Chapter 12.
Survival Analysis

Survival analysis is concerned with studying the time between entry to a study and a subsequent event. Originally the analysis was concerned with time from treatment until death, hence the name, but survival analysis is applicable to many areas as well as mortality. Recent examples include time to discontinuation of a contraceptive, maximum dose of bronchoconstrictor required to reduce a patient's lung function to 80% of baseline, time taken to exercise to maximum tolerance, time that a transdermal patch can be left in place, time for a leg fracture to heal.

When the outcome of a study is the time between one event and another, a number of problems can occur.

1. The times are most unlikely to be Normally distributed.
2. We cannot afford to wait until events have happened to all the subjects, for example until all are dead. Some patients might have left the study early - they are *lost to follow up* . Thus the only information we have about some patients is that they were still alive at the last follow up. These are termed *censored observations* .

Kaplan-Meier survival curve

We look at the data using a Kaplan-Meier survival curve[(1)]. Suppose that the survival times, including censored observations, after entry into the study (ordered by increasing duration) of a group of n subjects are $t_1, t_2, \ldots t_n$ The proportion of subjects, S(t), surviving beyond any follow up time (t_p) is estimated by

$$S(t) = \frac{(r_1 - d_1)}{r_1} \times \frac{(r_2 - d_2)}{r_2} \ldots \times \ldots \frac{(r_p - d_p)}{r_p}$$

where t_p is the largest survival time less than or equal to t and r_i is the number of subjects alive just before time t_i (the ith ordered survival time), d_i denotes the number who died at time t_i where i can be any value between 1 and p. For censored observations $d_i = 0$.

Method

Order the survival time by increasing duration starting with the shortest one. At each event (i) work out the number alive immediately before the event (r *i*). Before the first event all the

patients are alive and so S(t) = 1. If we denote the start of the study as t_0, where $t_0 = 0$, then we have $S^{(t_0)} = 1$. We can now calculate the survival times t_i, for each value of i from 1 to n by means of the following recurrence formula.

Given the number of events (deaths), d_i, at time t_i and the number alive, r_i, just before t_i calculate

$$S(t_i) = \frac{r_i - d_i}{r_i} \times S(t_{i-1})$$

We do this only for the events and not for censored observations. The survival curve is unchanged at the time of a censored observation, but at the next event after the censored observation the number of people "at risk" is reduced by the number censored between the two events.

Example of calculation of survival curve

McIllmurray and Turkie[2] describe a clinical trial of 69 patients for the treatment of Dukes' C colorectal cancer. The data for the two treatments, γ linoleic acid or control are given in Table 12.1[3]

Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment

Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

The calculation of the Kaplan-Meier survival curve for the 25 patients randomly assigned to receive 7 linoleic acid is described in Table 12.2 . The + sign indicates censored data. Until 6 months after treatment, there are no deaths, 50 S(t). The effect of the censoring is to remove from the alive group those that are censored. At time 6 months two subjects have been censored and so the number alive just before 6 months is 23. There are two deaths at 6 months.

Thus,

$$S(6)= \frac{1 \times (23-2)}{23} = 0.9130$$

We now reduce the number alive ("at risk") by two. The censored event at 9 months reduces the "at risk" set to 20. At 10 months there are two deaths, so the proportion surviving is 18/20 = 0.90 and the cumulative proportion surviving is 0.913 x 0.90 = 0.8217. The cumulative survival is conveniently stored in the memory of a calculator. As one can see the effect of the censored observations is to reduce the number at risk without affecting the survival curve S(t).

Table 12.2 Calculation of survival case for 25 patients randomly assigned to receive γ linoleic acid

Case (i)	Survival time (months) (t_i)	Number alive (r_i)	Deaths (d_i)	Proportion surviving $\frac{(r_1 - d_1)}{n_1}$	Cumulative proportion surviving S(t)
	0	0	0	-	1
1	1+	25	0	1	1
2	5+	24	0	1	1
3	6	23	2	0.9130	0.9130
4	6				
5	9+	21	0	1	0.9130
6	10	20	2	0.90	0.8217
7	10				
8	10+				
9	12	17	4	0.7647	0.6284
10	12				
11	12				
12	12				
13	12+				
14	13+	12	0	1	0.6284
15	15+	11	0	1	0.6284

16	16+	10	0	1	0.6284
17	20+	9	0	1	0.6284
18	24	8	1	0.875	0.5498
19	24+				
20	27+	6	0	1	0.5498
21	32	5	1	0.80	0.4399
22	34+				
23	36+				
24	36+				
25	44+				

Finally we plot the survival curve, as shown in Figure 12.1 The censored observations are shown as ticks on the line.

