



4-1992

A Progressive Disease Model for Doubly-Censored Bivariate Survival Data that Accommodates Covariate Information

Hilmi F. Yahya
Western Michigan University

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Statistics and Probability Commons

Recommended Citation

Yahya, Hilmi F., "A Progressive Disease Model for Doubly-Censored Bivariate Survival Data that Accommodates Covariate Information" (1992). *Dissertations*. 1986.

<https://scholarworks.wmich.edu/dissertations/1986>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



A PROGRESSIVE DISEASE MODEL FOR DOUBLY-CENSURED
BIVARIATE SURVIVAL DATA THAT ACCOMMODATES
COVARIATE INFORMATION

by

Hilmi F. Yahya

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Mathematics and Statistics

Western Michigan University
Kalamazoo, Michigan
April 1992

A PROGRESSIVE DISEASE MODEL FOR DOUBLY-CENSURED
BIVARIATE SURVIVAL DATA THAT ACCOMMODATES
COVARIATE INFORMATION

Hilmi F. Yahya, Ph.D.

Western Michigan University, 1992

A model for the natural history of a progressive disease is developed. The model has three disease states and can be expressed as the joint distribution of two survival random variables.

Covariate information is incorporated into the model using the proportional hazards model for the marginal distributions. The model will also accommodate data with observations which are censored on one or both of the survival random variables.

The likelihood function for censored data is exhibited for finding the maximum likelihood estimates of the parameters and their standard errors for testing the effects of the covariates. The method used to obtain these estimates is the maximum likelihood method. Typical epidemiological measures are written in terms of the parameters of this model. Potential application of this model to cancer and heart disease research is discussed.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9224026

**A progressive disease model for doubly-censored bivariate
survival data that accommodates covariate information**

Yahya, Hilmi Fadel, Ph.D.

Western Michigan University, 1992

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

ACKNOWLEDGEMENTS

I would like to thank my committee for their advice and guidance during this work and for serving on my committee, especially Professor Gerald L. Sievers, Professor Michael Stoline, Professor Dennis Pence, and Professor Thomas Vidmar. I would like to thank Professor Joseph McKean for his help and support during my study. I would like to thank David Zietler and Jorge Hidalgo for their help. I would like to thank my friends for their support during the years. I would like to thank my family for their support and understanding during the years, especially my wife, Farideh, without whose support, patience, and understanding this work could not have happened.

Hilmi Yahya

PLEASE NOTE

**Duplicate page number(s); text follows.
Filmed as received.**

iii

University Microfilms International

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER	
I. INTRODUCTION	1
II. DEFINITIONS	8
Louis et al. Model	8
Standard Epidemiologic Measures of a Disease State	12
Further Definitions	17
III. LIKELIHOOD FUNCTION FOR THE UNCENSORED CASE	21
Assumptions	21
Likelihood Function	24
IV. LIKELIHOOD FUNCTION FOR THE CENSORED CASE	30
Observed Data for the Model	30
Likelihood Function for the Censored Case	32
V. LOUIS ET AL. EPIDEMIOLOGICAL MEASURES IN TERMS OF OUR MODEL	39
The Epidemiological Measures in Terms of Our Model	39
The Asymptotic Distributions of the Standard Epidemiological Measures	44
VI. EXAMPLE	49

Table of Contents--Continued

Description of the Data	49
Estimation of the Parameters	51
Hypothesis Testing	57
Analysis and Discussion	58
BIBLIOGRAPHY	69

LIST OF TABLES

1.	Likelihood Contributions for Observed Data	38
2.	Key Relationships	39
3.	Variables Used in the Model for the Application of the FHSD	50
4.	Overall Preclinical Incidence and Its Standard Errors for Men and Women for the Two Models	61
5.	95% C.I. for the Overall Preclinical Incidence for Men and Women for the Two Models	65
6.	MLE Estimates of Baseline Hazards Associated With CHD Morbidity for the Two Models	65
7.	MLE Estimates of the Baseline Hazards Associated With CHD Mortality for the Two Mode	66
8.	Estimated Preclinical Incidence When Fitting Our Model With Covariate (SEX)	66
9.	Estimated Preclinical Incidence When Fitting Our Model (No Covariates) for Men Only	67
10.	Estimated Preclinical Incidence When Fitting Our Model (No Covariates) for Women Only	67

LIST OF FIGURES

1.	Progressive Disease State Model.....	9
2.	Age Changes With Time But X and Y Do Not	11
3.	Dividing the Positive XY Plane Into Rectangles.....	23

CHAPTER I

INTRODUCTION

The fundamental increase of research and activity in the statistical analysis of survival data over the past few decades was largely stimulated by the many problems arising in the analysis of clinical trials. The interest in this area has resulted in a large volume of writing on this topic.

Particular attention and advancement have been devoted to developing parametric and nonparametric methods for the univariate case for the analysis of failure times, which permit the comparison of survival curves in the presence of censoring. Kaplan and Meier (1958) considered nonparametric estimation methods for estimating a one-dimensional survival function in the presence of censored data. These estimates are the reduced-sample estimator (all censoring variables are observable) and the product-limit estimator. Breslow and Crowley (1974) supplied details for the asymptotic behavior of both estimates of Kaplan and Meier. Miller (1976) utilized the product-limit estimator to discuss linear regression with censored data. Cox (1972) was concerned with regression models and life

tables; he presented a life table regression model considered by Kaplan and Meier (1958) to incorporate covariate information into the product-limit life table. Cox (1972) modelled the hazard function as a function of covariates and unknown regression coefficients multiplied by an arbitrary unknown function of time. Chiang, Hardy, Hawkins, and Kapadia (1989) and Holford (1976) also considered a parametric approach for a life table regression model for the analysis of univariate survival data with covariates. Pettersen (1986) considered the estimation of a fully parametric failure time model for the analysis of univariate survival data with time dependent covariates by means of the maximum likelihood method.

Particular attention and advancement have been given to developing nonparametric methods for estimating bivariate survival functions. For example, Campbell (1981) considered the estimation of a bivariate survival function in the presence of censored data using a nonparametric approach. He developed two estimators: a reduced-sample estimator and a self-consistent estimator, focusing on the nonparametric approach of Kaplan and Meier (1958) and the related self-consistency technique of Efron (1967).

Albert, Gertman, and Louis (1978) presented a basic probabilistic formulation for the natural history of a progressive disease in a population through the use of a "disease state" model with the following states: disease-

free state, preclinical state, and clinical state. They presented statistical methods for estimating components of the disease natural history described by this model. The model and methodology of Albert, Gertman, and Louis (1978) are contained in a series of three papers; the other two are: Albert, Gertman, Louis, and Liu (1978) and Louis, Albert, and Heghinian (1978). Since Louis is an author in all three papers, we will refer to this model as the Louis et al. model.

Louis et al. (1978) characterized the natural history of a progressive disease in terms of the joint distribution of a person's age when he enters the preclinical state, the sojourn time in that disease state, and the person's present age. Information to be used in estimation is generated by an ongoing screening program. These data are used to provide the maximum likelihood estimates of the joint distribution of holding times in the various states.

Classical descriptors of the natural history of the disease state model can be computed from this joint distribution, such as incidence, prevalence, lifetime attack rate, and expected preclinical duration of the disease.

Louis et al. (1978) proposed the use of maximum likelihood estimation to estimate the bivariate survival function of two failure times by partitioning the positive quadrant of the plane into rectangles. The probabilities

of the two failure times falling into each of these rectangles are estimated using likelihood methods. They derived the likelihood function for both approaches (parametric and nonparametric), and adopted the nonparametric approach for a numerical example. They stopped short of specifying the form of the joint distribution for the parametric approach.

Little attention has been given to developing parametric methods which permit the estimation of the joint (multivariate) survival distribution function. Bemis, Bain, and Higgins (1972), Freund (1961), Gumbel (1960), Marshall and Olkin (1967), Hawkes (1972), and O'Neill (1985) considered a bivariate extension of the exponential distribution for the bivariate survival function. Hougaard (1986, 1987) suggested a bivariate Weibull distribution for the bivariate survival function. Clayton (1978, 1982), Clayton and Cuzick (1985b), and Oakes (1982) considered a bivariate survival distribution which is characterized by two marginal distributions and an association parameter. Clayton and Cuzick (1985b) considered methods for estimating and testing only the association parameter for right censored data. Oakes (1982) described a special case of the distribution of Clayton and Cuzick in which the marginals are exponential. Oakes discussed in detail maximum likelihood estimation for the uncensored case. He also considered estimation and testing of the association

parameter. Farlie (1960) considered another bivariate distribution function which is characterized by two marginal distributions and an association parameter. All of these models model the dependence between the two variables. However, the Clayton and Cuzick formula describes the positive dependence, while the Farlie distribution describes both positive and negative dependence.

A fundamental tool for the parametric approach is the proportional hazards model which was first applied in the two sample problem by Mantel (1966) and later was clearly formulated by Cox (1972). The proportional hazards model is of central importance in epidemiology (for more details, see Chapters II and VI).

In this paper we extend the Louis et al. (1978) work to include the estimation of covariate effects, and then test the effects of the various covariates on the time to entry to the preclinical state and the sojourn time in the preclinical state.

Here we will extend the Louis et al. (1978) model by specifying the joint distribution for the time to entry to the disease state and the sojourn time in the disease state. We will use the joint distribution introduced by Clayton and Cuzick (1985b). This joint distribution was introduced to model association in bivariate survival distributions and is characterized by an association

parameter and the two marginal hazard functions. We model each of the marginal hazard functions by Cox's (1972) proportional hazards model in order to include covariate information. The bivariate survival function of Clayton and Cuzick will help us in writing and estimating some standard epidemiologic descriptors of the natural history of the disease model in terms of covariate effects.

We will use maximum likelihood methods to estimate the parameters of this joint distribution. To make the estimation process easier, we use a method developed by Holford (1976) and Chiang et al. (1989) which divides the natural history time line "follow-up period" into intervals and then assumes constant baseline hazards over each interval for both marginal distributions. This assumption also implies that survival times within intervals are exponentially distributed. This assumption was first used by Holford (1976), who allowed the hazard functions to be functions of covariates and subdivided the period of follow up into fixed intervals. This assumption was motivated by life tables, where observations are taken at fixed points (beginnings of intervals) in order to analyze the data. Chiang et al. (1989) believe that this is a reasonable assumption and take this as an approach to the problem.

In Chapter II, we define some basic terms needed for this work. We introduce the Louis et al. model and define some basic epidemiological measures in terms of the Louis

et al. model. The bivariate survival function of Clayton and Cuzick (1985b) is also introduced along with the proportional hazards model.

In Chapter III, we make the necessary assumptions for our model in order to develop the likelihood function for the uncensored case.

In Chapter IV, we develop the likelihood function for the censored case.

In Chapter V, we write the Louis et al. epidemiologic measures in terms of our model. We also give the formulas for the asymptotic distribution of the preclinical incidence and the overall preclinical incidence.

In Chapter VI, we consider an example. The data used are from the Framingham Heart Study Data taken from Kahn and Sempos (1989). Our model is used to analyze a portion of the data. We employ an iterative method to obtain the maximum likelihood estimates of the parameters with their standard errors (covariance matrix) which are needed for making inferences about some of the parameters. We also calculate estimates of some of the classical descriptors of the natural history of the progressive disease obtained in Chapter V.

CHAPTER II

DEFINITIONS

Louis et al. Model

In this chapter we will introduce some definitions needed in this study. The following definitions were given by Louis et al. (1978):

1. The Natural History of a Disease Process: Louis et al. (1978) defined the natural history of a disease process as a specification of the states that a subject can be expected to pass through during the course of that disease.

2. Progressive Disease: The adjective "progressive" is intended to connote that once a state is visited and abandoned, it cannot be visited again (presumably because the disease has progressed to a more advanced stage). Figure 1 illustrates the disease state for a progressive disease model for one person.

3. Louis et al. Model: The Louis et al. (1978) model consists of three disease states defined (see Figure 1) as:

(1). The Disease Free State (DFS) is the state in which the subject is free of the disease; this state has duration from birth to age X.

In the Louis et al. (1978) model $X = \infty$ means that the subject is not susceptible to the disease, or the subject will never get the disease.

(3). The Clinical State (CS) is the state in which symptoms are present. It is entered at age $X + Y = T$. For example, in cancer studies, T could be the time when the symptoms surface. In heart disease studies, T could be the time of death of the coronary heart disease, or could

be the time of getting the second heart attack. In the Louis et al. model, $X < \infty$ and $Y = \infty$ is a person who contracts the disease but never progresses to the clinical state.

