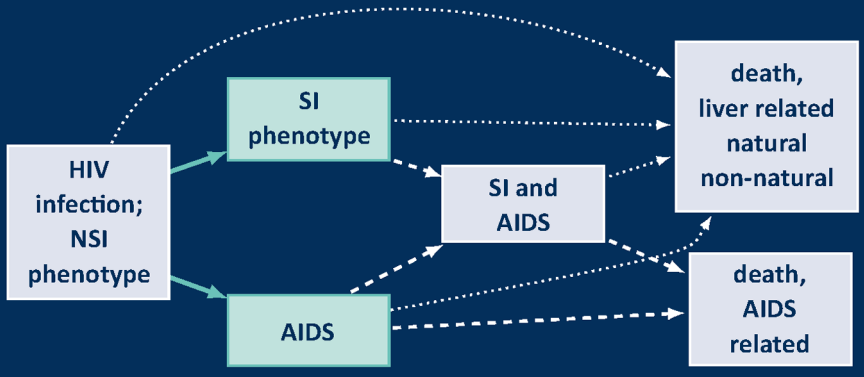
Chapman & Hall/CRC Biostatistics Series

Data Analysis with Competing Risks and Intermediate States



Ronald B. Geskus

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Data Analysis with Competing Risks and Intermediate States

Chapman & Hall/CRC Biostatistics Series

Published Titles

Adaptive Design Methods in Clinical Trials, Second Edition
Shein-Chung Chow and Mark Chang

Adaptive Designs for Sequential Treatment Allocation
Alessandro Baldi Antognini and Alessandra Giovagnoli

Adaptive Design Theory and Implementation Using SAS and R, Second Edition
Mark Chang

Advanced Bayesian Methods for Medical Test Accuracy
Lyle D. Broemeling

Advances in Clinical Trial Biostatistics
Nancy L. Geller

Applied Meta-Analysis with R
Ding-Geng (Din) Chen and Karl E. Peace

Basic Statistics and Pharmaceutical Statistical Applications, Second Edition
James E. De Muth

Bayesian Adaptive Methods for Clinical Trials
Scott M. Berry, Bradley P. Carlin, J. Jack Lee, and Peter Muller

Bayesian Analysis Made Simple: An Excel GUI for WinBUGS
Phil Woodward

Bayesian Methods for Measures of Agreement
Lyle D. Broemeling

Bayesian Methods in Epidemiology
Lyle D. Broemeling

Bayesian Methods in Health Economics
Gianluca Baio

Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation
Ming T. Tan, Guo-Liang Tian, and Kai Wang Ng

Bayesian Modeling in Bioinformatics
Dipak K. Dey, Samiran Ghosh, and Bani K. Mallick

Benefit-Risk Assessment in Pharmaceutical Research and Development
Andreas Sashegyi, James Felli, and Rebecca Noel

Biosimilars: Design and Analysis of Follow-on Biologics
Shein-Chung Chow

Biostatistics: A Computing Approach
Stewart J. Anderson

Causal Analysis in Biomedicine and Epidemiology: Based on Minimal Sufficient Causation
Mikel Aickin

Clinical and Statistical Considerations in Personalized Medicine
Claudio Carini, Sandeep Menon, and Mark Chang

Clinical Trial Data Analysis using R
Ding-Geng (Din) Chen and Karl E. Peace

Clinical Trial Methodology
Karl E. Peace and Ding-Geng (Din) Chen

Computational Methods in Biomedical Research
Ravindra Khattree and Dayanand N. Naik

Computational Pharmacokinetics
Anders Källén

Confidence Intervals for Proportions and Related Measures of Effect Size
Robert G. Newcombe

Controversial Statistical Issues in Clinical Trials
Shein-Chung Chow

Data Analysis with Competing Risks and Intermediate States
Ronald B. Geskus

Data and Safety Monitoring Committees in Clinical Trials
Jay Herson

Design and Analysis of Animal Studies in Pharmaceutical Development
Shein-Chung Chow and Jen-pei Liu

Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition
Shein-Chung Chow and Jen-pei Liu

Design and Analysis of Bridging Studies
Jen-pei Liu, Shein-Chung Chow, and Chin-Fu Hsiao

Design and Analysis of Clinical Trials for Predictive Medicine
Shigeyuki Matsui, Marc Buyse, and Richard Simon

Design and Analysis of Clinical Trials with Time-to-Event Endpoints
Karl E. Peace

Design and Analysis of Non-Inferiority Trials
Mark D. Rothmann, Brian L. Wiens, and Ivan S. F. Chan

