Annu. Rev. Public Health 1997. 18:83–104

CENSORING ISSUES IN SURVIVAL ANALYSIS

Kwan-Moon Leung[1,2], *Robert M. Elashoff*[1], *and Abdelmonem A. Afifi*[1]

[1]Department of Biostatistics, School of Public Health, University of California, Los Angeles, California 90095: e-mail, moon@ucla.edu; [2]Quality Initiatives Division, Health Net, Woodland Hills, California 91367

KEY WORDS: survival analysis, right censoring, interval censoring, informative censoring, ignorability

ABSTRACT

A key characteristic that distinguishes survival analysis from other areas in statistics is that survival data are usually censored. Censoring occurs when incomplete information is available about the survival time of some individuals. We define censoring through some practical examples extracted from the literature in various fields of public health. With few exceptions, the censoring mechanisms in most observational studies are unknown and hence it is necessary to make assumptions about censoring when the common statistical methods are used to analyze censored data. In addition, we present situations in which censoring mechanisms can be ignored. The effects of the censoring assumptions are demonstrated through actual studies.

INTRODUCTION

Survival analysis is used in various fields for analyzing data involving the duration between two events, or more generally the times of transition among several states or conditions. It is also known as *lifetime data analysis, reliability analysis, time to event analysis*, and *event history analysis* depending on the type of application. In this paper, the term *survival time* is used interchangeably with the terms *risk period, lifetime, failure time*, and *time to a certain event*. To determine the survival time, we need to define two time points: the time of origin, i.e. the time at which an original event, such as birth, occurs and the time of failure, i.e. the time at which the final event, such as death, occurs. A

0163-7525/97/0510-0083$08.00

subject is said to be at risk if the original event has occurred, but the final event has not.

A key characteristic that distinguishes survival analysis from other areas in statistics is that survival data are usually censored or incomplete in some way. We define censoring through some practical examples, then describe the common statistical methods used to analyze censored data and discuss the necessity of making assumptions about censoring when those methods are used. We also discuss situations in which the censoring mechanism can be ignored, and investigate the effects of different censoring assumptions in actual studies. Finally, we indicate some trends in the future of research on censoring.

ORIGINS OF CENSORING

Censoring occurs when incomplete information is available about the survival time of some individuals. In this section, we present a number of concrete examples extracted from the literature in various fields of public health. The purpose is to use real-life situations to illustrate types of censoring and to motivate the discussion presented in the later sections.

Examples

Example 1. Health insurance and mortality To examine the relationship between the status of insurance and the risk of subsequent mortality, adults older than 25 years who reported that they were uninsured or privately insured in the first National Health and Nutrition Examination Survey (NHANES I) (9) were followed prospectively from initial interviews between 1971 and 1975 until 1987 (end of the Epidemiologic Follow-up Study, NHEFS). There were a total of 4882 eligible subjects, of whom 669 subjects were uninsured.

The time period of interest was the time from the start of follow-up to death, and the research question is whether this variable is affected by whether or not the individual is insured. In general, an observation is said to be *right censored* if the person was alive at study termination or was lost to follow-up at any time during the study. By right censoring, it is meant that the survival time is only known to exceed a certain value. In this study, the analysis was adjusted for other factors such as baseline age, gender, race, smoking status, alcohol consumption, obesity, self-rated health, employment status, and so forth.

Example 2. Do men and women relapse into alcoholism for different reasons? A sample of 44 women and 50 men attending an alcohol treatment facility operated by the Western Australian Alcohol and Drug Authority were studied (29). A range of demographic, social, and psychological measures were observed to determine whether women and men relapse for different reasons. The length of follow-up was three months and the variable of interest was the time from

start of follow-up to relapse. After 3 months, 25 subjects were lost to follow-up (6 women and 19 men), and 42 subjects (22 women and 20 men) had relapsed. Here again, right-censoring occurred.

Note that the numbers of loss to follow-up are quite different between women and men. These unequal censoring rates can cause the analysis to lose power when assessing gender effect (see, for example, Reference 15).

Example 3. Survival with inoperable lung cancer A sample of 61 patients with inoperable lung cancer who were treated with the drug cyclophosphamide at the Eastern Cooperative Oncology Group were studied (22). The primary outcome variable was the time from diagnosis to death. In this sample, 28 right-censored observations represented patients whose treatment was terminated because of the appearance of metastatic disease or because of a significant increase in the size of their primary lesion. Since such disease progressions are usually associated with shortened residual survival time, the censoring may be *informative* (that is, censoring provides more information than the fact that survival time exceeded a certain time) and may cause serious problems if analysis of such data does not take this information into account.

Example 4. Long-term life-style intervention with high-normal blood pressure levels The study sample comprised 2382 participants (1566 men and 816 women), who were 110% to 165% of desirable body weight (11). Individuals were allocated at random to four treatment arms using a 2×2 factorial design (weight loss $\times$ Sodium reduction). This was a phase II study of the Trials of Hypertension Prevention (TOHP).

The primary outcome variable was the time from randomization to either diastolic blood pressure (DBP) becoming $\geq$90 mm Hg, systolic blood pressure (SBP) becoming $>$140 mm Hg, or taking antihypertensive medication. An observation was considered to be right censored if the subject's blood pressure was within normal limits, that is, had DBP $\leq$90 mm Hg and SBP $\leq$140 mm Hg, and did not take any antihypertensive medication by the end of the study, or if the subject was lost to follow-up. Note that had we not considered taking antihypertensive medication as an endpoint, then the observation would have been right censored at the time of taking antihypertensive medication. However, this right-censored mechanism would be informative since subjects are likely to start taking antihypertensive medication because their blood pressure exceeded the normal limits.

Example 5. Longitudinal study of respiratory symptoms in aluminum potroom workers A prospective study of respiratory health in aluminum potroom workers was initiated in the Nordic countries on January 1, 1986 (20, 28). The workers were followed until December 31, 1989, or until leaving potroom

work, whichever came first. The workers were supposed to have health examinations and to fill out questionnaires regarding respiratory symptoms at the start of employment, then yearly at a routine examination, if attending the plant's health clinic because of respiratory symptoms, or when leaving employment.

The study sample consisted of 1301 subjects who were employed during the study period and had at least two examinations. If the subjects reported wheezing and dyspnea, they were considered symptomatic. The time variable of interest was the time from employment to development of symptoms. Besides fluoride exposure, other potential covariates included age, smoking habits, and previous exposure to dust/gases.

