DOCUMENT RESUME

ED 173 209 SO 011 767

AUTHOR	Tuma, Nancy Brandon; Hannan, Michael T.
TITLE	Final Report for Dynamic Models for Causal Analysis of Panel Data. Approaches to the Censoring Problem in Analysis of Event Histories. Part III, Chapter 2.
INSTITUTION	Center for Advanced Study in the Behavioral Sciences, Stanford, Calif.; Stanford Univ., Calif.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	[77]
GRANT	NIE-G-76-0082
NOTE	34p.; For related documents, see SO 011 759-772
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; *Data Analysis; *Educational Change; Measurement Techniques; Models; *Research Methodology; *Research Problems; *Social Change; Social Science Research; Sociometric Techniques; *Statistical Bias; *Validity

ABSTRACT

The document, part of a series of chapters described in SO 011 759, considers the problem of censoring in the analysis of event-histories (data on dated events, including dates of change from one qualitative state to another). Censoring refers to the lack of information on events that occur before or after the period for which data are available. Unless censorship is dealt with, researchers are likely to make erroneous inferences about the change process. The report considers several approaches to estimation when event-histories are censored. A constant rate (Poisson) model is considered because the methodological issues are more easily understood. Models in which the rate of an event depends on exogenous variables or time and in which there are multiple kinds of events are also analyzed. The report then discusses approaches to estimation based on maximum likelihood (ML), pseudo-maximum likelihood, the method of moments, and recent work by statisticians on methods that make weak parametric assumptions. The conclusion is that an important advantage of the ML approach to the censoring problem is that it is easily extended to different data structures and different models. (Author/KC)

Final Report for

DYNAMIC MODELS FOR CAUSAL ANALYSIS OF PANEL DATA

APPROACHES TO THE CENSORING PROBLEM IN ANALYSIS OF EVENT HISTORIES*

by

Nancy Brandon Tuma
Stanford University

Michael T. Hannan
Center for Advanced Study in the Behavioral Sciences
and
Stanford University

*This research was supported by the National Institute of Education grant NIE-G-76-0082. The authors want to thank Glenn Carroll for his help in conducting the Monte Carlo simulations. Both authors contributed equally to this paper, and the order of names was determined by the flip of a coin. We wish to thank Burton Singer and an anonymous reviewer for their helpful comments. This paper is also published as a chapter in Sociological Methodology 1979, Karl Schuessler (ed.), pages 209-240. San Francisco: Jossey-Bass.

SO 011 767

9

APPROACHES TO THE CENSORING PROBLEM IN ANALYSIS OF EVENT HISTORIES

Nancy Brandon Tuma
STANFORD UNIVERSITY

Michael T. Hannan
CENTER FOR ADVANCED STUDY IN THE BEHAVIORAL SCIENCES
AND
STANFORD UNIVERSITY

Sociologists are increasingly engaged in collecting data on people, firms, schools, nations, and other social units over time.

This research was supported by the National Institute of Education Grant NIE-G-76-0082. The authors want to thank Glenn Carroll for his help in conducting the Monte Carlo simulations. Both authors contributed equally to this paper, and the order of names was determined by the flip of a coin. We wish to thank Burton Singer and an anonymous reviewer for their helpful comments.

Some sources of temporal data such as archives and newspapers contain information on the dates at which certain events occur. In other instances we can obtain retrospective histories on individuals and organizations. We refer to data on dated events—including dates of change from one qualitative state to another—as event histories.

Sociologists with a wide variety of substantive interests have access to such event histories. Many of the available histories characterize the careers of individuals; there are histories of births, marriages, jobs, illnesses (or hospitalizations), arrests and convictions, residences, and so on. Sociologists interested in formal organizations often have access to information on dates of internal reorganizations, mergers, starts of contracts (with unions), and failures. For those interested in system-level problems, there are documents giving the dates of episodes of collective violence (for example, riots, lynchings, insurrections), wars, changes in political regimes, revitalization movements, and so forth.

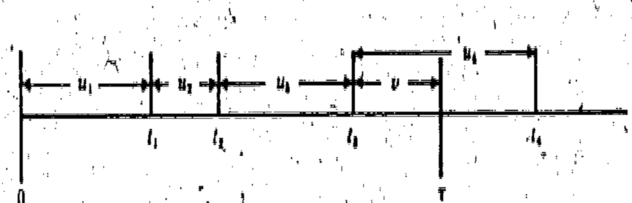
Sociologists rarely make full use of such data. Instead, we typically analyze only a portion of them. When the events in question have short durations (for example, strikes, riots), analysts frequently aggregate over time periods. They may conduct time-series analysis of counts of events in successive periods, usually coded in terms of the severity of events (see, for example, Snyder and Tilly, 1972; Shorter and Tilly, 1974; Chirot and Ragin, 1975). Or they may aggregate events over the entire period observed and analyze the data cross-sectionally (for example, see Spilerman, 1976). When the average duration between events is fairly long (for example, marriages, political regimes), analysts commonly simplify the data in another way. They take "pictures" at a series of discrete time points and characterize units in terms of the state occupied at each point—for example, each person is identified as married or not at each point. The series of cross-sections are then treated like panel observations. For example, studies of marital stability using data on marital histories commonly investigate changes in marital status over some fixed period, say a year. Then individuals are coded as either changing or not changing their initial marital status (see, for example, Bumpass and Sweet, 1972; Glick and Norton, 1971; Morgan and others, 1974).

Ignoring information on the timing of events may in some cases be justified on grounds of convenience. Occasionally there are substantive reasons for ignoring the timing of events. See, for example, White's 1970 arguments for ignoring the duration of vacancies of jobs. However, we suspect that in many cases sociologists are unaware of both the significance of the timing of events and statistical procedures for analyzing that timing. Event histories would probably be more widely available in sociology if their significance as data, and procedures for analyzing these data, were better known.

It seems obvious that analyses of event histories that use all information contained in the data—information on the number, sequence, and timing of events—are preferred to approaches that use only a portion of it. We have discussed elsewhere (Tuma and others, in press) both the value of analyzing event histories and various procedures for doing such analysis. Here we focus on a problem that arises in attempts to use all information in event histories: the problem of censored observations.

A diagrammatic representation of a typical event history may help in visualizing the problem. In Figure 1 the period of observation lies between the two vertical lines extending above and below the time axis at 0 and τ. The dates at which events occur are indicated by vertical lines above the time axis. The timing of the jth event is denoted by t_j. The time between the $(j-1)$th and jth events is denoted by u_j; later we refer to this as the jth spell or interval. In our example the fourth event happens after τ; therefore it is not observed. However, we do observe that no event occurs between t_3 and τ; we denote the length of this period by v. Whenever v is greater than zero, the record is said to be *censored* on the right. That is, the timing of events occurring after τ is not

Figure 1. A typical event history.



observed. Thus censoring is a characteristic of an observation plan; it describes a characteristic of a sample.[1]

In Figure 1 observation begins at the origin of the unit's history. Sometimes observation begins after the origin so that some early events may not be observed. In this case there is said to be censoring on the left. Event-history data available to sociologists are almost always censored on the right and sometimes on the left, too. The methodological problem is whether to use censored observations (intervals) such as v—and if so, *how* to use them.

