Chapter 17
Introduction to Survival Analysis

17.1 Background

Suppose that one wished to study the occurrence of some event in a population of subjects. If the time until the occurrence of the event were unimportant, the event could be analyzed as a binary outcome using the logistic regression model. For example, in analyzing mortality associated with open heart surgery, it may not matter whether a patient dies during the procedure or he dies after being in a coma for two months. For other outcomes, especially those concerned with chronic conditions, the time until the event is important. In a study of emphysema, death at eight years after onset of symptoms is different from death at six months. An analysis that simply counted the number of deaths would be discarding valuable information and sacrificing statistical power.

Survival analysis is used to analyze data in which the time until the event is of interest. The response variable is the time until that event and is often called a *failure time, survival time*, or *event time*. Examples of responses of interest include the time until cardiovascular death, time until death or myocardial infarction, time until failure of a light bulb, time until pregnancy, or time until occurrence of an ECG abnormality during exercise. Bull and Spiegelhalter[83] have an excellent overview of survival analysis. 1

The response, event time, is usually continuous, but survival analysis allows the response to be incompletely determined for some subjects. For example, suppose that after a five-year follow-up study of survival after myocardial infarction a patient is still alive. That patient's survival time is *censored* on the right at five years; that is, her survival time is known only to exceed five years. The response value to be used in the analysis is 5+. Censoring can also occur when a subject is lost to follow-up. 2

If no responses are censored, standard regression models for continuous responses could be used to analyze the failure times by writing the expected failure time as a function of one or more predictors, assuming that

F.E. Harrell, Jr., *Regression Modeling Strategies*, Springer Series in Statistics, DOI 10.1007/978-3-319-19425-7_17

the distribution of failure time is properly specified. However, there are still several reasons for studying failure time using the specialized methods of survival analysis.

1. Time to failure can have an unusual distribution. Failure time is restricted to be positive so it has a skewed distribution and will never be normally distributed.
2. The probability of surviving past a certain time is often more relevant than the expected survival time (and expected survival time may be difficult to estimate if the amount of censoring is large).
3. A function used in survival analysis, the hazard function, helps one to understand the mechanism of failure.[308]

Survival analysis is used often in industrial life-testing experiments, and it is heavily used in clinical and epidemiologic follow-up studies. Examples include a randomized trial comparing a new drug with placebo for its ability to maintain remission in patients with leukemia, and an observational study of prognostic factors in coronary heart disease. In the latter example subjects may well be followed for varying lengths of time, as they may enter the study over a period of many years.

When regression models are used for survival analysis, all the advantages of these models can be brought to bear in analyzing failure times. Multiple, independent prognostic factors can be analyzed simultaneously and treatment differences can be assessed while adjusting for heterogeneity and imbalances in baseline characteristics. Also, patterns in outcome over time can be predicted for individual subjects.

Even in a simple well-designed experiment, survival modeling can allow one to do the following in addition to making simple comparisons.

1. Test for and describe interactions with treatment. Subgroup analyses can easily generate spurious results and they do not consider interacting factors in a dose-response manner. Once interactions are modeled, relative treatment benefits can be estimated (e.g., hazard ratios), and analyses can be done to determine if some patients are too sick or too well to have even a relative benefit.
2. Understand prognostic factors (strength and shape).
3. Model absolute effect of treatment. First, a model for the probability of surviving past time t is developed. Then differences in survival probabilities for patients on treatments A and B can be estimated. The differences will be due primarily to sickness (overall risk) of the patient and to treatment interactions.
4. Understand time course of treatment effect. The period of maximum effect or period of any substantial effect can be estimated from a plot of relative effects of treatment over time.
5. Gain power for testing treatment effects.
6. Adjust for imbalances in treatment allocation in non-randomized studies.

17.2 Censoring, Delayed Entry, and Truncation

Responses may be left–censored and interval–censored besides being right–censored. *Interval–censoring* is present, for example, when a measuring device functions only for a certain range of the response; measurements outside that range are censored at an end of the scale of the device. Interval–censoring also occurs when the presence of a medical condition is assessed during periodic exams. When the condition is present, the time until the condition developed is only known to be between the current and the previous exam. *Left–censoring* means that an event is known to have occurred before a certain time. In addition, *left–truncation* and *delayed entry* are common. Nomenclature is confusing as many authors refer to delayed entry as left–truncation. Left–truncation really means that an unknown subset of subjects failed before a certain time and the subjects didn't get into the study. For example, one might study the survival patterns of patients who were admitted to a tertiary care hospital. Patients who didn't survive long enough to be referred to the hospital compose the left-truncated group, and interesting questions such as the optimum timing of admission to the hospital cannot be answered from the data set. 3

Delayed entry occurs in follow-up studies when subjects are exposed to the risk of interest only after varying periods of survival. For example, in a study of occupational exposure to a toxic compound, researchers may be interested in comparing life length of employees with life expectancy in the general population. A subject must live until the beginning of employment before exposure is possible; that is, death cannot be observed before employment. The start of follow-up is delayed until the start of employment and it may be right–censored when follow-up ends. In some studies, a researcher may want to assume that for the purpose of modeling the shape of the hazard function, time zero is the day of diagnosis of disease, while patients enter the study at various times since diagnosis. Delayed entry occurs for patients who don't enter the study until some time after their diagnosis. Patients who die before study entry are left-truncated. Note that the choice of time origin is very important.[53,83,112,133]

Heart transplant studies have been analyzed by considering time zero to be the time of enrollment in the study. Pre-transplant survival is right–censored at the time of transplant. Transplant survival experience is based on delayed entry into the "risk set" to recognize that a transplant patient is not at risk of dying from transplant failure until after a donor heart is found. In other words, survival experience is not credited to transplant surgery until the day of transplant. Comparisons of transplant experience with medical treatment suffer from "waiting time bias" if transplant survival begins on the day of transplant instead of using delayed entry.[209,438,570]

There are several planned mechanisms by which a response is right–censored. *Fixed type I* censoring occurs when a study is planned to end after two years of follow-up, or when a measuring device will only measure responses up to a certain limit. There the responses are observed only if they

fall below a fixed value C. In *type II censoring*, a study ends when there is a pre-specified number of events. If, for example, 100 mice are followed until 50 die, the censoring time is not known in advance.

We are concerned primarily with *random type I right-censoring* in which each subject's event time is observed only if the event occurs before a certain time, but the censoring time can vary between subjects. Whatever the cause of censoring, we assume that the censoring is *non-informative* about the event; that is, the censoring is caused by something that is independent of the impending failure. Censoring is non-informative when it is caused by planned termination of follow-up or by a subject moving out of town for reasons unrelated to the risk of the event. If subjects are removed from follow-up because of a worsening condition, the *informative censoring* will result in biased estimates and inaccurate statistical inference about the survival experience. For example, if a patient's response is censored because of an adverse effect of a drug or noncompliance to the drug, a serious bias can result if patients with adverse experiences or noncompliance are also at higher risk of suffering the outcome. In such studies, efficacy can only be assessed fairly using the *intention to treat principle*: all events should be attributed to the treatment *assigned* even if the subject is later removed from that treatment.

