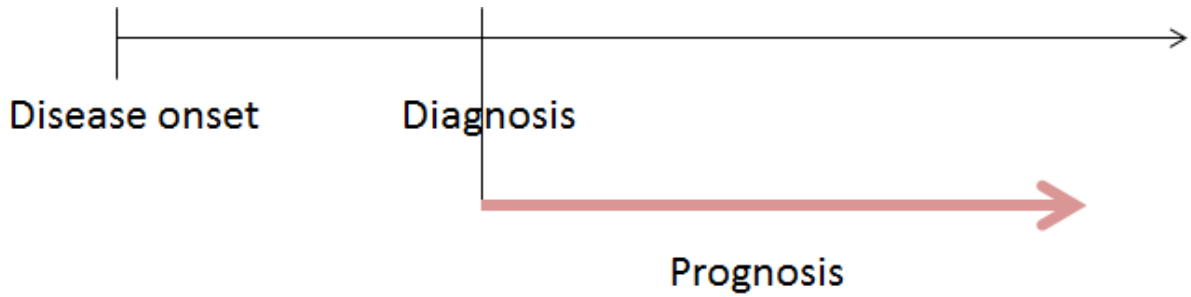
Measures of prognosis I: risk and survival

Milena Maule, Lorenzo Richiardi, Daniela Zugna

UNIVERSITÀ DEGLI STUDI DI TORINO
ALMA UNIVERSITAS TAURINENSIS

Prognosis is a prediction of the course of a disease following its onset: it refers to the possible outcomes of a disease and the frequency with which they can be expected to occur.



Prognostic factors can be identified with the characteristics of a particular patient and they can be used to predict outcome. Prognostic factors need not necessarily cause the outcomes, just be associated with them strongly enough to predict their development.

Prognostic factors can be any of several types, including:

- Demographic (e.g. gender, age, and so on)
- Disease-specific (e.g. tumour stage)
- Co-morbidities
- ...

1. Prognosis is predicting the progress or outcome of the disease
2. A person who was just diagnosed with a disease would be interested in the prognosis of the disease
3. Any measures used to quantify prognosis must be based on a set of people with a specified disease (denominator)

Potential end-points of interest:

- Death
- Recurrence
- Remission
- Cure
- ...

Incidence proportion: measures the proportion of new cases during a specified period of time (number of new cases in time period / number of people at risk)

What is the probability that patient will fail from the "event of interest" within a specified time interval?

Risk: is the probability of an individual to fail from the event of interest in a specified time interval: $P(T_i \leq t)$

The individual risk is estimated from averages taken from populations

$\Downarrow$

Risk and incidence proportion are synonymous

The risk (R) is calculated dividing the number of new cases of an event occurred in a population by the number of subjects at risk in the population at the beginning of the observation

$$\text{Risk} = \frac{\text{n° of cases of event}}{\text{Population at risk (at the beginning of the observation)}} = \frac{D}{N}$$

- Risk is also known as incidence proportion or cumulative incidence
- Uninterpretable without specification of the time period to which it applies

Mortality in 800 cancer patients with or without metastasis (5-year follow-up)

