Faculty of Health Sciences

Day 9: Survival analysis

Paul Blanche

Section of Biostatistics, University of Copenhagen

May 15, 2024

Disclaimer

"The readers are strongly encouraged encouraged to seek out and collaborate with statisticians with survival analysis expertise when considering time-to-event endpoints in their research."[1]

- Today's ILOs are ambitious but still, if you need to analyze survival data during your PhD, we encourage you to collaborate with a (competent) statistician and/or to follow the specific PhD course on this topic (7 full days).
- Today's topics are important to all as we will talk about the two most commonly used statistical methods in medical research: Kaplan-Meier and Cox regression[2]

[1] Le-Rademacher & Wang (2021). Time-to-event data: an overview and analysis considerations. Journal of Thoracic Oncology, 16(7), 1067-1074.

[2] According to citation impact, see. e.g. Van Noorden and Nuzzo (Nature News, 2014), in which the papers by Kaplan-Meier and Cox rank 11 and 24 in the ranking of the most most-cited research papers of all time.

Outline

Survival Data

- ILO: to recognize survival data and list contexts in which we meet them
- ILO: to define censoring and explain the challenges it creates
- ILO: to distinguish censoring from a competing risk

Simple & common analyses: possibilities and pitfalls

- ILO: to perform a Kaplan-Meier analysis and a log-rank test
- ILO: to fit and interpret a Cox model
- ILO: to list the main limitations of the Cox model
- ILO: to perform a Restricted Mean Survival Time (RMST) analysis
- ILO: to exemplify the difference between a risk ratio and a hazard ratio
- ILO: to recognize and avoid immortal time bias

Competing risks

- ILO: to exemplify competing risks data
- ILO: to describe a very common mistake
- ILO: to employ a basic (but appropriate!) method for competing risks data

What are time to event (survival) data?

"In many medical studies an outcome of interest is the time to an event. Such events may be adverse, such as death or recurrence of a tumour; positive, such as conception or discharge from hospital; or neutral, such as cessation of breast feeding. It is conventional to talk about survival data and survival analysis, regardless of the nature of the event." [3]

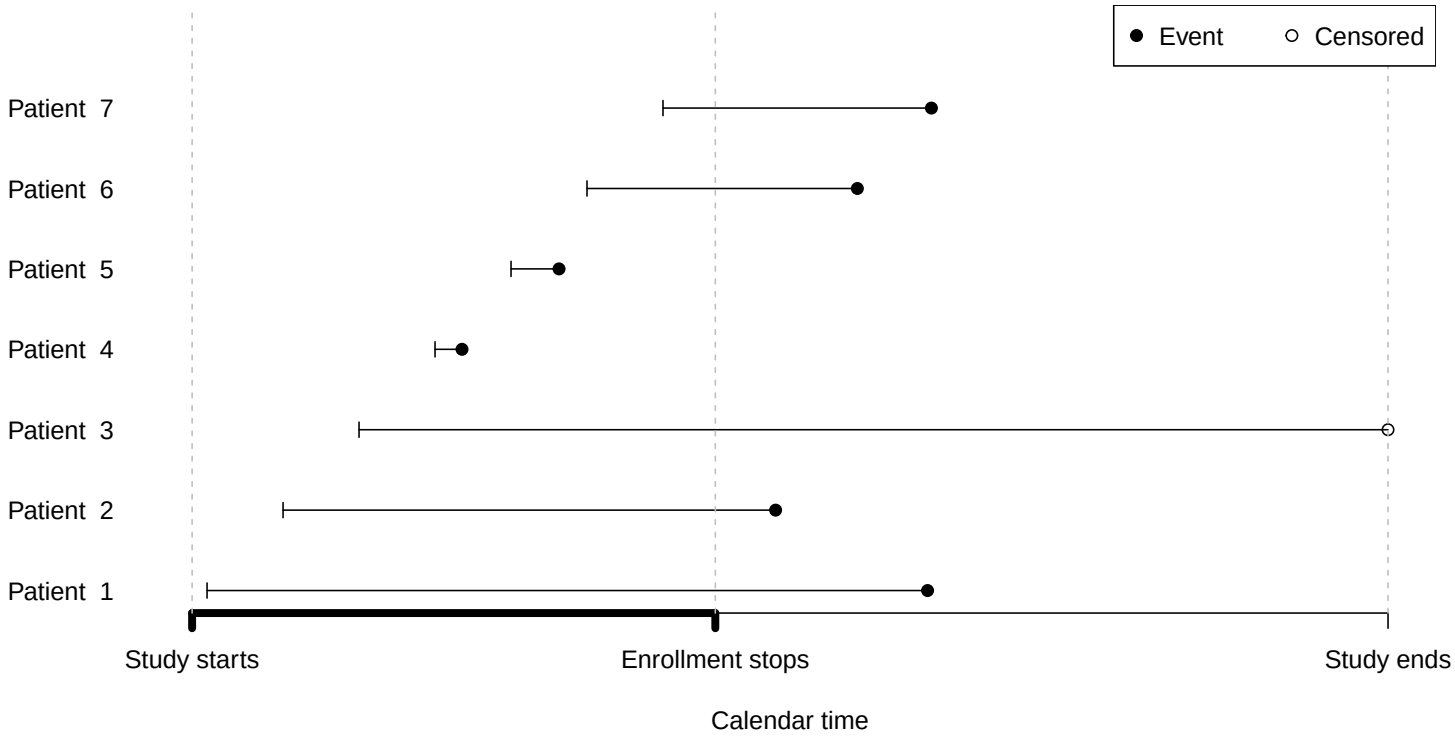
[3] Altman & Bland (1998). Time to event (survival) data. BMJ, 317(7156), 468-469.

