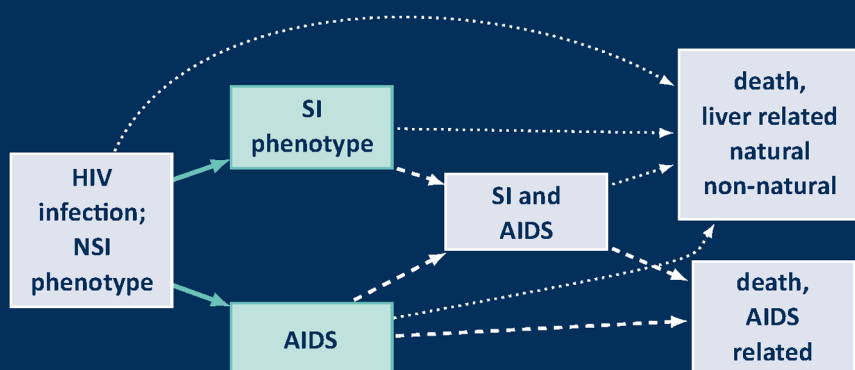


Chapman & Hall/CRC Biostatistics Series

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

Editor-in-Chief

Shein-Chung Chow, Ph.D., Professor, Department of Biostatistics and Bioinformatics,
Duke University School of Medicine, Durham, North Carolina

Series Editors

Byron Jones, Biometrical Fellow, Statistical Methodology, Integrated Information Sciences,
Novartis Pharma AG, Basel, Switzerland

Jen-pei Liu, Professor, Division of Biometry, Department of Agronomy,
National Taiwan University, Taipei, Taiwan

Karl E. Peace, Georgia Cancer Coalition, Distinguished Cancer Scholar, Senior Research Scientist
and Professor of Biostatistics, Jiann-Ping Hsu College of Public Health,
Georgia Southern University, Statesboro, Georgia

Bruce W. Turnbull, Professor, School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, New York

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. Bushmakina, 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

© 2016 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

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

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

To life



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

Preface	xiii
Acknowledgements	xvii
About the Author	xix
1 Basic Concepts	1
1.1 Introduction	1
1.2 Examples	2
1.2.1 Infection during a hospital stay	2
1.2.2 HIV infection	3
1.2.3 Bone marrow transplantation	6
1.3 Data structure	7
1.3.1 Time scales	7
1.3.2 Right censored data	8
1.3.3 Left truncated data	10
1.4 On rates and risks	13
1.5 Non-informative observation schemes?	15
1.5.1 Some possible solutions	20
1.6 The examples revisited	23
1.6.1 Infection during a hospital stay	23
1.6.2 HIV infection	24
1.6.3 Bone marrow transplantation	28
1.7 Notation	28
1.8 Basic techniques from survival analysis	30
1.8.1 Main concepts and theoretical relations	30
1.8.2 The Kaplan-Meier product-limit estimator	32
1.8.2.1 Confidence intervals	34
1.8.3 Nonparametric group comparisons	36
1.8.4 Cox proportional hazards model	37
1.8.5 Counting process format	40
1.9 Summary and preview	43
1.10 Exercises	44
1.11 R code for classical survival analysis	50
1.11.1 The <code>aidssi</code> data set	50

1.11.2	Define time and status information	51
1.11.3	Perform calculations	51
1.11.4	Summary of outcome	52
1.11.5	Log-rank test	55
1.12	Computer practicals	56
2	Competing Risks; Nonparametric Estimation	59
2.1	Introduction	59
2.2	Theoretical relations	60
2.2.1	The multi-state approach; cause-specific hazards . . .	60
2.2.2	The subdistribution approach	62
2.3	Estimation based on cause-specific hazard	63
2.4	Estimation: the subdistribution approach	68
2.4.1	Estimation with complete follow-up	70
2.4.2	A special choice for $\hat{\Gamma}$ and $\hat{\Phi}$	71
2.4.3	The ECDF and PL forms	73
2.4.3.1	The ECDF form	74
2.4.3.2	The PL form	74
2.4.4	Interpretation of the weighted estimators	76
2.5	Standard errors and confidence intervals	79
2.6	Log-rank tests and other subgroup comparisons	83
2.7	Summary; three principles of interpretability	85
2.8	Exercises	88
2.9	Software	92
2.9.1	Nonparametric estimation of F_k	92
2.9.1.1	The Aalen-Johansen form	92
2.9.1.2	The weighted product-limit form	96
2.9.2	Log-rank tests	100
2.10	Computer practicals	101
3	Intermediate Events; Nonparametric Estimation	105
3.1	Introduction; multi-state models	105
3.2	Main concepts and theoretical relations	107
3.2.1	Basic framework and definitions	107
3.3	Estimation	111
3.3.1	Data representation	111
3.3.2	Nelson-Aalen and Aalen-Johansen estimator	111
3.4	Example: HIV, SI, AIDS and death	115
3.4.1	The data	116
3.4.2	Analyses	117
3.5	Summary; some alternative approaches	125
3.6	Exercises	126
3.7	Software	127
3.7.1	The etm package	130
3.7.2	The msSurv package	134

