BOOK REVIEWS

EDITOR:
DONNA PAULER ANKERST

GERHARD TUTZ AND MATTHIAS SCHMID. **Modeling Discrete Time-to-Event Data**. Heidelberg: Springer.

Time is a continuous phenomenon, and consequently, the many textbooks on survival analysis focus on time-continuous techniques. On the other hand, data are discrete, and at least some rounding will be present in the observed event times. Continuous-time survival analysis allows for events to occur at any time, but in principle requires a kind of data accuracy that will not always be present in applications. This is most obvious for tied event times, but also the analysis of interval-censored data profits from discrete time survival methodology, to name just two fields. And it is not just about data accuracy: Causal reasoning, for instance, has deep roots in discrete time models. This being said, it is somewhat of a surprise that the textbook by Tutz and Schmid is, to the best of my knowledge, the first book entirely devoted to the analysis of discrete time-to-event data, and it consequently is a welcome addition to the literature.

This is an applied book with a focus on statistical modelling of discrete survival data. The book is accompanied by the R package discSurv written by Thomas Welchowski and the second author. Throughout, applications of the statistical methodology are illustrated with real life examples, and most of the data sets are freely available. After an introductory Chapter 1 and presentation of the time-honored life table approach in Chapter 2, central ideas are presented in Chapter 3 on Basic Regression Models. Here, the authors clearly demonstrate how to exploit that discrete-time hazards are just conditional probabilities, establish the connection to the powerful theory of generalized linear models and demonstrate how standard statistical software (including but not restricted to R) can be used to fit such models. The careful reader of Chapter 3 will be well equipped to undertake his or her (discrete-time) survival analysis in practice.

But the book does not stop there. The following Chapters 4–7 offer material not necessarily found in a standard survival textbook. To name some examples, Chapter 4 on Evaluation and Model Choice includes a presentation of the Brier Score for survival data, Chapter 5 considers more flexible modelling of the hazards using smoothing techniques, Chapter 6 presents tree-based approaches, such as Classification And Regression Trees (CART) or Bagging, and Chapter 7 explains penalized likelihood and boosting approaches for

high-dimensional covariate data. These chapters will be of interest to both researchers in the field and data analysts who wish to apply such methods. Chapters 8 and 9 deal with the important topics of competing risks and frailties, respectively, and the final Chapter 10 explains how the methods at hand extend to multiple spells, that is, event histories.

Finally, it should be noted that the book is not a rigorous mathematical treatment of the discrete-time approach. As the authors state on the very first page, the book's "mathematical level is moderate" and the "focus [is] on basic concepts and data analyses." While the omission of the mathematical details serves the presentation well (e.g., in Chapter 3, where use of standard generalized linear models software is explained), it whets the appetite for another book on discrete-time survival analysis. For example, Section 2.1 of the book by Aalen, Borgan, and Gjessing (2008), gives a flavor of how the mathematics of the fundamental counting process approach simplify in discrete time. The authors Gerhard Tutz and Matthias Schmid are to be congratulated for opening the doors in this often overlooked topic.

In summary, I am happy to recommend this clearly written text to anyone working with survival data. Working with survival data, one will sooner or later be confronted with a problem that can or should be attacked in discrete time. But even when the models stay time-continuous, the material offered in Chapters 4–7 makes the present book stand out and a worthwhile read irrespective of the approach taken, time-discrete or time-continuous.

References

Aalen, O., Borgan, Ø., and Gjessing, H. (2008). Survival and Event History Analysis. New York: Springer.

Jan Beyersmann
Institute of Statistics
Ulm University
Ulm, Germany
jan.beyersmann@uni-ulm.de

CHUL AHN, MOONSEONG HEO, AND SONG ZHANG. **Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research**. Boca Raton: CRC Press.

With the huge literature in technical reports, journal articles, blogs, multiple sample size and power packages, and specially written procedures in SAS, R, STATA, MatLab, and other statistical software, one might wonder why we need another book on sample size calculations. The answers will be evident upon reading this text. I write from the point of view of a biostatistician situated in a medical school biostatistics department, scientific reviewer, and consultant. I sometimes need sample size and power calculation routines for clustered and randomized designs not always found in commercially available sample size software or, if available, not in the form I need.