What are censoring & censored data?

"The distinguishing feature of survival data is that at the end of the follow up period the event will probably not have occurred for all patients. For these patients the survival time is said to be censored, indicating that the observation period was cut off before the event occurred. We do not know when (or, indeed, whether) the patient will experience the event, only that he or she has not done so by the end of the observation period." [4]

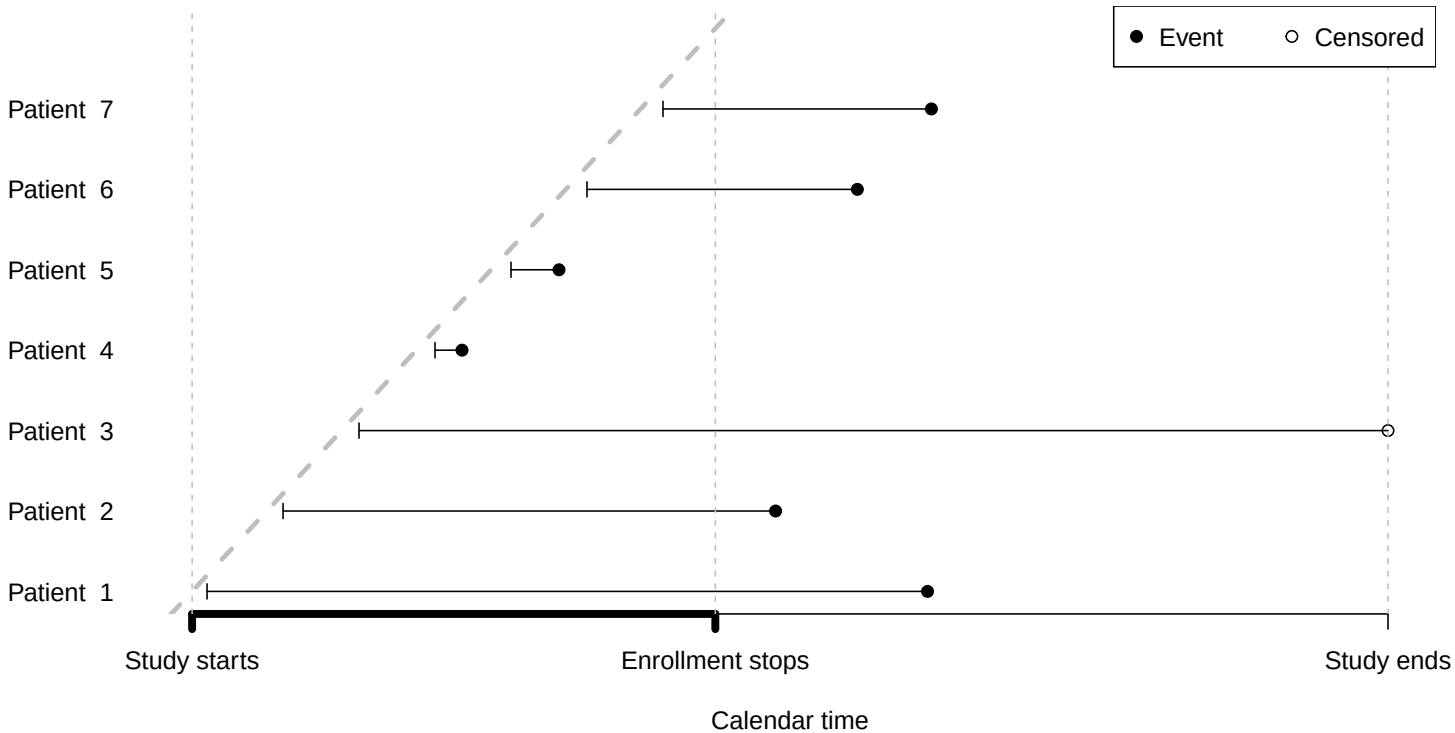
[4] Altman & Bland (1998). Time to event (survival) data. BMJ, 317(7156), 468-469.