The problem of censoring is not new to social scientists.[2] Sørensen (1976) has discussed it in the following context. Suppose a survey at a given point in time collects information on the dates on which the N individuals in the sample entered and left a certain state, such as employment. To avoid recall errors, we may want to analyze data pertaining only to either a recent time period (say the past 2 years) or the current state occupied by each individual. Our problem is that we do not know when the current state will end. In a study of the determinants of unemployment, for example, some individuals will be unemployed at the interview. We can ask how long they *have been* unemployed, but we cannot determine the completed duration of the current episode of unemployment.

Exactly the same problem arises with archival data. A sociologist studying the failure rate of organizations faces this problem because all public documents stop at some date. In any observed year, some organizations founded in a given previous year will "die" and others will not. Observations on the life span of surviving organizations are censored.

Panel studies that obtain retrospective data on the periods between waves of interviews lead to the same problem. In the four large-scale income maintenance experiments conducted in the past decade, for example, family heads were interviewed several

[1] According to Kendall and Buckland (1971, p. 20): "A sample is said to be censored when certain values are unknown (or deliberately ignored) although their existence is known."

[2] Demographers have been especially aware of the problem. For example, see Shops and Menken (1972a, 1972b). Economists have also begun to address the issue; see, for example, Heckman (1977).

times a year during the course of the experiment. At each interview the dates of all intervening changes in jobs and in marital status were collected. At the end of the experiment some heads had not yet changed jobs or marital status and others had dropped out of the experiment. Both types of observations involve censoring. What should be done with information on these cases?

There are at least three possible approaches to the censoring problem: (1) Ignore censored observations and analyze only those cases with an event during the observation period; (2) treat censored observations as though an event occurred at the time of the last observation; or (3) use a method of estimation that adjusts for censoring under the assumption that the same stochastic model applies to all cases, whether or not observations on them are censored. Although the second strategy may be used inadvertently, it probably has few defenders. To recode censored observations in this way is the same as recoding nonevents as events. Ignoring censored observations is more common, especially when caused by attrition. (Almost all panel studies of change in marital status do this. See the discussion in Hannan and others, 1976.) In this chapter we compare both these arbitrary procedures to a statistically sound method for several stochastic models. We find that these arbitrary procedures lead to biased estimates of parameters of the models considered.

Although there is a fairly large statistical literature on the censoring problem,[3] this literature seems largely unfamiliar to sociologists. Sørensen (1976) has discussed the censoring problem in a sociological context and proposed various ways of dealing with the problem in particular circumstances—without, however, referring to the statistical literature. We review his proposals and place them in a broader context. In particular, we show that all but one of his proposals are special cases of a common procedure—maximum-likelihood estimation of hazard functions—applied to a model in which the rate at which an event occurs (the hazard) is a time-independent constant. We compare the small-sample properties of various estimators for this model.

[3] For nonrepeatable events, Gross and Clark's (1975) recent text provides a useful introduction to this literature. For repeatable events, Cox and Lewis (1966) are helpful.

We pay special attention to the effects of the degree of censoring on the quality of the estimators. Then we show that the general approach can be generalized to deal with the effects of causal variables on rates, with the rates of multiple kinds of events, and with time dependence of rates. These extensions require parametric assumptions about the nature of the stochastic process generating the events. We conclude with a brief look at some recent statistical contributions on estimation from censored data; these estimators are based on weaker parametric assumptions, but their relative quality in small samples is not yet known.

MODELING THE STOCHASTIC PROCESS GENERATING EVENTS

It is useful to distinguish two types of events: nonrepeatable events (such as the death of an individual) and repeatable events (such as outbreaks of collective violence in a city). The statistical literature on censoring emphasizes the first type. When there is no need to specify the joint probability distribution of a *sequence* of events, it is usually fairly simple to model the stochastic process generating the event. When a unit can have repeated events, systematic analysis of event histories requires that one formulate a model giving the joint probability distribution for a sequence of events.

Let T_j be a random variable denoting the timing of the jth event and let Z_j be a random variable denoting the state entered when the jth event occurs. Because events and censoring rarely occur at any fixed interval of time, we assume that possible values of T_j belong to a continuous interval of time. We assume, moreover, that the possible values of Z_j belong to a set of positive integers. We must specify a model that gives the joint probability distribution of the T's and Z's:

$$\Pr[T_1 < t_1, T_2 < t_2, \ldots, T_M < t_M, Z_1 = z_1, Z_2 = z_2, \ldots, Z_M = z_M] \quad (1)$$

where $0 < t_1 < t_2 \cdots < t_M$ and M is the largest number of events under consideration.

Let $Y(t)$ be a random variable representing the state occupied at time t. If we know $Y(t)$ for all t, we can obviously infer $t_1, t_2, \ldots, t_M$ and $z_1, z_2, \ldots, z_M$. Similarly, if we know

$$p_k(t) = Pr[Y(t) = k] \quad (2)$$

for all k and t, we can determine the joint probability distribution in (1).

Thus we need a tractable model that lets us determine (2). In this chapter we consider only continuous-time discrete-state Markov processes as a means of simplifying the exposition of the basic methodological issues surrounding the censoring problem. However, the procedures that we recommend for analyzing event histories can be extended to models based on other stochastic processes (for example, a semi-Markov process).

The formal assumptions of a discrete-state, continuous-time Markov process may be specified as follows. Let $p_{jk}(v, t)$ denote the probability that a unit in state j at time v is in state k at time t, $v < u_1 < u_2 < \cdots < u_n < t$:[4]

$$p_{jk}(v, t) = \Pr[Y(t) = k \mid Y(v) = j, Y(u_1) = y_1, \ldots, Y(u_n) = y_n] \quad (3)$$

Under the Markov assumption

$$p_{jk}(v, t) = \Pr[Y(t) = k \mid Y(v) = j] \quad (4)$$

This means that we do not need to know which state is occupied at every moment between v and t in order to know the probability of being in state k at time t. To satisfy this assumption, the Chapman-Kolmogorov equation must hold:

$$p_{jk}(v, t) = \sum p_{jh}(v, u)\, p_{hk}(u, t) \quad (5)$$

where $v \le u \le t$. It is also natural to assume that

$$0 \le p_{jk}(v, t) \le 1 \quad (6)$$

$$p_{jk}(t, t) = \begin{cases} 0 & \text{for } j \ne k \\ 1 & \text{for } j = k \end{cases} \quad (7)$$

[4]In Equations (3) through (9), the symbols v, u, and u_m ($m = 1, \ldots, n$) merely represent points on the time axis. In the rest of this chapter they have the meaning we gave them in Figure 1.

$$\sum_{k} p_{jk}(v, t) = 1 \qquad (8)$$

Given these assumptions, it follows that

$$p_k(t) = \sum_j p_j(v) p_{jk}(v, t) \qquad (9)$$

Furthermore, note that by repeated application of the Chapman-Kolmogorov equation we can compute $p_k(t)$ if we know the probability of a transition between all states j and k for very small time intervals Δt. This relationship (9) is useful if we do not observe $Y(t)$ and would like to predict it. However, this prediction requires that we assume that

$$\lim_{\Delta t \to 0} p_{jk}(t, t + \Delta t)/\Delta t = r_{jk}(t) < \infty \qquad (10)$$

for all j and k, $j \neq k$. Note that $r_{jk}(t)$ cannot be negative.

We refer to $r_{jk}(t)$ as the rate of a change from j to k at time t. Others sometimes call it an instantaneous rate of transition, a transition intensity, or an infinitesimal generator. The rates are the fundamental parameters of a Markov process.

