16

Event History Analysis

PAUL D. ALLISON

Event history analysis is a term commonly used to describe a variety of statistical methods that are designed to describe, explain or predict the occurrence of events. Outside the social sciences, these methods are often called *survival analysis*, owing to the fact that they were originally developed by biostatisticians to analyze the occurrence of deaths. But despite their biomedical origin, these same methods are perfectly suitable for studying a vast array of social phenomena such births, marriages, divorces, job terminations, promotions, arrests, migrations, and revolutions. There also many other names for event history methods, including failure time analysis, hazard analysis, transition analysis, and duration analysis.

In general, an event may be defined as a *qualitative* change that occurs at some particular point in time. To apply event history methods you need event history data – a longitudinal record of when events occurred to some individual or sample of individuals. For example, if you ask a sample of women to report the birth dates of all their children, you will get a set of event history data that will allow you to analyze the occurrence of births. Of course, if you want to do a causal or predictive analysis, you would also want to measure possible explanatory variables, such as the woman's date of birth, race, education, family income, marital status, and so on. Some of these may be constant over time (like race or region of origin) while others (like marital status and income) may vary with time. As we shall see, the distinction between time-constant and time-varying explanatory variables can be very important in selecting a method of analysis.

Some kinds of event history analysis allow for repeated events and different kinds of events. But it is helpful to postpone these complications until we have dealt with the simpler situation in which each individual experiences no more than one event, and all events are assumed to be of the same type. The classic example is where the event of interest is a death and we do not distinguish different kinds of deaths.

PROBLEMS WITH CONVENTIONAL METHODS: RECIDIVISM EXAMPLE

To appreciate the virtues of event history analysis, it is helpful to consider the problems that arise in attempting to apply conventional methods (like linear regression) to the analysis of event history data. Here is an example. In the early 1970s, researchers conducted a field experiment on 432 inmates who were released from Maryland state penitentiaries (Rossi et al., 1980). Half of them were randomly assigned to receive financial aid (roughly equivalent to unemployment compensation) for the first three months after their release. The other half got no money. The goal was to determine whether financial aid would reduce the likelihood of arrest. At the time of their release, the subjects completed an extensive interview about their past history. They were then followed for one year after their release, and the event of interest was the first arrest, the date of which was obtained from police records. Only about 25% of the inmates were arrested. During the one-year follow-up, they were also interviewed at regular intervals to ascertain changes in employment status, marital status, and so on.

Now, how should one analyze such data? An obvious approach would be to do a logistic regression in which the dependent variable is whether or not a person was arrested. The independent variables could include such things as receipt of financial aid, years of schooling, age at release, and number of prior convictions. This would not be a bad method, but neither is it ideal. For one thing, the method does not make use of the timing of the arrests. It is reasonable to suppose that, on average, people who were arrested in the first week after being released had a higher propensity toward crime than those who were arrested near the end of the one-year period. But the logistic regression treats them as identical.

A potential solution to this problem is to use the length of time from release to first arrest as the dependent variable, and do an ordinary linear regression instead of a logistic regression. That might work well if all the released inmates were arrested but, as already noted, only 25% were arrested during the one-year follow-up. What should be done with the other 75% who were not arrested? In the parlance of event history analysis, cases that do not experience the event during the period of observation are called 'censored'. Virtually all event history data contain some censored observations, and all methods that claim the title of event history methods are designed to deal with censoring in one way or another.

If the number of censored cases is small, it is tempting to exclude them from the analysis. But few researchers would want to lose 75% of the cases. Alternatively, one could assign the maximum time observed – one year in this example – as the value of the dependent variable for censored cases. But clearly this is an underestimate. It has been shown that both of these *ad hoc* methods can lead to substantial biases (Sørensen, 1977; Tuma and Hannan, 1978).

Regardless of whether you do logistic regression of arrest status or linear regression of time to arrest, there is another problem that is even more serious than censoring: How do you include variables that vary over the one-year follow-up period? Take employment status, for instance. In each of the 52 weeks of observation, we know whether the person was employed full-time or not. One possibility would be to include 52 dummy variables for employment status in the regression model. Aside from being unwieldy, however, this would raise the possibility of reverse causation. If someone is arrested in the 10th week and incarcerated as a result, that could have a big impact on whether he is employed in the 12th week. The result would potentially be a large bias in estimating the effect of employment status on arrest. To my knowledge, there is no *ad hoc* solution to this problem that even comes close to the event history methods we will consider shortly.

To sum up, event history data typically have two characteristics that make conventional methods unsuitable: censoring and time-varying explanatory variables (also known as time-dependent covariates). All event history methods deal with censoring in some way. Some also deal with time-varying explanatory variables. Before examining these methods in more detail, we need to take a closer look at various kinds of censoring.

CENSORING

Censoring can take several different forms. In the recidivism example we have just considered, all the censored cases were *right censored*. Or more accurately, their *event times* were right censored. An event time is said to be right censored if all we know is that is greater than some number c, called the censoring time. For the recidivism example, if arrest times are measured in weeks from release, then $c = 52$. In this case, the censoring time is the same for everyone, so the data are described as *singly right censored*. In many other data sets, however, the censoring times (or potential censoring times) vary across observations. This could happen, for example, if prisoners are released at different points in calendar time, but everyone is followed up until some particular date in calendar time. Those released earlier have longer potential censoring times than those released later.

This variation in censoring times is relatively unproblematic if censoring occurs simply because the researcher stops the follow-up according to some prespecified rule. On the other hand, we also treat observations as right censored if the follow-up stops for reasons that are not under control of the researcher. For example, people may die, move away, or refuse to continue participating in the study. Censoring of this sort is called *random censoring*, but it is important to understand that in this context random does not mean that the censoring is unrelated to anything else. Rather, it means that the censoring is part of the phenomenon under investigation, not a part of the research design.

Random censoring *is* potentially problematic. Conventional event history methods implicitly assume that random censoring is *noninformative*. This means that the fact that an individual is censored at a certain point in time does not provide any information about that individual's risk of experiencing the event. In the recidivism example, suppose that some of the released prisoners died during the one-year follow-up period (in fact, none did). The usual approach would be to treat their arrest times as censored at the time of death. This censoring would be noninformative if those who were censored by death had the same risk of arrest as those who did not die.

Unfortunately, there are many situations in which it is not plausible to assume that censoring is noninformative. More unfortunate is the fact that there is no way to test this assumption. Worse still, even if one is certain that the assumption is violated, there is no generally acceptable way to correct for such violations. So we are stuck with using the conventional methods despite the fact that they may produce somewhat biased estimates. The lesson here is that, in designing and executing the collection of event history data, one should do everything possible to minimize random censoring.