This book offers unified notation, intuitively clear power, and sample size formulas in closed form, and well written explanations in each chapter, making the book easy to use, and especially after a first read, a quick and authoritative reference for the calculations needed to get the job done. Many worked examples are shown, but no computer code is included in the text. The sample size and power formulas are in closed form however, and follow a clear progression from the familiar two-sample t-test sample size to the more complex clustered, repeated measures and multilevel designs. The familiar probit model is still there but accompanied by variance inflation factors and other adjustments particular to the design. Being in closed form, the expressions are easily coded and intuitive.

The first chapter reviews sample size expressions for continuous and binary independent outcomes, one and two sample designs, and balanced and unbalanced two sided designs, setting the notation and pattern of presentation for the rest of the book. Clustered outcomes are addressed in Chapter 2 for clusters of equal fixed size and random size. Chapter 3 covers repeated measures designs, including power for treatment group differences in rates of change across time with and without adjustment for baseline covariates, and the time averaged difference (TAD), especially useful when the treatment has a rapid onset and repeated measures thereafter. Chapter 4 is directed to the generalized estimating equation approach in comparison with the mixed effects model for correlated outcomes, with two or more treatment groups, and displays the effects of missing data on sample size and power. Chapters 5 and 6 focus on correlated outcomes in two- and three-level randomized clinical trials with fixed and random slopes. They include demonstrations of the relative effects of the number of subjects and the number of measurements per subject in terms of the autocorrelation assumptions, compound symmetry (CS), and autoregressive lag 1 (AR(1)). They address alternative hypotheses pertaining to slopes, and designs with randomization at the second and third levels.

I find the displays of the effects of missing data important and useful, especially in grant writing. I normally look for books with computer code in the examples, the lack of code in this text is hardly a decrement because worked examples are given and the expressions are easily coded. This is a book that should be in the office, having it gives me the peace of mind that I can address complex designs with the rigor I want in manuscripts, reviews, and grant writing.

Joel Michalek
Department of Epidemiology and Biostatistics
University of Texas Health Science Center at San Antonio
San Antonio, Texas, U.S.A.
michalekj@uthscsa.edu

MARCO SCUTARI AND JEAN-BAPTISTE DENIS. **Bayesian Networks: With Examples in R**. Boca Raton: CRC Press.

Bayesian networks (BNs) have seen as varied applications as the broader subject Statistics itself and the research activity continues at full throttle over the past and current decade. The string search "Bayesian Network" at

www.scholar.google.com for the periods 1981–90, 1991–2000, 2001–2010, 2011–2017, respectively, yields 196, 3720, 25,800, and 22,200 results. Though the results do not necessarily translate into research articles published on the topic, it is a fairly good indication of the interest that has been evolving on the topic of BNs. It is an area of interest to both the Statistics and Machine Learning community and it has witnessed alike contributions from the two communities. With R as the popular software choice, the current book makes a compelling reading for anybody engaging with these networks. Both the authors have among themselves created some of the most useful R packages for developing the networks.

A BN is a graphical representation of nodes, denoting random variables, with conditional dependencies represented through uni-directional arcs. Some restrictions are placed on the network structure, such as there should be no loops, leading to the famous directed acyclic graph (DAG). The simplest example of a DAG is when we try to model whether the grass is wet (W) because of rain (R) or the sprinkler (S). Here, it is natural to specify that R and S lead into W. Since the sprinkler is rarely turned on during rain, R also explains whether S is used or not. Thus, nodes, arcs, and the network structure are the main constituents of a BN. A few intuitive questions that a beginner may have about BNs can be listed as follows.

1. How does one specify the network structure?
2. How are the weights of the network determined? Is it driven by intuition, experience, and/or statistical inference?
3. Are there analytical methods that will automatically detect the structure of the network?
4. How does one measure the strength of the relationship between the nodes? Is there an overall measure of network goodness/fit?
5. If one or more arcs are forcibly deleted/removed, how is the network affected?

We next look at how the book through its six chapters and three appendices provides the answers for the above set of questions.