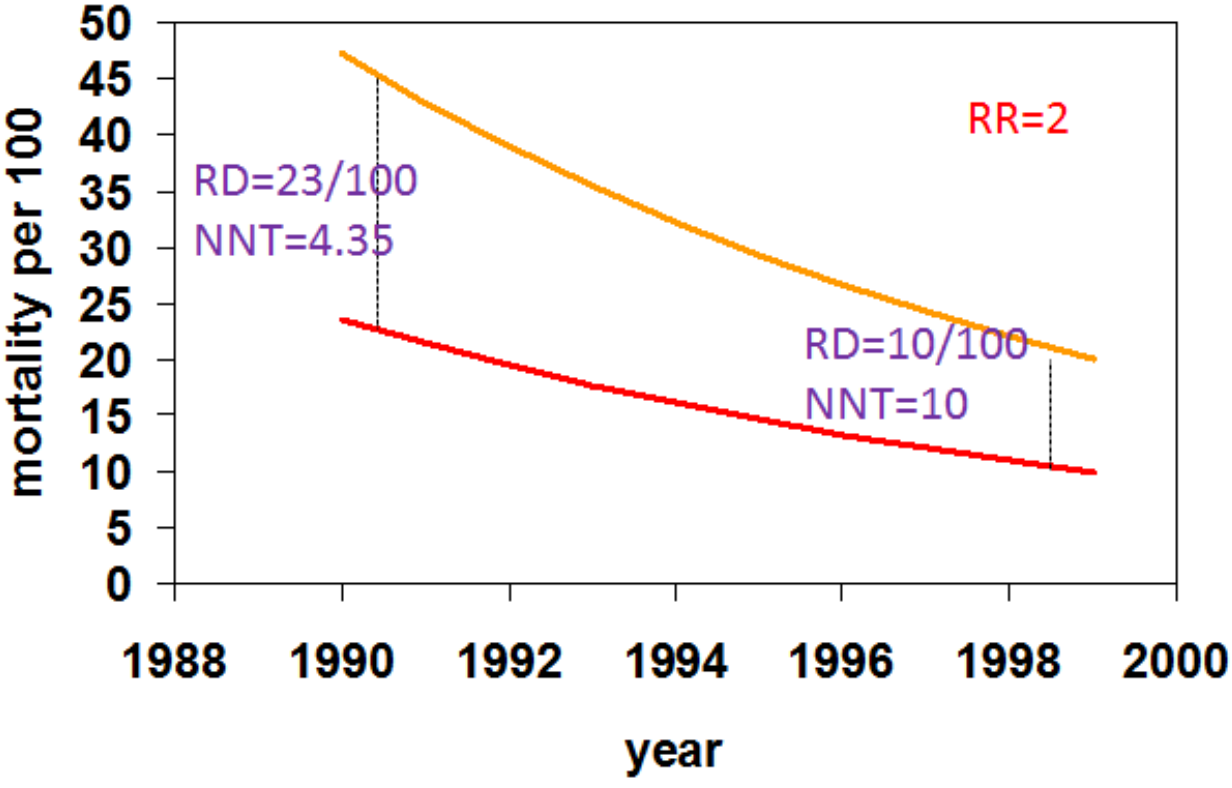
	Metastasis	
	Yes	No
No. initially at risk	400	400
Deaths	200	100
Person-years at risk	1500	1750
Risk	50%	25%
Rate	13 per 100 pyrs	5.7 per 100 pyrs

RELATIVE MEASURES:

Risk ratio	2.0
Rate ratio	2.3

ABSOLUTE MEASURES:

Risk difference	25 per 100 patients
Rate difference	7.3 per 100 pyrs
NNT	1/0.25 = 4

