

Chapter 15

Survival Analysis

This chapter reviews some of the fundamental concepts and basic methods in survival analysis. Frequently, event rates such as mortality or occurrence of nonfatal myocardial infarction are selected as primary response variables. The analysis of such event rates in two groups could employ the chi-square statistic or the equivalent normal statistic for the comparison of two proportions. However, when the length of observation is different for each participant, estimating an event rate is more complicated. Furthermore, simple comparison of event rates between two groups is not necessarily the most informative type of analysis. For example, the 5-year survival for two groups may be nearly identical, but the survival rates may be quite different at various times during the 5 years. This is illustrated by the survival curves in Fig. 15.1. This figure shows survival probability on the vertical axis and time on the horizontal axis. For Group A, the survival rate (or one minus the mortality rate) declines steadily over the 5 years of observation. For Group B, however, the decline in the survival rate is rapid during the first year and then levels off. Obviously, the survival experience of the two groups is not the same, although the mortality rate at 5 years is nearly the same. If only the 5-year survival rate is considered, Group A and Group B appear equivalent. Curves such as these might reasonably be expected in a trial of surgical versus medical intervention, where surgery might carry a high initial operative mortality.

Fundamental Point

Survival analysis methods are important in trials where participants are entered over a period of time and have various lengths of follow-up. These methods permit the comparison of the entire survival experience during the follow-up and may be used for the analysis of time to any dichotomous response variable such as a nonfatal event or an adverse event.

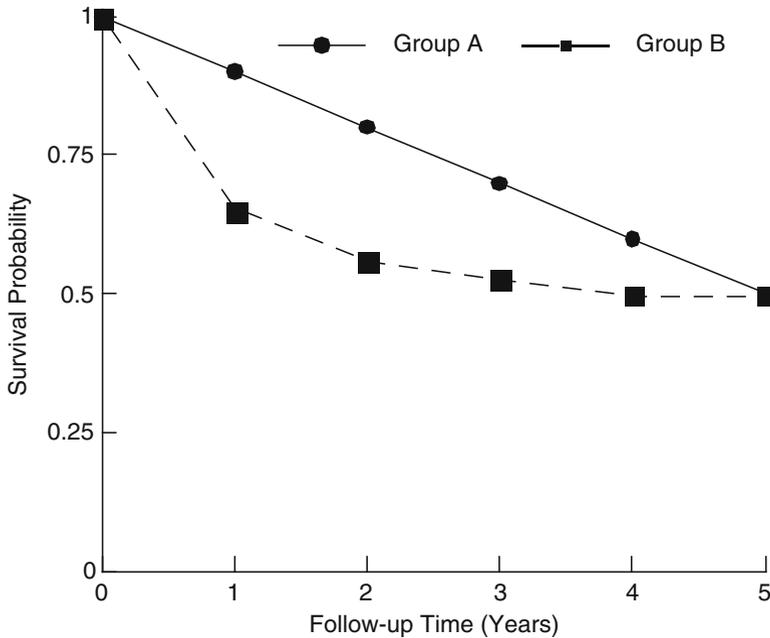


Fig. 15.1 Survival experience for two groups (*A* and *B*)

A review of the basic techniques of survival analysis can be found in elementary statistical textbooks [1–6] as well as in overview papers [7]. A more complete and technical review is in other texts [8–11]. Many methodological advances in the field have occurred and this book will not be able to cover all developments. The following discussion will concern two basic aspects: first, estimation of the survival experience or survival curve for a group of participants in a clinical trial and second, comparison of two survival curves to test whether the survival experience is significantly different. Although the term survival analysis is used, the methods can be applied to any dichotomous response variable when the time from enrollment to the time of the event, not just the fact of its occurrence, is an important consideration. For ease of communication, we shall use the term event, unless death is specifically the event.

Estimation of the Survival Curve

The graphical presentation of the total survival experience during the period of observation is called the survival curve, and the tabular presentation is called the lifetable. In the sample size discussion (Chap. 8), we utilized a parametric model to represent a survival curve, denoted $S(t)$, where t is the time of follow-up. A classic parametric form for $S(t)$ is to assume an exponential distribution

$S(t) = e^{-\lambda t} = \exp(-\lambda t)$, where λ is the hazard rate [11]. If we estimate λ , we have an estimate for $S(t)$. One possible estimate for the hazard ratio is the number of observed events divided by the total exposure time of the person at risk of the event. Other estimates are also available and are described later. While this estimate is not difficult to obtain, the hazard rate may not be constant during the trial. If λ is not constant, but rather a function of time, we can define a hazard rate $\lambda(t)$, but now the definition of $S(t)$ is more complicated. Specifically, $S(t) = \exp\left[\int_0^t \lambda(s) ds\right]$, that is, the exponential of the area under the hazard function curve from time 0 to time t . Furthermore, we cannot always be guaranteed that the observed survival data will be described well by the exponential model, even though we often make this assumption for computing sample size. Thus, biostatisticians have relied on parameter-free or non-parametric ways to estimate the survival curve.

This chapter will cover two similar non-parametric methods, the Cutler-Ederer method [12] and the Kaplan-Meier method [13] for estimating the true survival curve or the corresponding lifetable. We use the Cutler-Ederer method to motivate the more flexible Kaplan-Meier method which is the current standard. Before a review of these specific methods, however, it is necessary to explain how the survival experience is typically obtained in a clinical trial and to define some of the associated terminology.

The clinical trial design may, in a simple case, require that all participants be observed for T years. This is referred to as the follow-up or exposure time. If all participants are entered as a single cohort at the same time, the actual period of follow-up is the same for all participants. If, however, as in most clinical trials, the entry of participants is staggered over some recruitment period, then equal periods of follow-up may occur at different calendar times for each participant, as illustrated in Fig. 15.2.

A participant may have a study event during the course of follow-up. The event time is the accumulated time from entry into the study to the event. The interest is

Fig. 15.2 T year follow-up time for four participants with staggered entry

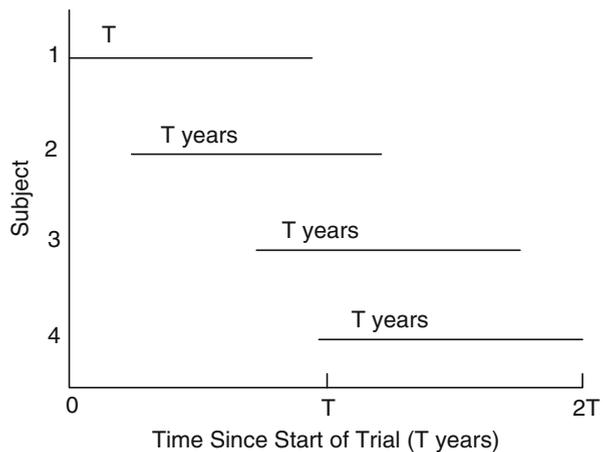


Fig. 15.3 Follow-up experience of four participants with staggered entry: two participants with observed events (*asterisk*) and two participants followed for time T without events (*open circle*).

