Survival analysis)

Daniela Zugna
Dipartimento di Scienze Biomediche ed Oncologia Umana, Università di Torino
daniela.zugna@unito.it

Definitions

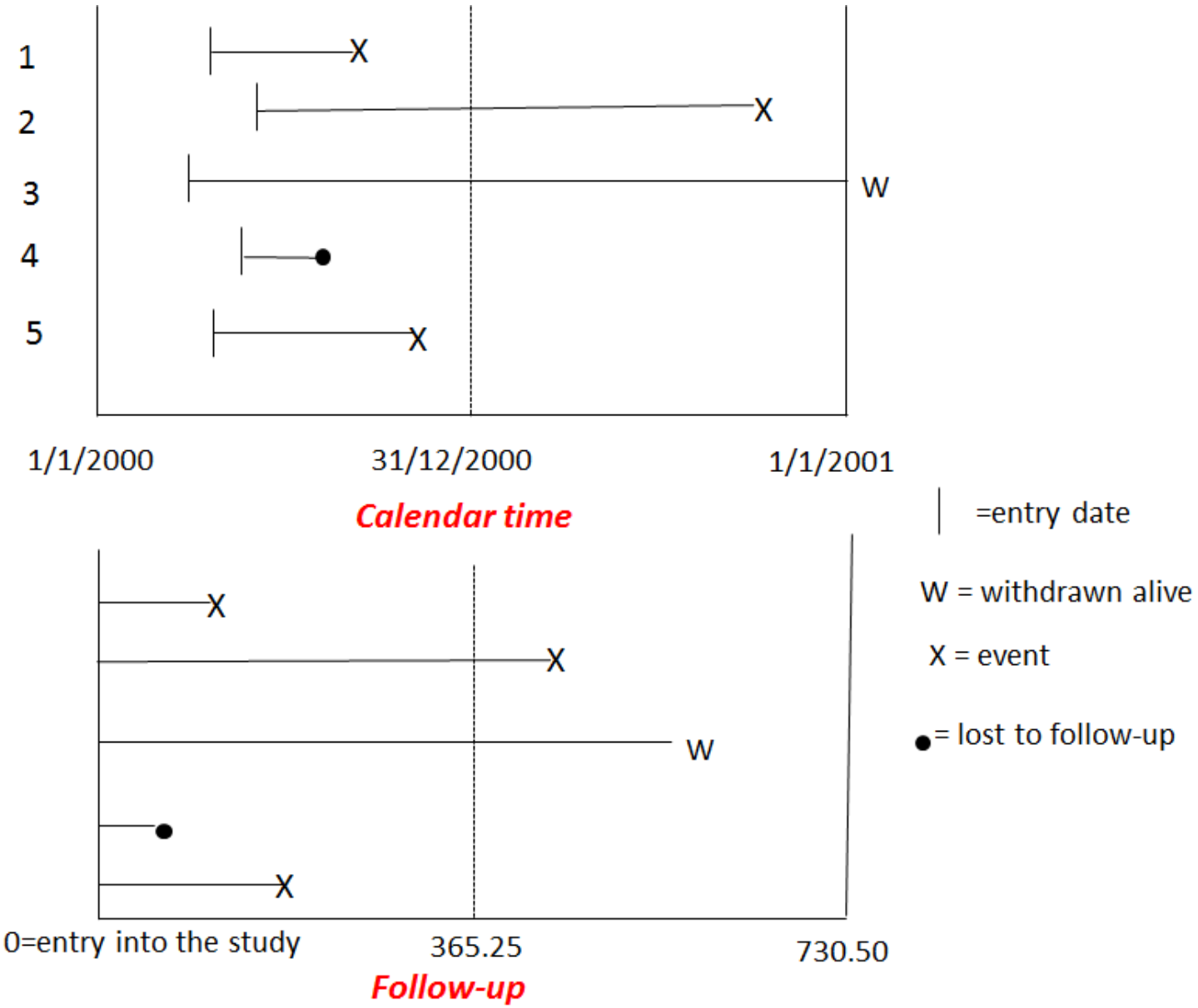
Time to event is a random variable, called failure or survival time, always non-negative: $X \geq 0$

X can either be discrete, i.e. taking a finite set of values $t_1, t_2, \ldots, t_n$ or continuous, i.e. defined on $(0, \infty)$.

To define a failure time, we need of:

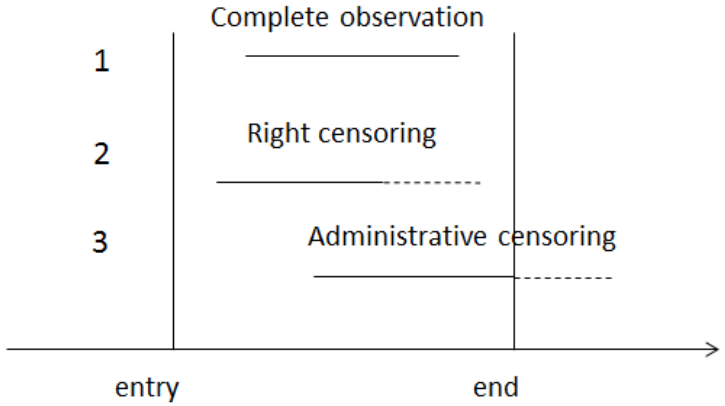
- a well-defined time origin, i.e time when the subject gets at risk of the event of interest (time of randomization, birth, diagnosis time, etc)
- a time scale (follow-up time, calendar time, age, etc)
- an event of interest (disease, recurrence of disease, death, etc)

Time scales


1
2
3
4
5
X
X
W
X
1/1/2000
31/12/2000
1/1/2001
Calendar time
=entry date
W = withdrawn alive
X = event
= lost to follow-up
X
X
W
X
0=entry into the study
365.25
730.50
Follow-up

Censoring

Censoring occurrs when the value of an observation is only partially known



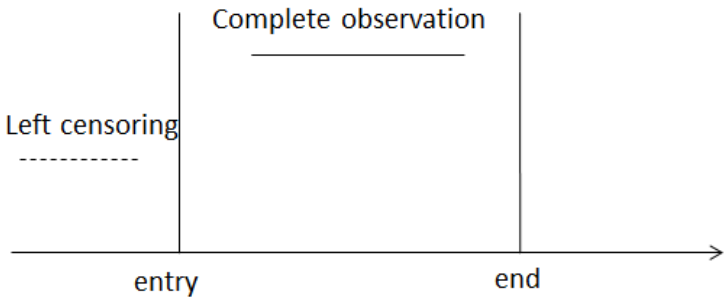
Right censoring: we do not know the time to event (subjects (2) and (3)), we only know the true unobserved time is to the right of censoring time

$T_i = min(X_i, C_i)$ is observed for each subject (X=time to event, C=time to censoring). In addition to T_i, we introduce the random variable δ_i:
$\delta_i = 1$ if the event occurred $(X_i \leq C_i)$
$\delta_i = 0$ otherwise $(X_i > C_i)$

For each subject we observe (T_i, δ_i): for subject (2), (3) $\delta = 0$ and for subject (1) $\delta = 1$

Ex: loss to follow-up, drop-out, study end (administrative censoring)

Left censoring



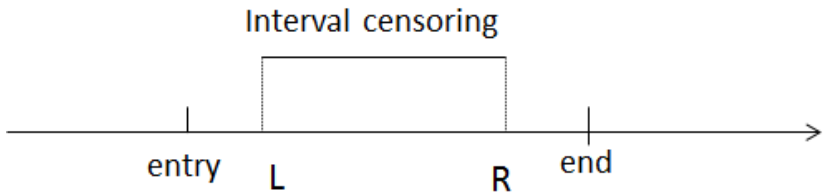
Left censoring: the event of interest has already occurred for the individual before he/she is observed in the study, but we do not know the time

$T_i = max(X_i, C_i)$ is observed for each subject (X=time to event, C=time to censoring)

$\delta_i = 1$ if $C_i \leq X_i$
$\delta_i = 0$ if $C_i > X_i$

Ex: (Miller) study of age at which African children learn a task. Some already knew (left-censored), some learnt during study (exact), some had not yet learnt by end of study (right-censored)

Censoring



Interval censoring: we know the event of interest has occurred in (R,L) but we do not know the exact time in this interval.

$T_i \in (L_i, R_i)$ for subject i

Ex: In Framingham Heart Study age at first diagnosis of coronary heart disease is known exactly. Age at first occurrence of angina pectoris can be known to be between 2 clinical examinations (2 years apart)

Ex: In studies with periodic follow-up: time to undetectable viral load in AIDS studies, based on measurements of viral load taken at each clinic visit; time to recurrence of colon cancer after surgery, where follow-up of patients is every 3 months after resection of primary tumor

Censoring

Is the distribution of censoring times C_i independent of the distribution of times to the event of interest X_i?