Staggered entry & study end lead to censored data



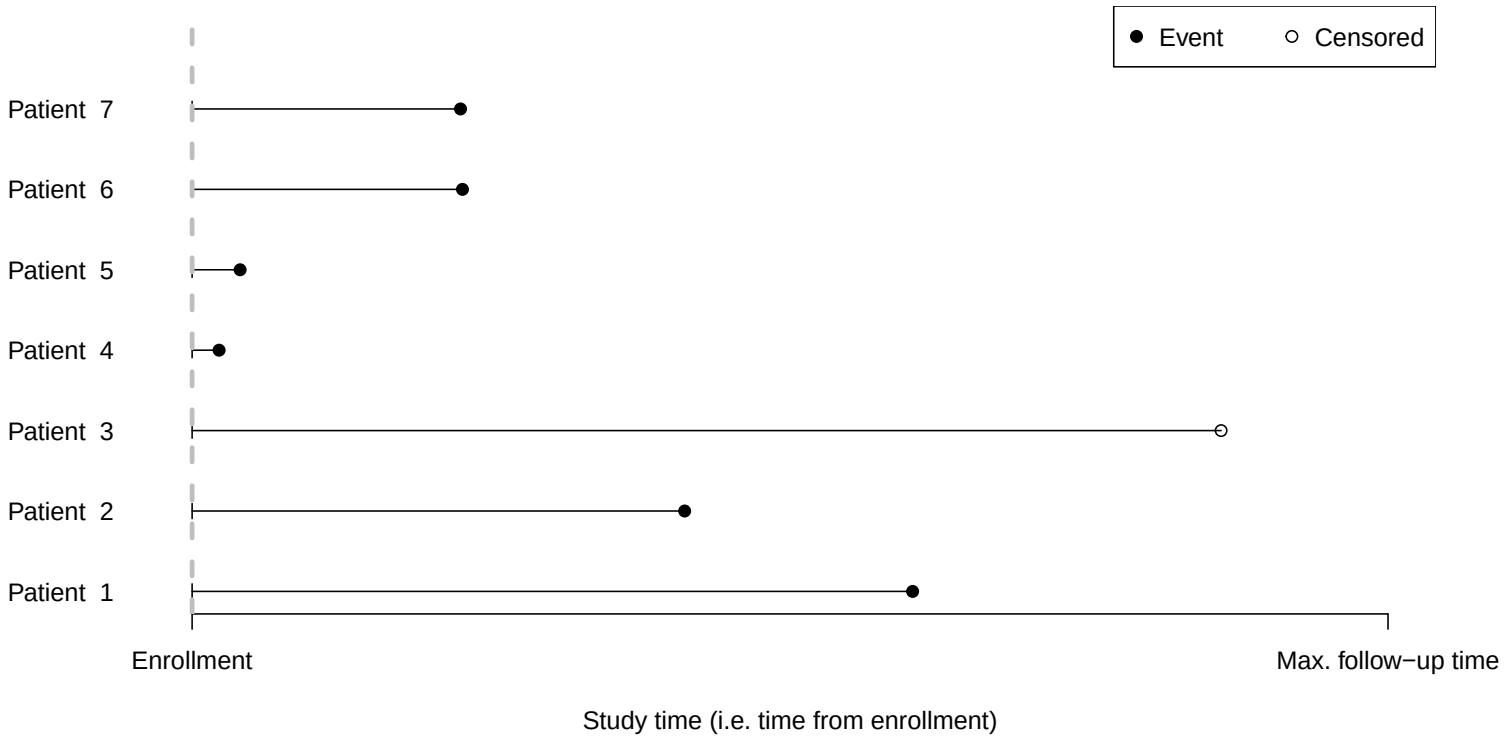
- often subjects enter the study during an accrual period, when e.g. they start a treatment, are diagnosed with something or undergo surgery;
- and are followed until a specific date, when follow-up ends.
- Those event-free at the end of follow-up are censored. We only know that the time to event is larger than the observed time from date of enrollment to date of end of follow-up.

From “calendar time” to “study time” (1/2)



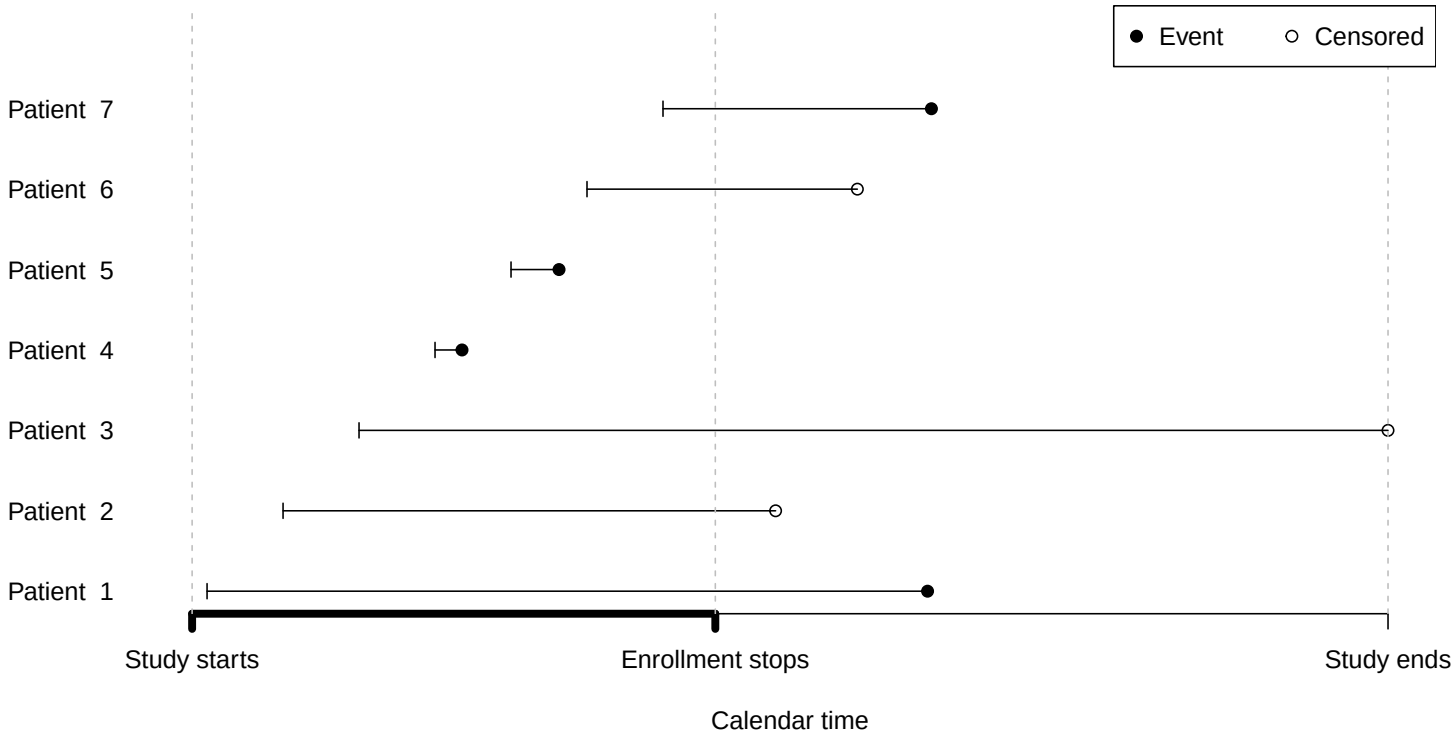
- Censored observations can happen at different “study times”, because subjects entered the trial at different dates.