Difference Equations with Public Health Applications
Lemuel A. Moyé and Asha Seth Kapadia

DNA Methylation Microarrays: Experimental Design and Statistical Analysis
Sun-Chong Wang and Arturas Petronis

DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments
David B. Allison, Grier P. Page, T. Mark Beasley, and Jode W. Edwards

Dose Finding by the Continual Reassessment Method
Ying Kuen Cheung

Elementary Bayesian Biostatistics
Lemuel A. Moyé

Empirical Likelihood Method in Survival Analysis
Mai Zhou

Frailty Models in Survival Analysis
Andreas Wienke

Generalized Linear Models: A Bayesian Perspective
Dipak K. Dey, Sujit K. Ghosh, and Bani K. Mallick

Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries
Daryl S. Paulson

Inference Principles for Biostatisticians
Ian C. Marschner

Interval-Censored Time-to-Event Data: Methods and Applications
Ding-Geng (Din) Chen, Jianguo Sun, and Karl E. Peace

Introductory Adaptive Trial Designs: A Practical Guide with R
Mark Chang

Joint Models for Longitudinal and Time-to-Event Data: With Applications in R
Dimitris Rizopoulos

Measures of Interobserver Agreement and Reliability, Second Edition
Mohamed M. Shoukri

Medical Biostatistics, Third Edition
A. Indrayan

Meta-Analysis in Medicine and Health Policy
Dalene Stangl and Donald A. Berry

Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools
Marc Lavielle

Modeling to Inform Infectious Disease Control
Niels G. Becker

Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies
Mark Chang

Multiple Testing Problems in Pharmaceutical Statistics
Alex Dmitrienko, Ajit C. Tamhane, and Frank Bretz

Noninferiority Testing in Clinical Trials: Issues and Challenges
Tie-Hua Ng

Optimal Design for Nonlinear Response Models
Valerii V. Fedorov and Sergei L. Leonov

Patient-Reported Outcomes: Measurement, Implementation and Interpretation
Joseph C. Cappelleri, Kelly H. Zou, Andrew G. Bushmakin, Jose Ma. J. Alvir, Demissie Alemayehu, and Tara Symonds

Quantitative Evaluation of Safety in Drug Development: Design, Analysis and Reporting
Qi Jiang and H. Amy Xia

Randomized Clinical Trials of Nonpharmacological Treatments
Isabelle Boutron, Philippe Ravaud, and David Moher

Randomized Phase II Cancer Clinical Trials
Sin-Ho Jung

Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research
Chul Ahn, Moonseong Heo, and Song Zhang

Sample Size Calculations in Clinical Research, Second Edition
Shein-Chung Chow, Jun Shao and Hansheng Wang

Statistical Analysis of Human Growth and Development
Yin Bun Cheung

Statistical Design and Analysis of Stability Studies
Shein-Chung Chow

Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis
Kelly H. Zou, Aiyi Liu, Andriy Bandos, Lucila Ohno-Machado, and Howard Rockette

Statistical Methods for Clinical Trials
Mark X. Norleans

Statistical Methods for Drug Safety
Robert D. Gibbons and Anup K. Amatya

Statistical Methods in Drug Combination Studies
Wei Zhao and Harry Yang

Statistics in Drug Research: Methodologies and Recent Developments
Shein-Chung Chow and Jun Shao

Statistics in the Pharmaceutical Industry, Third Edition
Ralph Buncher and Jia-Yeong Tsay

Survival Analysis in Medicine and Genetics
Jialiang Li and Shuangge Ma

Theory of Drug Development
Eric B. Holmgren

Translational Medicine: Strategies and Statistical Methods
Dennis Cosmatos and Shein-Chung Chow

Chapman & Hall/CRC Biostatistics Series

Data Analysis with Competing Risks and Intermediate States

Ronald B. Geskus
Academic Medical Center
and
Public Health Service of Amsterdam
Amsterdam, The Netherlands

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

ISBN 13: 978-1-4665-7035-1 (hbk)

To life

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Preface

In the end we all die. More interesting than the death event itself is the time component: at what age does one die and what characteristics make some individuals die earlier than others? Survival analysis provides the set of tools that help answer the question of which factors influence the time until the occurrence of some event, which is not restricted to be death.

