Statistics for Surgeons – Understanding Survival Analysis

Girdhar Gopal Agarwal

Received: 27 July 2011 / Accepted: 16 April 2012 / Published online: 6 June 2012
© Indian Association of Surgical Oncology 2012

Abstract Survival data stand out as a special statistical field. This paper tries to describe what survival data is and what makes it so special. Survival data concerns times to some events. A key point is the successive observation of time, which on the one hand leads to sometimes not being observed so that all that is known is that they exceed some given times (censoring), and on the other hand implies that predictions regarding the future course should be conditional on the present status (truncation). In the simplest case, this condition is that the individual is alive. The successive conditioning makes the hazard function, which describes the probability of an event happening during a short interval given that the individual is alive today, the most relevant concept. Here we discuss parametric as well as non-parametric methods. Examples are presented in a way that can be followed without the help of computers.

Keywords Censored data · Survivor function · Hazard function · Kaplan-Meier estimate · Logrank test

Introduction

First, we describe what survival data is in. What makes such data different from other types of data? In other words, why does the subject need a special statistical theory [1, 2]? The answer is given in section on “[Special Features of Survival Data](#)” [3, 4]. We also try to answer some other questions:

What is hazard function and why is it so important [5]? How are two survival functions compared [6]? What is the role of nonparametric methods?

Survival Data

Survival data is a term used for data measuring the time to some event. In the simplest case, the event is death, but the term also covers other events, like disease recurrence (e.g. cancer), occurrence of a disease, a complication, or, e.g., the time to occurrence of an epileptic seizure. The study participants are “at risk” for the event continuously over time from the start of the observation. The duration of observation may vary from one subject to another, e.g., staggered entry into a clinical trial. The observation may be shortened by lost contact or other non-relevant competing risks.

Survival data consider only a single or a few types of states and events, ruling out the possibility of continuous measurements of, e.g., quality of life. To study such a concept, it must be classified into a few well-defined groups. On the other hand, the time to event is considered in great detail in survival analysis. The biostatistical tradition for times for human beings is to count days and then analyze them by rank-based (non-parametric) methods. For example, for studies of cancer, this does not make medical sense because cancers can be developing for up to 3–4 years before being diagnosed [7]. Grouping cancers into years might be acceptable from a medical point of view but is unacceptable as the nonparametric methods might have too many coinciding observations (ties) [1, 2].

Special Features of Survival Data

Here the responses are times and thus are not measured the same way as the other variables. The larger observations

G. G. Agarwal (✉)
University of Lucknow,
Lucknow, Uttar Pradesh, India
e-mail: singhgyanendra147@rediffmail.com

G. G. Agarwal
e-mail: girdhar1751@gmail.com

take longer times to observe than smaller ones. Time is observed sequentially. A consequence of this is: presence of censored data SPSS16.0 [8, 9].

Censored Data

Censoring of data can occur due to loss to follow up, or because the subject withdraws from the study (e.g. because of an adverse drug reaction), or because the subject does not experience the event before the study ends. There are three kinds of censoring- right, left and random (Fig. 1). The first patient developed the event at 5 years. But in the second patient, the event did not occur until the end of study. The third patient entered the study during the second year and withdrew at 6 years. This patient's contribution to the study was 4 years of follow up. The second and third patients' survival time was incomplete at the end of study and were right censored. Left censored data can occur when a subject's survival time is incomplete at the beginning of follow up period. For example, if we were following subjects with HIV infection, we might start the follow up when a subject's first test was positive for the HIV virus, but we may not know the exact time of the first viral infection.

Analysis of Survival Data

Survivor Function

The most obvious and natural way of summarizing survival data is a survivor function (or survival curve) [10] which is the proportion of subjects remaining event-free at each point in time from the start of the observation. This would be

Fig. 1 Type I and type II censoring

extremely simple if all subjects were followed up to some common time, say 5 years. The problem arises when a subject's observation stops at say, 18 months, i.e. it is censored. Censored observations contribute to the estimation of the survival curve up to the time of censoring. The survivor curve is calculated by the successive multiplication of the probability of having an event in each earlier time period.