From “calendar time” vs to “study time” (2/2)



- Censored observations can happen at different “study times”, because subjects entered the trial at different dates.

Dropout and censoring (1/2)

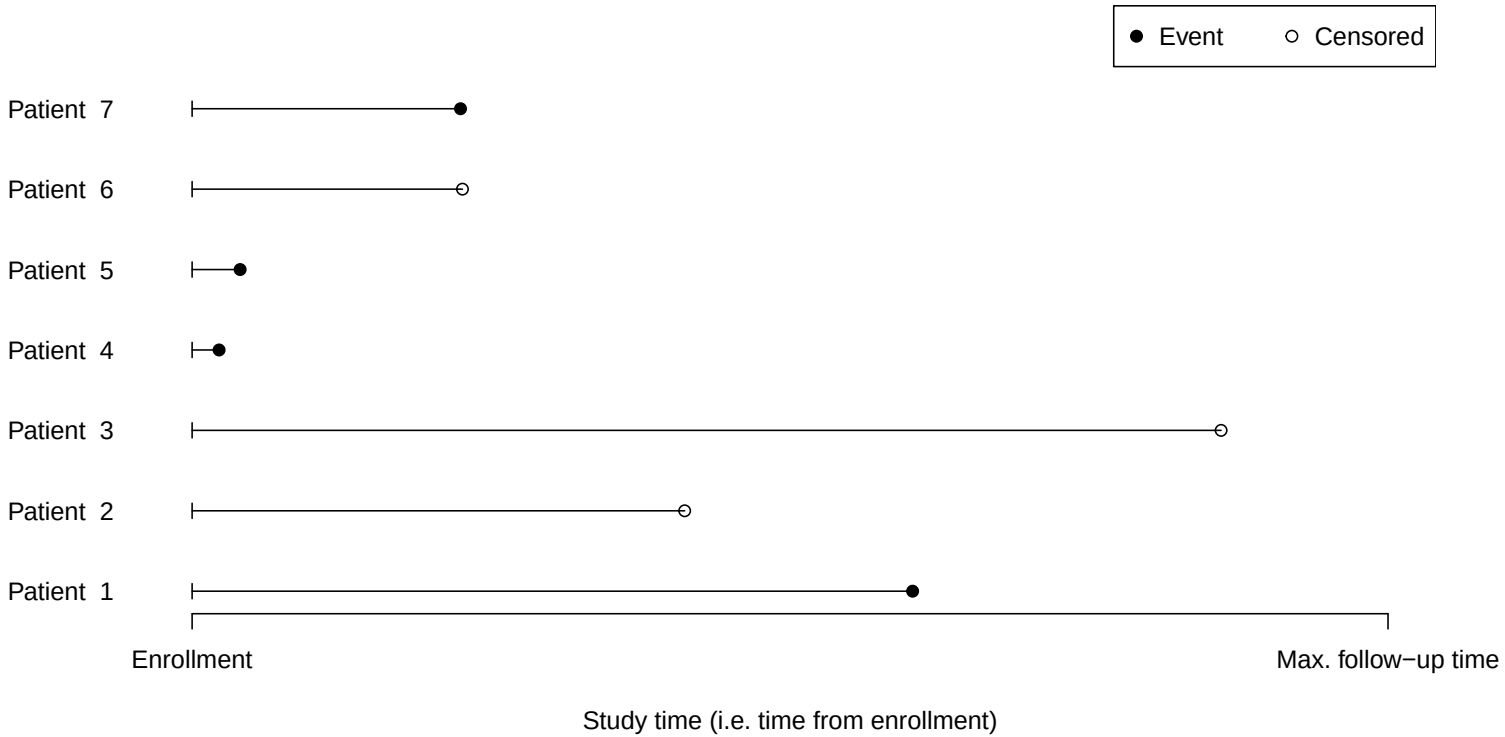


Often, some patients "drop out" or are "lost of follow-up" before the study ends (e.g. patients 2 and 6 here), e.g. because they:

- move to another region / hospital
- do no longer show up to biweekly follow-up visits (e.g. assume that the event is 50% decline in CD4 cell count from baseline or a positive drug test; here the time unit is half a week).