Chapter 1 naturally begins with discrete BNs, and hence the choice for the underlying statistical models is the multinomial distribution. DAG networks are intuitively specified through an artificial survey example for the mode of transport used to commute from home to office. Here, the node structure is carefully specified and using the factor levels of six variables and plausible probabilities, establishment of an expert system is completed. Given raw data, inference about the network parameters using maximum likelihood and Bayesian estimation is illustrated. Diagnostics for the inferred DAG structure are carried through conditional independence tests and network scores. The presentation is natural and richly performed as the questions asked earlier are clearly and pedagogically explained, specifically, questions 1, 2, 4, and 5 are addressed in the context of a multinomial BN.

The second chapter builds the networks for continuous variables, which explains its title containing "Gaussian Bayesian Networks." Here, the average impact of the parent node on the mean of the daughter node is specified through the simple linear regression model. In an illustrative example that runs throughout the chapter, crop yield is the terminal node, preceded by the number of seeds and their mean weight, whose two nodes in turn are determined by the vegetative organs node. The initial nodes of environmental potential and genetic potential lead into the vegetative organs node. The structure of this chapter is the continuous variable extension of the discrete BN, making it more readily understood for the reader who has proceeded through the beginning chapter. DAG structure and other diagrams make the flow clear and easier to follow.

The third chapter fits networks where some of the node variables may be discrete and the rest continuous. Such a scenario is more practical and hence its detailed treatment is a requisite. The complexity arising as a nature of the mixture type of variables is handled via a Monte Carlo Markov chain framework and the authors illustrate the inference through use of the JAGS software.

Chapter 4 forms the central and most important part of BN. Yes, it contains the theoretical details of the workings of the networks. The important concept of d-separation begins the chapter and other useful definitions are provided in the second section. The technical aspect of Markov blanket is then dealt with. The third question from the earlier list that asks whether we have methods that can automatically detect the BN structure is answered here in the form of the hill climbing algorithm. Crucially, the function bnlearn along the with the options of whitelist and blacklist give rich infrastructure for building BNs. The chapter concludes with the causality aspect of BN.

Chapter 5 provides an overview of the software infrastructure for analysis of BN in R. Table 5.1 especially lays out the comparison of seven important BN R packages:bnlearn, catnet, deal, pcalg, gRbase, gRain, and rbmnacross 10 software features. Some of the packages' deficiencies may not be valid anymore since the table was prepared back in 2014. Detailed descriptions of each of the packages follows the table. The third section of the chapter details software other than R and such comparison is required for the reader who is interested in BN.

The concluding chapter of the book considers two important real-world problems from life sciences. Network query is an additional measure elaborated in this chapter with an application. More importantly, the chapter exposes the reader to the suitability of BN for new scenarios and gives the required confidence for its use in their own world. Reference and complementary material is suggested in each chapter's "Further Reading" section. The Appendix is useful for some of the theory used in earlier parts of the book.

In summary, the book is undoubtedly useful for students and practitioners. The reviewer genuinely feels that this book is a must on the shelf of anybody that indulges in BN.

Prabhanjan Tattar
Global Data Insight & Analytics (GDI&A)
Ford Motor Company, Chennai, India
prabhanjannt@gmail.com

JINQIAO DUAN. **An Introduction to Stochastic Dynamics**. New York: Cambridge University Press.

Rigorous treatment of stochastic dynamical systems requires a high level of technical effort. This effort, however, discourages students and researchers of applied sciences. As many applied problems ask for this framework, there is a need for non-technical introductions to this topic, which is exactly the point the present work aims to accomplish. The book does not intend to only show simulation techniques and informal interpretations of simulated results, but rather to communicate the proper mathematical tools. A balance between informal explanations and precise mathematical statements is achieved by simultaneously introducing a wide range of the basic mathematical structures while quoting the fundamental theorems without proof (but with appropriate literature hints). The mathematical techniques become clear in many examples on the one hand, and in proving less demanding theorems on the other hand. The book focuses on stochastic differential equations and their treatment, while branching processes and interacting particle processes are not covered.

