

# Statistics for Biology and Health

*Series Editors*

K. Dietz, M. Gail, K. Krickeberg, J. Samet, A. Tsiatis

**Springer**

*New York*

*Berlin*

*Heidelberg*

*Hong Kong*

*London*

*Milan*

*Paris*

*Tokyo*

# **SURVIVAL ANALYSIS**

## **Techniques for Censored and Truncated Data**

Second Edition

John P. Klein

*Medical College of Wisconsin*

Melvin L. Moeschberger

*The Ohio State University Medical Center*

With 97 Illustrations



Springer

John P. Klein  
Division of Biostatistics  
Medical College of Wisconsin  
Milwaukee, WI 53226  
USA

Melvin L. Moeschberger  
School of Public Health  
Division of Epidemiology and Biometrics  
The Ohio State University Medical Center  
Columbus, OH 43210  
USA

*Series Editors*

K. Dietz  
Institut für Medizinische Biometrie  
Universität Tübingen  
Westbahnhofstrasse 55  
D-72070 Tübingen  
Germany

M. Gail  
National Cancer Institute  
Rockville, MD 20892  
USA

K. Krickeberg  
Le Chatelet  
F-63270 Manglieu  
France  
  
A. Tsiatis  
Department of Statistics  
North Carolina State University  
Raleigh, NC 27695  
USA

J. Samet  
School of Public Health  
Department of Epidemiology  
Johns Hopkins University  
615 Wolfe St.  
Baltimore, MD 21205-2103  
USA

Library of Congress Cataloging-in-Publication Data  
Klein, John P., 1950–

Survival analysis : techniques for censored and truncated data / John P. Klein, Melvin L. Moeschberger. — 2nd ed.

p. cm. — (Statistics for biology and health)

Includes bibliographical references and index.

ISBN 0-387-95399-X (alk. paper)

1. Survival analysis (Biometry) I. Moeschberger, Melvin L.

II. Title. III. Series.

R853.S7 K535 2003

610'.7'27—dc21

2002026667

ISBN 0-387-95399-X

Printed on acid-free paper.

© 2003, 1997 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not especially identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

SPIN 10858633

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg

A member of BertelsmannSpringer Science+Business Media GmbH

---

# Preface

The second edition contains some new material as well as solutions to the odd-numbered revised exercises. New material consists of a discussion of summary statistics for competing risks probabilities in Chapter 2 and the estimation process for these probabilities in Chapter 4. A new section on tests of the equality of survival curves at a fixed point in time is added in Chapter 7. In Chapter 8 an expanded discussion is presented on how to code covariates and a new section on discretizing a continuous covariate is added. A new section on Lin and Ying's additive hazards regression model is presented in Chapter 10. We now proceed to a general discussion of the usefulness of this book incorporating the new material with that of the first edition.

A problem frequently faced by applied statisticians is the analysis of time to event data. Examples of such data arise in diverse fields such as medicine, biology, public health, epidemiology, engineering, economics and demography. While the statistical tools we shall present are applicable to all these disciplines our focus is on applications of the techniques to biology and medicine. Here interest is, for example, on analyzing data on the time to death from a certain cause, duration of response to treatment, time to recurrence of a disease, time to development of a disease, or simply time to death.

The analysis of survival experiments is complicated by issues of censoring, where an individual's life length is known to occur only in a certain period of time, and by truncation, where individuals enter the study only if they survive a sufficient length of time or individuals are

included in the study only if the event has occurred by a given date. The use of counting process methodology has, in recent years, allowed for substantial advances in the statistical theory to account for censoring and truncation in survival experiments. The book by Andersen et al. (1993) provides an excellent survey of the mathematics of this theory. In this book we shall attempt to make these complex methods more accessible to applied researchers without an advanced mathematical background by presenting the essence of the statistical methods and illustrating these results in an applied framework. Our emphasis is on applying these techniques, as well as classical techniques not based on the counting process theory, to data rather than on the theoretical development of these tools. Practical suggestions for implementing the various methods are set off in a series of practical notes at the end of each section. Technical details of the derivation of these techniques (which are helpful to the understanding of concepts, though not essential to using the methods of this book) are sketched in a series of theoretical notes at the end of each section or are separated into their own sections. Some more advanced topics, for which some additional mathematical sophistication is needed for their understanding or for which standard software is not available, are given in separate chapters or sections. These notes and advanced topics can be skipped without a loss of continuity.