A third variable in the model is the subject's age at time t , $A(t)$. Thus at an instant time t , each person has ascribed three quantities: (1) A = current age, (2) X = age at time of entry to PCS, and (3) Y = sojourn time in the PCS.

For 2 examples: (1) an individual has $A = 35$, $X = 25$, $Y = 5$, this person is now 35 years old, entered PCS at age 25, and entered CS 5 years later (at age 30); (2) $A = 30$, $X = 45$, $Y = 10$, this person is now 30 years old, will enter PCS at age 45, and will enter CS at age 55. Since data will be analyzed at a specific time t , we will suppress the argument t , so that A will refer to the age of the subject at the time of interest.

In this model, X and Y are considered fixed points in a person's life, and do not change over time, although they may be determined by random events. This is a crucial consequence of Louis et al. formulation. For example, if an individual (now) has age $A = 30$, $X = 35$, $Y = 10$, then 10 years from now, $A = 40$, $X = 35$, $Y = 10$. Age will change with time, but the temporal history of the disease state model (represented by X and Y) will not, see Figure 2 below.

Consider a population of subjects at a specific time. Associated with this population is a set of triples, (X,Y,A) values. This set of triples has a probability density function denoted by $f(x,y,a)$.

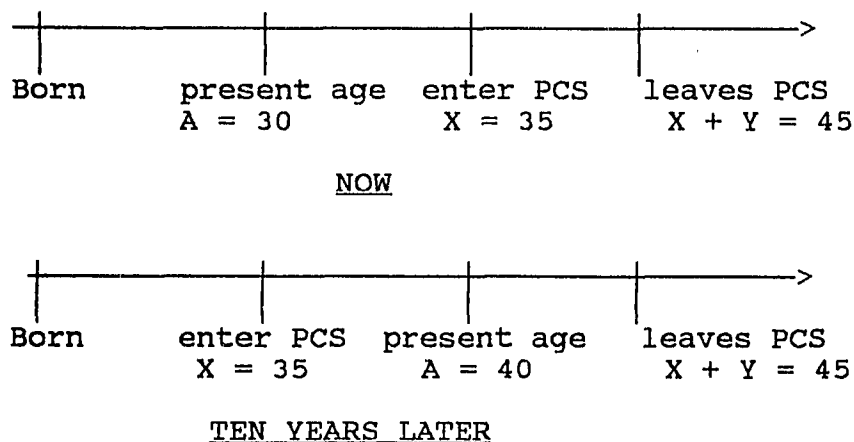


Figure 2. Age Changes With Time But X and Y Do Not.

Notice that allowing $X = \infty$ and $Y = \infty$ in the Louis et al. model means that $f(x,y,a)$ is a mixed density with a lump of probability at infinity points and the marginal densities for X and Y are generally defective (total integral is less than unity). In our model we will assume that $f(x,y,a)$ is continuous (so the lumps at infinity will have zero probability). To avoid getting those cases in our model, we make the assumption that if the patient lives long enough with no other competing risks intervening in the natural history of the disease state model, then he will enter the PCS and then will progress to the CS eventually. The joint density $f(x,y,a)$ represents a

probabilistic description of the natural history of the disease in the population of interest at time t .

The factors that determine the values for X and Y (age of entry into and sojourn time in PCS) for an individual are partly hereditary and partly due to environmental factors which operate on the individual. If strata of different ages experience different environmental factors which are relevant to the determination of (X,Y) , the end effect will be that the distribution of (X,Y) will vary from age stratum (or cohort) to age stratum. Therefore a cohort effect is said to exist in a population of interest, say Π , if the distribution of (X,Y) varies over age strata. Under the assumption of "no cohort effect," i.e., (X,Y) is independent of A , we have that

$$f_{xyA}(x,y,a) = f_{xy}(x,y) \cdot f_A(a),$$

where $f_{xy}(x,y)$ is the joint density of (X,Y) and $f_A(a)$ is the density of A (the age distribution of the subject population).

Standard Epidemiologic Measures of a Disease State

Louis et al. (1978) defined some classical descriptors of the natural history of the disease state model and derived also some formulas from those definitions. Some of those definitions are:

1. Nonsusceptible: A subject is nonsusceptible to the disease state if, for that subject, $X = \infty$. Such

subjects never enter PCS. But in our model, every subject is susceptible to the disease (if the subject lives long enough then he will get the disease eventually).

2. Chronic Habitue: In the Louis et al. model, a subject is a chronic habitue of the PCS if, for that subject, $X < \infty$ and $Y = \infty$. Such subjects never leave the PCS. According to our assumption, there will be no chronic habitues of the PCS (since if every person lives long enough then he will progress to the next state eventually).

3. Incidence: $I_{PC}(a)$ is the age specific preclinical incidence of PCS among those aged a . It is the instantaneous rate at which members of Π whose present age is a , enter PCS.

$$I_{PC}(a) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P[A < X \leq A + \delta | A = a] = f_{X|A}(a|a),$$

where $f_{X|A}(a|a)$ is the conditional density of X , given that $A = a$, evaluated at $X = a$.

The overall preclinical incidence, I_{PC} is the instantaneous rate at which subjects of all ages in Π enter PCS. Louis et al. showed

$$I_{PC} = \int_0^{\infty} I_{PC}(a) f_A(a) da.$$

If there is no cohort effect then Louis et al. showed that $I_{PC}(a) = f_X(a)$,

$$I_{PC} = \int_0^{\infty} I_{PC}(a) f_A(a) da = \int_0^{\infty} f_X(a) f_A(a) da,$$

and

$$\int_0^{\infty} I_{PC}(a) da = \int_0^{\infty} f_X(a) da = P[X < \infty],$$

in terms of our model, this quantity will be 1. The quantity $P[X < \infty]$ is called the preclinical lifetime attack rate (proportion of patients who enter the PCS) of the PCS among those in Π , provided that there is no cohort effect.

Similarly, $I_{CS}(a)$ is the age specific clinical incidence of CS among those aged a . It is the instantaneous rate at which members of Π whose present age is a , enter CS.

$$I_{CS}(a) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P[A < T \leq A + \delta | A = a] = f_{T|A}(a|a),$$

where $f_{T|A}(a|a)$ is the conditional density of T , given that $A = a$, evaluated at $T = a$, where $T = X + Y$.

Similarly, the overall clinical incidence, I_{CS} is the instantaneous rate at which subjects of all ages in Π enter CS. Louis et al. showed

$$I_{CS} = \int_0^{\infty} I_{CS}(a) f_A(a) da.$$

If there is no cohort effect then Louis et al. showed that $I_{CS}(a) = f_T(a)$,

$$I_{CS} = \int_0^{\infty} I_{CS}(a) f_T(a) da = \int_0^{\infty} f_T(a) f_A(a) da,$$

and

$$\int_0^{\infty} I_{CS}(a) da = \int_0^{\infty} f_T(a) da = P[T < \infty],$$

in terms of our model, this quantity will be 1. The quantity $P[T < \infty]$ is called the clinical lifetime attack rate (proportion of patients who enter the CS) of the CS among those in Π , provided that there is no cohort effect.

Incidence is interpreted as a measure of the rate at which new cases of the disease occur in a population previously without the disease (Armitage & Berry, 1987).

4. Prevalence: $\Phi_{PC}(a)$ is the age specific preclinical prevalence of the PCS among those aged a . It is the proportion of subjects among those aged a , who are then in the PCS. Equivalently, $\Phi_{PC}(a)$ is the probability that a subject chosen at random from those in Π who are aged a , will be in the PCS. Louis et al. wrote $\Phi_{PC}(a)$ in terms of the joint density of (X, Y, A) as

$$\begin{aligned} \Phi_{PC}(a) &= P[X \leq A < X+Y \mid A = a] \\ &= \int_0^{\infty} \int_{a-x}^{\infty} f_{X,Y|A}(x, y \mid a) dy dx, \end{aligned}$$

provided that there is no chronic habitues of the PCS. As we mentioned earlier, our assumption guarantees that there will be no chronic habitues of the PCS.

Φ_{PC} , the overall preclinical prevalence of PCS, is the proportion of those in Π (all ages) who are in the PCS. Equivalently, Φ_{PC} is the probability that a randomly selected subject from Π will be in the PCS. Therefore, Louis et al. showed that if there is no cohort effect and if there are no chronic habitues of the PCS (our assumption guarantees that there will be no chronic habitues of the PCS) then,

$$\Phi_{PC}(a) = \int_0^a \int_{a-x}^{\infty} f_{XY}(x, y) dy dx,$$

and therefore

$$\Phi_{PC} = \int_0^{\infty} \left(\int_0^a \int_{a-x}^{\infty} f_{XY}(x, y) dy dx \right) f_A(a) da.$$

Prevalence is interpreted as a measure of the frequency of existing disease at a given time. The prevalence rate and incidence rate are related since an incident case is immediately on occurrence a prevalent case and remains as such until recovery or death. Rate of change of preclinical prevalence equals the rate of change of the overall preclinical incidence minus the rate of change of the overall clinical incidence (Louis et al., 1978; Armitage & Berry, 1987).

5. Mean Duration of PCS (Sojourn Time): Louis et al. (1978) showed that the mean sojourn time in the PCS is related to incidence and prevalence both in the case of no

cohort effect and when the mean age entry to the preclinical state is finite, by

$$E(Y \mid X < \infty) = \frac{\int_0^{\infty} \Phi_{PC}(a) da}{\int_0^{\infty} I_{PC}(a) da}.$$

In our model, the left hand side of this equation will be $E(Y)$, since $P[X < \infty] = 1$.

Further Definitions

1. Follow up Period: The follow up period is the period which starts when a patient enters into observation and ends when one of three possible events occur: (1) failure times of interest occur, (2) patient leaves the study and cannot be seen any more (loss to follow up), or (3) termination of the study.

2. Survival Function: Let T be a nonnegative random variable representing a survival time. The survival function of T evaluated at time t is defined as (see Kalbfleisch & Prentice, 1980)

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(x) dx,$$

where $F(t)$ is the distribution function of T , defined as

$$F(t) = \int_0^t f(x) dx,$$

and f is

$$f(t) = -\frac{dS(t)}{dt}.$$

3. Hazard Function: A hazard function of T , $\lambda(t)$, is defined as the instantaneous rate of failure at time t , conditional upon survival to time t (see Kalbfleisch & Prentice, 1980), or

$$\begin{aligned}\lambda(t) &= \lim_{\Delta \rightarrow 0} \frac{P[t \leq T < t + \Delta | T \geq t]}{\Delta} = \frac{f(t)}{S(t)} \\ &= -\frac{\left(\frac{dS(t)}{dt}\right)}{S(t)} = -\frac{d \log_e[S(t)]}{dt}, \quad t > 0.\end{aligned}$$

Therefore

$$S(t) = e^{-\int_0^t \lambda(u) du}, \quad t > 0,$$

and

$$f(t) = \lambda(t) S(t) = \lambda(t) e^{-\int_0^t \lambda(u) du}, \quad t > 0.$$

The quantity

$$\Lambda(t) = \int_0^t \lambda(x) dx, \quad t > 0.$$

is called the integrated hazard or the cumulative hazard function.

4. Cox's Proportional Hazards Model: Let $\lambda(t|z)$ represent the hazard function at time t associated with the failure time T for a subject with the covariate vector z ;

the proportional hazards model which was suggested by Cox (1972) specifies that

$$\lambda(t|z) = \lambda_0(t) e^{\beta' z},$$

or

$$\log_e \left[\frac{\lambda(t|z)}{\lambda_0(t)} \right] = \beta' z,$$

where β is a vector of regression coefficients associated with z , and $\lambda_0(t)$ is the hazard function when all the covariates are equal to zero (Kalbfleisch & Prentice, 1980). The hazard function $\lambda_0(t)$ is an unknown quantity and is called the baseline hazard function. A proportional hazards model holds when the ratio of the hazard functions for two covariate vectors z_1 and z_2 does not vary with time t . Cox's proportional hazards model is formulated in terms of the effects of the covariates upon hazard rates rather than upon time to failure.

5. Clayton and Cuzick Bivariate Survival Function:
Clayton and Cuzick (1985b) described a class of bivariate distributions characterized by an association parameter and two arbitrary marginal distribution functions.

