Determining the Risk Set for Hazard Functions When Survival Methods are Applied to Social Science and Prevention Data

Lynne A. Malacane

Susan A. Murphy

Linda M. Collins

The Pennsylvania State University

The Methodology Center

Technical Report Series

#97-9

College of Health and Human Development

The Pennsylvania State University

Running head: Determining the Risk Set

Determining the Risk Set for Hazard Functions When Survival Methods Are Applied to Social Science and Prevention Data

Lynn A. Malacane

Susan A. Murphy

Department of Statistics

Linda M. Collins

The Methodology Center and

Department of Human Development and Family Studies

The Pennsylvania State University

Abstract

Survival analysis is becoming increasingly important in the social sciences and prevention. When using survival analysis methods in social science and prevention applications it is important to identify the type of censoring occuring in the data, so that the appropriate risk set can be used in computation of the hazard function. In this article we discuss the two most widely used approaches to computing the hazard function and the type of censoring these methods assume. We point out a type of censoring that occurs frequently in social science studies involving prospective longitudinal panel designs. This *retro-censoring* occurs when a subject's response is necessarily lost in the interval in which the subject is censored. When retro-censoring occurs, the popular methods used in biostatistics and other fields to estimate hazard probabilities will lead to biased estimators and inflated Type I error rates. We suggest a simple way of dealing with retro-censoring and demonstrate by means of a simulation that this approach results in unbiased estimators and correct Type I error rates.

Determining the Risk Set for Hazard Functions When Survival Methods Are Applied to Social Science and Prevention Data

Survival and event history analyses, statistical methods for modeling the amount of time until an event, have long been important analytic tools in biostatistics. Now these analyses are becoming increasingly important in a wide variety of areas within the social sciences and prevention (Willett & Singer, 1993). A few examples are Mensch and Kandel's (1988) study of the relationship between dropping out of high school and drug involvement; Willett and Singer's (1993) study of relapse into cocaine use; Chilcoat and Anthony's (1996) investigation of the effects of parent monitoring on initiation of drug use in children; Bolger, Downey, Walker, and Steininger's (1989) study of suicide ideation in children; and Long, Allison, and McGinnis' (1979) analysis of job changes in assistant professors.

The hazard function plays a central role in survival and event history analyses. In discrete-time survival analysis, the type of survival analysis used most frequently in the social sciences, the hazard function expresses the probability (i.e. risk) that an event will occur during a particular time interval, given that it has not already occured. The hazard function, which is estimated for each time interval in a study, can be inspected for patterns across time. For example, the hazard may increase as time progresses, implying that the subjects who have not yet experienced the event are at ever increasing risk of doing so. Or the hazard may decrease or fluctuate over time. The hazard function is usually plotted as a function of the time intervals, which helps make patterns of increasing and decreasing risk easier to recognize. An example of a hazard function taken from Bolger et al. (1989) appears in Figure 1. Bolger et al. questioned 419 undergraduate students about the age at which they had first contemplated suicide. In Figure

1, each hazard estimate is plotted against the age in years. From the hazard plot in Figure 1, it can be seen that the risk of contemplating suicide for the first time rises from approximately age 11 and peaks at about age 15.

Censoring occurs when the event cannot be observed for all individuals in a study. Usually censoring occurs because data collection ends before the event happens to everyone in the study, or because one or more individuals leave a study. There are several different forms that censoring can take, depending upon the exact characteristics of a study. It is important to identify the type of censoring taking place in a study, because the hazard function must be computed differently for different types of censoring. Intuitively, the hazard function can be estimated by the number of subjects experiencing the event in a given time interval divided by the size of the *risk set*, i.e. the number of subjects at risk of experiencing the event in the same time interval. Only those who have not yet experienced the event can be in the risk set. The numerator, the number of subjects observed to experience the event during a particular interval, is usually simple to determine from a study. However, the denominator of the hazard function may be hard to determine if some people drop out of the study. For example, when studying drug use relapse, suppose 20 subjects out of 100 are observed to relapse during the i^{th} interval and 40 subjects leave the study during the same i^{th} interval. Certainly, the numerator of the hazard estimate for the i^{th} interval should be 20. But should the denominator of the hazard estimate be 100 or 100-40? The choice makes a big difference. In the first case, the estimated hazard is 20/100 = 0.20, but in the second case, the estimated hazard is 20/60 = 0.33! Underestimation of the size of the risk set will overestimate the hazard, and overestimation of the size of the risk set will underestimate the hazard.

Although the various definitions of the risk set are simple, the choice has a profound impact on the computation of the hazard function and thus on the results of a survival analysis. The purpose of this article is to clarify how the risk set definition should be chosen in social science and prevention studies. First, we will discuss the general idea of censoring in survival data. Then, we will show some specific examples of censoring and show how the type of censoring that occurs in a study affects the estimation of the hazard function. We will introduce a form of censoring which we call retro-censoring. Although retro-censoring has not been identified previously, it occurs frequently in longitudinal panel designs like those common in the social sciences and prevention. We will suggest a simple method for dealing with retro-censoring. A simulation study will be presented, which shows that the Type I error rate is considerably inflated when retro-censoring is present and the hazard function is based on an inappropriate risk set.