3.7.3	The <code>mstate</code> package	138
3.8	Computer practicals	141
4	Regression; Cause-Specific/Transition Hazard	143
4.1	Introduction	143
4.2	Regression on cause-specific hazard: basic structure	144
4.3	Combined analysis and type-specific covariables	146
4.3.1	Same results in one analysis	147
4.3.2	Type-specific covariables	149
4.3.3	Effects equal over causes	151
4.3.4	Proportional baseline hazards	152
4.4	Why does the stacked approach work?	153
4.4.1	Cause as stratum variable	153
4.4.2	Effects equal over causes	155
4.4.3	Proportional baseline hazards	155
4.5	Multi-state regression models for transition hazards	156
4.5.1	Combined analyses: assume effects to be equal	159
4.5.2	Proportional baseline hazards	160
4.5.3	Dual role of intermediate states	164
4.5.4	Beyond the Markov model: effect of transition time	165
4.5.5	Standard error	167
4.6	Example: causes of death in HIV infected individuals	167
4.6.1	Analysis using well-defined contrasts	172
4.7	Summary	176
4.8	Exercises	177
4.9	Software	179
4.10	Computer practicals	180
5	Regression; Translation to Cumulative Scale	183
5.1	Introduction	183
5.2	From cause-specific/transition hazard to probability	184
5.2.1	Competing risks	184
5.2.2	Multi-state models	187
5.3	Regression on subdistribution hazard	188
5.3.1	Choice of weight function	191
5.3.2	Estimation of standard error	192
5.3.3	Time-varying covariables	192
5.3.4	Examples	195
5.4	Multinomial regression	199
5.5	Summary	200
5.6	Exercises	202
5.7	Software	202
5.7.1	From cause-specific/transition hazard to probability	202
5.7.2	Regression on subdistribution hazard	206
5.7.3	Proportional odds model	208

5.8	Computer practicals	209
5.8.1	Multi-state analysis	209
6	Epilogue	213
6.1	Which type of quantity to choose?	213
6.2	Exercises	217
	Bibliography	221
	Appendix: Answers to Exercises	231
	Index	245

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

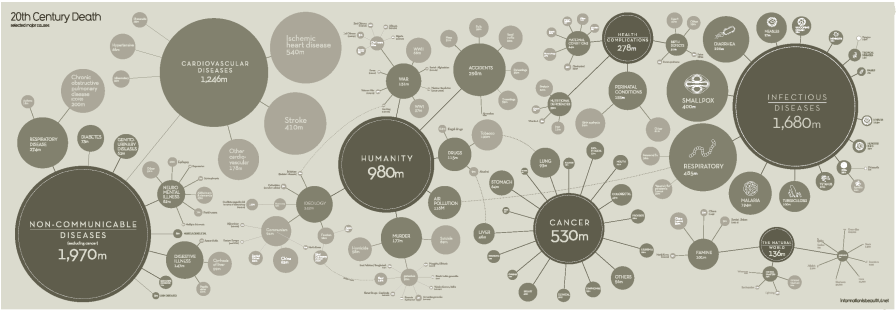


FIGURE 1
All causes of death in the 20th century.

subgroups may not be readable (have a look at the website if you want to read them¹), 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

¹The 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. There 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. Jacquélien, 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

1	All causes of death in the 20th century.	xiii
1.1	Schematic structure of the classical survival analysis setting. .	1
1.2	Example: staphylococcus infection during a hospital stay, with discharge without infection as competing event.	3
1.3	Example: progression from HIV infection to AIDS, with pre-AIDS mortality as competing event.	4
1.4	Example: spectrum in causes of death after HIV infection. . .	4
1.5	Example: progression from HIV infection to AIDS and death, with an SI switch as an intermediate event.	5
1.6	Example: events after bone marrow transplantation.	6
1.7	Events in the calendar time scale and the personal time scale.	8
1.8	Late entry in calendar time scale and personal time scale. . .	11
1.9	Graphical representation of left truncation via the event space. Grey triangle: unobserved period. Vertical line: individual from Figure 1.8; the dashed line becomes a solid line from his moment of entry onwards.	11
1.10	Graphical representation of non-informative censoring. Left panel: information beyond three months; individuals leaving the study at three months are in gray. Right panel: complete data until ten months.	17
1.11	Graphical explanation why we cannot decide from the observed data whether the censoring time and event time distributions are independent.	19
1.12	Graphical explanation why we cannot decide from the observed data whether the event time and entry time distributions are independent.	20
1.13	Typical situation of semi-competing risks: the event of interest, virological failure, does not terminate follow-up.	22
1.14	Graphical representation of semi-competing risks data.	22
1.15	Kaplan-Meier curves for progression to AIDS on the complementary scale “one-minus-survival”.	25
1.16	Estimates of cause-specific cumulative incidence for AIDS and pre-AIDS death.	26
1.17	Follow-up data for ten example individuals. The vertical dashed lines define the three calendar periods of follow-up.	42
1.18	Mortality in Dutch population.	45