In the end we all die, but not all from the same cause. Information on the spectrum of causes of death has added value. Figure 1 shows the number of individuals that died of different causes in the twentieth century. The smallest

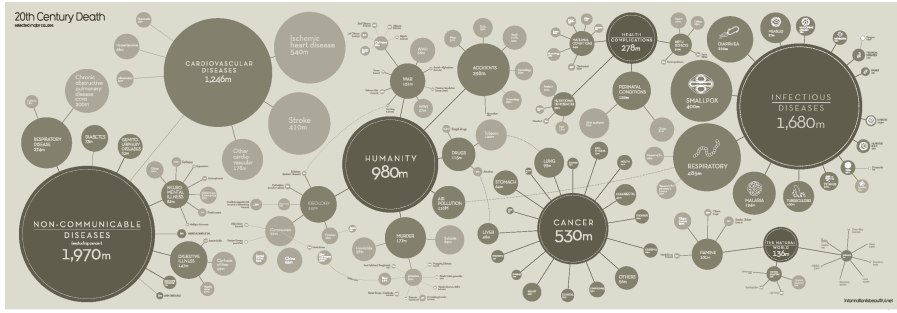


FIGURE 1
All causes of death in the 20th century.

subgroups may not be readable (have a look at the website if you want to read them[1]), but it is seen that the most frequent category is made up of the non-communicable diseases, within which cardiovascular diseases form the largest subgroup. The causes of death are competing: if one occurs, the others do not. Again, including the time component has added value. Some causes, like measles, tend to occur at a young age, whereas others, like prostate cancer, almost exclusively occur in older men. Measles mortality prevents prostate cancer from occurring, but if mortality due to measles is reduced, mortality due to prostate cancer, at a later age, may rise. A competing risks model extends the classical survival setting by considering a collection of mutually

[1]The figure was downloaded on September 1st, 2013 from the website `http://www.informationisbeautiful.net/visualizations/20th-century-death`. Author: David McCandless.

exclusive potential event types. Different causes of death are the classical example, but any set of event types can be considered.

The issues that come up when competing risks are present have often been ignored, even in top (medical) journals [61]. But the situation is changing rapidly. In the last decade, several papers have been published that explain when and why a standard time-to-event approach fails to provide the correct answer. Still, confusion remains with respect to the quantities that can be estimated and their interpretation. Fortunately, once the concepts are understood and the appropriate type of analysis has been chosen, techniques of estimation are not much different from the ones used in classical survival analysis with only one event type.

In the end we all die, but not all for the same reason nor with the same life histories. Even if two individuals die at the same age and of the same cause, their life courses have been different. Between birth and death, intermediate events occur that influence one's life course. The are several ways to model the occurrence of such events and their effect on later ones. One approach is via a multi-state model, in which events are seen as transitions from one state to another. Death may be the final one, but the process under investigation may terminate earlier. Under the frequently assumed Markov property, a multi-state model can be seen as a sequence of competing risks models.

Multi-state models can give a more detailed description of a disease process. Only few articles have explained the use of multi-state models to non-statisticians. Although the range of modeling choices and possible assumptions is larger than in a competing risks setting, there is little additional complexity with respect to interpretation. Also, some of the estimation techniques are a direct extension of those from the competing risks setting such that the same software can be used. Other computations are more complex and require software that has been written specifically for such models.

This book is divided into five chapters and an epilogue. The first chapter introduces the main concepts, with emphasis on the competing risks setting. For a number of examples that will be used throughout the book, we formulate the type of questions that may be of interest and define the corresponding estimands. We explain when the data at hand can be used to answer these questions. We also give an overview of the main definitions and techniques from classical survival analysis that are used and extended in later chapters. In Chapters 2 and 3 we more formally define the concepts that play a role in competing risks and multi-state models respectively. We address nonparametric estimation of the relevant quantities. In Chapters 4 and 5, we quantify the effect of covariables via regression models. In Chapter 4 we explain how and why the ideas and techniques of the classical Cox proportional hazards model extend to settings with competing risks and intermediate states. The difference is in the interpretation of the estimates and their translation to the cumulative scale. The latter is explained in Chapter 5. We also describe two approaches in which parameter estimates have a direct interpretation on the cumulative scale. One quantifies effects on another type of hazard, the sub-

distribution hazard. This model is often called the Fine and Gray model. The third uses a model that quantifies effects directly on the cumulative probability, the proportional odds model. Two of the main difficulties from a practical perspective are the translation of research questions into modeling choices and the interpretation of the results. Therefore, these issues are given a lot of attention throughout the book. The epilogue is completely devoted to this issue.