Although right censoring is by far the most common kind of censoring, some event history data may also have left censoring. An event time is said to be left censored if all we know about it is that it is less than some number c. For example, suppose we do a prospective study of intravenous drug users with the aim of determining when and if they contract HIV. At the onset of the study, however, some users are found to have already contracted the disease, and we have no way of knowing when that happened. These cases are left censored. Most event history methods, like Cox regression, are not designed to handle left censoring.

Finally, an observation is called *interval censored* if we know that an event occurred between time a and time b, but we do not know exactly when it happened within the interval. This kind of censoring is also quite common. For the intravenous drug user example, the study might administer blood tests at six-month intervals. If a particular person is HIV-negative at one screening but HIV-positive at the next, the time of the event is interval censored. If the intervals are regularly spaced for all observations, the data can often be analyzed by the discrete-time methods described later.

NONPARAMETRIC ESTIMATION OF SURVIVAL DISTRIBUTIONS

Without a doubt, the oldest method of event history analysis is the life table, with the first known example appearing in the seventeenth century. The life table can be regarded as a nonparametric method for estimating the probability distribution of event times even when some of the observed event times are right censored. More specifically, the goal is to estimate the *survivor function*, denoted by $S(t)$. This is the probability that an event has not yet occurred by time t. (If the event is death, we say that the individual has *survived* to time t). We would like to be able to estimate this probability for any value of t.

When there are no censored cases, this is an easy task. To estimate the probability of surviving to a specified time t, we simply calculate the proportion of cases that are still alive at time t. There is also little difficulty if all censoring occurs at the end of the study. Again, we calculate the proportion of cases surviving to each specified time t, except that we have to stop at the earliest censoring time. In the recidivism study, for example, all the censoring occurred at 52 weeks. So we can easily estimate the survivor function for weeks 1 to 52, but cannot go any further.

This simple approach does not work, however, if some censoring times are smaller than some event times. That is when a life table is necessary. Here is an example. The sample consisted of 1296 nursing home patients who were followed from the date of entry to the date of discharge or the date of censoring (Morris et al., 1994). (These data are available

Table 16.1 *Life table for discharge of nursing home patients*

Interval (Lower,	Upper)	Number failed	Number censored	Effective sample size	Conditional probability of failure	Survival
0	150	719	0	1296.0	0.5548	1.0000
150	300	172	0	577.0	0.2981	0.4452
300	450	78	40	385.0	0.2026	0.3125
450	600	40	56	259.0	0.1544	0.2492
600	750	17	72	155.0	0.1097	0.2107
750	900	7	42	81.0	0.0864	0.1876
900	1050	4	37	34.5	0.1159	0.1714
1050		0	12	6.0	0	0.1515

on the web at http://lib.stat.cmu.edu/datasets/csb/). About 20% of the discharge times were censored because patients were still in the nursing home when the study terminated. Our goal is to estimate the survivor distribution for length of stay, that is, the number of days between entry and discharge. For the uncensored patients, the length of stay varied between 0 and 942. For the censored patients, the censoring times varied between 365 and 1092 (by design, everyone was followed for at least one year).

Table 16.1 displays a typical life table for these data. The time scale has been divided into seven intervals, each 150 days long, plus a final open-ended interval. In practice, the number of intervals and their upper and lower boundaries are somewhat arbitrary. The last column, labeled 'Survival', is the goal of the calculations. It is the estimated probability of surviving (still being in the nursing home), at the start of each interval. Thus, we see that the estimated probability of still being in the nursing home at 300 days is 0.3125.

The preceding columns give the intermediate calculations. 'Number failed' is the number of people who experience the event during each interval. 'Number censored' is the number of people whose censoring time falls within the interval. For example, among those people who had not yet been discharged when the study terminated, 40 had been in the nursing home between 350 and 450 days. The next column, 'Effective sample size', is an estimate of the number of people 'at risk' of discharge during that interval. The presumption is that people who were censored within the interval were only at risk for half the interval. Therefore, the effective sample size is the number who had not yet been discharged at the start of the interval, minus half the number who were censored within the interval. For the third interval, 405 patients had not yet been discharged at 300 days. But 40 of those patients were censored between time 300 and 450. So the effective sample size is $405 - 40/2 = 385$. The 'Conditional probability of failure' is an estimate of the probability that someone who survived to the start of the interval (had not yet been discharged) was discharged during the interval. It is simply the number who failed divided by the effective sample size. In the first interval, the estimated conditional probability of failure is $719/1296 = 0.5548$.

Once we have the conditional probabilities of failure, it is easy to calculate the survivor probabilities. The probability of surviving to time 150 is just 1 minus the probability of failing in the first interval: $1\ 0.5548 = 0.4452$. What is the probability of surviving to time 300? To get to 300, you have to survive the first interval and then survive the second interval. The probability of doing that is the probability of surviving the first interval times the conditional probability of surviving the second interval, given that you have survived the first: $(1 - 0.5548)\ (1 - 0.2981) = 0.3125$. Simi1ilarly, the estimated probability of surviving to time 450 is $(1 - 0.5548)(1 - 0.2981)(1 - 0.2026) = 0.2492$. Continuing in this fashion, we get the survival probabilities for the starting times for each interval. These probabilities are graphed as a function of time in Figure 16.1. Graphs of this sort are often referred to as *survival curves*.

One problem with the life table method is that the division of time into intervals is arbitrary. We can avoid this by using the Kaplan – Meier method (Kaplan and Meier, 1958), which modifies the life table in two ways. First, the time intervals are defined by the smallest time units observed in the data

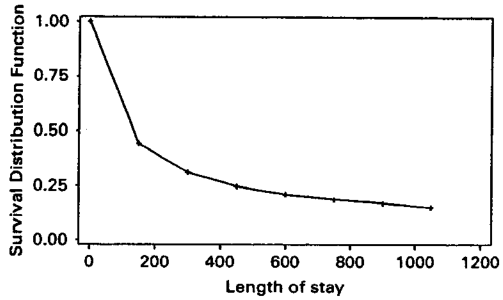

Figure 16.1 *Estimated survivor function using life table method*

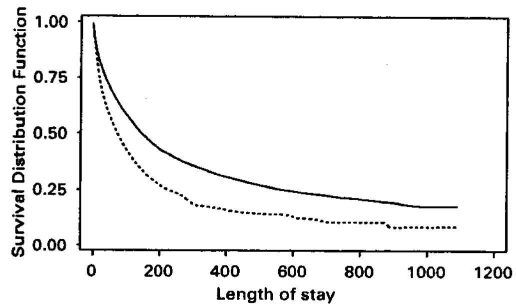

Figure 16.3 *Survival curves for men and women using Kaplan–Meier method*

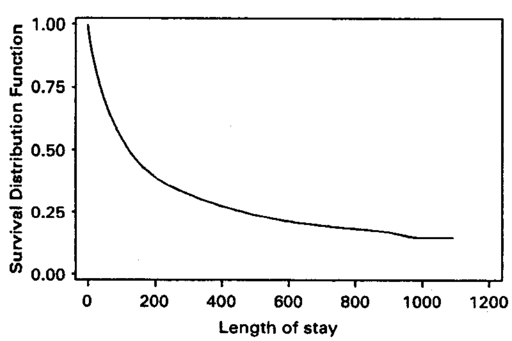

Figure 16.2 *Estimated survivor function using Kaplan–Meier method*

set. For the nursing home data, time is measured in days, so each day is a separate interval in the life table. (That can produce a very long table, but attention is usually focused on graphs). Second, the effective sample size for each interval is just the number of cases who have not yet had the event at the beginning of the interval. Thus, we do not subtract half the censored cases within the interval. Since the intervals are small, this usually makes little difference.