A CONSTANT-RATE MODEL

Sociologists may be more interested in modeling variable rates than constant rates. Nonetheless, there are two reasons for beginning with a constant-rate model. First, the simplicity of the model makes the main features of the general strategy easy to comprehend. Second, Sørensen (1976) restricted his attention to this model and we wish to clarify his proposals.

We still focus on a continuous-time, discrete-state Markov process. To compare with Sørensen's (1977) discussion, we need an additional simplification. Let $Y(t)$ record the number of events that occur on or before t, $0 \leq t < \infty$, and let the rates that define the Markov process be given by

$$r_{jk} = \begin{cases} \alpha & \text{if } k = j + 1 \\ 0 & \text{otherwise} \end{cases} \qquad (11)$$

Thus transitions can only occur to the next higher integer; that is, $Y(t)$ is a nondecreasing function of time and in any small moment of time can increase by at most 1. It may be proved that the Poisson model (discussed by Sørensen) results from these assumptions. (See Breiman, 1969, pp. 205–206 for a simple proof.) In particular, we find

$$\Pr[Y(\tau) = y] = [(\alpha\tau)^y/y!]e^{-\alpha\tau} \qquad (12)$$

where α represents the constant rate of occurrence. We wish to estimate α, the rate at which an event occurs. However, in our problem (by definition) event histories are censored: Some units will have experienced no event by τ. Others will have had one or more events, but, as shown in Figure 1, the last time segment is also censored because the end of the observation period rarely coincides with the last event.

Maximum-Likelihood Estimators

In this chapter we concentrate on maximum-likelihood estimators because they can be obtained for most stochastic processes, can handle censoring, and usually have good properties in a large sample (for example, see Dhrymes, 1970).[5] The principle of maximum likelihood is to choose those parameter values that make the observations in the sample most likely.

We begin with the case in which the sample observations consist of the values of $Y(\tau_i)$ for each individual i, $i = 1, \ldots, N$, where τ_i is the length of the observational period for i. Let y_i represent the observed value of $Y(\tau_i)$—that is, the number of events for i between 0 and τ_i. If observations on the N units are independent, then the likelihood function is

$$L[Y(\tau_1), Y(\tau_2), \ldots, Y(\tau_N)] = \prod_{i=1}^{N} [(\alpha\tau_i)^{y_i}/y_i!]e^{-\alpha\tau_i} \qquad (13)$$

This is the product of the probabilities of the observations $(y_1, y_2, \ldots, y_N)$ as given by (12). Differentiating the log of (13) with respect to α and setting the result equal to zero gives the maximum-likelihood estimator (MLE) for the parameter of a

[5] To the best of our knowledge, Boag (1949) is the first to have reported the use of maximum-likelihood estimation to deal with the censoring problem.

Poisson model:

$$\hat{\alpha} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} \tau_i \tag{14}$$

which is just the total number of observed events divided by the total observation time.

There are two methodological reasons for using data on the timing of events rather than the total number of events. First, respondents giving retrospective histories may accurately recall only recent events—at the extreme, only the date of the most recent event. Second, when several kinds of events can occur, it is difficult, if not impossible, to write probability densities for the number of visits to the set of states.

When the date of each event is known, the sample observations consist of both y_i—the number of events within τ_i—and the sequence of dates for these events $(t_1, \ldots, t_{y_i})$ for each individual i. These data contain the implicit information that no event occurred between t_y and τ_i.

Let $f(t_1, \ldots, t_y)$ denote the probability density that the y events would be observed on the dates indicated, and let $G(\tau - t_y | t_1, \ldots, t_y)$ denote the probability of no event between t_y and τ, conditional on $t_1, \ldots, t_y$. (G is usually called the survivor function.) Then a general formulation of the likelihood of the observations (not specific to the constant-rate model) is just

$$L = \prod_{i=1}^{N} f(t_1, \ldots, t_y) \cdot G(\tau - t_y | t_1, \ldots, t_y) \tag{15}$$

where subscript i on t's and y's is suppressed for clarity. Note that both completed spells and censored spells contribute to the likelihood function.

For Model (15) the joint probability of the dates of events is the product of the probability density of the intervals between events:

$$f(t_1, t_2, \ldots, t_y) = f(u_1) \cdot f(u_2) \cdots f(u_y) \tag{16}$$

where $u_j = t_j - t_{j-1}$ (see Figure 1). Moreover, it may be shown that for a constant-rate model (that is, a Poisson model) the

length of time between events is exponentially distributed (see, for example, Breiman, 1969, pp. 37–38) and has the probability density

$$f(u_j) = \alpha e^{-\alpha u_j}, \quad u \geq 0 \tag{17}$$

for all j. In addition, the probability of no event between t_y and τ is just

$$G(\tau - t_y) = G(v) = e^{-\alpha v} \tag{18}$$

Substituting these expressions in (15), we obtain the likelihood function for the constant-rate model when the data consist of the dates of all events within the observation period:

$$L = \prod_{i=1}^{N} \left[\prod_{j=1}^{\infty} (\alpha e^{-\alpha u_{ij}})^{y_{ij}} \right] \cdot e^{-\alpha v_i} \tag{19}$$

where y_{ij} is equal to 1 if the jth event occurs before τ_i for individual i and otherwise is zero. Since y_{ij} equals zero when the jth event for individual i occurs after the censoring point τ, this simplifies to

$$L = \prod_{i=1}^{N} \left[\prod_{j=1}^{y_i} (\alpha e^{-\alpha u_{ij}}) \right] \cdot e^{-\alpha v_i} \tag{20}$$

Henceforth we use this simplification in forming estimators. Note that y_i, the number of events before τ_i, is the sum of the y_{ij}:

$$y_i = \sum_{j=1}^{\infty} y_{ij} \tag{21}$$

The maximum-likelihood estimator of α obtained from (20) is

$$\hat{\alpha} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} \left[v_i + \sum_{j=1}^{y_i} u_{ij} \right] \tag{22}$$

But for every individual i the sum of the lengths of completed spells plus the length of the censored spell is just the total observation time τ_i (see Figure 1). Therefore, as before, the maximum-likelihood estimator of α is the ratio of the total number of observed events to the total observation time. For present purposes the important point is that for a constant-rate model the maximum-likelihood estimator of α using the dates of events (Equation 22) is identical to that obtained using information on the number of events (Equation 14).

Since we refer regularly to Expression (22), we simplify and partially alter our notation.[6] Let Y represent the total number of events, let U denote the total length of completed spells, and let V refer to the total length of censored spells. Then the expression in (22) can be written:

$$\hat{\alpha} = Y/(U + V) \tag{23}$$

We shall refer to this as the all-event MLE.

At one point below we refer to an MLE that uses only the date of the first event in the observation period. That is, we use only y_{1i} and t_{1i} (which equals u_{1i}). Then the likelihood function is

$$L = \prod_{i=1}^{N} (\alpha e^{-\alpha u_{1i}})^{y_{1i}} \cdot (e^{-\alpha \tau_i})^{(1-y_{1i})} \tag{24}$$

and the MLE is

$$\hat{\alpha} = \sum_{i=1}^{N} y_{1i} \Bigg/ \left[\sum_{i=1}^{N} y_{1i}u_{1i} + \sum_{i=1}^{N} (1 - y_{1i})\tau_i\right] \equiv Y_1/(U_1 + V_1) \tag{25}$$

where Y_1 denotes the number of individuals who have a first event, U_1 refers to the total time to the first observed event, and V_1 denotes the total observation time for individuals who have no events. The general form of the estimator is the same as in (23), but dates of events after the first event are ignored. This is, of course, the only possible specification when an event is nonrepeatable.