4

17.3 Notation, Survival, and Hazard Functions

In survival analysis we use T to denote the response variable, as the response is usually the time until an event. Instead of defining the statistical model for the response T in terms of the expected failure time, it is advantageous to define it in terms of the *survival function*, $S(t)$, given by

$$S(t) = \text{Prob}\{T > t\} = 1 - F(t), \tag{17.1}$$

where $F(t)$ is the cumulative distribution function for T. If the event is death, $S(t)$ is the probability that death occurs after time t, that is, the probability that the subject will survive at least until time t. $S(t)$ is always 1 at $t = 0$; all subjects survive at least to time zero. The survival function must be non-increasing as t increases. An example of a survival function is shown in Figure 17.1. In that example subjects are at very high risk of the event in the early period so that the $S(t)$ drops sharply. The risk is low for $0.1 \leq t \leq 0.6$, so $S(t)$ is somewhat flat. After $t = .6$ the risk again increases, so $S(t)$ drops more quickly. Figure 17.2 depicts the *cumulative hazard function* corresponding to the survival function in Figure 17.1. This function is denoted by $\Lambda(t)$. It describes the accumulated risk up until time t, and as is shown later, is the negative of the log of the survival function. $\Lambda(t)$ is non-decreasing as t increases; that is, the accumulated risk increases or remains the same. Another important function is the *hazard function*, $\lambda(t)$, also called the *force*

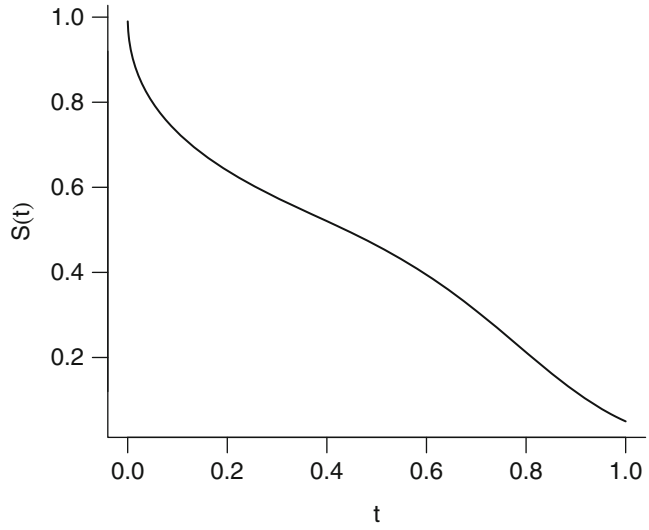


Fig. 17.1 Survival function

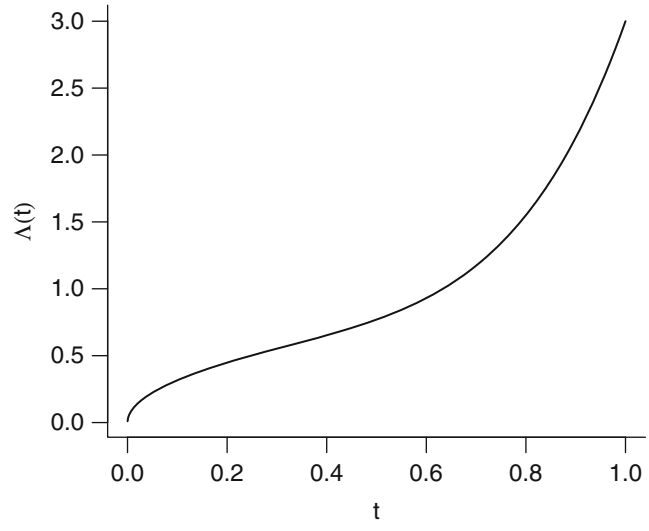


Fig. 17.2 Cumulative hazard function

of mortality, or *instantaneous event (death, failure) rate*. The hazard at time t is related to the probability that the event will occur in a small interval around t, given that the event has not occurred before time t. By studying the event rate at a given time conditional on the event not having occurred by

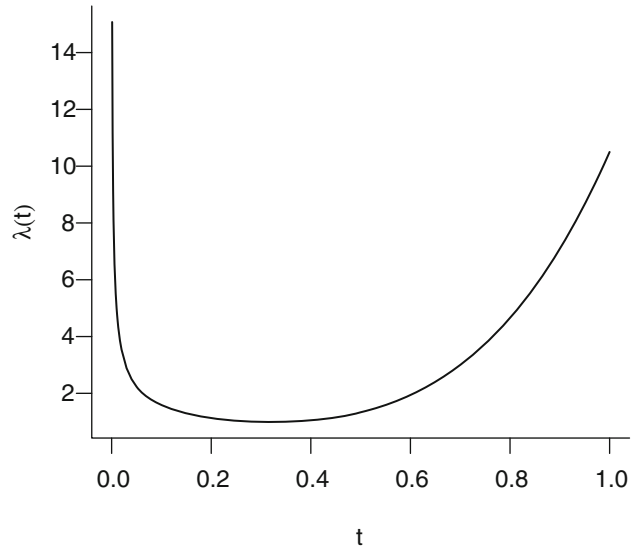


Fig. 17.3 Hazard function

that time, one can learn about the mechanisms and forces of risk over time. Figure 17.3 depicts the hazard function corresponding to $S(t)$ in Figure 17.1 and to $\Lambda(t)$ in Figure 17.2. Notice that the hazard function allows one to more easily determine the phases of increased risk than looking for sudden drops in $S(t)$ or $\Lambda(t)$.

The hazard function is defined formally by

$$\lambda(t) = \lim_{u \to 0} \frac{\text{Prob}\{t < T \leq t + u | T > t\}}{u}, \tag{17.2}$$

which using the law of conditional probability becomes

$$\begin{aligned} \lambda(t) &= \lim_{u \to 0} \frac{\text{Prob}\{t < T \leq t + u\}/\text{Prob}\{T > t\}}{u} \\ &= \lim_{u \to 0} \frac{[F(t+u) - F(t)]/u}{S(t)} \\ &= \frac{\partial F(t)/\partial t}{S(t)} \\ &= \frac{f(t)}{S(t)}, \end{aligned} \tag{17.3}$$

where $f(t)$ is the probability density function of T evaluated at t, the derivative or slope of the cumulative distribution function $1 - S(t)$. Since

$$\frac{\partial \log S(t)}{\partial t} = \frac{\partial S(t)/\partial t}{S(t)} = -\frac{f(t)}{S(t)}, \tag{17.4}$$

the hazard function can also be expressed as

$$\lambda(t) = -\frac{\partial \log S(t)}{\partial t}, \tag{17.5}$$

the negative of the slope of the log of the survival function. Working backwards, the integral of $\lambda(t)$ is:

$$\int_0^t \lambda(v)dv = -\log S(t). \tag{17.6}$$

The integral or area under $\lambda(t)$ is defined to be $\Lambda(t)$, the cumulative hazard function. Therefore

$$\Lambda(t) = -\log S(t), \tag{17.7}$$

or

$$S(t) = \exp[-\Lambda(t)]. \tag{17.8}$$

So knowing any one of the functions $S(t)$, $\Lambda(t)$, or $\lambda(t)$ allows one to derive the other two functions. The three functions are different ways of describing the same distribution.