Figure 16.2 shows the Kaplain – Meier graph for the nursing home discharge data. The shape is essentially the same as the graph in Figure 1, but the curve is considerably smoother because many more points are plotted. For those time points that are plotted in Figure 16.1, the estimated survivor probabilities are, in fact, very close to the corresponding probabilities in Figure 16.2.

Survival curves get more interesting when you compare them for different groups. For the nursing home data, Figure 16.3 shows separate curves for men and women (with men coded 1 and women coded 0). The curve for women is always higher than that for men, which tells us that at every point in time women have a higher probability of still being in the nursing home. Equivalently, for whatever reason, men are being discharged more rapidly than women.

There are many different statistical tests for the null hypothesis that two groups have the same survivor function, the most common of which is the log-rank test. For these data, the log-rank statistic is 37.2 which (under the null hypothesis) has a chi-square distribution with 1 degree of freedom. This is highly significant, so we reject the null hypothesis and conclude that the survivor functions are different for males and females.

PARAMETRIC REGRESSION MODELS

Simple comparison of survival curves may be informative but it is usually not sufficient. Typically, researchers will want to adjust for other variables via some kind of regression model. A fairly simple regression model for event history data is the accelerated failure time (AFT) model, one member of the more general class of parametric regression models. Assume, for the moment, that there is no censoring, and let T_i be the event time for the ith individual in the sample. Let $x_{i1},\ldots,x_{ik}$ be a set of explanatory variables for individual i. (These are not allowed to vary with time.) The AFT model is

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \sigma\varepsilon_i, \qquad (16.1)$$

where log is the natural logarithm, ε_i is a random disturbance with a fixed variance, and σ is a scale parameter that controls the variance

Table 16.2 *Estimates of a lognormal regression model for nursing home stays*

Variable	Estimate	Standard error	Chi-square	*p*
Intercept	4.689	0.637	54.16	<0.0001
Age	0.015	0.007	4.03	0.0446
Male	−0.602	0.134	20.09	<0.0001
Married	−0.153	0.160	0.91	0.3395
Treatment	0.190	0.111	2.92	0.0875
Health Status	−0.312	0.060	26.73	<0.0001
Scale (σ)	1.940	0.045		

of the random component in the equation. We further assume that ε_i is independent of the x variables and of any other εj. These assumptions imply the usual assumptions of the linear regression model. So, if there are no censored data, we can get best linear unbiased estimates of the β coefficients by ordinary least squares (OLS), using log T as the dependent variable.

Of course, most event history data have at least some censoring and OLS just will not work in that situation. Censoring is easily handled by the method of maximum likelihood (ML) which, under fairly broad conditions, produces coefficient estimates that are consistent, asymptotically efficient, and asymptotically normal. Many software packages (e.g., SAS, Stata, SYSTAT, BMDP, S-PLUS) have ML procedures for at least some versions of the AFT model.

To implement ML, it is necessary to specify the probability distribution of the random disturbance term ε. A standard normal distribution would be the most familiar choice. But there are three other distributions that are also commonly used for this kind of modeling: extreme value, logistic, and log-gamma. Each of these distributions has an implied distribution for the event time T, as shown in the following table:

Distribution of ε	Distribution of T
Normal	Lognormal
Logistic	Log-logistic
Extreme value	Weibull (or exponential)
Log-gamma	Gamma

Typically, we refer to the different submodels by the distribution of T. The gamma model is the most general model because it has both the lognormal and Weibull models as special cases.

What is the point of considering different distributions for ε? Each of these distributions has somewhat different implications for *hazard functions*, an important concept in event history analysis that we shall discuss in the next section.

Table 16.2 shows the results from fitting a lognormal regression model to the length-of-stay data for nursing home. In addition to the gender variable, the model includes variables for age, marital status, health status and 'treatment'. Health status was coded as integer values from 2 through 5, with higher scores indicating worse health. The treatment variable was a dummy variable for whether or not the patient was admitted to one of 18 nursing homes (out of a total of 36) that received special treatment: higher *per-diem* payments for accepting more disabled Medicaid patients, and bonuses for improving patient's health status and discharging patients within 90 days.

In Table 16.2, we see a highly significant effect of gender, with males having smaller lengths of stay. This is consistent with the survival curve comparison, but now we are controlling for several other variables. The coefficient of − 0.602 is the effect of gender on the logarithm of length of stay, which is not readily interpretable. We can get an interpretable number by exponentiating the coefficient: exp(− 0.602) = 0.55. This tells us that the expected length of stay for men is only 55% of the expected length of stay for women.

Older patients have somewhat longer lengths of stay, an effect that is just barely statistically significant (at the 0.05 level). More specifically, if we calculate 100[exp(0.015)−1] = 1.5, we can conclude that each additional year of age (at admission) is associated with a 1.5% increase in expected length of stay (controlling for other variables in the model). Contrary to expectation, patients in the treatment nursing homes had longer lengths of stay (not statistically significant) and patients with poorer health status had *shorter* lengths of stay (highly significant).

Although AFT models can be very useful, there are a few limitations with this approach. First, results can vary depending on the distribution chosen for ε, and it may be difficult to determine which is the more appropriate distribution. For example, if a Weibull model is fitted to the nursing home data, the p-value for treatment goes down to 0.04, which many would judge to be statistically significant. A second problem is that most software for fitting these models does not allow for time-varying explanatory variables (although some software for the Weibull model can do this).

These problems do not occur with the method of Cox regression, which will be discussed in the next section. Nevertheless, AFT models (along with other parametric regression models) have some advantages over Cox regression: they are much better at handling left censoring and irregular interval censoring; and they make it much easier to generate predicted times to events.