The preface indicates that a profound basic education in elementary stochastics is required, up to some ideas about stochastic differential equations. In my opinion, the book provides enough material to allow students and applied scientists to catch up without a deep prior insight into these topics, though some thoughts and ideas are surely more simple to grasp with deeper background knowledge. Physicists, theoretical chemists, and biologists with a sound mathematical education should be able to access this book, particularly, since the first four chapters briefly recall most of the relevant mathematical terms, including an introduction into stochastic differential equations.

The centerpiece of the text comprises Chapters 5–7. Here, nontrivial concepts are explained, starting with classical aspects of stochastic differential equations, as first exit time and exit probability in Chapter 5. Readers particularly interested in these topics are perhaps better off with the book by Grasmann and van Herwaarden (1999).

Chapter 6 discusses concepts of deterministic dynamical systems and how to lift them to stochastic dynamics. For example, invariant manifolds and the Hartman–Grobman theorem are explained in the context of stochastic differential equations. Therefore, random maps, Wiener shifts, and cocycles are introduced. The standard reference here is the classical book by Arnold (1998). The merit of the present book is to explain these sophisticated terms in a way accessible for the audience at hand, and to use these advanced concepts to show how the well-known deterministic concepts can be formulated and applied in the stochastic context.

The last chapter introduces Levy processes. Many applications require this more advanced approach, which is not well known in the applied community. The nice and intuitive explanation offers researchers in applied sciences a first insight into this flexible tool.

This work provides good supplementary literature for seminars and lectures, but due to the missing proofs of the central theorems, it is difficult to use this textbook as a main source in teaching. It is, however, an extraordinarily useful introduction and handbook for any applied scientist working with stochastic dynamical systems.

References

Arnold, L. (1998). *Random Dynamical Systems*. Heidelberg: Springer.

Grasmann, J. and van Herwaarden, O. A. (1999). *Asymptotic Methods for the Fokker–Planck Equation and the Exit Problem in Applications*. Heidelberg: Springer.

JOHANNES MÜLLER
Department of Mathematics
Technische Universität München
Munich, Germany
johannes.mueller@mytum.de

ALEX BOTTLE AND PAUL AYLIN. **Statistical Methods for Healthcare Performance Monitoring**. Boca Raton: CRC Press.

The need for performance monitoring has been driven by both scandals and science. Government, health care professionals, and patients want to know that the care provided is of high quality and good value for its considerable cost. While this is an obvious goal, it is difficult to implement in practice.

After giving a short introduction in Chapter 1, Chapter 2 describes some of the medical scandals, including a case that led to the death of dozens of children in a UK hospital, and a murderous physician in Australia. This chapter also illustrates the lessons learned from these incidents, and the subsequent policy changes in the respective countries. Chapter 3 discusses the difference between the unit of reporting and the unit of analysis. Analyses can be performed at the level of the physician, hospital, or even the whole health authority. And, as the authors show, they can all have different implications on statistical methods and practice. Reporting by physician implies that the physician alone is responsible for their patients' care, but of course they do not work in isolation. Nurses and other health providers play an important role as well.

The different indicators of quality of care are explained in Chapter 4. The definitions of quality are diverse, but common to all of them are indicators such as safety, effectiveness, and efficiency. In practice, and depending on the purpose, some are more important than others. After defining what to measure, Chapter 5 picks up on the assessment of the quality of the data. It gives an overview of the main types of data for performance monitoring and briefly covers related issues such as linkage and ethics. Administrative data, for example, are usually recorded without a specific research purpose in mind, while clinical registry data or survey data are purpose-built. Through a wealth of real world examples, this chapter discusses strengths and weaknesses of different data types as well as country-specific considerations. Chapter 6 distinguishes between risk adjustment and risk prediction and the principles behind them. This is also the first time the book

goes into more statistical depth by explaining issues with missing values, variable selection and model fitting metrics. Closely related with the previous risk-adjustment section is the description for producing summaries of the observed and expected performance in Chapter 7. This is the most technical chapter in the book, but very short and still with few formulas.