One property of $\Lambda(t)$ is that the expected value of $\Lambda(T)$ is unity, since if $T \sim S(t)$, the density of T is $\lambda(t)S(t)$ and

$$\begin{aligned} E[\Lambda(T)] &= \int_0^\infty \Lambda(t)\lambda(t)\exp(-\Lambda(t))dt \\ &= \int_0^\infty u\exp(-u)du \\ &= 1. \end{aligned} \tag{17.9}$$

Now consider properties of the distribution of T. The population qth quantile (100qth percentile), T_q, is the time by which a fraction q of the subjects will fail. It is the value t such that $S(t) = 1 - q$; that is

$$T_q = S^{-1}(1 - q). \tag{17.10}$$

The median life length is the time by which half the subjects will fail, obtained by setting $S(t) = 0.5$:

$$T_{0.5} = S^{-1}(0.5). \tag{17.11}$$

The qth quantile of T can also be computed by setting $\exp[-\Lambda(t)] = 1 - q$, giving

$$T_q = \Lambda^{-1}[-\log(1-q)] \quad \text{and as a special case,}$$
$$T_{.5} = \Lambda^{-1}(\log 2). \tag{17.12}$$

The mean or expected value of T (the expected failure time) is the area under the survival function for t ranging from 0 to ∞:

$$\mu = \int_0^\infty S(v)dv. \tag{17.13}$$

Irwin has defined *mean restricted life* (see [334,335]), which is the area under $S(t)$ up to a fixed time (usually chosen to be a point at which there is still adequate follow-up information).

The random variable T denotes a random failure time from the survival distribution $S(t)$. We need additional notation for the response and censoring information for the ith subject. Let T_i denote the response for the ith subject. This response is the time until the event of interest, and it may be censored if the subject is not followed long enough for the event to be observed. Let C_i denote the censoring time for the ith subject, and define the event indicator as

$$\begin{aligned} e_i &= 1 \quad \text{if the event was observed} \quad (T_i \leq C_i), \\ &= 0 \quad \text{if the response was censored} \quad (T_i > C_i). \end{aligned} \tag{17.14}$$

The observed response is

$$Y_i = \min(T_i, C_i), \tag{17.15}$$

which is the time that occurred first: the failure time or the censoring time. The pair of values (Y_i, e_i) contains all the response information for most purposes (i.e., the potential censoring time C_i is not usually of interest if the event occurred before C_i).

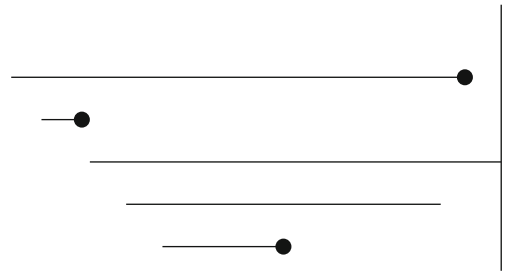
Figure 17.4 demonstrates this notation. The line segments start at study entry (survival time $t = 0$).

A useful property of the cumulative hazard function can be derived as follows. Let z be any cutoff time and consider the expected value of Λ evaluated at the earlier of the cutoff time or the actual failure time.

$$\begin{aligned} E[\Lambda(\min(T, z))] &= E[\Lambda(T)[T \leq z] + \Lambda(z)[T > z]] \\ &= E[\Lambda(T)[T \leq z]] + \Lambda(z)S(z). \end{aligned} \tag{17.16}$$

The first term in the right–hand side is

$$\begin{aligned} &\int_0^\infty \Lambda(t)[t \leq z]\lambda(t)\exp(-\Lambda(t))dt \\ &= \int_0^z \Lambda(t)\lambda(t)\exp(-\Lambda(t))dt \end{aligned} \tag{17.17}$$



T_i	C_i	Y_i	e_i
75	81	75	1
7	76	7	1
68+	68	68	0
52+	52	52	0
20	56	20	1

Termination of Study

Fig. 17.4 Some censored data. Circles denote events.

$$= -[u\exp(-u) + \exp(-u)]|_0^{\Lambda(z)}$$
$$= 1 - S(z)[\Lambda(z) + 1].$$

Adding $\Lambda(z)S(z)$ results in

$$E[\Lambda(\min(T, z))] = 1 - S(z) = F(z). \tag{17.18}$$

It follows that $\sum_{i=1}^{n} \Lambda(\min(T_i, z))$ estimates the expected number of failures occurring before time z among the n subjects. 5

17.4 Homogeneous Failure Time Distributions

In this section we assume that each subject in the sample has the same distribution of the random variable T that represents the time until the event. In particular, there are no covariables that describe differences between subjects in the distribution of T. As before we use $S(t)$, $\lambda(t)$, and $\Lambda(t)$ to denote, respectively, the survival, hazard, and cumulative hazard functions.

The form of the true population survival distribution function $S(t)$ is almost always unknown, and many distributional forms have been used for describing failure time data. We consider first the two most popular parametric survival distributions: the exponential and Weibull distributions. The exponential distribution is a very simple one in which the hazard function is constant; that is, $\lambda(t) = \lambda$. The cumulative hazard and survival functions are then

$$\Lambda(t) = \lambda t \quad \text{and}$$
$$S(t) = \exp(-\Lambda(t)) = \exp(-\lambda t). \tag{17.19}$$

The median life length is $\Lambda^{-1}(\log 2)$ or

$$T_{0.5} = \log(2)/\lambda. \tag{17.20}$$

The time by which 1/2 of the subjects will have failed is then proportional to the reciprocal of the constant hazard rate λ . This is true also of the expected or mean life length, which is $1/\lambda$.

The exponential distribution is one of the few distributions for which a closed-form solution exists for the estimator of its parameter when censoring is present. This estimator is a function of the number of events and the total *person-years* of exposure. Methods based on person-years in fact implicitly assume an exponential distribution. The exponential distribution is often used to model events that occur "at random in time."[323] It has the property that the future lifetime of a subject is the same, no matter how "old" it is, or

$$\text{Prob}\{T > t_0 + t | T > t_0\} = \text{Prob}\{T > t\}. \tag{17.21}$$

This "ageless" property also makes the exponential distribution a poor choice for modeling human survival except over short time periods.