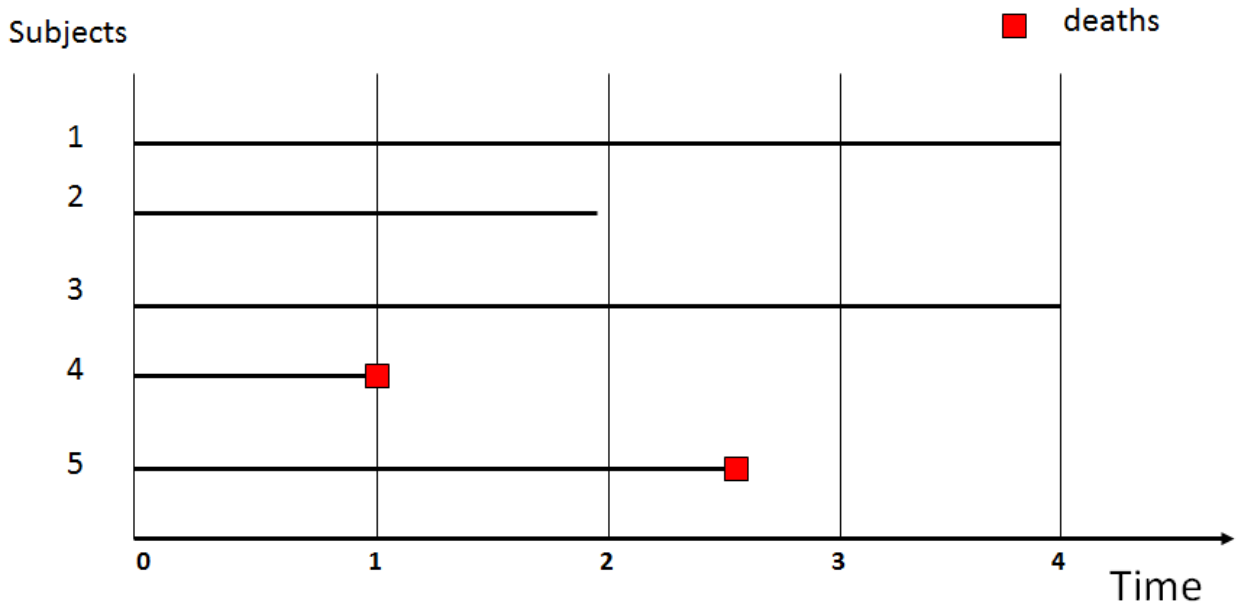


Number needed to treat(NNT): is the average number of patients who need to be treated to prevent one additional bad outcome and is used in assessing the effectiveness of a health-care intervention. It is defined as the inverse of the absolute risk reduction. The higher the NNT, the less effective is the treatment

Risk difference(RD): is the difference between the observed risks (proportions of individuals with the outcome of interest) in two groups

- risk difference for an individual describes the estimated difference in the probability of experiencing the event
- risk difference can be calculated for any study, even when there are no events in either groups
- the clinical importance of a risk difference may depend on the underlying risk of events: a risk difference of 0.02 (or 2%) may represent a small, clinically insignificant change from a risk of 58% to 60% or a proportionally much larger and potentially important change from 1% to 3%
- although the risk difference provides more directly relevant information than relative measures, it is important to be aware of the underlying risk of events and consequences of the events when interpreting a risk difference

To calculate risk, everyone being studied has to be followed for the complete specified time, but somebody may fail from some other cause or be lost to follow-up $\Rightarrow$ the incidence can be expressed as a rate (λ)



Numerator: 2 deaths
Denominator: (?)
5 subjects at risk but 1 subject lost to follow-up (it is only know that the event will happen but it is unknow the time)

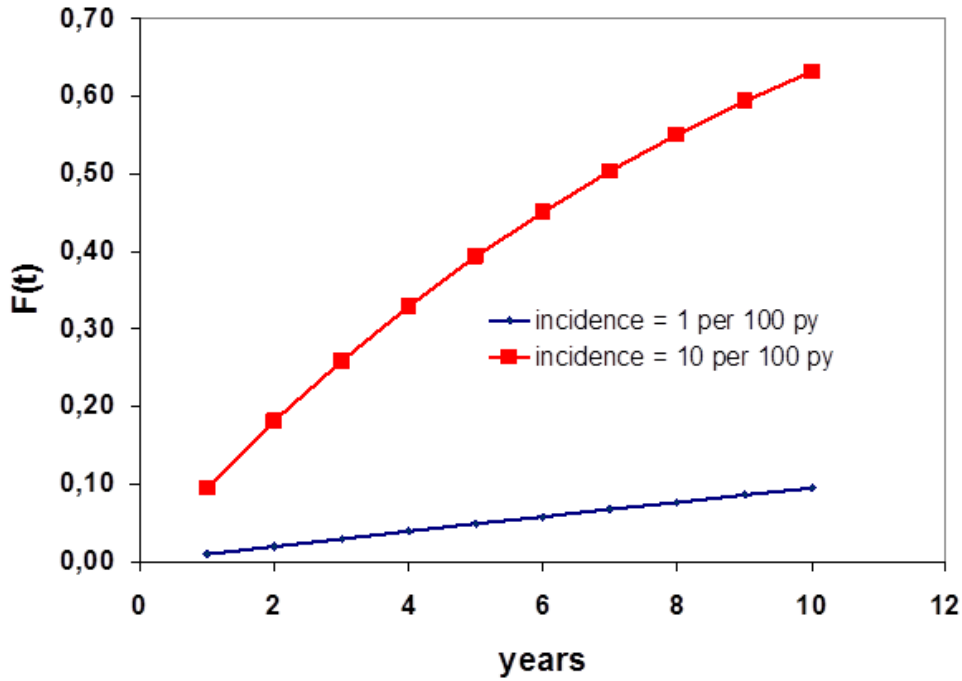
Rate: number of cases per time of observation

$$\text{Rate} = \frac{\text{n° of cases of event}}{\text{total time of observation}} = \frac{D}{Y}$$