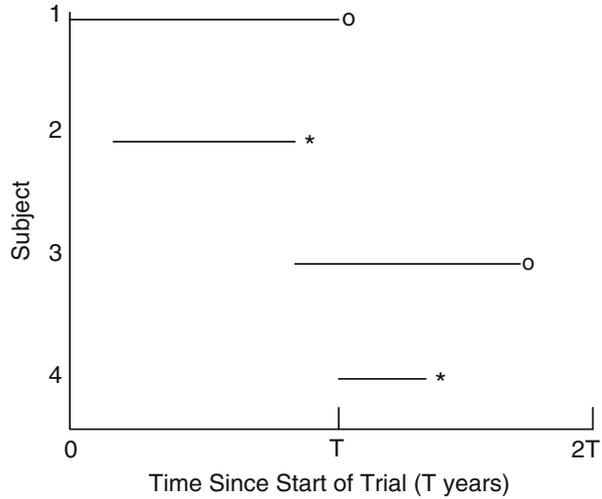
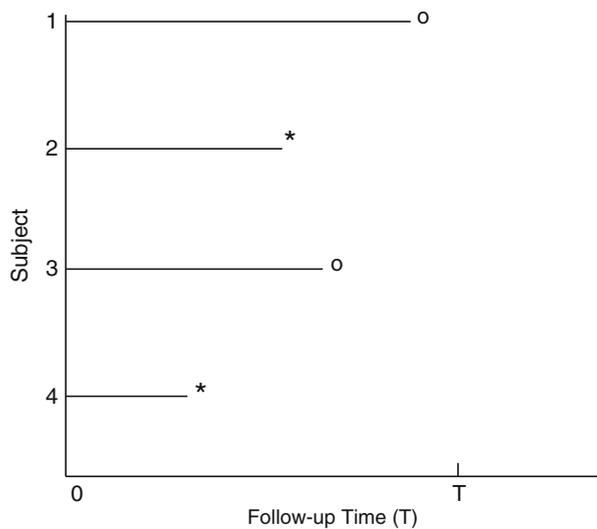


Fig. 15.4 Follow-up experience of four participants with staggered entry converted to a common starting time: two participants with observed events (*asterisk*) and two participants followed for time T without events (*open circle*).



not in the actual calendar date when the event took place but rather the interval of time from entry into the trial until the event. Figures 15.3 and 15.4 illustrate the way the actual survival experience for staggered entry of participants is translated for the analysis. In Fig. 15.3, participants 2 and 4 had an event while participants 1 and 3 did not during the follow-up time. Since, for each participant, only the time interval from entry to the end of the scheduled follow-up period or until an event is of interest, the time of entry can be considered as time zero for each participant. Figure 15.4 illustrates the same survival experience as Fig. 15.3, but the time of entry is considered as time zero.

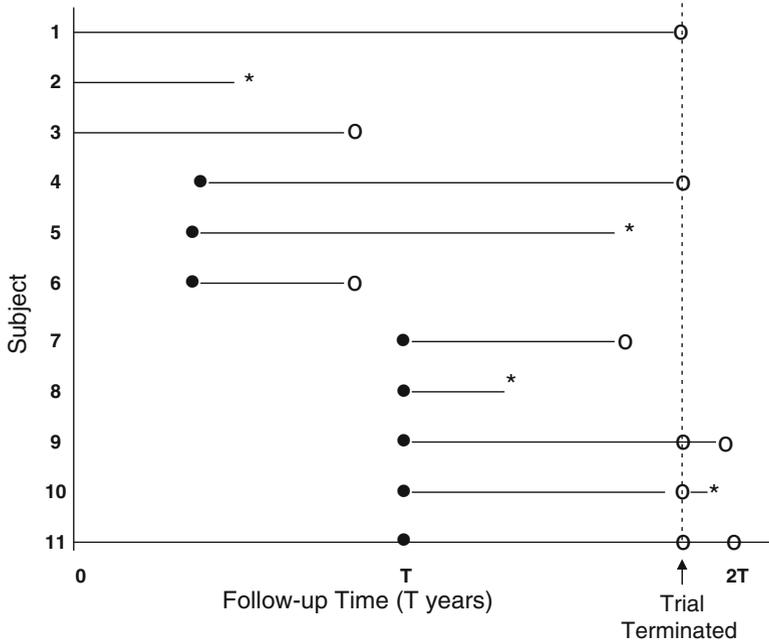


Fig. 15.5 Follow-up experience of 11 participants for staggered entry and a common termination time, with observed events (*asterisk*) and censoring (*open circle*). Follow-up experience beyond the termination time is shown for participants 9 through 11

Some participants may not experience an event before the end of observation. The follow-up time or exposure time for these participants is said to be *censored*; that is, the investigator does not know what happened to these participants after they stopped participating in the trial. Another example of censoring is when participants are entered in a staggered fashion, and the study is terminated at a common date before all participants have had at least their complete T years of follow-up. Later post-trial events from these participants are also unobserved, but the reason for censoring is administrative. Administrative censoring could also occur if a trial is terminated prior to the scheduled time because of early benefits or harmful effects of the intervention. In these cases, censoring is assumed to be independent of occurrence of events.

Figure 15.5 illustrates several of the possibilities for observations during follow-up. Note that in this example the investigator has planned to follow all participants to a common termination time, with each participant being followed for at least T years. The first three participants were randomized at the start of the study. The first participant was observed for the entire duration of the trial with no event, and her survival time was censored because of study termination. The second participant had an event before the end of follow-up. The third participant was lost to follow-up. The second group of three participants was randomized later during the course of the trial with experiences similar to the first group of three. Participants

7 through 11 were randomized late in the study and were not able to be followed for at least T years because the study was terminated early. Participant 7 was lost to follow-up and participant 8 had an event before T years of follow-up time had elapsed and before the study was terminated. Participant 9 was administratively censored but theoretically would have been lost to follow-up had the trial continued. Participant 10 was also censored because of early study termination, although she had an event afterwards which would have been observed had the trial continued to its scheduled end. Finally, the last participant who was censored would have survived for at least T years had the study lasted as long as first planned. The survival experiences illustrated in Fig. 15.5 would all be shifted to have a common starting time equal to zero as in Fig. 15.4. The follow-up time, or the time elapsed from calendar time of entry to calendar time of an event or to censoring could then be analyzed.

In summary then, the investigator needs to record for each participant the time of entry and the time of an event, the time of loss to follow-up, or whether the participant was still being followed without having had an event when the study is terminated. These data will allow the investigator to compute the survival curve.

Cutler-Ederer Estimate

Though the Cutler-Ederer estimate is still in use [14–18], it has been largely replaced as a method for estimation of survival curves by the Kaplan-Meier estimate. Nonetheless it is useful as an introduction to survival curve estimation.

In the Cutler-Ederer or actuarial estimate [12], the assumption is that the deaths and losses are uniformly distributed over a set of fixed-length intervals. On the average, this means that one half the losses will occur during the first half of each interval. The estimate for the probability of surviving the j^{th} interval, given that the previous intervals were survived, is \hat{p}_j , where

$$\hat{p}_j = \frac{n_j - \delta_j - 0.5\lambda_j}{n_j - 0.5\lambda_j}$$

The λ_j losses are assumed to be at risk, on the average, one half the time and thus should be counted as such. These conditional probabilities \hat{p}_j are then multiplied together to obtain an estimate, $\hat{S}(t)$, of the survival function at time t .

Kaplan-Meier Estimate

The Kaplan-Meier Estimate relaxes the assumption of events distributed uniformly across fixed length intervals. Using the time of death, observations can be ranked.

Table 15.1 Participants entered at two points in time (Group *I* and Group *II*) and followed to a common termination time^a

Years of follow-up		Group	
		<i>I</i>	<i>II</i>
1	Participants entered	100	100
	First year deaths	20	25
	First year survivors	80	75
2	Participants entered	80	
	Second year deaths	20	
	Second year survivors	60	

^aAfter Kaplan and Meier [13]

This is a useful improvement, since in a clinical trial with staggered entry of participants and censored observations, survival data will be of varying degrees of completeness.