The Weibull distribution is a generalization of the exponential distribution. Its hazard, cumulative hazard, and survival functions are given by

$$\begin{aligned} \lambda(t) &= \alpha\gamma t^{\gamma-1} \\ \Lambda(t) &= \alpha t^{\gamma} \\ S(t) &= \exp(-\alpha t^{\gamma}). \end{aligned} \tag{17.22}$$

The Weibull distribution with $\gamma = 1$ is an exponential distribution (with constant hazard). When $\gamma > 1$, its hazard is increasing with t, and when $\gamma < 1$ its hazard is decreasing. Figure 17.5 depicts some of the shapes of the hazard function that are possible. If T has a Weibull distribution, the median of T is

$$T_{0.5} = [(\log 2)/\alpha]^{1/\gamma}. \tag{17.23}$$

There are many other traditional parametric survival distributions, some of which have hazards that are "bathtub shaped" as in Figure 17.3.[243,323] The restricted cubic spline function described in Section 2.4.5 is an alternative basis for $\lambda(t)$.[286,287] This function family allows for any shape of smooth $\lambda(t)$ since the number of knots can be increased as needed, subject to the number of events in the sample. Nonlinear terms in the spline function can be tested to assess linearity of hazard (Rayleigh-ness) or constant hazard (exponentiality).

6

The restricted cubic spline hazard model with k knots is

$$\lambda_k(t) = a + bt + \sum_{j=1}^{k-2} \gamma_j w_j(t), \tag{17.24}$$

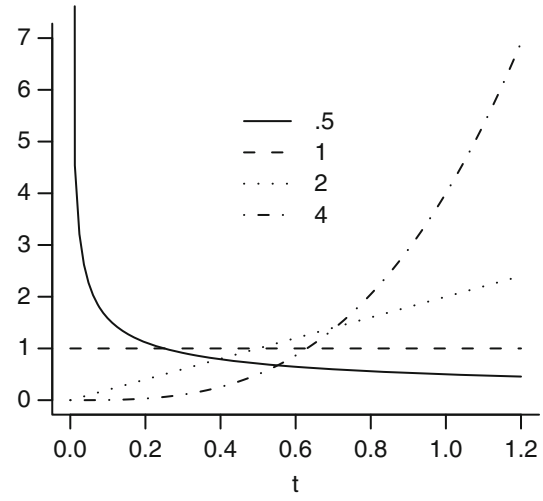


Fig. 17.5 Some Weibull hazard functions with $\alpha = 1$ and various values of γ.

where the $w_j(t)$ are the restricted cubic spline terms of Equation 2.25. There terms are cubic terms in t. A set of knots $v_1, \ldots, v_k$ is selected from the quantiles of the uncensored failure times (see Section 2.4.5 and [286]).

The cumulative hazard function for this model is

$$\Lambda(t) = at + \frac{1}{2}t^2 + \frac{1}{4} \times \text{ quartic terms in t.} \tag{17.25}$$

Standard maximum likelihood theory is used to obtain estimates of the k unknown parameters to derive, for example, smooth estimates of $\lambda(t)$ with confidence bands. The flexible estimates of $S(t)$ using this method are as efficient as Kaplan–Meier estimates, but they are smooth and can be used as a basis for modeling predictor variables. The spline hazard model is particularly useful for fitting steeply falling and gently rising hazard functions that are characteristic of high-risk medical procedures.

17.5 Nonparametric Estimation of S and Λ

17.5.1 Kaplan–Meier Estimator

As the true form of the survival distribution is seldom known, it is useful to estimate the distribution without making any assumptions. For many analyses, this may be the last step, while in others this step helps one select a statistical model for more in-depth analyses. When no event times are censored, a nonparametric estimator of $S(t)$ is $1 - F_n(t)$ where $F_n(t)$ is the usual

Table 17.1 Kaplan-Meier computations

Day	No. Subjects At Risk	Deaths	Censored	Cumulative Survival
12	100	1	0	$99/100 = .99$
30	99	2	1	$97/99 \times 99/100 = .97$
60	96	0	3	$96/96 \times .97 = .97$
72	93	3	0	$90/93 \times .97 = .94$
.	.	.	.	.
.	.	.	.	.

empirical cumulative distribution function based on the observed failure times $T_1, \ldots, T_n$. Let $S_n(t)$ denote this empirical survival function. $S_n(t)$ is given by the fraction of observed failure times that exceed t:

$$S_n(t) = [\text{number of } T_i > t]/n. \tag{17.26}$$

When censoring is present, $S(t)$ can be estimated (at least for t up until the end of follow-up) by the Kaplan–Meier[333] product-limit estimator. This method is based on conditional probabilities. For example, suppose that every subject has been followed for 39 days or has died within 39 days so that the proportion of subjects surviving at least 39 days can be computed. After 39 days, some subjects may be lost to follow-up besides those removed from follow-up because of death within 39 days. The proportion of those still followed 39 days who survive day 40 is computed. The probability of surviving 40 days from study entry equals the probability of surviving day 40 after living 39 days, multiplied by the chance of surviving 39 days.

The life table in Table 17.1 demonstrates the method in more detail. We suppose that 100 subjects enter the study and none die or are lost before day 12.

Times in a life table should be measured as precisely as possible. If the event being analyzed is death, the failure time should usually be specified to the nearest day. We assume that deaths occur on the day indicated and that being censored on a certain day implies the subject survived through the end of that day. The data used in computing Kaplan–Meier estimates consist of $(Y_i, e_i), i = 1, 2, \ldots, n$ using notation defined previously. Primary data collected to derive (Y_i, e_i) usually consist of entry date, event date (if subject failed), and censoring date (if subject did not fail). Instead, the entry date, date of event/censoring, and event/censoring indicator e_i may be specified.

The Kaplan–Meier estimator is called the product-limit estimator because it is the limiting case of actuarial survival estimates as the time periods shrink so that an entry is made for each failure time. An entry need not be in the table for censoring times (when no failures occur at that time) as long as the number of subjects censored is subtracted from the next number

Table 17.2 Summaries used in Kaplan-Meier computations

i	t_i	n_i	d_i	$(n_i - d_i)/n_i$
1	1	7	1	6/7
2	3	6	2	4/6
3	9	2	1	1/2

at risk. Kaplan–Meier estimates are preferred to actuarial estimates because they provide more resolution and make fewer assumptions. In constructing a yearly actuarial life table, for example, it is traditionally assumed that subjects censored between two years were followed 0.5 years.

The product-limit estimator is a nonparametric maximum likelihood estimator [331, pp. 10–13]. The formula for the Kaplan–Meier product-limit estimator of $S(t)$ is as follows. Let k denote the number of failures in the sample and let $t_1, t_2, \ldots, t_k$ denote the unique event times (ordered for ease of calculation). Let d_i denote the number of failures at t_i and n_i be the number of subjects *at risk* at time t_i; that is, n_i = number of failure/censoring times $\geq t_i$. The estimator is then

$$S_{\text{KM}}(t) = \prod_{i:t_i \leq t} (1 - d_i/n_i). \tag{17.27}$$

The Kaplan–Meier estimator of $\Lambda(t)$ is $\Lambda_{\text{KM}}(t) = -\log S_{\text{KM}}(t)$. An estimate of quantile q of failure time is $S^{-1}_{\text{KM}}(1-q)$, if follow-up is long enough so that $S_{\text{KM}}(t)$ drops as low as $1-q$. If the last subject followed failed so that $S_{\text{KM}}(t)$ drops to zero, the expected failure time can be estimated by computing the area under the Kaplan–Meier curve.