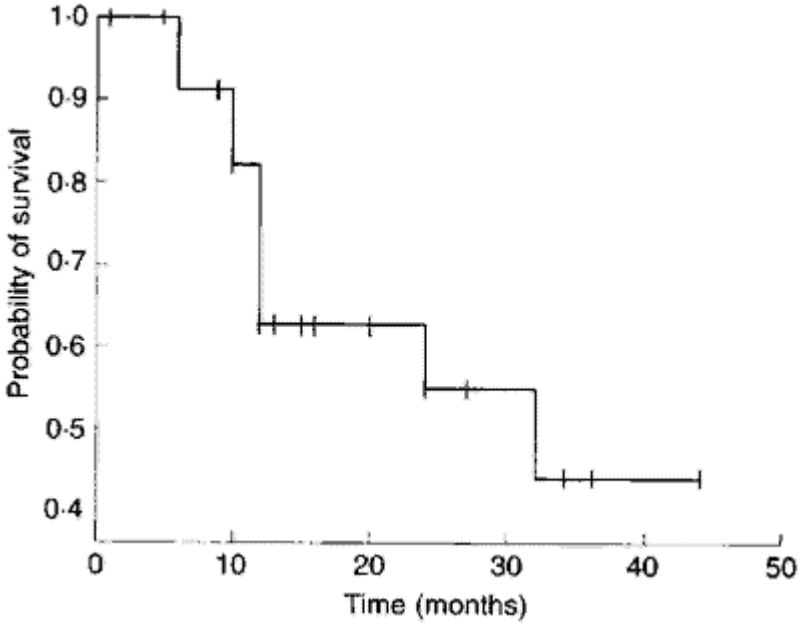

Figure 12.1 Survival curve of 25 patients with Dukes' C colorectal cancer treated with linoleic acid.

Log Rank Test

To compare two survival curves produced from two groups A and B we use the rather curiously named log rank test,1 so called because it can be shown to be related to a test that uses the logarithms of the ranks of the data.

The assumptions used in this test are:

1. That the survival times are ordinal or continuous.
2. That the risk of an event in one group relative to the other does not change with time. Thus if linoleic acid reduces the risk of death in patients with colorectal cancer, then this risk reduction does not change with time (the so called *proportional hazards assumption*).

We first order the data for the two groups combined, as shown in Table 12.3 . As for the Kaplan-Meier survival curve, we now consider each event in turn, starting at time t = 0.

Table 12.3 Calculation of log rank statistics for 49 patients randomly assigned to receive γ linoleic acid (A) or control (B)

Survival time (months) t_i	Group	Total at risk r	Number of events d_i	Total at risk in group A r_{Ai}	Expected number of events E_{Ai}
0		49			
1+	A	49	0	25	0
3+	B	48	0	24	0
5+	A	47	0	24	0
6	A	46	6	23	3.0
6	A				
6	B				
6	B				
6	B				
6	B				
8	B	40	2	21	1.05
8	B				
9+	A	38	0	21	0
10	A	37	2	20	1.0811
10	A				
10+	A				
12	A	34	6	17	3.0
12	A				
12	A				
12	A				
12	B				

12	B				
12+	A				
12+	B				
13+	A	26	0	12	0
15+	A	25	0	11	0
15+	B	24	0	10	0
16+	A	23	0	10	0
16+	B	22	0	9	0
18+	B	21	0	9	0
18+	B				
20	B	19	1	9	0.4736
20+	A				
22+	B	17	0	8	0
24	A	16	2	8	1.0
24	B				
24+	A				
27+	A	13	0	6	0
28+	B	12	0	5	0
28+	B				
28+	B				
30	B	9	1	5	0.5555
30+	B				
32	A	7	1	5	0.7143
33+	B	6	0	4	0
34+	A	5	0	4	0
36+	A	4	0	3	0
36+	A				
42	B	2	1	1	0.50
44+	A				

At each event (death) at time t_i we consider the total number alive r_i and the total number still alive in group A r_{Ai} up to that point. If we had a total of d_i events at time t_i then, under the null hypothesis, we consider what proportion of these would have been expected in group A.

Clearly the more people at risk in one group the more deaths (under the null hypothesis) we would expect.

Thus we obtain

$$E_{Ai} = r_{Ai} / r_i \times d_i$$

The effect of the censored observations is to reduce the numbers at risk, but they do not contribute to the expected numbers.