Example: λ=2/13.5=0.15 $\Rightarrow$ 15 deaths per 100 person-years

This calculation of the incidence rate assumes that incidence remains constant during the period of study

The underlying model, assuming a constant rate over time, is an exponential model:



The risk at time t is given by: $F(t) = P(T \leq t) = 1 - exp(-\lambda t)$

Example:

λ=2/13.5=0.15 $\Rightarrow$ 15 deaths per 100 person-years

4-year risk: $F(4) = 1 - exp(-0.15 * 4) = 0.45$

Approximated formula if the risk is low (say $< 20\%$): F(t)=$\lambda * t$

Rate of testicular cancer at age 20-24: 25 per 10^5 PY

5-year risk of testicular cancer for a men aged 20: 25*5 per 10^5 = 125 per 10^5

Mortality in 800 cancer patients with or without metastasis (5-year follow-up)

	Metastasis	
	Yes	No
No. initially at risk	400	400
Deaths	200	100
Person-years at risk	1500	1750
Risk	50%	25%
Rate	13 per 100 pyrs	5.7 per 100 pyrs

RELATIVE MEASURES:

Risk ratio	2.0
Rate ratio	2.3

ABSOLUTE MEASURES:

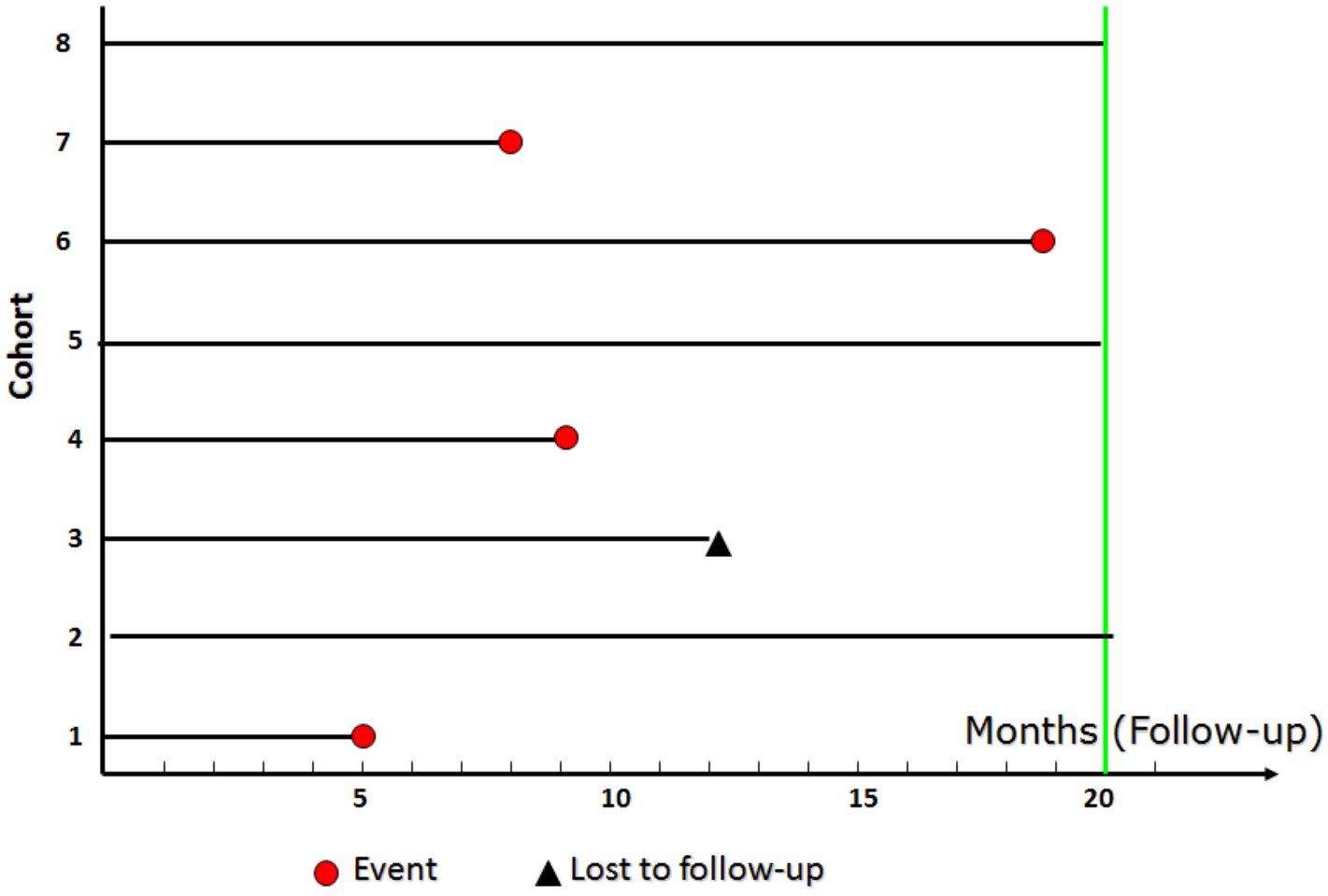
Risk difference	25 per 100 patients
Rate difference	7.3 per 100 pyrs
NNT	1/0.25 = 4