In addition to right censoring, that is, leaving the potroom or ending the survey without respiratory symptoms, some observations were *singly interval censored* because for them the study endpoint was established only by periodical examinations. By singly interval censoring it is meant that the outcome variable is not known exactly, rather it is known only up to a time interval.

Example 6. Multicenter AIDS Cohort Study (MACS) From April 1984 to September 1993, there were 4954 men between the ages of 18 and 70 who were recruited for the Multicenter AIDS Cohort Study (MACS) (5). The MACS is a longitudinal study of the natural history of human immunodeficiency virus type 1 (HIV-1) among homosexual and bisexual men. Subjects were recruited at four centers: Los Angeles, Chicago, Pittsburgh, and Baltimore.

An important time variable was the incubation period of AIDS (time from HIV-1 seroconversion to an AIDS-defining illness). An observation was right censored if the subject was AIDS free on September 1, 1993, or was lost to follow-up. The censoring issue becomes more complicated when we realize that both the time of HIV seroconversion and the time of AIDS onset are known only up to a time interval since those times are determined by periodical examinations. Such observations are called *doubly interval censored*, i.e. the survival time (incubation period of AIDS) is subject to interval censoring on the left and on the right.

Example 7. Survival with malignant melanoma In the period between 1971 and 1993, approximately 6000 patients with malignant melanoma were treated by the staffs of the John Wayne Cancer Institute (JWCI) (24). The primary objective of the study reported here was to examine the efficacy of a new polyvalent melanoma cell vaccine (MCV) in treating patients with metastatic disease. Such treated patients represent a subset of the JWCI patients. To provide appropriate treatments, patients were followed periodically after admission to detect any change in disease stages. Excluding patients with stage III disease when first seen at the JWCI, we had 1548 patients in the data set. By the time of analysis, 890 had advanced to stage III, 788 of whom died. Beside the indicator

of treatment (MCV or no MCV), other covariates include gender, distant site, Breslow's depth of patient's primary tumor, and the time interval between the first diagnosed stage II disease and disease metastasis.

Here we note that, by the time a stage III was diagnosed (regional bone recurrence or metastatic disease), metastasis had already occurred. Thus, the date of the first metastasis is only known to lie between the dates of the last diagnosed stage II disease (regional recurrence without involving bone) and the first diagnosed stage III disease. Such observations were thus *left interval censored.*

Types of Censoring Mechanisms

POINT CENSORING We now summarize the different types of censoring mechanisms illustrated in the above examples. We begin with observations that may be point-censored as depicted in Figure 1, where the subjects are under observation from time T_0 to time T_1, and the survival time or censoring time are known exactly. Such situations arise when subjects are continuously monitored or when the time of event is well documented, for example, when the subjects are hospitalized or when the event is death. We call censoring arising from these situations point censoring to distinguish it from interval censoring, to be discussed shortly, in which the survival time is known only up to a time interval. The solid line represents the risk period for each subject. The line ending with an asterisk ($*$) indicates an occurrence of the event of interest, and the line ending with an open point (o) indicates an occurrence of an event other than the event of interest, e.g. loss to follow-up or death due to causes other than the one under study.

In Figure 1, the entire risk period for Subject A falls within the observation period and the time of occurrence of the event is known; hence there is no censoring for this observation. For subject B, the risk period starts during the observation period and the event occurs after follow-up is terminated at T_1.

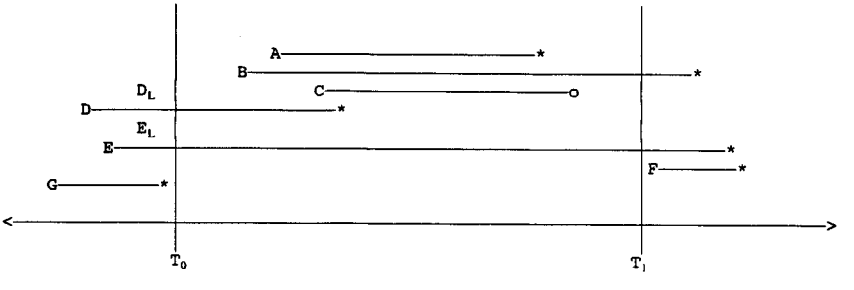

Figure 1 Types of point-censored observations.

The observation of Subject B is therefore right censored at T_1. This type of censoring occurred in all the above examples. For Subject C, the observation is also right censored but it is so because an event other than the one of interest occurs during the observation period and takes the subject out of the risk set (the set of subjects who are at risk). To distinguish between these two types of right-censored data, we call censoring due to study termination (Subject B) *end-of-study censoring* and censoring due to other reasons (Subject C) *loss-to-follow-up censoring*. Under the assumption that the time of entry to the study is independent of the risk period, it can be easily shown that end-of-study censoring is independent of survival time, and hence it poses no problem to the analysis. However, it is not always reasonable to make such an independence assumption about loss-to-follow-up censoring, and hence we need to make further assumptions about the censoring mechanism when analyzing the data. Statistical methods for handling censoring mechanisms are discussed in detail in the next section.

While cases B and C in Figure 1 represent right censoring, Subject D represents a case with left truncation. Such a problem could happen, for example, when a subject in the AIDS study was already HIV-1 seropositive prior to enrollment and the time variable of interest is the incubation period of AIDS. There are other situations, such as for Subject E, in which the observation is both left and right censored; we call such observations *doubly censored.* An example of this sort exists in the AIDS study example where a subject was already HIV-1 seropositive when enrolled but was still AIDS-free at the end of the study. A crucial question here is whether the time (D_L) from the beginning of the risk period to the beginning of the observation period is known. If D_L is observed, then one could apply the methodologies developed for no censoring (or right censoring) for analysis with proper adjustment of the risk set. However, if D_L is not observed, then we cannot specify the origin of the survival time. In this case external information such as the distribution, over chronological time, of the original event is needed (14, 17, 32).

Finally, in most applications there are cases where the origin and the event both occur prior to the start of follow-up or after follow-up ends. Such cases generally do not affect the analysis but they can affect the generalizablility of the findings. Subjects F and G in Figure 1 represent such cases known as *completely right censored* and *completely left censored*, respectively.

INTERVAL CENSORING In many applications, for instance, in Examples 5–7, the time of the event may be known only up to a time interval, especially when the time is established by periodical examinations. Figure 2 defines the relevant time points and illustrates how interval-censored data arise.