Approaches to Estimating the Hazard Function Under Various Assumptions about Censoring

In this section we discuss three different approaches to computing the hazard function, each appropriate for a different type of censoring. In our discussion of censoring we will refer several times to Figure 2, which illustrates the various possibilities for an individual in a single time interval in a survival study. The time interval may be defined as beginning and ending at specified calendar dates that are the same for all subjects, or it may begin at different calendar dates for different subjects, although the length of each interval is the same across subjects. Figure 2(a) illustrates a time interval in which the event occurs at some time during the interval and there is no censoring. Figure 2(b) illustrates a time interval where the individual is censored

at some time during the interval, but no event occurs. In Figure 2(c), the event occurs at some time during the interval, and in the remaining time a censoring occurs. In Figure 2(d), the individual is censored at some time during the interval, after which the event occurs in the remaining time. Figure 2(e) illustrates a time interval where neither an event nor a censoring occurs. Note that in some studies observations are timed to coincide with the beginning and end of the intervals, while in other studies observations do not correspond to the beginning and end of the intervals. The timing of the observations in relation to the beginning and end of the intervals is an important consideration determining the type of censoring, as will be shown below.

For the three approaches discussed here, let n_i be the number of subjects who had not yet experienced the response at the beginning of the i^{th} interval. Let d_i be the number of subjects who experience the response during the i^{th} interval. Let c_i be the number of subjects who are censored during the i^{th} interval. All of the methods make the following two assumptions: (1) all time intervals are of equal length; (2) the censoring is independent. By independent censoring we mean that the true hazard probability for individuals who remain in the study during the interval is the same as the true hazard probability for individuals who leave the study during or prior to the interval.

The Kalbfleisch and Prentice Approach

The Kalbfleisch and Prentice approach assumes that censoring can occur only at an end point of an interval. This kind of censoring characterizes Long et al. (1979), who studied male biochemists for five years beginning with their first year as assistant professors until they changed employers for the first time. Their data appear in Table 1. Each interval is a year,

starting at the beginning of the summer and ending at the end of the spring term for each subject. At the end of each interval, Long et al. determined which of the faculty had changed jobs in the previous year. The study ended at the end of the fifth year. The hazard can be estimated to discover when male biochemists who had not yet changed employers were most and least likely to change employers for the first time (Allison, 1982).

Most of the intervals illustrated in Figure 2 can be seen in the Long et al. (1979) study. In this case, no subjects left the study, but censoring occurs because at the end of the study (which is also the end of the fifth interval), the event had not occurred for 129 biochemists who had not yet changed employers. It is known that up to the end of the fifth year the biochemists had not changed employers, but it is not known whether they changed employers later. Thus, for example, the situation in Figure 2(a), an event and no censoring, occurs in year one to eleven people. An example of the situation in Figure 2(e), no event and no censoring, occurs in year two to 189 minus 25 people, or 164 people. The remaining possibilities, Figures 2(b), 2(c), and 2(d), involve censoring. In Long et al. (1979), all censoring takes place at the end of year five, so these three possibilities can occur only in year five. The situation in Figure 2(b), where there is censoring with no event, occurs to 129 individuals at the end of year five (with C placed at the far right end of the interval). The situation in Figure 2(c), where there is an event followed by censoring, occurs to the twelve individuals who change jobs during year five. Because in this study censoring occurs only at the end of the study and thus at the end of an interval, the situation in Figure 2(d) does not appear.

The appropriate method for estimating the hazard function with this type of censoring is discussed by Kalbfleisch and Prentice (1980) and Efron (1988). This approach is commonly

used in discrete-time survival analysis studies in social science contexts such as those of Singer and Willett (1993; Willett & Singer, 1993). This method assumes that each censoring in the i^{th} interval occurs only at the endpoint of the i^{th} interval. That is, the possibilities illustrated in Figure 2(b) and 2(c) cannot occur unless the censoring occurs at the end of the interval as it does in Long et al. (1979). In other words, this method considers censored subjects as at risk for the entire interval.

The risk set is the number of subjects who have not yet experienced the response at the beginning of the i^{th} interval. So the size of the risk set is n_i, and therefore, the estimate of the hazard probability for the i^{th} interval is simply

$$\frac{d_i}{n_i}. \tag{1}$$

Allison (1982, 1984) uses this estimator for the hazard in his discussion of the Long et al. (1979) study. For example, to estimate the hazard probability for the fifth year note that the risk set for the fifth year is 141 biochemists and therefore,

$$\frac{d_1}{n_i} = \frac{12}{141} = 0.0851.$$

This hazard probability estimates that 8.51 percent of male biochemists working as assistant professors change employers for the first time in their fifth year given that they have not changed during their first four years.

The Actuarial Approach

The actuarial approach is commonly used when individuals can be censored in the middle of an interval. For example, Berkson and Gage (1950) conducted a randomized trial to compare the effects of two treatments, A and B, on patients with head-and-neck cancer. The event is death occuring after treatment randomization. The survival data for the 51 patients in arm A of the head-and-neck cancer study conducted by the Northern California Oncology Group, discretized into half-year time intervals following treatment, appears in Table 2. Each time interval runs from one six month anniversary to the next six month anniversary of the treatment. The hazard function based on these data may be used to determine when the surviving patients in arm A were least and most at risk of dying in the half years following treatment.

