

# Introduction to survival analysis

Paul Dickman

September 2020  
[www.pauldickman.com](http://www.pauldickman.com)

## Overview of this lecture

Target audience is students and researchers in biomedical sciences without extensive training in statistics.

1. Introduction to survival analysis.
  2. Estimating the survival function using the Kaplan-Meier method.
  3. Testing for differences in survival using the log rank test.
- ▶ Slides available at  
<http://www.pauldickman.com/video/survival-intro/>
  - ▶ R, Stata and SAS code available on the same page as the slides.

## About Paul Dickman

- ▶ Professor of Biostatistics at Karolinska Institutet, Stockholm, Sweden.
- ▶ Research interests in register-based epidemiology, with particular focus on cancer epidemiology.
- ▶ Primary research interests are development and application of statistical methods for studying the survival of cancer patients using registry data.
- ▶ Stata is my primary software (<http://www.pauldickman.com/#software>).

## Terminology

- ▶ In the strictest sense, a *ratio* is the result of dividing one quantity by another. In the sciences, however, it is mostly used in a more specific sense, that is, when the numerator and the denominator are two separate and distinct quantities [1].
- ▶ A *proportion* is a type of ratio in which the numerator is included in the denominator.
- ▶ A *rate* is a measure of change in one quantity per unit of another quantity.
- ▶ The 'survival rate' of a group of patients over a specified time period is defined as the 'number of patients surviving' divided by 'the total alive at start'. It is therefore not strictly a rate, but a proportion.
- ▶ We will estimate both proportions (e.g., survival proportions) and rates (e.g., mortality rates) and should recognise that these are conceptually different.

## Analysis of time-to-event data (survival analysis)

- ▶ Survival analysis concerns analysing the time to the occurrence of an event, e.g., time from diagnosis of a disease to death.
- ▶ Time-to-event data are generated when the response measurement of interest is the time from a well-defined origin of measurement to occurrence of an event of interest.
- ▶ The outcome can be thought of as comprising two dimensions
  - a. event indicator (binary), and
  - b. time at risk (continuous)
- ▶ Time-to-event analysis is also known as failure time analysis (primarily in engineering), lifetime analysis, and survival analysis.

## What can we estimate from time-to-event data?

- ▶ Survival probability, i.e., the proportion who have not experienced the event (e.g., death) at a given time point during follow-up
- ▶ Summary measures (e.g., mean or median) of survival time
- ▶ Event rates (hazard functions); the event rate at time  $t$  conditional on survival until time  $t$  or later
- ▶ Hazard ratios; ratios of hazard functions between different groups (e.g., exposed vs. unexposed) while adjusting for confounders
- ▶ The characteristic that complicates the use of standard statistical methods is *censoring* — which occurs when the survival time is only partially known (details to follow).

## Why do we need survival analysis methods?

- ▶ If everyone has complete follow-up, then it is OK to ignore time.
- ▶ For example, when studying 30-day mortality following surgery we can just study the binary outcome (did the patient die within 30 days) provided everyone is (potentially) followed for 30 days.
- ▶ However, if not all individuals have the same potential time at risk then we need special methods (survival analysis).
- ▶ For example, if we recruit all patients diagnosed with cancer during 2015–2019 and follow them up for death until the end of 2019 then those diagnosed later will have a shorter follow-up (and therefore be less likely to die during follow-up).

## Censoring

- ▶ Censoring (right censoring) refers to the situation where an individual can no longer be followed up and the event of interest has not occurred.
- ▶ In studying the survival of cancer patients, for example, patients enter the study at the time of diagnosis (or the time of treatment in randomised trials) and are followed up until the event of interest is observed. Censoring may occur in one of the following forms:
  - ▶ Termination of the study before the event occurs (administrative censoring);
  - ▶ Death due to a cause not considered to be the event of interest (in cause-specific survival analyses); and
  - ▶ If the patient emigrates or withdraws from the study.
- ▶ These are examples of right censoring, which is the most common form of censoring in medical studies.
- ▶ Left censoring, interval censoring, left truncation, and right truncation will not be considered here.

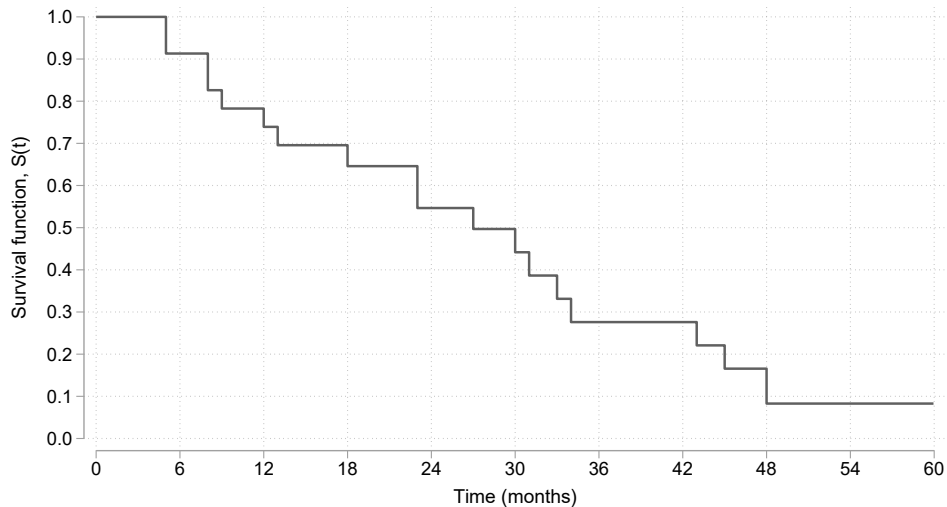


## Sample data: AML (acute myeloid leukemia)

- ▶ Data set is distributed with the R survival package.
- ▶ Research question at the time was whether the standard course of chemotherapy should be extended ('maintained') for additional cycles.
- ▶ Data on 23 patients; 3 variables
  - `time` time at risk (months)
  - `status` 1 if dead; 0 if censored
  - `x` maintenance chemotherapy; 0=yes; 1=no