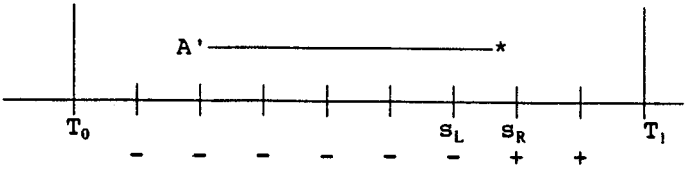

Figure 2 An example of singly-interval censored observations.

For subject A′ in Figure 2 (Subject A in Figure 1 but the time of event is interval censored), instead of knowing that the event of interest occurred at time s, we only know that the event occurred in the time interval $(s_L, s_R]$, i.e. after S_L but up to and possibly at S_R. An example of this kind exists in the AIDS study where subjects were periodically examined to determine the HIV-1 status. In Figure 2, "−" means an HIV-1 negative result was reported and "+" means an HIV-1 positive result was reported. In fact, observations from most studies with a nonlethal outcome are interval censored since we usually cannot monitor subjects continuously. The issue is whether we should analyze the data as interval censored or point censored. For instance, if the median survival time is 5 years and the intervals are about 3 to 6 months wide, then we have no reason to complicate the analysis by considering interval censoring. On the other hand, if the intervals are about 1 year or longer, then we should account for such uncertainty in the analysis. It is also interesting to note that interval censoring theoretically includes both left and right censoring.

Figure 3 extends Figure 2 to represent more complicated but practical situations in which we continue to periodically monitor Subject A′ after an immediate event (denoted by "∗" in Figure 3), for example disease metastasis or HIV-1 seroconversion up to a final event (denoted by "#" in Figure 3), for example death or AIDS onset. The time between the event "∗" and the event "#" is known up to an interval on both ends.

We call such an observation *doubly interval censored* to distinguish it from *singly interval censored* observations. To analyze doubly interval censored data,

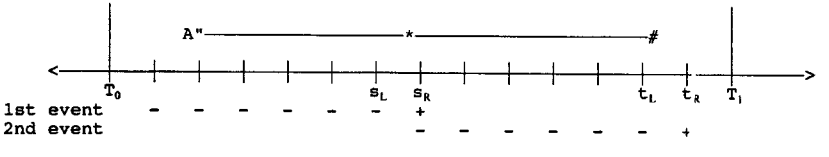

Figure 3 An example of doubly interval censored observations.

it is tempting to transform the observations to the singly interval censoring form, that is, for Subject A″ we create the interval $(t_L - s_R, t_R - s_L]$, and then apply the methods developed for singly interval censored data. However, De Gruttola & Lagakos (4) pointed out that this approach is not valid.

COMMON STATISTICAL METHODS FOR CENSORED DATA

In this section we discuss the common statistical methods used to analyze censored data. As for other types of incomplete data, several approaches have been proposed (see 25 for discussion of statistical methods for incomplete data). As described below, there are four basic approaches for the analysis of censored data: complete data analysis, the imputation approach, analysis with dichotomized data, and the likelihood-based approach.

Complete-data analysis As many researchers and statistical packages do when faced with incomplete data, one can simply ignore the censored observations and analyze only the uncensored complete observations. The main advantage of this approach is simplicity. However, there are some disadvantages. (*a*) Loss of efficiency: The loss in sample size can be considerable since it is not unusual, especially in medical or epidemiological studies, that 50% or more observations are censored. (*b*) Estimation bias: Inferences based on analyzing the uncensored observations only may be biased. It is a common misconception that one need not make any assumptions about the censoring mechanism when performing a complete-data analysis. In reality, such an analysis requires a strong assumption regarding the censoring mechanism: As in the incomplete data situations, complete-data analysis produces unbiased estimates only if the missing (censored observations) are missing (censored) completely at random (25).

Imputation approach Although imputation is one of the popular approaches for handling incomplete data, it may not be appropriate for censored data. In the context of right censoring, there are two extreme ways to impute the missing survival times: (*a*) assuming all censored cases fail right after the time of censoring, that is, left-point imputation or (*b*) assuming all censored cases never fail, that is, right-point imputation. It is clear that neither of these approaches is appropriate since the survival probabilities would be underestimated and overestimated, respectively. Another approach is to assume that the failure time after censoring follows a specific model and estimate the model parameters in order to impute the residual survival time (time from censoring to failure). However, this approach depends on the model assumptions, which are very difficult to check without information on survival after censoring (the missing information).

In the context of interval censoring, the inappropriateness of imputation is less clear. Many researchers use imputation techniques, especially right-point or mid-point imputation, when the observations are interval censored. This may be due to lack of statistical software packages for analysis of interval-censored data. However, many authors (23, 24) have pointed out that both right-point and mid-point imputations may generate seriously biased results, as will be seen in later sections.

Analysis based on dichotomized data The problem of right censoring and interval censoring may be avoided if one analyzes the incidence of occurrence versus nonoccurrence of the event within a fixed period of time and disregards the survival times. In this case, the dichotomized data can be easily analyzed by the standard techniques for binary outcomes, such as contingency tables and logistic regression. However, there are some disadvantages of this approach:

1. It cannot distinguish between loss-to-follow-up and end-of-study censoring.

2. Variability in the timing of the event among those who had the event within the observation period cannot be modeled. Let us consider an extreme example: Suppose we are studying the effect of a new drug on patients who underwent surgery for a particular disease. Eighty percent of the subjects who have the placebo have recurrence shortly after the surgery, while 80% of the subjects who took the new drug remained disease free for at least 5 years but had recurrence within 10 years. All other patients did not have recurrence after 10 years. If we analyze these data with dichotomized outcomes, we may find no difference between treatment the groups when the observation period is 10 years. However, there would likely be a significant difference between the treatment groups when the observation period is 5 years.

3. No time-dependent covariates (such as age, smoking status or alcohol consumption status) can be used in modeling.

However, the approach of analyzing dichotomized data may be acceptable when the risk of failure is low, the risk periods (survival times) are long, and the covariates are associated with preventing the event rather than with prolonging the survival time. Such situations are common in many epidemiologic studies.