As a very simple example, suppose that 100 participants were entered into a study and followed for 2 years. One year after the first group was started, a second group of 100 participants was entered and followed for the remaining year of the trial. Assuming no losses to follow-up, the results might be as shown in Table 15.1. For Group *I*, 20 participants died during the first year and of the 80 survivors, 20 more died during the second year. For Group *II*, which was followed for only 1 year, 25 participants died. Now suppose the investigator wants to estimate the 2-year survival rate. The only group of participants followed for 2 years was Group *I*. One estimate of 2-year survival, $S(2)$, would be $\hat{S}(2) = 60/100$ or 0.60. Note that the first-year survival experience of Group *II* is ignored in this estimate. If the investigator wants to estimate 1 year survival rate, $S(1)$, she would observe that a total of 200 participants were followed for at least 1 year. Of those, 155 (80 + 75) survived the first year. Thus, $\hat{S}(1) = 155/200$ or 0.775. If each group were evaluated separately, the survival rates would be 0.80 and 0.75. In estimating the 1-year survival rate, all the available information was used, but for the 2-year survival rate the 1-year survival experience of Group *II* was ignored.

Another procedure for estimating survival rates is to use a conditional probability. For this example, the probability of 2-year survival, $S(2)$, is equal to the probability of 1-year survival, $S(1)$, times the probability of surviving the second year, given that the participant survived the first year, $pr(2|1)$. That is, $S(2) = S(1)pr(2|1)$. In this example, $\hat{S}(1) = 0.775$. The estimate for $pr(2|1)$ is $60/80 = 0.75$ since 60 of the 80 participants who survived the first year also survived the second year. Thus, the estimate for $\hat{S}(2) = 0.775 \times 0.75$ or 0.58, which is slightly different from the previously calculated estimate of 0.60.

Kaplan and Meier [13] described how this conditional probability strategy could be used to estimate survival curves in clinical trials with censored observations. Their procedure is usually referred to as the Kaplan-Meier estimate, or sometimes the product-limit estimate, since the product of conditional probabilities leads to the survival estimate. This procedure assumes that the exact time of entry into the trial is known and that the exact time of the event or loss of follow-up is also known.

For some applications, time to the nearest month may be sufficient, while for other applications the nearest day or hour may be necessary. Kaplan and Meier assumed that a death and loss of follow-up would not occur at the same time. If a death and a loss to follow-up are recorded as having occurred at the same time, this tie is broken on the assumption that the death occurred slightly before the loss to follow-up.

In this method, the follow-up period is divided into intervals of time so that no interval contains both deaths and losses. Let p_j be equal to the probability of surviving the j^{th} interval, given that the participant has survived the previous interval. For intervals labeled j with deaths only, the estimate for p_j , which is \hat{p}_j , is equal to the number of participants alive at the beginning of the j^{th} interval, n_j , minus those who died during the interval, δ_j , with this difference being divided by the number alive at the beginning of the interval, i.e. $\hat{p}_j = (n_j - \delta_j)/n_j$. For an interval j with only l_j losses, the estimate \hat{p}_j is one. Such conditional probabilities for an interval with only losses would not alter the product. This means that an interval with only losses and no deaths may be combined with the previous interval.

Example Suppose 20 participants are followed for a period of 1 year, and to the nearest tenth of a month, deaths were observed at the following times: 0.5, 1.5, 1.5, 3.0, 4.8, 6.2, 10.5 months. In addition, losses to follow-up were recorded at: 0.6, 2.0, 3.5, 4.0, 8.5, 9.0 months. It is convenient for illustrative purposes to list the deaths and losses together in ascending time with the losses indicated in parentheses. Thus, the following sequence is obtained: 0.5, (0.6), 1.5, 1.5, (2.0), 3.0, (3.5), (4.0), 4.8, 6.2, (8.5), (9.0), 10.5. The remaining seven participants were all censored at 12 months due to termination of the study.

Table 15.2 presents the survival experience for this example as a lifetable. Each row in the lifetable indicates the time at which a death or an event occurred. One or more deaths may have occurred at the same time and they are included in the same

Table 15.2 Kaplan-Meier lifetable for 20 participants followed for 1 year

Interval	Interval number	Time of death	n_j	δ_j	l_j	\hat{p}_j	$S(t)$	$Var \hat{S}(t)$
[0.5, 1, 5)	1	0.5	20	1	1	0.95	0.95	0.0024
[1.5, 3.0)	2	1.5	18	2	1	0.89	0.85	0.0068
[3.0, 4.8)	3	3.0	15	1	2	0.93	0.79	0.0089
[4.8, 6.2)	4	4.8	12	1	0	0.92	0.72	0.0114
[6.2, 10.5)	5	6.2	11	1	2	0.91	0.66	0.0133
[10.5, ∞)	6	10.5	8	1	7 ^a	0.88	0.58	0.0161

n_j : number of participants alive at the beginning of the j^{th} interval

δ_j : number of participants who died during the j^{th} interval

l_j : number of participants who were lost or censored during the j^{th} interval

\hat{p}_j : estimate for p_j , the probability of surviving the j^{th} interval given that the participant has survived the previous intervals

$S(t)$: estimated survival curve

$Var \hat{S}(t)$: variance of $\hat{S}(t)$

^aCensored due to termination of study

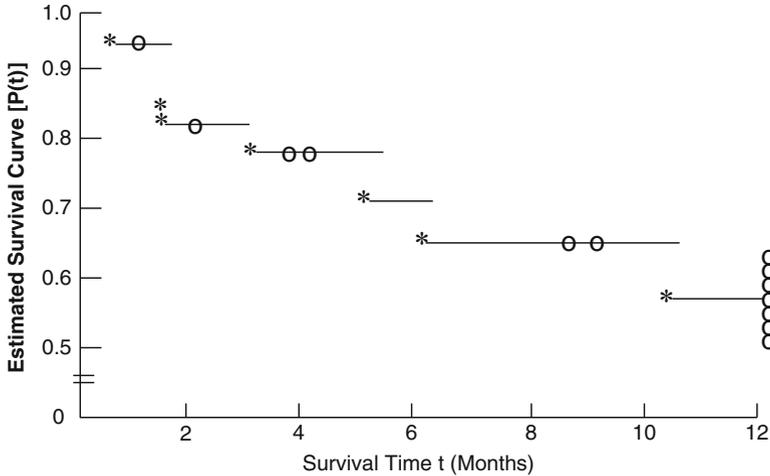


Fig. 15.6 Kaplan-Meier estimate of a survival curve, $\hat{S}(t)$, from a 1-year study of 20 participants, with observed events (asterisk) and censoring (open circle).

row in the lifetable. In the interval between two consecutive times of death, losses to follow-up may have occurred. Hence, a row in the table actually represents an interval of time, beginning with the time of a death, up to but not including the time of the next death. In this case, the first interval is defined by the death at 0.5 months up to the time of the next death at 1.5 months. The columns labeled n_j , δ_j , and l_j correspond to the definitions given above and contain the information from the example. In the first interval, all 20 participants were initially at risk, one died at 0.5 months, and later in the interval (at 0.6 months) one participant was lost to follow-up. In the second interval, from 1.5 months up to 3.0 months, 18 participants were still at risk initially, two deaths were recorded at 1.5 months and one participant was lost at 2.0 months. The remaining intervals are defined similarly. The column labeled \hat{p}_j is the conditional probability of surviving the interval j and is computed as $(n_j - \delta_j)/n_j$ or $(20 - 1)/20 = 0.95$, $(18 - 2)/18 = 0.89$, etc. The column labeled $\hat{S}(t)$ is the estimated survival curve and is computed as the accumulated product of the \hat{p}_j ($0.85 = 0.95 \times 0.89$, $0.79 = 0.95 \times 0.89 \times 0.93$, etc).