To demonstrate computation of $S_{\text{KM}}(t)$, imagine a sample of failure times given by

$$1 \quad 3 \quad 3 \quad 6^+ \quad 8^+ \quad 9 \quad 10^+,$$

where + denotes a censored time. The quantities needed to compute S_{KM} are in Table 17.2. Thus

$$\begin{aligned} S_{\text{KM}}(t) &= 1, \quad 0 \leq t < 1 \\ &= 6/7 = .85, \quad 1 \leq t < 3 \\ &= (6/7)(4/6) = .57, \quad 3 \leq t < 9 \\ &= (6/7)(4/6)(1/2) = .29, \quad 9 \leq t < 10. \end{aligned} \tag{17.28}$$

Note that the estimate of $S(t)$ is undefined for $t > 10$ since not all subjects have failed by $t = 10$ but no follow-up extends beyond $t = 10$. A graph of the Kaplan–Meier estimate is found in Figure 17.6.

```
require(rms)
```

```
tt ← c(1,3,3,6,8,9,10)
stat ← c(1,1,1,0,0,1,0)
S ← Surv(tt, stat)
survplot(npsurv(S ∼ 1), conf="bands", n.risk=TRUE,
         xlab=expression(t))
survplot(npsurv(S ∼ 1, type="fleming-harrington",
                  conf.int=FALSE), add=TRUE, lty=3)
```

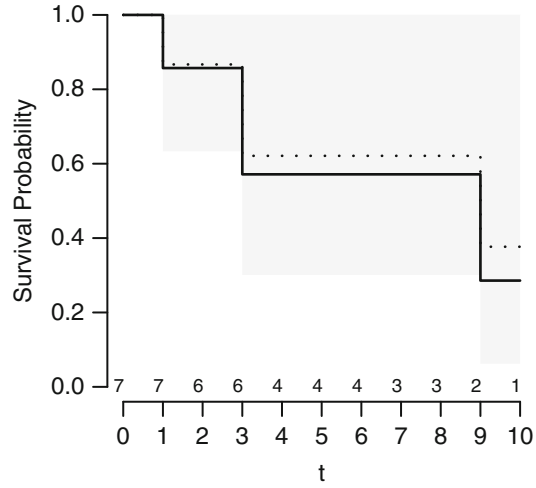


Fig. 17.6 Kaplan–Meier product–limit estimator with 0.95 confidence bands. The Altschuler–Nelson–Aalen–Fleming–Harrington estimator is depicted with the dotted lines.

The variance of $S_{\mathrm{KM}}(t)$ can be estimated using Greenwood's formula [331, p. 14], and using normality of $S_{\mathrm{KM}}(t)$ in large samples this variance can be used to derive a confidence interval for $S(t)$. A better method is to derive an asymmetric confidence interval for $S(t)$ based on a symmetric interval for $\log \Lambda(t)$. This latter method ensures that a confidence limit does not exceed one or fall below zero, and is more accurate since $\log \Lambda_{\mathrm{KM}}(t)$ is more normally distributed than $S_{\mathrm{KM}}(t)$. Once a confidence interval, say $[a, b]$ is determined for $\log \Lambda(t)$, the confidence interval for $S(t)$ is computed by $[\exp\{-\exp(b)\}, \exp\{-\exp(a)\}]$. The formula for an estimate of the variance of interest is [331, p. 15]:

$$\mathrm{Var}\{\log \Lambda_{\mathrm{KM}}(t)\} = \frac{\sum_{i:t_i \leq t} d_i/[n_i(n_i - d_i)]}{\{\sum_{i:t_i \leq t} \log[(n_i - d_i)/n_i]\}^2}. \tag{17.29}$$

Letting s denote the square root of this variance estimate, an approximate $1-\alpha$ confidence interval for $\log \Lambda(t)$ is given by $\log \Lambda_{\mathrm{KM}}(t) \pm zs$, where z is the $1-\alpha/2$ standard normal critical value. After simplification, the confidence interval for $S(t)$ becomes

$$S_{\mathrm{KM}}(t)^{\exp(\pm zs)}. \tag{17.30}$$

Even though the $\log \Lambda$ basis for confidence limits has theoretical advantages, on the $\log - \log$ scale the estimate of $S(t)$ has the greatest instability where much information is available: when $S(t)$ falls just below 1.0. For that reason, the recommended default confidence limits are on the $\Lambda(t)$ scale using

$$\mathrm{Var}\{\Lambda_{\mathrm{KM}}(t)\} = \sum_{i:t_i \leq t} \frac{d_i}{[n_i(n_i - d_i)]}. \tag{17.31}$$

Letting s denote its square root, an approximate $1-\alpha$ confidence interval for $S(t)$ is given by

$$\exp(\pm zs) S_{\mathrm{KM}}(t), \tag{17.32}$$

truncated to $[0, 1]$.

17.5.2 Altschuler–Nelson Estimator

Altschuler[19], Nelson[472], Aalen[1] and Fleming and Harrington[196] proposed estimators of $\Lambda(t)$ or of $S(t)$ based on an estimator of $\Lambda(t)$:

$$\hat{\Lambda}(t) = \sum_{i:t_i \leq t} \frac{d_i}{n_i}$$

$$S_{\Lambda}(t) = \exp(-\hat{\Lambda}(t)). \tag{17.33}$$

$S_{\Lambda}(t)$ has advantages over $S_{\mathrm{KM}}(t)$. First, $\sum_{i=1}^{n} \hat{\Lambda}(Y_i) = \sum_{i=1}^{n} e_i$ [605, Appendix 3]. In other words, the estimator gives the correct expected number of events. Second, there is a wealth of asymptotic theory based on the Altschuler–Nelson estimator.[196]

See Figure 17.6 for an example of the $S_{\Lambda}(t)$ estimator. This estimator has the same variance as $S_{\mathrm{KM}}(t)$ for large enough samples.

17.6 Analysis of Multiple Endpoints

Clinical studies frequently assess multiple endpoints. A cancer clinical trial may, for example, involve recurrence of disease and death, whereas a cardiovascular trial may involve nonfatal myocardial infarction and death. Endpoints may be combined, and the new event (e.g., time until infarction or

death) may be analyzed with any of the tools of survival analysis because only the usual censoring mechanism is used. Sometimes the various endpoints may need separate study, however, because they may have different risk factors.

When the multiple endpoints represent multiple causes of a terminating event (e.g., death), Prentice et al. have developed standard methods for analyzing cause-specific hazards[513] [331, pp. 163–178]. Their methods allow each cause of failure to be analyzed separately, censoring on the other causes. They do not assume any mechanism for cause removal nor make any assumptions regarding the interrelation among causes of failure. However, analyses of competing events using data where some causes of failure are removed in a different way from the original dataset will give rise to different inferences.