The censoring is independent if the censored subject at C_i is rapresentative of all subjects surviving to C_i. If the subject drops out from the study because of a cause associated to X_i, the censoring is dependent

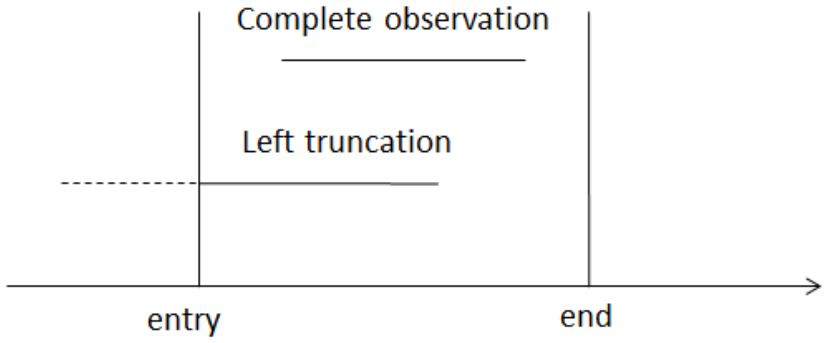
The censored subject can be different from other participants to the study in his/her baseline characteristics, but his/her probability of censoring prior to or at time C_i should be the same among all those with the same baseline characteristics

Distribution of censoring times C_i is usually assumed to be independent of the distribution of times to the event of interest X_i

Censoring

- Administrative censoring $\rightarrow$ independent censoring
- The subject is sicker than the remaining sample in studying the disease progression $\rightarrow$ dependent censoring
- The subject stops the treatment depending on his health status $\rightarrow$ dependent censoring
- Migration: depends on the aim of the study
- Death from other causes: competing events
- ...
- ...

Left truncation



Left truncation: delayed entry into the study

The subject is observed after the time in which he gets at risk of occurrence of event of interest

The investigator only observe the individuals if they are event-free after a certain follow-up time: there can be individuals that had the event but we do not know about them

Ex: In AIDS studies, we usually observe the HIV seroconverters after seroconversion time, and hence we do not know what happened in the elapsed time between seroconversion and entry into the study

Data

The American bone marrow transplantation data

During the period from 1985 and 1991, 1715 patients with acute or chronic leukemia had bone marrow transplanted from a donor who was either an HLA-identical sibling, an HLA-matched unrelated donor, or an HLAmismatched unrelated donor. The data came from more than 80 institutions reporting to the International Bone Marrow Transplantation Registry in Milwaukee, USA. Outcomes of primary interest included relapse and death in remission

Some variables:
disease 10=ALL, 20=AML, 30=CML
timedxtx time from diagnosis to transplant (months)
sex 0=males 1=females
karnofsky =1 if the Karnofsky index is > 90, =0 otherwise
stage 1=early, 2=intermediate, 3=advanced
time time from transplant to event/cens. (months)
donor 1=HLA-identical sibling, 2=HLA-matched unrelated,
3=HLA-mismatched unrelated
event 0=censored, 1=relapse, 2=death in remission

In Stata: stset time, fail(event!=0)

Data

MAC Prevention Clinical Trial

ACTG 196 was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (mycobacterium avium complex), one of the most common opportunistic infections in AIDS patients. The treatment regimens were: clarithromycin (new), rifabutin (standard) and clarithromycin plus rifabutin. Patients enrolled between April 1993 and February 1994 and the follow-up ended August 1995. The main intent-to-treat analysis compared the 3 treatment arms.

Some variables:
ID Subject ID
ENTDATE Entry date
ENDDATE Date follow-up ended due to MAC or censoring
CENSOR Death Indicator (1=death, 0=censor)
AGE Age of subject in years
TREATMENT Therapy

In Stata: stset ENDDATE, fail(CENSOR) enter(ENTDATE)id(ID)

Data

CASCADE: Concerted Action on SeroConversion to AIDS and Death in Europe

CASCADE was established in 1997 as a collaboration between the investigators of European cohorts of people with well-estimated dates of HIV seroconversion known as seroconverters. CASCADE's main aim is to monitor newly-infected individuals and those already enrolled in studies, covering the entire duration of HIV infection. Seroconverters are recruited into the individual cohorts locally and nationally and are typically followed up for life. CASCADE' s aims focus on identification of early HIV infection and research questions requiring knowledge of the time of seroconversion.

Some variables:
ID Subject ID
SERODATE Seroconversion date
ENTDATE Entry date
ENDDATE Date follow-up ended due to death or censoring
CENSOR Death Indicator (1=death, 0=censor)
AGE Age of subject in years
SEX Gender
DRUG History of IV Drug Use (0=no,1=yes)

In Stata:stset ENDDATE, fail(CENSOR) origin(SERODATE) enter(ENTDATE) id(ID)

Definitions

The probability distribution of T can be specified in many ways:

1. probability density function ($f(t)$)
2. survivor function ($S(t)$)
3. hazard function ($\lambda(t)$)
4. cumulative hazard function ($\Lambda(t)$)

Interrelations between these functions are defined for both discrete and continuous distributions

Density function

T continuous:

$$f(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t)}{\Delta t} = \frac{d}{dt} F(t)$$

$f(t)\Delta t$=probability that failure is between t and $t + \Delta t$

T discrete and taking values $t_1 < t_2 < \ldots, j = 1, \ldots$:

$$f(t_j) = P(T = t_j) \qquad j = 1, 2, \ldots$$

Survivor function

T continuous:

$$S(t) = P(T > t) = \int_t^\infty f(u)du$$

T discrete and taking values $t_1 < t_2 < \ldots$

$$S(t) = \sum_{j:t_j > t} f(t_j)$$

- $S(t)$ monotone nonincreasing continuous function
- $S(0) = 1$ and $lim_{t\to\infty} S(t) = 0$
- $S(t) = 1 - F(T)$ with $F(t) = P(T \leq t)$=cumulative distribution function

Hazard function

T continuous:

$$\lambda(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T > t)}{\Delta t}$$

$\lambda(t)\Delta t$:probability that failure is between t and $t + \Delta t$ conditioned on having survived until t

- $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} log(S(t))$
- $S(t) = exp(-\int_0^t \lambda(u)du)$
- $f(t) = \lambda(t)exp(-\int_0^t \lambda(u)du)$

Hazard function

T discrete and taking values $t_1 < t_2 < \ldots$

$$\lambda_j = P(T = t_j | T > t_j) = \frac{f(t_j)}{S(t_j)} \qquad i = 1, 2, \ldots$$

- $S(t) = P(T > t_1, T > t_2, T > t_{j+1}) =$

  $P(T > t_1)P(T > t_2 | T > t_1) \ldots P(T > t_{j+1} | T > t_j) =$

  $\prod_{j:t_j<t}(1 - \lambda_j),$ $\qquad t_j < t \leq t_{j+1}$

- $f(t_j) = \lambda_j \prod_1^{j-1}(1 - \lambda_j)$

Cumulative hazard function

T continuous:

$$\Lambda(t) = \int_0^t \lambda(u)du$$

- $\Lambda(t) = \int_0^t -\frac{d}{du} log(S(u))du = -logS(t) + logS(0) = -logS(t)$
- $S(t) = exp(-\Lambda(t))$