1.19	Graphical representation of individual follow-up of 171 MSM from the ACS.	47
1.20	Relation between pregnancy rate and dropout rate per cycle.	48
1.21	Three possible situations with respect to relative position of late entry, AIDS and death.	49
1.22	Estimate of cumulative incidence function for combined end point “AIDS/SI”.	53
2.1	The probability to progress to event k at time s is a product of the probabilities $P(T \geq s)$ and $\lambda_k(s)$	61
2.2	Example of calculations with competing risks: graphical representation of the complete information. Information that is not observed is shown via dotted lines (late entry) and dashed lines (right censoring).	65
2.3	Example of calculations with competing risks: graphical representation of the observed data. Lines not ending with a closed circle correspond to right censored observations.	66
2.4	Estimated cause-specific cumulative incidence function for AIDS and SI as first events, alternate display format (solid black lines). The “marginal” Kaplan-Meier curves are shown in gray.	67
2.5	Estimate of cause-specific cumulative incidence for AIDS and SI as first events, in a stacked display format.	67
2.6	Example of calculations with competing risks: graphical representation of the data in which censorings are treated as events and events as censorings.	72
2.7	Example of calculations with competing risks: graphical representation of the data with entry treated as event and events and censorings inducing right truncation.	73
2.8	Comparison of different scales for calculation of 95% confidence intervals around estimate of SI-specific cumulative incidence. Dotted: log; solid: log-log; dashed: linear. Black: based on $\widehat{F}_k$; gray: based on $\widehat{F}_k^{\text{PL}}$	80
2.9	Estimate of SI-specific cumulative incidence and 95% confidence intervals on the log scale (1.20). Gaynor/Betensky: thick grey line; PL form: thin solid black line.	81
2.10	Comparison of variance estimators based on the Aalen-Johansen and the product-limit form. Method: $\widehat{F}_k^{\text{AJ}} = +$; $\widehat{F}_k^{\text{PL}} = \circ$	82
2.11	Estimated cause-specific (left panel) and subdistribution (right panel) cumulative SI-specific hazard. Solid lines: wild type CCR5. Dashed lines: CCR5- Δ 32 deletion.	85
2.12	Estimate of CE-specific cumulative incidence (black solid line), Kaplan-Meier (dashed line) and DOC-specific cumulative incidence (gray line).	90

2.13	Estimate of cause-specific cumulative incidence for relapse (left panel) and death (right panel) for females and males. Solid lines: males. Dashed lines: females.	91
2.14	Estimate of relapse-specific cumulative incidence (black lines) and Kaplan-Meier (grey lines) for males (solid lines) and females (dashed lines).	91
3.1	Graphical representation of illness-death model without recovery.	107
3.2	Graphical representation of acyclic illness-death model with disease recovery.	108
3.3	Directed graph of disease course from HIV infection to death, with SI and AIDS as transient states.	115
3.4	Nelson-Aalen estimates of the cumulative transition hazard for all six direct transitions.	118
3.5	Aalen-Johansen estimates of the transition probability for all six transitions with a direct path.	120
3.6	Aalen-Johansen estimates of the transition probability for all six transitions with a direct path, from four years onwards.	120
3.7	Aalen-Johansen estimates of all state occupation probabilities.	121
3.8	Probability to be alive with AIDS (thick solid curve), with 95% CI (dotted lines). Thin solid curve is estimated probability to be in state 1, AIDS.	122
3.9	Distribution of time from HIV infection to death. Grey curve and area: Kaplan-Meier estimator, with 95% CI. Black lines: Aalen-Johansen estimator based on a multi-state model, with 95% CI.	122
3.10	Fixed horizon prediction at ten years after HIV infection, starting in states HIV and SI.	123
3.11	State-AIDS entry distribution (black) and the probability of being alive with AIDS (grey).	124
3.12	Estimate of probability to die over time for an individual in state SI at time $s_k = 2, 3, 4, 5, 6, 7, 8$ years.	128
4.1	Schematic representation of the effect of the CCR5- Δ 32 deletion on each of the competing events.	145
4.2	Estimated baseline hazards for all six transitions.	161
4.3	Illness-death model for the effect of the SI switch on the pathway to AIDS.	164
4.4	Cause-specific cumulative mortality by HCV status and calendar period of follow-up. Dashed line: HIV mono-infected individuals; solid line: HCV/HIV coinfecting individuals. 95% confidence intervals in grey.	169