This results in no follow-up data after the dropout date, hence the time to the event of interest (e.g. death) is censored at that time.

Dropout and censoring (2/2)



Often, some patients "drop out" or are "lost of follow-up" before the study ends (e.g. patients 2 and 6 here), e.g. because they:

- move to another region / hospital
- do no longer show up to biweekly follow-up visits (e.g. assume that the event is 50% decline in CD4 cell count from baseline or a positive drug test: here the time unit is half a week).

This results in no follow-up data after the dropout date, hence the time to the event of interest (e.g. death) is censored at that time.

Definitions of time zero and event matter!

When talking about survival data and survival analysis, the time "zero" and the event of interest should always be clearly defined.

Examples of time "zero":

- start of treatment or randomization
- diagnosis (e.g. of cancer, diabetes)
- birth

Examples of events:

- death
- diagnosis (e.g. of cancer, diabetes)
- major adverse cardiovascular events (MACE), e.g. death, myocardial infarction or stroke.
- death or cancer progression (progression-free survival)

The two last are "composite endpoints".

Censoring $\neq$ competing risk !

Composite endpoints are popular because traditional/simple analyzes assume that there is no competing risks, which means that nothing else than the restricted follow-up (i.e. censoring) prevents us from observing the event of interest.

This is not the case if we study events such as "stroke" or "cancer progression", or non-fatal events among elderly people, because some subjects die without experiencing those.

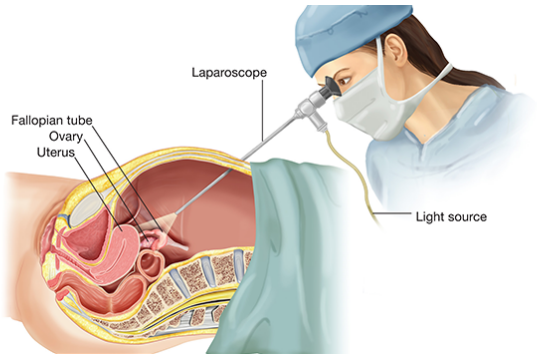
Note: some competing risks are not a problem because they are so rare that they are irrelevant. E.g., if we study time to pregnancy within 2 years, it seems safe to ignore the competing risk of death.

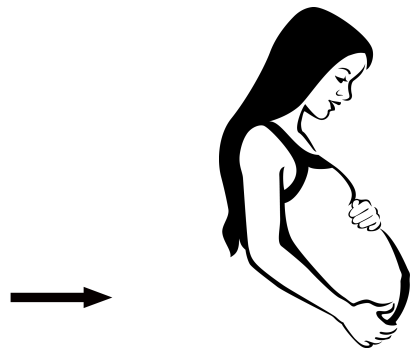
Case: pregnancy in subfertile women

Data, n=38:

time	status
4	0
4	1
9	1
2	1
1	1
24	0

The usual censored survival data format!





Outcome:

- time: time in months (menstrual cycles) since laparoscopy
- status: pregnancy (1) or no pregnancy by the end of the follow-up (0)

Research aim: estimate/describe the proportion women not pregnant yet, by any given time after laparoscopy.

Reference: Bland, J. M., & Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). Bmj, 317(7172), 1572-1580.

Censored data: the challenge!

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0

What is the estimated probability of a woman not being pregnant yet 1 month after laparoscopy?

Censored data: the challenge!

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0

What is the estimated probability of a woman not being pregnant yet 1 month after laparoscopy?

- Easy! $1 - 6/38 = 84\%$

Censored data: the challenge!

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0

What is the estimated probability of a woman not being pregnant yet 1 month after laparoscopy?

- Easy! $1 - 6/38 = 84\%$

And 2 months after laparoscopy?

- Easy! $1 - 11/38 = 71\%$

Censored data: the challenge!

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0

What is the estimated probability of a woman not being pregnant yet 1 month after laparoscopy?

- Easy! $1 - 6/38 = 84\%$

And 2 months after laparoscopy?

- Easy! $1 - 11/38 = 71\%$

And 3 months after laparoscopy?

Censored data: the challenge!

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0

What is the estimated probability of a woman not being pregnant yet 1 month after laparoscopy?

- Easy! $1 - 6/38 = 84\%$

And 2 months after laparoscopy?

- Easy! $1 - 11/38 = 71\%$

And 3 months after laparoscopy?

- Not easy! We don't know whether woman number 9 became pregnant during the third month. But we do know that she was not pregnant yet at the end of the second month... How should we use this "incomplete", but not "completely missing", information?