The graphical display of the next to last column of Table 15.2, $\hat{S}(t)$, is given in Fig. 15.6. The step function appearance of the graph is because the estimate of $S(t)$, $\hat{S}(t)$ is constant during an interval and changes only at the time of a death. With very large sample sizes and more observed deaths, the step function has smaller steps and looks more like the usually visualized smooth survival curve. If no censoring occurs, this method simplifies to the number of survivors divided by the total number of participants who entered the trial.

Because $\hat{S}(t)$ is an estimate of $S(t)$, the true survival curve, the estimate will have some variation due to the sample selected. Greenwood [19] derived a formula for

estimating the variance of an estimated survival function which is applicable to the Kaplan-Meier method. The formula for the variance of $\hat{S}(t)$, denoted $V[\hat{S}(t)]$ is given by

$$V[\hat{S}(t)] = \hat{S}^2(t) \sum_{j=1}^K \frac{\delta_j}{n_j(n_j - \delta_j)}$$

where n_j and δ_j are defined as before, and K is the number of intervals. In Table 15.2, the last column labeled $V[\hat{S}(t)]$ represents the estimated variances for the estimates of $S(t)$ during the six intervals. Note that the variance increases as one moves down the column. When fewer participants are at risk, the ability to estimate the survival experience is diminished.

Other examples of this procedure, as well as a more detailed discussion of some of the statistical properties of this estimate, are provided by Kaplan and Meier [13]. Computer programs are available [20–23] so that survival curves can be obtained quickly, even for very large sets of data.

The Kaplan-Meier curve can also be used to estimate the hazard rate, λ , if the survival curve is exponential. For example, if the median survival time is estimated as T_M , then $0.5 = S(T_M) = \exp(-\lambda T_M)$ and thus $\hat{\lambda} = \ln(0.5)/T_M$ as an estimate of λ . Then the estimate for $S(t)$ would be $\exp(-\hat{\lambda}t)$. In comparison to the Kaplan-Meier, another parametric estimate for $S(t)$ at time t_j , described by Nelson [24], is

$$\hat{S}(t_j) = \exp\left\{-\sum_{i=1}^j \delta_i/n_j\right\}$$

where δ_i is the number of events in the i^{th} interval and n_j is the number at risk for the event. While this is a straightforward estimate, the Kaplan-Meier does not assume an underlying exponential distribution and thus is used more than this type of estimator.

Comparison of Two Survival Curves

We have just discussed how to estimate the survival curve in a clinical trial for a single group. For two groups, the survival curve would be estimated for each group separately. The question is whether the two survival curves $S_C(t)$ and $S_I(t)$, for the control and intervention groups respectively, are different based on the estimates $\hat{S}_C(t)$ and $\hat{S}_I(t)$.

Point-by-Point Comparison

One possible comparison between groups is to specify a time t^* for which survival estimates have been computed using the Kaplan-Meier [13] method. At time t^* , one can compare the survival estimates $\hat{S}_C(t^*)$ and $\hat{S}_I(t^*)$ using the statistic

$$Z(t^*) = \frac{\hat{S}_C(t^*) - \hat{S}_I(t^*)}{\{V[\hat{S}_C(t^*)] + V[\hat{S}_I(t^*)]\}^{1/2}}$$

where $V[\hat{S}_C(t^*)]$ and $V[\hat{S}_I(t^*)]$ are the Greenwood estimates of variance [19]. The statistic $Z(t^*)$ has approximately a normal distribution with mean zero and variance one under the null hypothesis that $\hat{S}_C(t^*) = \hat{S}_I(t^*)$. The problem with this approach is the multiple looks issue described in Chap. 16. Another problem exists in interpretation. For example, what conclusions should be drawn if two survival curves are judged significantly different at time t^* but not at any other points? The issue then becomes, what point in the survival curve is most important.

For some studies with a T year follow-up, the T year mortality rates are considered important and should be tested in the manner just suggested. Annual rates might also be considered important and, therefore, compared. One criticism of this suggestion is that the specific points may have been selected *post hoc* to yield the largest difference based on the observed data. One can easily visualize two survival curves for which significant differences are found at a few points. However, when survival curves are compared, the large differences indicated by these few points are not supported by the overall survival experience. Therefore, point-by-point comparisons are not recommended unless a few points can be justified and specified in the protocol prior to data analysis.

Comparison of Median Survival Times

One summary measure of survival experience is the time at which 50% of the cohort has had the event. One common and easy way to estimate the median survival time is from the Kaplan-Meier curve. (See for example, Altman [1].) This assumes that the cohort has been followed long enough so that over one-half of the individuals have had the event. Confidence intervals may be computed for the median survival times [25]. If this is the case, we can compare the median survival times for intervention and control M_I and M_C , respectively. This is most easily done by estimating the ratio of the estimates M_I/M_C . A ratio larger than unity implies that the intervention group has a longer median survival and thus a better survival experience. A ratio less than unity would indicate the opposite.

We can estimate 95% confidence intervals for M_I/M_C by

$$(M_I/M_C)e^{-1.96S}, (M_I/M_C)e^{+1.96S}$$

where the standard deviation, SD, of M_I/M_C is computed as

$$SD = \sqrt{1/(O_I + O_C)}$$

for cases where the survival curves are approximately exponential, and O_I = the total number of events in the intervention group (i.e., $\sum \delta_i$) and O_C = the total number of events in the control group.

Total Curve Comparison

Because of the limitations of comparison of point-by-point estimates, Gehan [26] and Mantel [27] originally proposed statistical methods to assess the overall survival experience. These two methods were important steps in the development of analytical methods for survival data. They both assume that the hypothesis being tested is whether two survival curves are equal, or whether one is consistently different from the other. If the two survival curves cross, these methods should be interpreted cautiously. Since these two original methods, an enormous literature has developed on comparison of survival curves and is summarized in several texts [8–11]. The basic methods described here provide the fundamental concepts used in survival analysis.

Mantel [27] proposed the use of the procedure described by Cochran [28] and Mantel and Haenszel [29] for combining a series of 2×2 tables. In this procedure, each time, t_j , a death occurs in either group, a 2×2 table is formed as follows:

	Death at time t_j	Survivors at time t_j	At risk prior to time t_j
Intervention	a_j	b_j	$a_j + b_j$
Control	c_j	d_j	$c_j + d_j$
	$a_j + c_j$	$b_j + d_j$	n_j

The entry a_j represents the observed number of deaths at time t_j in the intervention group and c_j represents the observed number of deaths at time t , in the control group. At least a_j or c_j must be non-zero. One could create a table at other time periods (that is, when a_j and c_j are zero), but this table would not make any contribution to the statistic. Of the n_j participants at risk just prior to time t_j , $a_j + b_j$ were in the intervention group and $c_j + d_j$ were in the control group. The expected number of deaths in the intervention group, denoted $E(a_j)$, can be shown to be

$$E(a_j) = (a_j + c_j)(a_j + b_j)/n_j$$

and the variance of the observed number of deaths in the intervention group, denoted as $V(a_j)$ is given by

$$V(a_j) = \frac{(a_j + c_j)(b_j + d_j)(a_j + b_j)(c_j + d_j)}{n_j^2(n_j - 1)}$$

These expressions are the same as those given for combining 2×2 tables in the Appendix of Chap. 17. The Mantel-Haenszel (*MH*) statistic is given by

$$MH = \left\{ \sum_{j=1}^K a_j - E(a_j) \right\}^2 / \sum_{j=1}^K V(a_j)$$

and has approximately a chi-square distribution with one degree of freedom, where K is the number of distinct event times in the combined intervention and control groups. As an asymptotic approximation,

$$Z_{MH} = \left\{ \sum_{j=1}^K a_j - E(a_j) \right\} / \sqrt{\sum_{j=1}^K V(a_j)}$$

the (signed) square root of *MH*, can be compared to a standard normal distribution [30, 31].