The censoring in this prospective study occurs for two reasons. First, some living patients leave the study. For example, Table 2 shows that at some point during the third half-year following treatment, a patient left the study. It is known that this patient had not died up to the beginning of the third half-year. It can also be assumed that up to the time the patient left the study during the third half-year following treatment, the patient had not died, because it would be very unusual for physicians to lose track of the death record of a patient who died while in their care. Thus, the situation illustrated in Figure 2(c) is very unlikely. As a result, we can be sure that the one person who left the study experienced either the situation in 2(b) or 2(d) in the third half-year. That is, the individual either died sometime after leaving the study but before the end

of the third half-year (Figure 2(d)) or sometime after the end of the third half-year (Figure 2(b)). The second cause of censoring in this study is the data collection design itself. The intervals begin and end at different calendar times for different subjects, but the study was declared ended and data collection ceased on a particular date for all subjects. For most subjects, this date fell sometime in the middle of an interval. Any subject alive at the beginning of the interval was censored when the study was declared ended. If the subject lived to the end of the entire interval, this corresponds to Figure 2(b). If the subject died before the date the study ended and in the same interval as the end of the study, this corresponds to Figure 2(c). If the subject died after the study ended but before the end of the interval, this corresponds to Figure 2(d). The situations depicted in Figures 2(a) and 2(e), which do not involve censoring, also occur in these data.

A different example of this type of censoring is provided by Bolger et al. (1989), who questioned undergraduate students to determine the first age at which they had contemplated suicide. In this retrospective study, the event is first contemplation of suicide. Each time interval is one year, running from birthday to birthday for an individual. However, the subjects were not interviewed on their birthdays, so because the interview took place in the middle of an interval, censoring took place immediately after the interview for those subjects who had not had suicidal thoughts up to the time of the interview. The data appear in Table 3. For example, at the time of questioning thirty-one 19 year old students had not yet contemplated suicide. Thus censoring occurs because the event has not yet happened at the time of the data collection. It is known that the 31 students had not contemplated suicide up to the time of data collection, but it is unknown whether the students contemplated suicide at a later date. Indeed for these 31

students, either situation 2(b) or 2(d) occurs in their 19th year. In this study, all of the intervals depicted in Figure 2 are possible. Assuming that data collection does not occur on an individual's birthday, Figure 2(b) occurs when an individual has never contemplated suicide up to the time of data collection, and does not contemplate it up to the end of the interval (i.e. until his or her next birthday); Figure 2(c) occurs when an individual contemplates suicide for the first time in the interval containing the data collection, and before the data collection; Figure 2(d) occurs when an individual contemplates suicide for the first time in the interval containing the data collection, but between the data collection and his or her next birthday.

In both of the above studies an interval with a censoring is observed to contain an event if the event occurred prior to the censoring. If the event occurs after censoring in the interval then the event is not observed. This is because in the first study it is implausible that the subject would "disappear" after experiencing the event of death and in the second study the researcher would not be aware of an event occurring after the censoring (time of data collection). Because a subject censored in an interval was at risk of experiencing the event prior to the censoring, the subject should be included in the risk set in some way. Many authors (e.g. Berkson & Gage, 1950; Blossfield, Hamerle, & Mayer, 1989; Cutler & Ederer, 1958; Marubini & Valsecchi, 1995; Thompson, 1977) make the simplifying assumption that the censoring occurs uniformly throughout the interval, and as a result, *on average* the subject is at risk of an observable event for about one half of the interval.

As discussed by the above authors, this type of censoring can be handled by means of the actuarial method. The actuarial method assumes that censoring occurs uniformly in the interval and therefore on average at the midpoint of each interval. Hence, censored subjects can only

contribute to the risk set for approximately half of the interval because they can only contribute to the numerator of the hazard for approximately half of the interval. This method assumes that if Figure 2(c) occurs, the event is observed and recorded. The risk set is the number of subjects who have not yet experienced the response at the beginning of the i^{th} interval minus half of the censored subjects from this interval. Hence, the size of the risk set is $n_i - \frac{1}{2}c_i$, and the actuarial method estimates the hazard probability for the i^{th} interval as

$$\frac{d_i}{n_i - \frac{1}{2}c_i}. \tag{2}$$

Because the response in the Berkson and Gage cancer survival example is death, one may reasonably assume that if a subject is planning to leave the study in an interval or the study ends during a six month interval, a death prior to this will be known to the researcher. Furthermore, censoring does not have to occur at the end of the interval but may occur throughout each interval, so to estimate the hazard probabilities the actuarial method should be used. To estimate the hazard probability for the second half-year, the size of the risk set is

$$n_i - \frac{1}{2}c_i = 25 - \frac{1}{2}(3) = 23.5.$$

The estimated hazard probability is then

$$\frac{d_i}{n_i - \frac{1}{2}c_i} = \frac{7}{23.5} = 0.2979.$$

We estimate that 29.79 percent of the head-and-neck cancer patients in arm A will die during the second half-year given that they have lived for the first half-year.

The actuarial method is also appropriate for use in the Bolger et al. suicide contemplation example. The retrospective nature of the suicide contemplation example implies that if the student contemplates suicide in the same interval as the interview date but prior to the interview then the researcher learns of this. Recall that censoring occurs immediately after the interview rather than at the end of each age interval. Hence, the actuarial method should be used to estimate the hazard probabilities. This method gives the size of the risk set for age 19 as

$$n_i - \frac{1}{2}c_i = 109 - \frac{1}{2}(31) = 93.5$$

and the corresponding hazard probability as

$$\frac{d_i}{n_i - \frac{1}{2}c_i} = \frac{11}{93.5} = 0.1176.$$

We estimate that 11.76 percent of the students will contemplate suicide for the first time while they are 19 years old given that they have not contemplated suicide previously.