Two Pseudo-MLEs

The maximum-likelihood framework permits us to show the consequences of dealing inadequately with the censoring problem. We consider the two possibilities introduced at the outset. The most common procedure is to ignore censored observations. The pseudo-likelihood expression formed under this

[6] Subsequently capital letters usually denote sample sums rather than random variables.

procedure is

$$L = \prod_{i=1}^{N} \prod_{j=1}^{\infty} (\alpha e^{-\alpha u_{ji}})^{y_{ji}} \tag{26}$$

and the "pseudo-MLE" is

$$\tilde{\alpha} = Y/U \tag{27}$$

The second alternative is to recode censored observations so that events are assumed to occur at the end of the observational period. A procedure similar to the one presented above leads to a "pseudo-MLE" with the form

$$\tilde{\alpha} = (Y + N)/(U + V) \tag{28}$$

The true MLE in (23) is lower than either of these estimators for any given set of data. This is seen from the ratio of the MLE to the pseudo-MLE. In the case of the estimator (27) that ignores censored observations, this ratio ($\hat{\alpha}/\tilde{\alpha}$) is $U/(U + V)$, the proportion of the total observation time ($U + V$) spent in uncensored spells (U). In the case of the estimator that recodes censored spells (28), this ratio ($\hat{\alpha}/\tilde{\alpha}$) is $Y/(Y + N)$, the fraction of spells ($Y + N$) that are not censored (Y). Thus the greater the number of censored spells and the longer the observation time covered by them, the larger are these estimators relative to the MLE. Since the MLE is unbiased in large samples, these arbitrary procedures give upwardly biased estimates in large samples. Furthermore, as we discuss below, the MLE is slightly upwardly biased in small samples. Since the ad hoc procedures give still higher estimates, they cannot improve on the MLE even in small samples. We return to this issue when we discuss models that incorporate causal effects on rates.

Sørensen's Moment Estimators

Sørensen (1976) applied the method of moments to several of Feller's (1971, pp. 11–14) results on the "waiting time paradox" to derive estimators for the constant-rate model. He considered two situations.

In the first case the investigator trusts only the accuracy of the date of the most recent event. Sørensen's goal was to show that the parameter of the constant-rate model may be estimated using only information on censored spells. The estimator depends on whether the origin of the process is far (with a small probability of no event prior to the consoring) or near. In the former case (Sørensen's eq. 14 written in our notation):

$$\alpha = 1/E(v) \tag{29}$$

The moment estimator for α is obtained by replacing E(v) with V/N, the sample mean of the censored spells:

$$\hat{\alpha}_{S1} = N/V \tag{30}$$

where the subscript on $\hat{\alpha}$ stands for Sørensen's first estimator.

In fact, this estimator can be interpreted as a maximum likelihood estimator. In the special case of the constant rate model, time is reversible. That is, beginning at any arbitrary chosen point, the probability distribution of time to an event is the same working both forward and backward in time. So this estimator is formally equivalent to one that follows a sample of units backward in time from the censoring point until each has an event. Because the interruption is far from the start of the process the probability that a unit has no event within the observational period is essentially zero. Consequently, there is no consoring problem, so the likelihood function is

$$L = \prod_{i=1}^{N} \alpha e^{-\alpha v_i} \tag{31}$$

and the MLE is

$$\hat{\alpha} = N/V = \hat{\alpha}_{S1} \tag{32}$$

This is just the MLE for first events when there is no censoring (Equation 25).

Near the origin of the process the relevant equation (Sørensen's eq. 17 is

$$\alpha = (1 - e^{-\alpha r})/E(v) \tag{33}$$

Again Sørensen proposed a moment estimator, replacing $e^{-\alpha r}$ with Po, the sample proportion with no event, and E(v) with the observed mean length of the censored spells. We have already denoted the number of units with an event as Y1. Let us indicate the

number with no event as Y_0 so that $N = Y_0 + Y_1$. Then P_0 is $(N - Y_1)/N$ or Y_0/N. For the moment we denote the mean length of censored spells by $\bar{w}$. The Sørensen's estimator is

$$\hat{\alpha}_{S2} = [1 - (Y_0/N)]/\bar{w}, \tag{34}$$

This estimator may also be interpreted as a first-event backward MLE, with r the origin of the backward process. Individuals are followed backward in time until they have an event as to the original starting point, zero, if they have no event. The observed mean duration $\bar{w}$ reflects these two types of observations. For the Y_0 individuals who do not have an event in (0, r), the duration is merely r. These observations contribute $Y_0 r$ to the mean. For those who have an event in (0, r), the contribution is v in Figure 1. But since we are now taking a backward perspective, the duration v ends in an event and should be called u. That is, taking a backward perspective,

$$\bar{w} = (Y_0 r + \sum_{i=1}^{N} y_{1i} u_i)/N = (Y_0 r + U_1)/N \tag{35}$$

Then

$$\hat{\alpha}_{S2} = [1 - (Y_0/N)]/\bar{w} = Y_1/(U_1 + Y_0 r) \tag{36}$$

Since Sorensen assumed $r_i = r$ for all i, $Y_0 r = \Sigma_{i=1}^{N}(1 - y_{1i}) r = V_1$ the Equation 25). Then

$$\hat{\alpha}_{S2} = Y_1/(U_1 + V_1) \tag{37}$$

which is identical to the first-event MLE in (25).

Our analysis of Sorensen's estimators demonstrates that maximum likelihood applied only to censored observations can be used to estimate the parameter of a Poisson process. This result depends on the indifference of the model to time; once can work either forward or backward and obtain the same result.

In the second case considered by Sorensen, the investigator needs to use all spells during the observation period in the analysis. The logic underlying the choice between his proposed estimators is less clear than in the previous case. First Sorensen conferred and rejected the inverse of the mean length of all spells, (Y + N) /(U + V), as upwardly biased (see Equation 28). Instead he proposed that the length of all censored spells be doubled.

That is, he proposed the estimator

$$\hat{\alpha}_{S3} = (Y + N)/(U + 2V) \qquad (38)$$

His reasoning depended on the fact that the expected length of the censored spells is twice the expected length of spells ending in an event (see Feller, 1971). Doubling the observed length of censored spells (that is, v) is intended to compensate for the difference in the expected length of censored and uncensored spells.

Finally, Sørensen (1976) proposed the following estimator in his equation 22:

$$\hat{\alpha}_{S4} = (\bar{\tau} + V/N)/(U/N)\tau = Y/N\tau = Y/(U + V) \qquad (39)$$

This is just the estimator in (23), which is MLE. Clearly it is not the same as the estimator in (38). Sørensen's discussion does not offer any grounds for choosing between the estimators in (38) and (39). Below we investigate the differences between the two estimators.

A Monte Carlo Study of Constant-Rate Estimators

Though the estimators discussed above use different information about events within a period of time, all but one (38) are maximum-likelihood estimators, which are asymptotically normal, unbiased, and efficient under quite general conditions. In other words, in large samples one cannot improve on these estimators. However, for small samples the MLE in (25) is known to be slightly upwardly biased and to have a nonnormal distribution (Bartholomew, 1957, 1963; Mendenhall and Lehman, 1960).