T discrete and taking values $t_1 < t_2 < \ldots$

$$\Lambda_j = \sum_{j:t_j<t} \lambda_j$$

Estimation of the survivor function

Survival function: $S(t) = P(T > t)$

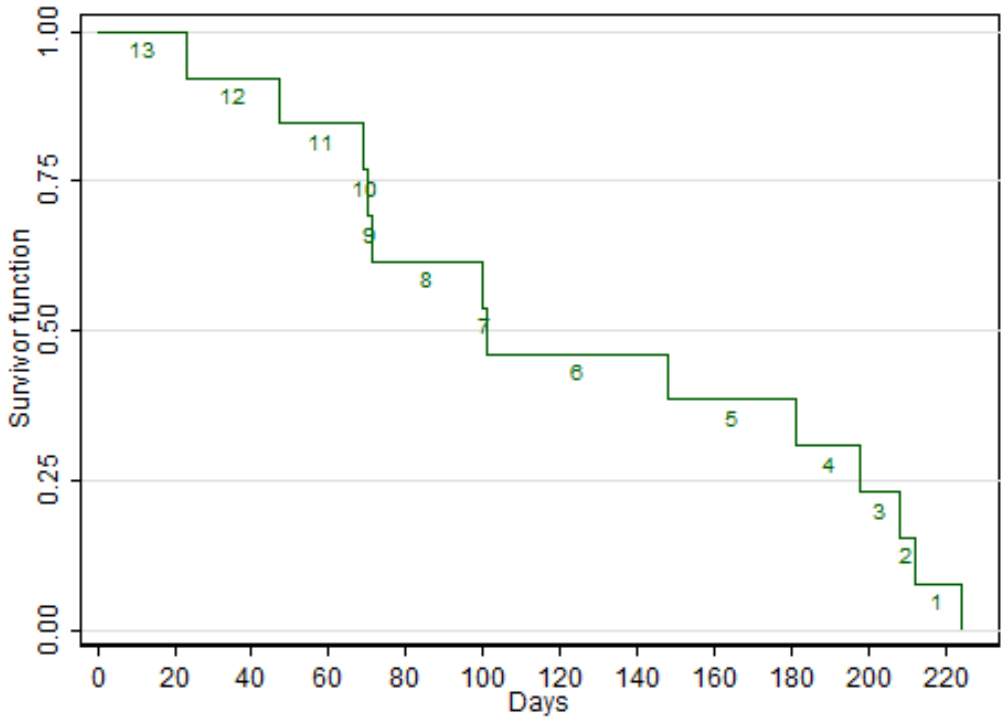
Non-parametric methods do not make any assumptions about the distribution of the process (suited for first exploratory data analyses)

1. Kaplan-Meier method

2. Life table

In absence of censoring:

The following are times to death (days) for 13 women affected by breast cancer:
23 47 69 70 71 100 101 148 181 198 208 212 224



If an uncensored sample of n distinct failure times is observed:

$$S_n(t) = \frac{no.ofsamplevalues > t}{n}$$

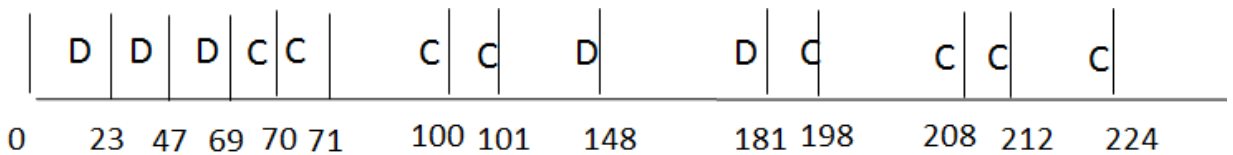
$S_n(t)$ is a step function decreasing by n^{-1} immediately following each observed failure time

In presence of censoring:

The following are times to death (days) for 13 women affected by breast cancer:
$23\ 47\ 69\ 70^+\ 71^+\ 100^+\ 101^+\ 148\ 181\ 198^+\ 208^+\ 212^+\ 224^+$
($^+$ indicating right censoring)

S(50)=13-2/13=0.85
S(80)=? (we do not know what happens to individuals censored at 70 and 71 days: observed censoring time tells us that the unobserved failure time is greater than 70 and 71 respectively)

Intuitively suppose to split the observed timespan of the study into intervals defined by the failures/censoring times:



$P(T > 80) = P(T > 23)P(T > 47|T > 23)P(T > 69|T > 47)P(T > 70|T > 69)P(T > 70|T > 69)P(T > 71|T > 70) = \frac{13-1}{13}\frac{12-1}{12}\frac{11-1}{11}\frac{10-0}{10}\frac{9-0}{9}\frac{8-0}{8} = 0.77$

Kaplan-Meier estimator

Product of conditional probabilities:

Let be $t_j < t < t_{j+1}$. Then

$$P(T > t) = P(T > t_j) = P(T > t_1, T > t_2, \ldots, T > t_j) =$$

$$= P(T > t_1) \prod_{j=2}^{m} P(T > t_j | T > t_{j-1}) = \prod_{j=1}^{m} (1 - P(T = t_j | T > t_{j-1})) =$$

$$= \prod_{j=1}^{m} (1 - \lambda_j) = \prod_{j=1}^{m} (1 - \frac{d_j}{n_j})$$

d_j=number of failures at time t_j
n_j=number of subjects at risk just prior of time t_j

Kaplan-Meier estimator

Likelihood estimator:

Suppose that d_j items fail at $t_j (j = 1, \ldots, m)$ and let n_j be the number of items at risk at a time just prior to t_j.

Being λ_j the probability of failure in the jth interval conditional on survival at the start of the interval, the likelihood is given by the product of independent binomials:

$$L(\lambda) = \prod_{j=1}^{m} [(\lambda_j)^{d_j} (1 - \lambda_j)^{n_j - d_j}]$$

$$\widehat{\lambda}_j = \frac{d_j}{n_j}$$

$$\widehat{S}(t) = \prod_{j:t_j<t} (1 - \widehat{\lambda}_j) = \prod_{j:t_j<t} (1 - \frac{d_j}{n_j})$$

Kaplan-Meier estimator

Let $t_1 < t_2 < \ldots < t_m$ represent the observed failure times in a sample of size n from an homogeneous population with survivor function $S(t)$

$$\widehat{S}(t) = \prod_{j:t_j<t} \left(\frac{n_j - d_j}{n_j}\right) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right)$$

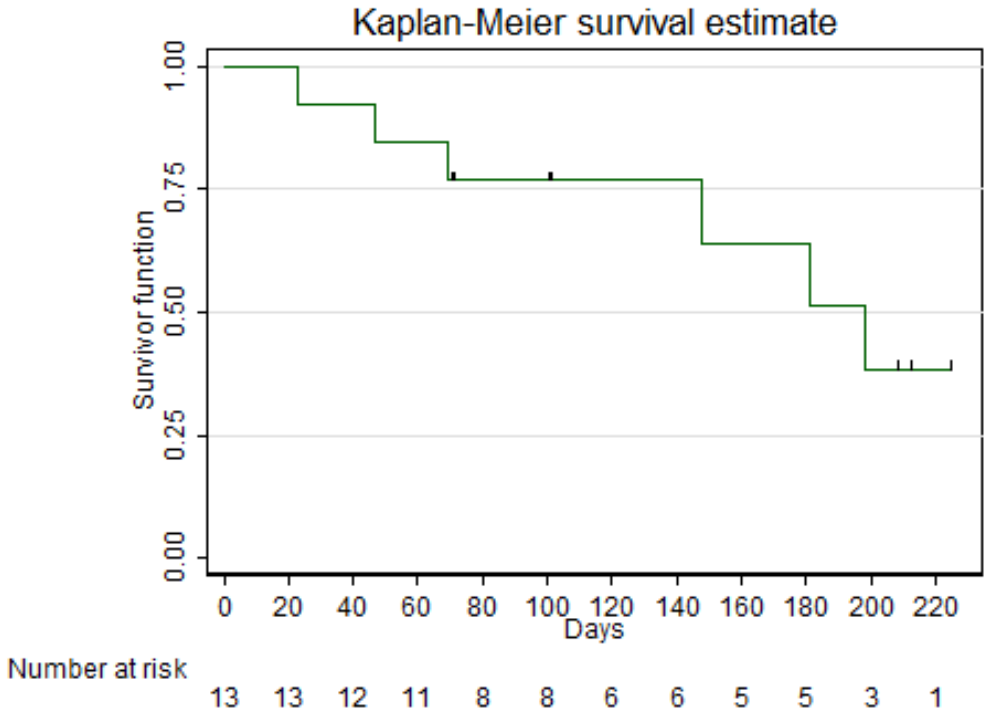
- d_j=number of failures at time t_j
- n_j=number of items at risk just prior of time t_j, including failures and censored items at or after t_j
- c_j=number of censored items between t_j and t_{j+1}
- $n_j = n_{j-1} - d_{j-1} - c_{j-1}$
- $n_j = \sum_{l \geq j}(c_l + d_l)$

Example

Survival of 13 women affected by breast cancer

t_j	n_j	d_j	c_j	d_j/n_j	$1-d_j/n_j$	$S(t_j)$
23	13	1	0	1/13	12/13	12/13
47	12	1	0	1/12	11/12	12/13*11/12
69	11	1	0	1/11	10/11	11/13*10/11
70	10	0	1	0	10/10	10/13*1
71	9	0	1			
100	8	0	1			
101	7	0	1			
148	6	1	0	1/6	5/6	10/13*5/6
181	5	1	0	1/5	4/5	10/13*5/6*4/5
198	4	0	1			
208	3	0	1			
212	2	0	1			
224	1	0	1			