- The risk (cumulative incidence) is defined as $F(t) = P(T \leq t)$
- The incidence rate or hazard is defined as:

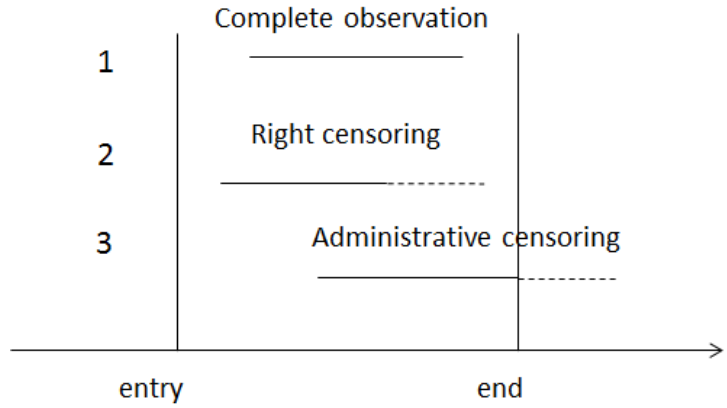
  $\lambda(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t+\Delta t | T > t)}{\Delta t}$

  $\lambda(t)\Delta t$: probability that failure is between t and $t + \Delta t$ conditioned on having survived until t
- By assuming constant rate: $\lambda(t) = \lambda \rightarrow F(t) = 1 - exp(-\lambda t)$
- If a long period of study is used, the risk of failure may change over time $\Rightarrow$ it becomes necessary to calculate incidence rates over shorter periods of time (during which they are relatively constant) and then aggregating them $\Rightarrow$ survival analysis

Cohort study



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

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

The censoring is independent if the censored subject at a given time is rapresentative of all subjects surviving to that time. If the subject drops out from the study because of a cause associated to the event of interest, the censoring is dependent

Survivor function: $S(t) = P(T > t) = 1 - F(t)$

- In absence of censoring, $S(t)$ is estimated as the ratio of the number of survivors at time t to the number of subjects at risk at the beginning of the observation
- In presence of censoring and by assuming constant rate over time, $S(t)$ can be estimated as: $\widehat{S}(t) = 1 - \widehat{F}(t) = exp(-\widehat{\lambda}t)$
- In presence of censoring, the survival function can be non-parametrically estimated by Kaplan-Meier method

Cohort study

1° battle

Prob.=a

death

1-a

1-a=A

2° battle

Prob.=b

death

1-b

A*(1-b)=B

3° battle

Prob.=c

death

1-c

B*(1-c)=C

4° battle

Prob.=d

death

1-d

C*(1-d)=D

Survived

Cumulative survival

Time

Survival proportion

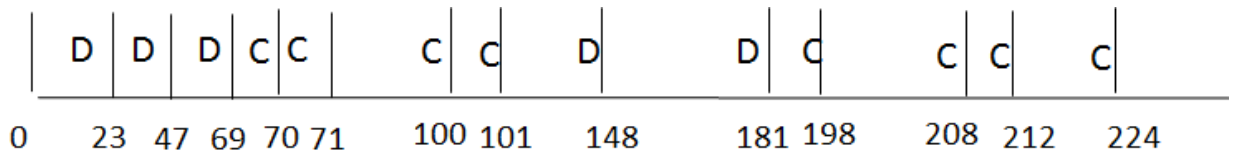
1. Divide time into short bands
2. Calculate the period specific survival proportion (numbers survived/ numbers at risk)
3. Multiply the conditional probabilities