Likelihood-based approach Perhaps the most effective approach to censoring problems is to use methods of estimation that adjust for whether or not an individual observation is censored. Many of these approaches can be viewed as maximizing the likelihood under certain model assumptions, including assumptions about the censoring mechanism. Likelihood-based approaches include,

for example, the Kaplan-Meier estimator of the survival function in a one-sample problem, the log-rank test for testing equality of two survival functions in a two-sample problem, and the Cox-regression and accelerated-failure-time models for analysis of time to event data with covariates. The main advantage of the likelihood-based approach is that it utilizes all the information available. However, as in the other approaches, assumptions about the censoring mechanism are still required. In the following sections, we present these assumptions, examine how important they are, and discuss when they can be ignored.

NECESSITY OF MAKING ASSUMPTIONS ABOUT CENSORING

With some exceptions, the censoring mechanisms in most observational studies are unknown. Two such exceptions are the so-called type I and type II censoring designs. Type II designs (often used in engineering) are studies in which a total of n subjects are under observation but, instead of continuing until all subjects fail, the study is terminated when the rth subject fails. In principle, the analysis of type II censoring data poses no problem to the investigator. However, type II censoring designs are rarely used by biomedical or public health researchers and, therefore, we do not discuss them further in this paper.

By a type I censoring design we mean a study in which every subject is under observation for a specified period C_0 or until failure. A slightly more complicated type I censoring design is one in which each subject has his/her own *fixed* censoring time C_i, instead of a common censoring time C_0. In this study design, the likelihood function for each subject can be represented by one of the following two probabilities: the probability that the event occurred in a small interval including time t_i [denoted by $f_i(t_i)$] or the probability that the subject did not have the event at C_i [denoted by $S_i(C_i)$]. Mathematically, the likelihood function of the ith subject can be written as

$$L_i = f_i(t_i)^{\delta_i} S_i(C_i)^{1-\delta_i}, \qquad 1.$$

where δ_i is an indicator of whether the actual survival time of the ith subject is observed or not, that is, $\delta_i = 1$ if observed and $= 0$ if not. Having $\delta_i = 0$ means that the subject survived the observation period C_i without the event occurring, and hence the observation was censored. Note also that S_i is the survival function of interest and f_i is the corresponding density function. The likelihood function of the whole sample is the product of the individual likelihood function of all the subjects. Many statistical procedures dealing with right censoring are based on the likelihood function (Eq. 1).

Unfortunately, in most applications the censoring mechanisms are more complicated than in the type I design, especially when both end-of-study and

loss-to-follow-up censoring are involved. Thus, we usually cannot apply the likelihood function (Eq. 1) directly without making further assumptions about the censoring mechanism. In medial and epidemiological studies, censoring times C_i are often random rather than fixed. For instance, in example 3, patients may enter the study in more or less a random fashion according to their times of diagnosis. If the study is terminated at a preassigned date, then the end-of-study censoring times (times from subjects' entry to study termination) are effectively random. Similarly, the loss-to-follow-up censoring times are also effectively random since the times at which loss to follow-up occurs are not known in advance.

To analyze data of this kind, we may proceed by considering the joint distribution of T (survival time) and C (censoring time), that is, the likelihood function

$$L = \prod_{i=1}^{n} f_i(t_i, c_i), \qquad 2.$$

where $f_i(t_i, c_i)$ is the density function of survival time $T = t_i$ and censoring time $C = c_i$. However, we are often not interested in modeling both the survival time and censoring time. Instead, we are only interested in the distribution of the survival time or the effects of certain covariates on the survival time. Furthermore, for the ith subject, we only observe $y_i = \min\{t_i, c_i\}$ and the censoring indicator δ_i instead of (t_i, c_i). Under these conditions, the distribution function of T is described by statisticians as being nonidentifiable (33) unless we make further assumptions. Here nonidentifiability means that there exist more than one distribution function of T that are compatible with the data.

A simple but commonly used assumption to resolve this problem is *independent censoring*, that is, we assume that the survival time T and the censoring time C are independent. Note that some authors use the term random censoring when they actually mean independent censoring. Under the independent censoring assumption, analyses can be simply based on the likelihood function (Eq. 1).

For left truncation, data can be handled in a similar manner to right censoring (for example, 14, 32). For interval censoring, the situation is slightly different in that it requires the knowledge of the *examination scheme* (prospective study) or *sampling plan* (retrospective study) as explained below.

Suppose the disease process is denoted by $X(t)$ indicating the disease state at time t. We assume that a particular subject is observed at times $t_0 < t_1 < \cdots < t_m$ to be in states $s_0, s_1, \ldots, s_m$, respectively. Here s_j, $j = 1, \ldots, m$ could represent the disease stages. For these observations, the likelihood is given by

$$L = \Pr\{X(t_0) = s_0, \ldots, X(t_m) = s_m;\ T_0 = t_0, \ldots, T_m = t_m;\ M = m\} \qquad 3.$$

Here, not only are the examination times $T_0, T_1, \ldots, T_m$ assumed random, the

number of examinations M is also assumed to be a random variable. However, we generally are not interested in modeling the examination times or their number. Rather, we are usually interested only in the disease states and therefore, for interval-censored data, inference is usually based on

$$L = \Pr\{X(t_0) = s_0, \ldots, X(t_m) = s_m\}. \quad 4.$$

Authors often assume that the examination scheme (or the sampling plan) has been prespecified or has at least been chosen completely independently of the disease process. In this case, analysis can be based on the likelihood function (Eq. 4).

IGNORABILITY OF THE CENSORING MECHANISM

RIGHT CENSORING In the statistical inference procedures included in most statistical packages, the assumptions used in the likelihood function ($*$) or its equivalent are often either not stated explicitly or are stated in vague terms. As a result, it is likely that many users of survival analysis are either unaware that they are making certain assumptions about the censoring mechanism or are unclear about precisely what these assumptions are. We think it is important that such censoring assumptions be clearly understood since only then can we identify those situations in which these assumptions can be ignored.

As Kaplan & Meier (17) noted: "In practice this assumption (independent censoring) deserves special scrutiny." However, the Kaplan-Meier estimator may overestimate the survival function of T if the survival time and the censoring time are positively correlated, and underestimate the survival function if the times are negatively correlated.

Independent censoring may not hold for all situations but under some dependence conditions we can still use the likelihood function (Eq. 1). Lagakos (21) presented two such situations:

(a) Nonprognostic censoring A censored observation at time C_i indicates that the survival time exceeds C_i and carries no prognostic information about subsequent survival times for either the same individual or other individuals.

(b) Noninformative censoring (*also known as constant sum condition*) The instantaneous probability of failure in a small interval about $y = \min\{t, c\}$ given survival to y is unchanged by the additional information that the subject was uncensored up to time y (see also 16).