We envision two complementary uses for this book. The first is as a reference book for investigators who find the need to analyze censored or truncated life time data. The second use is as a textbook for a graduate level course in survival analysis. The minimum prerequisite for such course is a traditional course in statistical methodology. The material included in this book comes from our experience in teaching such a course for master's level biostatistics students at The Ohio State University and at the Medical College of Wisconsin, as well as from our experience in consulting with investigators from The Ohio State University, The University of Missouri, The Medical College of Wisconsin, The Oak Ridge National Laboratory, The National Center for Toxicological Research, and The International Bone Marrow Transplant Registry.

The book is divided into thirteen chapters that can be grouped into five major themes. The first theme introduces the reader to basic concepts and terminology. It consists of the first three chapters which deal with examples of typical data sets one may encounter in biomedical applications of this methodology, a discussion of the basic parameters to which inference is to be made, and a detailed discussion of censoring and truncation. New to the second edition is Section 2.7 that presents a discussion of summary statistics for competing risks probabilities. Section 3.6 gives a brief introduction to counting processes, and is included for those individuals with a minimal background in this area who wish to have a conceptual understanding of this methodology. This section can be omitted without jeopardizing the reader's understanding of later sections of the book.

The second major theme is the estimation of summary survival statistics based on censored and/or truncated data. Chapter 4 discusses estimation of the survival function, the cumulative hazard rate, and measures of centrality such as the median and the mean. The construction of pointwise confidence intervals and confidence bands is presented. Here we focus on right censored as well as left truncated survival data since this type of data is most frequently encountered in applications. New to the second edition is a section dealing with estimation of competing risks probabilities. In Chapter 5 the estimation schemes are extended to other types of survival data. Here methods for double and interval censoring; right truncation; and grouped data are presented. Chapter 6 presents some additional selected topics in univariate estimation, including the construction of smoothed estimators of the hazard function, methods for adjusting survival estimates for a known standard mortality and Bayesian survival methods.

The third theme is hypothesis testing. Chapter 7 presents one-, two-, and more than two-sample tests based on comparing the integrated difference between the observed and expected hazard rate. These tests include the log rank test and the generalized Wilcoxon test. Tests for trend and stratified tests are also discussed. Also discussed are Renyi tests which are based on sequential evaluation of these test statistics and have greater power to detect crossing hazard rates. This chapter also presents some other censored data analogs of classical tests such as the Cramer–Von Mises test, the  $t$  test and median tests are presented. New to this second edition is a section on tests of the equality of survival curves at a fixed point in time.

The fourth theme, and perhaps the one most applicable to applied work, is regression analysis for censored and/or truncated data. Chapter 8 presents a detailed discussion of the proportional hazards model used most commonly in medical applications. New sections in this second edition include an expanded discussion of how to code covariates and a section on discretizing a continuous covariate. Recent advances in the methodology that allows for this model to be applied to left truncated data, provides the investigator with new regression diagnostics, suggests improved point and interval estimates of the predicted survival function, and makes more accessible techniques for handling time-dependent covariates (including tests of the proportionality assumption) and the synthesis of intermediate events in an analysis are discussed in Chapter 9.

Chapter 10 presents recent work on the nonparametric additive hazard regression model of Aalen (1989) and a new section on Lin and Ying's (1994) additive hazards regression models. One of these models may be the model of choice in situations where the proportional hazards model or a suitable modification of it is not applicable. Chapter 11 discusses a variety of residual plots one can make to check the fit of the Cox proportional hazards regression models. Chapter 12 discusses parametric models for the regression problem. Models presented in-

clude those available in most standard computer packages. Techniques for assessing the fit of these parametric models are also discussed.

The final theme is multivariate models for survival data. In Chapter 13, tests for association between event times, adjusted for covariates, are given. An introduction to estimation in a frailty or random effect model is presented. An alternative approach to adjusting for association between some individuals based on an analysis of an independent working model is also discussed.