Application of this procedure is straightforward. First, the times of events and losses in both groups are ranked in ascending order. Second, the time of each event, and the total number of participants in each group who were at risk just before the death ($a_j + b_j$, $c_j + d_j$) as well as the number of events in each group (a_j , c_j) are determined. With this information, the appropriate 2×2 tables can be formed.

Example Assume that the data in the example shown in Table 15.2 represent the data from the control group. Among the 20 participants in the intervention group, two deaths occurred at 1.0 and 4.5 months with losses at 1.6, 2.4, 4.2, 5.8, 7.0, and 11.0 months. The observations, with parentheses indicating losses, can be summarized as follows:

Intervention: 1.0, (1.6), (2.4), (4.2), 4.5, (5.8), (7.0), (11.0)

Control: 0.5, (0.6), 1.5, 1.5, (2.0), 3.0, (3.5), (4.0), 4.8, 6.2, (8.5), (9.0), 10.5.

Using the data described above, with remaining observations being censored at 12 months, Table 15.3 shows the eight distinct times of death, (t_j), the number in each group at risk prior to the death, ($a_j + b_j$, $c_j + d_j$), the number of deaths at time t_j , (a_j , c_j), and the number of participants lost to follow-up in the subsequent interval (l_j). The entries in this table are similar to those given for the Kaplan-Meier lifetable shown in Table 15.2. Note in Table 15.3, however, that the observations from two groups have been combined with the net result being more intervals. The entries in Table 15.3 labeled $a_j + b_j$, $c_j + d_j$, $a_j + c_j$, and $b_j + d_j$ become the entries in the eight 2×2 tables shown in Table 15.4.

Table 15.3 Comparison of survival data for a control group and an intervention group using the Mantel-Haenszel procedures

Rank <i>j</i>	Event times <i>t_j</i>	Intervention			Control			Total	
		<i>a_j + b_j</i>	<i>a_j</i>	<i>l_j</i>	<i>c_j + d_j</i>	<i>c_j</i>	<i>l_j</i>	<i>a_j + c_j</i>	<i>b_j + d_j</i>
1	0.5	20	0	0	20	1	1	1	39
2	1.0	20	1	0	18	0	0	1	37
3	4.5	19	0	2	18	1	1	2	35
4	3.0	14	0	1	15	2	2	1	31
5	4.5	16	1	0	12	0	0	1	27
6	4.8	15	0	1	12	0	0	1	26
7	6.2	14	0	1	11	2	2	1	24
8	10.5	13	0	13	8	7	7	1	20

$a_j + b_j$ = number of participants at risk in the intervention group prior to the death at time t_j
 $c_j + d_j$ = number of participants at risk in the control group prior to the death at time t_j
 a_j = number of participants in the intervention group who died at time t_j
 c_j = number of participants in the control group who died at time t_j
 l_j = number of participants who were lost or censored between time t_j and t_{j+1}
 $a_j + c_j$ = number of participants in both groups who died at time t_j
 $b_j + d_j$ = number of participants in both group who are alive minus the number who died at time t_j

Table 15.4 Eight 2×2 tables corresponding to the event times used in the Mantel-Haenszel statistic in survival comparison of intervention (*I*) and control (*C*) groups

1. (0.5 mo) ^a	D^\dagger	A^\ddagger	R^\S	5. (4.5 mo)	<i>D</i>	<i>A</i>	<i>R</i>
<i>I</i>	0	20	20	<i>I</i>	1	15	16
<i>C</i>	1	19	20	<i>C</i>	0	12	12
	1	39	40		1	27	28
2. (1 mo)	<i>D</i>	<i>A</i>	<i>R</i>	6. (4.8 mo)	<i>D</i>	<i>A</i>	<i>R</i>
<i>I</i>	1	19	20	<i>I</i>	0	15	15
<i>C</i>	0	18	18	<i>C</i>	1	11	12
	1	37	38		1	26	27
3. (1.5 mo)	<i>D</i>	<i>A</i>	<i>R</i>	7. (6.2 mo)	<i>D</i>	<i>A</i>	<i>R</i>
<i>I</i>	0	19	19	<i>I</i>	0	14	14
<i>C</i>	2	16	18	<i>C</i>	1	10	11
	2	35	37		1	24	25
4. (3 mo)	<i>D</i>	<i>A</i>	<i>R</i>	8. (10.5 mo)	<i>D</i>	<i>A</i>	<i>R</i>
<i>I</i>	0	17	17	<i>I</i>	0	13	13
<i>C</i>	1	14	15	<i>C</i>	1	7	8
	1	31	32		1	20	21

^aNumber in parenthesis indicates time, t_j , of a death in either group
 D^\dagger = number of participant who died at time t_j
 A^\ddagger = number of participants who are alive between time t_j and time t_{j+1}
 R^\S = number of participants who were at risk before death at time t_j ($R = D + A$)

The Mantel-Haenszel statistic can be computed from these eight 2×2 tables (Table 15.4) or directly from Table 15.3. The term $\sum_{j=1}^8 a_j = 2$ since there are only two deaths in the intervention group. Evaluation of the term $\sum_{j=1}^8 E(a_j) = 20/40 + 20/38 + 2 \times 19/37 + 17/32 + 16/28 + 15/27 + 14/25 + 13/21$ or $\sum_{j=1}^8 E(a_j) = 4.89$. The value for $\sum_{j=1}^8 V(a_j)$ is computed as

$$\sum_{j=1}^8 V(a_j) = \frac{(1)(39)(20)(20)}{(40)^2(39)} + \frac{(1)(37)(20)(18)}{(38)^2(37)} + \dots$$

This term is equal to 2.21. The computed statistic is $MH = (2 - 4.89)^2 / 2.21 = 3.78$. This is not significant at the 0.05 significance level for a chi-square statistic with one degree of freedom. The *MH* statistic can also be used when the precise time of death is unknown. If death is known to have occurred within an interval, 2×2 tables can be created for each interval and the method applied. For small samples, a continuity correction is sometimes used. The modified numerator is

$$\left\{ \left| \sum_{j=1}^K [a_j - E(a_j)] \right| - 0.5 \right\}^2$$

where the vertical bars denote the absolute value. For the example, applying the continuity correction reduces the *MH* statistic from 3.76 to 2.59.

Gehan [26] developed another procedure for comparing the survival experience of two groups of participants by generalizing the Wilcoxon rank statistic. The Gehan statistic is based on the ranks of the observed survival times. The null hypothesis, $S_I(t) = S_C(t)$, is tested. The procedure, as originally developed, involved a complicated calculation to obtain the variance of the test statistic. Mantel [32] proposed a simpler version of the variance calculation, which is most often used.