Outline

Survival Data

- ILO: to recognize survival data and list contexts in which we meet them
- ILO: to define censoring and explain the challenges it creates
- ILO: to distinguish censoring from a competing risk

Simple & common analyses: possibilities and pitfalls

- ILO: to perform a Kaplan-Meier analysis and a log-rank test
- ILO: to fit and interpret a Cox model
- ILO: to list the main limitations of the Cox model
- ILO: to perform a Restricted Mean Survival Time (RMST) analysis
- ILO: to exemplify the difference between a risk ratio and a hazard ratio
- ILO: to recognize and avoid immortal time bias

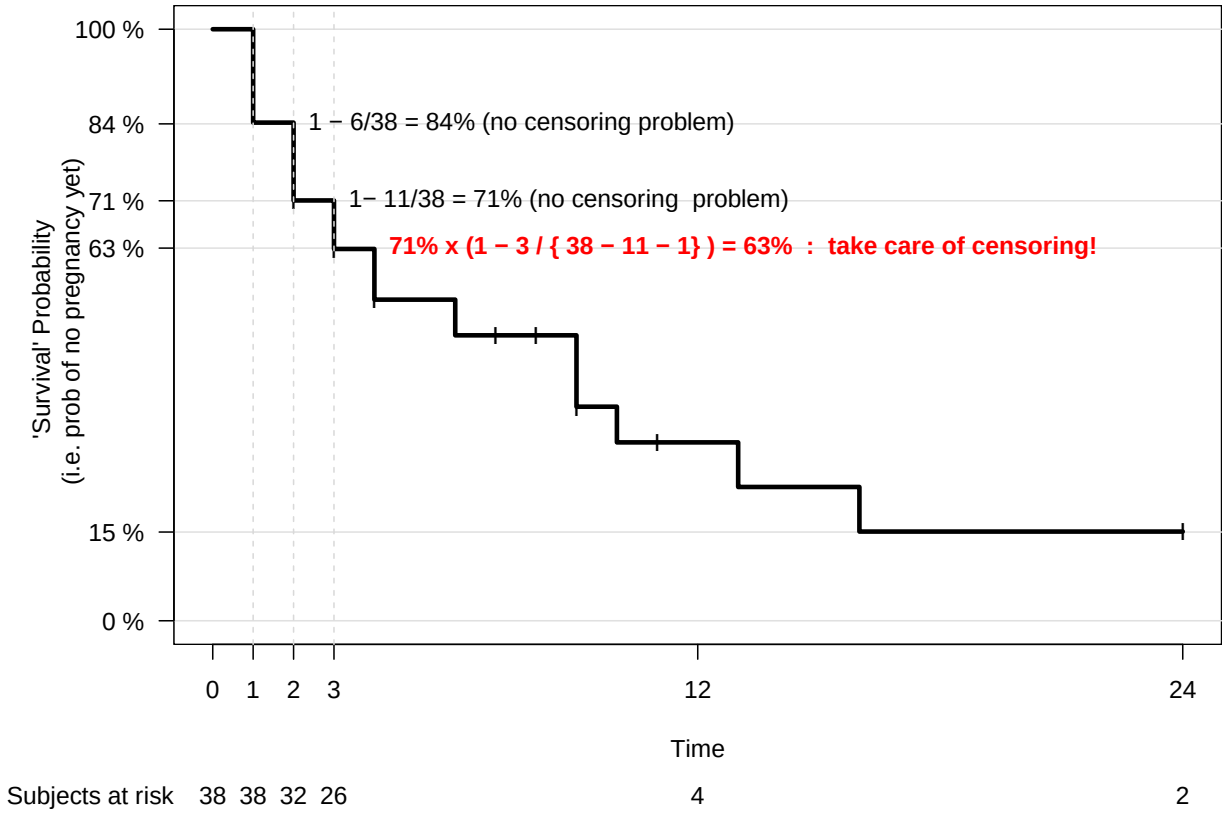
Competing risks

- ILO: to exemplify competing risks data
- ILO: to describe a very common mistake
- ILO: to employ a basic (but appropriate!) method for competing risks data

Kaplan-Meier: main idea

Data ordered by time, n=38:

id	time	status
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	2	1
8	2	1
9	2	0
10	2	1
11	2	1
12	2	1
13	3	1
14	3	1
15	3	0
16	3	1
17	4	0



$$\widehat{P}(T>3) = \underbrace{\widehat{P}(T>3|T>2)}_{=1-3/26} \times \underbrace{\widehat{P}(T>2)}_{=71\%} = 63\%$$

Here $\widehat{P}(T>3|T>2) = 1-3/26$ because at the third month we observe 3 pregnancies among the 26 women for whom we can observe a pregnancy. These women are called women "at risk". Note that 26 (at risk) = 38 (all) - 11 (already pregnant) - 1 (already lost of follow-up).

Kaplan-Meier: main assumption (our case, 1/2)

To be valid, the computation requires that the women lost of follow-up by any time point are not different from those who remain in the study.
In other words, the "at risk" women need to be representative of all those who have not experienced the event yet.

This means that the women lost of follow-up should have the same chances of becoming pregnant that those who remain in the study.

- Realistic if e.g. loss of follow-up is due to study end and staggered entries (especially if the accrual period is short).
- Not realistic if e.g. loss of follow-up is due to women dropping out because they have received a new diagnosis of infertility that discourages them from further trying to become pregnant.