The last four sections of each chapter have the same structure. We first summarize the concepts that have been introduced, place them in a broader perspective and refer to issues that come up in subsequent chapters. The next section contains exercises that are intended to help understand and reflect on the concepts. Technical exercises are denoted by an asterisk. Answers to the exercises are provided at the end of the book. The exercises are followed by a section devoted to software. We briefly explain options in SAS and Stata. We give a detailed description of the functionality in the R statistical program [83]. Each chapter ends with computer practicals, in which you are asked to practice the concepts in R, using an existing data set on patients that underwent a bone marrow transplantation. In principle, all of the practicals on competing risks and some on multi-state models can also be made using Stata or SAS, but we do not provide any suggestions or answers.

Since we use examples from the biomedical and epidemiological field, the intended readership primarily consists of medical statisticians and epidemiologists. However, the book is useful for any researcher that has some experience with the analysis of standard time-to-event data and wants to extend knowledge and skills to the competing risks or multi-state setting. It should be easy to translate "individuals" and "diseases" to units and phenomena from one's own research area.

With respect to the topics covered and the intended readership, our book is most closely related to the book *Competing Risks and Multistate Models with R* [12]. That book takes a more theoretical perspective and relies fairly heavily on the description via counting processes, whereas our focus is more on interpretation. Yet, we also explain the techniques of estimation and inference and how they extend the setting of classical survival analysis with a single event type. We explain some of the more theoretical aspects in separate sections. Two important topics that are not covered in this book are the imputation of missing event types and estimation via the use of pseudo-values.

With respect to the use of R, we assume that one knows how to install a package, how to execute a script, how to select rows and columns, and we assume some familiarity with the use of functions and help files. In the example R code, we write down the full function name because this helps in finding the appropriate help file. For example, we write `summary.coxph`. Because of R's object orientation, it is sufficient to write `summary` when performing the analyses.

The book has a website, `http://www.competingrisks.org`, where you can find additional information. It has a file with hints for making

the computer practicals (`ComputerExercisesRHints.pdf`) as well as a file with suggested R code, the resulting outcomes and explanatory text (`ComputerPracticalsAnswers.pdf`). It also contains a script file with all example code that is used in the book (`ExampleCode.R`) and gives information on upcoming courses.

Acknowledgements

The writing of this book has been a challenging and inspiring process, to which several persons have contributed. The book originates from the courses that I have been teaching together with Hein Putter. I borrowed several ideas on multi-state models from his slides and computer practicals. Johannes Mertsching provided some of the results in Chapter 2 while doing his internship. I thank the reviewers for their useful comments and suggestions, as well as my colleagues Marta Fiocco, Amy Matser and Jannie van der Helm for reading the first chapter of the book. Koos Zwinderman, you were right in saying that writing this book was going to be a weekend and evening job. Yet, I am grateful that you also gave me the opportunity to write parts of my book during office hours. I am convinced that it has been a profitable investment for the department as well. I also like to thank John Kimmel for allowing me to miss my deadline so many times, and Laurie Oknowsky and Michael Davidson for the editorial process.

I thank the Bikram yoga teachers for their tough classes, which filled me with new energy and creativity. Dear friends, you gave me mental support because you kept asking me about my progress. I hope I have been able to give you some idea of what the book is about. Yet, having finished the book does not mean that I will leave my laptop at home the next time. Jacquelien, I hope to open another bottle with white bubbles soon. Roberto, I have finally finished my "facebookie".

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

About the Author

Ronald Geskus received his Ph.D. in mathematical statistics at the Delft Technical University in 1997, based on a thesis entitled "Asymptotically efficient estimation with interval censored data" (supervisor Piet Groeneboom). Since 1995 he has been affiliated with the public health service of Amsterdam (PHS), where he performed and supervised studies on HIV and other sexually transmitted infections. There he specialized in the statistical analysis of data collected in cohort studies. In 2014 he was appointed associate professor at the Academic Medical Center (AMC) in Amsterdam. He has worked in several other medical and statistical research environments in the Netherlands and has spent sabbaticals in HIV/AIDS research groups in Paris and Madrid.

His research interests include: i) models for complex time-to-event data (competing risks, multi-state models), ii) models for complex longitudinal data, iii) prediction based on time-updated marker values, iv) causal inference. He contributed to and supervised many medical, epidemiological and statistical studies, performed both at the PHS and the AMC as well as within international collaborations.

He has published around 150 peer reviewed scientific articles, applied as well as methodological. He published methodological papers on i) the estimation of time from HIV infection to AIDS if the time of infection is unknown, ii) the development of markers in relation to disease progression, and iii) the analysis of competing risks with left truncated and right censored data. He is the joint first author of a highly cited tutorial on competing risks and multi-state models.

He has been teaching courses on the analysis of competing risks data in Brazil, Spain, Belgium, Sweden, Austria and the Netherlands.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

List of Figures

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

List of Tables