Lagakos (21) proved that nonprognostic censoring models and independent censoring models are special cases of the noninformative censoring model.

In many situations, censoring can be recognized to be noninformative (for example, end-of-study censoring), and hence standard procedures assuming the

likelihood function (Eq. 1) can be used. However, in other situations, it is not clear whether censoring is noninformative; in fact, it is sometimes clear that censoring is related to the survival times as, for instance, in examples 3 and 4.

In spite of its crucial importance, the noninformative censoring assumption is not possible to test without making additional restrictions, for example, restrictions on the joint distribution of T and C. As mentioned above, the problem is that given only the sample data (y_i, δ_i), the survival function of T is not identifiable.

Several authors have proposed models for informative censoring and tests for noninformative censoring under various conditions. For instance, Moeschberger (26) suggested modeling the joint distribution of (T, C) as in (Eq. 2). Lagakos & Williams (22) introduced a semiparametric model, called the *cone model*, which includes an exponential survival function of T (with parameter λ), an unspecified function $c(y)$ that measures the relative odds of observing a failure at $y = \min(t, c)$ and a scalar parameter $\theta \in [0, 1]$. Here the parameter θ reflects the degree to which censoring affects survival with $\theta = 1$ indicating noninformative. Moeschberger (26) and Lagakos & Williams (22) presented the maximum likelihood estimates and large-sample test for noninformative censoring. In addition, several nonparametric estimates of the survival function were presented by various authors. Peterson (27) suggested an estimate of the bounds for the survival function without assuming any censoring mechanism or parametric distribution of T. Fisher & Kanarek (8) proposed a nonparametric estimate of the survival function under a "stretch/contract" model in which censoring coincides with an event that alters subsequent survival times by a known scale parameter α, that is, they assume $\Pr\{T > t \mid C = c < t\} = \Pr\{T > c + \alpha(t - c) \mid C > c + \alpha(t - c)\}$. Here $\alpha < 1$ means censoring is favorable for survival whereas $\alpha > 1$ means censoring is unfavorable for survival. In particular, $\alpha = 1$ corresponds to noninformative censoring. Slud & Rubinstein (31) and Klein & Moeschberger (19) generalized the Kaplan-Meier estimator with informative censoring situations where certain measures of the dependence of survival and censoring times are known. In (31) they assume knowledge of $\rho(t) = h(t \mid C < t)/h(t \mid C \geq t)$ which measures the relative risk of failure at time T among previously censored subjects as compared with subjects not yet censored, where $h(t \mid \cdot)$ is the conditional hazard function of T. In (19) they assume knowledge of the parameter θ that relates to Kendall's coefficient of concordance τ of (T, C) by $\tau = (\theta - 1)/(\theta + 1)$.

Interval Censoring

Compared with right censoring situations, relatively few articles were devoted to the problem of ignorability of interval censoring. Similar to the definition of noninformative censoring for right-censored data, Gruger et al (10)

defined noninformative examination schemes (or sampling plans) for interval censoring problems. Basically, an examination scheme (including the number of examinations and their times) is called noninformative if the likelihood function given the examination scheme is proportional to the likelihood function obtained when the examination scheme is fixed in advance. They considered the following examination schemes that are often seen in practice and concluded that they satisfy the noninformative examination scheme assumption.

1. Examination at regular intervals Under this scheme, all subjects are examined or observed at preassigned intervals. This situation occurs frequently in medical studies since examination times are often preassigned.

2. Random sampling Under this scheme, all subjects are examined or observed in a more or less random fashion and the examination times are independent of the subjects' disease history.

3. Doctor's care Under this scheme, the next examination time is chosen on the basis of the subject's current observed status. For example, if a patient is in a critical stage, then the time of the next examination will be chosen to be in the very near future.

Under the above examination schemes, Gruger et al (10) showed that the likelihood function (Eq. 4) can be used to obtain an estimate of the survival function (4, 23) or estimates of the regression coefficients of survival times on the covariates (7, 18, 24). Gruger et al (10) considered another situation called *patient self-selection*. In this scheme a patient's examination is initialized when the patient feels unwell and/or when symptoms suggest that the disease is advancing. Alternatively, a patient who feels unwell might refuse to appear for examination because of loss of confidence in the efficacy of the treatment. Under this scheme, the examination times are no longer noninformative and must be taken into account when analyzing the data. That task, however, requires modeling the joint distribution of the disease state and the examination times [that is, the likelihood function (Eq. 3)]. Gruger et al (10) suggested that investigators should plan in advance in order to avoid this difficulty.

Unlike the case of right censoring, there are only a very few articles on testing the assumption of noninformative examination schemes for interval-censored data. Heitjan (12) and Heitjan & Rubin (13) introduced the concept of "coarse" data that have right censoring and interval censoring as special cases. They provided the conditions under which it is appropriate to ignore the stochastic nature of the coarsening (censoring), and called such conditions coarsening at random. However, parametric assumptions are generally required in order to test the coarsening at random conditions.

EFFECTS OF THE CENSORING ASSUMPTIONS

In this section, we analyze the data from Examples 3, 5, and 7 (discussed in the section on Origins of Censoring) to illustrate the importance of censoring assumptions for right censoring, singly interval censoring, and doubly interval censoring, respectively.

Right censoring: Survival with lung cancer (Example 3 continued) The problem and the data were described in the first section. Recall that the observations were censored because the patients experienced metastasic disease or a significant increase in the size of their primary lesion. Such disease progression usually indicates a shortened residual survival time. Thus, we cannot apply the standard procedure (assuming noninformative censoring) to analyze the data. Table 1 lists the survival times and the censoring times of those 61 patients (33 uncensored observations and 28 censored observations). For these 28 patients with originally censored observations, the ultimate survival times were obtained later. Notice that those ultimate survival times, in general, cannot be treated as the actual survival times since those patients were removed from the study at the time of censoring. However, for the purpose of illustration we will treat those times as if they were the actual survival times.