COX REGRESSION

By far the most popular method for analyzing event history data is Cox regression. First proposed by the British statistician, Sir David Cox, in 1972, this method has many attractive features: minimal assumptions about the distribution of event times; ease of incorporating time-varying explanatory variables; ability to handle both continuous- and discrete-time data; capacity for semiparametric stratification; and allowance for left truncation. Cox (1972) actually did two things: he proposed a new model called the *proportional hazards* (PH) model, and he devised a new estimation method now known as *partial likelihood*.

In the PH model, the dependent variable is $h(t)$, the hazard of an event at time t. Specifically, if there are no time-varying explanatory variables, the PH model may be written as

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}. \quad (16.2)$$

Obviously, to understand this model, it is essential to have a clear understanding of what $h(t)$ is. Roughly speaking, $h(t)$ can be interpreted as the instantaneous probability that an event will occur at time t. That is not quite accurate, however, because unlike a probability, the hazard can be greater than one (although it can never be less than zero).

Here is a formal definition. Let $P(t, t + \Delta t)$ be the conditional probability that an event occurs in the time interval $(t, t + \Delta t)$, given that it has not already occurred prior to t. To get the hazard function, we divide this probability by the length of the interval Δt, and take the limit as Δt goes to 0:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t, t + \Delta t)}{\Delta t} \quad (16.3)$$

The hazard is allowed to be different at every point in time t, which is why we call it a hazard *function*. In equation (16.2) the hazard has an i subscript to indicate that it can vary across individuals. If $h(t)$ has a constant value r, it can be interpreted as the expected number of events in a one-unit interval of time. Alternatively, $1/r$ is the expected length of time until the next event. Suppose, for example, that the events are residence changes, time is measured in years, and the estimated hazard of a residence change is 0.20. That would imply that, for a given individual, the expected number of changes in a year is 0.20 and the expected length of time between changes is $1/0.20 = 5$ years.

Like a probability (from which it is derived), the hazard is never directly observed. Nevertheless, it governs both the occurrence and timing of events, and models formulated in terms of the hazard may be estimated from observed data. Going back to (16.2) for the PH model, notice that on the right-hand side there is an unspecified function of time $\alpha(t)$. We could make this more specific by assuming, for example, that, $\alpha(t) = \alpha_0 + \alpha_1 t$, where α_0 and α_1 are constants to be estimated. This would give us a parametric PH model known as the Gompertz model. Alternatively, if we specified $\alpha(t) = \alpha_0 + \alpha_1 \log t$, we would get a Weibull model (which also happens to be an AFT model, although expressed in different form). But the beauty of Cox's PH model is that it is not necessary to decide what $\alpha(t)$ is. We can estimate the β coefficients for any function $\alpha(t)$ without restriction. The partial likelihood method is what makes this possible.

Why is (16.2) called the proportional hazards model? Because if we take the ratio of the hazards for any two individuals at the same time $h_i(t)/h_j(t)$, where $i \neq j$, that ratio is a constant that does not depend on time. One

Table 16.3 *Estimates of a proportional hazards regression model for nursing home stays*

Variable	Estimate	Standard error	Chi-square	p
Age	−0.006	0.004	1.79	0.1806
Male	0.362	0.075	23.52	<0.0001
Married	0.115	0.088	1.71	0.1905
Treatment	−0.081	0.064	1.64	0.2005
Health status	0.166	0.035	22.62	<0.0001

implication of this property is that the 'effect' of any explanatory variable is invariant over time. While this may seem restrictive, the PH model is much more flexible than the parametric models we have already considered. Furthermore, as we shall see, the model can easily be extended to allow for nonproportional hazards.

Now let us consider the partial likelihood estimation method, which is what makes it possible estimate the βs without specifying $\alpha(t)$. The method is very much like maximum likelihood but instead of maximizing the full likelihood function, one works with a portion of the likelihood function that depends on the βs but not on $\alpha(t)$. This part of the likelihood depends only on the order in which events occur, not on the exact times of the events. Some information is lost in the process, but what is gained is a great deal of robustness. In practice, the partial likelihood is treated almost exactly the same as if it were an ordinary likelihood function. Nearly all full-featured statistical packages (SAS, SPSS, Stata, S-PLUS, SYSTAT, BMDP) include a procedure for doing partial likelihood estimation of the PH model.

Let us apply the method to our nursing home discharge data. Table 16.3 shows results of a Cox regression with the same explanatory variables as in Table 16.2 for the lognormal model. To specify the model, virtually all Cox regression software requires that the dependent variable be listed in two parts: a variable containing the time of the event or censoring, and a variable indicating whether or not that time was a censoring time or an event time. A list of independent variables is the only other information that is required.

Comparing Table 16.3 with Table 16.2, there are a couple of noteworthy differences. First, unlike most regression methods, there is no intercept reported in the computer output. That is because the intercept (if there really is one) is part of the $\alpha(t)$ function in (16.2), which drops out of the estimation process. The other big difference is that the signs of the coefficients are all the opposite of what they were in Table 16.2. Again, that is no accident, and it stems from the fact that the dependent variable in the AFT model is the event time, while the dependent variable in the PH model is the hazard for the event. If the hazard for some event is low, then the event is unlikely to occur and the expected time until an event occurs will be large. On the other hand, if the hazard is high, the event is very likely to occur and the expected time to the event will be small. So hazards and event times are inversely related.

Despite these apparent differences, the results in Tables 16.2 and 16.3 are reasonably consistent. Both tables show highly significant effects of gender and health status. For the other variables, the p-values from the Cox regression are a bit higher than those for the AFT regression. One consequence is that Age, which was marginally significant in the AFT regression, is no longer significant in the Cox regression.

To interpret the magnitudes of the coefficients, it is helpful to first transform them using the same formula as for the AFT models: $100[\exp(\beta)-1]$. This gives the percent change in the *hazard* of an event for each one-unit increase in a particular explanatory variable (holding other variables constant). For Male, we have $100[\exp(0.362)-1] = 44\%$. This says that males have a hazard of discharge that is 44% higher than the hazard for females (after adjusting for the other variables in the model). For 'Health Status', we have $100[\exp(0.166)-1] = 18\%$. This says that for each one-unit increase in the health status scale, the hazard for discharge goes up by 18%.

TIME-DEPENDENT COVARIATES

There are no time-dependent covariates in the nursing home discharge data – all the independent variables were measured at the time of admission. In principle, however, some of these variables, like health status, could change during the nursing home stay. If we had measurements on these changing variables, how could we incorporate them into the analysis? It is easy to build them into the PH model. For example, suppose that $x_{i1}(t)$ denotes a patient's health status at time t, where t is the length of time since admission to the nursing home. Then we could write the model:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1}(t) + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}. \tag{16.4}$$

This equation says that the hazard of discharge at time t depends on the patient's health status at the same time t. According to this model, any change in health status produces an immediate change in the hazard of discharge.