ID	time	status	x
1.	5	1	Nonmaintained
2.	5	1	Nonmaintained
3.	8	1	Nonmaintained
4.	8	1	Nonmaintained
5.	9	1	Maintained
6.	12	1	Nonmaintained
7.	13	0	Maintained
8.	13	1	Maintained
9.	16	0	Nonmaintained
10.	18	1	Maintained
11.	23	1	Maintained
12.	23	1	Nonmaintained
13.	27	1	Nonmaintained
14.	28	0	Maintained
15.	30	1	Nonmaintained
16.	31	1	Maintained
17.	33	1	Nonmaintained
18.	34	1	Maintained
19.	43	1	Nonmaintained
20.	45	0	Maintained
21.	45	1	Nonmaintained
22.	48	1	Maintained
23.	161	0	Maintained

## Estimated survival function (Kaplan-Meier method) for the AML data

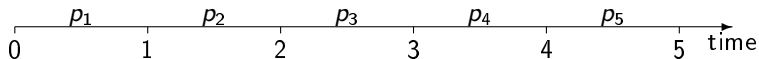


## The survival function, $S(t)$

- ▶  $S(t)$  is a nonincreasing function with a value 1 at time zero and value 0 as  $t$  approaches infinity.
- ▶  $S(t)$  is a **function** which depends on  $t$  and should not be referred to as the survival rate.
- ▶ Nonparametric methods for estimating  $S(t)$  (described next) involve estimating the survival proportion at discrete values of  $t$  and then interpolating these to obtain an estimate of  $S(t)$ .

## General approach to nonparametric estimation of $S(t)$

- ▶ Assume we wish to estimate five-year survival,  $S(5)$ .
- ▶ 18/23 patients dies within 5 years, but not all 23 could be followed for 5 years. Cannot use this as the estimate of 5-year survival.
- ▶ We recognise that, in order to survive 5 years, one must survive each and every year (up to 5).



- ▶ We start by estimating the following conditional survival probabilities:
  - $p_1$ , the probability of surviving at least 1 year from time 0
  - $p_2$ , the probability of surviving at least 2 years conditional on surviving 1 year
  - $p_3$ , the probability of surviving at least 3 years conditional on surviving 2 years
  - $p_4$ , the probability of surviving at least 4 years conditional on surviving 3 years
  - $p_5$ , the probability of surviving at least 5 years conditional on surviving 4 years

## General approach to nonparametric estimation of $S(t)$ (2)

- ▶ The probability of surviving at least 5 years (from time zero) is then given by the product of these conditional survival probabilities.

$$S(5) = \prod_{i=1}^5 p_i$$

- ▶ That is, to survive five years one must survive year 1 and year 2 and year 3, and year 4, and year 5.
- ▶ The advantage of this approach is that we can appropriately account for censoring when estimating the probability of surviving a small time interval (i.e., when estimating the conditional survival probabilities).
- ▶ This approach is employed by the Kaplan-Meier (also known as product-limit) method [2] and the life table method (not covered here).

## Overview: Kaplan-Meier method for estimating $S(t)$

- ▶ Split follow-up into intervals, with a new interval at every event time.
- ▶ Estimate conditional probabilities of surviving each interval

$$p_i = 1 - d_i/n_i$$

where  $d_i$  is the number of events and  $n_i$  number at risk for interval  $i$ .

- ▶  $S(t)$  is the product of the conditional probabilities up to time  $t$ .

$$S(t_k) = \prod_{i=1}^k p_i$$

- ▶ If there are both events and censorings at the same time (as at times 13 and 45), assume the events precede the censorings. That is, censored observations are assumed to be at risk when the events occur.