Table 1 Survival times in weeks of 61 patients with inoperable adenocarcinoma of the lung

33 Uncensored observations		28 Censored observations	
0.43	15.71	0.14 (3.00)*	8.71 (20.43)
2.86	18.43	0.14 (12.43)	10.57 (25.00)
3.14	18.57	0.29 (1.14)	11.86 (17.29)
3.14	20.71	0.43 (17.14)	15.57 (21.57)
3.43	29.14	0.57 (4.43)	16.57 (45.00)
3.43	40.57	0.57 (5.43)	17.29 (24.14)
3.71	48.57	1.86 (12.14)	18.71 (29.43)
3.86	49.43	3.00 (7.86)	21.29 (26.71)
6.14	53.86	3.00 (13.86)	23.86 (29.00)
6.86	61.86	3.29 (10.57)	26.00 (53.86)
9.00	66.57	3.29 (34.43)	27.57 (49.71)
9.43	68.71	6.00 (7.86)	32.14 (63.86)
10.71	68.96	6.14 (9.29)	33.14 (99.00)
10.86	72.86		47.29 (48.71)
11.14	72.86		
13.00			
14.43			

*In parentheses are the eventual failure times of the 28 censored patients.

Lagakos & Williams (22) obtained the maximum likelihood estimates (standard errors) of the cone model discussed in the last section as $\hat{\lambda} = 0.0409$ (0.0123) and $\hat{\theta} = 0.25$ (0.36). Notice that $\hat{\theta}$ is significantly different from one (recall that $\theta = 1$ means censoring is noninformative) based on a large-sample test of H_0: $\theta = 1$, and hence the noninformative censoring assumption is not satisfied. We also obtained the estimates of the upper and the lower bounds of the survival function based on Peterson's procedure (27), and the estimated survival function based on the procedures proposed by Fisher & Kanarek (8), Slud & Rubinstein (31), and Klein & Moeschberger (19) with various model parameters. Figure 4 displays the estimates of the survival function together with the empirical distribution function derived from all 61 complete observations. Each of the three models that account for nonignorable censoring involves some parameters whose values must be guessed by the investigator. In our analysis, we attempted a number of such values, but, in order to save space, we only present the estimates of the survival function that are closest to the empirical distribution function. In practice, however, we can only estimate the bounds of the survival function with a range of model parameters. Figure 4 suggests the following conclusions. First, the Kaplan-Meier estimate (based on the noninformative censoring assumption) overestimates the "actual" survival functions. Second, Peterson's bounds of the survival function are too wide to be useful. In fact, we can use the Kaplan-Meier estimate as an upper bound of the survival function. Third, the estimates based on Lagakos & Williams (22), Fisher & Kanarek (8), Slud & Rubinstein (31), and Klein & Moeschberger (19) agree quite well overall, and they agree with the empirical distribution function well through about 36 months. After 36 months, only the Slud-Rubinstein's estimate continues to agree well with the empirical distribution function. This situation may suggest that the conditions required by the other models are not satisfied beyond 36 months. Our general conclusion from this example is that assuming ignorable censoring can lead to a biased (in this case an over-optimistic) estimate of the survival function. Furthermore, most of the methods that account for noninformative censoring produce reasonable estimates of the survival functions.

Singly interval censoring: Respiratory symptoms in aluminum potroom workers (Example 5 continued) Recall from example 5 that the disease status of the aluminum workers can only be determined at the time of the health examinations, and hence the time at which a symptom first occurs is only known in the time interval between the last examination without a symptom and the first examination with a symptom. To illustrate the consequence of falsely assuming that the occurrence of symptoms coincided with reporting, in which case the data can be analyzed as right censored, Samuelsen & Kongerud (28) fitted

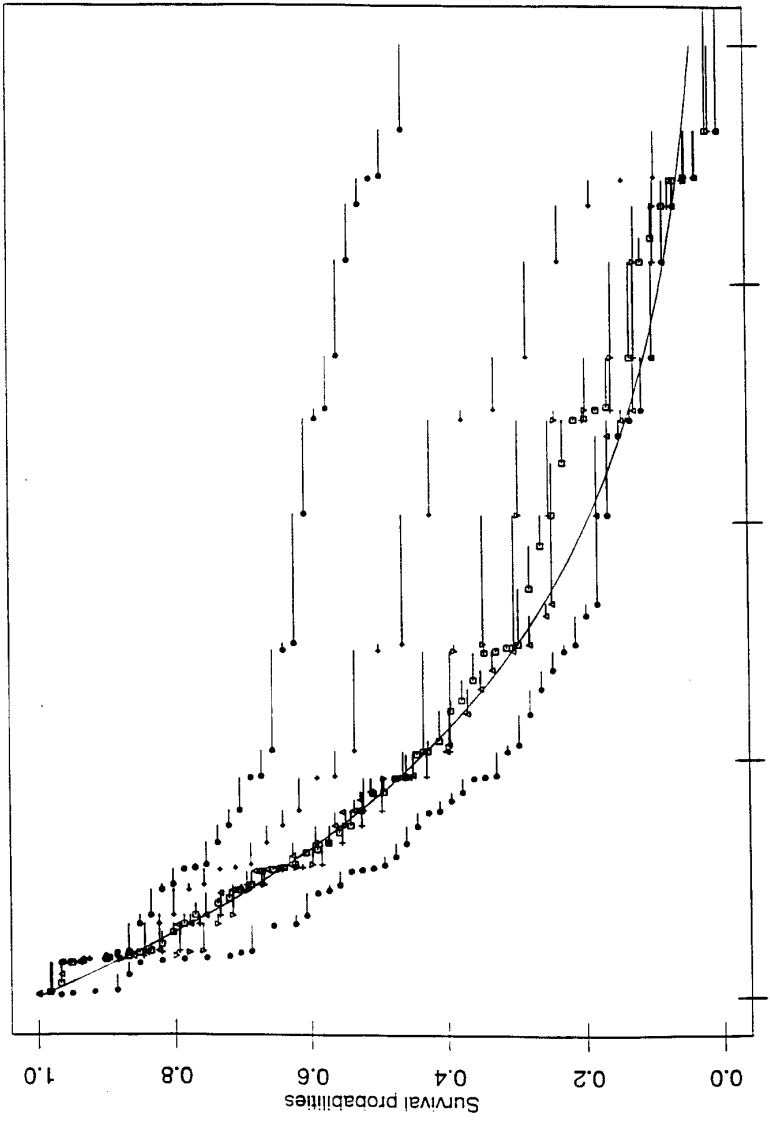

Figure 4 Estimates of the survival function–lung cancer example. Kaplan-Meier estimate (**filled diamond 3 dashes**); Lagakos-Williams estimate (**four dashes**); Peterson estimated bounds (**• two dashes**); Fisher-Kanarek estimate (**open triangle two dashes**); Slud-Rubinstein's estimate (**+ two dashes**); Klein-Moeschberger's estimate (**∇ two dashes**).

the data with techniques for right censoring and for interval censoring using Turbull's estimate (34). Notice that we call such an approach, i.e. replacing an interval-censored observation by its right-endpoint, as right imputation.