Let T_1 and T_2 be two nonnegative random variables (failure times). Let

$$F(t_1, t_2) = P[T_1 > t_1, T_2 > t_2], \quad t_1, t_2 > 0,$$

be the joint survival function of T_1 and T_2 . Then the bivariate survival function given by Clayton and Cuzick is

defined as

$$F(t_1, t_2) = (e^{\gamma \Lambda_1(t_1)} + e^{\gamma \Lambda_2(t_2)} - 1)^{-\frac{1}{\gamma}},$$

where $\gamma > 0$, $t_1 > 0$, $t_2 > 0$ and Λ_1 , and Λ_2 are the cumulative hazard functions of T_1 and T_2 respectively, and γ is a parameter that reflects the association between T_1 and T_2 .

The joint density function of T_1 and T_2 is

$$f(t_1, t_2) = (\gamma + 1) \lambda_1(t_1) \lambda_2(t_2) e^{\gamma \Lambda_1(t_1)} e^{\gamma \Lambda_2(t_2)} [e^{\gamma \Lambda_1(t_1)} + e^{\gamma \Lambda_2(t_2)} - 1]^{(-\frac{1}{\gamma} - 2)},$$

where $\gamma > 0$, $t_1 > 0$, $t_2 > 0$, and $\lambda_1(t_1)$ and $\lambda_2(t_2)$ are the hazard functions for T_1 and T_2 respectively.

It can easily be shown that the marginal densities of T_1 and T_2 are

$$f_{T_1}(t_1) = \lambda_1(t_1) e^{-\Lambda_1(t_1)}, \quad t_1 > 0,$$

$$f_{T_2}(t_2) = \lambda_2(t_2) e^{-\Lambda_2(t_2)}, \quad t_2 > 0.$$

It should be noted also that neither T_1 nor T_2 can be infinity with positive probability.

CHAPTER III

LIKELIHOOD FUNCTION FOR THE UNCENSORED CASE

Assumptions

To ease the notation, we begin by looking at the uncensored case. The modifications needed to accommodate the censored case will be studied in the next chapter. The likelihood function for the uncensored case helps in writing and understanding the likelihood function for the censored case.

Recall that the joint survival function for two nonnegative random variables (X,Y) , given by Clayton and Cuzick (1985b) is

$$F(x,y) = [e^{\gamma\Lambda_1(x)} + e^{\gamma\Lambda_2(y)} - 1]^{-\frac{1}{\gamma}}, \quad \gamma > 0, x > 0, y > 0,$$

where γ is an association parameter between X and Y , and Λ_1 and Λ_2 are the cumulative hazard functions for X and Y respectively. The joint density function of (X,Y) is

$$f(x,y) = (\gamma+1) \lambda_1(x) \lambda_2(y) e^{\gamma\Lambda_1(x)} e^{\gamma\Lambda_2(y)} D(x,y)^{(-\frac{1}{\gamma}-2)},$$

where $\gamma > 0, x > 0, y > 0, \lambda_1$ and λ_2 are the hazard functions associated with X and Y respectively and

$$D(x,y) = e^{\gamma\Lambda_1(x)} + e^{\gamma\Lambda_2(y)} - 1,$$

Consider a sample of n observations (n independent pairs), $(X_1, Y_1), (X_2, Y_2), \dots, (X_k, Y_k)$, of the two random variables (X, Y) whose joint density function is given by Clayton and Cuzick. We partition the X -axis into intervals I_1, I_2, \dots, I_M and partition the Y -axis into intervals J_1, J_2, \dots, J_N .

As we mentioned earlier in Chapter I, we will use the Chiang et al. (1989) assumption of constant baseline hazards (i.e., $\lambda_{1i}(x) = \mu_{1i}$, $x \in I_i$ and $\lambda_{2j}(y) = \mu_{2j}$, $y \in J_j$) in the i^{th} and j^{th} intervals respectively. We model the hazard functions for the k^{th} individual whose (X, Y) values fall in rectangle $I_i \times J_j$, by assuming Cox's (1972) proportional hazards model holds for each of X and Y in each interval I_i and J_j respectively, where $i = 1, \dots, M$ and $j = 1, \dots, N$.

The proportional hazards model will allow us to include covariates in the model in order to study the effects of the covariates on X and Y . We assume that the vector of covariates z is p -dimensional and the same for both X and Y .

Thus the hazard functions λ_1 and λ_2 in the i^{th} and j^{th} intervals for the k^{th} individual whose observed (X, Y) value is (x_k, y_k) are defined as

$$\lambda_{1i}(x_k) = \mu_{1i} e^{a_i' z_k}, \quad x_k \in I_i = (a_i, a_{i+1}],$$

and

$$\lambda_{2j}(y_k) = \mu_{2j} e^{B' z_k}, \quad y_k \in I_j = (b_j, b_{j+1}],$$

where $I_M = (a_M, a_{M+1}]$, $a_1 = 0$, $a_{M+1} = \infty$, $J_N = (b_N, b_{N+1}]$, $b_1 = 0$, $b_{N+1} = \infty$, μ_{1i} is the baseline hazard for X in I_i , $i = 1, 2, \dots, M$ and μ_{2j} is the baseline hazard for Y in J_j , $j = 1, 2, \dots, N$. Note that the μ_{1i} 's and μ_{2j} 's are unknown parameters to be estimated. z_k is the value of z for the k^{th} individual. $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_p)$ are the coefficients associated with z for the failure time X , and $B' = (\beta_1, \beta_2, \dots, \beta_p)$ are the coefficients associated with z for the failure time Y . We will assume that the regression parameters α and B for the covariates z are constant (the same) for all intervals.

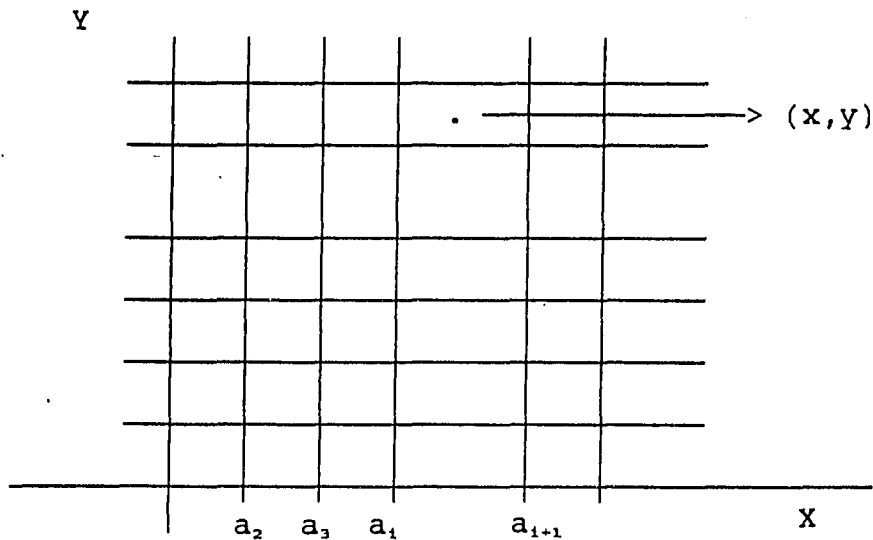


Figure 3. Dividing the Positive XY Plane Into Rectangles.

After we divide the first quadrant of XY-plane into rectangles as shown in Figure 3, then each individual observed values (x, y) will fall in one and only one

rectangle.

We need to compute the cumulative hazard functions associated with X and Y . First, we calculate the cumulative hazard function for the k^{th} individual whose X value falls in the i^{th} interval (assuming constant hazard over each interval) as follows

$$\begin{aligned}
 \Lambda_{1i}(x_k) &= \int_0^{x_k} \lambda_1(u) du \\
 &= \int_{I_1} \lambda_{11}(u) du + \int_{I_2} \lambda_{12}(u) du + \dots + \int_{I_{i-1}} \lambda_{1i-1}(u) du + \int_{a_i}^{x_k} \lambda_{1i}(u) du \\
 &= \int_{I_1} \mu_{11} e^{a'x_k} du + \int_{I_2} \mu_{12} e^{a'x_k} du + \dots + \int_{I_{i-1}} \mu_{1i-1} e^{a'x_k} du + \int_{a_i}^{x_k} \mu_{1i} e^{a'x_k} du \\
 &= [\mu_{11}(a_2 - a_1) + \dots + \mu_{1i-1}(a_i - a_{i-1}) + \mu_{1i}(x_k - a_i)] e^{a'x_k} \\
 &= \left[\sum_{r=1}^{i-1} \mu_{1r}(a_{r+1} - a_r) + \mu_{1i}(x_k - a_i) \right] e^{a'x_k}.
 \end{aligned}$$

Similarly, the cumulative hazard function for the k^{th} individual whose Y value falls in the j^{th} interval (assuming constant hazard over each interval) is

$$\Lambda_{2j}(y_k) = \left[\sum_{r=1}^{j-1} \mu_{2r}(b_{r+1} - b_r) + \mu_{2j}(y_k - b_j) \right] e^{b'y_k}.$$

where $(b_{r+1} - b_r)$ is the length of the r^{th} interval.

Likelihood Function

In this section we will build the likelihood function for the uncensored case, but with the failure times of

interest (X,T) , instead of (X,Y) . The progressive disease model (P.D.M) we are considering here has two nonnegative failure time random variables (X,T) , with $X \leq T$ (see Figure 1 of Chapter II). The distribution of (X,T) is the distribution which contributes directly to the likelihood function for the censored case (to be discussed in the next chapter). To write the likelihood function for the uncensored case, where (X,T) are the observed variables, we cannot apply the Clayton and Cuzick (1985b) joint density function in this case, since $X \leq T$.

To apply the Clayton and Cuzick formula we model it for (X,Y) first, where $X > 0$ and $Y > 0$, as we have done earlier, then we make the transformation $X = X$ and $T = X + Y$. To get the joint density function $g(x,t)$ of (X,T) (the Jacobian is 1) as

$$g(x, t) = f(x, t-x) = (\gamma+1) \lambda_1(x) \lambda_2(t-x) e^{\gamma[\Lambda_1(x) + \Lambda_2(t-x)]} D^{(-\frac{1}{\gamma}-2)},$$

where

$$D = e^{\gamma\Lambda_1(x)} + e^{\gamma\Lambda_2(t-x)} - 1,$$

$$\Lambda_{2j}(t_k - x_k) = \left[\sum_{r=1}^{j-1} \mu_{2r}(b_{r+1} - b_r) + \mu_{2j}(t_k - x_k - b_j) \right] e^{\beta' z_k}.$$

$\gamma > 0$, $0 < x \leq t$, λ_1 and λ_2 the hazard functions for X and $(T - X)$ and $\Lambda_1(x)$ and $\Lambda_2(t-x)$ are the cumulative hazard functions for X and $(T - X)$ respectively.

To write the likelihood function for the uncensored

case, we assume that we have n observations (n independent pairs) $(X_1, T_1), (X_2, T_2), \dots, (X_k, T_k)$ of the two random variables (X, T) whose joint density function $g(x, t)$ given above. The contribution to the likelihood function for the k^{th} individual whose (X, T) values falls in the $I_i \times J_n$ rectangle (then $T - X$ will fall in some interval, say j , $j \leq m$) as

$$\begin{aligned} f(x_k, t_k - x_k) &= \sum_{r=1}^M \sum_{q=1}^N f_{rq}(x_k, t_k - x_k) I[(x_k, t_k - x_k) \in I_r \times I_q] \\ &= \sum_{r=1}^M \sum_{q=1}^N (\gamma + 1) \mu_{1r} \mu_{2q} e^{(\alpha' + \beta') z_k} e^{\gamma \Lambda_{1r}(x_k)} e^{\gamma \Lambda_{2q}(t_k - x_k)} \\ &\quad \cdot D_{rqk}(x_k, t_k - x_k)^{(-\frac{1}{\gamma} - 2)} I[(x_k, t_k - x_k) \in I_r \times I_q], \quad (1) \end{aligned}$$

where

$$I[(x, y) \in I_r \times J_n] = \begin{cases} 1 & \text{if } x \in I_r, y \in J_n \\ 0 & \text{otherwise.} \end{cases}$$

$$f_{ij}(x_k, t_k - x_k) = (\gamma + 1) \mu_{1i} \mu_{2j} e^{(\alpha' + \beta') z_k} e^{\gamma [\Lambda_{1i}(x_k) + \Lambda_{2j}(t_k - x_k)]} D^{(-\frac{1}{\gamma} - 2)},$$

$$D = e^{\gamma \Lambda_{1i}(x_k)} + e^{\gamma \Lambda_{2j}(t_k - x_k)} - 1,$$

and

$$\Lambda_{2j}(t_k - x_k) = \left[\sum_{r=1}^{j-1} \mu_{2r} (b_{r+1} - b_r) + \mu_{2j} (t_k - x_k - b_j) \right] e^{\beta' z_k}.$$