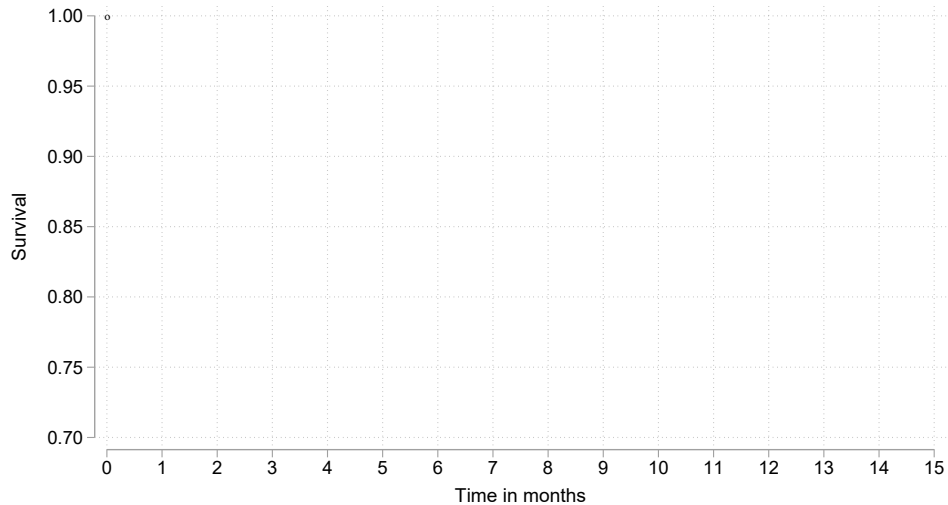
## Data

## Kaplan-Meier

ID	time	status	x	Tabulation of number at risk and number of deaths at each event time				
				t	atrisk	dead	p	S(t)
1.	5	1	Nonmaintained	0	23			1
2.	5	1	Nonmaintained					
3.	8	1	Nonmaintained	5	23	2	21/23	1*21/23
4.	8	1	Nonmaintained	8	21	2	19/21	1*21/23*19/21
5.	9	1	Maintained	9	19	1	18/19	1*21/23*19/21*18/19
6.	12	1	Nonmaintained	12	18	1	17/18	1*21/23*19/21*18/19*17/18
7.	13	0	Maintained	13	17	1	16/17	
8.	13	1	Maintained	16	15	0	15/15	
9.	16	0	Nonmaintained	18	14	1	13/14	
10.	18	1	Maintained	23	13	2	11/13	
11.	23	1	Nonmaintained	27	11	1	10/11	
12.	23	1	Nonmaintained	28	10	0	10/10	
13.	27	1	Nonmaintained	30	9	1	8/9	
14.	28	0	Maintained	31	8	1	7/8	
15.	30	1	Nonmaintained	33	7	1	6/7	
16.	31	1	Maintained	34	6	1	5/6	
17.	33	1	Nonmaintained	43	5	1	4/6	
18.	34	1	Maintained	45	4	1	3/4	
19.	43	1	Nonmaintained	48	2	1	1/2	
20.	45	0	Maintained	161	1	0	1/1	
21.	45	1	Nonmaintained					
22.	48	1	Maintained					
23.	161	0	Maintained					

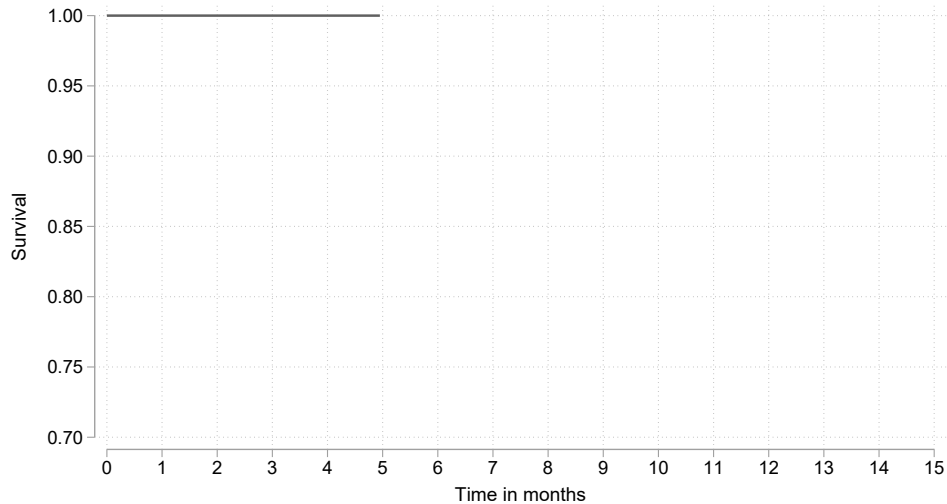


Survivor function at time 0 is  $S(t)=1$



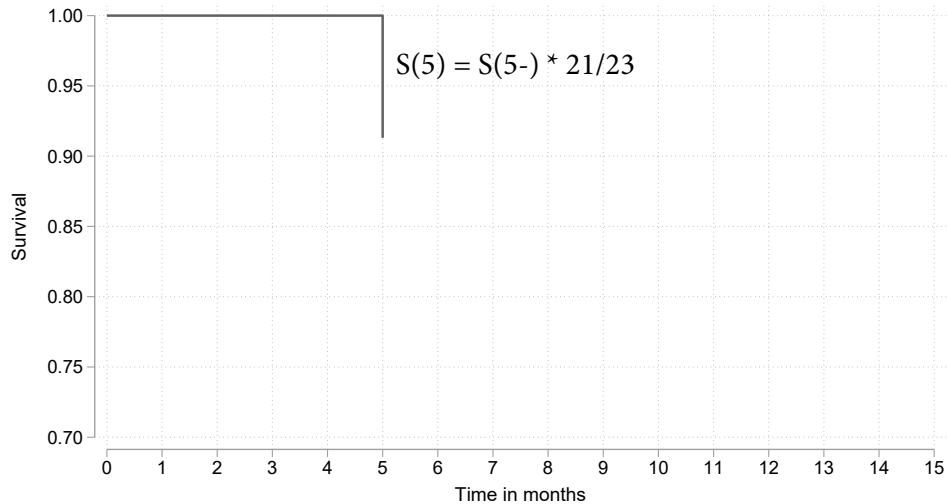
## Survivor function up to time 5- (just before time 5)