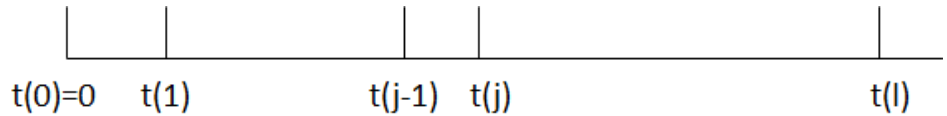
Kaplan-Meier estimate



- $\widehat{S}(t)$ is a step function with steps corresponding to failure times
- $\widehat{S}(t)$ is right continuous: $\widehat{S}(t) = \widehat{S}(t_+)$
- censoring influences only the height of the steps, depending on the risk set (denominator)
- $\widehat{S}(t)$ goes to 0 only if the last observed event is a failure

Life table

Life tables: traditional procedure applied to grouped survival data, i.e. the time interval when an event occurrs is known but the exact time is unknown



- Time axis is partitioned into fixed intervals $[t_{j-1}, t_j), j = 1, \ldots, l$
- d_j=number of failures during j-th interval $[t_{j-1}, t_j)$
- n_j=number of items at risk entering into the j-th interval
- c_j=number of censored items during j-th interval
- $n_j^{'}$=number of items at risk during j-th interval

$n_1 = n,\ n_j = n_{j-1} - d_{j-1} - c_{j-1}$

$n_j^{'} = n_j - wc_j, \qquad 0 \leq w \leq 1,$ ($w = 1/2$ censorings assumed to occur uniformly throughout the interval)

Life table

$q_j = \frac{d_j}{n'_j}$=conditional probability for having an event in $[t_{j-1}, t_j)$

$p_j = 1 - q_j$=conditional probability for surviving in $[t_{j-1}, t_j)$

$\widehat{S}_j = S(t_{j-1})p_{j-1} = \prod_{l:1}^{j-1} p_l = \prod_{l:1}^{j-1}(1 - \frac{d_l}{n'_l}) = \prod_{l:1}^{j-1}(1 - \frac{d_l}{n_l - wc_l})$

$\widehat{f}_j = \frac{S(t_j) - S(t_{j+1})}{t_{j+1} - t_j}$, calculated at middle point of interval

$\widehat{\lambda}_j = \frac{f(t_j)}{\frac{S(t_j) + S(t_{j+1})}{2}} = \frac{1}{t_{j+1} - t_j} \frac{d_j}{n'_j - d_j/2}$, calculated at middle point of interval

$\widehat{\lambda}_j$=number of failures in I_j/total count of person time of observation in I_j:

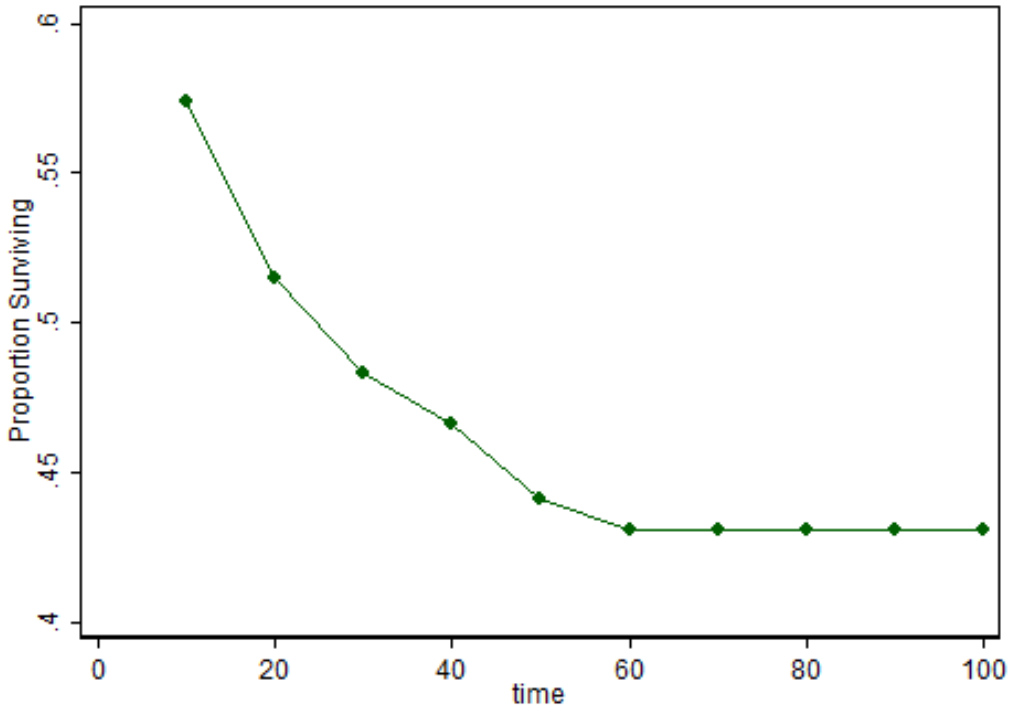
- it allows for the rate to change from an interval to another one
- it assumes a constant rate in each interval

Example

The American bone marrow transplantation data

t_j	l_j	n_j	d_j	c_j	n_j'	q_j	p_j	$S(t_j)$
0	[0,1)	1715	705	123	1653.5	0.43	0.57	0.57
1	[1,2)	887	87	86	844	0.10	0.90	0.51
2	[2,3)	714	40	128	650	0.06	0.94	0.48
3	[3,4)	546	16	175	458.5	0.03	0.97	0.47
4	[4,5)	355	16	117	296.5	0.05	0.95	0.44
5	[5,6)	222	4	93	175.5	0.02	0.98	0.43
6	[6,7)	125	0	58	96	0.00	1.00	0.43
7	[7,8)	67	0	49	42.5	0.00	1.00	0.43
8	[8,9)	18	0	14	11	0.00	1.00	0.43
9	[9,10)	4	0	4	2	0.00	1.00	0.43

Example



- Usually the first interval starts with $t_0 = 0$
- Stata estimates the survival function at the right-hand endpoint of each interval

Life table

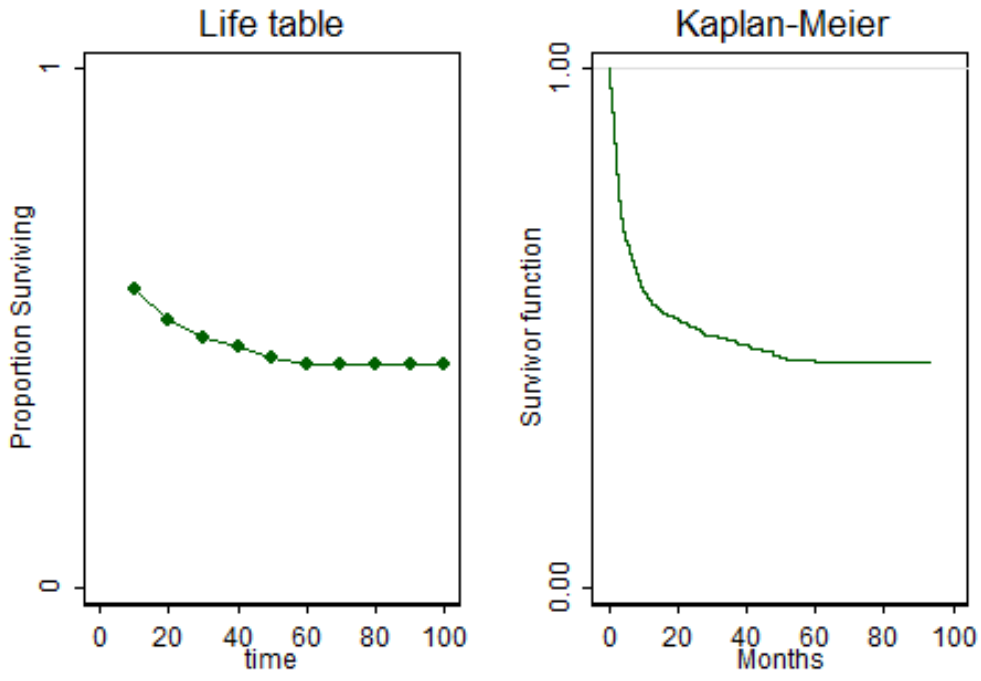
Life tables are used with grouped data or in the case of large dataset because it needs less computing time and space, but

