PREFACE TO

SURVIVAL ANALYSIS: TECHNIQUES FOR CENSORED AND TRUNCATED DATA

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 text book 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. 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. 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. A final section of this chapter presents some other censored data analogs of classical tests such as the Cramer -Von Mises test, the t test and median tests are presented.

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. 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 additive hazard regression model, which 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 these two semi-parametric regression models. Chapter 12 discusses parametric models for the regression problem. Models presented include 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 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 Chapter 8 Chapter 9 Sections 1-4 Chapter 11 Sections 1-6 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. The data used in the examples and in most of the exercises is available at this Web site.