Let n_{ij} = the number of individuals whose $(x_k, t_k - x_k) \in I_i \times J_j$, then

$$\sum_{i=1}^M \sum_{j=1}^N n_{ij} = n.$$

Then the overall likelihood for the n individuals is

$$L(\alpha, \beta, \gamma, \mu_1, \mu_2, X, t) = \prod_{k=1}^n f(x_k, t_k - x_k) \\ = \prod_{k=1}^n \left(\sum_{i=1}^M \sum_{j=1}^N f_{ij}(x_k, t_k - x_k) I[(x_k, t_k - x_k) \in I_i \times J_j] \right).$$

The log-likelihood becomes

$$\log L = \sum_{k=1}^n \log \left[\sum_{i=1}^M \sum_{j=1}^N f_{ij}(x_k, t_k - x_k) I[(x_k, t_k - x_k) \in I_i \times J_j] \right].$$

If for example, the k^{th} individual observed values (x_k, t_k) fall in rectangle $I_p \times J_q$ (then $(x_k, t_k - x_k)$ fall in $I_p \times J_r$, for some $r \leq q$), then the likelihood contribution for the k^{th} individual given in Equation (1), simplifies to $f_{pr}(x_k, t_k - x_k)$, i.e.,

$$f(x_k, t_k - x_k) = \sum_{p=1}^M \sum_{r=1}^N f_{pr}(x_k, t_k - x_k) I[(x_k, t_k - x_k) \in I_p \times J_r] \\ = f_{pr}(x_k, t_k - x_k) = (\gamma + 1) \mu_{1p} \mu_{2r} e^{(\alpha' + \beta') z_k} e^{\gamma \Lambda_{1p}(x_k)} \\ \cdot e^{\gamma \Lambda_{2r}(t_k - x_k)} D_{prk}(x_k, t_k - x_k)^{(-\frac{1}{\gamma} - 2)},$$

where

$$D_{pmk}(x_k, t_k - x_k) = e^{\gamma \Lambda_{1p}(x_k)} + e^{\gamma \Lambda_{2m}(t_k - x_k)} - 1.$$

Therefore

$$\begin{aligned}
\log L &= \sum_{k \in R_{11}} \log f_{11}(x_k, t_k - x_k) + \sum_{k \in R_{12}} \log f_{12}(x_k, t_k - x_k) + \dots \\
&+ \sum_{k \in R_{1N}} \log f_{1N}(x_k, t_k - x_k) + \dots + \sum_{k \in R_{MN}} \log f_{MN}(x_k, t_k - x_k) \\
&= \sum_{i=1}^M \sum_{j=1}^N \sum_{k \in R_{ij}} \log f_{ij}(x_k, t_k - x_k),
\end{aligned}$$

where R_{ij} is the set of indices for those $(X, T-X)$ s' in rectangle $I_i \times J_j$. Substituting for $f_{ij}(x_k, t_k - x_k)$ in the above equation, we get

$$\begin{aligned}
\log L &= \sum_{i=1}^M \sum_{j=1}^N \left(n_{ij} \log [(\gamma + 1) \mu_{1i} \mu_{2j}] + \sum_{k \in R_{ij}} \left[(\alpha' + \beta') z_k + \gamma [\Lambda_{1i}(x_k) \right. \right. \\
&\quad \left. \left. + \Lambda_{2j}(t_k - x_k)] + \left(-\frac{1}{\gamma} - 2\right) \log D_{ijk}(x_k, t_k - x_k) \right] \right),
\end{aligned}$$

where

$$D_{ijk}(x_k, t_k - x_k) = e^{\gamma \Lambda_{1i}(x_k)} + e^{\gamma \Lambda_{2j}(t_k - x_k)} - 1.$$

This likelihood function can be maximized with respect to the parameter vector

$$\mathbf{Q} = (\gamma, \mu_{11}, \dots, \mu_{1M}, \mu_{21}, \dots, \mu_{2N}, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p),$$

where p is the number of covariates.

The likelihood function for the uncensored case helps in understanding and writing the likelihood function for the censored case. The likelihood function for the uncensored case has little practical applications, since we may not be able to observe all (X, T) values for our model and for progressive diseases in general. For example, in

cancer studies, X could be the time for tumor onset (can be detected by a screening test, that gives us information about X) which is seldom observed; of course, if you can not observe X then you cannot determine Y , but T which could be the time when symptoms surface, can be observed often. Therefore we will deal with the case when (X, T) may or may not be completely observed in Chapter IV.

CHAPTER IV

LIKELIHOOD FUNCTION FOR THE CENSORED CASE

Observable Data for the Model

The approach of Louis et al. (1978) is to use non-parametric likelihood methods to estimate the joint distribution of (X,T) . In this approach the joint distribution is described by a finite collection of probabilities associated with regions (rectangles) in the plane.

In our study the joint distribution of (X,T) is specified by using the joint distribution given by Clayton and Cuzick (1985b), henceforth parametric likelihood methods will be employed to estimate the parameters of interest.

In the progressive disease model we have introduced earlier, (X,T) may or may not be completely observed. For example in cancer studies X (tumor onset time which can be detected by a screening test, that gives us information about X) is not observable, while T (time when symptoms surface) may or may not be observed. In the case of heart diseases, X (time to coronary heart disease, for example) may or may not be observed, while T (time to death from the disease) may or may not be observed also. Using the model

for the bivariate survivor function for the progressive disease model described in Chapter II and used for uncensored (X, T) in Chapter III, we investigate the likelihood function when X or T or both may be censored.

We will treat the situation in which X or T or both may be censored on the right. The observed data for each individual can be represented by the following data vector

$$w = \begin{bmatrix} e \\ U \\ V \\ z \\ R \\ C \end{bmatrix},$$

where

e = age of the subject when he enters the study.

$U = \min \{ X, \text{age at loss to follow up or time of analysis} \}.$

If the patient is lost to follow up, then $X > U$, otherwise $X = U$. We assume $X \geq e$, i.e., the subjects are disease free when they enter the study.

$V = \min \{ T, \text{age at loss to follow up or time of analysis} \}.$

If the patient is lost to follow up, then $T > V$, otherwise $T = V$.

z = is the vector of covariates measured on each subject at age e .

$$R = \begin{cases} 1 & , \quad X \text{ is censored; } X > U. \\ 0 & , \quad X \text{ is uncensored; } X = U. \end{cases}$$

$$C = \begin{cases} 1 & , \quad T \text{ is censored; } T > V. \\ 0 & , \quad T \text{ is uncensored; } T = V. \end{cases}$$

Likelihood Function for the Censored Case

In developing the likelihood function for the censored case, it is helpful to divide the data into four categories (cases) according to the pair of indicators (R,C); their contribution to the overall likelihood will be discussed in detail below. The k^{th} individual generates data vector \mathbf{w}_k and $L(\mathbf{w}_k)$ represents the likelihood contribution for the k^{th} individual.

To build the likelihood function for the censored case, the likelihood of observing a vector (e,U,V,R,C) for the k^{th} individual satisfying ($e_k \in I_n, u_k \in I_1, v_k \in J_n, R = r, C = c$) must be expressed in terms of the joint density function of (X,T), $h \leq i \leq m$.

Case 1: When $R = 0, C = 0$.

$R = 0$ implies that $X = U$; that is, the observed U is actually the age at which the subject enters the preclinical state before the end of the study. Also $C = 0$ implies that $T = V$; that is, the observed V is actually the age at which the subject enters the clinical state before the end of the study. So the contribution to the likelihood function for the k^{th} subject whose (U,V) values fall in $I_1 \times J_n$ rectangle (with (U,V-U) fall in $I_1 \times J_j, j \leq m$) is obtained by evaluating $f(x, t-x)$ at the observed point ($u_k,$

v_k) for the k^{th} individual. Therefore the likelihood contribution for the k^{th} individual in this case is

$$L(\mathbf{w}_k) = f_{ij}(u_k, v_k - u_k) = (\gamma + 1) \mu_{1i} e^{\alpha' z_k} \mu_{2j} e^{\beta' z_k} \cdot e^{\gamma \Lambda_{1i}(u_k)} e^{\gamma \Lambda_{2j}(v_k - u_k)} [D_{ijk}(u_k, v_k - u_k)]^{(-\frac{1}{\gamma} - 2)}. \quad (1)$$

Case 2: When $R = 0$, $C = 1$.

$R = 0$ implies that $X = U$; that, is the observed U is actually the age at which the subject enters the preclinical state before the end of the study, while $C = 1$ implies that $T > V$; that is, V is a censoring time on T , and T is not observed before the end of study, but we assume that it will occur at sometime after V . So the contribution to the likelihood function for the k^{th} subject whose (U, V) values fall in $I_i \times J_m$ rectangle (with $(U, V - U)$ fall in $I_i \times J_j$, $j \leq m$) can be obtained by integrating $f(x, t - x)$ with respect to t from $t = v_k$ to $t = \infty$, evaluated at $x = u_k$. Therefore the likelihood contribution for the k^{th} individual in this case is

$$\begin{aligned}
L(\underline{w}_k) &= \int_{v_k}^{\infty} f(u_k, t_k - u_k) dt_k \\
&= \int_{v_k}^{b_{m+1}} f_{im}(u_k, t_k - u_k) dt_k + \sum_{r=m+1}^N \int_{b_r}^{b_{r+1}} f_{ir}(u_k, t_k - u_k) dt_k \\
&= \int_{v_k}^{b_{m+1}} (\gamma + 1) \mu_{1i} \mu_{2m} e^{[\alpha' + \beta'] z_k} e^{\gamma [\Lambda_{1i}(u_k) + \Lambda_{2m}(t_k - u_k)]} D_{imk}^{(-\frac{1}{\gamma} - 2)} dt_k \\
&\quad + \sum_{r=m+1}^N \int_{b_r}^{b_{r+1}} (\gamma + 1) \mu_{1i} e^{\alpha' z_k} \mu_{2r} e^{\beta' z_k} e^{\gamma \Lambda_{1i}(u_k)} e^{\gamma \Lambda_{2r}(t_k - u_k)} D_{irk}^{-\frac{1}{\gamma} - 2} dt_k \\
&= \mu_{1i} e^{\alpha z_k} e^{\gamma \Lambda_{1i}(u_k)} \left[D(u_k, v_k - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{m+1} - u_k)^{(-\frac{1}{\gamma} - 1)} \right] \\
&\quad + \sum_{r=m+1}^N \mu_{1i} e^{\alpha z_k} e^{\gamma \Lambda_{1i}(u_k)} \left[D(u_k, b_r - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{r+1} - u_k)^{(-\frac{1}{\gamma} - 1)} \right] \\
&= \mu_{1i} e^{\alpha z_k} e^{\gamma \Lambda_{1i}(u_k)} \left[D(u_k, v_k - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{m+1} - u_k)^{(-\frac{1}{\gamma} - 1)} \right] \\
&\quad + \mu_{1i} e^{\alpha' z_k} e^{\gamma \Lambda_{1i}(u_k)} \left[D(u_k, b_{m+1} - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{m+2} - u_k)^{(-\frac{1}{\gamma} - 1)} + \right. \\
&\quad \cdot D(u_k, b_{m+2} - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{m+3} - u_k)^{(-\frac{1}{\gamma} - 1)} + \dots + \\
&\quad + \dots + D(u_k, b_{N-1} - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_N - u_k)^{(-\frac{1}{\gamma} - 1)} \\
&\quad \left. + D(u_k, b_N - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{N+1} - u_k)^{(-\frac{1}{\gamma} - 1)} \right] \\
&= \mu_{1i} e^{\alpha' z_k} e^{\gamma \Lambda_{1i}(u_k)} \left[D(u_k, v_k - u_k)^{(-\frac{1}{\gamma} - 1)} - D(u_k, b_{N+1} - u_k)^{(-\frac{1}{\gamma} - 1)} \right] \\
&= \mu_{1i} e^{\alpha' z_k} e^{\gamma \Lambda_{1i}(u_k)} D(u_k, v_k - u_k)^{(-\frac{1}{\gamma} - 1)}. \tag{2}
\end{aligned}$$

Since $b_{N+1} = \infty$, then

$$\Lambda_{2N}(b_{N+1}) = \infty, \quad D(u_k, b_{N+1} - u_k) = \infty.$$

and therefore

$$D(u_k, b_{N+1} - u_k)^{(-\frac{1}{\gamma} - 1)} = 0.$$

Case 3: When $R = 1$, $C = 0$.