1. it is necessary to group the durations into fixed intervals $\rightarrow$ the results depend more or less on these arbitrarily defined time-intervals
2. it is needed to observe a relatively large number of events, so that estimates conditional for each interval are reliable

Kaplan-Meier method is preferred to life table with continuous data because it does not require to define intervals, but it is based on the calculation of a risk set at every point in time where at least an event occurred

Example

ltable timevar statusvar [freq=freqvar], by(varlist) intervals(width/numlist))
stset time,failure(status)
sts list, by(treat)
sts graph, by(treat) lost



Properties of K-M estimator

In absence of censoring: $\widehat{S}_n(t) = \frac{no.ofsamplevalues > t}{n}$

Since it is an estimated probability from a binomial distribution, for large values of n:

$\widehat{S}_n(t) \sim N(S(t), S(t)(1 - S(t))/n$

In presence of censoring:

- $\widehat{S}_{KM}(t)$ is approximately normal
- The mean of $\widehat{S}_{KM}(t)$ converges to the true $S(t)$
- The variance can be estimated by Greenwood formula
- $(1 - \alpha)\%$ confidence interval can be estimated by:
  $(\widehat{S}_{KM}(t) - z_{1-\alpha/2} se(\widehat{S}_{KM}(t), \widehat{S}_{KM}(t) + z_{1-\alpha/2} se(\widehat{S}_{KM}(t)))$

Greenwood formula

$$\widehat{S}(t) = \prod_{j:t_j<t} (1 - \widehat{\lambda} j) = \prod_{j:t_j<t} (1 - \frac{d_j}{n_j})$$

Since the $\widehat{\lambda} j$ are binomial proportions, $\widehat{\lambda} j$ is approximately normal, with mean the true λ_j, and $var(\widehat{\lambda} j) = \widehat{\lambda} j(1 - \widehat{\lambda} j)/n_j$ and $\widehat{\lambda} j$ are independent for large samples

By applying the Delta method (if Y is normal with mean μ and variance σ^2, then $g(Y)$ is approximately normally distributed with mean $g(\mu)$ and variance $(g^{'}(\mu))^2\sigma^2)$):

$$Var(log(\widehat{S}(t))) = \sum_{j:t_j<t} Var(log(1 - \widehat{\lambda} j)) = \sum_{j:t_j<t} \frac{1}{1 - \widehat{\lambda} j}^2 var(\widehat{\lambda} j) =$$

$$\sum_{j:t_j<t} \frac{1}{1 - \widehat{\lambda} j}^2 \widehat{\lambda} j \frac{1 - \widehat{\lambda} j}{n_j} = \sum_{j:t_j<t} \frac{\widehat{\lambda} j}{(1 - \widehat{\lambda} j)n_j} = \sum_{j:t_j<t} \frac{d_j}{(n_j - d_j)n_j}$$

Greenwood formula

By Delta method: $var(\widehat{S}(t)) = var(exp(log(\widehat{S}(t)))) = \widehat{S}^2(t) var(log(\widehat{S}(t)))$

Greenwood formula: $var(\widehat{S}(t)) = \widehat{S}^2(t) \sum_{j:t_j<t} \frac{d_j}{(n_j - d_j)n_j}$

The precision of the survival estimate tends to decrease as the number of subjects at risk decreases (and the variance increase): tails of survival curve are unstable

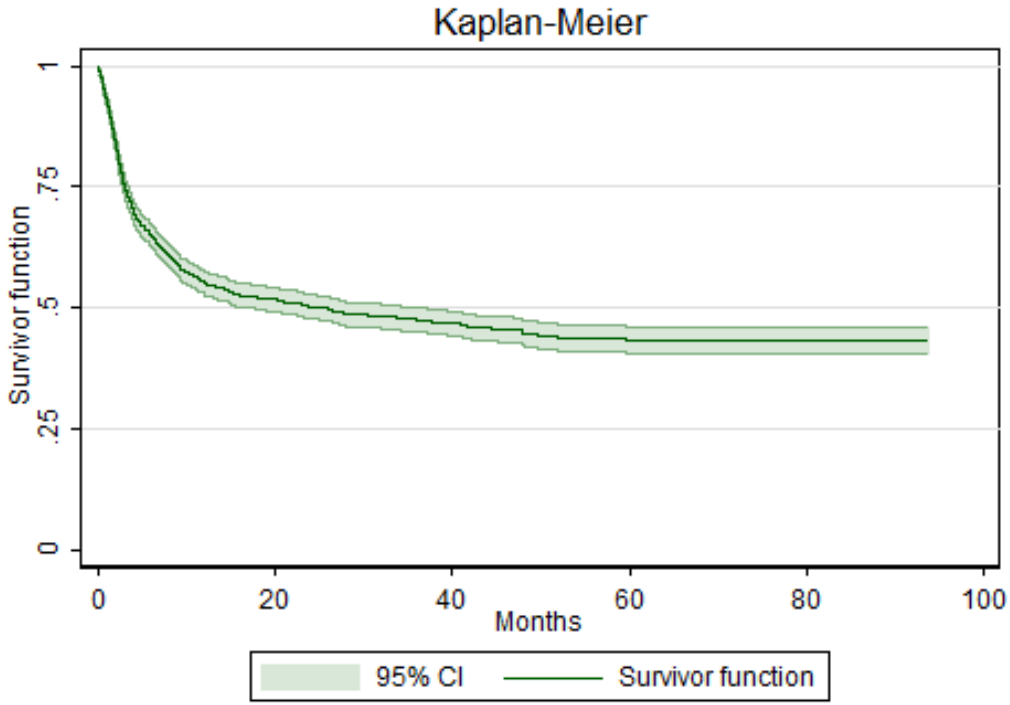
$(1-\alpha)\%$ confidence interval estimated by using standard error obtained by Greenwood formula can include values out of range $[0,1]$. Hence $(1-\alpha)\%$ confidence interval is usually calculated for $log(-log(S(t))$ and then it is transformed back: $S(t) = exp(-exp(log(-logS(t))))$

$$var(log(-log(\widehat{S}(t)))) = \frac{1}{(log(\widehat{S}(t)))}^2 \sum_{j:t_j<t} \frac{d_j}{(n_j - d_j)n_j}$$

Greenwood formula

The American bone marrow transplantation data

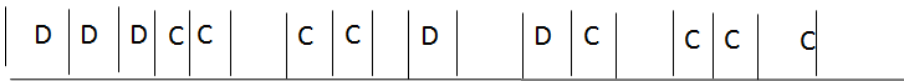
sts graph,ci



Nelson-AAlen estimator

$$\Lambda(t) = \int_0^t \lambda(u)du$$

The observed timespan of the study can be splitted into a series of intervals (of width Δ) so that there is only one event per interval:



$$\widehat{\Lambda}(t) = \sum_{j:t_j<t} \lambda_j \Delta$$

Since $\lambda_j \Delta$ is approximately the probability of dying in the j-th interval:

Nelson-AAlen estimator: $\widehat{\Lambda}_{NA}(t) = \sum_{j:t_j<t} \frac{d_j}{n_j}$

Fleming-Harrington estimator: $\widehat{S}_{FH}(t) = exp(-\widehat{\Lambda}_{NA}(t))$
$\widehat{S}_{FH}(t)$ is close to $\widehat{S}_{KM}(t)$ as well as $\widehat{\Lambda}_{NA}(t)$ is close to $\widehat{\Lambda}_{KM}(t) = -log(\widehat{S}_{KM}(t))$

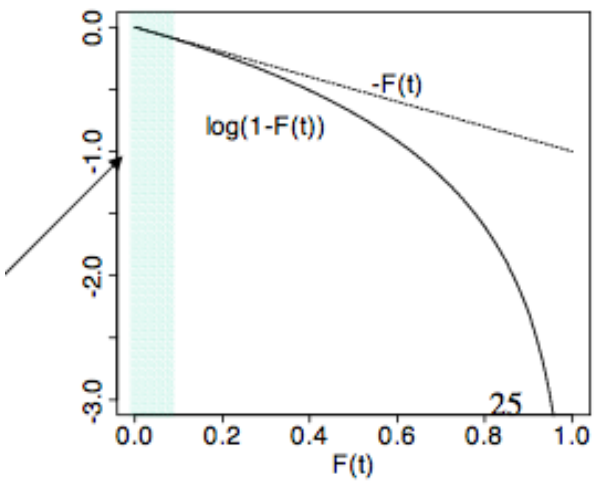
Nelson-AAlen estimator

Cumulative hazard estimate is not equivalent to cumulative probability of death $F(t) = 1 - S(t)$

