Biostatistics in Oncology Trials: Survival Analysis

“Why clinicians hope for survivors and statisticians for deaths”

Geert Verbeke

I-BioStat: Interuniversity Institute for Biostatistics and statistical Bioinformatics

K.U.Leuven & Hasselt University, Belgium

geert.verbeke@med.kuleuven.be

http://perswww.kuleuven.be/geert_verbeke

1 Overview

- ▷ Example
- ▷ The survival curve
- ▷ Estimation of survival curve
- ▷ The problem of censoring
- ▷ Kaplan-Meier estimate of survival curve
- ▷ Comparison of survival curves
- ▷ Power issues
- ▷ Examples from biomedical literature

2 Example: Survival times of cancer patients

- Cameron and Pauling [1]; Hand et al. [2] p. 255
- Patients with advanced cancer of the stomach, bronchus, colon, ovary, or breast were treated (in addition to standard treatment) with ascorbate.
- The outcome of interest is the survival time (days)
- Research question(s):

What is the prognosis for a patient with specific type of cancer ?

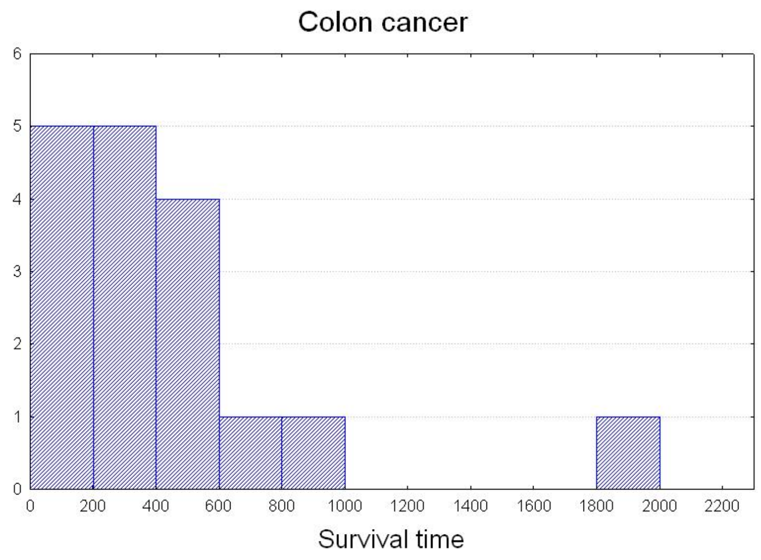
Do survival times differ with organ affected ?

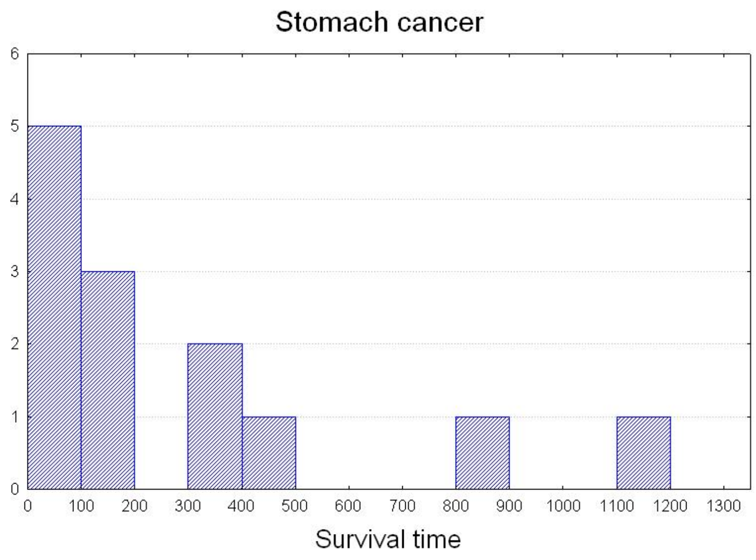
- Dataset 'Cancer':

Stomach	Bronchus	Colon	Ovary	Breast
124	81	248	1234	1235
42	461	377	89	24
25	20	189	201	1581
45	450	1843	356	1166
412	246	180	2970	40
51	166	537	456	727
1112	63	519		3808
46	64	455		791
103	155	406		1804
876	859	365		3460
146	151	942		719
340	166	776		
396	37	372		
	223	163		
	138	101		
	72	20		
	245	283		

	Average (days)	Median (days)
Stomach:	286	124
Bronchus:	211.6	155
Colon:	457.4	372
Ovary:	884.3	406
Breast:	1395.9	1166

- Note the severe differences between averages and medians, due to the skewness of the distribution:





- Comparisons between groups is therefore based on parametric tests after appropriate transformation (e.g., logarithmic), or based on non-parametric tests (e.g., Wilcoxon test).

3 The survival curve

- Often it is of interest to make a prognosis for specific patients, i.e., it is of interest to estimate the probability of 'surviving' a specific amount of time

- In other contexts, the response is not **'survival'**, but still a **'time to event'**:
    - ▷ Progression free 'survival'
    - ▷ How long will a bulb 'survive'
    - ▷ Time untill first tooth is affected with caries
    - ▷ Time a rat needs to find the exit of a maze
    - ▷ . . .

- Terminology: **Survival** and **Failure**

- In the cancer example, it may be of interest to estimate how likely it is that a patient with colon cancer, treated (in addition to standard treatment) with ascorbate, will survive 1 year, 2 years, . . .

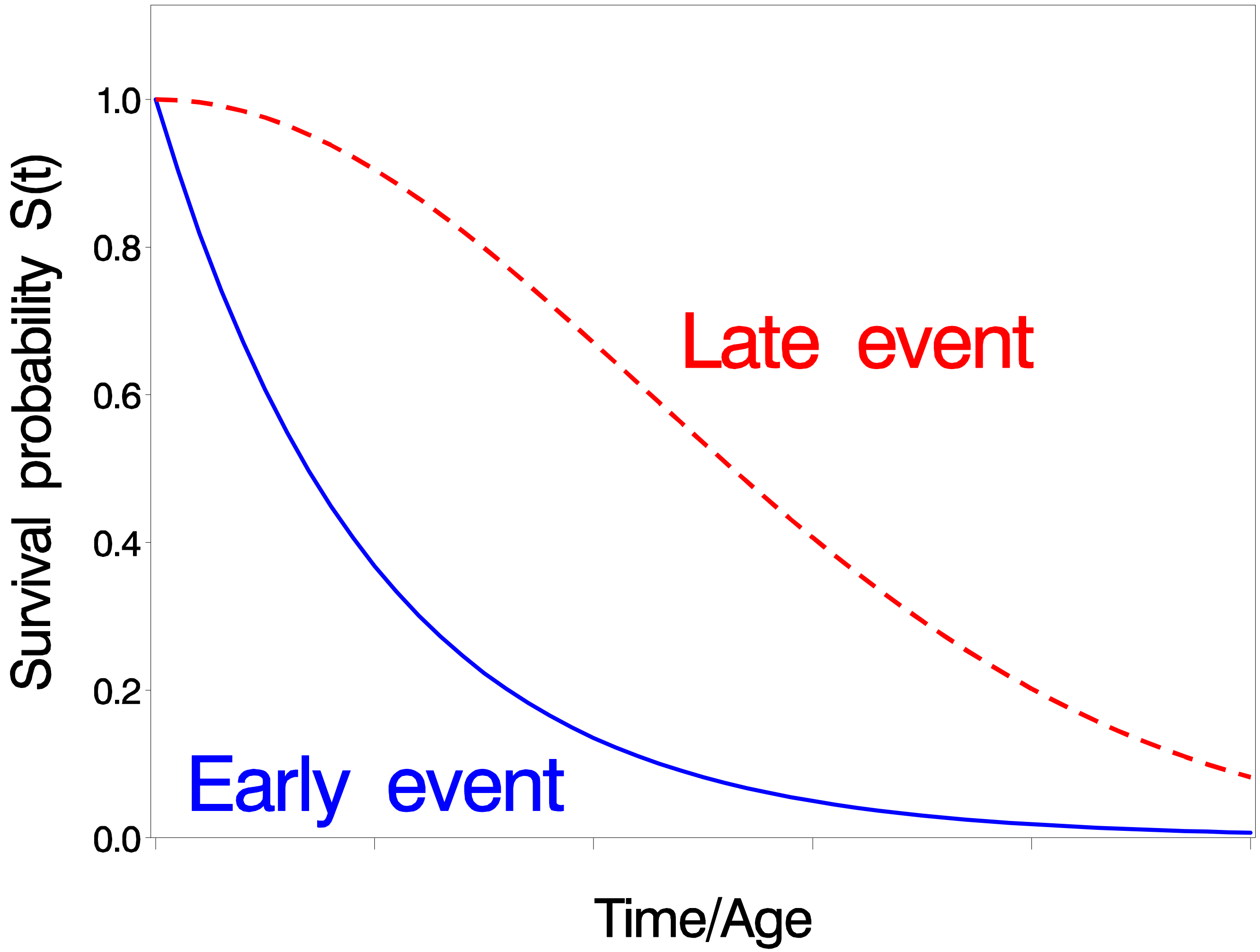
- Interest is then in the **survival function / curve**:

$$S(t) = P(\textbf{Outcome} > t)$$

“The probability of surviving time point t”

- Properties of $S(t)$:
  - ▷ $S(0) = 1$: There is absolute certainty to ‘survive’ $t = 0$
  - ▷ $S(+\infty) = 0$: There is absolute certainty to ‘fail’ eventually
  - ▷ $S(t)$ is a decreasing function

- Examples of survival curves:



4 Estimation of survival curve

- As $S(t)$ can be interpreted as a proportion, it can easily be estimated by the observed proportion of subjects surviving time point t:

$$S(t) = P(\text{Outcome} > t) \quad \longrightarrow \quad \widehat{S}(t) = \frac{\#\ \text{subjects surviving } t}{N}$$

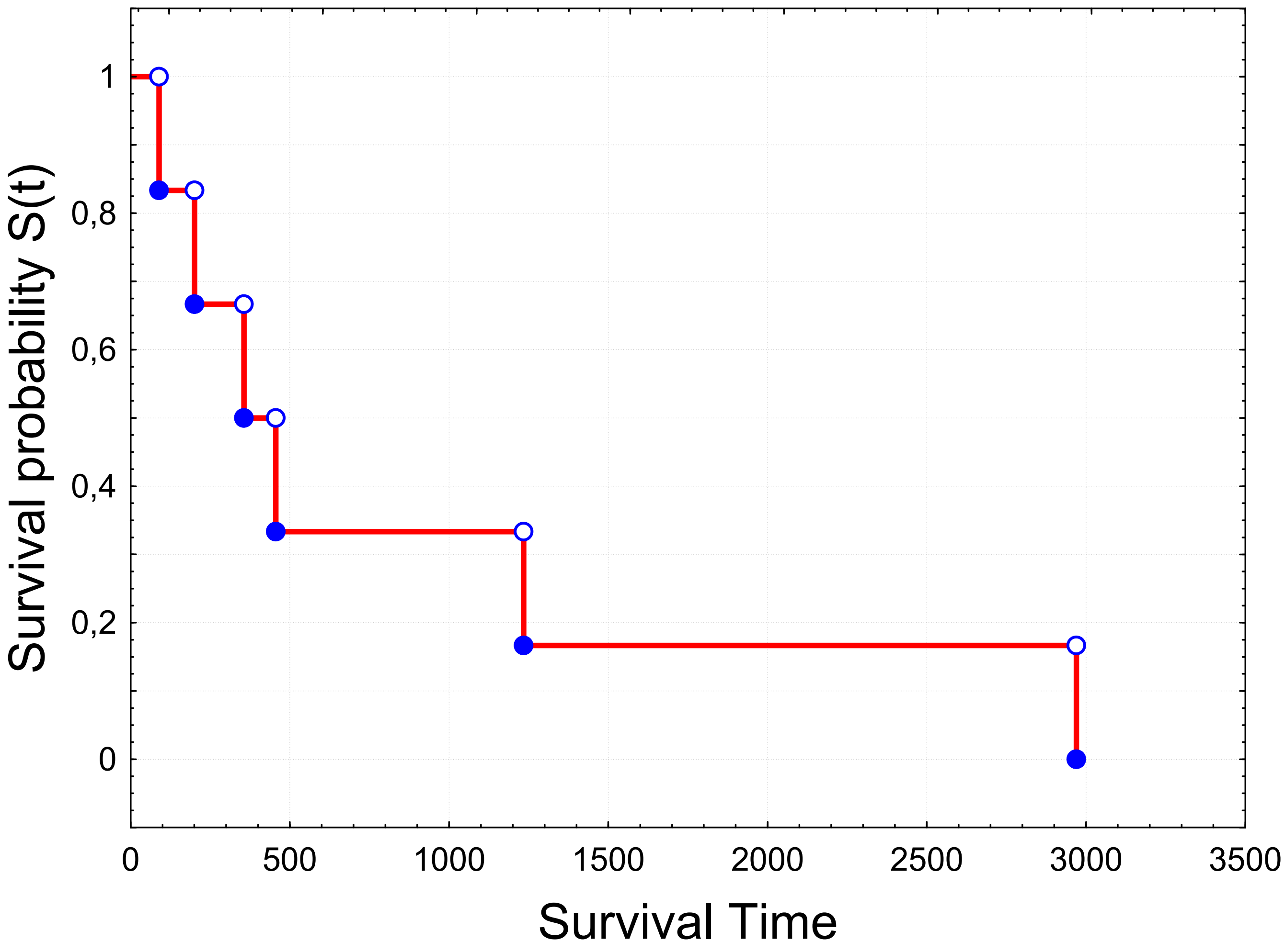
- As an example, we estimate the survival curve for ovary cancer patients
- The following 6 event times were recorded:

1234 89 201 356 2970 456

- Calculations:

Time (t)	# Surving t	$\widehat{S}(t)$
0	6	$6/6 = 1.00$
30	6	$6/6 = 1.00$
89	5	$5/6 = 0.83$
100	5	$5/6 = 0.83$
201	4	$4/6 = 0.67$
356	3	$3/6 = 0.50$
400	3	$3/6 = 0.50$
556	2	$2/6 = 0.33$
1234	1	$1/6 = 0.17$
2970	0	$0/6 = 0.00$

- Graphically:



5 The problem of censoring

Event time cannot always be measured !