There should be ample material in this book for a one or two semester course for graduate students. A basic one semester or one quarter course would cover the following sections:

Chapter 2  
Chapter 3, Sections 1–5  
Chapter 4  
Chapter 7, Sections 1–6, 8  
Chapter 8  
Chapter 9, Sections 1–4  
Chapter 11  
Chapter 12

In such a course the outlines of theoretical development of the techniques, in the theoretical notes, would be omitted. Depending on the length of the course and the interest of the instructor, these details could be added if the material in section 3.6 were covered or additional topics from the remaining chapters could be added to this skeleton outline. Applied exercises are provided at the end of the chapters. Solutions to odd numbered exercises are new to the second edition. The data used in the examples and in most of the exercises is available from us at our Web site which is accessible through the Springer Web site at <http://www.springer-ny.com> or <http://www.biostat.mcw.edu/homepgs/klein/book.html>.

Milwaukee, Wisconsin  
Columbus, Ohio

John P. Klein  
Melvin L. Moeschberger

---

# Contents

Preface	v
Chapter 1 — Examples of Survival Data	1
1.1 Introduction	1
1.2 Remission Duration from a Clinical Trial for Acute Leukemia	2
1.3 Bone Marrow Transplantation for Leukemia	3
1.4 Times to Infection of Kidney Dialysis Patients	6
1.5 Times to Death for a Breast-Cancer Trial	7
1.6 Times to Infection for Burn Patients	8
1.7 Death Times of Kidney Transplant Patients	8
1.8 Death Times of Male Laryngeal Cancer Patients	9
1.9 Autologous and Allogeneic Bone Marrow Transplants	10
1.10 Bone Marrow Transplants for Hodgkin's and Non-Hodgkin's Lymphoma	11
1.11 Times to Death for Patients with Cancer of the Tongue	12
	<b>ix</b>



1.12	Times to Reinfection for Patients with Sexually Transmitted Diseases	13
1.13	Time to Hospitalized Pneumonia in Young Children	14
1.14	Times to Weaning of Breast-Fed Newborns	14
1.15	Death Times of Psychiatric Patients	15
1.16	Death Times of Elderly Residents of a Retirement Community	16
1.17	Time to First Use of Marijuana	17
1.18	Time to Cosmetic Deterioration of Breast Cancer Patients	18
1.19	Time to AIDS	19

## Chapter 2 — Basic Quantities and Models 21

2.1	Introduction	21
2.2	The Survival Function	22
2.3	The Hazard Function	27
2.4	The Mean Residual Life Function and Median Life	32
2.5	Common Parametric Models for Survival Data	36
2.6	Regression Models for Survival Data	45
2.7	Models for Competing Risks	50
2.8	Exercises	57

## Chapter 3 — Censoring and Truncation 63

3.1	Introduction	63
3.2	Right Censoring	64
3.3	Left or Interval Censoring	70
3.4	Truncation	72
3.5	Likelihood Construction for Censored and Truncated Data	74
3.6	Counting Processes	79
3.7	Exercises	87

---

Chapter 4 — Nonparametric Estimation of Basic Quantities for Right-Censored and Left-Truncated Data	91
4.1 Introduction	91
4.2 Estimators of the Survival and Cumulative Hazard Functions for Right-Censored Data	92
4.3 Pointwise Confidence Intervals for the Survival Function	104
4.4 Confidence Bands for the Survival Function	109
4.5 Point and Interval Estimates of the Mean and Median Survival Time	117
4.6 Estimators of the Survival Function for Left-Truncated and Right-Censored Data	123
4.7 Summary Curves for Competing Risks	127
4.8 Exercises	133
Chapter 5 — Estimation of Basic Quantities for Other Sampling Schemes	139
5.1 Introduction	139
5.2 Estimation of the Survival Function for Left, Double, and Interval Censoring	140
5.3 Estimation of the Survival Function for Right-Truncated Data	149
5.4 Estimation of Survival in the Cohort Life Table	152
5.5 Exercises	158
Chapter 6 — Topics in Univariate Estimation	165
6.1 Introduction	165
6.2 Estimating the Hazard Function	166
6.3 Estimation of Excess Mortality	177
6.4 Bayesian Nonparametric Methods	187
6.5 Exercises	198