Kaplan-Meier Estimate

Assume we have events occurring at times $t_1, t_2, \dots, t_k$ which are ordered in increasing time. Then from the Multiplication Law of Probability:

$$\text{Prob}(\text{alive beyond } t_j) = \text{Prob}(\text{alive just before } t_j) \times \text{Prob}(\text{not dying at } t_j)$$

or in a shorter notation:

$$S(t_j^+) = S(t_j^-) \times P(t_j)$$

The survival curve is then constructed as a step function with downward steps at each event times t_j . Thus $S(t_j^+)$ equals $S(t_{j+1}^-)$ and so the curve becomes a series of multiplications.

$$\begin{aligned} S(t_1^+) &= S(t_1^-) \times P(t_1) = 1.0 \times P(t_1) \\ S(t_2^+) &= S(t_2^-) \times P(t_2) = S(t_1^-) \times P(t_1) \times P(t_2) \\ &= 1.0 \times P(t_1) \times P(t_2) \\ S(t_3^+) &= S(t_3^-) \times P(t_3) = 1.0 \times P(t_1) \times P(t_2) \times P(t_3) \quad \text{etc.} \end{aligned}$$

At each event time t_j

$$\begin{aligned} P(\text{No Event at } t_j) &= P(t_j) \\ &= \frac{\text{Number of subjects not having an event at } t_j}{\text{Number of subjects at risk at } t_j} \\ &= \frac{n_j - d_j}{n_j} \end{aligned}$$

where n_j subjects are at risk at time t_j and d_j subjects have an event, at t_j . This procedure is called the Kaplan-Meier product moment estimator [11] of the survivor function and assumes that the individual time of each event or each end of observation is known for all subjects.

Example 1: Calculation of Survival Curve (Kaplan-Meier)

The following survival time (in days) were recorded in one group of patients in a clinical trial; the + indicates a censored observation.

2+, 2+, 5, 9+, 14, 16+, 16+, 17+, 29, 30+, 37, 37+, 39, 44+, 44+, 58+, 60, 67, 68, 82, 82, 86+, 86+, 89, 93+, 97, 100+, 100+, 100+, 100+.

The calculations are organized in columns,

t_j	n_j	d_j	$P(t_j)$	$S(t_j)$	SE
0	29	1		1.0	
5	27	1	$26/27=0.9630$	$1.0 \times 0.9630=0.9630$	0.0363
14	25	1	$24/25=0.9600$	$0.9630 \times 0.9600=0.9244$	0.0517
29	21	1	$20/21=0.9524$	$0.9244 \times 0.9524=0.8804$	0.0651
37*	19	1	$18/19=0.9474$	$0.8804 \times 0.9474=0.8341$	0.0764
39	17	1	$16/17=0.9412$	$0.8341 \times 0.9412=0.7850$	0.0863
60	13	1	$12/13=0.9231$	$0.7850 \times 0.9231=0.7246$	0.0985
67	12	1	$11/12=0.9167$	$0.7246 \times 0.9167=0.6642$	0.1072
68	11	1	$10/11=0.9091$	$0.6642 \times 0.9091=0.6039$	0.1132
82	10	2**	$8/10=0.8000$	$0.6039 \times 0.8000=0.4831$	0.1185
89	6	1	$5/6=0.8333$	$0.4831 \times 0.8333=0.4026$	0.1231
97	4	1	$3/4=0.7500$	$0.4026 \times 0.7500=0.3018$	0.1270

N.B. * Censored observation assumed to “outlive” the death at 37.

** Two patients died on the same day of follow-up.

The standard error in the far right hand column is computed from Greenwood’s formula as:

$$SE(S(t_j)) = S(t_j) \left[\sum_{i=1}^{i=j} \frac{d_i}{(n_i(n_i - d_i))} \right]^{1/2}$$

and can be used to place an approximate 95 % confidence interval on $S(t_j)$ in the usual way as,

$$S(t_j) \pm 1.96 \times SE(S(t_j)).$$

Fig. 2 Survival curve for Example 1 (‘+’ indicates an censored observation)

The estimated survival curve is plotted in Fig. 2.