The N_I observations from the intervention group and the N_C observations from the control group must be combined into a sequence of $N_C + N_I$ observations and ranked in ascending order. Each observation is compared to the remaining $N_C + N_I - 1$ observation and given a score U_i which is defined as follows:

$$U_i = (\text{number of observations ranked definitely less than the } i^{\text{th}} \text{ observation}) - (\text{number of observations ranked definitely greater than the } i^{\text{th}} \text{ observation.})$$

The survival outcome for the i^{th} participant will certainly be larger than that for participants who died earlier. For censored participants, it cannot be determined whether survival time would have been less or greater than the i^{th} observation. This is true whether the i^{th} observation is a death or a loss. Thus, the first part of the score U_i assesses how many deaths definitely preceded the i^{th} observation. The second part of the U_i score considers whether the current, i^{th} , observation is a death or a loss.

Table 15.5 Example of Gehan statistics scores U_i for intervention (I) and control (C) groups

Observation I	Ranked observed time	Group	Definitely less	Definitely more	U_i
1	0.5	C	0	39	-39
2	(0.6) ^a	C	1	0	1
3	1.0	I	1	37	-36
4	1.5	C	2	35	-33
5	1.5	C	2	35	-33
6	(1.6)	I	4	0	4
7	(2.0)	C	4	0	4
8	(2.4)	I	4	0	4
9	3.0	C	4	31	-27
10	(3.5)	C	5	0	5
11	(4.0)	C	5	0	5
12	(4.2)	I	5	0	5
13	4.5	I	5	27	-22
14	4.8	C	6	26	-20
15	(5.8)	I	7	0	7
16	6.2	C	7	24	-17
17	(7.0)	I	8	0	8
18	(8.5)	C	8	0	8
19	(9.0)	C	8	0	8
20	10.5	C	8	20	-12
21	(11.0)	I	9	0	9
22-40	(12.0)	12I,7C	9	0	9

^aParentheses indicate censored observations

If it is a death, it definitely precedes all later ranked observations regardless of whether the observations correspond to a death or a loss. If the i^{th} observation is a loss, it cannot be determined whether the actual survival time will be less than or greater than any succeeding ranked observation, since there was no opportunity to observe the i^{th} participant completely.

Table 15.5 ranks the 40 combined observations ($N_C=20$, $N_I=20$) from the example used in the discussion of the Mantel-Haenszel statistic. The last 19 observations were all censored at 12 months of follow-up, 7 in the control group and 12 in the intervention group. The score U_1 is equal to the zero observations that were definitely less than 0.5 months, minus the 39 observations that were definitely greater than 0.5 months, or $U_1 = -39$. The score U_2 is equal to the one observation definitely less than the loss at 0.6 months, minus none of the observations that will be definitely greater, since at 0.6 months the observation was a loss, or $U_2 = 1$. U_3 is equal to the one observation (0.5 months) definitely less than 1.0 month minus the 37 observations definitely greater than 1.0 month giving $U_3 = 36$. The last

19 observations will have scores of 9 reflecting the nine deaths which definitely precede censored observations at 12.0 months.

The Gehan statistic, G , involves the scores U_i and is defined as

$$G = W^2/V(W)$$

where $W = \sum U_i$, (for U_i 's in control group only) and

$$V(W) = \frac{N_C N_I}{(N_C + N_I)(N_C + N_I - 1)} \sum_{i=1}^{N_C + N_I} (U_i^2)$$

The G statistic has approximately a chi-square distribution with one degree of freedom [26, 32]. Therefore, the critical value is 3.84 at the 5% significance level and 6.63 at the 1% level. In the example, $W = -87$ and the variance $V(W) = 2,314.35$. Thus, $G = (-87)^2/2,314.35 = 3.27$ for which the p -value is equal to 0.071. This is compared with the p -value of 0.052 obtained using the Mantel-Haenszel statistic.

The Gehan statistic assumes the censoring pattern to be equal in the two groups. Breslow [33] considered the case in which censoring patterns are not equal and used the same statistic G with a modified variance. This modified version should be used if the censoring patterns are radically different in the two groups. Peto and Peto [34] also proposed a version of a censored Wilcoxon test. The concepts are similar to what has been described for Gehan's approach. However, most software packages now use the Breslow or Peto and Peto versions.

Generalizations

The general methodology of comparing two survival curves using this methodology has been further evaluated [35–40]. These two tests by Mantel-Haenzel and Gehan, can be viewed as a weighted sum of the difference between observed number of events and the expected number at each unique event time [7, 40]. Consider the previous equation for the logrank test and rewrite the numerator as

$$W = \sum_{j=1}^K w_j [a_j - E(a_j)]$$

where

$$V(W) = \sum_{j=1}^K w_j^2 \frac{(a_j + c_j)(b_j + d_j)(a_j + b_j)(c_j + d_j)}{n_j^2(n_j - 1)}$$

and w_j is a weighting factor. The test statistic $W^2/V(W)$ has approximately a chi-square distribution with one degree of freedom or equivalently $W/\sqrt{V(W)}$ has approximately a standard normal distribution. If $w_i = 1$, we obtain the Mantel-Haenszel or logrank test. If $w_i = n_j/(N + 1)$, where $N = N_C + N_I$ or the combined sample size, we obtain the Gehan version of the Wilcoxon test. Tarone and Ware [40] pointed out that the Mantel-Haenszel and Gehan are only two possible statistical tests. They suggested a general weight function $w_i = [n_j/(N + 1)]^\theta$ where $0 < \theta < 1$. In particular, they suggested that $\theta = 0.5$. Prentice [38] suggested a weight $w_j = \prod_{i=1}^j n_i / (n_i + d_i)$ where $d_i = (a_i + c_i)$ which is related to the product limit estimator at t_j as suggested by Peto and Peto [34]. Harrington and Fleming [35] generalize this further by suggesting weights $w_j = \left\{ \prod_{i=1}^j n_i / (n_i + d_i) \right\}^\rho$ for $\rho > 0$.

All of these methods give different weights to the various parts of the survival curve. The Mantel-Haenszel or logrank statistic is more powerful for survival distributions of the exponential form where $\lambda_I(t) = \theta \lambda_C(t)$ or $S_I(t) = \{S_C(t)\}^\theta$ where $\theta \neq 1$ [32]. The Gehan type statistic [26], on the other hand, is more powerful for survival distributions of the logistic form $S(t, \theta) = e^{t+\theta} / (1 + e^{t+\theta})$. In actual practice, however, the distribution of the survival curve of the study population is not known. When the null hypothesis is not true, the Gehan type statistic gives more weight to the early survival experience, whereas the Mantel-Haenszel weights the later experience more. Tarone and Ware [40] indicate other possible weighting schemes could be proposed which are intermediate to these two statistics [35, 40]. Thus, when survival analysis is done, it is certainly possible to obtain different results using different weighting schemes depending on where the survival curves separate, if they indeed do so. The logrank test is the standard in many fields such as cancer and heart disease. The condition $\lambda_I(t) = \theta \lambda_C(t)$ says that risk of the event being studied in the intervention is a constant multiple of the hazard $\lambda_C(t)$. That is, the hazard rate in one arm is proportional to the other and so the logrank test is best for testing proportional hazards. This idea is appealing and is approximately true for many studies.

There has been considerable interest in asymptotic (large sample) properties of rank tests [37, 39] as well as comparisons of the various analytic methods [36]. While there exists an enormous literature on survival analysis, the basic concepts of rank tests can still be appreciated by the methods described above.

Earlier, we discussed using an exponential model to summarize a survival curve where the hazard rate λ determines the survival curve. If we can assume that the hazard rate is reasonably constant during the period of follow-up for the intervention and the control group, then comparison of hazard rates is a comparison of survival curves [1]. The most commonly used comparison is the ratio of the hazards, $R = \lambda_I / \lambda_C$. If the ratio is unity, the survival curves are identical. If R is greater than one, the intervention hazard is greater than control so the intervention survival curve falls below the standard curve. That is, the intervention is worse. On the other hand, if R is less than one, the control group hazard is larger, the control group survival curve falls below the intervention curve, and intervention is better.