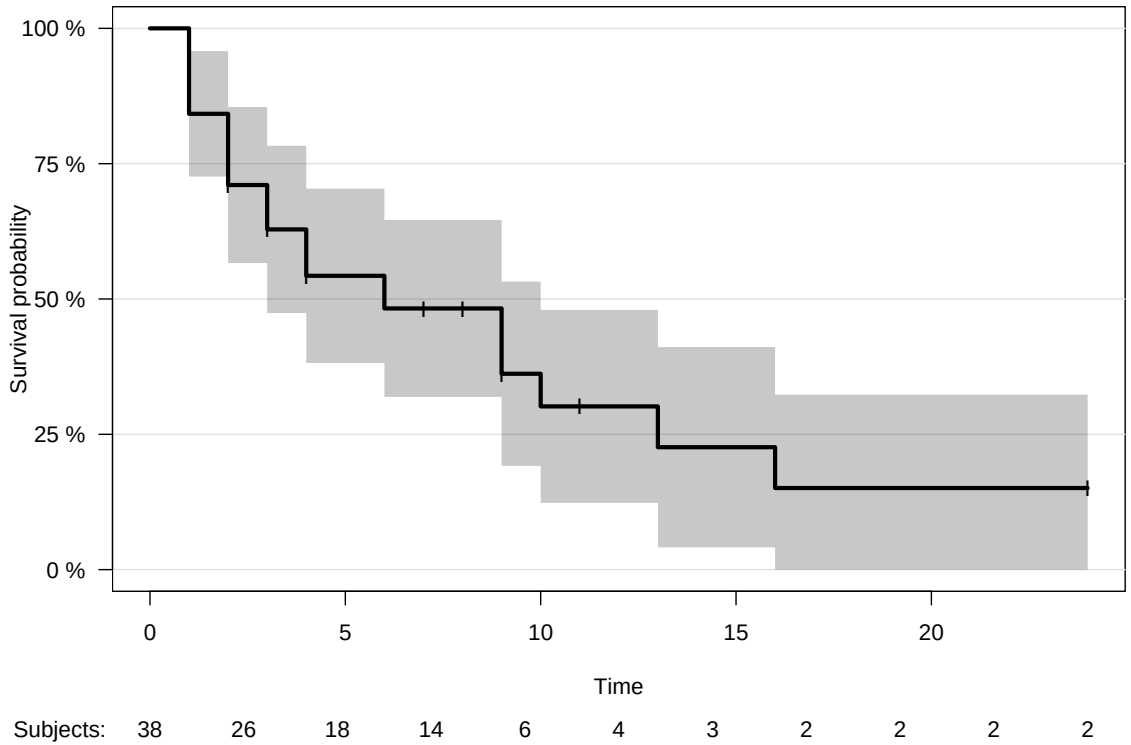
Kaplan-Meier: main assumption (in general, 2/2)

The "exact" definition and meaning of independent censoring is complex.

But, the concept and interpretation of independent right censoring is essentially that among those who are still alive, additional information of being uncensored should provide no further insight into the future risk of event.[5]

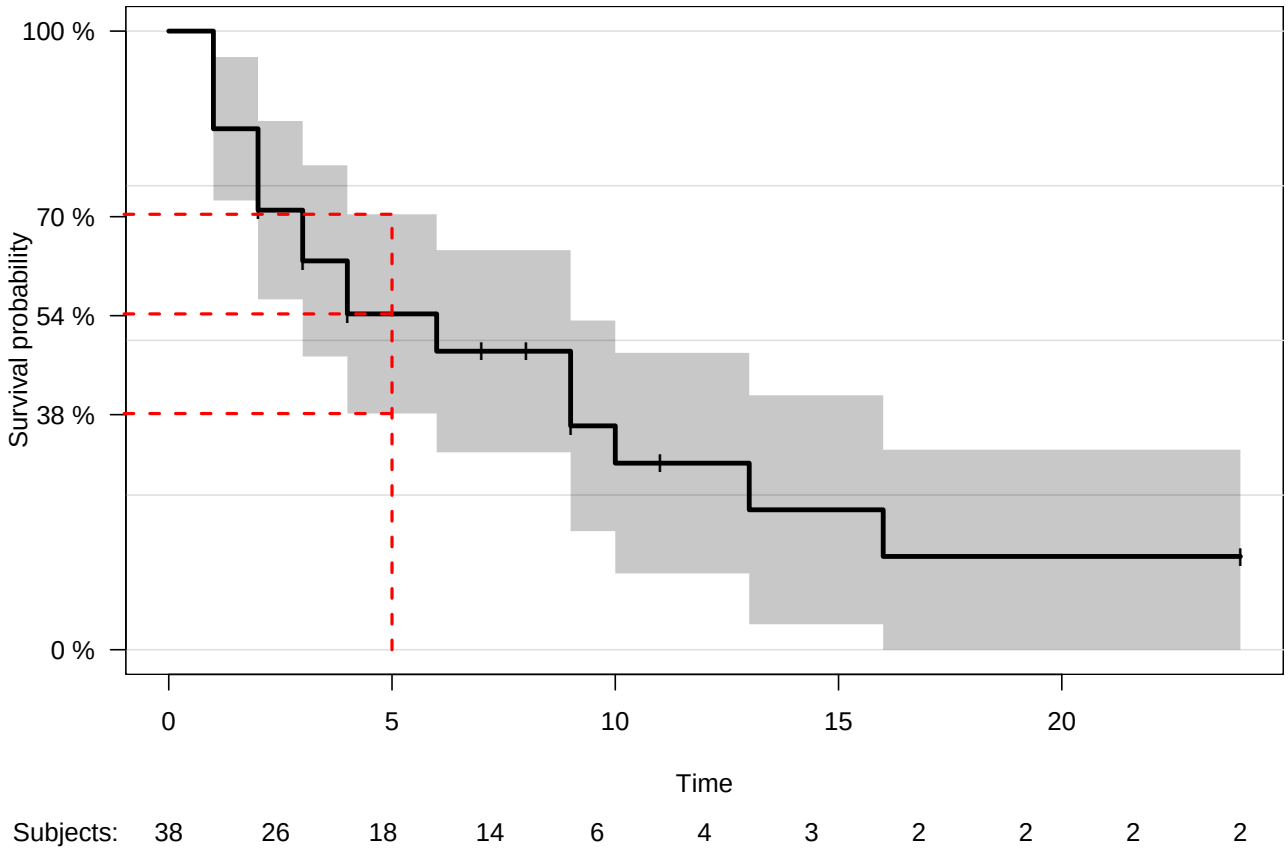
[5] See e.g. Andersen, P. K., & Keiding, N. (2012). Statistics in Medicine, 31(11-12), 1074-1088.