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)

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



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

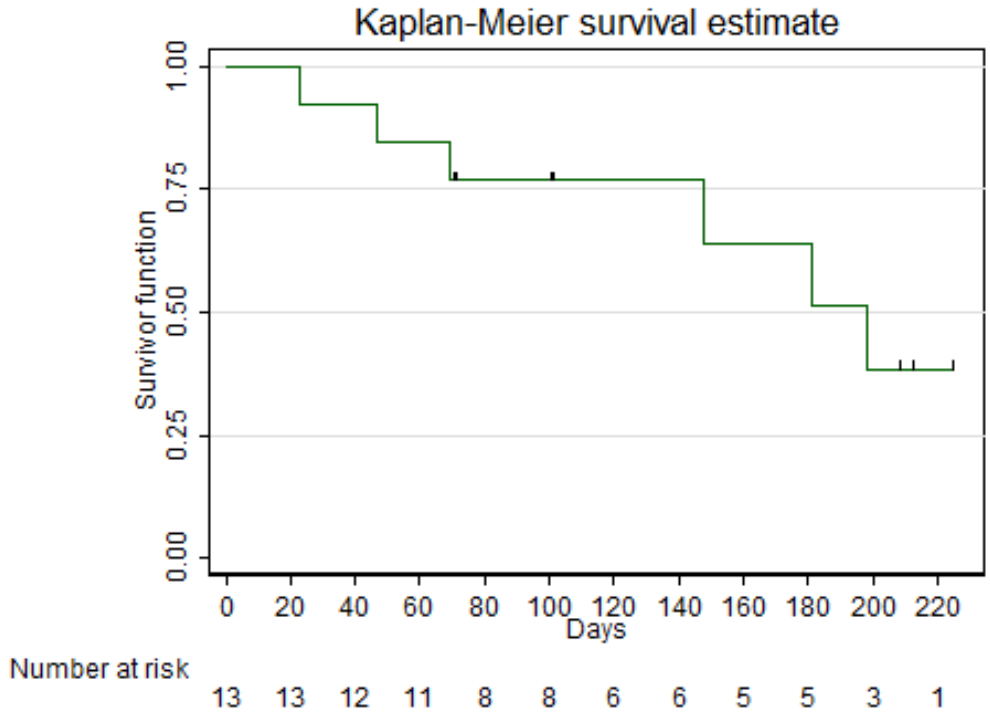
Kaplan-Meier estimator: $\widehat{S}(t) = \prod_{j=1}^{m}(1 - \frac{d_j}{n_j})$

- $j = 1, \ldots, m$ are the short bands defined by failures/censoring times
- d_j is the number of failures at time t_j
- n_j is the number of subjects at risk just prior of time t_j

Example

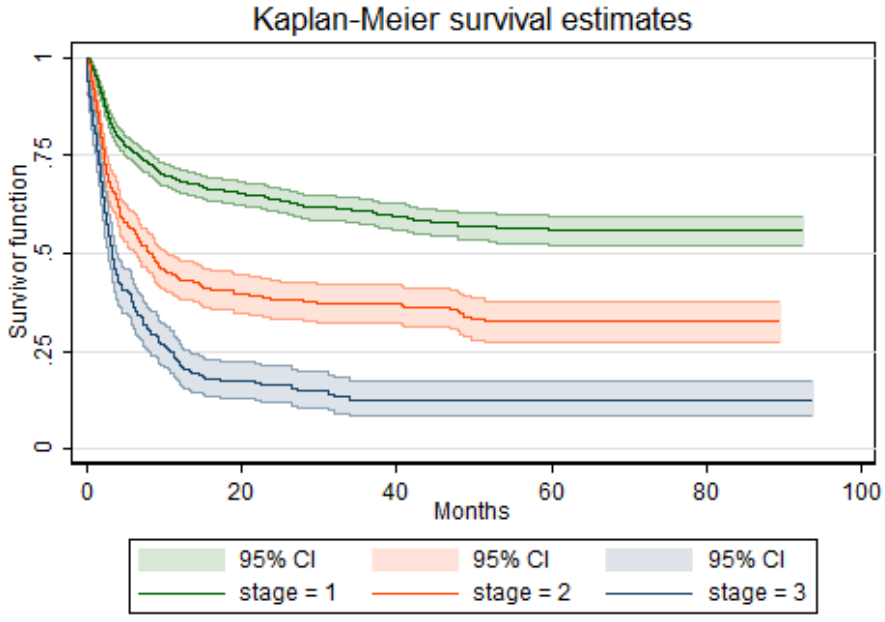
Survival of 13 women affected by breast cancer