$C = 0$ implies that $T = V$; that is, the observed V is actually the age at which the subject enters the clinical state before the end of the study. However, $R = 1$ implies that $X > U$; that is, U is a censoring time on X ; X is not observed, but we know that X occurred before the subject entered the CS, that is, X occurred between the observed u_k and v_k , ($u_k < X \leq v_k$). So the likelihood contribution for the k^{th} subject whose (U, V) values fall in $I_i \times J_a$ rectangle (with $(U, V-U)$ fall in $I_i \times J_j$, $j \leq m$) is obtained by integrating $f(x, t-x)$ with respect to x from u_k to v_k , evaluated at $t = v_k$. Therefore the likelihood contribution for the k^{th} individual in this case is

$$\begin{aligned} L(\underline{w}_k) &= \int_{u_k}^{v_k} f(x_k, v_k - x_k) dx_k = \int_{u_k}^{a_{i+1}} f_{ij}(x_k, v_k - x_k) dx_k \\ &+ \sum_{r=1+1}^{c-1} \int_{a_r}^{a_{r+1}} f_{rj}(x_k, v_k - x_k) dx_k + \int_{a_c}^{v_k} f_{cj}(x_k, v_k - x_k) dx_k, \quad (3) \end{aligned}$$

where v_k falls in interval c on the x -axis,

$$f_{ij}(u_k, v_k - u_k) = (\gamma + 1) \mu_{1i} e^{a' z_k} \mu_{2j} e^{\beta' z_k} e^{\gamma \Delta_{1i}(u_k)} e^{\gamma \Delta_{2j}(v_k - u_k)}$$

$$\cdot D_{ijk}(u_k, v_k - u_k)^{(-\frac{1}{\gamma} - 2)},$$

and

$$D_{ijk}(u_k, v_k - u_k) = e^{\gamma \Lambda_{1i}(u_k)} + e^{\gamma \Lambda_{2j}(v_k - u_k)} - 1.$$

The above integrals in equation (9) cannot be obtained in a closed form; therefore, they must be evaluated numerically.

Case 4: When $R = 1$, $C = 1$.

$R = 1$ implies that $X > U$; that is, U is a censoring time on X , and X is not observed. $C = 1$ implies that $T > V$; that is, V is a censoring time on T , and T is not observed. For this case, we say that this subject has survived without entering the PCS or CS, which means that by the time v_k neither X nor T is observed, that is, $v_k < X \leq T$. So the likelihood contribution for the k^{th} subject whose (U, V) values fall in $I_i \times J_j$ rectangle (with $(U, V - U)$ fall in $I_i \times J_j$, $j \leq m$) is obtained by integrating $f(x, t - x)$ with respect to x and t over the region $(X > v_k, T > x_k)$. Therefore the likelihood contribution for the k^{th} individual in this case is

$$\begin{aligned}
L(\underline{w}_k) &= \int_{v_k}^{\infty} \int_{x_k}^{\infty} f(x_k, t_k - x_k) dt_k dx_k \\
&= \int_{v_k}^{\infty} \int_{x_k}^{\infty} (\gamma + 1) \mu_{1c} \mu_{2j} e^{(\alpha' + \beta')x} e^{\gamma [\Lambda_{1c}(x_k) + \Lambda_{2j}(t_k - x_k)]} D^{(-\frac{1}{\gamma} - 2)} dt_k dx_k \\
&= \int_{v_k}^{\infty} \mu_{1c} e^{\frac{\alpha'}{\gamma} x_k} e^{\gamma \Lambda_{1c}(x_k)} D_{cjk}(x_k, x_k - x_k)^{(-\frac{1}{\gamma} - 1)} dx_k \\
&= \int_{v_k}^{\infty} \mu_{1c} e^{\frac{\alpha'}{\gamma} x_k} e^{-\Lambda_{1c}(x_k)} dx_k \\
&= e^{-\Lambda_{1c}(v_k)}. \tag{4}
\end{aligned}$$

where v_k falls in interval c on the x -axis.

Now the k^{th} individual generates data vector w_k , therefore the overall likelihood for the n individuals is

$$L(w_1, w_2, \dots, w_n) = \prod_{k=1}^n L(w_k).$$

This likelihood is to be maximized with respect to the unknown parameters. In Table 1 below, we summarize the likelihood contributions for the observed data.

Table 1
Likelihood Contributions for Observed Data

R	C	likelihood contribution for data
0	0	$(\gamma+1) \mu_{1i} \mu_{2j} e^{(\alpha'+\beta')z_k} e^{\gamma(\Lambda_{1i}+\Lambda_{2j})} D(u_k, v_k-u_k)^{(-\frac{1}{\gamma}-2)}$
0	1	$\mu_{1i} e^{\alpha'z} e^{\gamma\Lambda_{1i}(u_k)} D_{ijk}(u_k, v_k-u_k)^{(-\frac{1}{\gamma}-1)}$
1	0	$\int_{u_k}^{v_k} f(x_k, v_k-x_k) dx_k$
1	1	$e^{-\Lambda_{1i}(v_k)}$

CHAPTER V

LOUIS ET AL. EPIDEMIOLOGICAL MEASURES IN TERMS OF OUR MODEL

The Epidemiological Measures in Terms of Our Model

We have introduced the Louis et al. (1978) model in Chapter II. We also introduced some standard epidemiological measures defined by Louis et al. in terms of their model. In this chapter we will define those measures in terms of our model. In Table 2 below, we summarize some of the relationships between incidence, prevalence and mean sojourn times.

Table 2

Key Relationships	
relationship	Sufficient conditions for validity
$I_{PC}(a)$	no cohort effect
I_{PC}	no cohort effect, $f_A(.)$ is uniform
$I_{CS}(a)$	no cohort effect
I_{CS}	no cohort effect, $f_A(.)$ is uniform
$\Phi_{PC}(a)$	no cohort effect
Φ_{PC}	no cohort effect $f_A(.)$ is uniform
$E(Y X<\infty)$	no cohort effect

1. Incidence:

If there is no cohort effect then we can write the preclinical incidence among those aged a , defined in terms of our model as

$$I_{PC}(a) = f_X(a) = \sum_{i=1}^M \mu_{1i} e^{a'x} e^{-\Lambda_{1i}(a)} I[aeI_i], \quad (1)$$

Louis et al. showed in Chapter II that the overall preclinical incidence is

$$I_{PC} = \int_0^{\infty} f_X(x) f_A(x) dx,$$

In order to define I_{PC} in terms of our model we need to define the distribution of A . We assume that A is uniformly distributed as

$$f_A(a) = \frac{1}{Md_i}, \quad a \in I_i,$$

where M is the number of intervals on the X -axis and d_i is the length of interval i . Notice that

$$\int_0^{\infty} f_A(a) da = \sum_{i=1}^M \int_{I_i} \frac{1}{Md_i} da = 1.$$

Since we have (a possible) different scale of intervals on the Y -axis in general, then $f_A(a)$ can be defined in this case as

$$f_A(a) = \frac{1}{Ns_j}, \quad a \in I_j,$$

where N is the number of intervals on the Y -axis and s_j is

the length of interval j .

The overall preclinical incidence in terms of our model becomes

$$\begin{aligned} I_{PC} &= \int_0^{\infty} f_X(x) f_A(x) dx = \sum_{i=1}^M \frac{1}{Md_i} \int_{T_i} \mu_{1i} e^{x/z} e^{-\Lambda_{1i}(x)} dx \\ &= \sum_{i=1}^M \frac{1}{Md_i} [e^{-\Lambda_{1i}(a_i)} - e^{-\Lambda_{1i}(a_{i+1})}], \end{aligned} \quad (4)$$

Similarly, if there is no cohort effect, then the clinical incidence among those aged a becomes

$$\begin{aligned} I_{CS}(a) &= f_T(a) = \int_0^a f(x, a-x) dx \\ &= \int_0^a (\gamma+1) \mu_{1i} \mu_{2j} e^{(\alpha'+\beta')x} e^{\gamma[\Lambda_{1i}(x) + \Lambda_{2j}(a-x)]} D^{(-\frac{1}{\gamma}-2)} dx \end{aligned} \quad (5)$$

where

$$D = e^{\gamma\Lambda_{1i}(x)} + e^{\gamma\Lambda_{2j}(a-x)} - 1.$$

Since the above integral cannot be obtained in a closed form, it must be evaluated numerically.

The overall clinical incidence in terms of our model becomes

$$\begin{aligned} I_{CL} &= \int_0^{\infty} f_T(t) f_A(t) dt = \int_0^{\infty} \left(\int_0^t f(x, t-x) dx \right) f_A(t) dt \\ &= \sum_{j=1}^N \frac{1}{Ns_j} \int_{T_j} \left(\int_0^t f(x, t-x) dx \right) dt, \end{aligned} \quad (6)$$

where

$$f_{ij}(x, t-x) = (\gamma+1) \mu_{1i} \mu_{2j} e^{(\alpha'+\beta')x} e^{\gamma[\Lambda_{1i}(x) + \Lambda_{2j}(a-x)]} D^{(-\frac{1}{\gamma}-2)},$$

and

$$D = e^{\gamma\Lambda_{1i}(x)} + e^{\gamma\Lambda_{2j}(a-x)} - 1.$$

This integral must be evaluated numerically.

1. Prevalence:

If there is no cohort effect, then the preclinical prevalence among those aged a , given by Louis et al. in Chapter II is

$$\Phi_{PC}(a) = \int_0^a \left(\int_{a-x}^{\infty} f(x, y) dy \right) dx,$$

then the preclinical prevalence among those aged a , defined in terms of our model becomes

$$\Phi_{PC}(a) = \int_0^a \mu_{1i} e^{\alpha'x} e^{\gamma\Lambda_{1i}(x)} D(x, a-x)^{(-\frac{1}{\gamma}-1)} dx. \quad (7)$$

This integral cannot be obtained in a closed form and must be evaluated numerically.

The overall preclinical prevalence becomes

$$\begin{aligned} \Phi_{PC} &= \int_0^{\infty} \Phi_{PC}(a) f_A(a) da = \sum_{j=1}^N \frac{1}{NS_j} \int_{J_j} \Phi_{PC}(a) da \\ &= \sum_{j=1}^N \frac{1}{NS_j} \int_{J_j} \left(\int_0^a \mu_{1i} e^{\alpha'x} e^{\gamma\Lambda_{1i}(x)} D(x, a-x)^{(-\frac{1}{\gamma}-1)} dx \right) da, \end{aligned} \quad (8)$$

where

$$D(x, a-x) = e^{\gamma \Lambda_{1i}(x)} + e^{\gamma \Lambda_{2j}(a-x)} - 1.$$

This integral cannot be obtained in a closed form and must be evaluated numerically.

3. Mean Duration of Preclinical State (Sojourn Time):

Recall that the mean duration time of the preclinical state defined by Louis et al. in Chapter II is

$$E(Y|X < \infty) = \frac{\int_0^{\infty} \Phi_{PC}(a) da}{\int_0^{\infty} I_{PC}(a) da}, \quad (9)$$

according to our model, the quantity $E[Y | X < \infty]$ will be $E(Y)$, since $P[X < \infty] = 1$.

Substituting for $\Phi_{PC}(a)$ and $I_{PC}(a)$ which are given in equations (1) and (7) in equation (9), the mean duration time in terms of our model becomes

$$E(Y) = \frac{\sum_{j=1}^N \int_{I_j} \left(\int_0^a \mu_{1i} e^{\alpha' x} e^{\gamma \Lambda_{1i}(x)} D(x, a-x)^{(-\frac{1}{\gamma}-1)} dx \right) da}{\sum_{i=1}^M [e^{-\Lambda_{1i}(a_i)} - e^{-\Lambda_{1i}(a_{i+1})}]}. \quad (10)$$

The denominator of this equation is 1, since

$$\int_0^{\infty} I_{PC}(a) da = \int_0^{\infty} f_X(x) dx = P[X < \infty] = 1,$$

according to our assumption, if every patient lives long enough then he will get the disease eventually.

The Asymptotic Distributions of the Standard Epidemiologic Measures

In order to make inferences about the epidemiological measures obtained in the last section, we need to find their distributions. To find the mean and variance of each of those measures we will apply the delta method (Bishop [1973]). Define the parameter vector to be estimated as

$$\underline{\theta} = (\gamma, \mu_{11}, \dots, \mu_{1M}, \mu_{21}, \dots, \mu_{2N}, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p).$$

Since $\hat{\underline{\theta}}$ is the MLE of $\underline{\theta}$, and from the properties of the maximum likelihood estimates, $\hat{\underline{\theta}}$ is approximately normal with mean $\underline{\theta}$ and covariance matrix $I(\underline{\theta})^{-1}$.