$\Lambda(t) = -log(S(t) = -log(1 - F(t)) \neq F(t)$

$\Lambda(t)$ and $F(t)$ have similar values when $F(t)$ is very small:

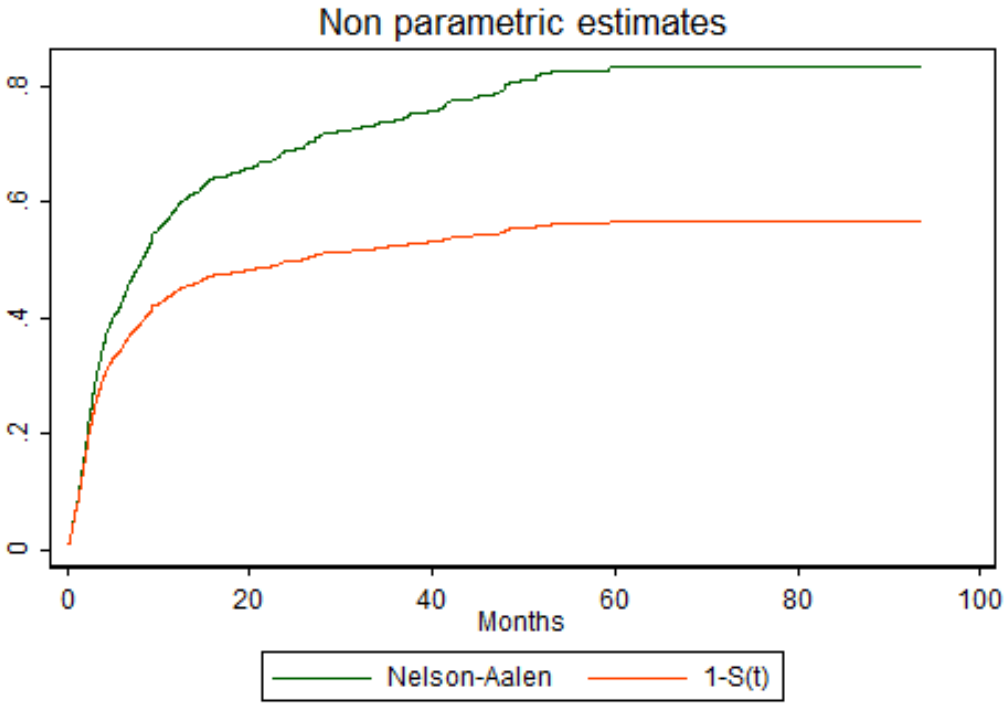
$-\Lambda(t) = log(1 - F(t)) \approx -F(t)$



Nelson-Aalen estimator

The American bone marrow transplantation data

sts graph,na



Summary

In reporting a survival curve:

- definition of entry point and end-point
- dates of start and end recruitment
- when follow-up was up-dated
- the percentage of censoring, specifying the percentage of lost-to-follow-up

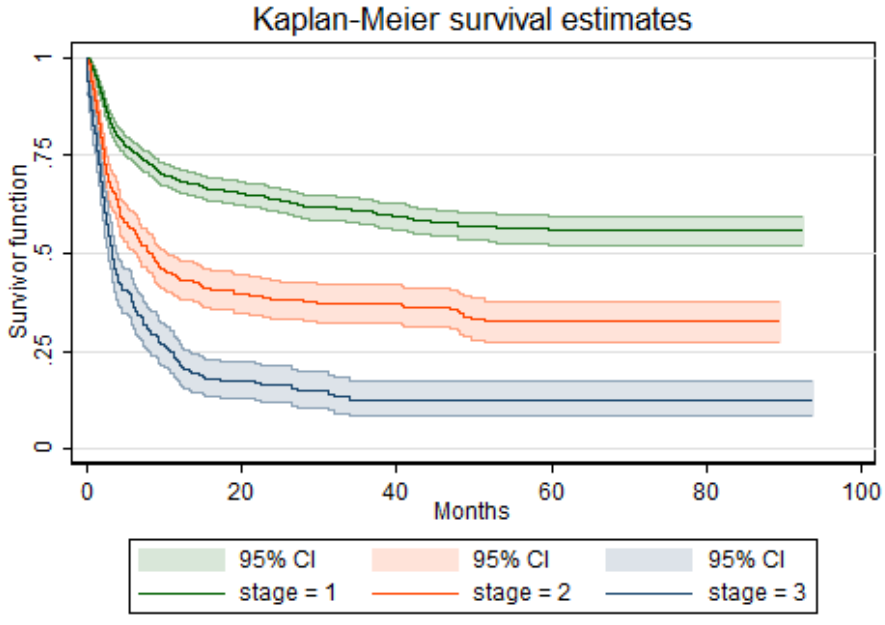
These informations allow the reader to understand the maturity of data (recruitment and minimum potential follow-up) and the quality of data

Summary

In reading a survival curve:

- Observe the shape of the curve more than details
- Useful refer to estimated survival value (and standard error) at important points in time
- Not consider the curve when there are less then 10-20 subjects at risk left
- Keep in mind that we assumed that censored subject would not have a survival experience different from the others

Comparison between survival curves



- How can we compare individuals diagnosed at different stages in terms of survival (to death or remission after transplant?)
- How can we compare individuals with different treatment or with different clinical/biological characteristics?
- Are the single time-point or the overlap between confidence bands an appropriate measures?

Comparison between survival curves

- Comparison of a single time point is not efficient and mainly it is based on an arbitrary choice (how much unstable are the tails of distribution?)

- Overlap of confidence intervals is not appropriate for an overall comparison because they correspond to confidence intervals for $\widehat{S}(t)$ at a single time point t (we cannot say that the true survival function $S(t)$ is contained between the pointwise confidence intervals with 95% confidence):

$$H_0 : S_1(t) = S_2(t), \qquad H_1 : S_1(t) \neq S_2(t)$$

$$Z = \frac{\widehat{S}_1(t) - \widehat{S}_2(t)}{se(\widehat{S}_1(t)) + se(\widehat{S}_2(t))}$$

Under $H0$, $Z \sim N(0,1)$. This test has to be repeated for each $t \rightarrow$ multiple tests

Comparison between survival curves

Non-parametric tests to compare survival times

1. test based on the distribution of the maximum observed difference (ex. Kolmogorov-Smirnov type)
2. test based on permutations (depends on the censoring distribution in a complex way)
3. test based on the median survival (often precision of the estimates is low)
4. rank test

Tests based on parametric assumptions to compare survival times

Log-rank test

- Consider two treatment groups A and B

  $H_0 : S_A(t) = S_B(t),\ H_1 : S_A(t) = (S_B(t))^{\theta}$

  If $0 < \theta < 1$, $S_A(t) > S_B(t)$, if $\theta > 1$, $S_A(t) < S_B(t)$, if $\theta = 1$, $S_A(t) = S_B(t)$
- Order the distinct failure times observed in the two groups in ascending order
- At each $t_{(j)}$ consider a 2x2 contingency table

t(j)

	A	B	
death	djA	djB	dj
alive	njA-djA	njB-djB	nj-dj
	njA	njB	

Log-rank test

- Consider the distribution of the observed cell frequencies conditional on the observed marginal totals under the null hypothesis $H_0 : \lambda_A(t) = \lambda_B(t)$ for each t: if the margins of the table are considered fixed, then d_j follows a hypergeometric distribution
- Generate a 2x2 contingency table of expected under H_0:

t(j)

	A	B	
death	njA*dj/nj	njB*dj/nj	dj
alive			nj-dj
	njA	njB	nj

Log-rank test

Observed deaths in A: $O(d_{jA}) = O_{jA} = d_{jA}$
Expected deaths in A: $E(d_{jA}) = E_{jA} = \frac{d_j n_{jA}}{n_j}$

Observed deaths in B: $O(d_{jB}) = O_{jB} = d_{jB}$
Expected deaths in B: $E(d_{jB}) = E_{jB} = \frac{d_j n_{jB}}{n_j}$

The statistic test is built on:

$$O_{jA} - E_{jA}$$

$$Var(d_{jA}) = \left[n_{jA} \frac{d_j}{n_j} \left(1 - \frac{d_j}{n_j} \right) \right] \left[\frac{n_j - n_{jA}}{n_j - 1} \right]$$

Log-rank test

Considering all J times with at least one event, the statistic is:

$$Q_{MH} = \frac{\left[\sum_j (d_{jA} - E(d_{jA}))\right]^2}{\sum_j Var(d_{jA})}$$

equivalently

$$Q_{MH} = \frac{\left[O_A - E_A\right]^2}{Var(O_A)}$$

$O_A = \sum_j O_{jA}, E_A = \sum_j E_{jA}$

- under $H0$, Q_{MH} is asymptotically distributed as a χ^2_1
- The higher is Q_{MH} the smaller is the probability that the sample is consistent with $H0$