Composite performance measures combine several indicators into a single index on the basis of an underlying model. They capture multidimensional concepts that single models cannot, and are explained in Chapter 8. Again, this is a bit more technical, but many examples make it easier to understand. Chapter 9 is one of the longer chapters and deals with different ways to compare units in terms of their performance. This is probably the most crucial part of health care performance monitoring as we want to know whether a unit's performance is "bad," "acceptable," or even "excellent." Graphical indicators like funnel plots are covered, but also Bayesian statistics, multiple testing problems, and variation. They are generally not discussed in great detail, but references to theory and application are always provided. Chapter 10 extends these comparisons across borders and demonstrates the rationale for multi-country comparisons. Presenting the results to stakeholders is described in Chapter 11. This is nicely illustrated with plots and screenshots from real cases. Health care performance monitoring is difficult and expensive, so Chapter 12 discusses evaluation of the monitoring system itself, asking questions such as: does it achieve in practice what we hope that it achieves? Again, this chapter is more heavy on the statistics, it covers concepts such as interrupted time-series, differences-in-differences, and instrumental variables. A subchapter outlines economic evaluation methods.

Overall, the book provides an interesting and easily accessible overview on health care performance monitoring and the statistical methods associated with it. Each chapter has a short overview at the beginning and sometimes a conclusion at the end, so it can also serve as a reference book. According to the authors, this book is not primarily aimed at statisticians, but all who want to compare and measure health care performance. The level of statistics obtained from an undergraduate nursing or medical degree is enough to follow. Furthermore, the authors marked several chapters and subchapters more heavy on statistics that can be skipped without missing the big picture. The rich examples make the book enjoyable to read and it has my unconditional recommendation to all interested in the topic.

Christoph F. Kurz
Institute of Health Economics and Health Care Management
Helmholtz Zentrum München
Neuherberg, Germany
christoph.kurz@helmholtz-muenchen.de

NIELS G. BECKER, **Modeling to Inform Infectious Disease Control**. Boca Raton: CRC Press.

Infectious disease control is certainly a relevant and timely topic worthy of comprehensive treatment in a dedicated text. This new book might be more suited for students and study courses than for experienced researchers, and for such a teaching focus or for an easily accessible reference, it is perfectly suited.

After the introduction, where some motivation and background is given, the book starts with a basic model dealing with minor outbreaks and homogeneous infective agents. Step by step, in the following chapters, complexity of the approaches increases and more effects are included. This stepwise approach increases the accessibility for the reader, helping them to understand, under which assumptions, one needs to consider refined models versus simpler versions that might be sufficient. It emphasizes that no model is perfect, that choosing a model is always predicated on the assumption of certain properties, rendering it an approximation to reality.

Among aspects addressed are populations and their underlying structures, such as households, transmission patterns of disease, and the central role of vaccination. Practical questions, such as what defines a successful vaccination campaign, are answered throughout. A very interesting point is taken up in Chapter 10, which is how to use infectious disease data to inform model choice, that is, within statistical models. As one typically tries to keep models as simple as possible, it is important to deal with this problem. In general, population reproduction numbers run like a thread through the whole books, making it easy to compare the influences of model details on the resulting formulas.

Individual chapter contents are thoroughly prepared. Starting with a short motivation, the models and analyses are displayed in single steps, avoiding the use of humbling phrases such as "as easily checked by the reader." The chapters end with a short discussion summarizing what has been learned, a nice collection of well-suited interesting exercises helping to deepen understanding of the content by "doing yourself," and supplementary material. The latter help the reader decide whether they would like to further research important topics in other sources. Many real-life examples in the main text and exercises are included, including larger data sets, such as the Hagelloch measles epidemic. By doing so, the book helps to understand how to apply the introduced methods not only to "nice academic examples," but to real-life problems, and to learn about the special difficulties there.

One minor critique might be that the content of the book is somewhat narrow. An example is that it exclusively deals with "step by step" models, and does not mention at all other possibilities, such as the classical Kermack–McKendrick model in the form of a system of ordinary differential equations. While this focus limits the comparison of different types of model approaches, it offers the advantage of avoiding confusion for the unexperienced reader. Choice of model type, which is an art of its own, is left to beyond the scope of the book.