t	atrisk	dead	p
0	23		
5	23	2	21/23



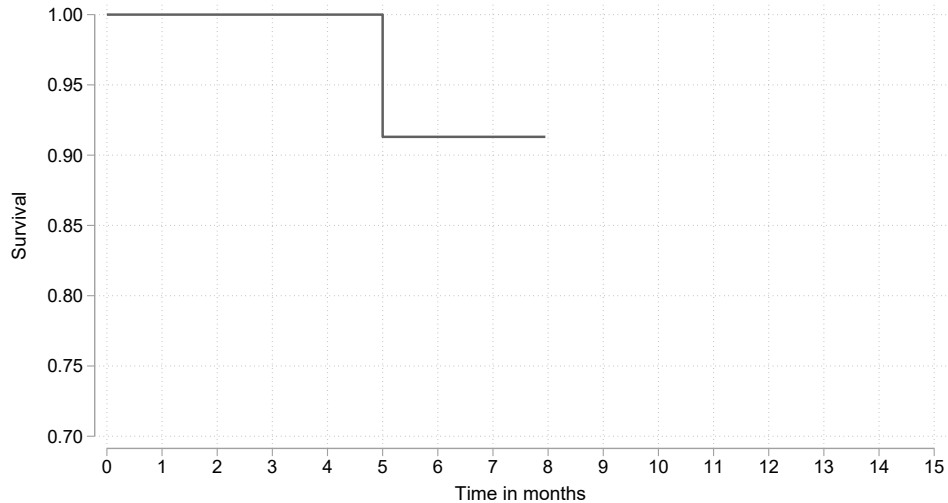
## Survivor function up to time 5

t	atrisk	dead	p
0	23		
5	23	2	21/23



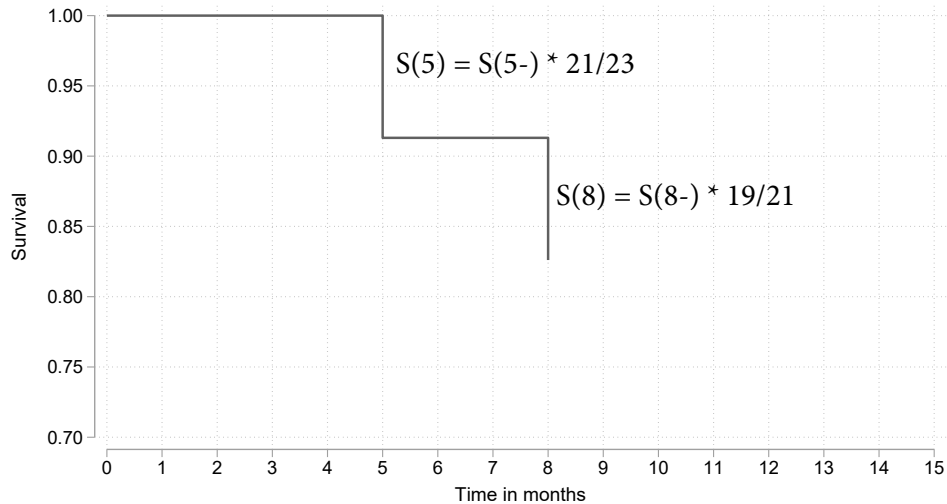
## Survivor function up to time 8- (just before time 8)

t	atrisk	dead	p
5	23	2	21/23
8	21	2	19/21



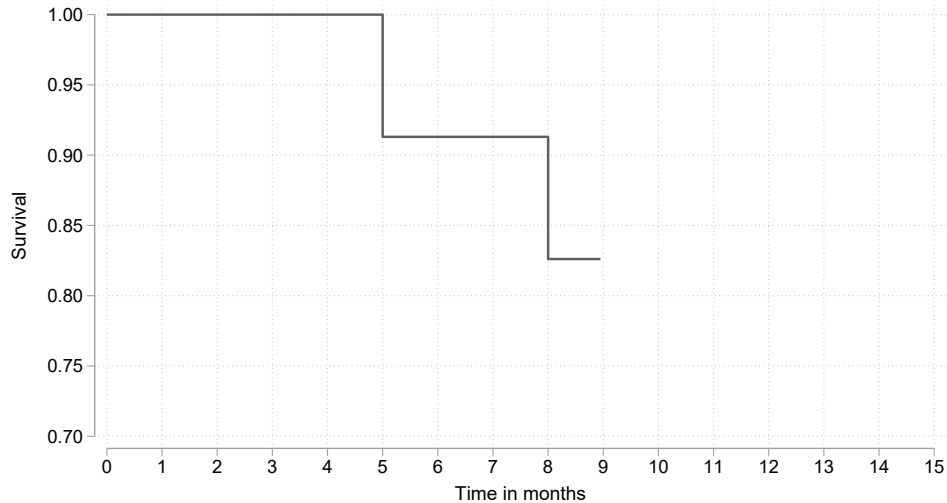
## Survivor function up to time 8

t	atrisk	dead	p
5	23	2	21/23
8	21	2	19/21



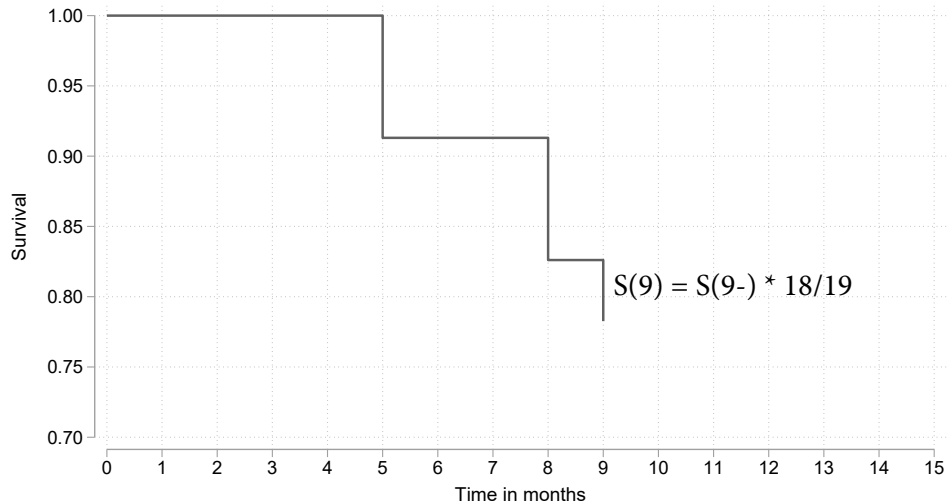
## Survivor function up to time 9- (just before time 9)