4.5	Person-years of follow-up (before colon) and number of deaths (after colon) for each combination of HCV status and calendar period, overall and by values of risk group and gender. Rectangles with the same pattern denote combinations for which covariable effects were assumed equal in the regression model for overall survival [104].	170
4.6	Number of cause-specific deaths within subgroups by combination of HCV and calendar period. Patterns in rectangles have same meaning as in Figure 4.5. L: liver related; A: AIDS related; N: natural; NN: non-natural.	171
4.7	Covariable combinations chosen in the regression model for the four CODs. L: liver related; A: AIDS related; N: natural; NN: non-natural. P-values: $0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < < 1$.	173
5.1	Cause-specific cumulative incidence for AIDS (on survival scale) and SI (on event scale). Grey: nonparametric estimate; black: estimate based on proportional cause-specific hazards model.	186
5.2	Explanation of the difference between effects on cause-specific hazard and on cause-specific cumulative incidence.	187
5.3	Transition probability from HIV (lower thick lines) and SI (upper thin lines) to death for WW individuals (solid lines) and WM individuals (dashed lines), aged 25 years at HIV infection. Error bars: 95% CI at specific time points.	189
5.4	Transition probability from HIV (lower thick lines) and SI (upper thin lines) to death for individuals aged 27 years (solid lines) and 52 years (dashed lines) that have the wild type for CCR5. Error bars: 95% CI at specific time points.	189
5.5	Cause-specific cumulative incidence for AIDS (on survival scale) and SI (on event scale). Grey: nonparametric estimate; black: estimate based on proportional subdistribution hazards model.	197
5.6	Effect estimates and 95% confidence intervals based on cause-specific hazards model (upper line) and subdistribution hazards model (lower line).	198
5.7	Cause-specific cumulative incidence for AIDS (on survival scale) and SI (on event scale). Grey: nonparametric estimate; black: estimate based on proportional odds model.	200
6.1	Cumulative AIDS-specific hazard for IDU (dashed line) and MSM (solid line).	216
6.2	Estimate of cumulative incidence of combined end point death and AIDS.	218

6.3 Cause-specific cumulative incidence (with 95% CI) by exposure group. Solid lines: early starters; dashed lines: late starters. 219

A.1 Kaplan-Meier plots. A: grey solid line; B: black solid thin line; C: grey dashed line; D: black dashed line; E:black solid thick line. 232

A.2 Multi-state model; SI, AIDS and death are possible states and the sequence in which SI and AIDS occur can influence mortality. 237



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

--- *List of Tables* ---

1.1	Hazard types and corresponding cumulative measures in the presence of multiple end points	15
1.2	Observed causes of death before AIDS diagnosis	25
1.3	Four example individuals, aidssi data set	51
2.1	Detailed calculation of $\widehat{F}_A^{AJ}(2)$ and $\widehat{F}_A^{AJ}(6)$	66
2.2	Competing risks data with complete follow-up during the first seven weeks	70
2.3	Detailed calculation of $\widehat{F}_A^{PL}(2)$ and $\widehat{F}_A^{PL}(6)$	75
2.4	P-values log-rank tests for effect of CCR5 genotype	84
2.5	Estimators of hazard and cumulative incidence	86
2.6	Three example individuals	96
2.7	Same three individuals with time-varying weights	97
2.8	The weighted aidssi data set as created by the crprep function	99
3.1	Time and status information from four individuals in the aidssi2 data set	116
3.2	Time and status information in transition-based long format for the individuals from Table 3.1	117
3.3	Number of observed transitions (from row to column), number of censored observations and total number at risk	118
3.4	Event time and number at risk at the first four transitions, stratified by transition type	119
3.5	Nonparametric estimation in the R packages mvna , etm , msSurv and mstate	129
4.1	Four example individuals from aidssi data set	145
4.2	Four example individuals in the stacked long format	147
4.3	Four example individuals in the stacked long format with compound and type-specific variables	149
4.4	Hazard for each combination of variable and event type	151
4.5	Time and status information in stacked long format for four individuals from aidssi2 data set	157
4.6	Effects of CCR5- Δ 32 deletion and age at infection (per 10 years)	158