While it is simple to modify the model to allow for time-dependent covariates, it may not be so easy to estimate that model. One of the attractions of the partial likelihood method is that it is relatively straightforward to incorporate time-dependent covariates into the estimation process. To do that, however, you need appropriate data. Ideally, one should know the values of the time-dependent covariates at every point during the observation period. That may not be difficult for some variables, like marital status. If we know the dates of any changes in marital status and we know the status before and after the change, then we can assign a marital status for every day of observation.

On the other hand, it may be difficult to get daily measurements of health status. Instead, we may only know health status at weekly or monthly intervals. In such cases, it is necessary to devise some plausible rule for assigning values to the days in between. One possible rule would be to presume that each measured value remains in effect until the next measurement. Another possibility is to do some kind of linear interpolation. Such choices must be made by the investigator, often on the basis of substantive considerations.

Once these issues have been resolved, the next question is how to implement the estimation process. It turns out that there are two rather different computational roads that lead to the same result. In the first approach – let us call it the programming method – there is a single record for each individual. The changing values of the time-dependent covariate are coded in multiple variables on that record. For example, if health status is measured monthly and patients are observed for a maximum of three years, one would need 36 variables to describe the health status measurements.

To specify the model, it is necessary to write a small program that assigns the appropriate value of health status to each time at which a discharge occurs. This program must be executed as part of the estimation process, not before. The reason is subtle but extremely important. If someone is discharged on day 125, that person is compared to all the other persons who were still in the nursing home 125 days after admission – the 'risk set' for day 125. In doing that comparison, it is necessary to retrieve the values of health status on day 125 for all the people at risk of discharge. But many of those people might also be in other risk sets. If a discharge occurred on day 100, for example, the program must retrieve the values of health status on day 100 for all the people in that risk set. But all the people in the nursing home on day 125 will also have been there on day 100. So, for a given individual, different values of the time-dependent covariates are retrieved at different points in time.

The other computational approach to time-dependent covariates is known as *episode splitting* (also called the counting process method). In this method, each individual may be represented in the data set by more than one record. Each record corresponds to an interval of time during which *all the covariates are constant*. The records must contain the following variables: the starting time of the interval (measured as time since admission), the stopping time (also measured as time since admission), the values of all independent variables (both time-constant and time-dependent), and a censoring indicator. Any record that does not end in an event is treated as censored. Time-constant variables are replicated across the multiple records for each individual. In doing the analysis, one simply specifies the variables for starting time, stopping time, censoring, and the covariates. At the analysis stage, there is no distinction between time-varying and time-constant explanatory variables. That is because, within each record, both are constant.

The choice between these two computational approaches depends greatly on the

Table 16.4 *Estimates of a proportional hazards regression model for nursing home stays, with treatment by time interaction*

Variable	Estimate	Standard error	Chi-square	*p*
Age	−0.005	0.004	1.59	0.2066
Male	0.361	0.075	23.36	<0.0001
Married	0.109	0.088	1.52	0.2173
Treatment	−0.240	0.085	7.94	0.0048
Health status	0.165	0.035	22.34	<0.0001
Treat. × Time	0.0013	0.0004	7.70	0.0055

software being used. Some programs (like SPSS and BMDP) use only the programming method. Other programs (like Stata) use only the episode splitting method. And some (like SAS) allow for both methods, either separately or in combination. If correctly implemented, both approaches should produce exactly the same numerical output. I tend to favor the episode splitting approach because it makes it easier to avoid mistakes. After you have created the multiple records, you can examine them carefully to see if they conform to your intentions. And once you have constructed the data set, you can specify the models in a very simple form.

On the other hand, there are certain situations where the programming method is much simpler. Here is one example. For the model displayed in Table 16.3, an implicit assumption is that each independent variable has the same effect at all points in time. This is a crucial implication of the proportional hazards assumption. But what if that assumption is not true? Perhaps there is a large effect of the treatment on the hazard of discharge when people are first admitted to the nursing home, but the effect becomes progressively smaller as time goes on. The Cox model can be modified to express this idea by including an interaction between treatment and time. Specifically, if x_k is the treatment variable, we can include the product of x_k and t:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \beta_{k+1} x_{ik} t.$$

This model says that the 'effect' of x_k is a linear function of t. Alternatively, letting $z_i(t) = x_{ik}t$, we can write this model as

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \beta_{k+1} z_i(t).$$

In short, we now have a model with a time-dependent covariate z. But the episode splitting method will not work for this model because there are no intervals during which $z(t)$ is constant. Hence, we must use the programming method. We will skip the details because the implementation varies greatly across software packages.

Table 16.4 shows the results of estimating the interaction model using the programming method. We see that the interaction term is statistically significant beyond the 0.01 level, indicating that the effect of treatment does vary with time since admission. More specifically, the 'main effect' of treatment represents the effect of treatment at time 0, the date of admission. It is negative (− 0.240) and highly significant. Applying the $100[\exp(\beta) - 1]$ transformation, we may say that at the time of admission, the treated group has a 21% lower hazard of discharge than the control group. However, for each additional day since admission, the effect of treatment goes up by 0.0013, implying that the effect is zero at 185 days. After that, the treatment effect becomes steadily more positive. At one year, it is equal to the coefficient at admission, only in the opposite direction. This explains why, overall, we do not see much effect of treatment in Table 16.3. The early negative effect is balanced by the later positive effect.

This example illustrates one way to test the proportional hazards assumption in the Cox regression model: check to see if variables have significant interactions with time. In this case, the method of diagnosis is also the cure. By including interactions with time, we extend the Cox model to allow for nonproportional hazards.

STRATIFICATION

Another useful feature of Cox regression is the ability to control for one or more variables in a completely nonparametric manner. This is called *stratification*, although the meaning of this term is somewhat different than in other contexts. For example, in the nursing home study, suppose that we want

Table 16.5 *Estimates of a proportional hazards regression model for nursing home stays, with stratification by health status*

Variable	Estimate	Standard error	Chi-square	p
Age	−0.006	0.004	2.11	0.1457
Male	0.356	0.075	22.64	<0.0001
Married	0.119	0.088	1.82	0.1768
Treatment	−0.251	0.086	8.62	0.0033
Treat. × Time	0.0014	0.0005	8.91	0.0028

estimate the effect of the treatment controlling for health status, but we do not want to assume that the effect of health status satisfies the proportional hazards assumption. We specify the following model:

$$\log h_i(t) = \alpha_H(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \beta_{k+1} z_i(t). \qquad (16.5)$$

The only difference between this equation and the one in the preceding section is that the unspecified function of time $\alpha_H(t)$ is now subscripted with an H. This indicates that there may be a different unspecified function of time for each value of the health status variable. Consequently, the model allows for differences in the discharge rate across the four possible values of health status, but these differences may vary with time. On the other hand, the coefficients of the other variables are assumed to be the same within each health stratum.