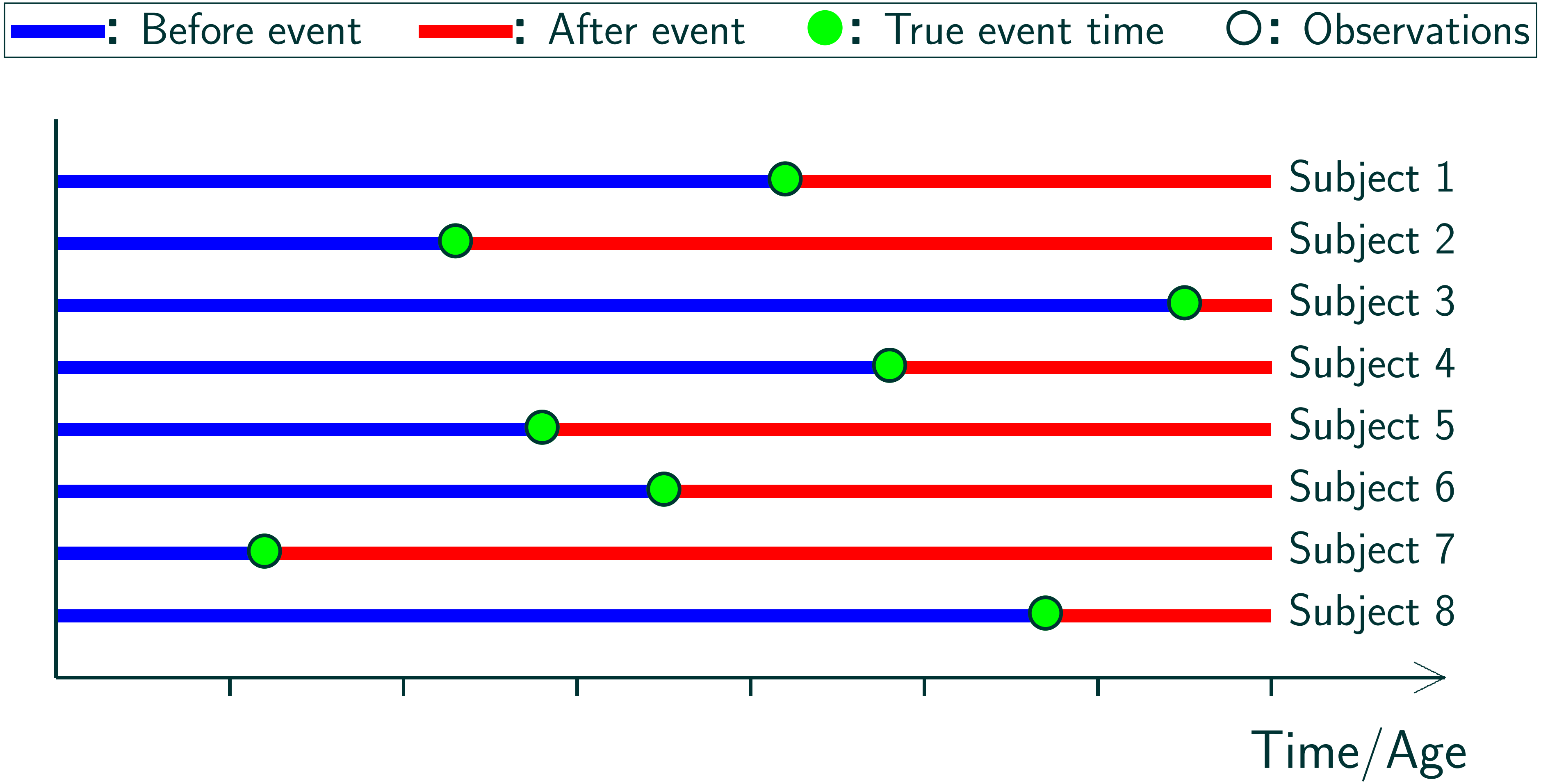
$$\Downarrow$$

Censored observations

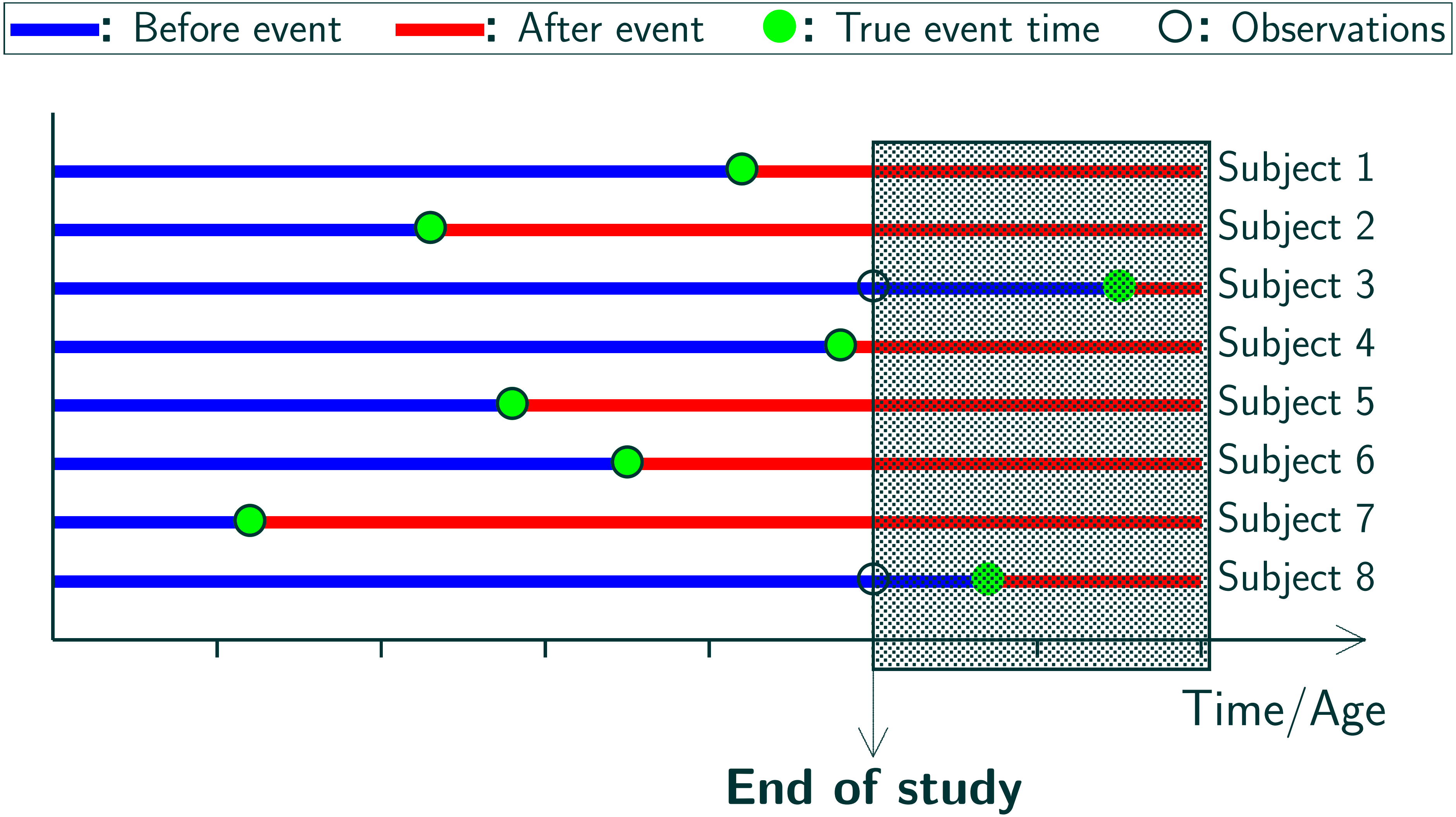
Various types of censoring:

- ▷ **Right**
- ▷ Left
- ▷ Interval
- ▷ Mixture of the above

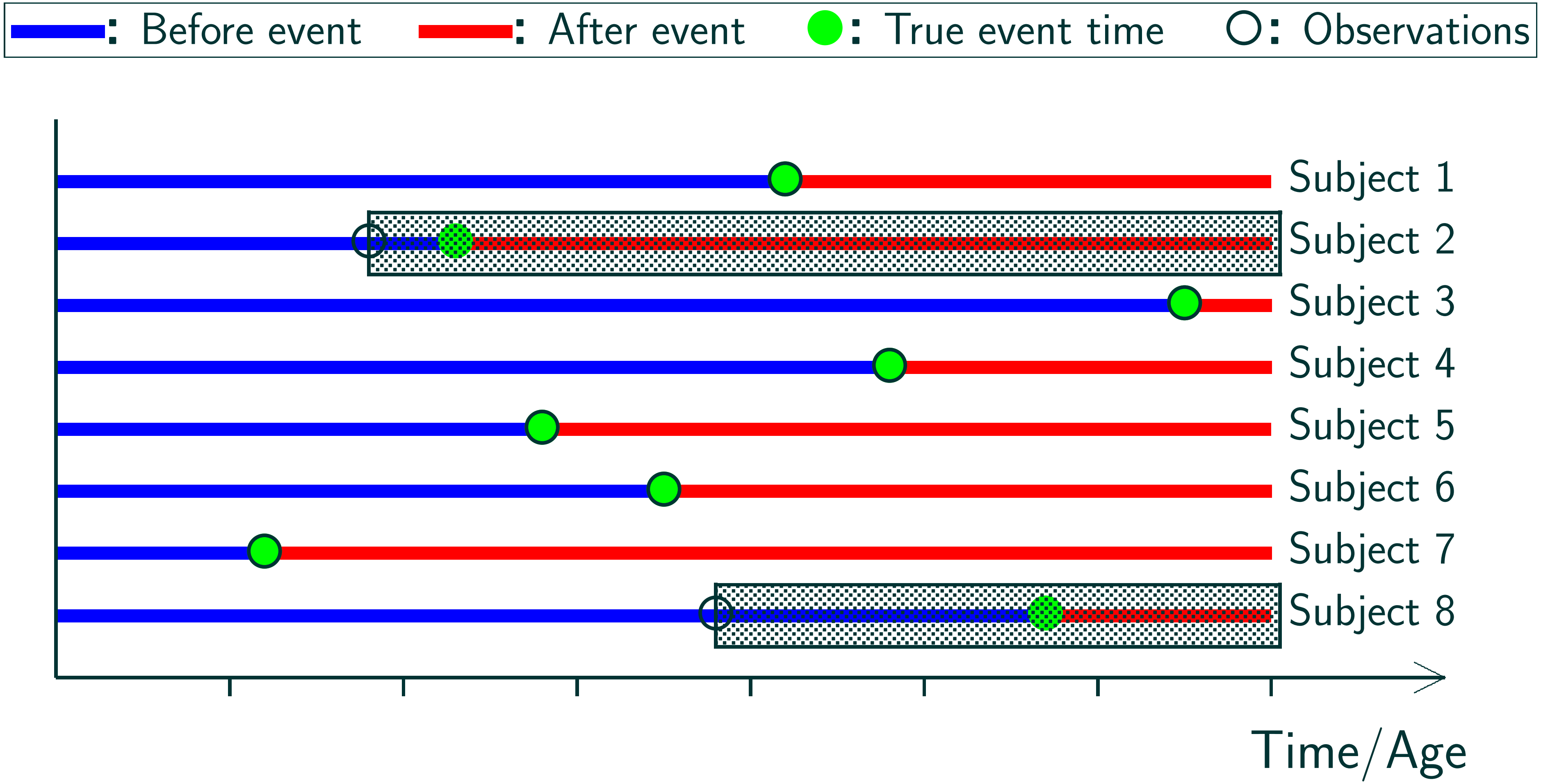
No censoring



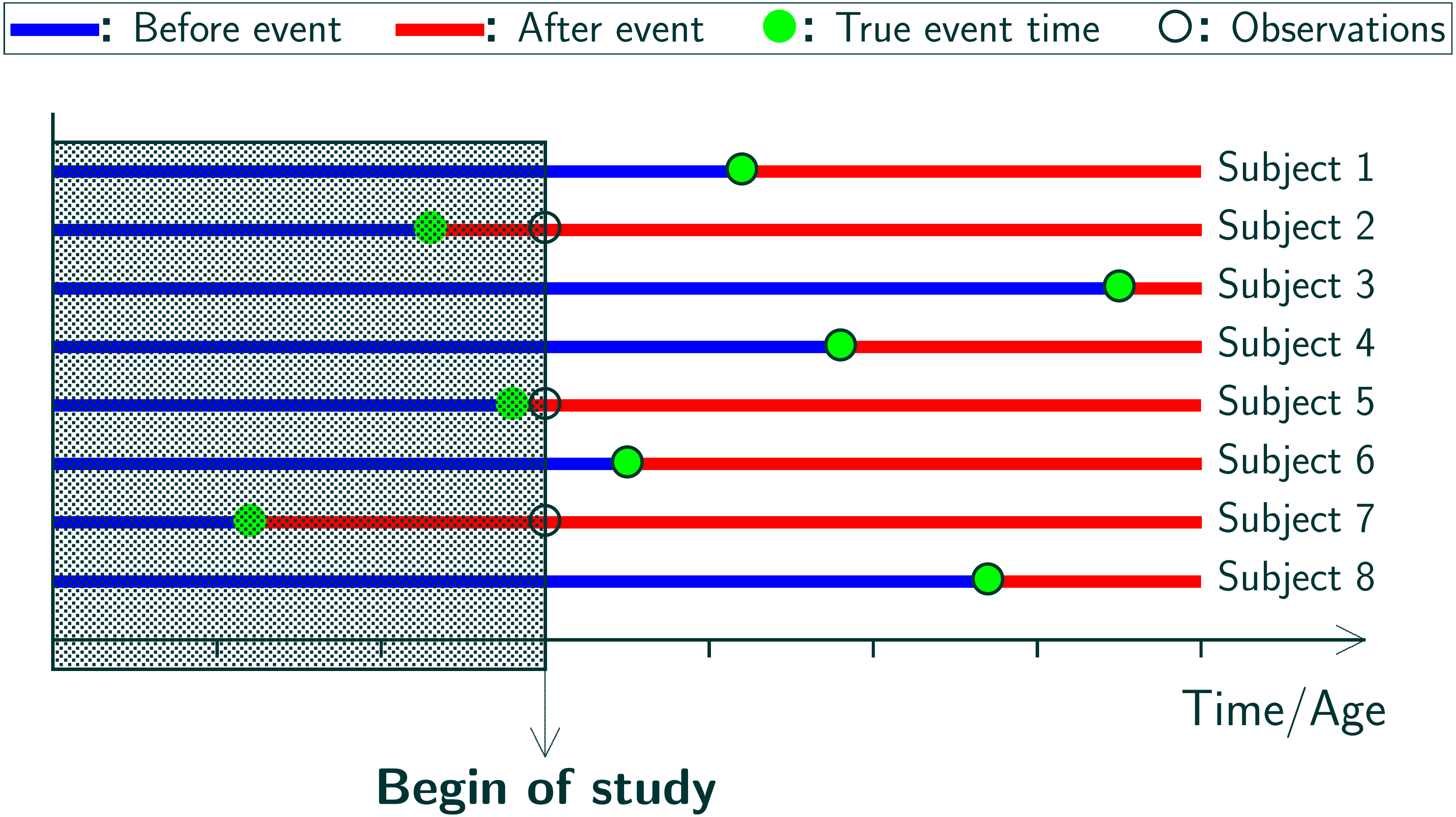
Right censoring due to study end



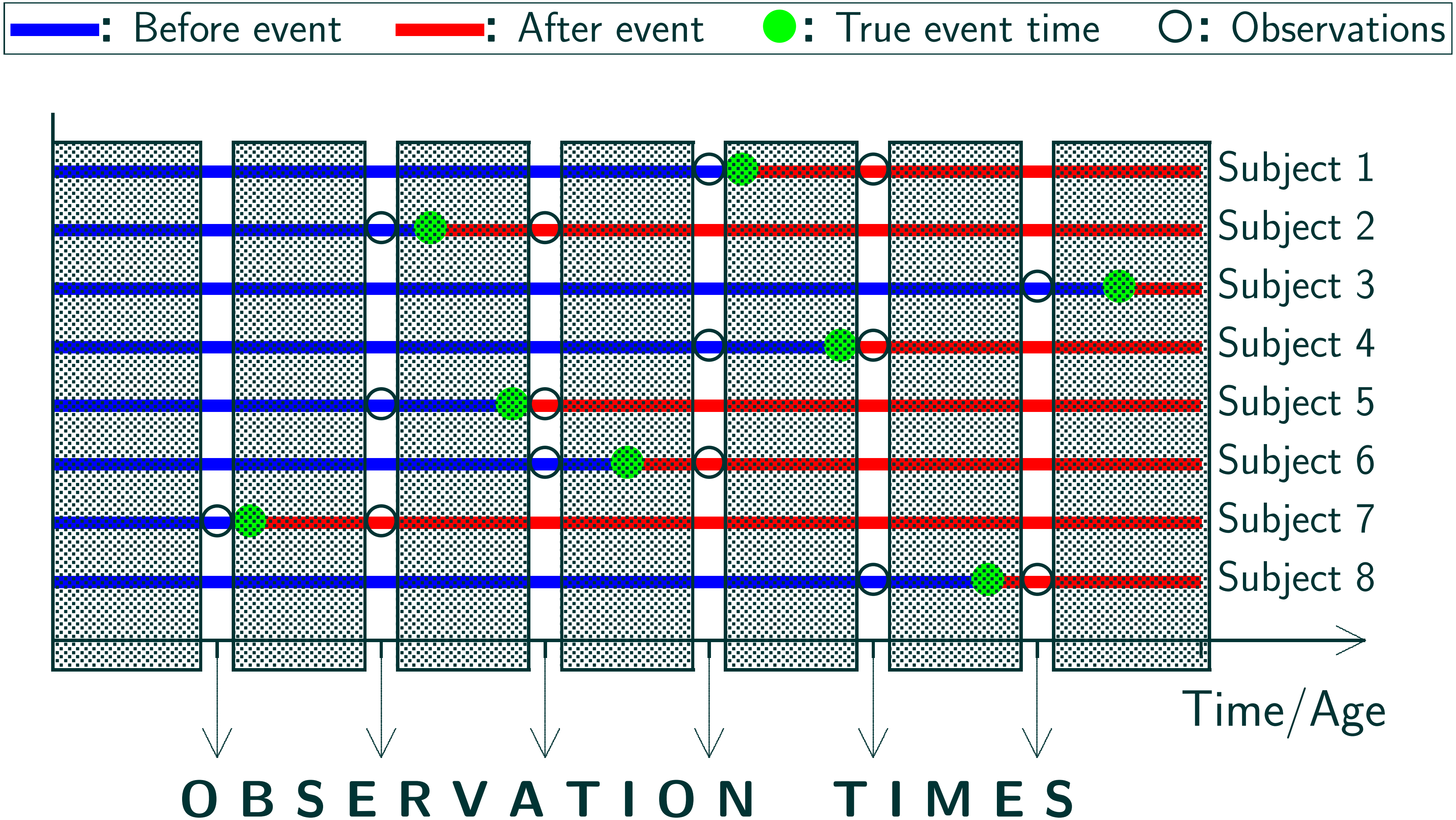
Right censoring due to dropout



Left censoring due to late study onset



Interval censoring due to discrete observation times



- Our focus will be on **right censoring**, i.e., either the true event time or a lower bound of it is observed

- Standard statistical tools for the analysis of censored observations assume random censoring:

Event time and censoring time are independent

- Counter examples:
  - ▷ Patients entering the study later have a better prognosis due to increased experience of surgeon
    $\Longrightarrow$ Negative association between censoring and event time
  - ▷ Patients leaving the study because they get worse
    $\Longrightarrow$ Positive association between censoring and event time

6 Example: Myelomatosis

- Peto et al. [3]; Allison [4] p.26