t	atrisk	dead	p
8	21	2	19/21
9	19	1	18/19



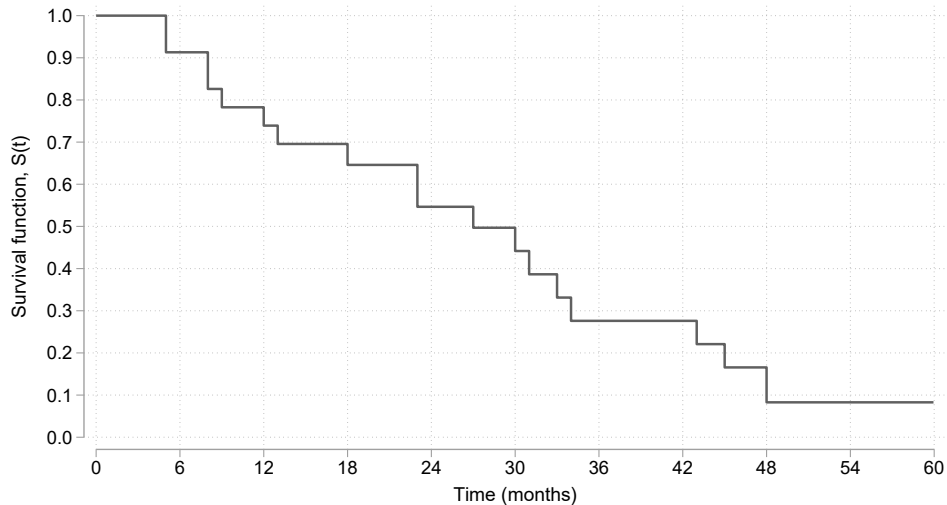
## Survivor function up to time 9

t	atrisk	dead	p
8	21	2	19/21
9	19	1	18/19



## Survivor function does not change at times 16 and 28

t	atrisk	dead	p
16	15	0	15/15
18	14	1	13/14





## Details: Kaplan-Meier method for estimating $S(t)$

- ▶ To obtain Kaplan-Meier estimates of survival, the patient survival times are first ranked in increasing order.
- ▶ The times where events (deaths) occur are denoted by  $t_j$ , where  $t_1 < t_2 < t_3 < \dots$
- ▶ The number of deaths occurring at  $t_j$  is denoted by  $d_j$ .
- ▶ If both censoring(s) and death(s) occur at the same time, then the censoring(s) are assumed to occur immediately after the death time.
- ▶ That is, individuals with survival times censored at  $t_j$  are assumed to be at risk at  $t_j$ .
- ▶ The Kaplan-Meier estimate of the cumulative survivor function at time  $t$  is given by

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_j \leq t} (1 - \frac{d_j}{n_j}) & \text{if } t \geq t_1 \end{cases} \quad (1)$$

where  $n_j$  is the number of persons at risk.

## Details: Kaplan-Meier method for estimating $S(t)$ (2)

- ▶ A plot of the Kaplan-Meier estimate of the survivor function takes the form of a step function, in which the survival probabilities decrease at each death time and are constant between adjacent deaths times.
- ▶ Censorings do not affect the estimate of  $S(t)$ , but contribute in Equation 1 by decreasing  $n_i$  at the next death time.
- ▶ If the largest observed survival time (which we will call  $t_z$ ) is a censored survival time, then  $\hat{S}(t)$  is undefined for  $t > t_z$ , otherwise  $\hat{S}(t) = 0$  for  $t > t_z$ .
- ▶ The standard error of the estimate can be obtained using Greenwood's method [3].
- ▶ At  $t = 5$  months we observed 2 deaths among the 23 patients at risk, so  $p_1 = 1 - 2/23 = 0.913$ .
- ▶ At  $t = 8$  months we observed 2 deaths among the 21 patients at risk, so  $p_2 = 1 - 2/21 = 0.905$ .
- ▶ Subsequently,  $\hat{S}(t) = 0.913 \times 0.905 = 0.826$  for  $8 \leq t < 9$ .

## Kaplan-Meier estimates in R

```
> kmfit <- survfit(Surv(time, status) ~ 1, data = aml)
> summary(kmfit)
```

```
Call: survfit(formula = Surv(time, status) ~ 1, data = aml)
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
5	23	2	0.9130	0.0588		0.8049		1.000
8	21	2	0.8261	0.0790		0.6848		0.996
9	19	1	0.7826	0.0860		0.6310		0.971
12	18	1	0.7391	0.0916		0.5798		0.942
13	17	1	0.6957	0.0959		0.5309		0.912
18	14	1	0.6460	0.1011		0.4753		0.878
23	13	2	0.5466	0.1073		0.3721		0.803
27	11	1	0.4969	0.1084		0.3240		0.762
30	9	1	0.4417	0.1095		0.2717		0.718
31	8	1	0.3865	0.1089		0.2225		0.671
33	7	1	0.3313	0.1064		0.1765		0.622
34	6	1	0.2761	0.1020		0.1338		0.569
43	5	1	0.2208	0.0954		0.0947		0.515
45	4	1	0.1656	0.0860		0.0598		0.458
48	2	1	0.0828	0.0727		0.0148		0.462

```
> ## Plot the Kaplan-Meier estimates
> plot(kmfit, xmax=60)
```

## Kaplan-Meier estimates in Stata – stset

```
. stset time, fail(status)
```

```
      failure event:  status != 0 & status < .  
obs. time interval:  (0, time]  
exit on or before:  failure
```