We can estimate the hazard ratio by comparing the ratio of total observed events (O) divided by expected number of events (E) in each group; that is, the estimate of R can be expressed as

$$\hat{R} = \frac{O_I/E_I}{O_C/E_C}$$

That is, $O_I = \sum a_i$, $O_C = \sum b_i$, $E_I = \sum E(a_i)$, and $E_C = \sum E(b_i)$. Confidence intervals for the odds ratio R are most easily determined by constructing confidence intervals for the log of the odds ratio $\ln R$ [41]. The 95% confidence interval for $\ln R$ is $K - 1.96/\sqrt{V}$ to $K + 1.96/\sqrt{V}$ where $K = (O_I - E_I)/V$ and V is the variance as defined in the logrank or Mantel-Haenszel statistics. (That is, V equals $V(a_i)$.) We then connect confidence intervals for $\ln R$ to confidence intervals for R by taking antilogs of the upper and lower limit. If the confidence interval excludes unity, we could claim superiority of either intervention or control depending on the direction. Hazard ratios not included in the interval can be excluded as likely outcome summaries of the intervention. If the survival curves have relatively constant hazard rates, this method provides a nice summary and complements the Kaplan-Meier estimates of the survival curves.

Covariate Adjusted Analysis

Previous chapters have discussed the rationale for taking stratification into account. If differences in important covariates or prognostic variables exist at entry between the intervention and control groups, an investigator might be concerned that the analysis of the survival experience is influenced by that difference. In order to adjust for these differences in prognostic variables, she could conduct a stratified analysis or a covariance type of survival analysis. If these differences are not important in the analysis, the adjusted analysis will give approximately the same results as the unadjusted.

Three basic techniques for stratified survival analysis are of interest. The first compares the survival experience between the study groups within each stratum, using the methods described in the previous section. By comparing the results from each stratum, the investigator can get some indication of the consistency of results across strata and the possible interaction between strata and intervention.

The second and third methods are basically adaptations of the Mantel-Haenszel and Gehan statistics, respectively, and allow the results to be accumulated over the strata. The Mantel-Haenszel stratified analysis involves dividing the population into S strata and within each stratum j , forming a series of 2×2 tables for each K_j event, where K_j is the number of events in stratum j . The table for the i^{th} event in the j^{th} stratum would be as follows:

	Event	Alive	
Intervention	a_{ij}	b_{ij}	$a_{ij} + b_{ij}$
Control	c_{ij}	d_{ij}	$c_{ij} + d_{ij}$
	$a_{ij} + c_{ij}$	$b_{ij} + d_{ij}$	n_{ij}

The entries a_{ij} , b_{ij} , c_{ij} , and d_{ij} are defined as before and

$$E(a_{ij}) = (a_{ij} + c_{ij})(a_{ij} + b_{ij})/n_{ij}$$

$$V(a_{ij}) = \frac{(a_{ij} + c_{ij})(b_{ij} + d_{ij})(a_{ij} + b_{ij})(c_{ij} + d_{ij})}{n_{ij}^2(n_{ij} - 1)}$$

Similar to the non-stratified case, the Mantel-Haenszel statistic is

$$MH = \left\{ \sum_{j=1}^S \sum_{i=1}^{K_j} a_{ij} - E(a_{ij}) \right\}^2 / \sum_{j=1}^S \sum_{i=1}^{K_j} V(a_{ij})$$

which has a chi-square distribution with one degree of freedom. Analogous to the Mantel-Haenszel statistic for stratified analysis, one could compute a Gehan statistic W_j and $V(W_j)$ within each stratum. Then an overall stratified Gehan statistic is computed as

$$G = \left\{ \sum_{j=1}^S W_j \right\}^2 / \sum_{j=1}^S V(W_j)$$

which also has chi-square statistic with one degree of freedom.

If there are many covariates, each with several levels, the number of strata can quickly become large, with few participants in each. Moreover, if a covariate is continuous, it must be divided into intervals and each interval assigned a score or rank before it can be used in a stratified analysis. Cox [42] proposed a regression model which allows for analysis of censored survival data adjusting for continuous as well as discrete covariates, thus avoiding these two problems.

One way to understand the Cox regression model is to again consider a simpler parametric model. If one expresses the probability of survival to time t , denoted $S(t)$, as an exponential model, then $S(t) = e^{-\lambda t}$ where the parameter, λ , is called the force of mortality or the hazard rate as described earlier. The larger the value of λ , the faster the survival curve decreases. Some models allow the hazard rate to change with time, that is $\lambda = \lambda(t)$. Models have been proposed [43–45] which attempt to incorporate the hazard rate as a linear function of several baseline covariates, x_1, x_2, \dots, x_p that is, $\lambda(x_1, x_2, \dots, x_p) = b_1x_1 + b_2x_2 + \dots + b_px_p$. One of the covariates, say x_1 , might represent the intervention and the others, for example, might represent age, sex, performance status, or prior medical history. The coefficient, b_1 , then would indicate whether intervention is a significant prognostic factor,

i.e., remains effective after adjustment for the other factors. Cox [42] suggested that the hazard rate could be modeled as a function of both time and covariates, denoted $\lambda(t, x_1, x_2, \dots, x_p)$. Moreover, this hazard rate could be represented as the product of two terms, the first representing an unadjusted force of mortality $\lambda_0(t)$ and the second the adjustment for the linear combination of a particular covariate profile. More specifically, the Cox proportional hazard model assumes that

$$\lambda(t, x_1, x_2, \dots, x_p) = \lambda_0(t) \exp(b_1x_1 + b_2x_2 + \dots + b_px_p)$$

That is, the hazard $\lambda(t, x_1, x_2, \dots, x_n)$ is proportional to an underlying hazard function $\lambda_0(t)$ by the specific factor $\exp(b_1x_1 + b_2x_2 \dots)$. From this model, we can estimate an underlying survival curve $S_0(t)$ as a function of $\lambda_0(t)$. The survival curve for participants with a particular set of covariates X , $S(t, x)$ can be obtained as $S(t, x) = [S_0(t)]^{\exp(b_1x_1 + b_2x_2 + \dots)}$. Other summary test statistics from this model are also used. The estimation of the regression coefficients b_1, b_2, \dots, b_p is complex, requiring non-linear numerical methods, and goes beyond the scope of this text. Many elementary texts on biostatistics [1, 3, 5, 46] or review articles [7] present further details. A more advanced discussion may be found in Kalbfleish and Prentice [10] or Fleming and Harrington [9]. Programs exist in many statistical computing packages which provide these estimates and summary statistics to evaluate survival curve comparisons [20–23]. Despite the complexity of the parameter estimation, this method is widely applied and has been studied extensively [47–55]. Pocock, Gore, and Kerr [52] demonstrate the value of some of these methods with cancer data. For the special case where group assignment is the only covariate, the Cox model is essentially equivalent to the Mantel-Haenszel statistic.