Figure 5 displays the Kaplan-Meier estimate of the probability of symptoms based on right-imputed data and the Trunbull estimate (a generalization of the Kaplan-Meier estimator for interval censoring). Note that the Kaplan-Meier estimate underestimates the probability of symptoms in early follow-up and overestimates it in late follow-up, thus resulting in an overestimate of the length of survival probabilities. This over-optimistic estimate occurs because the time to respiratory symptoms appears to be longer than it actually is when the right imputation approach is used.

Doubly interval censoring: Survival with malignant melanoma (Example 7 continued) So far we have considered only the problem of survival function estimation. In this example we illustrate the effect of different censoring assumptions on the estimates of regression coefficients for doubly interval-censored data. We consider the data from the malignant melanoma example described in the first section. Recall that the data structure is similar to that in Figure 3. To be more specific, the initial event is the first diagnosed stage

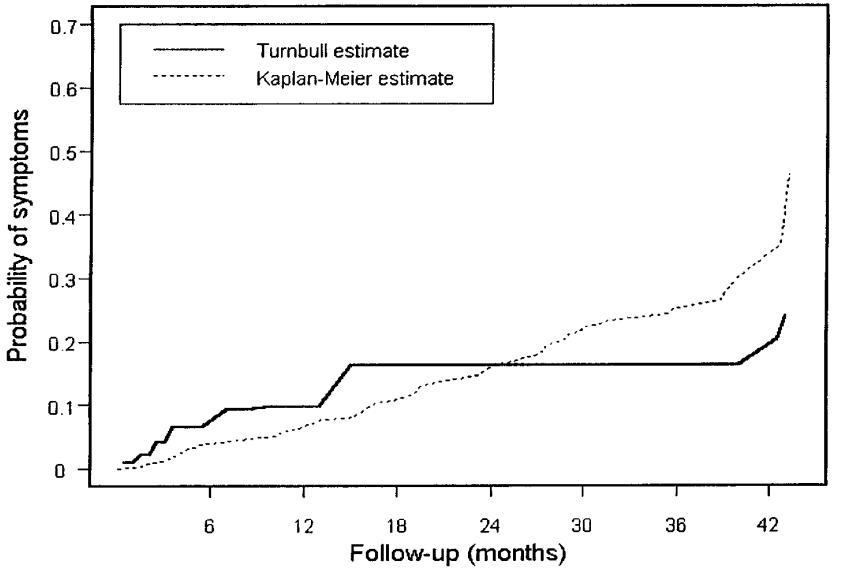

Figure 5 Estimate of the probability of symptoms–respiratory symptoms example.

II disease, the intermediate event is disease metastasis, and the final event is death. Since the time of disease metastasis is known only to be between the time of the last stage II disease and the time of the first stage III disease, the time of the intermediate event is left interval-censored when one computes survival time for post-disease metastasis. In this example, the times of the final events are either known exactly or they are right censored (i.e. there is no right interval-censoring).

Because of the left interval-censoring, we cannot directly apply the standard approaches for right-censored data for analysis. As mentioned in the second section, a simple analytic approach is to impute the time of the intermediate event (disease metastasis) by the right-point or the mid-point of the time interval and then apply the standard techniques for right-censored data. However, this approach may not be appropriate, as will be seen below. Table 2 presents the estimates of the regression coefficients from the imputation approaches and from an approach that takes the interval censoring into account as proposed by Leung & Elashoff (24). Basically, they assumed that the time of the event within the censored interval is governed by an unknown distribution, and proposed an estimate of the distribution.

Comparison of these estimates suggests that large differences exist among methods, especially for the estimates related to the interval censored time (the time from the first diagnosed stage II disease to disease metastasis). First, the imputation approaches led to very different estimates on the effect of site and the effect of time between the first diagnosed stage II disease and disease metastasis (metastases) as compared to the "correct" method. Second, the imputation approaches underestimate the standard errors of the regression

Table 2 Parameter estimates of the Weibull proportional hazards model—Melanoma Study (example 7)*

Coefficient	Leung & Elashoff† method	Mid-point imputation	Right-point imputation
MCV treatment (1 = treated, 0 = control)	−0.194 (0.097)	−0.197 (0.076)	−0.311 (0.076)
Breslow Depth (1 = depth ≥ 1.8 mm, 0 = depth < 1.8 mm)	−0.113 (0.104)	−0.068 (0.075)	−0.046 (0.075)
Gender (1 = male, 0 = female)	0.241 (0.100)	0.271 (0.076)	0.241 (0.075)
Metastasis site (1 = distant, 0 = others)	0.298 (0.126)	0.452 (0.088)	0.541 (0.088)
I{S ≥ 2 years}**	−0.991 (0.111)	−0.687 (0.082)	−0.087 (0.097)
λ_1 (scale parameter of Weibull dist.)	0.0345 (0.0069)	0.0499 (0.0073)	0.0887 (0.0131)
λ_2 (shape parameter of Weibull dist.)	1.161 (0.041)	1.043 (0.026)	0.854 (0.022)

*The survival time is defined as the time from disease metastasis to death.
**S represents the time from the first diagnosed stage II disease to disease metastasis and I{S ≥ 2 years} = 1 if S ≥ 2 years and = 0 otherwise.
† See Reference 24.

coefficient estimates. In summary, this example shows that one might obtain biased estimates and incorrect statistical inferences by falsely assuming that the time of event is equal to the right-point or the mid-point of the time interval.

DISCUSSION AND FUTURE DIRECTIONS

As we have demonstrated in the last section, assuming incorrect censoring mechanisms may lead to a serious bias. In practice, there are two common types of misassumptions regarding censoring mechanisms: For right censoring, researchers usually assume independent censoring (or noninformative censoring). For interval censoring (both singly and doubly interval censoring), researchers often assume that the occurrence of an event coincides with the reporting time (that is, right-imputation). In writing this paper, our intention was to direct investigators' attention to the dangers involved in analyzing censored survival data and point out some techniques to avoid these pitfalls.