Retro-Censoring

Retro-censoring, which has not been identified previously, is different from the other types of censoring we have discussed. In retro-censoring, a subject may leave the study in the same interval in which he or she experiences the event. This corresponds to Figures 2(c) and 2(d). However, this censoring is very different from that occuring in Berkson and Gage (1950) and Bolger et al. (1989), because there is no possibility of observing an event that occurs in the same interval as the censoring but before the censoring, as illustrated in Figure 2(c). This type of censoring is particularly common in longitudinal panel studies like those often carried out in the social sciences. For example, in the AAPT study, Panel 2 (Hansen & Graham, 1991), students were given a questionnaire yearly from sixth grade through ninth grade, and the information from these questionnaires was used to determine when they began drug experimentation (see Table 4). Each time interval begins and ends with a yearly questionnaire. The event is defined to be initial use of alcohol, cigarettes, or marijuana. The hazard for this study may be used to determine when initiation of drug use is most likely to occur among never-users. This information may then be used to determine the best time to conduct a drug prevention program.

In this example, 98 of the children who had not reported prior drug use and were present for the interview in seventh grade were not present in eighth grade for the interview. These 98 children could have experienced either 2(b), 2(c) or 2(d) in the 7^{th} grade. This is a very different kind of censoring from the Cancer Survival example. In the Cancer Survival example, more is known about the individual who drops out of a study. In that case it is known that up to the time

the individual leaves the study during an interval, the event has not occurred. In contrast, it *is* possible that some of the 98 children in the AAPT study initiated drug use before they left the study (Figure 2(c)). So in the Cancer Survival example, the patient censored in the third half-year contributes some information about the third half-year, but in the Drug Use Initiation example, the 98 children censored in the seventh grade contribute *no* information about initial drug use in seventh grade.

Both the Kalbfleisch and Prentice method and the actuarial method are inappropriate for estimating the hazard function in data involving retro-censoring. The Kalbfleisch and Prentice method assumes that censoring takes place only at the end of the interval, which is not true of retro-censoring. The actuarial method allows censoring to occur in the middle of the interval, but it is assumed that if Figure 2(c) occurs, the event is observed by the researcher, which is an invalid assumption when retro-censoring is present. Because neither of these methods is suitable for situations where retro-censoring is possible, a new method for estimating the hazard is needed.

A simple solution to the problem of retro-censoring. An individual is retro-censored in an interval, if the individual leaves the study during the interval and the researcher does not know if an event occurred in the interval either prior to the individual leaving or after the individual leaves. To deal with the problem of retro-censoring, we may code the censoring as occurring at the end of the previous interval. Technically this means that a subject who is present for the i^{th} interview, is not present for the $i+1^{th}$, and does not report experiencing the event at the i^{th} interview, is considered censored at the i^{th} interview. This method allows all five possibilities in Figure 2 to occur, but assumes that in the situations illustrated in Figure 2(c) and 2(d) the

response is unknown. The risk set is the number of subjects who have not yet experienced the response at the beginning of the i^{th} interval minus all the censored subjects from the i^{th} interval. The size of the risk set is $n_i - c_i$ because censored subjects can contribute no information to the study about their responses. The estimated hazard probability for the i^{th} interval according to this method is

$$\frac{d_i}{n_i - c_i} \tag{3}$$

It is tempting to try to use the censored subjects in the risk set somehow, but this will lead to incorrect hazard estimates. The risk set, n_i, is not used because censored subjects are not at risk for the entire interval, and therefore, should not be included in the risk set. And if a subject responds and then is censored, the response cannot be observed. Therefore, $n_i - \frac{1}{2}c_i$ cannot be used as the risk set. Both of these risk sets will result in underestimation of the hazard probability.

This third method for retro-censoring should be used to estimate the hazards for the Drug Use Initiation example. In this example, it is possible to begin using drugs and then be censored without the drug use being observed. There are 98 students who are present at the 7th grade data collection but not at the 8th grade data collection. The size of the risk set for the interval between seventh and eighth grade using the retro-censoring method is

$$n_i - c_i = 401 - 98 = 303$$

So the estimated hazard for the seventh grade is

$$\frac{d_i}{n_i - c_i} = \frac{142}{303} = 0.4686$$

This hazard probability estimates that 46.86 percent of the students will begin using drugs during the interval between seventh and eighth grade given they have not begun using drugs earlier. Similarly, the 76 students who are present at the 6th grade but not at the 7th grade should be coded censored at the 6th grade, the 13 students who are present at the 8th grade but not at the 9th grade should be coded as censored at the 8th grade, and the 102 students who are present at the last interview in 9th grade and who have not initiated drug use should be coded as censored at the 9th grade. With this coding the SAS program given by Willet and Singer (1993) to estimate the hazard probabilities may be used.