Let $g(\underline{\theta})$ be a function of $\underline{\theta}$; we want to approximate the distribution of $g(\underline{\theta})$. By using the delta method, we get

$$g(\hat{\underline{\theta}}) \approx N\left(g(\underline{\theta}), \left[\frac{dg(\underline{\theta})}{d\underline{\theta}}\right] [I(\underline{\theta})]^{-1} \left[\frac{dg(\underline{\theta})}{d\underline{\theta}}\right]'\right),$$

where

$$\frac{dg(\underline{\theta})}{d\underline{\theta}} = \left(\frac{dg}{d\theta_1}, \frac{dg}{d\theta_2}, \dots, \frac{dg}{d\theta_v}\right), \quad \dim(\underline{\theta}) = v = 2p + M + N + 1,$$

The estimated variance of $g(\hat{\underline{\theta}})$ is

$$\left(\left[\frac{dg(\underline{\theta})}{d\underline{\theta}}\right] [I(\underline{\theta})]^{-1} \left[\frac{dg(\underline{\theta})}{d\underline{\theta}}\right]'\right)_{\underline{\theta}=\hat{\underline{\theta}}};$$

this quantity is evaluated at $\theta = \hat{\theta}$.

As an example, we will derive the formulas for the derivatives of the preclinical incidence and the overall preclinical incidence. The estimate of the preclinical incidence of those who are aged a

$$g_{\theta}(a) = I_{PC}(a) = \sum_{i=1}^M \mu_{1i} e^{a's} e^{-\Lambda_{1i}(a)} \cdot I[aeI_i]$$

is

$$g_{\theta}(a) = \hat{I}_{PC}(a) = \sum_{i=1}^M \hat{\mu}_{1i} e^{a's} e^{-\hat{\Lambda}_{1i}(a)} \cdot I[aeI_i],$$

where

$$\hat{\Lambda}_{1i}(a) = \left(\sum_{r=1}^{i-1} \hat{\mu}_{1r} (a_{i+1} - a_r) + \hat{\mu}_{1i} (a - a_i) \right) e^{a's}.$$

If we differentiate g with respect to θ , we get

$$\frac{dg}{d\gamma} = \frac{dg}{d\beta_m} = \frac{dg}{d\mu_{2r}} = 0, \quad (11)$$

$$\frac{dg}{d\alpha_m} = \sum_{i=1}^M \mu_{1i} z_k e^{a'z_k} e^{-\Lambda_{1i}(a)} (1 - \Lambda_{1i}(a)) \cdot I[aeI_i], \quad (12)$$

$$\begin{aligned} & \sum_{s=1}^M e^{a's} e^{-\Lambda_{1i}(a)} [1 - \mu_{1s} (a - a_s) e^{a's}] \cdot I[aeI_i], \quad s=i \\ dg(\theta)/d\mu_{1s} = & \begin{cases} & (13) \\ -\sum_{i=1}^M \mu_{1i} e^{2a's} (a_{s+1} - a_s) e^{-\Lambda_{1i}(a)} \cdot I[aeI_i], & s < i. \end{cases} \end{aligned}$$

$$dg(\theta)/d\mu_{1s} = 0, \text{ for } s > i.$$

is

$$\left(\left[\frac{dg(\underline{\theta})}{d\underline{\theta}} \right] [I(\underline{\theta})]^{-1} \left[\frac{dg(\underline{\theta})}{d\underline{\theta}} \right]' \right),$$

where $I(\underline{\theta})^{-1}$ is the inverted covariance matrix of $\underline{\theta}$ obtained from maximizing the log-likelihood for the censored case, and $dg/d\underline{\theta}$ is given by

$$\frac{dg(\underline{\theta})}{d\underline{\theta}} = \left(\frac{dg}{d\theta_1}, \frac{dg}{d\theta_2}, \dots, \frac{dg}{d\theta_v} \right), \quad \dim(\underline{\theta}) = v = 2p + M + N + 1,$$

which is obtained from Equations (11), (12) and (13).

The estimate of the overall preclinical incidence

$$g(\underline{\theta}) = I_{PC} = \sum_{i=1}^M \frac{1}{Md_i} (e^{-\Lambda_{1i}(a_i)} - e^{-\Lambda_{1i}(a_{i+1})}),$$

is

$$g(\underline{\theta}) = \hat{I}_{PC} = \sum_{i=1}^M \frac{1}{Md_i} (e^{-\hat{\Lambda}_{1i}(a_i)} - e^{-\hat{\Lambda}_{1i}(a_{i+1})}), \quad (14)$$

where

$$\Lambda_{1i}(a_i) = \sum_{r=1}^{i-1} \mu_{1r}(a_{r+1} - a_r) e^{a'_r/s}, \quad \Lambda_{1i}(a_{i+1}) = \sum_{r=1}^i \mu_{1r}(a_{r+1} - a_r) e^{a'_r/s}.$$

If we differentiate g with respect to $\underline{\theta}$, we get

$$\frac{dg(\underline{\theta})}{d\gamma} = \frac{dg(\underline{\theta})}{d\beta_m} = \frac{dg(\underline{\theta})}{d\mu_{2r}} = 0, \quad (15)$$

$$\frac{dg}{d\alpha_m} = \sum_{i=1}^M \frac{1}{Md_i} z_k [\Lambda_{1i}(a_{i+1}) e^{-\Lambda_{1i}(a_{i+1})} - \Lambda_{1i}(a_i) e^{-\Lambda_{1i}(a_i)}]. \quad (16)$$

$$\begin{aligned}
 & \sum_{i=1}^M \frac{1}{M} e^{\alpha' z} e^{-\Lambda_{1i}(a_{i+1})}, \quad s=i \\
 dg(\underline{\theta})/d\mu_{1s} = & \begin{cases} \\ \sum_{i=1}^M \frac{1}{M} e^{\alpha' z} [e^{-\Lambda_{1i}(a_{i+1})} - e^{-\Lambda_{1i}(a_i)}], & s < i. \end{cases} \quad (17)
 \end{aligned}$$

$dg(\underline{\theta})/d\mu_{1s} = 0$, for $s > i$.

The covariance matrix for I_{PC} is

$$\left(\left[\frac{dg(\underline{\theta})}{d\theta} \right] [I(\underline{\theta})]^{-1} \left[\frac{dg(\underline{\theta})}{d\theta} \right]' \right), \quad (18)$$

where $I(\underline{\theta})^{-1}$ is the inverted covariance matrix of $\underline{\theta}$ obtained from maximizing the log-likelihood for the censored case, and $dg/d\theta$ is given by

$$\frac{dg(\underline{\theta})}{d\theta} = \left(\frac{dg}{d\theta_1}, \frac{dg}{d\theta_2}, \dots, \frac{dg}{d\theta_v} \right), \quad \dim(\underline{\theta}) = v = 2p + M + N + 1,$$

which is obtained from the above derived formulas.

In this chapter we were able to write Louis et al. epidemiological measures in terms of covariates and that is one of our objectives in this study.

The derivation of the means and variances of the above measures is one of our objectives in this study too. We can make inferences about those measures. For example, one can compare the incidence rates of two groups (males versus females or treatment versus controlled), or even can calculate confidence intervals about those measures (see Table 5).

The mean sojourn time in the PCS plays a fundamental role in the evaluation of the screening program for early detection of the disease. Zelen (1974) has suggested that the survival time for certain types of cancer patients may be positively correlated with preclinical sojourn duration time Y . Individuals with large Y values tend to have more slowly developing disease and therefore can be expected to live longer than those with small Y values (and rapidly developing diseases). Zelen argued that positive screenees tend to have longer than the average preclinical sojourn times. If this is so, he concludes, then the prognosis of patients who are discovered via screening is, on the average, better than those patients who present with clinically "manifest" symptoms.

CHAPTER VI

EXAMPLE

Description of the Data

The data which we are using are extracted from the Framingham Heart Study Data Set (FHSD) extracted from Kahn and Sempos (1989) which contains 13 variables including an 18-year follow-up. We used a subset of these data and we chose a subset of the available covariates for our example. The covariates in this set are sex, cholesterol, diastolic blood pressure, systolic blood pressure, and cigarette smoking.

The listing and summary included in Table 3 below contain 10 variables for Coronary Heart Disease (CHD) incidence and total mortality with an 18-year follow up period. A brief description of the definition and range of values for each variable are presented in Table 3 below.

A complete data set without missing values for the covariates of any of the patients is required for the application of the model. Cases were excluded from the analysis if there were missing values for any of the variables or if there is a preexisting CHD at the first examination, (since we assume that $X \geq e$). A definite CHD

Table 3

Variables Used in the Model for the
Application of the FHSD

Variable Name	Variable description
SEX	0 Men 1 Women
AGE	Age at first examination 45-62, Age in years.
SBP	Systolic blood pressure at first examination 90-300 mm Hg
DBP	Diastolic blood pressure at first examination 50-160 mm Hg
CHOL	Serum cholesterol at first examination. 96-430 Mg/100 ml.
CIG	Number of cigarettes smoked per day at first examination. 0-60.
CHD	Age at withdrawal or first CHD event.
DEATH	Age at withdrawal or death.
R	Censoring indicator for CHD 0 means exact CHD time is observed (uncensored) 1 means withdrawn before CHD is observed (censored)
C	Censoring indicator for death 0 means exact death time is observed (uncensored) 1 means withdrawn alive.

is first diagnosed as an incidence case at any examination if, after a review of all available information, a panel of investigators agreed upon a definite diagnosis of myocardial infarction, coronary insufficiency, angina pectoris, or CHD death. There are causes of death in the study other than CHD death, such as cancer, other cardiovascular diseases, stroke and others. These other causes of death are considered as censoring times and we assume that CHD death would have occurred some time after that.

The Appendix found in Khan and Sempos (1989) is also a subset of the original study. This Appendix is not a random sample of the Framingham data.

For more a complete description and discussion of the FHSD, see Gordon and Shurtleff (1973) and Shurtleff (1974).

Estimation of the Parameters

The method used to estimate the parameters is the maximum likelihood method. A first glance at the log-likelihood function for the censored case seems very difficult to maximize. In estimating the parameters using the MLE method, we ran into many problems, and in fact they are problems to be faced in many parametric estimation problems (see Bard, 1974; Nash & Walker-Smith, 1987).

First, the first and second derivatives of the log-likelihood function for the censored case with respect to the parameters are long and not easy to write in a closed

form. In fact, if we have a large number of parameters to be estimated, then writing the first and second derivatives will not be a convenient way and this will make the estimation process longer and even inefficient.

Second, the log-likelihood function for the censored case is nonlinear in the parameters. Nonlinear parametric estimation is a difficult task (see Nash & Walker-Smith, 1987).

Third, the choice of intervals is another problem. Using finer and finer intervals makes it possible for the model to provide closer approximation to the parameters, but because the number of observations (number of those exposed to risk) in each interval becomes smaller and smaller, the accuracy of the estimates is more likely to decrease (Elandt-Johnson & Johnson, 1980; Holford 1976). Also, choosing more intervals will add more parameters to be estimated. This will make convergence slower, and since we are using an iterative method (which requires initial values) to maximize the log-likelihood function, this will make it harder to find good reasonable initial points.

Fourth, the choice of initial values is another problem. The method used to maximize the log-likelihood function is a quasi-Newton method, which is an iterative method that is sensitive to initial values. The more parameters we have, the harder it is to get reasonable initial values. If we do not choose good initial points,

then this will either slow convergence or will cause divergence (overflow, underflow, divide by zero or logarithm of a negative number). We chose zeroes for the regression coefficients (alphas and betas). One way of finding good initial values for the baseline hazards associated with CHD morbidity and CHD mortality is by calculating the percentages of CHD morbidity and mortality in each interval.

Fifth, the algorithm used for the maximization of the log-likelihood function requires bounds on the parameters. For example, the baseline hazards are constrained to be positive (since a hazard function is a rate). Alphas and betas are not constrained. Gamma was constrained to be positive (since it is constrained to be positive in Clayton and Cuzick formula).