Below we report results of a Monte Carlo study that investigated how properties of the various MLEs depend on sample size and the level of censoring. These results are useful for showing that in small samples the upward bias is quite small and that the large-sample theory is a good approximation for the situations confronted by most sociologists. It also shows the costs of ignoring information on certain spells. Finally, it lets us contrast the ML estimators with the non-ML estimator that doubles the length of censored spells.

In all phases of the Monte Carlo study we examined samples of sizes 25, 50, 100, 250, and 500 and arbitrarily set α equal to 1 per unit of time. (For convenience we assume that time is measured in years.) For each unit in a sample we used the relationship

$$\ln[G(u_j)] = \ln[e^{-\alpha u_j}] = -\alpha u_j = -u_j \qquad (40)$$

to simulate u_j, the time between events $j - 1$ and j. For each unit we drew a pseudo-random number from a (0,1) uniform distribution and called this $G(u_j)$; then we used (40) to calculate u_j. We repeated the entire process 200 times for each sample size.

We begin with the results for the first-event MLE (25). Table 1 reports the means of the estimates of α over the 200 samples for each sample size. If the estimator were unbiased, the mean over the 200 samples would approximately equal unity, the true value of α. With no censoring and $N = 25$, the mean was 5.5 per-

TABLE 1
Effect of Censoring on the Behavior of a Maximum-Likelihood Estimator for a Constant-Rate Model
($\alpha = 1$; Number of Samples = 200)

Sample Size	Degree of Censoring	Mean $\hat{\alpha}$	Variance of $\hat{\alpha}$
25	0	1.055	0.058
	0.20	1.059	0.091
	0.50	1.034	0.110
	0.80	1.018	0.206
	0.90	1.019	0.426
50	0	1.015	0.023
	0.20	1.017	0.043
	0.50	0.998	0.057
	0.80	0.991	0.094
	0.90	0.974	0.168
100	0	1.006	0.012
	0.20	1.006	0.023
	0.50	0.997	0.030
	0.80	0.996	0.055
	0.90	0.990	0.103
250	0	1.005	0.004
	0.20	1.004	0.010
	0.50	0.999	0.012
	0.80	1.006	0.028
	0.90	1.010	0.049
500	0	1.001	0.002
	0.20	0.999	0.004
	0.50	0.995	0.006
	0.80	1.000	0.013
	0.90	0.996	0.023

cent higher than the true value. However, the upward bias declined to 1.5 percent of the true value for $N = 50$ and was negligible for the larger sa[illegible] sizes.

Next we co[illegible] the effect of censoring. First we chose a censoring point τ [illegible] that the probability of no event prior to τ—that is, the level of censoring—was 0.2, 0.5, 0.8, and 0.9; we used an expression like (40) to calculate τ (τ replaced u). If the uncensored time for a case was greater than τ, then the case was censored. If it was less than τ, it was uncensored.

Table 1 reports the means across 200 samples for each level of censoring and for each sample size. Three results are of interest. First, for medium-sized samples censoring had a negligible effect on the bias. Second, for small samples, especially $N = 25$, censoring actually reduced the upward bias in the estimator. Finally, as one would expect, censoring increased the variance of the estimator considerably. A shift from no censoring to 20 percent censoring approximately doubled the variance; so did a shift from 80 to 90 percent censoring. Insofar as the mean squared error is concerned, the impact of censoring on the variance actually swamped its effect on bias. For the combination of small samples and high levels of censoring, which implies few observed events, the estimator was very imprecise. In relatively large samples, however, the effects of even extreme levels of censoring were rather modest. These results suggest that the first-event MLE has good properties for small samples with slight censoring and for moderately large samples with even high degrees of censoring.

To this point we have only used the first event for each unit to estimate the rate at which an event occurs. In many research contexts there are multiple events within a given observational period τ. To investigate estimators that use data on multiple events, we chose τ to be 1 year, so that the average number of events within the period would be unity. However, during this period some units would have more than one event and others none.

We wish to contrast three estimators of the rate for τ equal to 1 year: (1) the MLE using data on all events (23); (2) Sørensen's "backward" MLE, which uses information only on the length of the last spell when the process is near the origin, $\hat{\alpha}_{S2}$ (37); and (3) Sørensen's non-ML estimator that doubles the length of censored spells, $\hat{\alpha}_{S3}$ (38). The findings are given in Table 2.

TABLE 2

Comparison of Alternative Estimators for a Constant-Rate Model ($\alpha = 1$; Number of Samples = 200)

Estimator	Sample Size	Mean	Variance
(1) All-event MLE for first-year $\hat{\alpha}$	25	1.005	0.035
	50	1.005	0.021
	100	0.999	0.010
(2) Last-spell MLE for first-year $\hat{\alpha}_{S2}$	25	1.047	0.060
	50	1.040	0.041
	100	1.017	0.017
(3) Last spell doubled for first-year $\hat{\alpha}_{S3}$	25	1.239	0.025
	50	1.236	0.016
	100	1.229	0.007
(4) All-event MLE for fifth-year $\hat{\alpha}$	25	0.98[illegible]	0.042
	50	0.975	0.018
	100	0.988	0.009
(5) Last-spell MLE for fifth-year $\hat{\alpha}_{S1}$	25	1.112	0.003
	50	1.110	0.001
	100	1.112	0.001

The results for the all-event MLE were very good, much better than the first-event MLE (compare with Table 1). For example, the average bias for $N = 25$ was only 0.5 percent compared with 5.5 to 1.8 percent for the first-event MLE with the same sample size. This is remarkably good for such a small sample.

Next consider Sørensen's "backward" MLE. Both the upward bias and variance were larger than for the all-event MLE. So the backward MLE is noticeably inferior to the all-[illegible]nt MLE. However, the backward MLE is still a reasonably good estimator and could be used if only the starting date of the last spell is known or can be accurately recalled.

Now consider the non-ML estimator that doubles the length of censored spells. It was upwardly biased and had a much larger bias (approximately 23 percent) than any other estimator studied. Clearly it cannot be recommended.

Finally, we chose τ equal to 5 years. We wish to compare two estimators: the MLE that uses data on all events in the fifth year (23) and the backward MLE that uses only the last spell in the fifth year (30). The former lets us examine the effects of having censoring on both the left and the right. The latter lets us study $\hat{\alpha}_{51}$, which presumes that all units have had at least one event during the observation period. Given the assumptions of the model and $\alpha = 1$, the probability of no event in 5 years is less than 0.01, so $\hat{\alpha}_{51}$ is the appropriate estimator.

First consider the results for the all-event MLE for the fifth year, in which data were censored on both the left and right. The all-event MLE was slightly downwardly biased by approximately 2 percent. For a given sample size the variance was about the same as for the all-event MLE for the first year. On the other hand, the estimator that uses only data on the last spell during the fifth year, $\hat{\alpha}_{51}$, had a fairly large upward bias (about 11 percent) but a very small variance. For small samples (N = 25 and 50), this estimator had a smaller mean squared error than the all-event MLE for the fifth year. With larger sample sizes, however, it had a larger mean squared error than the all-event MLE. Consequently, the choice between (23) and (30) is less clear than in the case of data on the first year, where data are censored on the right but not the left.