Hazard Function

While the survivor function focuses on the probability of not failing, the hazard function [12, 13] focuses on failing. In many situations, it is important to know how the hazard (risk) of a particular outcome changes with time. For example, it is well known that infant mortality is highest in the few days following birth and thereafter declines very rapidly. The hazard function, $h(t)$, gives the instantaneous risk per unit time for the event (death, relapse, etc.) to occur, given that the individual has survived up to time t . This is defined as;

$$h(t) = \text{Prob}(\text{Event in } t, t + \delta t | \text{no event prior to } t)$$

In other words given that a subject has not had an event up to time t the hazard is the probability that the subject will have an event in the next small time interval. We often want to look at the way hazard changes over time. To do this we must divide time up to intervals and calculate an average hazard for each interval as,

$$h(t) = \frac{\text{Number of events in interval}}{\text{Sum of time at risk for subjects in interval}}$$

Using an interval of 30 days with the earlier data results in Table 1.

Table 1 Calculation of hazards for data of Example 1

Time Interval	Mid-point	Number of events	Time at risk in interval	h(t)
1-30	15	3	$2+2+5+9+14+16+16+17+29+20 \times 30=710$	$3/710=0.0042$
31-60	45	3	$7+7+9+14+14+28+13 \times 30=469$	$3/469=0.0064$
61-90	75	5	$7+8+22+22+26+26+29+5 \times 30=290$	$5/290=0.0172$
91-100	95	1	$3+7+10+10+10=40$	$1/40=0.0250$

Table 2 Data for 30 resected melanoma patients

Patient number	Age	Sex ^a	Initial stage	Treatment received ^b	Remission duration ^c	Survival time	Remission status ^d	Survival status ^e
1	59	2	3B	1	33.7	33.7	0	0
2	50	2	3B	1	3.8	3.9	1	1
3	76	1	3B	1	6.3	10.5	1	1
4	66	2	3B	1	2.3	5.4	1	1
5	33	1	3B	1	6.4	19.5	1	1
6	23	2	3B	1	23.8	23.8	0	0
7	40	2	3B	1	1.8	7.9	1	1
8	34	1	3B	1	5.5	16.9	1	0
9	34	1	3B	1	16.6	16.6	0	0
10	38	2	2	1	33.7	33.7	0	0
11	54	2	2	1	17.1	17.1	0	0
12	49	1	3B	2	4.3	8.0	1	1
13	35	1	3B	2	26.9	26.9	0	0
14	22	1	3B	2	21.4	21.4	0	0
15	30	1	3B	2	18.1	18.1	0	0
16	26	2	3B	2	5.8	16.0	1	0
17	27	1	3B	2	3.0	6.9	1	1
18	45	2	3B	2	11.0	11.0	0	0
19	76	2	3A	2	22.1	24.8	1	0
20	48	1	3A	2	23.0	23.0	0	0
21	91	1	4A	2	6.8	8.3	1	1
22	82	2	4A	2	10.8	10.8	0	0
23	50	2	4A	2	2.8	12.2	1	0
24	40	1	4A	2	9.2	12.5	1	0
25	34	1	3A	2	15.9	24.4	1	1
26	38	1	4A	2	4.5	7.7	1	1
27	50	1	2	2	9.2	14.8	1	0
28	53	2	2	2	8.2	8.2	0	0
29	48	2	2	2	8.2	8.2	0	0
30	40	2	2	2	7.8	7.8	0	0

^a 1 male; 2, female.^b 1, BCG; 2, *C.parvum*.^c Remission and survival times are in months.^d 1 Remission achieved; 0 censored^e 1 Death; 0 censored.

Constant Hazard

Plotting the hazards above against the mid point of the time interval shows a hazard increasing with time. However, the constant hazard situation is important to recognize if present since it implies *exponential* survival. In this special case the mathematical form of the survival function is,

$$S(t) = e^{-ht}$$

Thus,

$$\text{Log}_e S(t) = -ht$$

Exponential survival can thus be seen more directly by plotting the survival curve on a natural logarithmic scale. The survival curve would then become a straight line with downward slope equal to h .