Figure 3 compares the hazard estimates using the retro-censoring method, the actuarial method, and the Kalbfleisch-Prentice method. The Kalbfleisch-Prentice method assumes that censoring occurs after all responses in an interval and uses n_i as the risk set. This assumption is inappropriate because it is possible for a student to begin drug use and move or drop out of school before their response is observed. Similarly, the actuarial method is also inappropriate for the Drug Use Initiation example because censoring may occur after an unknown response. Both the Kalbfleisch-Prentice and the actuarial method overstate the risk set and underestimate the hazard for all three grades.[1]

Effects of Using an Incorrect Method on Retro-Censored Data

A simulation study was performed to demonstrate the effects of using an incorrect method on retro-censored data. If the data are subject to retro-censoring and the

Kalbfleisch-Prentice or actuarial method is used, then as discussed earlier, the estimator of the true hazard probability will be negatively biased due to the overestimated risk set. This bias becomes more negative as either the true hazard probability or the true proportion of censorings increases.

Method

We focus on one typical interval and evaluate the bias in the estimated hazard probability as a function of the true hazard probability and the true censoring probability. At the start of the interval n individuals are present. The true hazard probability for the interval is denoted by p, and the probability of being censored at any time in the interval is denoted by pc. We only allow retro-censoring. That is, if an individual is censored in the interval we have no knowledge of the presence of an event. This description is equivalent to n Bernoulli random variables, each with probability of success equal to p and then simulating n further Bernoulli random variables, each with probability of success pc. We pair these, resulting in n pairs of Bernoulli random variables. If the second member of the pair is a one, the subject is censored and we discard the first member of the pair. If the second member of the pair is zero and the first member of the pair is a one then the subject is observed to experience the event. If the second member of the pair is zero and the first member of the pair is zero then the subject is in the study for the entire interval, but no event occurs. In this case, d is the number of ones in the remaining first member of the pairs, and c is the number of ones in the second member of the pairs. This was done ten thousand times in SAS. The sample size, n, took on values 50, 100, 200, 300, 400, 500, and 600. Possible values of p were 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 and possible values of pc were 0.1, 0.2, 0.3, and 0.4.

Results

For each of the ten thousand samples, the hazard was estimated using the retro-censoring, Kalbfleisch-Prentice, and actuarial methods. From each method's estimate, a 95 percent confidence interval was calculated. Then, each confidence interval was checked to see if it captured the true hazard probability. The error rate or the percent of confidence intervals out of 10,000 that did not capture the true hazard probability was calculated for each method. Of course, a 95 percent confidence interval should have an error rate of 5 percent.

Table 5 gives error rates for a representative selection of sample sizes, true hazard probabilities and true censoring probabilities. Results were highly consistent across the conditions not shown. Note that in general, the error rate for the retro-censoring method is close to .05, whereas the error rates for the other two methods are often wildly different from .05. Table 6 shows the sample sizes for which the confidence intervals based on the Kalbfleisch-Prentice method resulted in at least a doubling of the error rate from 0.05 to 0.1. Once *pc* is greater than or equal to 0.3, any sample size greater than 50 will result in at least a doubling of the error rate no matter what *p* is. Similarly, once *p* is greater than or equal to 0.3, any sample size greater than 50 will result in at least a doubling of the error rate no matter what *pc* is. Blank entries correspond to cases in which the error rate was not double in the largest simulated sample size of 600. Table 7 shows the sample sizes for which the confidence intervals based on the actuarial method resulted in at least a doubling of the error rate from 0.05 to 0.1. Once *pc* is greater than or equal to 0.4, any sample size greater than 50 will result in at least a doubling of the error rate no matter what *p* is. Also, once *p* is greater than or equal to 0.9, any sample size greater than 50 will result in at least a doubling of the error rate. As before blank

entries correspond to cases in which the error rate was not at least double in the largest simulated sample size of 600.

Tables 5, 6, and 7 show that in general, for a fixed sample size and a fixed true proportion of censorings, as the true hazard probability increases, the error rate increases. This increase in the error rate occurs because the bias becomes more and more negative as the true hazard probability or p increases. Using similar reasoning, for a fixed sample size and a fixed true hazard probability, as the true proportion of censorings increases, the error rate increases. Finally, for a fixed true hazard probability and a fixed true proportion of censorings, as the sample size increases, the error rate increases. This increase in the error rate occurs because as the sample size or n increases the estimated variance is reduced allowing the bias to show more.

Discussion

Survival analysis is a technique that has proven highly useful in biostatistics and now is becoming increasingly popular in the social sciences and prevention. The choice of method for computing the hazard function in a study using survival analysis is a very important one. In order to make this decision, the researcher must have a clear idea about the assumptions each alternative makes about censoring, and about how censoring is operating in the study in question. It is important to identify whether censoring can occur throughout each interval and whether a subject's response will be lost if the subject is censored in a paticular interval. In this article we clarify these issues for social scientists, and point out a type of censoring that occurs frequently in social science data, called retro-censoring.

The Kalbfleisch and Prentice method is appropriate for studies where censoring occurs only at the end of an interval, as illustrated in Long et al. (1970). Where censoring occurs only at

the end of an interval, there can be no retro-censoring. The actuarial method is appropriate for studies where censoring occurs in the middle of the interval, as long as there is no retro-censoring, as illustrated in Berkson and Gage (1950) and Bolger et al. (1989). Retro-censoring, as illustrated in Hansen and Graham (1991), violates the assumptions of these two commonly used methods for estimating the hazard. This violation is potentially serious. In a simulation study, we showed that when the Kalbfleisch and Prentice and actuarial methods are applied to data where retro-censoring is present, the true Type I error rate can be well above the nominal rate. This is particularly a problem when N is large, the hazard probability is large, and/or the censoring probability is large. However, if the procedure for computing the hazard function recommended in this article is followed, the Type I error rate remains very close to the nominal rate.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), Sociological methodology (pp. 61-98). San Francisco: Jossey-Bass.