Models with stratification can be easily estimated using the partial likelihood method. Table 16.5 shows the results of stratifying by health status for the nursing home data. Although, health status does not show up in the table, it *is* being controlled, and in a way that is less restrictive than in Table 16.4. Nevertheless, the results for the other variables are only slightly different in Table 16.5 than in Table 16.4.

LEFT TRUNCATION AND LATE ENTRY

Conventional Cox regression programs presume that there is some common origin time (time 0) at which everyone begins to be at risk of an event. For the nursing home study, the origin time was the day on which each person was admitted to the nursing home. The event time was then the number of days between admission and discharge. The implication is that if a person was discharged on, say, day 279, that person was *at risk* of a discharge every day between 0 and 279.

That is not always a plausible assumption. Suppose the nursing home study was conducted in the following way. On a certain date, all patients currently residing in the nursing home were recruited into the study and then followed forward until discharge or censoring. One could set the origin time to be the date of recruitment, and then the event time would be the difference between recruitment date and discharge date. But the recruitment date is a purely arbitrary point in time, and there is no reason to think the hazard for discharge would depend on time since that arbitrary point. Instead, one could, as before, record discharge times as time since admission. But that raises a new problem. By design, it is not really possible for a person to be discharged between the date of admission and the date of recruitment. If the person had been discharged during that interval, he or she would not be available for recruitment and would not have been in the study.

This problem is known as left truncation or late entry to the risk set. The solution is to continue to measure discharge times from the date of admission, but remove individuals from the risk set during the interval from admission to recruitment. Most Cox regression programs cannot do this, but many now have this capability. To implement these procedures, all that is necessary is to specify an entry time and an event time for each individual.

COMPETING RISKS

All the techniques we have discussed so far presume that events are indistinguishable: all deaths are the same, all arrests are the same, all nursing home discharges are the same. For many applications, however, events can be

Table 16.6 *Estimates of a Cox regression model for exits from power*

Variable	Estimate	Standard error	Chi-square	*p*
Manner	0.250	0.124	4.07	0.0437
Year	−0.018	0.008	4.83	0.0280
Age	0.023	0.005	18.33	<0.0001
Conflict	0.094	0.128	0.54	0.4608
Income	−0.177	0.082	4.64	0.0312
Literacy	0.0007	0.003	0.05	0.8212

separated into different types, and the different types may have potentially different causes. For example, if the event is a job termination, we might expect that job performance would have very different effects on the hazard of quitting and the hazard of being fired.

Here is a more detailed example. The data consist of information on 438 principal leaders of developing countries during the past 60 years (Bienen and van de Walle, 1991). (This data set is available on the web at www.ssc.upenn.edu/~allison under the name leaders.dat.) For each leader, we have information on the number of years in power, and the manner in which he was removed from power: constitutional means (146 cases), nonconstitutional means (*coup d'etat*, assassination, etc.; 154 cases), or death from natural causes (27 cases). Another 111 of the leaders were still in power (censored) when the study terminated in 1987. These three modes of exit from office are appropriately regarded as *competing risks* because the occurrence of any one of them removes the individual from risk of the other two.

The goal is to estimate the effects of several explanatory variables on the hazard of leaving office. The covariates we shall examine are:

Manner	1 = nonconstitutional entry to power, 0 = constitutional).
Year	Year of entry into power.
Age	Age on assuming power.
Conflict	1 = medium or high ethnic conflict, otherwise 0.
Income	Natural logarithm of GNP per capita, 1973.
Literacy	Literacy rate.
Region	0 = Middle East, 1 = Africa, 2 = Asia, 3 = Latin America.

Table 16.6 presents results from a Cox regression predicting the hazard of leaving office, without distinguishing the manner of departure. (For a more detailed analysis of these data, see Allison, 1995.) Age at entry has a strong and unsurprising effect: leaders who are older when they assumed power are more likely to leave quickly. There is also some evidence that higher rates of departure are associated with lower country income, earlier starting years, and the seizure of power by nonconstitutional means.

But this analysis lumps together three different kinds of events that may, in fact, be quite different. To disaggregate these event types, we first specify a separate hazard function for each event type. Specifically, $h_{ij}(t)$ is the hazard for person i experiencing an event of type j at time t. Then we write a separate proportional hazards equation for each event type:

$$\log h_{ij}(t) = \alpha_j(t) + \beta_{1j}x_{i1} + \cdots + \beta_{kj}x_{ik}, \quad j = 1,\ldots,3. \quad (16.6)$$

Estimating these three equations is a simple task that can be done with conventional software for Cox regression. The basic rule is this: estimate each equation separately by estimating a Cox model for that specific event type, treating all other events as though the individual was censored at the time of event occurrence. In practice, this is easily accomplished by repeatedly estimating the same model, each time specifying different sets of events to be treated as censored.

Table 16.7 displays results from doing this for the three modes of departure from office. What is of interest here is that many variables have quite different effects on the different modes. Age, for example, is the only variable that has a significant effect on the hazard of a natural death. Age also affects exits by constitutional means, but has no apparent effect on nonconstitutional exits. On the other hand, the hazard of a nonconstitutional exit decreases with calendar time (year), but the hazards for the other two modes are unaffected. Increases in income appear to reduce the hazard of nonconstitutional exit, but do

Table 16.7 *Estimates of Cox regression models for three types of exit*

Variable	Estimate	Standard error	Chi-square	*p*
Constitutional exits				
Manner	−0.336	0.217	2.39	0.1223
Year	0.003	0.012	0.06	0.8123
Age	0.025	0.008	8.66	0.0032
Conflict	−0.023	0.194	0.01	0.9067
Income	−0.119	0.116	1.06	0.3038
Literacy	0.014	0.005	6.21	0.0127
Nonconstitutional exits				
Manner	0.686	0.175	15.45	<0.0001
Year	−0.036	0.012	8.85	0.0029
Age	0.009	0.008	1.19	0.2745
Conflict	0.437	0.201	4.73	0.0296
Income	−0.435	0.144	9.17	0.0025
Literacy	−0.006	0.005	1.49	0.2214
Death from natural causes				
Manner	0.089	0.536	0.03	0.8679
Year	−0.055	0.035	2.53	0.1117
Age	0.084	0.019	19.32	<0.0001
Conflict	−0.474	0.494	0.92	0.3380
Income	0.141	0.271	0.27	0.6014
Literacy	−0.010	0.013	0.66	0.4156

not affect constitutional exits. If a leader gained power by nonconstitutional means, he is far more likely to be ousted by nonconstitutional means. In short, the factors affecting an exit from power seem to depend rather heavily on the type of exit.