When the multiple endpoints represent a mixture of fatal and nonfatal outcomes, the analysis may be more complex. The same is true when one wishes to jointly study an event-time endpoint and a repeated measurement.

9

17.6.1 Competing Risks

When events are independent, each event may also be analyzed separately by censoring on all other events as well as censoring on loss to follow-up. This will yield an unbiased estimate of an easily interpreted cause-specific $\lambda(t)$ or $S(t)$ because censoring is non-informative [331, pp. 168–169]. One minus $S_{\mathrm{KM}}(t)$ computed in this manner will correctly estimate the probability of failing from the event in the absence of other events. Even when the competing events are not independent, the cause-specific hazard model may lead to valid results, but the resulting model does not allow one to estimate risks conditional on removal of one or more causes of the event. See Kay[340] for a nice example of competing risks analysis when a treatment reduces the risk of death from one cause but increases the risk of death from another cause.

10

Larson and Dinse[376] have an interesting approach that jointly models the time until (any) failure and the failure type. For r failure types, they use an r-category polytomous logistic model to predict the probability of failing from each cause. They assume that censoring is unrelated to cause of event.

17.6.2 Competing Dependent Risks

In many medical and epidemiologic studies one is interested in analyzing multiple causes of death. If the goal is to estimate cause-specific failure probabilities, treating subjects dying from extraneous causes as censored and then computing the ordinary Kaplan–Meier estimate results in biased (high) survival estimates[212,225]. If cause m is of interest, the cause-specific hazard

function is defined as

$$\lambda_m(t) = \lim_{u \to 0} \frac{\Pr\{\text{fail from cause } m \text{ in } [t, t+u) | \text{alive at } t\}}{u}. \tag{17.34}$$

The *cumulative incidence function* or probability of failure from cause m by time t is given by

$$F_m(t) = \int_0^t \lambda_m(u) S(u) du, \tag{17.35}$$

where $S(u)$ is the probability of surviving (ignoring cause of death), which equals $\exp[-\int_0^u (\sum \lambda_m(x))dx]$ [212]; [444, Chapter 10]; [102,408]. As previously mentioned, $1 - F_m(t) = \exp[-\int_0^t \lambda_m(u)du]$ only if failures due to other causes are eliminated and if the cause-specific hazard of interest remains unchanged in doing so.[212]

Again letting $t_1, t_2, \ldots, t_k$ denote the unique ordered failure times, a nonparametric estimate of $F_m(t)$ is given by

$$\hat{F}_m(t) = \sum_{i:t_i \leq t} \frac{d_{mi}}{n_i} S_{\text{KM}}(t_{i-1}), \tag{17.36}$$

where d_{mi} is the number of failures of type m at time t_i and n_i is the number of subjects at risk of failure at t_i.

Pepe and others[494,496,497] showed how to use a combination of Kaplan–Meier estimators to derive an estimator of the probability of being free of event 1 by time t given event 2 has not occurred by time t (see also [349]). Let T_1 and T_2 denote, respectively, the times until events 1 and 2. Let $S_1(t)$ and $S_2(t)$ denote, respectively, the two survival functions. Let us suppose that event 1 is not a terminating event (e.g., is not death) and that even after event 1 subjects are followed to ascertain occurrences of event 2. The probability that $T_1 > t$ given $T_2 > t$ is

$$\begin{aligned} \text{Prob}\{T_1 > t | T_2 > t\} &= \frac{\text{Prob}\{T_1 > t \text{ and } T_2 > t\}}{\text{Prob}\{T_2 > t\}} \\ &= \frac{S_{12}(t)}{S_2(t)}, \end{aligned} \tag{17.37}$$

where $S_{12}(t)$ is the survival function for $\min(T_1, T_2)$, the earlier of the two events. Since $S_{12}(t)$ does not involve any informative censoring (assuming as always that loss to follow-up is non-informative), S_{12} may be estimated by the Kaplan–Meier estimator $S_{\text{KM}_{12}}$ (or by S_Λ). For the type of event 1 we have discussed above, S_2 can also be estimated without bias by S_{KM_2}. Thus we estimate, for example, the probability that a subject still alive at time t will be free of myocardial infarction as of time t by $S_{\text{KM}_{12}}/S_{\text{KM}_2}$.

Another quantity that can easily be computed from ordinary survival estimates is $S_2(t) - S_{12}(t) = [1 - S_{12}(t)] - [1 - S_2(t)]$, which is the probability that event 1 occurs by time t and that event 2 has not occurred by time t.

The ratio estimate above is used to estimate the survival function for one event given that another has not occurred. Another function of interest is the *crude survival function* which is a *marginal* distribution; that is, it is the probability that $T_1 > t$ whether or not event 2 occurs:[362]

$$\begin{aligned} S_c(t) &= 1 - F_1(t) \\ F_1(t) &= \text{Prob}\{T_1 \leq t\}, \end{aligned} \tag{17.38}$$

where $F_1(t)$ is the *crude incidence function* defined previously. Note that the $T_1 \leq t$ implies that the occurrence of event 1 is part of the probability being computed. If event 2 is a terminating event so that some subjects can never suffer event 1, the crude survival function for T_1 will never drop to zero. The crude survival function can be interpreted as the survival distribution of W where $W = T_1$ if $T_1 < T_2$ and $W = \infty$ otherwise.[362]

11

17.6.3 State Transitions and Multiple Types of Nonfatal Events

In many studies there is one final, absorbing state (death, all causes) and multiple live states. The live states may represent different health states or phases of a disease. For example, subjects may be completely free of cancer, have an isolated tumor, metastasize to a distant organ, and die. Unlike this example, the live states need not have a definite ordering. One may be interested in estimating *transition probabilities*, for example, the probability $\pi_{ij}(t_1, t_2)$ that an individual in state i at time t_1 is in state j after an additional time t_2. Strauss and Shavelle[596] have developed an extended Kaplan–Meier estimator for this situation. Let $S^i_{KM}(t|t_1)$ denote the ordinary Kaplan–Meier estimate of the probability of not dying before time t (ignoring distinctions between multiple live states) for a cohort of subjects beginning follow-up at time t_1 in state i. This is an estimate of the probability of surviving an additional t time units (in any live state) given that the subject was alive and in state i at time t_1. Strauss and Shavelle's estimator is given by

$$\pi_{ij}(t_1, t_2) = \frac{n_{ij}(t_1, t_2)}{n_i(t_1, t_2)} S^i_{KM}(t_2|t_1), \tag{17.39}$$

where $n_i(t_1, t_2)$ is the number of subjects in live state i at time t_1 who are alive and uncensored t_2 time units later, and $n_{ij}(t_1, t_2)$ is the number of such subjects in state j t_2 time units beyond t_1.

