions for the survival part include different types of parameterizations between the longitudinal and survival outcomes, stratified relative risk models, handling of multiple failure times, and the consideration of accelerated failure time models. With respect to the longitudinal part we first present joint models with categorical longitudinal markers, and following we extend to multivariate joint models with multiple longitudinal outcomes. Finally, as an alternative to the standard joint model we present the latent class joint model which assumes that the association between the longitudinal and event time processes is due to the existence of latent heterogeneity in the population.

In Chapter 6 we present several diagnostic tools to assess the assumptions behind joint models based on residuals. We focus on separate types of residuals for the survival and longitudinal parts, respectively, and special attention is given on how these residuals can be affected by the nonrandom dropout induced by the occurrence of events. In addition, we also refer to misspecification of the random-effects distribution and how this affects the derived inferences.

Chapter 7 focuses on prediction and discrimination. More specifically, we illustrate how joint models can be used to estimate survival probabilities for the event time outcome and predictions for the longitudinal outcome, and illustrate how these are dynamically updated as additional information is collected for each subject. Following, we turn our attention to prospective accuracy measures for the longitudinal marker, and assess its capability in distinguishing between subjects who are about to experience the event and subjects who have a much lower risk. In particular, under a general definition of prediction rules, we present suitable definitions of sensitivity and specificity measures, and we determine the longitudinal marker's accuracy using receiver operating characteristic methodology.

Finally, Appendix A provides a brief introduction to the R language such that readers with no or little experience with this software package obtain the

minimal required background knowledge to enable them to apply the joint modeling techniques presented in this text in their own datasets.