Allison, P. D. (1984). Event history analysis: Regression for longitudinal event data. Beverly Hills, CA: Sage.

Berkson, J., & Gage, R. R. (1950). Calculation of survival rate for cancer. In Proceedings of Staff Meetings, 25 (pp. 270-286). Mayo Clinic.

Blossfield, H. P., Hamerle, A., & Mayer, K. U. (1989). Event history analysis: Statistical theory and application in the social science. Hillsdale, NJ: Lawrence Earlbaum Associates.

Bolger, N., Downey, G., Walker, E., & Steininger, P. (1989). The onset of suicide ideation in childhood and adolescence. Journal of Youth and Adolescence, 18, 175-189.

Chilcoat, H. D., & Anthony, J. C. (1996). Impact of parent monitoring on initiation of drug use through late childhood. Journal of the American Academy of Child and Adolescent Psychiatry, 35, 91-100.

Cutler, S. J., & Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. Journal of Chronic Disease, 8, 699-712.

Efron, B. (1988). Logistic regression, survival anlaysis, and the kaplan-meier curve. Journal of the American Statistical Association, 83, 414-425.

Gehan, E. A. (1969). Estimating survival functions from the life table. Journal of Chronic Diseases, 21, 629-644.

Hansen, W. B., & Graham, J. W. (1991). Preventing alcohol, marijuana, and cigarette use among adoelscents: Peer pressure resistance training versus establishing conservative norms. Preventive Medicine, 20, 414-430.

Kalbfleisch, J. D., & Prentice, R. L. (1980). The statistical analysis of failure time data. New York: John Wiley & Sons.

Lee, E. T. (1992). Statistical methods for survival data analysis. New York: John Wiley & Sons.

Long, J. S., Allison, P. D., & McGinnis, R. (1979). Entrance into the academic career. American Sociological Review, 44, 816-830.

Marubini, E., & Valsecchi, M. G. (1995). Analysing survival data from clinical trials and observational studies. Chichester, England: John Wiley & Sons.

Mensch, B. S., & Kandel, D. B. (1988). Dropping out of high school and drug involvement. Sociology of Education, 61, 95-113.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18, 155-195.

Thompson, W. A., Jr. (1977). On the treatment of grouped observations in life studies. Biometrics, 33, 463-470.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. Journal of Consulting and Clinical Psychology, 61, 952-965.

Yamaguchi, K. (1991). Event history analysis. Newbury Park, CA: Sage.

Author Notes

Lynn A. Malacane, Department of Statistics, The Pennsylvania State University.

Susan A. Murphy, Department of Statistics, The Pennsylvania State University.

Linda M. Collins, The Methodology Center and the Department of Human Development and Family Studies, The Pennsylvania State University.

This research was supported by grant, A P50 DA 10075-01 from the National Institute on Drug Abuse.

Correspondence concerning this article should be addressed to Susan A. Murphy, 326 Thomas Building, Department of Statistics, The Pennsylvania State University, University Park, PA 16802.

Footnotes

[1]A fourth method for defining the risk set, the clinical life table method, is discussed by Gehan (1969), Lee (1992), and Yamaguchi (1991). In this method, the goal is to estimate the hazard *rate* at the midpoint of the interval rather than to estimate the hazard probability for the interval as in the above methods. Because our goal is to estimate the hazard *probability* for an interval rather than the hazard rate at the midpoint of an interval, we do not discuss this method.

Table 1

Data from Long et al. (1979)

Year	Number of male biochemists who had not changed employers for the first time up to the beginning of the i^{th} year	Number of male biochemists who change employers for the first time during the i^{th} year	Number of male biochemists who are absent at the end of the i^{th} year
1	200	11	0
2	189	25	0
3	164	10	0
4	154	13	0
5	141	12	0
>5	129		

Table 2

Data from Berkson and Gage (1950)

Half-Year	Number of subjects who are alive at the beginning of the i^{th} half-year	Number of subjects who die during the i^{th} half-year	Number of subjects who are absent at the end of the i^{th} half-year
1	51	25	1
2	25	7	3
3	15	5	1
4	9	2	0
5	7	0	0
6	7	0	0
7	7	2	2
8	3	1	2

Table 3

Data from Bolger et al. (1989)

Age in Years	Number of students who had not contemplated suicide up to the beginning of the i^{th} age	Number of students who first contemplated suicide during the i^{th} age	Number of students who were the i^{th} age at the time of the interview and had not yet contemplated sucide
6	417	2	0
7	415	3	0
8	412	13	0
9	399	8	0
10	391	24	0
11	367	9	0
12	358	45	0
13	313	44	0
14	269	31	0
15	238	37	0
16	201	21	2
17	178	17	11
18	150	18	23
19	109	11	31
20	67	3	23
21	41	1	40

Table 4

Data from Hansen and Graham (1991)

Grade	Number of students who do not report drug use up to the beginning of the i^{th} grade	Number of students who begin drug use during the i^{th} grade	Number of students who are absent at the end of the i^{th} grade
6	711	234	76
7	401	142	98
8	161	46	13
>8	102		