UNOBSERVED HETEROGENEITY

If you look closely at equation (16.2) for Cox's PH model, you will notice that there is no random disturbance term. That does not mean that the model is deterministic because the dependent variable – the hazard – represents only the *propensity* for events to occur. Two people with the same hazard can end up with very different event times. Nevertheless, the model does say that all variations in the hazard are completely explained by the covariates that are included in the model, an unlikely assumption for virtually any real application.

Suppose we expand the model to include a random disturbance term ε which represents unobserved heterogeneity:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i.$$

As in conventional linear models, we might further assume that ε_i has normal distribution with mean 0 and constant variance, and is uncorrelated with any of the *x*s. What would such a model imply? There are two major implications of unobserved heterogeneity:

1. There will be an artifactual tendency for the observed hazard function to decrease with time. Thus, it will appear that the longer people survive, the lower their risk of death.
2. Estimates of the β coefficients using conventional methods will be biased toward zero, a phenomenon known as heterogeneity shrinkage (Gail et al., 1984). Fortunately, standard error estimates will still be valid, as will tests of hypotheses that coefficients are 0. Note that heterogeneity shrinkage is a potential problem for many other nonlinear models, including logistic regression (Allison, 1987).

Although there have been extensive efforts to devise methods to overcome these problems

Table 16.8 *Cox regression for 395 jobs, with and without standard error corrections*

Variable	Estimate	Standard error	Chi-square	Robust standard error	Robust Chi-square
Schooling	0.187	0.029	41.80	0.046	16.43
Prestige	−0.079	0.006	154.30	0.006	163.40
Salary (logged)	−0.597	0.114	27.21	0.142	17.55
Time	0.269	0.029	83.45	0.031	77.50

(Heckman and Singer, 1984; Elbers and Ridder, 1982; Hougaard 1986), I believe that the attempt is futile when no more than one event is observed for each individual, as in the case of death. There simply is not enough information in the data to effectively distinguish between heterogeneity and time dependence. On the other hand, when multiple events are observed for each individual, models with unobserved heterogeneity can be reliably estimated (Yamaguchi, 1986).

REPEATED EVENTS

All the methods discussed so far presume that each individual experiences no more than one event. Yet most events that are interesting to social scientists are repeatable: births, marriages, divorces, arrests, job terminations, residential moves, etc. Here is a simple example. For 100 persons, we have data on the lengths of 395 jobs they held over a ten-year period (data set jobmult.dat, available at www.ssc.upenn.edu/~allison). The number of jobs held by each person ranged from 1 to 10. The goal is to estimate a model in which the hazard of job termination (which may be voluntary or involuntary) depends on four explanatory variables: years of schooling, occupational prestige, salary (logged) at the beginning of each job, and time at the start of each job (in years since the beginning of the first job).

One simple approach is to treat each job as a separate observation, and estimate a Cox regression model for length of the job. The first three columns of Table 16.8 report results from such an analysis. All four variables have highly significant coefficients. More schooling increases the hazard of termination, higher prestige and higher salary reduce the hazard of termination, and the rate of termination increases with time.

A major problem with this analysis, however, is that it ignores the potential dependence among the several jobs for the same person. If a person's first job is very short, we might expect that later jobs would be short as well. If job lengths are positively correlated, treating them as if they were independent observations will result in standard errors that are underestimated and test statistics that are too high. There are several methods available to deal with the problem of dependence. We will consider two that are readily available: robust standard errors and fixed effects models.

The method of robust standard errors does not involve any changes in the coefficient estimates, but modifies the standard errors to correct for dependence. Based on work of Huber (1967) and White (1980), the calculation of robust standard errors is sometimes called the 'sandwich' method because of the structure of the matrix formula. For Cox regression, several major software packages (SAS, Stata, S-PLUS) now provide robust standard errors as an option. The last two columns of Table 16.8 give robust standard errors and the associated chi-squares (the squared ratio of the coefficient to its standard errors). The chi-squares for education and salary are substantially smaller after the correction, although still highly significant.

The fixed-effects method is a rather different approach to repeated events. First, it is based on a model that explicitly allows for unobserved heterogeneity across individuals,

$$\log h_{ij}(t) = \alpha(t) + \beta_1 x_{ij1} + \cdots + \beta_k x_{ijk} + \varepsilon_i, \qquad (16.7)$$

where $h_{ij}(t)$ is the hazard for event j for individual i. The term ε_i represents unobserved heterogeneity that is specific to individual i but constant across events j. Rather than assuming that ε_i is a random variable, we assume that it is just some constant value that differs across individuals (hence the term 'fixed effects'). To estimate this model with conventional software, we must combine the ε_i with $\alpha(t)$ to get

$$\log h_{ij}(t) = \alpha_i(t) + \beta_1 x_{ij1} + \cdots + \beta_k x_{ijk},$$

Table 16.9 *Fixed-effects Cox regression for 395 jobs*

Variable	Estimate	Standard error	Chi-square	*p*
Schooling	0	.	.	.
Prestige	–0.056	0.010	32.64	<.0001
Salary (logged)	–0.866	0.174	24.66	<.0001
Time	0.021	0.039	0.28	0.5988

which says that each individual has a distinct baseline hazard function. In this form, the model is easily estimated using the method of stratification described earlier (Allison, 1996). Each individual constitutes a distinct stratum.

Results for a fixed-effects analysis of the job termination data are shown in Table 16.9. The most obvious feature of these results is the null estimate for schooling, which illustrates a peculiar disadvantage of fixed effects methods – you cannot estimate the effects of any variables that are constant over time for the individual. That is because the method only uses information about within-individual variation. In effect, we are asking the question, 'For each person, why are some jobs longer and others shorter?'. Variables that do not vary within persons cannot provide any answers to this question.

That does not mean, however, that schooling is not controlled in this analysis. In fact, fixed effects methods control for *all* variables that are constant within individuals, whether they can be measured or not. Hence, the analysis reported in Table 16.9 controls for race, sex, region of birth, family background, stable personality characteristics, and so on. For nonexperimental designs, where the control of such factors is a major issue, this is a very attractive feature.

Because of the controls for unmeasured factors, the fixed effects method often produces results that are markedly different from a conventional analysis, even with robust standard errors. That is not the case in this example, where the coefficients for prestige and salary are similar in Tables 16.8 and 16.9. On other hand, the coefficient for time is markedly lower in Table 16.9 than in Table 8, and is no longer statistically significant. The other thing to keep in mind about the fixed effects method is that, when unobserved factors are uncorrelated with the measured variables, there may be a substantial loss of power, that is the standard errors may be appreciably higher than in a conventional analysis. That is because the between-individual variation contributes nothing to the analysis.