Example

Leukemia example (Cox & Oakes)

Time of failure t(j)	Events			Patients at risk			*E(dAj)*	*Var(dAj)*
weeks	*dAj*	*dBj*	*dj*	*nAj*	*nBj*	*nj*		
1	0	2	2	21	21	42	1.00	0.49
2	0	2	2	21	19	40	1.05	0.49
3	0	1	1	21	17	38	0.55	0.25
4	0	2	2	21	16	37	1.14	0.48
5	0	2	2	21	14	35	1.20	0.47
6	3	0	3	21	12	33	1.91	0.65

Example

Time of failure t(j)	Events			Patients at risk			*E(dAj)*	*Var(dAj)*
weeks	*dAj*	*dBj*	*dj*	*nAj*	*nBj*	*nj*		
1	0	2	2	21	21	42	1.00	0.49
2	0	2	2	21	19	40	1.05	0.49
3	0	1	1	21	17	38	0.55	0.25
4	0	2	2	21	16	37	1.14	0.48
5	0	2	2	21	14	35	1.20	0.47
6	3	0	3	21	12	33	1.91	0.65
.								
.								
.								
.								
Total	**9**	**21**	**30**				**19.25**	**6.26**

$$Q_{MH} = \frac{(9 - 19.25)^2}{6.26} = 16.79$$

A class of rank test

$$Q = \frac{\left[\sum_j w_j(d_{jA} - E(d_{jA}))\right]^2}{\sum_j w_j^2 Var(d_{jA})}$$

Test	Weight
Logrank (Mantel-Haenzel 1959)	$w_j = 1$
Gehan's Wilcoxon (Gehan 1965)	$w_j = n_j$
Peto/Prentice (Prentice 1978)	$w_j = n\widehat{S}(t_j)$
Tarone-Ware (1977)	$w_j = \sqrt{n_j}$
Fleming-Harrington (1982)	$w_j = \widehat{S}(t_j)^\alpha$

A more general class is in Harrington-Fleming (1982) with weights:

$$w_{jpq} = \widehat{S}(t_{j-1})^p(1 - \widehat{S}(t_{j-1}))^q, p, q \geq 0$$

$p = q = 0 \rightarrow$ log-rank test
$p > 0, q = 0 \rightarrow$ weight to early differences
$p = 0, q > 0 \rightarrow$ weight to late differences

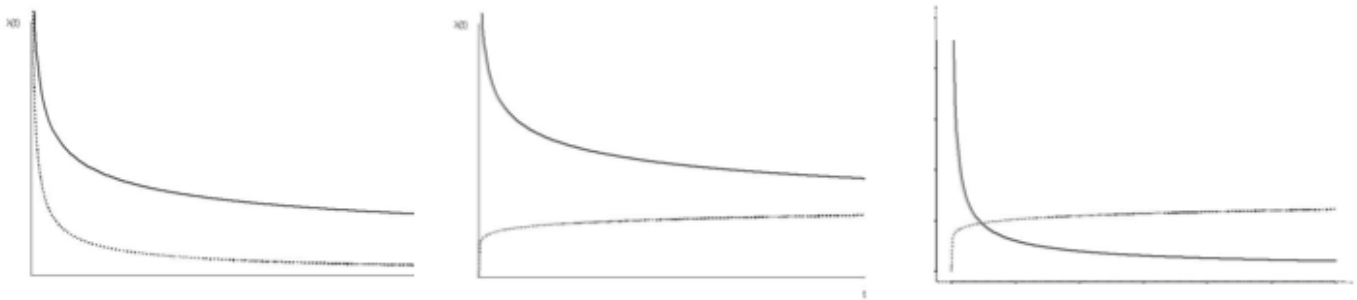
Test's choice

$$H_0 : S_A(t) = S_B(t) \qquad H_1 : S_A(t) = S_B^{\theta(t)}(t)$$

or equivalently

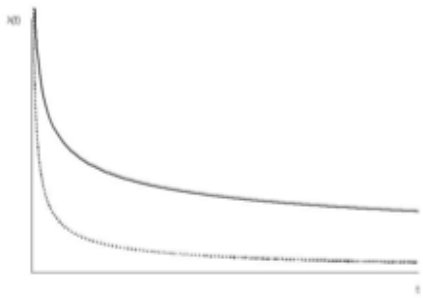
$$H_0 : \lambda_A(t) = \lambda_B(t) \qquad H_1 : \lambda_A(t) = \theta(t)\lambda_B(t)$$

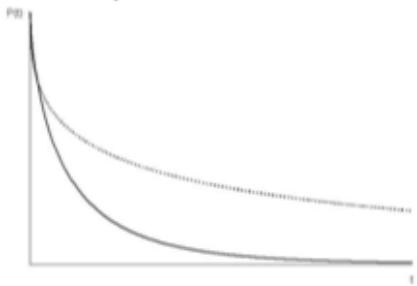
$$\theta(t) \neq 1, \theta(t) > 0$$



The choice of which test has to be done "a priori", depending on the alternative hypothesis in order to increase the power of the test

Test's choice



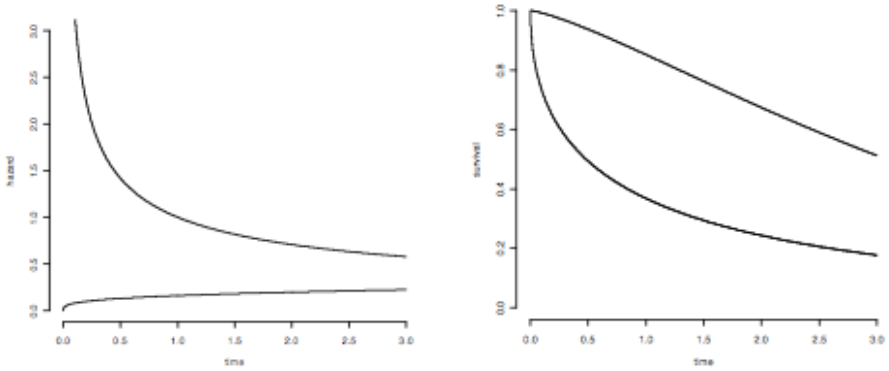


$$H_0 : S_A(t) = S_B(t) \qquad H_1 : S_A(t) = S_B^{\theta(t)}(t)$$

$$\theta(t) = \theta$$

- The log-rank test gives equal weight to all time points
- The log-rank test is most powerful under the assumption of proportional hazards (the ratio of hazard functions is the same at all time points)

Test's choice

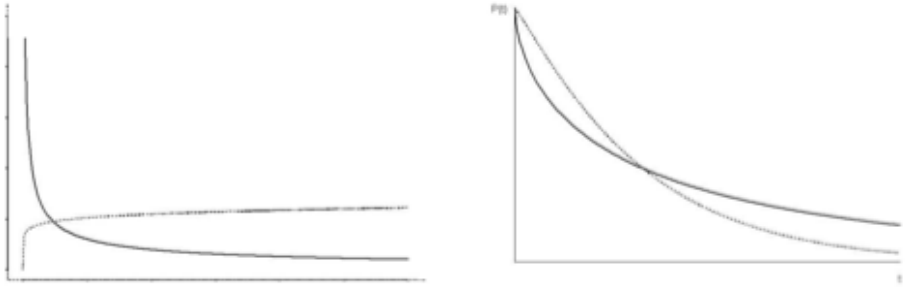


$$H_0 : S_A(t) = S_B(t) \qquad H_1 : S_A(t) = S_B^{\theta(t)}(t)$$

$$\theta(t_{(1)}) > \theta(t_{(2)}) > \theta(t_{(3)}) > \dots$$

- The Gehan-Breslow-Wilcoxon test gives more weight to deaths at early time points (it is sensitive to early differences between survival)
- The Gehan-Breslow-Wilcoxon test has high power when the failure times are lognormally distributed, with equal variance in both groups but a different mean (accelerated failure time model): it has most power when $\theta(t)$ is bigger for smaller t's.
- The Tarone-Ware e Prentice have the same property of the Gehan-Breslow-Wilcoxon test but they give a smaller weight to the beginning of follow-up time.