- Data on 25 patients diagnosed with myelomatosis (Kahler's disease), multiple malign tumours in the bone marrow
- Patients randomly assigned to two drug treatments
- Event time is the time from moment of randomization to death
- Some event times are censored due to study termination
- Patients with normal and patients with impaired renal functioning at moment of randomization

- Data:

Treat	Duration	Status	Renal	Treat	Duration	Status	Renal
1	8	1	1	2	180	1	0
1	852	0	0	2	632	1	0
1	52	1	1	2	2240	0	0
1	220	1	0	2	195	1	0
1	63	1	1	2	76	1	0
1	8	1	0	2	70	1	0
1	1976	0	0	2	13	1	1
1	1296	0	0	2	1990	0	0
1	1460	0	0	2	18	1	1
1	63	1	1	2	700	1	0
1	1328	0	0	2	210	1	0
1	365	0	0	2	1296	1	0
				2	23	1	1

Status:

- ▷ 0: Censored
- ▷ 1: Death

Renal:

- ▷ 0: Normal
- ▷ 1: Impaired

- Interest is in estimating and comparing the survival curves for patients with different treatments and for patients with different renal functioning at baseline

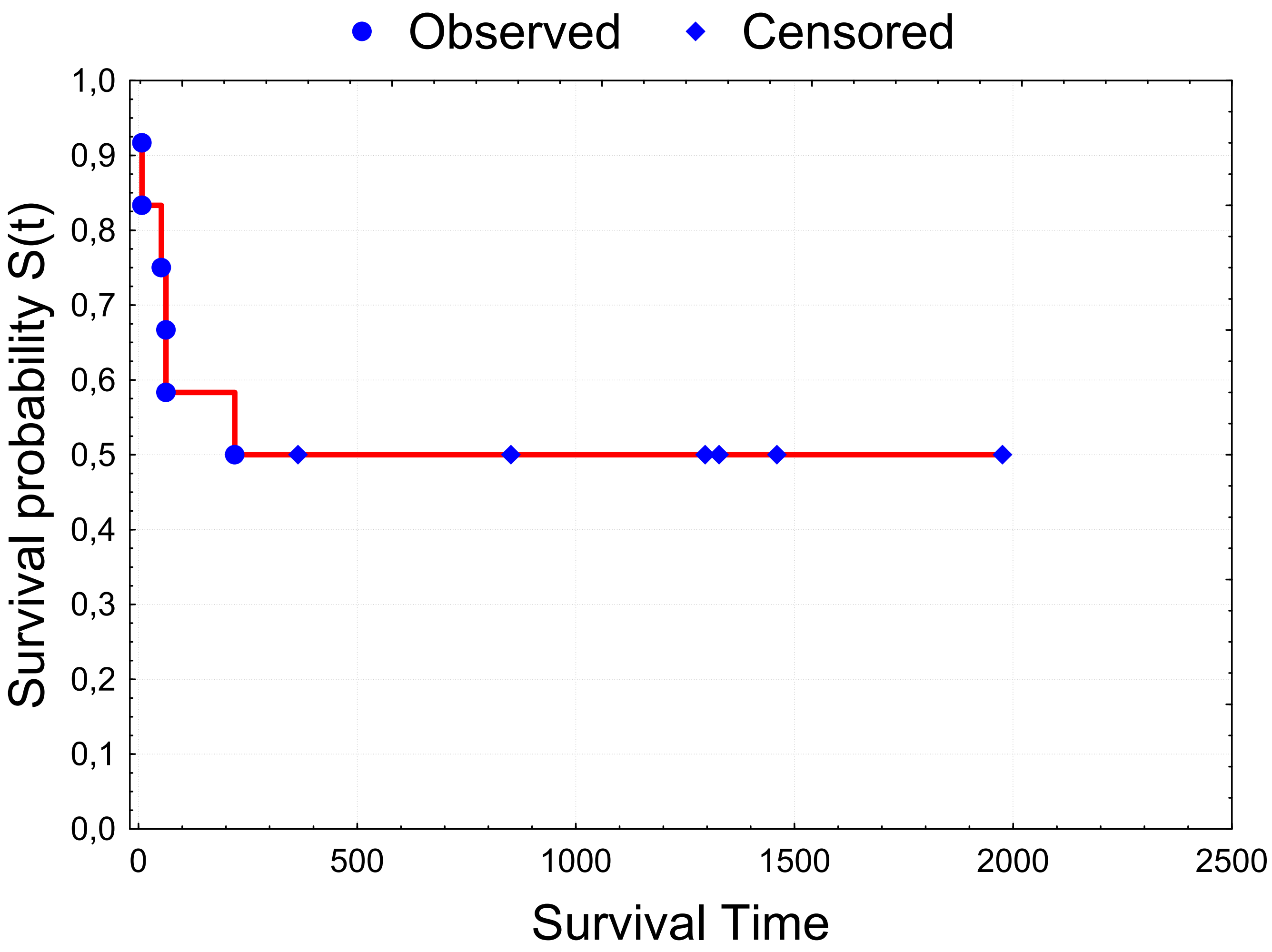
7 Kaplan-Meier estimate of survival curve