Table 5

Error Rates for the Retro-Censoring, Kalbfleisch-Prentice, and Actuarial Methods Applied to Data Subject to Retro-Censoring

Sample Size	True Hazard Probability	True Censoring Probability								
		.1			.2			.3		
		R-C	K-P	A	R-C	K-P	A	R-C	K-P	A
50	.1	.07	.07	.07	.08	.09	.09	.05	.13	.13
	.3	.05	.06	.11	.05	.09	.21	.06	.13	.38
	.5	.06	.07	.13	.05	.12	.34	.05	.20	.62
300	.1	.05	.07	.10	.05	.11	.23	.05	.17	.47
	.3	.05	.09	.20	.06	.24	.64	.05	.48	.95
	.5	.05	.15	.43	.05	.45	.94	.05	.82	1.0
500	.1	.05	.07	.12	.06	.14	.45	.06	.25	.67
	.3	.05	.12	.33	.05	.36	.92	.05	.69	1.0
	.5	.05	.20	.62	.05	.67	1.0	.05	.96	1.0

Note. R-C denotes retro-censoring; K-P denotes Kalbfleisch-Prentice; A denotes actuarial.

Table 6

Sample Sizes for which the 95% Confidence Intervals Based on the Kalbfleisch-Prentice Estimator Result in at Least a Doubling of the Error Rate in Data Involving Retro-Censoring

		p									
		.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
pc	.1		≥400	≥200	≥50	≥50	≥50	≥50	≥50	≥50	≥50
	.2	≥300	≥200	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50
	.3	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50
	.4	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50

Table 7

Retro-Censoring Sample Sizes for which the 95% Confidence Intervals Based on the Actuarial Method Result in at Least a Doubling of the Error Rate

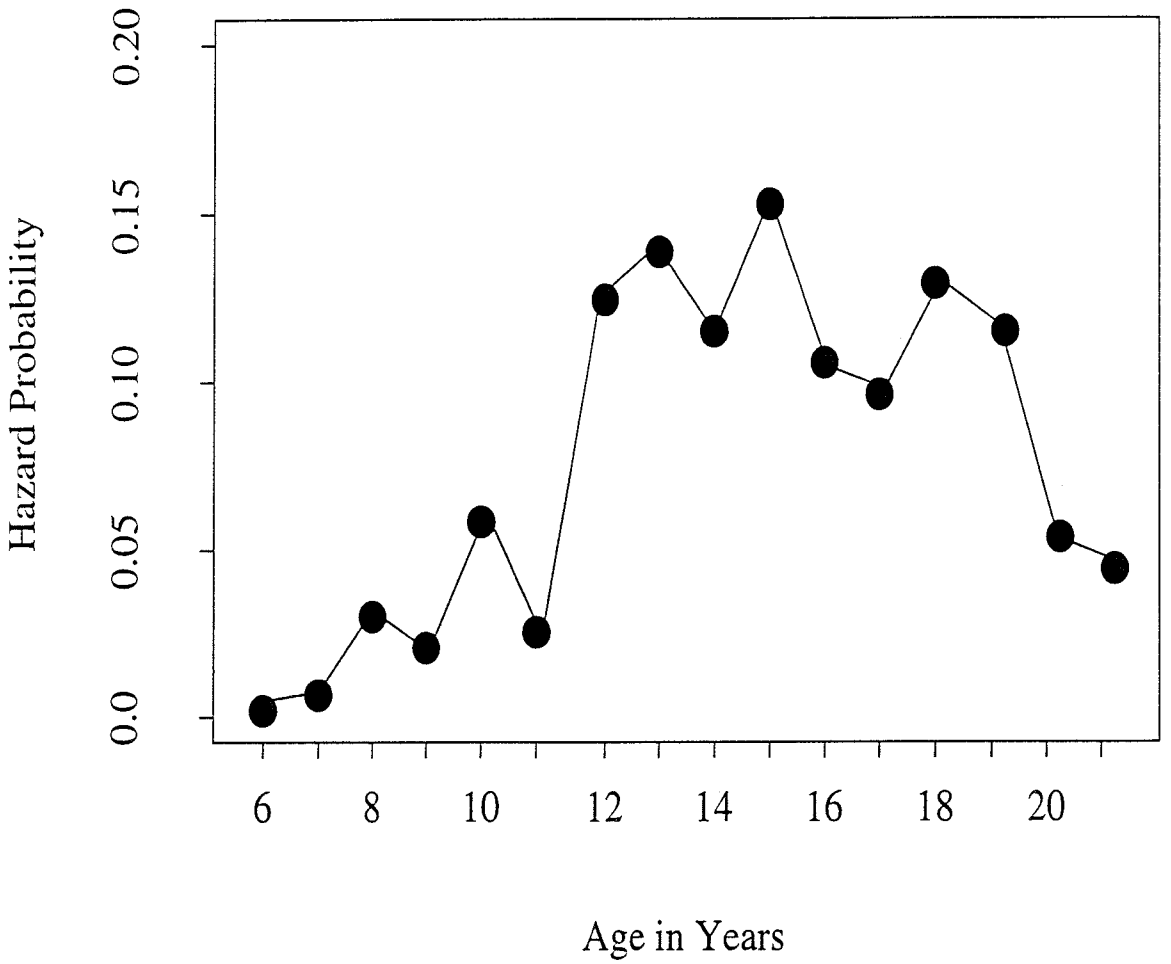
		p									
		.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
	.1			≥600	≥400	≥300	≥200	≥200	≥100	≥100	≥50
pc	.2	≥500	≥300	≥200	≥100	≥50	≥50	≥50	≥50	≥50	≥50
	.3	≥300	≥200	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50
	.4	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50	≥50