```
-----  
23  total observations  
0   exclusions  
-----
```

```
23  observations remaining, representing  
18  failures in single-record/single-failure data  
678 total analysis time at risk and under observation  
                                     at risk from t =          0  
earliest observed entry t =          0  
last observed exit t =          161
```

## Kaplan-Meier estimates in Stata

```
. sts list
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int.]	
5	23	2	0	0.9130	0.0588	0.6949	0.9775
8	21	2	0	0.8261	0.0790	0.6006	0.9309
9	19	1	0	0.7826	0.0860	0.5542	0.9032
12	18	1	0	0.7391	0.0916	0.5092	0.8734
13	17	1	1	0.6957	0.0959	0.4656	0.8417
16	15	0	1	0.6957	0.0959	0.4656	0.8417
18	14	1	0	0.6460	0.1011	0.4140	0.8053
23	13	2	0	0.5466	0.1073	0.3193	0.7264
27	11	1	0	0.4969	0.1084	0.2756	0.6842
28	10	0	1	0.4969	0.1084	0.2756	0.6842
30	9	1	0	0.4417	0.1095	0.2274	0.6371
31	8	1	0	0.3865	0.1089	0.1828	0.5875
33	7	1	0	0.3313	0.1064	0.1418	0.5353
34	6	1	0	0.2761	0.1020	0.1044	0.4803
43	5	1	0	0.2208	0.0954	0.0710	0.4222
45	4	1	1	0.1656	0.0860	0.0421	0.3604
48	2	1	0	0.0828	0.0727	0.0070	0.2868
161	1	0	1	0.0828	0.0727	0.0070	0.2868

```
. sts graph // Plot the Kaplan-Meier estimates (does not require sts list)
```

## The hazard function, $\lambda(t)$

- ▶ The term 'hazard' is the generic term used in survival analysis to describe the 'event rate'. If, for example, the event of interest is disease incidence then the hazard represents the incidence rate.
- ▶ The hazard function,  $\lambda(t)$ , is the instantaneous event rate at time  $t$ , conditional on survival up to time  $t$ . The units are events per unit time.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(\text{event in } (t, t + \Delta t] \mid \text{alive at } t)}{\Delta t} \quad (2)$$

- ▶ In contrast to the survivor function, which describes the probability of *not* failing before time  $t$ , the hazard function focuses on the failure rate at time  $t$  among those individuals who are alive at time  $t$ .
- ▶ That is, a lower value for  $\lambda(t)$  implies a higher value for  $S(t)$  and vice-versa.
- ▶ Note that the hazard is a rate, not a proportion or probability, so  $\lambda(t)$  can take on any value between zero and infinity, as opposed to  $S(t)$  which is restricted to the interval  $[0, 1]$ .



## Relation between the survivor and hazard functions (for completeness)

$$\begin{aligned}h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(\text{event in } (t, t + \Delta t] \mid \text{alive at } t)}{\Delta t} \\&= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{S(t) \times \Delta t} \text{ where } F(t) = 1 - S(t) \\&= \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \times \frac{-1}{S(t)} \\&= \frac{dS(t)}{dt} \times \frac{-1}{S(t)} \text{ by definition of a derivative} \\&= -\frac{d \ln S(t)}{dt} \text{ since } d/dx \ln(f(x)) = f'(x)/f(x)\end{aligned}$$

 Other relationships (for completeness)

$$-\frac{d \log S(t)}{dt} = h(t)$$

$\Updownarrow$

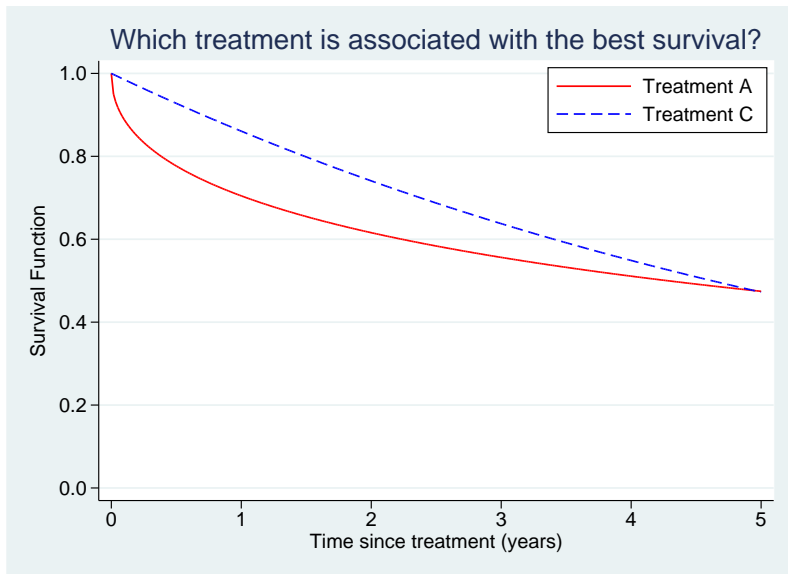
$$S(t) = \exp\left(-\int_0^t h(u) du\right) = \exp(-H(t))$$

$H(t) = \int_0^t h(u) du$  is called the *integrated hazard* or *cumulative hazard*.