A usefull single summary statistic in this situation is the average rate of events. This is calculated as,

$$h = \frac{\text{Total number of events}}{\text{Total observation time}} = \frac{d}{V}$$

For the data above this is,

$$h = 12/1509 = 0.00795 (\text{i.e. } 0.8\% \text{ per day})$$

An associated approximate standard error for h is,

$$SE(h) \approx h/\sqrt{d}$$

Comparison of Two Survival Functions

Example 2: Comparison of Two Treatments

Thirty melanoma patients (Stages 2–4) were studied to compare the immunotherapies BCG (*Bacillus Calmette-Guérin*) and *Corynebacterium parvum* for their abilities to prolong remission and survival time. The age, sex, disease stage, treatment received, remission duration, and survival times are given in following Table 2.

All the patients were resected before treatment began and thus have no evidence of melanoma at the time of first treatment.

The usual objective of this type of data is to determine the length of remission and survival and to compare the distributions of remission and survival time in each group. Before comparing the remission and survival distributions, we must examine if the two treatment groups are comparable with respect to prognostic factors. Let us use survival time to illustrate the steps. (The remission time could be similarly analyzed).

For such samples with progressively censored observations, the Kaplan-Meier product-limit (PL) method is

Table 3 Kaplan-Meier product-limit estimate of survival function $S(t)$

BCG Patients					
Death time(t)	3.9	5.4	7.9	10.5	19.5
S(t)	0.909	0.818	0.727	0.636	0.477
<i>C. parvum</i> patients					
Death time(t)	6.9	7.7	8.0	8.3	24.4
S(t)	0.947	0.895	0.839	0.774	0.516

appropriate for estimating the survival function. It does not require any assumptions about the form of the function that is being estimated. We have used SPSS 16.0 to find survival curves. Table 3 gives the PL estimate of the survival function for the two treatment groups.

The median survival time for BCG patients was 19.5 months. The median survival time for the *C. parvum* group can not be calculated since 14 of the 19 patients were still alive. Figure 3 plots the estimated survival function $S(t)$ for patients receiving the two treatments. The median survival time for BCG group can also be determined graphically. The survival curves clearly show that *C. parvum* patients had slightly better survival experience than BCG patients. For example, 50 % of the BCG patients survived at least 19.5 months while about 61 % of the *C. parvum* patients survived that long.

Fig. 3 Survival curves of patients receiving BCG and *C. parvum*

Compare the Two Survival Distributions

Want to ask whether the difference in survival between the two treatment groups is statistically significant. Before applying a statistical test, we should ensure that the two groups of patients are homogenous with respect to prognostic factors. We have checked the homogeneity of the two treatment groups with respect to age, sex, and disease stage. It is found that the data do not show heterogeneity between the two treatment groups. If heterogeneity is found, the groups can be divided into subgroups of members who are similar in their prognoses.

There are several parametric and nonparametric tests to compare two survival distributions. Since we have no information of the survival distribution that the data follow, we would continue to use nonparametric methods to compare the two survival distributions. There are several nonparametric tests: (i) Gehan's generalized Wilcoxon test, (ii) the Cox-Mantel test, (iii) the logrank test, (iv) Peto and Peto's generalized Wilcoxon test, and (v) Cox's *F*-test. All the tests are designed to handle censored data.

Using Logrank test, the difference in survival distribution of the two treatment groups is found to be insignificant ($p=0.387$). Therefore the data do not provide enough evidence to reject the hypothesis that the two survival distributions are same.