An approach to investigating the situation was explored by Fisher & Kanarek (8): "In some situations a subset of the loss-to-follow-up cases may in fact be followed although at considerable expense." With such information, the investigator can either test the assumption of noninformative censoring (36) or estimate the risk of informative censoring (1). However, this strategy can be very expensive and can only be done when censoring is nonlethal.

All the methods dealing with informative censoring discussed in the literature assume that all censored cases are either all informative or all noninformative. In practice, there are many situations in which all three kinds of censoring (positive dependence, negative dependence, and noninformative censoring) are present in one sample. Thus, it would be useful to extend the existing methods to deal with all these situations. Furthermore, the methods described here for estimating the survival function under various conditions assume a fixed model parameter (Fisher-Kanarek's α, Slud-Rubinstein's ρ and Klein-Moeschberger's θ; see the last two sections for details). In practice, however, these values are rarely known. Thus, it would also be useful if there are some guidelines for investigators to determine the value, or at least a reasonable range of the model parameters, based on a sample data.

When covariates are available, sometimes it may be possible to recover some of the information lost by identifying a surrogate response variable measured on the censored subjects and using it to predict the residual survival time. Although Cox (3) initialized this idea, no follow-up on this subject has appeared in the literature.

Finally, the best way to handle censoring is to prevent it from happening by a good design; no matter how effective the statistical methods are, some information will be lost when analyzing censored data (2, 6, 35). These design

strategies for preventing, or at least minimizing, censoring would make a very useful contribution to the literature.

Literature Cited

1. Baker SG, Wax Y, Patterson BH. 1993. Regression analysis of grouped survival data: informative censoring and double sampling. *Biometrics* 49:379–89
2. Brook RJ. 1982. On the loss of information through censoring. *Biometrika* 69:137–44
3. Cox DR. 1983. A remark on censoring and surrogate response variable. *J. R. Statist. Soc. B* 45:391–93
4. De Gruttola V, Lagakos SW. 1989. Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* 45:1–11
5. Dudley J, Jin S, Hoover D, Metz S, Thackeray R, et al. 1995. The multicenter AIDS cohort study: retention after $9\frac{1}{2}$ years. *Am. J. Epidemiol.* 142:323–30
6. Ebrahimi N, Soofi ES. 1990. Relative information loss under type II censored exponential data. *Biometrika* 77:429–35
7. Finkelstein DM, Wolfe RA. 1985. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* 41:933–45
8. Fisher L, Kanarek P. 1974. Presenting censored survival data when censoring and survival times may not be independent. In *Reliability and Biometry*, ed. F Proschan, RJ Serfling, pp. 303–26. Philadelphia: SIAM
9. Franks P, Clancy CM, Gold MR. 1993. Health insurance and mortality. *JAMA* 270:737–41
10. Grüger J, Kay R, Schumacher M. 1991. The validity of inferences based on incomplete observations in disease state models. *Biometrics* 47:595–605
11. Hebert PR, Bolt RJ, Borhani NO, Cook NR, Cohen JW, et al. 1994. Design of a multicenter trial to evaluate long-term life-style intervention in adults with high-normal blood pressure levels: trials of hypertension prevention (Phase II). *Ann. Epidemiol.* 5:130–39
12. Heitjan DF. 1993. Ignorability and coarse data: some biomedical examples. *Biometrics* 49:1099–109
13. Heitjan DF, Rubin DB. 1991. Ignorability and coarse data. *Ann. Stat.* 19:2244–53
14. Hyde J. 1977. Testing survival under right censoring and left truncation. *Biometrika* 64:225–30
15. Jennrich R. 1983. A note on the behavior of the log rank permutation test under unequal censoring. *Biometrika* 70:133–37
16. Kalbfleisch JD, McKay RJ. 1979. On constant-sum models for censored survival data. *Biometrika* 66:87–90
17. Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–81
18. Kim MK, De Gruttola VD, Lagakos SW. 1993. Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics* 49:13–22
19. Klein JP, Moeschberger ML. 1988. Bounds on net survival probabilities for dependent competing risks. *Biometrics* 44:529–38
20. Kongerud J, Samuelsen SO. 1991. A longitudinal study of respiratory symptoms in aluminum potroom workers. *Am. Rev. Respir. Dis.* 144:10–16
21. Lagakos SW. 1979. General right censoring and its impact on the analysis of survival data. *Biometrics* 35:139–56
22. Lagakos SW, Williams JS. 1978. Models for censored survival analysis: a cone class of variable-sum models. *Biometrika* 65:181–89
23. Law CG, Brookmeyer R. 1992. Effects of mid-point imputation on the analysis of doubly censored data. *Stat. Med.* 11:1569–78
24. Leung KM, Elashoff RM. 1996. A three-state disease model with interval-censored data: Estimation and applications to AIDS and cancer. *Lifetime Data Anal.* 2:175–94
25. Little RJA, Rubin DB. 1987. *Statistical Analysis with Missing Data*. New York: Wiley. 278 pp.
26. Moeschberger ML. 1974. Life tests under competing causes of failure. *Technometrics* 16:39–47
27. Peterson AV. 1976. Bounds for a joint distribution function with fixed subdistribution functions: application to competing risks. *Proc. Natl. Acad. Sci.* USA 73:11–13
28. Samuelsen SO, Kongerud J. 1994. In-

terval censoring in longitudinal data of respiratory symptoms in aluminium potroom workers: a comparison of methods. *Stat. Med.* 13:1771–80
29. Saunders B, Baily S, Phillips M, Allsop S. 1993. Women with alcohol problems: Do they relapse for reasons different to their male counterparts? *Addiction* 88:1413–22
30. Slud E, Byar D. 1988. How dependent causes of death can make risk factors appear protective. *Biometrics* 44:265–69
31. Slud EV, Rubinstein LV. 1983. Dependent competing risks and summary survival curves. *Biometrika* 70:643–49
32. Tsai WY, Jewell NP, Wang MC. 1987. A note on the product-limit estimator under right censoring and left truncation. *Biometrika* 74:883–86
33. Tsiatis A. 1975. A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.* USA 72:20–22
34. Turnbull BW. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Stat. Soc. B* 38:290–95
35. Turrero A. 1989. On the relative efficiency of grouped and censored survival data. *Biometrika* 76:125–31
36. Wax Y, Baker SG, Patterson BH. 1993. A score test for noninformative censoring using doubly sampled grouped survival data. *Appl. Stat.* 42:159–72

Annual Review of Public Health
Volume 18, 1997

CONTENTS