Sixth, since some of the integrals in the likelihood function for the censored case are not obtained in a closed form, we have to evaluate those integrals numerically in order to estimate the parameters. Numerical integration requires an enormous number of evaluations and computations during the maximization process. This, of course, will accumulate rounding errors, which may slow convergence significantly (especially if we have a high censoring percentage) and may affect the accuracy of the estimates. In order to minimize the approximation errors and to cut short on the computations and make them faster, easier, and

more efficient, we will consider one covariate (SEX) and will take 5 intervals. In this case we will have 13 parameters to estimate. We take a subset of the data in the Appendix of Khan and Sempos (1989) of 614 men and 687 women.

As we noticed from the above problems, the number of parameters to be estimated has an impact on the rate of convergence. A large number of parameters can slow convergence rapidly. For example, when printing the parameters for every evaluation of the log-likelihood function within each iteration, we observe that the IMSL subroutine used to maximize the log-likelihood (B2ONF) works on improving one or few parameters at a time, leaving the others unchanged during that iteration. So the more parameters we have, the slower the convergence will be.

We need to look for an iterative method that maximizes a nonlinear function, derivative-free, with constraints (bounds), is less sensitive to initial points, and handles many parameters.

We tried first to use BMDP, since it has programs for handling maximum likelihood estimation and survival analysis. It turned out that BMPD is inefficient for this problem, essentially because the program we used (LE) is for maximizing univariate functions. We tried also to implement the EM algorithm (Estimation-Maximization algorithm) to obtain the MLE estimates. There were

setbacks to applying the EM algorithm:

1. Each iteration for the EM algorithm involves two steps (an estimation step and a maximization step) which involves many evaluations which will slow convergence.

2. In general, it is very hard to get convergence with a high censoring percentage.

3. For the maximization step, we need an iterative method to maximize the log-likelihood function for the censored case (IMSL subroutines). This means that, for every maximization step, we need to call or use a maximization subroutine.

4. Most of the formulas involved in the application of the EM algorithm have to be evaluated numerically. As we mentioned earlier, this will add more computations and numerical errors; in addition to this, more computer time is needed which could be very expensive.

5. We need a separate EM algorithm to get the standard errors of the estimated parameters which means that we have to face the same problems as above. Louis et al. (1978) did not get the standard errors of the MLE estimates, but in a later work, Louis (1982) used the EM algorithm to get the standard errors of the MLE estimates.

For more information about the EM algorithm, see Clayton and Cuzick (1985b), Cox and Oakes (1984), Dempster, Laird, and Rubin (1977), Louis (1982), Louis, Albert and Heghinian (1978) and Wu (1983).

IMSL Inc. (1987) has subroutines that maximize (minimize) functions with descriptions given above. The subroutine used to maximize the negative log-likelihood function is B2ONF (FUNCTION, NP, GUESS, ITYPE, LOWER, UPPER, SCALE, FSCALE, IPARM, RPARM, THETA, FVALUE, COVAR, ICOVAR).

The output is contained in THETA, FVALUE and COVAR, where THETA is the vector containing the estimates of the parameters. FVALUE is the maximum of the function to be maximized. COVAR contains the approximation to the Hessian matrix (which is the matrix of the second derivatives of the function to be maximized with respect to the parameters, Bard [1974]). This subroutine (B2ONF) uses a quasi Newton method and an active set strategy to solve minimization (maximization) problems subject to simple bounds on the parameters.

According to the IMSL documentation, for a given starting point, an active set, which contains the indices of the parameters at their bounds, is built. The routine then computes the search direction according to the formula $d = -B^{-1}g^c$, where B is a positive definite approximation of the Hessian matrix, and g^c is the gradient evaluated at the starting point. The search direction for the variables (parameters) in the active set is set to zero. A line search is used to find a new point. Finally, the routine checks the optimality conditions (internal). The gradient

is estimated by using a finite difference method. Other IMSL subroutines are used, e.g., (LINRG) to invert the Hessian matrix and (SCOPY) to print the inverse of the Hessian matrix. (QDAG) and (QDAGI) are used for numerical integration. For more information on the above method (see Dennis & Schnabel, 1983; Gill & Murray, 1976; IMSL Inc., 1987).

Hypothesis Testing

The parameter vector we are estimating is $\theta = (\gamma, \underline{\alpha}, \underline{\beta}, \underline{\mu}_1, \underline{\mu}_2)$, where γ is a scalar parameter that measures the association between X and Y. $\underline{\mu}_1, \underline{\mu}_2$ are unknown baseline hazards associated with X and Y respectively, with $\dim(\underline{\mu}_1) = M$, $\dim(\underline{\mu}_2) = N$. $\underline{\alpha}$ and $\underline{\beta}$ are regression coefficients associated with X and Y respectively, with $\dim(\underline{\alpha}) = \dim(\underline{\beta}) = p = \text{number of covariates}$.

One of our objectives in this study is to test the hypothesis concerning the effect of the covariate (SEX) on CHD morbidity and CHD mortality, i.e., we want to test the following hypotheses

$$H_{0\alpha} : \alpha = 0 \quad \text{vs} \quad H_{1\alpha} : \alpha \neq 0,$$

$$H_{0\beta} : \beta = 0 \quad \text{vs} \quad H_{1\beta} : \beta \neq 0.$$

This means that we want to test if the covariate (SEX) has an effect on developing CHD or not; also we want to test if the covariate (SEX) has an effect on dying of CHD or not.

The standard errors of the estimates are obtained from

the inverse of information matrix, $I^{-1}(\hat{\theta})$, ($\hat{\theta}$ is the MLE of θ) which is the inverted Hessian matrix obtained from using the IMSL subroutine (B2ONF), as mentioned in the previous section. From the properties of the MLE estimates, the test statistic for each estimate θ_i with standard error $SE(\theta_i)$ is

$$Z_{\theta_i} = \frac{\theta_i}{SE(\theta_i)}.$$

This test statistic can be used to test the effect of the covariate (SEX) on CHD morbidity and CHD mortality, and can be used also to obtain interval estimation of θ_i based on the asymptotic normality of θ_i , i.e., under the null hypothesis that $\theta_i = 0$, θ_i will be normally distributed with mean zero and standard deviation $SE(\theta_i)$.

Analysis and Discussion

As mentioned earlier, we consider one covariate (SEX), and we will fit two models and compare the preclinical incidence rates for those two models.

In Model 1, we fit our model with one covariate (SEX); in this case we will have 13 parameters to estimate. After estimating those parameters, we will calculate the estimated preclinical incidence rates and their standard

errors in terms of our model which are obtained for men and women in Chapter V. Also, we will test if (SEX) has an effect on developing coronary heart disease (CHD), and test also if (SEX) has an effect on dying of CHD.

In Model 2, we fit our model with no covariates for men and women separately; in this case we will have 11 parameters to estimate. We will compare the incidence rates for Model 1 and Model 2, and will compare the incidence rates for men and women within each model. After estimating those parameters, we will compute the estimated preclinical incidence rates for men and women separately (see Tables 9 and 10 below). The numbers in brackets are the estimated incidence in each interval calculated by multiplying the sample size with the probability of incidence in that interval.

Recall that our objective is to test if the covariate (SEX) has an effect on developing CHD or not, and if the covariate (SEX) has an effect on dying of CHD or not, i.e.,

$$H_{0\alpha} : \alpha = 0 \text{ vs } H_{1\alpha} \neq 0,$$

$$H_{0\beta} : \beta = 0 \text{ vs } H_{1\beta} \neq 0.$$

The MLE estimate of the regression coefficient associated with CHD morbidity (α) is -0.8048 with an estimated standard error of 0.14221, and the MLE estimate of the regression coefficient associated with death of CHD (β) is -0.69196 with an estimated standard error of 0.1531. From the asymptotic normality of the MLE, we get

$$\hat{\alpha} \approx N(\alpha, \sigma_{\alpha}^2) , \hat{\beta} \approx N(\beta, \sigma_{\beta}^2) .$$

Therefore, the test statistic is the z-test, and the p-value for testing if alpha is zero or not is 0.0 and the p-value for testing if beta is zero or not is 0.0 also. This means that alpha and beta are statistically significant at 5% level of significance, i.e., the covariate (SEX) has a significant effect on developing CHD, and (SEX) has a significant effect on dying of CHD. This means that there is a significant difference between men and women CHD morbidity and CHD mortality, and women have a smaller risk than men (see Table 4).

Based on the asymptotic normality of the maximum likelihood estimates, a 95% confidence interval for alpha is (-1.0835, -0.526088) and a 95% confidence interval for beta is (-0.9921, -0.391884).

Comparing the estimated incidence values in Table 8 (page 66) for men and women in each interval, we see a difference between the incidence in men and the incidence in women which supports the above conclusion.

Table 4 (page 61) summarizes the overall clinical incidence and its standard error for men and women obtained from the formulas in Chapter V.

Comparing the estimated incidence values for men in Table 9 (page 67) and the estimated incidence values for women in Table 10 (page 67) when we fit our model (no covariates) for men and women separately, we see a dif-

ference between them. To see if this difference is statistically significant or not, we can test if the mens' incidence is different than the womens' incidence or not (since we have large sample sizes). The overall incidence for men obtained from the analysis is 0.010553 with a standard error 0.00536656, and the overall incidence for women is 0.005988 with a standard error 0.003271. From the asymptotic normality of the overall incidence obtained in Chapter V, the test statistic for comparing mens' and womens' incidence has a p-value of 0.0. This indicates that there is a significant difference between men's and women's incidence at 5% level of significance. This result supports the previous conclusions also.

Table 4

Overall Preclinical Incidence and Its Standard Error
for Men and Women for the Two Models

Sex	Our Model(w. cov.)	Our Model(no cov.)
Men	.0111142 (.02449)	.010553 (.00536)
Women	.00557 (.01732)	.005988 (.00327)

We can obtain an estimate for the preclinical incidence of men (women) among those who aged a , by substituting the MLE estimates in the formula

$$I_{PC}(a) = \hat{\mu}_{11} e^{az} e^{-\hat{\lambda}_{11}(a)}, \quad a \in I_1,$$

where

$$\hat{\Lambda}_{1i}(a) = \left(\sum_{r=1}^{i-1} \hat{\mu}_{1r} d_r + \hat{\mu}_{1i} [a - a_i] \right) e^{az}.$$

If $z = 0$, then the preclinical incidence of men among those who are aged a , is

$$I_{PC}(a) = \hat{\mu}_{1i} \exp \left(- \left(\sum_{r=1}^{i-1} \hat{\mu}_{1r} d_r + \hat{\mu}_{1i} [a - a_i] \right) \right), \quad a \in I_i.$$

If $z = 1$, then the preclinical incidence of women among those who are aged a , is

$$I_{PC}(a) = \hat{\mu}_{1i} e^a \exp \left[- \left(\sum_{r=1}^{i-1} \hat{\mu}_{1r} d_r + \hat{\mu}_{1i} [a - a_i] \right) e^a \right], \quad a \in I_i.$$

For example, the preclinical incidence of men among those who are aged 48 is approximately 0.0057096 with standard error of 0.01 and the preclinical incidence of women among those who are aged 48 is approximately 0.0025779 with a standard error of 0.0131. The standard errors of the preclinical incidence are obtained from Equations (11)-(13) of Chapter V. From the asymptotic normality of the preclinical incidence obtained in Chapter V, the test statistic for comparing men's and women's incidence among those who are aged 48 has a p-value of 0.0. This indicates that there is a significant difference between men's and women's incidence among those who are aged 48 at 5% level of significance and women have smaller incidence than men.

As we mentioned earlier, we can obtain confidence

intervals for the overall preclinical incidence, by obtaining the mean and variance of the overall preclinical incidence from the formulas in Chapter V based on the asymptotic normality of the overall preclinical incidence (see Tables 4 and 5). To see how we obtained the overall incidence for women and men when we fit our model, first, the MLE estimates are

$$\hat{\alpha} = -0.8048, \hat{\beta} = -0.69196, \hat{\mu}_{11} = 0.00581, \hat{\mu}_{12} = 0.01764, \\ \hat{\mu}_{13} = 0.2236, \hat{\mu}_{14} = 0.02342, \hat{\mu}_{15} = 0.02441.$$

Since we are using 5 intervals, with $a_1 = 45$, $a_2 = 52$, $a_3 = 59$, $a_4 = 66$, $a_5 = 73$ and $a_6 = \infty$, the length of those intervals are $d_1 = d_2 = d_3 = d_4 = 7$ and $d_5 = \infty$. We substitute those estimates in Equation (14) of Chapter V to get the overall preclinical estimate which simplifies to

$$\hat{I}_{PC} = \frac{[1 - e^{-\Lambda_{11}(a_5)}]}{Md_1}.$$

But note that we substitute $z = 1$ for women and $z = 0$ for men. For example,

$$e^{-\Lambda_{11}(45)} = 1, e^{\Lambda_{12}(52)} = 0.98198, e^{\Lambda_{13}(59)} = 0.926176, \\ e^{\Lambda_{14}(66)} = 0.8664, e^{\Lambda_{15}(73)} = 0.805166, e^{\Lambda_{16}(\infty)} = 0.0;$$

those values are for women, and the similar values for men are

$$e^{-\Lambda_{11}(45)} = 1, e^{\Lambda_{12}(52)} = 0.959, e^{\Lambda_{13}(59)} = 0.846, \\ e^{\Lambda_{14}(66)} = 0.722, e^{\Lambda_{15}(73)} = 0.611, e^{\Lambda_{16}(\infty)} = 0.0.$$

Therefore $I_{PC}(\text{men}) = 0.0111142$ and $I_{PC}(\text{women}) = 0.00557$.