Chapter 7 — Hypothesis Testing	201
7.1 Introduction	201
7.2 One-Sample Tests	202
7.3 Tests for Two or More Samples	205
7.4 Tests for Trend	216
7.5 Stratified Tests	219
7.6 Renyi Type Tests	223
7.7 Other Two-Sample Tests	227
7.8 Test Based on Differences in Outcome at a Fixed Point in Time	234
7.9 Exercises	238
Chapter 8 — Semiparametric Proportional Hazards Regression with Fixed Covariates	243
8.1 Introduction	243
8.2 Coding Covariates	246
8.3 Partial Likelihoods for Distinct-Event Time Data	253
8.4 Partial Likelihoods When Ties Are Present	259
8.5 Local Tests	263
8.6 Discretizing a Continuous Covariate	272
8.7 Model Building Using the Proportional Hazards Model	276
8.8 Estimation of the Survival Function	283
8.9 Exercises	287
Chapter 9 — Refinements of the Semiparametric Proportional Hazards Model	295
9.1 Introduction	295
9.2 Time-Dependent Covariates	297
9.3 Stratified Proportional Hazards Models	308

9.4	Left Truncation	312
9.5	Synthesis of Time-varying Effects (Multistate Modeling)	314
9.6	Exercises	326
Chapter 10 — Additive Hazards Regression Models		329
10.1	Introduction	329
10.2	Aalen's Nonparametric, Additive Hazard Model	330
10.3	Lin and Ying's Additive Hazards Model	346
10.4	Exercises	351
Chapter 11 — Regression Diagnostics		353
11.1	Introduction	353
11.2	Cox–Snell Residuals for Assessing the Fit of a Cox Model	354
11.3	Determining the Functional Form of a Covariate: Martingale Residuals	359
11.4	Graphical Checks of the Proportional Hazards Assumption	363
11.5	Deviance Residuals	381
11.6	Checking the Influence of Individual Observations	385
11.7	Exercises	391
Chapter 12 — Inference for Parametric Regression Models		393
12.1	Introduction	393
12.2	Weibull Distribution	395
12.3	Log Logistic Distribution	401
12.4	Other Parametric Models	405
12.5	Diagnostic Methods for Parametric Models	409
12.6	Exercises	419
Chapter 13 — Multivariate Survival Analysis		425
13.1	Introduction	425
13.2	Score Test for Association	427

13.3	Estimation for the Gamma Frailty Model	430
13.4	Marginal Model for Multivariate Survival	436
13.5	Exercises	438
Appendix A — Numerical Techniques for Maximization		443
A.1	Univariate Methods	443
A.2	Multivariate Methods	445
Appendix B — Large-Sample Tests Based on Likelihood Theory		449
Appendix C — Statistical Tables		455
C.1	Standard Normal Survival Function $P[Z \geq z]$	456
C.2	Upper Percentiles of a Chi-Square Distribution	457
C.3a	Confidence Coefficients $c_{10}(a_L, a_U)$ for 90% EP Confidence Bands	459
C.3b	Confidence Coefficients $c_{05}(a_L, a_U)$ for 95% EP Confidence Bands	463
C.3c	Confidence Coefficients $c_{01}(a_L, a_U)$ for 99% EP Confidence Bands	465
C.4a	Confidence Coefficients $k_{10}(a_L, a_U)$ for 90% Hall–Wellner Confidence Bands	468
C.4b	Confidence Coefficients $k_{05}(a_L, a_U)$ for 95% Hall–Wellner Confidence Bands	471
C.4c	Confidence Coefficients $k_{01}(a_L, a_U)$ for 99% Hall–Wellner Confidence Bands	474
C.5	Survival Function of the Supremum of the Absolute Value of a Standard Brownian Motion Process over the Range 0 to 1	477
C.6	Survival Function of $W = \int_0^\infty [B(t)]^2 dt$ , where $B(t)$ is a Standard Brownian Motion	478
C.7	Upper Percentiles of $R = \int_0^k  B^o(u)  du$ , where $B^o(u)$ is a Brownian Bridge	479

Appendix D — Data on 137 Bone Marrow Transplant Patients	483
Appendix E — Selected Solutions to Exercises	489
Bibliography	515
Author Index	527
Subject Index	531