12

17.6.4 Joint Analysis of Time and Severity of an Event

In some studies, an endpoint is given more weight if it occurs earlier or if it is more severe clinically, or both. For example, the event of interest may be myocardial infarction, which may be of any severity from minimal damage to the left ventricle to a fatal infarction. Berridge and Whitehead[52] have provided a promising model for the analysis of such endpoints. Their method assumes that the severity of endpoints which do occur is measured on an ordinal categorical scale and that severity is assessed at the time of the event. Berridge and Whitehead's example was time until first headache, with severity of headaches graded on an ordinal scale. They proposed a joint hazard of an individual who responds with ordered category j:

$$\lambda_j(t) = \lambda(t)\pi_j(t), \tag{17.40}$$

where $\lambda(t)$ is the hazard for the failure time and $\pi_j(t)$ is the probability of an individual having event severity j given she fails at time t. Note that a shift in the distribution of response severity is allowed as the time until the event increases.

13

17.6.5 Analysis of Multiple Events

It is common to choose as an endpoint in a clinical trial an event that can recur. Examples include myocardial infarction, gastric ulcer, pregnancy, and infection. Using only the time until the first event can result in a loss of statistical information and power.[a] There are specialized multivariate survival models (whose assumptions are extremely difficult to verify) for handling this setup, but in many cases a simpler approach will be efficient.

The simpler approach involves modeling the marginal distribution of the time until each event.[407,495] Here one forms one record per subject per event, and the survival time is the time to the first event for the first record, or is the time from the previous event to the next event for all later records. This approach yields consistent estimates of distribution parameters as long as the marginal distributions are correctly specified.[655] One can allow the number of previous events to influence the hazard function of another event by modeling this count as a covariable.

The multiple events within subject are not independent, so variance estimates must be corrected for intracluster correlation. The clustered sandwich covariance matrix estimator described in Section 9.5 and in [407] will provide

[a] An exception to this is the case in which once an event occurs for the first time, that event is likely to recur multiple times for any patient. Then the latter occurrences are redundant.

consistent estimates of variances and covariances even if the events are dependent. Lin[407] also discussed how this method can easily be used to model multiple events of differing types.

14

17.7 R Functions

The `event.chart` function of Lee et al.[394] will draw a variety of charts for displaying raw survival time data, for both single and multiple events per subject. Relationships with covariables can also be displayed. The `event.history` function of Dubin et al.[166] draws an event history graph for right-censored survival data, including time-dependent covariate status. These function are in the `Hmisc` package.

The analyses described in this chapter can be viewed as special cases of the Cox proportional hazards model.[132] The programs for Cox model analyses described in Section 20.13 can be used to obtain the results described here, as long as there is at least one stratification factor in the model. There are, however, several R functions that are pertinent to the homogeneous or stratified case. The R function `survfit`, and its particular renditions of the print, plot, lines, and points generic functions (all part of the `survival` package written by Terry Therneau), will compute, print, and plot Kaplan–Meier and Nelson survival estimates. Confidence intervals for $S(t)$ may be based on S, Λ, or $\log \Lambda$. The `rms` package's front-end to the `survival` package's `survfit` function is `npsurv` for "nonparametric survival". It and other functions described in later chapters use Therneau's `Surv` function to combine the response variable and event indicator into a single R "survival time" object. In its simplest form, use `Surv(y, event)`, where `y` is the failure/right–censoring time and `event` is the event/censoring indicator, usually coded `T/F`, 0 = censored 1 = event or 1 = censored 2 = event. If the event status variable has other coding (e.g., 3 means death), use `Surv(y, s==3)`. To handle interval time-dependent covariables, or to use Andersen and Gill's *counting process* formulation of the Cox model,[23] use the notation `Surv(tstart, tstop, status)`. The counting process notation allows subjects to enter and leave risk sets at random. For each time interval for each subject, the interval is made up of `tstart`$< t \leq$`tstop`. For time-dependent stratification, there is an optional `origin` argument to `Surv` that indicates the hazard shape time origin at the time of crossover to a new stratum. A `type` argument is used to handle left– and interval–censoring, especially for parametric survival models. Possible values of `type` are `"right","left","interval","counting","interval2","mstate"`.

The `Surv` expression will usually be used inside another function, but it is fine to save the result of `Surv` in another object and to use this object in the particular fitting function.

`npsurv` is invoked by the following, with default parameter settings indicated.

```
require(rms)
units(y) ← "Month"
# Default is "Day" - used for axis labels, etc.
npsurv(Surv(y, event) ~ svar1 + svar2 + ... , data, subset,
       type=c("kaplan-meier", "fleming-harrington", "fh2"),
       error=c("greenwood", "tsiatis"), se.fit=TRUE,
       conf.int=.95,
       conf.type=c("log","log-log","plain","none"), ...)
```

If there are no stratification variables (`svar1`, ...), omit them. To print a table of estimates, use

```
f ← npsurv(...)
print(f)    # print brief summary of f
summary(f, times, censored=FALSE)   # in survival
```

For failure times stored in days, use

```
f ← npsurv(Surv(futime, event) ~ sex)
summary(f, seq(30, 180, by=30))
```

to print monthly estimates.

There is a `plot` method To plot the object returned by `survfit` and `npsurv`. This invokes `plot.survfit`.

Objects created by `npsurv` can be passed to the more comprehensive plotting function `survplot` (here, actually `survplot.npsurv`) for other options that include automatic curve labeling and showing the number of subjects at risk at selected times. See Figure 17.6 for an example. Stratified estimates, with four treatments distinguished by line type and curve labels, could be drawn by

```
units(y) ← "Year"
f ← npsurv(Surv(y, stat) ~ treatment)
survplot(f, ylab="Fraction Pain-Free")
```

The `groupkm` in `rms` computes and optionally plots $S_{\mathrm{KM}}(u)$ or $\log \Lambda_{\mathrm{KM}}(u)$ (if `loglog=TRUE`) for fixed u with automatic stratification on a continuous predictor `x`. As in `cut2` (Section 6.2) you can specify the number of subjects per interval (default is `m=50`), the number of quantile groups (`g`), or the actual cutpoints (`cuts`). `groupkm` plots the survival or log–log survival estimate against mean `x` in each `x` interval.

The `bootkm` function in the `Hmisc` package bootstraps Kaplan–Meier survival estimates or Kaplan–Meier estimates of quantiles of the survival time distribution. It is easy to use `bootkm` to compute, for example, a nonparametric confidence interval for the ratio of median survival times for two groups.

See the Web site for a list of functions from other users for nonparametric estimation of $S(t)$ with left–, right–, and interval–censored data. The adaptive linear spline log-hazard fitting function `heft`[361] is freely available.