Finally, we add the total number of expected events in group A, $E_{Ai} = \Sigma E_{Ai}$. If the total number of events in group B is E_B we can deduce E_B from $E_B = n - E_A$. We do not calculate the expected number beyond the last event, in this case at time 42 months. Also, we would stop calculating the expected values if any survival times greater than the point we were at were found in one group only.

Finally, to test the null hypothesis of equal risk in the two groups we compute

$$X^2 = (O_A - E_A)^2 / E_A + (O_B - E_B)^2 / E_B$$

where O_A and O_B are the total number of events in groups A and B. We compare X^2 to a χ^2 distribution with one degree of freedom (one, because we have two groups and one constraint, namely that the total expected events must equal the total observed).

The calculation for the colorectal data is given in Table 12.3. The first non-censored event occurs at 6 months, at which there are six of them. By that time 46 patients are at risk, of whom 23 are in group A. Thus we would expect 6 x 23/46 = 3 to be in group A. At 8 months we have 46 - 6 = 40 patients at risk of whom 23 - 2 = 21 are in group A. There are two events, of which we would expect 2 x 21/40 = 1.05 to occur in group A.

The total expected number of events in A is E_A = 11.3745. The total number of events is 22, O_A = 10, O_B = 12. Thus E_B =10.6255.

Thus

$$X^2 = \frac{(10 - 11.37)^2}{11.37} + \frac{(12 - 10.63)^2}{10.63} = 0.34$$

We compare this with the χ^2 Table given in Appendix E, to find that P>0.10.

The relative risk can be estimated by $r = (O_A/E_A)/(O_B/E_B)$. The standard error of the log risk is given by [4]

$$SE(\log(r)) = \sqrt{(1/E_A + 1/E_B)}$$

Thus we find r = 0.78 and so log(r) = -0.248.

SE(log(r)) = 0.427, and so an approximate 95% confidence interval for log(r) is

-1.10 to 0.605 and so a 95% confidence interval for r is $e^{-1.10}$ to $e^{0.605}$, which is

0.33 to 1.83.

This would imply that linoleic acid reduced mortality by about 78% compared with the control group, but with a very wide confidence interval. In view of the very small χ^2 statistic, we have little evidence that this result would not have arisen by chance.

Further methods

In the same way that multiple regression is an extension of linear regression, an extension of the log rank test includes, for example, allowance for prognostic factors. This was developed by DR Cox, and so is called *Cox regression* . It is beyond the scope of this book, but is described elsewhere[4, 5].

Common questions

Do I need to test for a constant relative risk before doing the log rank test?

This is a similar problem to testing for Normality for a *t* test. The log rank test is quite "robust" against departures from proportional hazards, but care should be taken. If the Kaplan-Meier survival curves cross then this is clear departure from proportional hazards, and the log rank test should not be used. This can happen, for example, in a two drug trial for cancer, if one drug is very toxic initially but produces more long term cures. In this case there is no simple answer to the question "is one drug better than the other?", because the answer depends on the time scale.

If I don't have any censored observations, do I need to use survival analysis?

Not necessarily, you could use a rank test such as the Mann-Whitney U test, but the survival method would yield an estimate of risk, which is often required, and lends itself to a useful way of displaying the data.

References

1. Peto R, Pike MC, Armitage P *et al* . Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples. *Br J Cancer* l977; *35* :l-39.
2. McIllmurray MB, Turkie W. Controlled trial of ϒlinoleic acid in Dukes' C colorectal cancer. *BMJ* 1987; *294* :1260, *295* :475.
3. Gardner MJ, Altman DG (Eds). In: *Statistics with Confidence, Confidence Intervals and Statistical Guidelines* . London: BMJ Publishing Group, 1989; Chapter 7.
4. Armitage P, Berry G. In: *Statistical Methods in Medical Practice* . Oxford: Blackwell Scientific Publications, 1994:477-81.
5. Altman DG. *Practical Statistics for Medical Research* .. London: Chapman & Hall, 1991.

Exercises

Exercise 12.1 Twenty patients, ten of normal weight and ten severely overweight underwent an exercise stress test, in which they had to lift a progressively increasing load for up to 12 minutes, but they were allowed to stop earlier if they could do no more. On two occasions the equipment failed before 12 minutes. The times (in minutes) achieved were:

Normal weight: 4, 10, 12*, 2, 8, 12*, 8**, 6, 9, 12*

Overweight: 7**, 5, 11, 6, 3, 9, 4, 1, 7, 12*

*Reached end of test; **equipment failure. What are the observed and expected values? What is the value of the log rank test to compare these groups?

Exercise 12.2 What is the risk of stopping in the normal weight group compared with the overweight group, and a 95% confidence interval?