Two other approaches to repeated events are generalized estimating equations (GEE) and random effects (mixed) models. The GEE method uses the robust standard errors method, but also produces more efficient estimates of the coefficients. Random-effects models may be based on equation (16.7), but ε_i is treated as a random variable with a specified probability distribution. Random effects estimates are potentially more efficient than those produced by the fixed effects method, but they do not control for unobserved factors that are correlated with the variables in the model. These two methods are currently only available in specialized packages, but may see more widespread implementation in the near future.

DISCRETE-TIME AND TIED DATA

So far we have assumed that events can occur at any point in time and event times are measured with perfect precision. Under that assumption, it is impossible for two events to occur at *exactly* the same time. In practice, however, event times are often measured quite coarsely. We may only know the week, month, or year in which an event occurred. In such situations, many individuals may have 'tied' event times. That is, two or more individuals may have events at the same measured times.

The Cox regression method, in its classic form, is not appropriate when the data contain tied event times. When such data occur, most Cox regression programs invoke an approximate partial likelihood method proposed by Peto (1972) and Breslow (1974). However, this approximation may be poor when there are lots of tied event times. A much better approximation (Efron, 1977) is available in some programs. Better still are two exact methods, one that assumes that event times are truly continuous (a reasonable assumption for most applications) and another that assumes that event times are truly discrete (which may be appropriate for some special applications). Both of these exact methods are computationally intensive, however, and may be impractical for larger data sets with many ties.

There is another approach to coarse event times that is quite different in implementation but actually estimates the same underlying models as the exact Cox regression methods (Allison, 1982, 1984, 1995). The basic idea is very simple. For each unit of time that an individual is observed, create one observational record. For each record, create a dependent variable that is coded 1 if an event occurred during this unit of time, otherwise 0. Independent variables take on whatever values were measured at the beginning of that time unit. After pooling all the records, do a logistic regression predicting the dependent variable from the independent variables. This will produce maximum likelihood estimates of the 'truly discrete' model that can also be estimated by some Cox regression programs. Alternatively, to estimate the 'truly continuous' model, one may specify a complementary log-log link for the binary regression model. Unlike the exact partial likelihood methods, this maximum likelihood approach can readily handle large data sets with large number of tied event times.

BIBLIOGRAPHIC NOTE

There are many textbooks on event history (survival analysis) but the best tend to have a biomedical orientation. Of these, my favorites are Collett (1994) and Hosmer and Lemeshow (1998). Other useful texts from this perspective include Klein and Moeschberger (1997) and Kleinbaum (1996), the latter focusing exclusively on Cox regression. For texts by social scientists, see Blossfeld et al. (1989) and Yamaguchi (1991).

REFERENCES

Allison, Paul D. (1982) 'Discrete-time methods for the analysis of event histories', in Samuel Leinhardt (ed.), *Sociological Methodology 1982*. San Francisco: Jossey-Bass, pp. 61–98.

Allison, Paul D. (1984) *Event History Analysis*. Thousand Oaks, CA: Sage.

Allison, Paul D. (1987) 'Introducing a disturbance into logit and probit regression models', *Sociological Methods and Research*, 15: 355–74.

Allison, Paul D. (1995) *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute.

Allison, Paul D. (1996) 'Fixed effects partial likelihood for repeated events', *Sociological Methods and Research*, 25: 207–22.

Bienen, Henry S. and van de Walle, Nicholas (1991) *Of Time and Power*. Stanford, CA: Stanford University Press.

Blossfeld, Hans-Peter, Hamerle, Alfred and Mayer, Karl Ulrich (1989) *Event History Analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Breslow, Norman (1974) 'Covariance analysis of censored survival data', *Biometrics*, 30: 89–99.

Collett, David (1994) *Modelling Survival Data in Medical Research*. London: Chapman & Hall.

Cox, David R. (1972) 'Regression models and life tables', *Journal of the Royal Statistical Society, Series B*, 34: 187–202.

Efron, Bradley (1977) 'The efficiency of Cox's likelihood function for censored data', *Journal of the American Statistical Association*, 76: 312–19.

Elbers, C. and Ridder, G. (1982) 'True and spurious duration dependence: The identifiability of the proportional hazard model', *Review of Economic Studies*, 49: 403–10.

Gail, Mitchell H., Wieand, S. and Piantadosi, S. (1984) 'Biased estimates of treatment effect in randomized experiments with nonlinear regression and omitted covariates', *Biometrika*, 71: 431–44.

Heckman, James J. and Singer, Burton (1984) 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica*, 52: 63–132, 271–320.

Hosmer, David W. and Lemeshow, Stanley (1999) *Applied Survival Analysis*. New York: Wiley.

Hougaard, P. (1986) 'Survival models for heterogeneous populations derived from stable distributions', *Biometrika*, 73: 387–96.

Huber, P.J. (1967) 'The behavior of maximum likelihood estimates under nonstandard conditions', in L. Le Cam and J. Neyman (eds), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, pp. 221–33.

Kaplan, E.L. and Meier, P. (1958) 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association*, 53: 457–81.

Klein, John P. and Moeschberger, Melvin L. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Kleinbaum, David G. (1996) *Survival Analysis: A Self-Learning Text*. New York: Springer-Verlag.

Morris, Carl N., Norton, Edward C. and Zhou, Xiao H. (1994) 'Parametric duration analysis of nursing home usage', in N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse (eds), *Case Studies in Biometry*. New York: Wiley, pp. 231–48.

Peto, R. (1972) Contribution to the discussion of Cox (1972), *Journal of the Royal Statistical Society Series B*, 34: 205–7.

Rossi, Peter H., Berk, Richard A. and Lenihan, Kenneth J. (1980) *Money, Work and Crime*. New York: Academic Press.

Sørensen, Aage B. (1977) 'Estimating rates from retrospective questions', in D.R. Heise (ed.), *Sociological Methodology 1977*. San Francisco: Jossey-Bass, pp. 209–23.

Tuma, Nancy B. and Hannan, Michael T. (1978) 'Approaches to the censoring problem in analysis of event histories', in Karl F. Schuessler (ed.), *Sociological Methodology 1979*. San Francisco: Jossey-Bass, pp. 209–240.

White, Halbert A. (1980) 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica*, 48: 817–38.

Yamaguchi, Kazuo (1991) *Event History Analysis*. Thousand Oaks, CA: Sage Publications.

Yamaguchi, Kazuo (1986) 'Alternative approaches to unobserved heterogeneity in the analysis of repeatable events', in Nancy B. Tuma (ed.), *Sociological Methodology 1986*. Washington, DC: American Sociological Association, pp. 213–49.