As a professor of mathematical modeling, in addition to using the book for instruction, I could envision enlisting its service as the basis for a reading course for master level students, as the explanations are sufficiently detailed to facilitate independent learning. In conclusion, the presented book

provides useful and graspable material for interested readers to enter a relevant and important scientific area, whether they be undergraduate statistics students or advanced researchers in interdisciplinary fields.

CHRISTINA KUTTLER
Mathematics in Life Sciences
Technical University Munich
Garching, Germany
kuttler@ma.tum.de

MAX KUHN AND KJELL JOHNSON, **Applied Predictive Modeling**. New York: Springer.

Understanding the world and predicting its future outcomes has, in a sense, always been in the heart of scientific research. This has become even more true in the current day, where the volume and quality (not in all cases) of data have greatly improved, making it possible to attack even harder prediction problems than a few years ago. The availability of these data sources and the related demand for utilizing them for prediction purposes have led to a fierce development of new analytic techniques aiming to provide accurate predictions. New branches of data analysis have emerged focusing on such problems, with new names, such as *Predictive Analytics*. Two major scientific communities have been driving these developments forward, and even (forcefully) competing with each other, Computer Science and Statistics. The book by Max Kuhn and Kjell Johnson is somewhere on the ridge between these fields, introducing and explaining with practical examples several techniques utilized to tackle the prediction problem.

The book begins with an introductory chapter that sets the scene and presents the datasets that are used throughout the book. The rest of the book is split in four parts, namely, Part I on General Strategies, Part II on Regression Models, Part III on Classification Models, and Part IV on Other Considerations. Part I covers basic concepts on predictive modeling, including among others, how to pre-process the available data by considering transformations, removal or binning of predictors, and how to deal with over-fitting and tuning of model parameters. Part II focuses on regression models. It starts by the basic linear regression model, and then presents nonlinear models, including neural networks and support vector machines, and closes with regression trees and rule-based models. Part III, which focuses on classification models, follows an analogous presentation of the different classification methods. In particular, it first starts with linear discriminant analysis and its cousins, then it turns to nonlinear models with nonlinear discriminant analysis and again neural networks and support vector machines, and finally presents classification trees that include among others random forests, bagged trees and boosting. It closes with a nice discussion on remedies when class imbalance occurs. Finally, Part IV presents approaches for measuring the importance of the predictors considered in a particular problem, then provides a short introduction to feature selection, and closes with a list of factors that may affect model performance and predictions. In the Appendix, a brief introduction to R is given, with special mention to the Applied Predictive Modeling and caret packages.

What is great about this book is that it takes a practitioner's perspective. This is achieved with a combination of three aspects. First, readers are provided with the backbone of the theory one should know before utilizing any of the presented tools. Special attention is given to situations in which key assumptions may be violated that can influence the quality of the derived predictions. Second, there is a plethora of examples demonstrating how each predictive modeling technique can be used or can fail in practice. Motivated from these examples, quite often the authors give recommendations of specific modeling strategies one should follow to analyze the data. And finally, for each of the presented methods and modeling techniques, there is detailed R code available illustrating step-by-step how the analysis could be performed. The availability of the presented R packages makes this analysis a relatively easy task. Two topics, I wished the book had discussed in more detail, are extensions of simple regression models (i.e., linear and logistic regression) to account for non-linearities using (penalized) splines, fractional polynomials or other techniques, and penalized regression. The authors do mention these extensions but arguably in a bit limited manner compared to the proportion of the book devoted to the other approaches. However, from personal experience, often such extensions make simple linear and logistic regression models very competitive against the more modern techniques. In addition, coming from a Department of Biostatistics, I also missed extensions of the presented techniques in the case of censored data that constitute more than half of the applications encountered in biomedical sciences. All in all, I would definitely recommend this book to readers interested in learning about classic and modern predictive modeling techniques.

DIMITRIS RIZOPOULOS
Department of Biostatistics
Erasmus University Medical Center
Rotterdam, the Netherlands
d.rizopoulos@eramusmc.nl

ROBERT ELASHOFF, GANG LI, AND NING LI. **Joint Modeling of Longitudinal and Time-to-Event Data**. Boca Raton: CRC Press.