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			



- $\widehat{S}(t)$ 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

Comparison between survival curves



- How can we compare individuals diagnosed at different stages in terms of survival
- 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 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?)

- Test based on the median survival: not always the median exists and often precision of the estimates is low

Taking into account the total survival experience:

- The log-rank test (non-parametric hypothesis test)

- Point estimates: hazard ratio

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	

- 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

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: $E(d_{jB}) = E_{jB} = \frac{d_j n_{jB}}{n_j}$

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

$O_B = \sum_j O_{jB}, E_B = \sum_j E_{jB}$

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

- Under $H0$, X^2 is asymptotically distributed as a χ^2_1
- The higher is X^2 the smaller is the probability that the sample is consistent with $H0$

T	N	Experiment 1			Experiment 2		
		n_1	Obs_1	Exp_1	n_2	Obs_2	Exp_2
5	49	21	0	**1*21/49=0.429**	28	1	**1*28/49=0.571**
6*	48	21	0	**-**	27	0	**-**
11	47	21	0	**2*21/47=0.894**	26	2	**2*26/47=1.106**
13	45	21	0	**0.467**	24	1	**1-0.467=0.533**
24	44	21	0	**0.477**	23	1	**0.523**
30	43	21	1	**0.488**	22	0	**0.512**
50	42	20	1	**0.476**	22	0	**0.524**
50*	41	19	0	-	22	0	**-**
...		...			...		

$$X^2 = ([5 - 8.86]^2/8.86) + ([14 - 10.14]^2/10.14) = 3.152$$

$$p = 0.07$$

Stratified log-rank test

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

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 potential confounder, 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.

Summary

- 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
- The log-rank test gives equal weight to all time points irrespective of numbers at risk at the time of failure
- Different non-parametric tests have been proposed according to aim of the study

Hazard (rate) ratio

- The hazard is the instantaneous probability of failure within next interval of time, having already survived up to that time: $\lambda(t)$
- The hazard ratio (HR) is the relative hazard, when two groups are compared: HR = 2 doubled risk of having the outcome at any point in time during follow up
- The HR inherently assumes proportional hazards, that is constant ratio of the hazards over time (note: the log-rank test also assumes proportional hazards)

From previous example: HR = (O1/E1) / (O2/E2)= (8/13.43) / (14/8.57) = 0.35

By fitting Cox model: HR=0.35, 95% CI:0.15,0.85

Summary

In estimating a survival curve:

- Exposure status (e.g. treatment, no treatment)
- Follow-up time: Exit time entry time
- Outcome status upon exit (subject having outcome or censored)
- Confounders (age, sex, stage and grade of disease at entry, etc)

Statistical power is related to number of outcomes, not the number of subjects

Summary

In reading a survival curve:

- Observe the shape of the curve more than details (log-rank test and hazard ratio are reliable under proportionality assumption)
- 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
- Consider the percentage of censoring

Incorrect survival analysis

- Mean survival time is useless, unless
  1. all subjects are followed until outcome
  2. no subjects are censored

- Median survival time is not sensible, unless it is derived from the KM-curve (to take censoring into account)

- Survival proportion at a certain point in time
  1. discards a lot of the available information on survival
  2. is wrong, unless the time was specified a priori

Some limits of K-M survival curve

- Confounding cannot be taken into account
- Predictors are categorical
- Cannot be used directly in case of competing risks