Test's choice

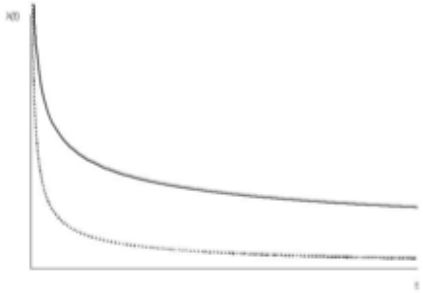


$$H_0 : S_A(t) = S_B(t) \qquad H_1 : S_A(t) = S_B^{\theta(t)}(t)$$

$$\theta(t) > 1 \qquad t < \tau, \theta(t) < 1 \qquad t > \tau$$

- If the two survival curves cross, then one group has a higher risk at early time points and the other group has a higher risk at late time points. Neither the log-rank nor the Wilcoxon-Gehan tests are helpful when the survival curves cross near the middle of the time course (however this could just be a coincidence of random sampling, and the assumption of proportional hazards could still be valid)
- There is not any global test in the class with sufficient power against an alternative hypothesis of crossing hazards

Sample size for log-rank test



$$H_0 : S_A(t) = S_B(t) \qquad H_1 : S_A(t) \neq S_B(t)$$

The difference between the two groups is expressed by the hazard ratio:
$\phi = \frac{\lambda_A(t)}{\lambda_B(t)}$ or $\phi = \frac{lnS_A(t)}{lnS_B(t)}$

To detect a difference ϕ with power $1 - \beta$, by a test at a significance level α, we need a total number of events equal to

Freedman (1982): $d = (z_{\alpha/2} + z_\beta)^2 \left(\frac{1+\phi}{1-\phi}\right)^2$

Schoenfeld (1981): if ϕ is about 1 $d = 4\left(\frac{(z_{\alpha/2}+z_\beta)}{-log\phi}\right)^2$

Example

Trial on ovarian cancer at stage I of chemotherapy after surgery versus surgery alone (control)

The expected survival probabilities at 5 years are:

$$S_C(5) = 60\%, S_T(5) = 75\%$$

$$\phi = log(0.75)/log(0.60) = 0.56$$

If $\alpha = 0.05$ and $1 - \beta = 0.80$:

$$d = (1.96 + 0.84)^2 \left(\frac{1 + 0.56}{1 - 0.56}\right)^2 \approx 100$$

Total sample size:

$$n = \frac{d}{1 - 1/2(S_C(t) + S_T(t))} = \frac{100}{1 - 1/2(0.60 + 0.75)} \approx 308$$

Extension of log-rank test

1. Comparison between two groups, by stratifying on prognostic factors or confounding factors
2. Comparison between two groups in presence of subjects changing group in time

Stratified log-rank test

The aim of the stratified analysis is to adjust for imbalances on important prognostic variables

Male patients usually have a worse prognosis then females for a given disease. If we want to compare the effect of treatments A and B we should consider the gender composition of the 2 groups treated with A and B. If it is different the simple test is not correct, because the result will be influenced not only by the treatment effect but also by the gender effect

The test statistic is the same as the log-rank but is within strata of the prognostic variable, so that the comparison is within homogenous groups, then an average measure, suitably weighted between strata, of the relative effect of the 2 treatments is obtained.

Stratified log-rank test

Females

	A	B	
death	djAf	djBf	djf
alive	njAf-djAf	njBf-djBf	njf-djf
	njAf	njBf	njf

Males

	A	B	
death	djAm	djBm	djm
alive	njAm-djAm	njBm-djBm	njm-djm
	njAm	njBm	njm

Consider a factor with S levels, on which to stratify the log-rank test

Within group s, build Js 2x2 tables corresponding to distinct failure times in strata s as the strata are independent

Stratified log-rank test:

$$Q^*_{MH} = \frac{\left\{ \sum_{s=1}^{S} \left[\sum_{j=1}^{Js} (d_{jAs} - E(d_{jAs})) \right] \right\}^2}{\sum_{s=1}^{S} \left(\sum_{j=1}^{Js} Var(d_{jAs}) \right)}$$

Time-dependent covariates

Comparison of 2 groups A and B when a subject can change group at the occurrence in time of a defined event

All subjects entered into the study as non responders: non responder to treatment(A) $\rightarrow$ responder to treatment (B)

A subject is transplanted after waiting time in which a suitable donor is made available: non transplanted status *vs* transplanted status

A seroconverter starts a specific treatment after reaching a well-defined viral load: untreated *vs* treated

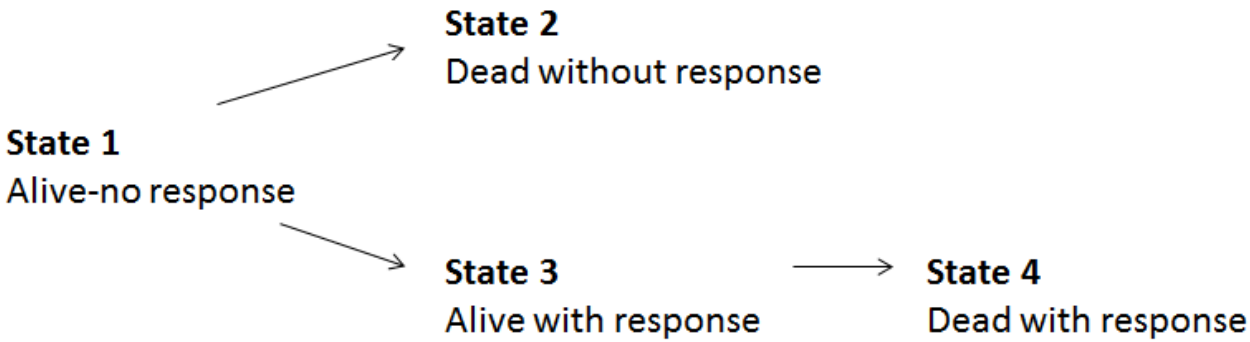
The classification in group A and B is time-dependent $\rightarrow$ the log-rank test applied according to the classification of subjects in group A and B at entry into the study is not correct

Example

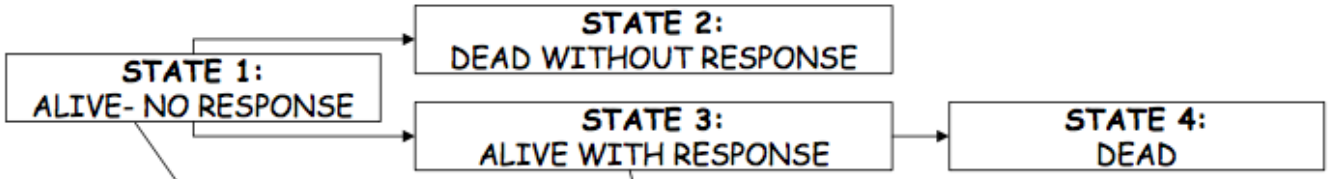
Anderson et al. 1983: Results of a II phase clinical trial on patients with a multiple myeloma (n=35), with response to treatment evaluated in time

Comparison between survival in responders and not responders

At the end of observation there were 10 non responders and 25 responders and the response had been achieved in a 3 month range



Example



Time (gg) from treatment start	"non respondent" States 1, 2		"respondent" States 3 e 4		$E(d_{Aj})$	$Var(d_{Aj})$
	n_{Aj}	d_{Aj}	n_{Bj}	d_{Bj}		
8	**33**	**1**	**2**	**0**	**0.94**	**0.054**
9	31	1	3	0	0.91	0.080
11	30	1	3	0	0.91	0.083
23	**21**	**0**	**11**	**1**	**0.66**	**0.226**
30	19	1	12	0	0.61	0.237
34	13	0	17	1	0.43	0.246
...	...	...	...	...	...	...
Totals		9		20	8.56	4.028

Mantel Byar test: $Q_{M-B} = (9 - 8.56)^2/4.028 = 0.05$

Bibliography

- Cox, D. and Oakes, D. (1984). Analysis of survival data. Chapman and Hall: London
- Kalbfleisch, J. and Prentice, R. (2002). The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Statistics
- E. Marubini, M.G. Valsecchi (1995). Analysing survival data from clinical trials and observational studies, Chichester: John Wiley & Sons
- Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457481
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22(4):719-48.
- Gehan EA. A generalized two-sample Wilcoxon test for doubly censored data. Biometrika. 1965 Dec;52(3):650-3.
- Tarone RE, Ware J. On distribution free tests for equality of survival distributions. Biometrika 197;64:156-160.
- Prentice RL. Linear rank tests with right-censored data. Biometrika 1978; 65:167-179.
- Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. Biometrika. 1982;69: 553-566.
- Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. J Clin Oncol. 1983;1(11):710-9.