17.8 Further Reading

1. Some excellent general references for survival analysis are [57, 83, 114, 133, 154, 197, 282, 308, 331, 350, 382, 392, 444, 484, 574, 604]. Govindarajulu et al.[229] have a nice review of frailty models in survival analysis, for handling clustered time-to-event data.
2. See Goldman,[220] Bull and Spiegelhalter,[83], Lee et al.[394], and Dubin et al.[166] for ways to construct descriptive graphs depicting right–censored data.
3. Some useful references for left–truncation are [83,112,244,524]. Mandel[435] carefully described the difference between censoring and truncation.
4. See [384, p. 164] for some ideas for detecting informative censoring. Bilker and Wang[54] discuss *right–truncation* and contrast it with right–censoring.
5. Arjas[29] has applications based on properties of the cumulative hazard function.
6. Kooperberg et al.[361, 594] have an adaptive method for fitting hazard functions using linear splines in the log hazard. Binquet et al.[56] studied a related approach using quadratic splines. Mudholkar et al.[466] presented a generalized Weibull model allowing for a variety of hazard shapes.
7. Hollander et al.[299] provide a nonparametric *simultaneous* confidence band for $S(t)$, surprisingly using likelihood ratio methods. Miller[459] showed that if the parametric form of $S(t)$ is known to be Weibull with known shape parameter (an unlikely scenario), the Kaplan–Meier estimator is very inefficient (i.e., has high variance) when compared with the parametric maximum likelihood estimator. See [666] for a discussion of how the efficiency of Kaplan–Meier estimators can be improved by interpolation as opposed to piecewise flat step functions. That paper also discusses a variety of other estimators, some of which are significantly more efficient than Kaplan–Meier.
8. See [112,244,438,570,614,619] for methods of estimating S or Λ in the presence of left–truncation. See Turnbull[616] for nonparametric estimation of $S(t)$ with left–, right–, and interval–censoring, and Kooperberg and Clarkson[360] for a flexible parametric approach to modeling that allows for interval–censoring. Lindsey and Ryan[413] have a nice tutorial on the analysis of interval–censored data.
9. Hogan and Laird[297, 298] developed methods for dealing with mixtures of fatal and nonfatal outcomes, including some ideas for handling outcome-related dropouts on the repeated measurements. See also Finkelstein and Schoenfeld.[193] The 30 April 1997 issue of *Statistics in Medicine* (Vol. 16) is devoted to methods for analyzing multiple endpoints as well as designing multiple endpoint studies. The papers in that issue are invaluable, as is Therneau and Hamilton[606] and Therneau and Grambsch.[604] Huang and Wang[311] presented a joint model for recurrent events and a terminating event, addressing such issues as the frequency of recurrent events by the time of the terminating event.
10. See Lunn and McNeil[429] and Marubini and Valsecchi [444, Chapter 10] for practical approaches to analyzing competing risks using ordinary Cox proportional hazards models. A nice overview of competing risks with comparisons of various approaches is found in Tai et al.[599], Geskus[214], and Koller et al.[358]. Bryant and Dignam[78] developed a semiparametric procedure in which competing risks are adjusted for nonparametrically while a parametric cumulative incidence function is used for the event of interest, to gain precision. Fine and Gray[192] developed methods for analyzing competing risks by estimating subdistribution functions. Nishikawa et al.[478] developed some novel approaches to competing risk analysis involving time to adverse drug events competing with time to withdrawal from therapy. They also dealt with different severities of events in an interesting way. Putter et al.[517] has a nice tutorial on competing risks, multi-state models, and associated R software. Fiocco et al.[194] developed

an approach to avoid the problems caused by having to estimate a large number of regression coefficients in multi-state models. Ambrogi et al.[22] provide clinically useful estimates from competing risks analyses.

11 Jiang, Chappell, and Fine[322] present methods for estimating the distribution of event times of nonfatal events in the presence of terminating events such as death.

12 Shen and Thall[568] have developed a flexible parametric approach to multi-state survival analysis.

13 Lancar et al.[372] developed a method for analyzing repeated events of varying severities.

14 Lawless and Nadeau[384] have a very good description of models dealing with recurrent events. They use the notion of the *cumulative mean function*, which is the expected number of events experienced by a subject by a certain time. Lawless[383] contrasts this approach with other approaches. See Aalen et al.[3] for a nice example in which multivariate failure times (time to failure of fillings in multiple teeth per subject) are analyzed. Francis and Fuller[204] developed a graphical device for depicting complex event history data. Therneau and Hamilton[606] have very informative comparisons of various methods for modeling multiple events, showing the importance of whether the analyst starts the clock over after each event. Kelly and Lim[343] have another very useful paper comparing various methods for analyzing recurrent events. Wang and Chang[650] demonstrated the difficulty of using Kaplan–Meier estimates for recurrence time data.

17.9 Problems

1. Make a rough drawing of a hazard function from birth for a man who develops significant coronary artery disease at age 50 and undergoes coronary artery bypass surgery at age 55.
2. Define in words the relationship between the hazard function and the survival function.
3. In a study of the life expectancy of light bulbs as a function of the bulb's wattage, 100 bulbs of various wattage ratings were tested until each had failed. What is wrong with using the product-moment linear correlation test to test whether wattage is associated with life length concerning (a) distributional assumptions and (b) other assumptions?
4. A placebo-controlled study is undertaken to ascertain whether a new drug decreases mortality. During the study, some subjects are withdrawn because of moderate to severe side effects. Assessment of side effects and withdrawal of patients is done on a blinded basis. What statistical technique can be used to obtain an unbiased treatment comparison of survival times? State at least one efficacy endpoint that can be analyzed unbiasedly.
5. Consider long-term follow-up of patients in the `support` dataset. What proportion of the patients have censored survival times? Does this imply that one cannot make accurate estimates of chances of survival? Make a histogram or empirical distribution function estimate of the *censored* follow-up times. What is the typical follow-up duration for a patient in the study

who has survived so far? What is the typical survival time for patients who have died? Taking censoring into account, what is the median survival time from the Kaplan–Meier estimate of the overall survival function? Estimate the median graphically or using any other sensible method.

6. Plot Kaplan–Meier survival function estimates stratified by `dzclass`. Estimate the median survival time and the first quartile of time until death for each of the four disease classes.
7. Repeat Problem 6 except for tertiles of `meanbp`.
8. The commonly used log-rank test for comparing survival times between groups of patients is a special case of the test of association between the grouping variable and survival time in a Cox proportional hazards regression model. Depending on how one handles tied failure times, the log-rank χ^2 statistic exactly equals the score χ^2 statistic from the Cox model, and the likelihood ratio and Wald χ^2 test statistics are also appropriate. To obtain global score or LR χ^2 tests and P-values you can use a statement as the following, where `cph` is in the `rms` package. It is similar to the `survival` package's `coxph` function.

```
cph(Survobject ~ predictor)
```

   Here `Survobject` is a survival time object created by the `Surv` function. Obtain the log-rank (score) χ^2 statistic, degrees of freedom, and P-value for testing for differences in survival time between levels of `dzclass`. Interpret this test, referring to the graph you produced in Problem 6 if needed.
9. Do preliminary analyses of survival time using the Mayo Clinic primary biliary cirrhosis dataset described in Section 8.9. Make graphs of Altschuler–Nelson or Kaplan–Meier survival estimates stratified separately by a few categorical predictors and by categorized versions of one or two continuous predictors. Estimate median failure time for the various strata. You may want to suppress confidence bands when showing multiple strata on one graph. See [361] for parametric fits to the survival and hazard function for this dataset.