To obtain the standard errors for the overall preclinical incidence for men and women, we use equations (15)-(18) of Chapter V. We need the derivatives of the overall preclinical incidence with respect to the parameters for women evaluated at the MLE estimates of those parameters, i.e.,

$$\frac{dI_{PC}}{d\theta} \big|_{\theta=\hat{\theta}} = (0, .00498, 0, .072, .072, .072, .072, .072, 0, 0, 0, 0, 0).$$

Similarly, the derivatives of the overall preclinical incidence with respect to the parameters for men are evaluated at the MLE estimates of those parameters, i.e.,

$$\frac{dI_{PC}}{d\theta} \big|_{\theta=\hat{\theta}} = (0, 0, 0, .1222, .1222, .1222, .1222, .1222, 0, 0, 0, 0, 0).$$

Using equation (18) of Chapter V and the computer, we get the standard error for women's incidence to be 0.01732 and the standard error for men's incidence to be 0.02449.

Tables 6 (page 65) and 7 (page 66) summarize the MLE estimates for the baseline hazards associated with X and T for men and women for the two models.

Table 8 (page 66) summarizes the estimated preclinical incidence for men and women when we fit our model with covariate (SEX). The numbers in brackets in Tables 8, 9 and 10 are the estimated counts of the preclinical incidence in each interval for men and women when we fit the two models. These numbers are obtained by multiplying the

preclinical incidence in each interval with the sample size. Tables 9 and 10 (page 67) summarize the estimated preclinical incidence for men and women respectively, when we fit our model with no covariates.

Table 5

95% C.I. for the Overall Preclinical Incidence
for Men and Women for the Two Models

Sex	Our Model(w. cov.)	Our Model(no cov.)
Men	(-.036896, .05912)	(-.00004, .02107)
Women	(-.021750, .03952)	(.000423, .012399)

Table 6

MLE Estimates of Baseline Hazards Associated
With CHD Morbidity for the Two Models

B.L.H.	Our Model w. Cov.	Our Model (Men) no Covariates	O. Model Women no Covariates
Mu11	0.00581	0.00691	0.00236
Mu12	0.01764	0.02131	0.00513
Mu13	0.02236	0.01876	0.01262
Mu14	0.02342	0.01960	0.01350
Mu15	0.02441	0.01912	0.01439

From the above results and according to the data obtained, we can conclude that CHD incidence among men is higher than CHD incidence among women.

Table 7

MLE Estimates of Baseline Hazards Associated
With CHD Mortality for the Two Models

B.L.H.	Our Model w. Cov.	Our Model (Men) no Covariates	O. Model Women no Covariates
Mu21	0.13391	0.12832	0.04979
Mu22	0.02585	0.02771	0.00654
Mu23	0.02881	0.03098	0.00680
Mu24	0.01566	0.02035	0.00000
Mu25	0.00000	0.00000	0.00000

Table 8

Estimated Preclinical Incidence When Fitting
Our Model With Covariate (SEX)

age groups	preclinical incidence(men)	preclinical incidence(women)
[45 - 52)	0.041 (25.174)	.01802 (12.397)
[52 - 59)	0.133 (69.382)	.00558 (38.337)
[59 - 66)	0.124 (76.136)	.05976 (41.055)
[66 - 73)	0.111 (68.154)	.06124 (42.075)
73 or more	0.611 (375.15)	.80517 (553.14)
Total	(614)	(687)

Our model can accommodate time-dependent covariates,
where the hazard functions in the i^{th} and j^{th} intervals for
the k^{th} individual will be of the form

$$\mu_{1i} e^{(\alpha' z_k + \eta_1' Y_{ik})} , \quad \mu_{2j} e^{(\beta' z_k + \eta_2' Y_{jk})} ,$$

Table 9

Estimated Preclinical Incidence When Fitting
Our Model (no cov.) for Men Only

age groups	preclinical incidence (men)	
[45 - 52)	0.04241 (26.04)	
[52 - 59)	0.13300 (81.166)	
[59 - 66)	0.10152 (62.333)	
[66 - 73)	0.09274 (56.940)	
73 or more	0.63064 (387.21)	
Total	(614)	

Table 10

Estimated Preclinical Incidence When Fitting
Our Model (no cov.) for Women Only

age groups	preclinical incidence (women)	
[45 - 52)	0.0164 (11.267)	
[52 - 59)	0.0347 (23.834)	
[59 - 66)	0.0802 (55.097)	
[66 - 73)	0.0783 (53.792)	
73 or more	0.7904 (543.01)	
Total	(687)	

where \underline{Y}_{1k} and \underline{Y}_{jk} are the vectors of time-dependent covariates associated with X and Y respectively, and η_1 and η_2 are the coefficients associated with \underline{Y}_{1k} and \underline{Y}_{jk} for X and Y. In this case, if we can add time-dependent covariates to the model, then we will be able to eliminate the Louis et al.

assumption of no cohort effect.

A possible future work will be applying this work to AIDS research by modifying the assumptions of the model.

Another possible future work is to include a fourth disease state, i.e., a disease free state, preclinical state, clinical state and a death state, using a parametric approach. For example, in cancer studies, X is the time for tumor onset (patient enters the PCS), T is the time when symptoms surface (patient enters the CS), and D is the time when the patient dies of cancer (patient enters the death state).

BIBLIOGRAPHY

- Albert, A., Gertman, P.M., & Louis, T.A. (1978). Screening for the early detection of cancer I. The temporal natural history of a progressive disease state. Mathematical Biosciences, 40, 1-59.
- Albert, A., Gertman, P.M., Louis, T.A. & Liu, S. (1978). Screening for the early detection of cancer II. The impact of screening on natural history of the disease. Mathematical Biosciences, 40, 61-109.
- Armitage, P., & Berry, G. (1987). Statistical methods in medical research. Boston, MA: Blackwell Scientific Publications.
- Bard, Y. (1974). Nonlinear parameter estimation. New York: Academic Press.
- Beck, G. (1979). Stochastic models with competing risks and covariates. Biometrics, 35, 427-438.
- Bemis, B., Bain, L., & Higgins, J. (1972). Estimation and hypothesis testing for parameters of a bivariate exponential distribution. Journal of the American Statistical Association, 67, 927-929.
- Bishop, M., Feinberg, S., & Holland, P. (1975). Discrete multivariate analysis. Cambridge, MA: MIT Press.
- Breslow, N., & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics, 2, 437-453.
- Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. Biometrika, 68, 417-422.
- Chiang, Y.K., Hardy, R.J., Hawkins, C.M., & Kapadia, A.S. (1989). An illness-death process with time-dependent covariates. Biometrics, 45, 669-681.
- Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65, 141-151.

- Clayton, D.G. (1982). The analysis of prospective studies of disease aetiology. Communications in Statistics, 11, 2129-2155.
- Clayton, D., & Cuzick, J. (1985a). The EM algorithm for Cox's data. Biometrika, 69, 351-364.
- Clayton, D.G., & Cuzick, J. (1985b). Multivariate generalizations of proportional hazards model. Journal of the Royal Statistical Society Series A, 82-117.
- Cox, D.R. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society Series B, 34, 187-220.
- Cox, D.R., & Hinkley, D.V. (1974). Theoretical statistics. London: Chapman and Hall.
- Cox, D.R., & Oakes, D. (1984). Analysis of survival data. New York: Chapman and Hall.
- Cuzick, J. (1982). Rank tests for association with right-censored data. Biometrika, 69, 351-364.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B, 39, 1-38.
- Dennis, J.E., & Schnabel, R. (1983). Numerical methods for unconstrained optimization and nonlinear equations. Englewood Cliffs, NJ: Prentice Hall.
- Efron, B. (1967). The two sample problem with censored data. In L. Lecam & J. Neyman (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 4, (pp. 831-853). Berkeley, CA: University of California Press.
- Elandt-Johnson, R.C., & Johnson, N.L. (1980). Survival models and data analysis. New York: Wiley.
- Farlie, D. (1960). The performance of some correlation coefficients for a general bivariate distribution. Biometrika, 47, 307-323.
- Feigel, P., & Zelen, M. (1965). Exponential survival probabilities. Biometrics, 21, 826-838.
- Fidler, H., Boyes, D., & Worth, A. (1968). Cervical cancer in British Columbia. Journal of Obstetrics and Gynaecology in British Commonwealth, 75, 392-404.

- Fruend, J. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association, 56, 971-977.
- Gill, P.E., & Murray, W. (1976). Minimization subject to bounds on the variables. NPL Report NAC, 72, National Physical Laboratory. London, England.
- Glasser, M. (1967). Exponential survival with covariance. Journal of the American Statistical Association, 62, 561-568.
- Gordon, T., & Shurtleff, D. (1973). The Framingham Study. NIH Publication 74-478, section 29. Washington, DC: Department of Public Health, Education and Welfare.
- Gumbel, E. (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55, 698-707.
- Hawkes, A. (1972). A bivariate exponential distribution with application in reliability. Journal of the Royal Statistical Society Series B, 34, 129-131.
- Holford, T.R. (1976). Life-tables with concomitant information. Biometrics, 32, 587-598.
- Hougaard, P. (1986). A class of multivariate failure time distributions. Biometrika, 73, 671-678.
- Hougaard, P. (1987). Modeling multivariate survival. Scandinavian Journal of Statistics, 14, 291-304.
- IMSL, INC. (1987). FORTTRAN subroutines for mathematical applications: Math/Library. Houston, TX: Author.
- Kahn, H.A., & Sempos, C.T. (1989). Statistical methods in epidemiology. New York: Oxford University Press.
- Kalbfleisch, J.D., & Prentice, R.L. (1980). The statistical analysis of failure time data. New York: Wiley.
- Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Society, 53, 457-481.
- Lawless, J.F. (1982). Statistical models and methods for lifetime data. New York: Wiley.
- Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. Journal of the

Royal Statistical Society Series B, 44, 226-233.

- Louis, T.A., Albert, A., & Heghinian, S. (1978). Screening for the early detection of cancer III. Estimation of disease natural history. Mathematical Biosciences, 40, 111-144.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports, 50, 163-170.
- Marshall, A., & Olkin, I. (1967). A multivariate exponential distribution. Journal of the American Statistical Association, 62, 30-44.
- Miller, R. (1976). Least squares regression with censored data. Biometrika, 63, 449-463.
- Miller, R. (1981). Survival analysis. New York: Wiley.
- Nash, J.C., & Walker-Smith, M. (1987). Non linear parametric estimation. New York: Marcel Dekker Inc.
- Oakes, D. (1982). A model for association in bivariate survival data. Journal of the Royal Statistical Society Series B, 44, 412-422.
- O'Neill, T. (1985). Testing for symmetry and independence in a bivariate exponential distribution. Statistics and Probability Letters, 3, 269-274.
- Pettersen, J. (1986). Fitting parametric survival models with time-dependent covariates. Applied Statistics, 35(3), 281-288.
- Shurtleff, D. (1974). The Framingham Study. NIH Publication 74-599, section 30. Washington, DC: Department of Public Health, Education and Welfare.
- Wu, C.F.J. (1983). On the convergence properties of the EM algorithm. Annals of Statistics, 11, 95-103.
- Zelen, M. (1974). Problems in cell kinetics and the early detection of disease. Reliability and Biometry S.I.A.M., 3, 701-726.
- Zelen, M., & Feinleib, M. (1969). On the theory of screening for chronic diseases. Biometrika, 56, 601-614.
- Zipf, C., & Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the

estimation of an exponential survival parameter.
Biometrics, 22, 665-672.