This book is a comprehensive state-of-the-art treatment of joint models for time-to-event and longitudinal data with numerous applications to real-world problems. Chapter 1 describes 11 data sets that are used as illustration throughout the book. Chapter 2 is an introduction to standard methods for longitudinal data that begins with a reminder of Rubin's classification of missingness mechanisms. It follows with brief presentations of linear and generalized linear mixed models, and generalized estimating equations and their weighting counterparts for missing data, before terminating with a description of multiple imputation. These methods are illustrated by quite complex analyses of several data sets.

Chapter 3 describes methods for survival data analyses with special emphasis on accelerated failure time and competing risk models. Several small examples presenting special cases of these models help the reader to understand their interpretations and interrelationships, but surprisingly, this chapter does not include a real data analysis.

Chapter 4 is the core chapter of the book that introduces joint models for one longitudinal marker and one time-to-event outcome by successively describing numerous published joint models. This chapter is quite different from other textbook or review articles on joint models as the authors focus mainly on the non-ignorable missing data problem. They describe various models proposed for jointly modeling longitudinal data and dropout using either continuous or discrete event time models, possibly with intermittent missing data combining linear mixed, logistic mixed, and Cox models. They also discuss informative visiting processes using frailty models for repeated events. For each model, at least one estimation method is detailed, most often using the Expectation-Maximization(EM) algorithm, but also the Bayesian approach. Most of the presented models are briefly illustrated on real data analyses. Throughout the chapter, many dependence structures between the outcome and the time-to-event are considered, including shared and correlated random effect models, outcome dependent survival models, latent class models and both parametric and semi-parametric estimation procedures. Due to the special focus on non-ignorable missing data, models such as pattern mixtures and terminal decline models,that are generally not treated in monographs on joint models, are described here. On the other hand, this chapter does not contain an intuitive introduction of joint models with a discussion of the different frameworks of use. Section 4.4 presents joint models considered as an extension of time-to-event models with time dependent covariates. This subsection includes a quite complete review of parametric and semi-parametric estimation methods for these models. A joint model using an accelerated failure time component is also described and illustrated. This type of joint model is rarely used because parameter interpretation with time-dependent covariates is difficult; more thorough comments on this point would have been useful. The chapter closes with a brief description of the use of these models for dynamic predictions of events.

Chapters 5 and 6 describe two kinds of extended joint models, with competing events and multiple longitudinal markers and events, respectively. Chapter 5 focuses entirely on correlated random effects models, considering both robust models with t-distributed random effects and models with heterogeneous random effects distributions given covariates, as well as models for ordinal longitudinal markers. As previously, EM and Bayesian estimation algorithms are described for each model and each model is applied to real data sets. Interestingly, a unique data set is used for most of the illustrations in this chapter, allowing comparisons of results.

In Chapter 6, the authors describe different approaches for jointly modeling time-to-event outcomes and several longitudinal markers, using either multiple mixed models with correlated random effects or a common latent process model for the markers. Two models for repeated events are then described using each of these two approaches. Finally, an example of shared random effect models for multiple markers and multiple events with cure fraction is described and illustrated on real data. In this chapter, less emphasis is devoted to estimation algorithms and more on parameter interpretation and underlying model assumptions. A useful list of available software is presented in the Appendix.

Chapter 7 briefly overviews several additional topics. First are indices built from selection models either to identify most influential subjects in case of perturbation toward nonignorability and global sensitivity indices of model parameters to departures from ignorability. Then a graphical method for assessing the goodness of fit of joint models and a method for identifying most influential subjects are illustrated. Two Bayesian methods using mixtures as priors for selecting fixed and random effects are described and illustrated. The book terminates by a brief literature review of joint multi state models and joint models with cure fractions.

In summary, this book is a comprehensive review of the existing literature on joint models, including most extensions of these models, fully parametric or not, for multiple events and multiple markers with a special focus on missingness problems and details about various estimation methods. By emphasizing the most advanced methods, this book usefully completes existing monographs on joint models and will be a helpful reference book for researchers in biostatistics and experienced statisticians, while applied statisticians could also be interested thanks to the numerous examples of real data analyses.

Hélène Jacqmin-Gadda
University of Bordeaux
Bordeaux, France
Helene.Jacqmin-Gadda@isped.u-bordeaux2.fr