- Suppose interest is in estimating the survival curve for patients with treatment 1
- Observed data:

Duration:	**8**	**852**	**52**	**220**	**63**	**8**	**1976**	**1296**	**1460**	**63**	**1328**	**365**
Status:	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**

- Simple **'naive'** solutions:
  - ▷ Ignoring the censored observations: **Over-optimistic**
  - ▷ Treating censored observations as event times: **Over-pessimistic**

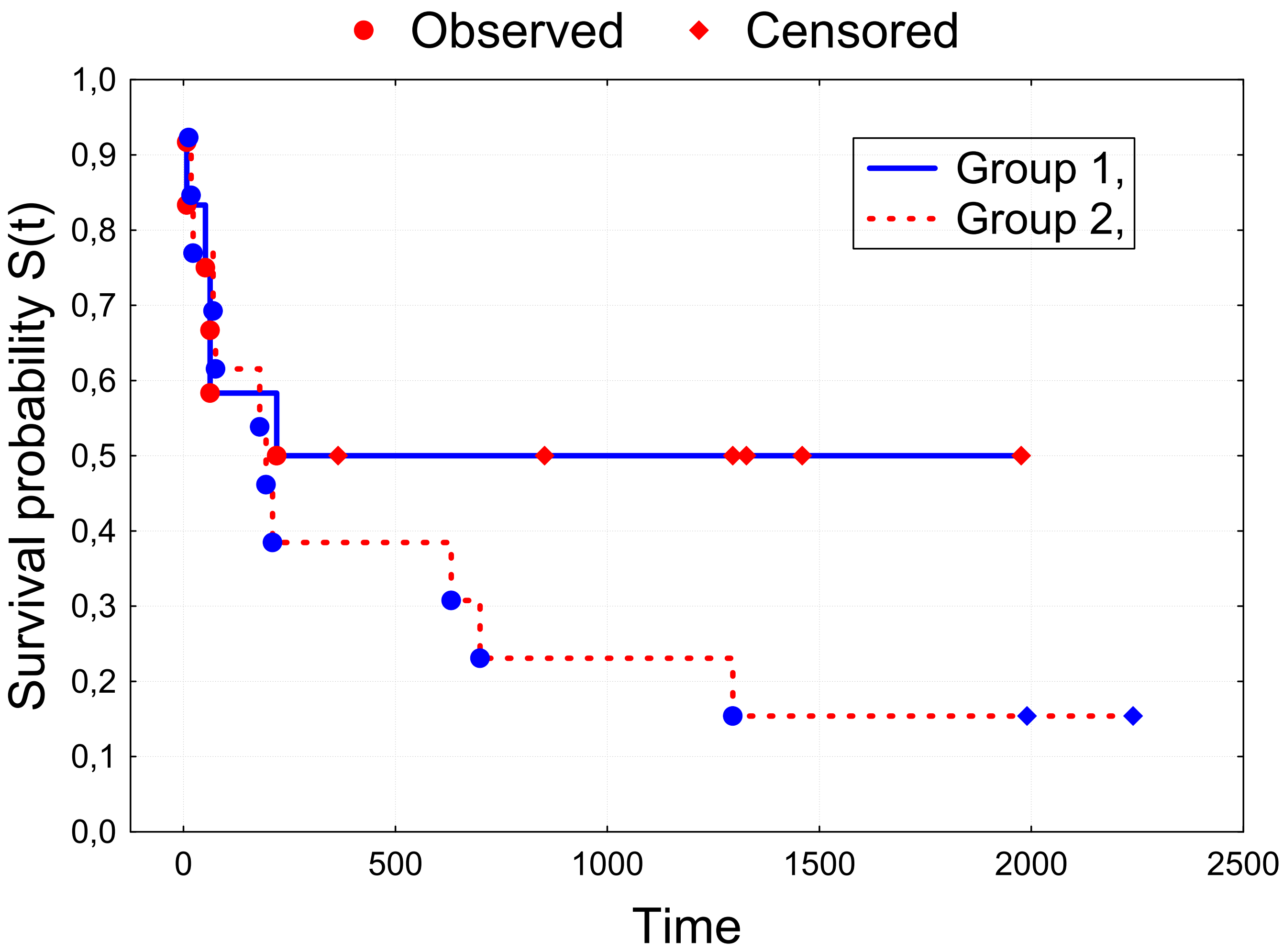
- The so-called **Kaplan-Meier** estimate $\widehat{S}(t)$ correctly accounts for the censoring:



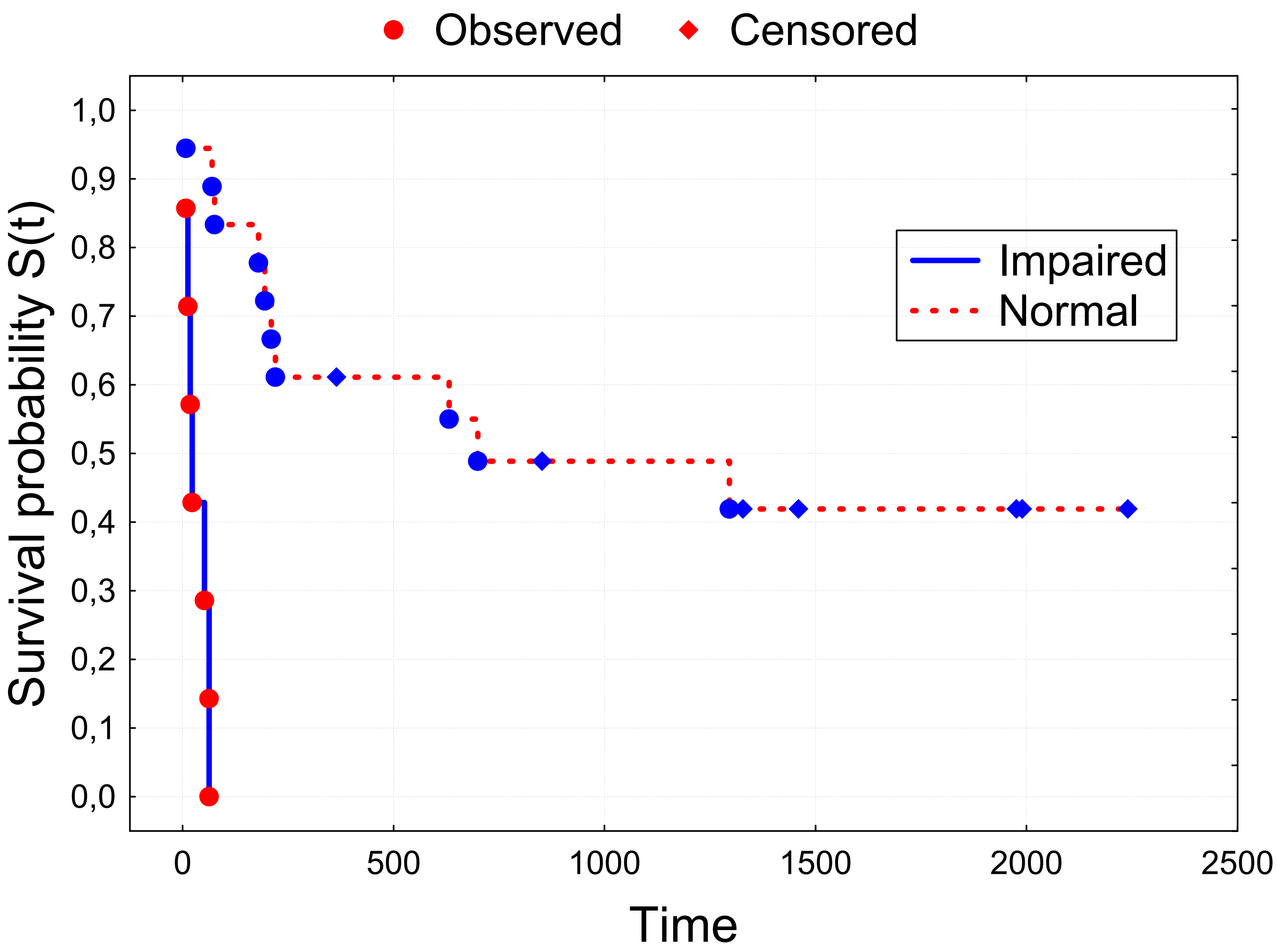
8 Comparison of survival curves

- Often, interest is in the comparison of survival curves of different groups
- For the Myelomatosis data, interest may be to compare survival between the two treatment goups
- Also of interest is the comparison of survival for patients with impaired renal functioning with survival for patients with normal renal functioning.
- We will focuss on the comparison of two groups, but extensions are available for the comparison of multiple groups
- For each group separately, the Kaplan-Meier estimate for the survival curve can be calculated.

- Kaplan-Meier estimates for both treatment groups:



- Kaplan-Meier estimates for patients with normal and impaired renal functioning, respectively:



- Due to the censoring, classical tests such as t-test and Wilcoxon test cannot be used for the comparison of the survival times

- Various tests have been designed for the comparison of survival curves, when censoring is present

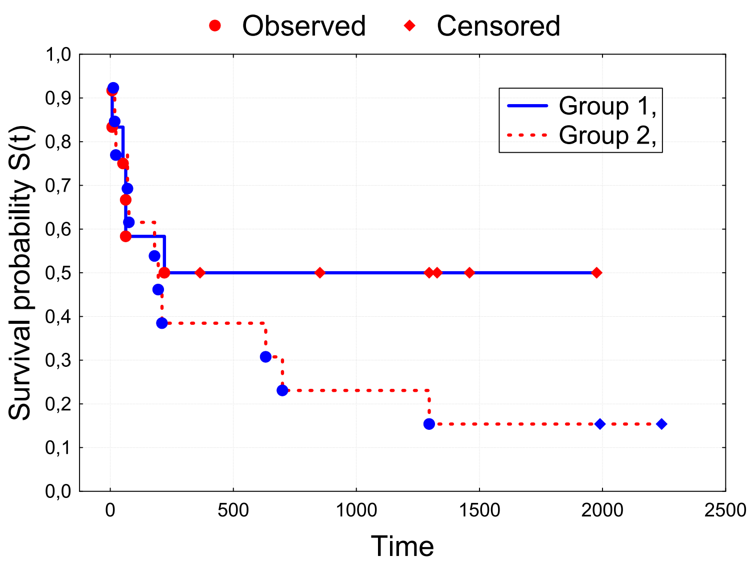
- The most popular ones are:
  - ▷ **Logrank** test
  - ▷ **Wilcoxon** (Gehan) test

- The Logrank test has **more** power than Wilcoxon for detecting **late** differences

- The Logrank test has **less** power than Wilcoxon for detecting **early** differences

- Test results:

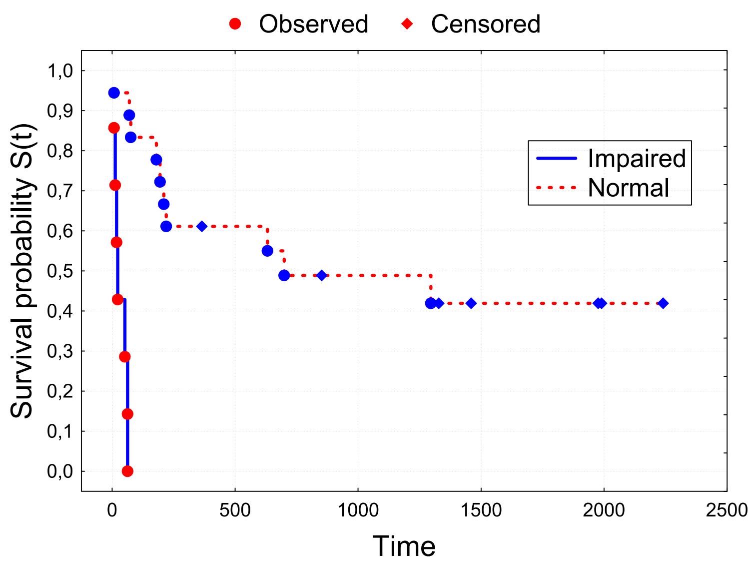
Effect of treatment



Logrank: p=0.2468

Wilcoxon: p=0.6260

Effect of renal functioning



Logrank: p=0.0029

Wilcoxon: p=0.0005

9 Power issues

The power of the tests depends on the number of events, not on the number of subjects

$\Downarrow$

Long-lasting huge studies needed to show small improvements versus successful therapies

10 Examples from biomedical literature

- Shatari et al. [5]:
  - ▷ Methods, p.439:

> The rate and duration of the recurrence requiring re-operation for obstructive symptoms were analysed by the Kaplan-Meier plot and log-rank test. χ^2 test and

▷ Figure 1, p.440:

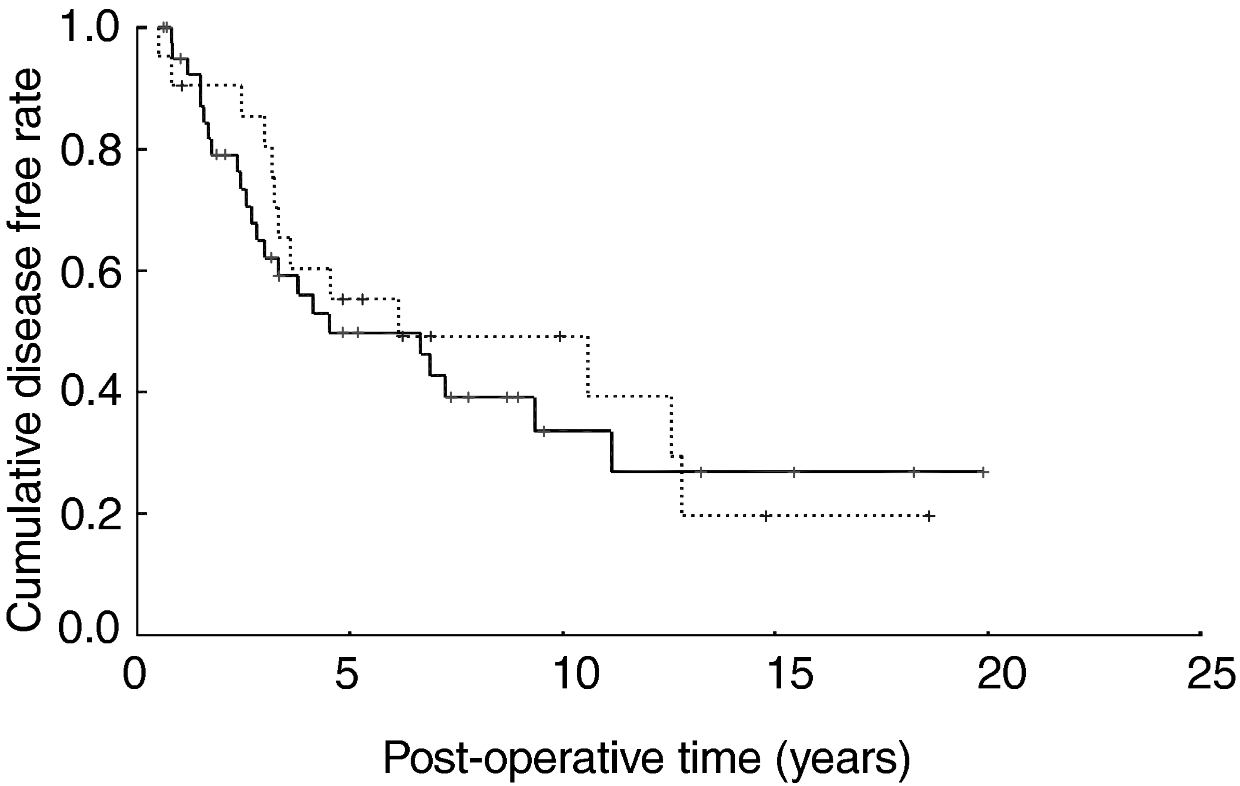


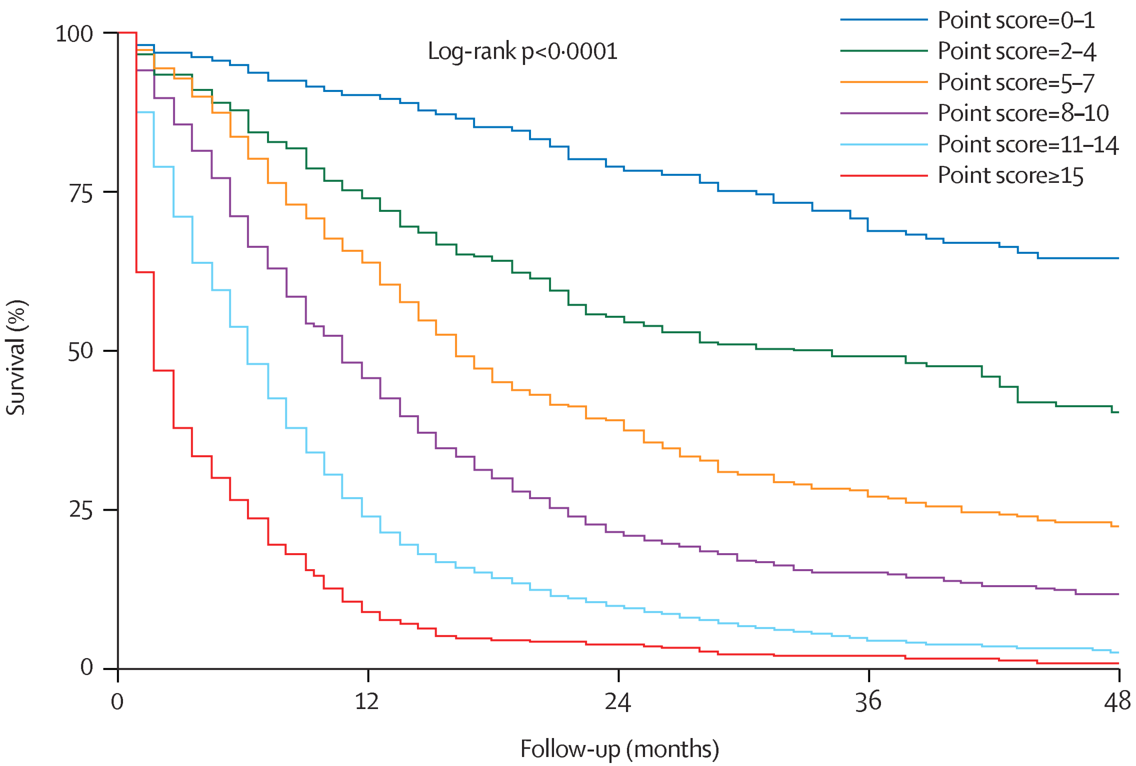
Figure 1 Kaplan-Meier nonrecurrent curves for short strictureplasty group (——) and long strictureplasty group (- - - - -). There is no significant difference between them (log rank test: $P = 0.702$).

- Blanchon et al. [6]:
  - ▷ Statistical Methods, p.831:

Statistical analyses

Mortality was used as the dichotomous outcome variable. Kaplan-Meier survival estimates were plotted over the follow-up period according to risk factors and were compared by the log-rank test. For multivariate analysis, a

▷ Figure 2, p.834:



Numbers at risk

0–1	162	146	130	119	107
2–4	210	158	120	105	90
5–7	372	245	154	109	91
8–10	590	284	142	93	79
11–14	948	253	105	55	34
≥15	515	54	21	10	8

Figure 2: **Kaplan-Meier survival curves from mortality in patients with NSCLC according to point score categories in development cohort**

Bibliography

[1] E. Cameron and L. Pauling. Supplemental ascorbate in the supportive treatment of cancer: re-evaluation of prolongation of survival times in terminal human cancer. *Proceedings of the National Academy of Science U.S.A.*, 75:4538–4542, 1978.

[2] D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway, and E. Ostrowski. *A handbook of small datasets*. Chapman & Hall, first edition, 1989.

[3] R. Peto, M.C. Pike, P. Armitage, N.E. Breslow, D.R. Cox, S.V. Howard, N. Mantel, K. McPherson, J. Peto, and P.G. Smith. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. *British Journal of Cancer*, 35:1–35, 1977.

[4] P.D. Allison. *Survival analysis using the SAS system: A practical guide*. NC: SAS Institute, 1995.

[5] T. Shatari, M.A. Clark, T. Yamamoto, A. Menon, C. Keh, J.Alexander-Williams, and M. Keighley. Long strictureplasty is as safe and effective as short strictureplasty in small-bowel crohn's disease. *Colorectal Disease*, 6:438–441, 2004.

[6] F. Blanchon, M. Grivaux, B. Asselain, et al. 4-year mortality in patients with non-small-cell lunc cancer: development and validation of a prognostic index. *Lancet Oncology*, 7:829–836, 2006.