Overall the results of the Monte Carlo study indicate several things. All maximum-likelihood estimators adjust for censoring reasonably well. In fact, except for very small samples, the quality of these estimators appears good even with extreme levels of censoring. This finding is very important for the analysis of data when the observational period is short relative to the average length of time between events, which often occurs in studies of marriage and divorce, failure of organizations, and so forth. The quality of estimators based on forward and backward first events was generally not as good as that of estimators that use all events during some period. Even first-event estimators did reasonably well, however, especially in moderately large samples. The only non-MLE investigated, one that doubles the length of censored spells, was much poorer than any MLE; we recommend that it not be used.

EXTENSIONS TO MORE REALISTIC MODELS

Analysis of event histories with estimators obtained by the method of moments (Sørensen's approach) requires that new esti-

mators be derived for each data structure. Moreover, it is not clear that this method can be readily generalized to handle models in which rates vary over units and over time. The maximum-likelihood approach that we recommend is easily generalized. We have shown in the previous section that the same analytic framework can be applied to a variety of data structures. In this section we outline how it can be extended to permit analysis when (1) rates depend on exogenous variables (allowing for a heterogeneous population), (2) multiple kinds of events can occur, and (3) rates vary over time.

Causal Models of Rates

Sociologists analyzing event histories want to know how the rate at which events occur depends on properties of units and social structure. For example, analysts studying the outbreak of riots in cities have focused on the effects of population size, governmental structure, industrial composition, and the like (see, for example, Spilerman, 1971). Suppose that for each unit we not only know its event history but also its level on a set of observable causal variables, $X_1, X_2, \ldots, X_H$.

To analyze the way in which the X's affect the rates, we require a model linking the two. A log-linear model is a natural candidate because it constrains rates to be nonnegative:

$$\alpha = e^{b_0+b_1X_1+b_2X_2+\cdots+b_HX_H} \tag{41}$$

Then we can evaluate the effects of the X's on the rate in terms of the b's. For a sociological application of this model, see Hannan and others (1977).

For simplicity we concentrate on the special case in which (41) contains only a dummy variable X:

$$\alpha = e^{b_0+b_1X} \tag{42}$$

where X equals 1 if the unit has property A and otherwise equals zero. For example, A could refer to receiving an experimental treatment. Then $X = 1$ for experimentals and $X = 0$ for controls.

Note that (42) can be written as

$$\alpha = e^{b_0}e^{b_1X} = \alpha_0(\alpha_1)^X \tag{43}$$

where $\alpha_0 = e^{b_0}$ and $\alpha_1 = e^{b_1}$. When $X = 0$, $\alpha = \alpha_0$; when $X = 1$, $\alpha = \alpha_0\alpha_1$. So $\alpha_1 = \alpha/\alpha_0$ is the ratio of the rate for those with $X = 1$ to the rate for those with $X = 0$. Given the relationship in (43), in this special case it is easier to estimate α_0 and α_1 than b_0 and b_1.

To obtain an all-events MLE, we merely replace α in (19) with (43):

$$\hat{\alpha}_0 = Y^{(0)}/(U^{(0)} + V^{(0)}) = Y^{(0)}/T^{(0)} \tag{44}$$

$$\hat{\alpha}_1 = (Y^{(1)}/T^{(1)})/(Y^{(0)}/T^{(0)}) \tag{45}$$

where the superscripts 0 and 1 denote values for those with $X = 0$ and $X = 1$, respectively.

Note that the estimated "control" rate (that is, for $X = 0$) is the same as before: the ratio of the total number of events to the total observation time for this group. The estimated multiplier, the relative effect of X, is just the ratio of two terms like (44). In fact, according to (43) the estimated rate for those with $X = 1$ is just

$$\hat{\alpha}_0 \cdot \hat{\alpha}_1 = Y^{(1)}/T^{(1)} \tag{46}$$

This simple extension of the constant-rate model permits an interesting further analysis of the effects of ignoring censored spells. Comparable to the pseudo-MLE in (27), we obtain:

$$\tilde{\alpha}_0 = Y^{(0)}/U^{(0)} \tag{47}$$

$$\tilde{\alpha}_1 = (Y^{(1)}/U^{(1)})/(Y^{(0)}/U^{(0)}) \tag{48}$$

Comparing the pseudo-MLE estimators with the all-event MLE, we see that the situation is just as before. The pseudo-MLE for the "control" rate (that is, for $X = 0$) is upwardly biased:

$$\tilde{\alpha}_0/\hat{\alpha}_0 = T^{(0)}/U^{(0)} \tag{49}$$

The pseudo-MLE for the effect of X, $\tilde{\alpha}_1$, is also biased:

$$\tilde{\alpha}_1/\hat{\alpha}_1 = (T^{(1)}/U^{(1)})/(T^{(0)}/U^{(0)}) \tag{50}$$

This is just the ratio of two ratios, each being the total time to the time spent in spells that end in an event. The greater the degree of censoring, the larger the two ratios. In fact, the ratio in (50) is a measure of the degree of censoring for the group with $X = 1$ to

that for $X = 0$. If X has a positive effect on the ratio (that is, $\alpha_1 > 1$), those with $X = 1$ have a shorter expected duration between events and a smaller degree of censoring. Thus the pseudo-MLE will overstate the causal effect of X. Similarly, when the effect of X is negative (that is, $\alpha_1 < 1$), the bias in $\tilde{\alpha}_1$ is negative. Hence the bias in $\tilde{\alpha}_1$ due to censoring will tend to overstate causal effects.

In the general case when some or all of the causal variables are continuous, the maximum-likelihood estimator does not have an explicit solution. In that case one must obtain the ML estimates of causal effects through some iterative procedure. Tuma and Crockford (1976) have developed a very efficient program for the estimation of general causal models for rates. We refer interested readers to this document for a discussion of the program and its use; see Tuma (1976) and Hannan and others (1977) for substantive applications to censored data.

Models with Multiple Kinds of Events

The problem in most research applications is more complex than we have indicated. In particular, multiple kinds of events may occur. We consider two examples.

Sørensen's (1976) work on censoring was motivated by an interest in the rate at which people voluntarily change jobs. So implicitly there are at least two kinds of events: voluntary and involuntary job shifts. In his analysis Sørensen (1975) ignored those with a theoretically uninteresting event (an involuntary job shift).

Our second example concerns the problem of attrition in panel studies that obtain event histories between successive waves. Usually some people interviewed in earlier waves do not answer questions in later waves. So, for example, in analyzing rates of change in marital status from such data, we must consider three events: marriage, marital dissolution, and attrition from the study. Virtually all reports from panel studies ignore those who leave the study.

From our foregoing analysis, it should not be surprising that excluding the observations on those with an involuntary job shift or on those who drop from a panel study leads to biased estimates of the rates that are of substantive interest: the rate of a voluntary job shift and rates of change in marital status. Consider a three-state model in which all N units begin in the first state. During the period of study they may make transitions between the three states. For simplicity we consider only the first change of state and assume that transition rates are constants. Instead of a single parameter α, we consider α_{12} and α_{13}, which denote the rate of changes from state 1 to 2 and from state 1 to 3, respectively. The rate of leaving state 1 is just $(\alpha_{12} + \alpha_{13})$.