One issue that is sometimes raised is whether the hazard rates are proportional over time. Methods such as the Mantel-Haenszel logrank test or the Cox Proportional Hazards model are optimal when the hazards are proportional [9]. However, though there is some loss of power, these methods perform well as long as the hazard curves do not cross, even if proportionality does not hold [56]. When the hazards are not proportional, which intervention is better depends on what time point is being referenced. If a significant difference is found between two survival curves using the Mantel-Haenszel logrank test or the Cox Proportional Hazards model when the hazards are not proportional, the two curves are still significantly different. For example, time to event curves are shown in Chap. 18. Figure 18.2a shows three curves for comparison of two medical devices with best medical or pharmacologic care. These three curves do not have proportional hazards but the comparisons are still valid and in fact the two devices demonstrate statistically significant superiority over the best medical care arm. The survival curves do not cross although are close together in the early months of follow-up.

The techniques described in this chapter as well as the extensions or generalizations referenced are powerful tools in the analysis of survival data. Perhaps none is exactly correct for any given set of data but experience indicates they are fairly robust and quite useful.

References

1. Altman DG. Practical Statistics for Medical Research. Chapman and Hall, 2015.
2. Armitage P, Berry G, Mathews J. Statistical Methods in Medical Research, ed 4th. Malden MA, Blackwell Publishing, 2002.
3. Breslow NE. Comparison of survival curves; in Buyse B, Staquet M, Sylvester R (eds): The Practice of Clinical Trials in Cancer. Oxford, Oxford University Press, 1982.
4. Brown BW, Hollander M. Statistics: A Biomedical Introduction. Wiley, 2009.
5. Fisher L, Van Belle G, Heagerty PL, Lumley TS. Biostatistics—A Methodology for the Health Sciences. New York, John Wiley and Sons, 2004.
6. Woolson RF, Clarke WR. Statistical Methods for the Analysis of Biomedical Data. Wiley, 2011.
7. Crowley J, Breslow N. Statistical Analysis of Survival Data. *Annu Rev Public Health* 1984;5:385–411.
8. Cox DR, Oakes D. Analysis of Survival Data. Taylor & Francis, 1984.
9. Fleming TR, Harrington DP. Counting Processes and Survival Analysis. Wiley, 2011.
10. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. Wiley, 2011.
11. Miller RG. Survival Analysis. Wiley, 2011.
12. Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 1958;8:699–712.
13. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* 1958;53:457–481.
14. Chan MCY, Giannetti N, Kato T, et al. Severe tricuspid regurgitation after heart transplantation. *J Heart Lung Transplant* 2001;20:709–717.
15. Kumagai R, Kubokura M, Sano A, et al. Clinical evaluation of percutaneous endoscopic gastrostomy tube feeding in Japanese patients with dementia. *Psychiatry Clin Neurosci* 2012;66:418–422.
16. Lara-Gonzalez JH, Gomez-Flores R, Tamez-Guerra P, et al. In Vivo Antitumor Activity of Metal Silver and Silver Nanoparticles in the L5178Y-R Murine Lymphoma Model. *Br J Med Med Res* 2013;3:1308–1316.
17. Miyamoto K, Aida A, Nishimura M, et al. Gender effect on prognosis of patients receiving long-term home oxygen therapy. The Respiratory Failure Research Group in Japan. *Am J Respir Crit Care Med* 1995;152:972–976.
18. Sarris GE, Robbins RC, Miller DC, et al. Randomized, prospective assessment of bioprosthetic valve durability. Hancock versus Carpentier-Edwards valves. *Circulation* 1993;88:II55-II64.
19. Greenwood M. The natural duration of cancer. Reports on Public Health and Medical Subjects 1926;33:1–26.
20. Everitt BS, Rabe-Hesketh S. Handbook of Statistical Analyses Using Stata, Fourth Edition. Taylor & Francis, 2006.
21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2013.
22. SAS Institute: SAS/STAT 12.1 User's Guide: Survival Analysis. SAS Institute, 2012.
23. TIBCO Software I: SPLUS. TIBCO Software Inc., 2008.
24. Nelson W. Hazard plotting for incomplete failure data. *Journal of Quality Technology* 1969;1:27–52.
25. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. *Biometrics* 1982;38:29–41.
26. Gehan EA. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. *Biometrika* 1965;52:203–223.
27. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer chemotherapy reports Part 1 1966;50:163–170.
28. Cochran WG. Some Methods for Strengthening the Common- χ^2 Tests. *Biometrics* 1954;10:417–451.

29. Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *J Natl Cancer Inst* 1959;22:719–748.
30. Crowley J, Breslow N. Remarks on the Conservatism of $\Sigma(O-E)^2/E$ in Survival Data. *Biometrics* 1975;31:957–961.
31. Peto R, Pike MC. Conservatism of the Approximation $\Sigma(O-E)^2/E$ in the Logrank Test for Survival Data or Tumor Incidence Data. *Biometrics* 1973;29:579–584.
32. Mantel N. Ranking Procedures for Arbitrarily Restricted Observation. *Biometrics* 1967;23:65–78.
33. Breslow NE. A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. *Biometrika* 1970;57:579–594.
34. Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *J R Stat Soc Ser A* 1972;135:185–207.
35. Harrington DP, Fleming TR. A Class of Rank Test Procedures for Censored Survival Data. *Biometrika* 1982;69:553–566.
36. Leurgans S. Three Classes of Censored Data Rank Tests: Strengths and Weaknesses under Censoring. *Biometrika* 1983;70:651–658.
37. Oakes D. The Asymptotic Information in Censored Survival Data. *Biometrika* 1977;64:441–448.
38. Prentice RL. Linear Rank Tests with Right Censored Data. *Biometrika* 1978;65:167–179.
39. Schoenfeld DA. The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. *Biometrika* 1981;68:316–319.
40. Tarone RE, Ware J. On Distribution-Free Tests for Equality of Survival Distributions. *Biometrika* 1977;64:156–160.
41. Simon R. Confidence Intervals for Reporting Results of Clinical Trials. *Ann Intern Med* 1986;105:429–435.
42. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Series B Stat Methodol* 1972;34:187–220.
43. Feigl P, Zelen M. Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics* 1965;21:826–838.
44. Prentice RL, Kalbfleisch JD. Hazard Rate Models with Covariates. *Biometrics* 1979;35:25–39.
45. Zelen M. Application of Exponential Models to Problems in Cancer Research. *J R Stat Soc Ser A* 1966;129:368–398.
46. Fisher L, Van Belle G. Biostatistics—A Methodology for the Health Sciences. New York, John Wiley and Sons, 1993.
47. Breslow NE. Covariance Analysis of Censored Survival Data. *Biometrics* 1974;30:89–99.
48. Breslow NE. Analysis of Survival Data under the Proportional Hazards Model. *International Statistical Review/Revue Internationale de Statistique* 1975;43:45–57.
49. Efron B. The Efficiency of Cox's Likelihood Function for Censored Data. *J Am Stat Assoc* 1977;72:557–565.
50. Kalbfleisch JD, Prentice RL. Marginal Likelihoods Based on Cox's Regression and Life Model. *Biometrika* 1973;60:267–278.
51. Kay R. Proportional Hazard Regression Models and the Analysis of Censored Survival Data. *J R Stat Soc Ser C Appl Stat* 1977;26:227–237.
52. Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 1982;38:153–162.
53. Prentice RL, Gloeckler LA. Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data. *Biometrics* 1978;34:57–67.
54. Schoenfeld DA. Chi-Squared Goodness-of-Fit Tests for the Proportional Hazards Regression Model. *Biometrika* 1980;67:145–153.
55. Tsiatis AA. A Large Sample Study of Cox's Regression Model. *Ann Stat* 1981;9:93–108.
56. Lagakos SW, Schoenfeld DA. Properties of Proportional-Hazards Score Tests under Misspecified Regression Models. *Biometrics* 1984;40:1037–1048.