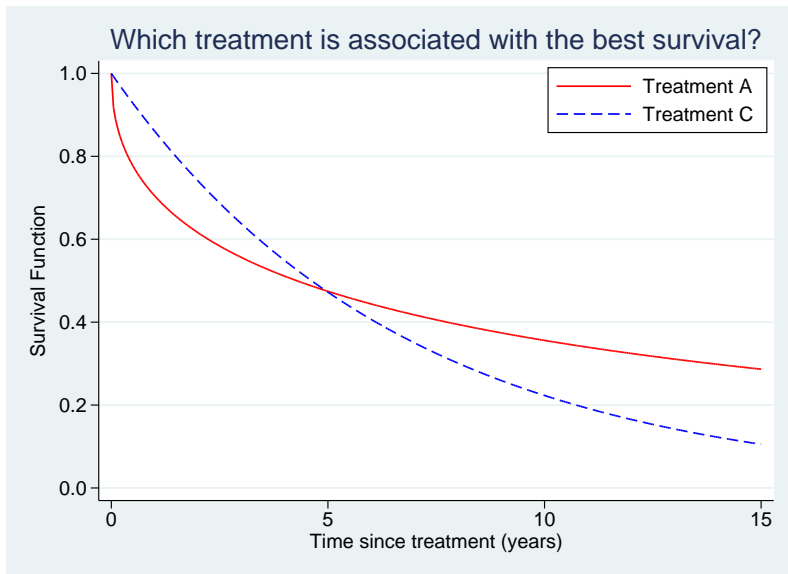
$$h(t) = -\frac{d \log(S(t))}{dt} = -\frac{S'(t)}{S(t)} = \frac{F'(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$



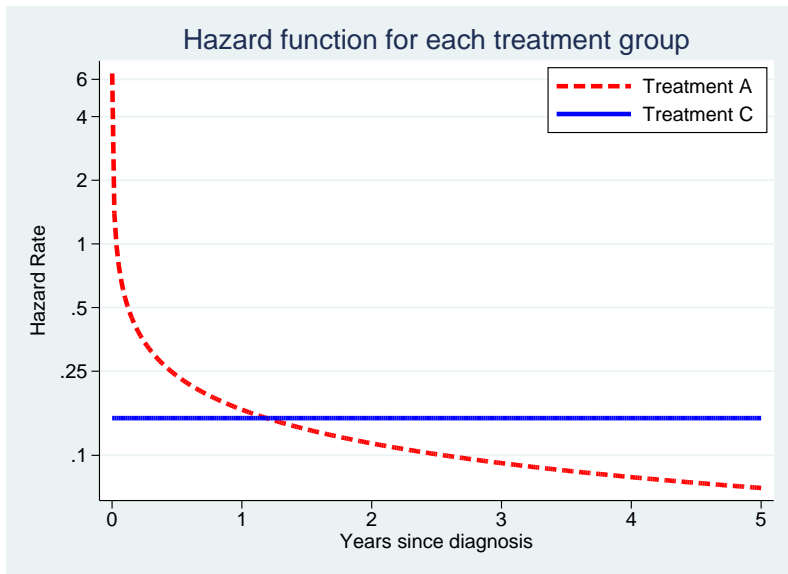
Which group (A or C) has the best survival?



What if we extend the follow-up?



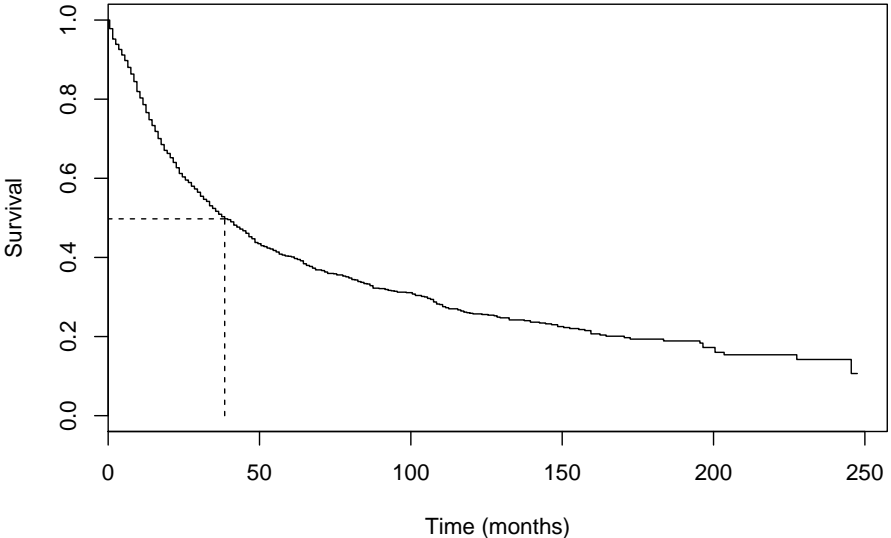
Now plot the two hazard functions



## Median survival time

- ▶ The median survival time can be used to summarise the distribution of survival times.
- ▶ The median survival time is the time at which  $S(t) = 0.5$ . That is, the time beyond which 50% of the individuals in the population are expected to survive.
- ▶ It is estimated by the time at which the estimate of  $S(t)$  falls below 0.5.
- ▶ The median survival time for the example shown on the next slide is approximately 40 months.

# An illustration of median survival time



## Testing for differences in survival between groups

- ▶ Comparing survival at a fixed time point (e.g., five years) wastes available information.
- ▶ It is invalid to compare the proportion surviving at a given time, based on the comparison of two binomial proportions, where the time point for comparison is chosen after viewing the estimated survivor functions (e.g. testing for a difference at the point where the Kaplan-Meier curves show the largest difference).
- ▶ Various tests are available (parametric and non-parametric) for testing equality of survival curves. The most common is the log rank test, which is non-parametric.
- ▶ Start by tabulating the number at risk in each group and the total number of events (deaths) at every time point when one or more deaths occur.
- ▶ Under the null hypothesis that the two survival curves are the same, the expected number of deaths in each group will be proportional to the number at risk in each group.