The likelihood function for first events with censored observations for this situation is

$$L = \prod_{i=1}^{N} f(u_i)^{(y_i+z_i)} \cdot G(v_i)^{(1-y_i-z_i)} \cdot \left(\frac{\alpha_{12}}{\alpha_{12}+\alpha_{13}}\right)^{y_i}\left(\frac{\alpha_{13}}{\alpha_{12}+\alpha_{13}}\right)^{z_i} \quad (51)$$

where $f(u_i)$, the probability density of a change after u_i, is just $(\alpha_{12} + \alpha_{13}) \cdot \exp[-u(\alpha_{12} + \alpha_{13})]$; y_i and z_i are (0, 1) variables indicating whether there was a move to state 2 ($y_i = 1$) or state 3 ($z_i = 1$); $G(v_i)$ is the probability of not having had an event over the observation period of length v; and $\alpha_{1j}/(\alpha_{12} + \alpha_{13})$ is the conditional probability of a move to state j, $j = 2$ or 3. Since the rates are constants, the likelihood simplifies to

$$L = \prod_{i=1}^{N} \alpha_{12}^{y_i} \alpha_{13}^{z_i} \left[e^{-(\alpha_{12}+\alpha_{13})u_i}\right]^{(y_i+z_i)} \cdot \left[e^{-(\alpha_{12}+\alpha_{13})v_i}\right]^{(1-y_i-z_i)} \quad (52)$$

The MLEs are

$$\hat{\alpha}_{12} = Y_1/(U_{12} + U_{13} + V_1) \quad (53)$$

$$\hat{\alpha}_{13} = Z_1/(U_{12} + U_{13} + V_1) \quad (54)$$

where U_{1j} is the total time until the first move when the change is from 1 to j; Y_1 is the total number of moves from 1 to 2; Z_1 is the total number of moves from 1 to 3; and V_1 is the total length of censored spells. (The subscript 1 is a reminder that all N cases begin in state 1 by assumption.) The main point of (53) and (54) is that estimates of *both* rates depend on the sum of *all* time prior to the first event or censoring, including the observation time on those who move to what might be considered the theoretically uninteresting state (for example, attrition).

As before we contrast the MLE with a pseudo-MLE that ignores certain observations. This time we consider the consequences of ignoring cases that move to state 3 in estimating α_{12}, the rate of a move from 1 to 2. This is just

$$\tilde{\alpha}_{12} = Y_1/(U_{12} + V_1) \quad (55)$$

Clearly the pseudo-MLE gives higher estimates of the rate than does the MLE in (53). Since we now know that the latter has good statistical properties and, if anything, is upwardly biased, we again conclude that the pseudo-MLE yields upwardly biased estimates of the rate.

This result has important implications for the causal analysis of the rate at which a change in state occurs. We have seen that ignoring censored observations gives biased estimates of the effects of causal variables in a Poisson model. The result generalizes to the multiple-state model. The effects of any causal variables that affect movement from one state to another will be estimated poorly when observations are excluded from the analysis on the basis of the state to which they move. This is an instance of the general phenomenon that selection of units in terms of endogenous variables biases estimates of causal effects. The strategy we have proposed avoids this problem.

Time-Dependent Rates

So far we have assumed that rates are time-independent, though possibly depending on exogenous causal variables. This assumption may often be inappropriate, however. For example, there are historical trends in rates of collective violence, organizational failure, and marital dissolution. Rates may depend on the duration since the previous event: The rate of a voluntary job shift may decline with increasing job tenure, and the rate of marital dissolution may decline with increases in the duration of the marriage. Rates may also depend on age. For instance, organizational death rates are higher for new organizations than old organizations. And most experimental interventions trigger adjustment processes that are time-dependent.

The general strategy that we have discussed can easily be extended to incorporate specific forms of time dependence in rates. One form that we have used involves specifying that rates vary from one delimited period to another but are time-independent within a period. In our application of this form, we estimated a rate of marital dissolution in the first 6 months of an income maintenance experiment and another rate for the subsequent 18 months of data. Experimental effects were allowed to vary from one period to another, but other causal variables were constrained to have the same effect in both periods (see Tuma and others, in press).

Alternatively one can specify a parametric form of time dependence. For example, Tuma (1976) estimated a model in which the rate of leaving a job was a quadratic function of elapsed time on the job. One could also use this procedure to estimate Sørensen's (1975) model in which the rate of voluntary job shifts declines exponentially with labor market experience (or age). Of course, other specifications are possible. A major advantage of the maximum-likelihood approach is the ease with which it is extended to a wide variety of data structures and substantive specifications.

An analyst may expect strong time dependence of rates but have no *a priori* hypothesis about its parametric form or the time points at which rates might change. Then the two strategies mentioned above do not apply. The statistical literature offers some nonparametric alternatives, however. These procedures ignore information on the exact timing of events and instead use information only on the ordering in time of events in the sample. This amounts to assuming that rates vary from period to period—where a period is defined in terms of successive events in the sample—but are constant within any period.

Kaplan and Meier (1958) introduced this approach. They formed a maximum-likelihood estimator of the survivor function $G(t)$. Suppose that in a sample of N units j of them have a nonrepeatable event with times $t_1, \ldots, t_j$ (ordered in time) and $N - j$ have histories censored on the right. When the units are subject to the same stochastic process, elementary probability considerations show that a maximum-likelihood estimator of $G(t)$ for the family of *all* possible distributions is

$$\hat{G}(t) = \prod_{t_i < t} (N - i)/(N - i + 1) \qquad (56)$$

It is straightforward to recover an estimate of $\alpha(t)$ once $\hat{G}(t)$ has been obtained. Turnbull (1974) has generalized the Kaplan-Meier estimator to the case in which histories are censored on both the left and the right. Thus the Kaplan-Meier type of estimator provides a nonparametric alternative to the procedures that we discussed earlier when nonrepeatable events occur to homogeneous units.

This approach has recently been generalized to handle situations in which rates depend on causal variables as well as time. Cox (1972, 1975) has developed a procedure based on a *partial-likelihood* function. The underlying idea can be sketched as follows.[7] Suppose that unit i has a rate proportional to $h(t)\exp(\mathbf{bX}_i)$, where $\mathbf{b}$ and $\mathbf{X}_i$ are vectors of parameters and variables, respectively. Suppose we again arrange data on the timing of events according to Kaplan and Meier's procedure. Then the (complete) likelihood is a product of three parts: (1) the likelihood that events occurred on the observed dates; (2) the likelihood of no events during the periods between the observed events; and (3) the likelihood that an event happened to each individual i, given that an event occurred and individual i was still at risk of the event. If the rate is time-dependent in an unknown way, the first two parts cannot be written explicitly. However, the third part can. Given the set of units at risk (the risk set), the conditional probability that unit i had an event at time t_i is simply

$$e^{\mathbf{bX}_i} \Big/ \sum_{k \in R(t_i)} [e^{\mathbf{bX}_k}] \qquad (57)$$

where $R(t_i)$ is the risk set at t_i. The partial likelihood is obtained by multiplying together j terms like this:

$$\prod_{i=1}^{j} e^{\mathbf{bX}_i} \Big/ \sum_{k \in R(t_i)} [e^{\mathbf{bX}_k}] \qquad (58)$$

Cox has suggested that (58) be treated like an ordinary likelihood function (though it is not) for purposes of estimation and testing. He showed that this leads to consistent estimates of causal effects (the b's). Efron (1977) proved that, under condi-

[7]For a clear and detailed treatment, see Efron (1977).

tions that appear quite general, the partial-likelihood estimator of effects of causal variables is asymptotically normal and reaches the Cramer-Rao lower bound.