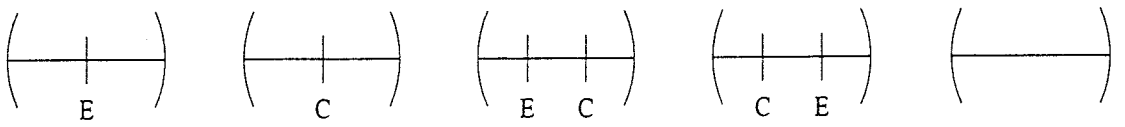
Figure Captions

Figure 1. Hazard estimates versus age in years.

Figure 2. Five possibilities for the i^{th} interval.

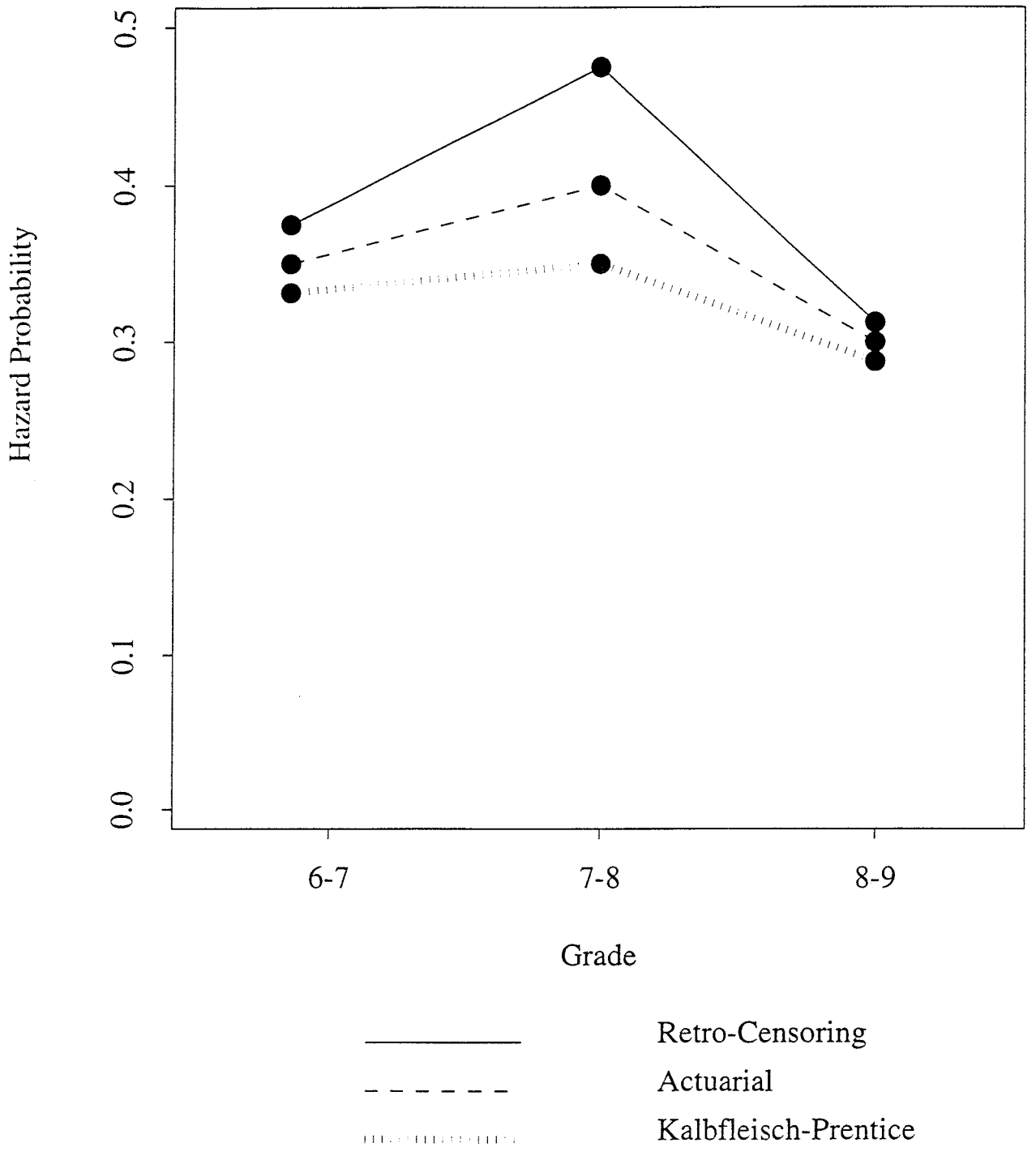
Figure 3. Retro-censoring, actuarial and Kalbfleisch-Prentice methods compared for the drug use initiation example.


Hazard Probability
0.0
0.05
0.10
0.15
0.20
6
8
10
12
14
16
18
20
Age in Years



(a) (b) (c) (d) (e)

E = Event C = Censored


Hazard Probability
0.0
0.1
0.2
0.3
0.4
0.5
6-7
7-8
8-9
Grade
Retro-Censoring
Actuarial
Kalbfleisch-Prentice