## Testing for differences in survival between groups (2)

- ▶ We will compare the survival functions between the two treatment groups.
- ▶ At  $t = 5$  months we observed 2 deaths (both in the 'non maintained' group).
- ▶ Conditional on 2 deaths being observed, we would expect  $2 \times 11/23 = 0.957$  deaths among the 11 people in the maintained group and  $2 \times 12/23 = 1.043$  deaths among the 12 people in the non-maintained group .
- ▶ We calculate the totals of the observed and expected deaths for each group, denoted  $O_1$ ,  $O_2$ ,  $E_1$ , and  $E_2$ , and calculate the following test statistic

$$\theta = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}. \quad (3)$$

- ▶ Under the null hypothesis,  $\theta$  will approximately follow a  $\chi^2$  distribution with 1 degree of freedom. That is, if  $\theta$  is greater than 3.84 then we reject the null hypothesis and conclude that there is a statistically significant difference between the two survival curves.

## Log rank test for comparing survival by treatment

event time	maintained			not maintained		
	at risk	obs	exp	at risk	obs	exp
5	11	0	0.957	12	2	1.043
8	11	0	1.048	10	2	0.952
9	11	1	0.579	8	0	0.421
12	10	0	0.556	8	1	0.444
13	10	1	0.588	7	0	0.412
18	8	1	0.571	6	0	0.429
23	7	1	1.077	6	1	0.923
27	6	0	0.545	5	1	0.455
30	5	0	0.556	4	1	0.444
31	5	1	0.625	3	0	0.375
33	4	0	0.571	3	1	0.429
34	4	1	0.667	2	0	0.333
43	3	0	0.600	2	1	0.400
45	3	0	0.750	1	1	0.250
48	2	1	1.000	0	0	0.000
Total		7	10.689		11	7.311

Totals:  $O_1 = 7$ ,  $E_1 = 10.69$ ,  $O_2 = 11$ ,  $E_2 = 7.31$

Test statistic:

$$\frac{(7-10.69)^2}{10.69} + \frac{(11-7.31)^2}{7.31}$$

$$= 3.4$$



## Log rank test for comparing survival by treatment

- ▶ The test statistic is  $\theta = 3.4$ , with an associated p-value of 0.07 implying no strong evidence of a difference in survival between the two groups.
- ▶ For  $k$  groups, the log rank test statistic is

$$\theta = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (4)$$

which has an approximate  $\chi_{k-1}^2$  distribution under the null hypothesis.

- ▶ The log rank test is designed to be sensitive to departures from the null hypothesis in which the two hazards (instantaneous death rates) are proportional over time. It is very insensitive to situations in which the hazard functions cross.

## Log rank test for comparing survival by treatment (2)

- ▶ It is difficult to apply the log rank test while simultaneously controlling for potential confounding variables (a regression approach is preferable).
- ▶ In a randomised clinical trial, however, potential confounders are controlled for in the randomisation, so we can use the log rank test to compare survival curves for the different treatment groups.
- ▶ The log rank test provides nothing more than a test of statistical significance for the difference between the survival curves, it tells us nothing about the size of the difference. A regression approach allows us to both determine statistical significance and to estimate the size of the effect.

## Log rank test in R

```
> survdiff(Surv(time, status) ~ x, data = aml)
```

```
Call:
```

```
survdiff(formula = Surv(time, status) ~ x, data = aml)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
x=Maintained	11	7	10.69	1.27	3.4
x=Nonmaintained	12	11	7.31	1.86	3.4

```
Chisq= 3.4 on 1 degrees of freedom, p= 0.07
```

## Log rank test in Stata

```
. sts test x
```

```
      failure _d:  status  
analysis time _t:  time
```

Log-rank test for equality of survivor functions

x	Events observed	Events expected
Maintained	7	10.69
Nonmaintained	11	7.31
Total	18	18.00

chi2(1) = 3.40  
Pr>chi2 = 0.0653

## The same test as a Cox model (described in a separate lecture)

```
. stcox x
```

```
No. of subjects =          23          Number of obs   =          23
No. of failures =          18
Time at risk    =          678
Log likelihood  = -41.250114          LR chi2(1)       =          3.30
                                          Prob > chi2      =          0.0694
```

```
-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
       x |   2.470004   1.265254     1.77   0.078     .9050476   6.740993
-----+-----
```

- ▶ Provides (conceptually) the same test as the log rank test but also provides a measure of association (hazard ratio).
- ▶ Mortality rate among patients without maintenance therapy is an estimated 2.47 times higher than the mortality rate in patients who did not.
- ▶ Can test for differences while adjusting for other variables.

## Now the Cox model in R

```
> summary(coxph(Surv(time, status) ~ x, data = aml))
```

```
n= 23, number of events= 18
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
xNonmaintained	0.9155	2.4981	0.5119	1.788	0.0737

	exp(coef)	exp(-coef)	lower .95	upper .95
xNonmaintained	2.498	0.4003	0.9159	6.813

```
Likelihood ratio test= 3.38 on 1 df, p=0.07
```

```
Wald test = 3.2 on 1 df, p=0.07
```

```
Score (logrank) test = 3.42 on 1 df, p=0.06
```

- ▶ Score test from Cox model is mathematically equivalent to the log rank test.

## References

- [1] Elandt-Johnson EC. Definition of rates: Some remarks on their use and misuse. *American Journal of Epidemiology* 1975;**102**:267–271.
- [2] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958;**53**:457–481.
- [3] Greenwood M. *The Errors of Sampling of the Survivorship Table*, vol. 33 of *Reports on Public Health and Medical Subjects*. London: Her Majesty's Stationery Office, 1926.