Comparison of Tests

When sample sizes are small (<50), Cox's *F*-test is more powerful than Gehan's generalized Wilcoxon's test if samples are from exponential or Weibull distributions and if there are no censored observations or the observations are singly censored. When samples are from exponential distributions, with or without censoring the Cox-Mantel and logrank tests are more powerful and more efficient than the generalized Wilcoxon tests of Gehan and Peto and Peto. There is little difference between the Cox-Mantel and the logrank tests and the two generalized Wilcoxon tests. When the samples are taken from Weibull distributions with constant hazard ratio (i.e., the ratio of the two hazard functions does not vary with time), the results are essentially the same as in the exponential case. However, when the hazard ratio is nonconstant, the two generalizations of the Wilcoxon test have more power than the other tests. Thus, the logrank test is more powerful than the Wilcoxon tests in detecting departures when the two hazard functions are parallel (proportional hazard) or there is random but equal censoring and when there is no censoring in the samples.

The generalized Wilcoxon tests give more weight to early failures than later failures whereas the logrank test gives equal weight to all failures. Therefore, the generalized Wilcoxon tests are more likely to detect early differences in the

two survival distributions whereas the logrank test is more sensitive to differences at the right tails. If heavy censoring exists, Gehan's statistic is dominated by a small number of early failures and has low power.

There are situations in which neither the logrank nor the Wilcoxon tests are very effective. When the two distributions differ but their hazard functions or survivorship functions cross, neither the Wilcoxon nor the logrank test is very powerful and it will be sensible to consider other tests. For example, Tarone and Ware [14] discuss general statistics of similar form (using scores) and Fleming et al. [15] present a two-sample test based on a Smirnov-type statistic designed to measure the maximum distance between estimates of two distributions. The latter approach is shown to be more effective than the logrank or Wilcoxon tests when the two survival distributions differ substantially for some range of t values but not necessarily elsewhere.

SPSS 16.0 version can be used to find Kaplan-Meier estimates and the tests such as : (i) logrank test, (ii) Tarone-Ware test and (iii) Breslow test.

Discussion

Here we have presented the survival analysis of the data in a way that it can be followed simply and without the help of the computers as well as using computer softwares. The detailed calculations will give an insight into the understanding of the subject matter. Of course, the softwares, such as SPSS and STATA are available for the purpose, but their rationale and appropriate use cannot be accomplished without the proper knowledge of the subject.

References

1. Cox DR, Oakes D (1984) Analysis of survival data. Chapman and Hall, New York
2. Armitage P, Gehan EA (1974) Statistical methods for the identification and use of prognostic factors. *Int J Cancer* 13:16–35
3. Pocock SJ, Gore SM, Kerr GM (1982) Long term survival analysis: the curability of cancer. *Stat Med* 1:93–104
4. Abdi EA, Hanson J, McPherson TA (1987) Adjuvant chemoimmunotherapy after regional lymphadenectomy for malignant melanoma. *Am J Clin Oncol* 10:117–122
5. Bonadonna G et al (1976) Combination chemotherapy as an adjuvant treatment in operable breast cancer. *N Engl J Med* 294:405–410
6. Armitage P (1959) The comparison of survival curves. *Journal of the Royal Statistical Society, Series A* 122:279–300
7. Soler-Vila H, Kasl SV, Jones BA (2005) Cancer-specific beliefs and survival: a population-based study of African-American and white breast cancer patients. *Cancer Causes Control* 16:105–114
8. Brookmeyer R, Goedert JJ (1989) Censoring in an epidemic with an application to hemophilia-associated AIDS. *Biometrics* 45:325–335

9. Cohen AC Jr (1963) Progressively censored sample in life testing. *Technometrics* 3:535–541
10. Ederer F, Axtell LM, Cutler SJ (1961) The relative survival rate- a statistical methodology. National Cancer Institute Monographs 6:101–121
11. Breslow NE, Day NE (1980) Statistical Methods in Cancer Research, Vol 1-The Analysis of Case-control Studies. International Agency for Research on Cancer, Lyon
12. Nelson W (1972) Theory and applications of hazard plotting for censored failure data. *Technometrics* 14:945–966
13. Nadas A (1970) On proportional hazard functions. *Technometrics* 12:413–416
14. Tarone RE, Ware J (1977) On distribution-free tests for equality of survival distributions. *Biometrics* 64:156–160
15. Fleming TR, Harrington DP, O'Brien PC (1984) Design for group sequential tests. *Control Clin Trials* 5:348–361