Breslow (1974) and Miller (1976) have also extended the Kaplan-Meier approach. They have estimated causal effects within this framework by nonlinear least squares. These estimators also appear to have good properties.

The Kaplan-Meier approach has undergone considerable development in the past few years, and we do not yet know much about the behavior of these estimators in practice. Miller (1976) contrasted the results of partial-likelihood and nonlinear least-squares Kaplan-Meier estimates of the death rate following heart transplants in a small sample. The two estimates gave somewhat different qualitative findings concerning the effects of causal variables. More must be learned about the small-sample behavior of these estimators before choosing among them and the parametric procedures we have described.

SUMMARY

Data on event histories, which record the dates of events, are likely to become increasingly common as sociologists become more aware of their value in studying change over time. These data are almost always censored; that is, they lack information on events that occur before or after the period for which data are available. Unless investigators deal with censoring in a sound way, they are likely to make erroneous inferences about the change process.

We have considered several models of the occurrence of events and several approaches to estimation when event histories are censored. We considered a constant rate (or Poisson) model at length because the methodological issues are more easily understood for this model. We also considered models in which the rate of an event depends on exogenous variables or time and in which there are multiple kinds of events. We discussed approaches to estimation based on maximum likelihood, pseudo-maximum likelihood, the method of moments, and recent work by statisticians on methods that make weak parametric assumptions.

We showed analytically that pseudo-ML approaches to the censoring problem give biased estimates of rates and of causal effects of exogenous variables on rates. We also demonstrated that all but one of the moment estimators of the constant-rate model suggested by Sørensen (1976) are ML estimators, which are known to have excellent large-sample properties. We conducted a Monte Carlo study of the small-sample properties of ML estimators for censored event histories. Our results showed that the various ML estimators have very good properties in small samples when the degree of censoring is small or in medium-sized samples even when the degree of censoring is large. Sørensen's non-ML estimator had a much larger bias than any ML estimator, however, so we advise that it not be used.

An important advantage of the ML approach to the censoring problem is that it is easily extended to different data structures and different models, including those with multiple kinds of events, causal effects on rates, and time-dependent rates. However, it does require specific parametric assumptions. We have reviewed recent statistical developments that deal with censoring and make weak parametric assumptions. There is proof that these new estimators have excellent large-sample properties, but little is yet known about their behavior in small samples or when there is a high degree of censoring.

REFERENCES

BARTHOLOMEW, D. J.

1957 "A problem in life testing." *Journal of the American Statistical Society*, series A, 122:232–239.

1963 "The sampling distribution of an estimate arising in life testing." *Technometrics* 5:361–374.

BOAG, J. W.

1949 "Maximum likelihood estimates of the properties of patients cured by cancer therapy." *Journal of the Royal Statistical Society*, series B, 11:15–53.

BREIMAN, L.

1969 *Probability and Stochastic Processes*. Boston: Houghton Mifflin.

BRESLOW, N.

1974 "Covariance analysis of censored data." *Biometrika* 30:89–99.

BUMPASS, L. L., AND SWEET, J. A.
1972 "Differentials in marital instability: 1970." *American Sociological Review* 37:754–766.

CHIROT, D., AND RAGIN, C.
1975 "The market, tradition, and peasant rebellion: The case of Romania in 1907." *American Sociological Review* 40:428–444.

COX, D. R.
1972 "Regression models and life tables." *Journal of the Royal Statistical Society* series B, 34:187–220.
1975 "Partial likelihood." *Biometrika* 62:269–276.

COX, D. R., AND LEWIS, P. A. W.
1966 *The Statistical Analysis of Series of Events*. London: Methuen.

DHRYMES, P.
1970 *Econometrics: Statistical Foundations and Applications*. New York: Harper & Row.

EFRON, B.
1977 "The efficiency of Cox's likelihood function of censored data." *Journal of the American Statistical Association* 72:557–565.

FELLER, W.
1971 *An Introduction to Probability Theory and Its Applications*. Vol. 2. New York: Wiley.

GLICK, P. C., AND NORTON, A. J.
1971 "Frequency, duration, and probability of marriage and divorce." *Journal of Marriage and the Family* 33:307–317.

GROSS, A. J., AND CLARK, V. A.
1975 *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

HANNAN, M. T., TUMA, N. B., AND GROENEVELD, L. P.
1976 "The impact of income maintenance on the making and breaking of marital unions: An interim report." Center for the Study of Welfare Policy. Research Memorandum 28. Menlo Park, Calif.: Stanford Research Institute.
1977 "Income and marital events: Evidence from an income maintenance experiment." *American Journal of Sociology* 82:1186–1211.

HECKMAN, J. J.
1977 "Sample selection bias as a specification error." Rand Corporation. Report R-1984 (April 1976). Revised 1977.

KAPLAN, E. L., AND MEIER, P.
1958 "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association* 53:457–481.

KENDALL, D., AND BUCKLAND, W. R.
1971 *A Dictionary of Statistical Terms*. New York: Hafner.

MENDENHALL, W., AND LEHMAN, E. H.
1960 "An approximation to the negative moments of the positive binomial useful in life testing." *Technometrics* 2:227–242.

MILLER, R. G.
1976 "Least squares regression with censored data." *Biometrika* 63:449–464.

MORGAN, J. N., DICKINSON, K., DICKINSON, J., BENUS, J., AND DUNCAN, G.
1974 *Five Thousand American Families—Patterns of Economic Progress*. Vol. 1. Ann Arbor: Institute for Social Research.

SHEPS, M. C., AND MENKEN, J. A.
1972a *Mathematical Models of Conception and Birth*. Chicago: University of Chicago Press.
1972b "On estimating the risk of conception from censored data." In T. N. E. Greville (Ed.), *Population Dynamics*. New York: Academic Press.

SHORTER, E., AND TILLY, C.
1974 *Strikes in France*. Cambridge, England: Cambridge University Press.

SNYDER, D., AND TILLY, C.
1972 "Hardship and collective violence in France, 1830–1960." *American Sociological Review* 37:520–532.

SØRENSEN, A. B.
1975 "The structure of intragenerational mobility." *American Sociological Review* 40:456–471.
1976 "Estimating rates from retrospective questions." In D. R. Heise (Ed.), *Sociological Methodology 1977*. San Francisco: Jossey-Bass.

SPILERMAN, S.
1971 "The causes of racial disturbances: Tests of an explanation." *American Sociological Review* 36:427–442.
1976 "Structural characteristics of cities and the severity of racial disorders." *American Sociological Review* 41:771–792.

TUMA, N. B.
1976 "Rewards, resources and the rate of mobility: A nonstationary multivariate stochastic model." *American Sociological Review* 41:338–360.

TUMA, N. B., AND CROCKFORD, D.
1976 "Invoking RATE." Center for the Study of Welfare Policy. Menlo Park, Calif.: Stanford Research Institute.

TUMA, N. B., HANNAN, M. T., AND GROENEVELD, L. P.
In press "Dynamic analysis of event histories." *American Journal of Sociology.*

TURNBULL, B. W.
1974 "Nonparametric estimation of a survivorship function with doubly censored data." *Journal of the American Statistical Association* 69:169–173.

WHITE, H. C.
1970 *Chains of Opportunity: System Models of Mobility in Organizations.* Cambridge, Mass.: Harvard University Press.