Typical KM plot (same pregnancy data)



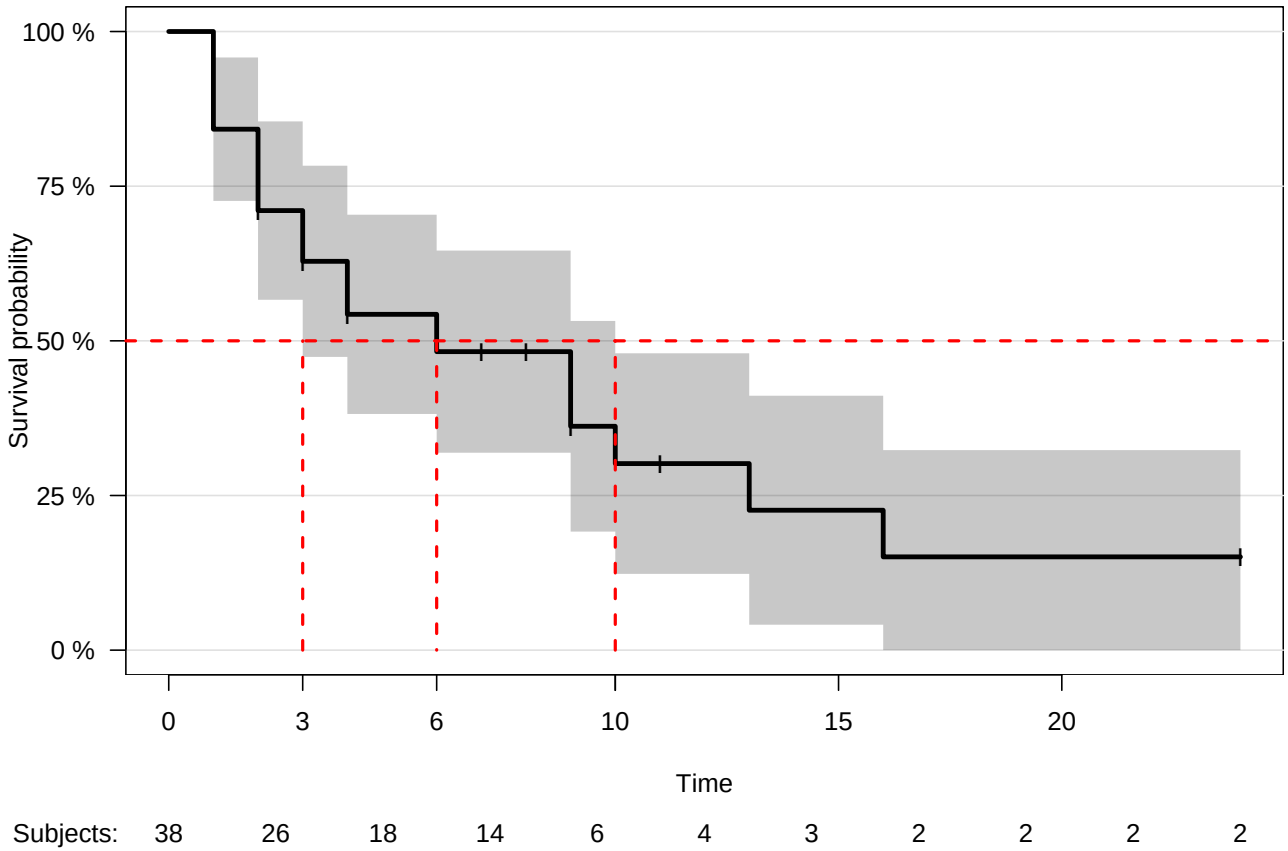
- 95% CI, "ticks" to display the censored times, N. of subjects at risk.
- Does not go down to 0 when the largest observed time is a censored observation (i.e., `status`=0).
- Plot stops where there is no longer subjects at risk (here 24 months)

KM plot Interpretation



- The survival probability at t=5 months is estimated to be 54% (95%-CI= [38%;70%]).
- Similar results can be read for any time.

Median survival time

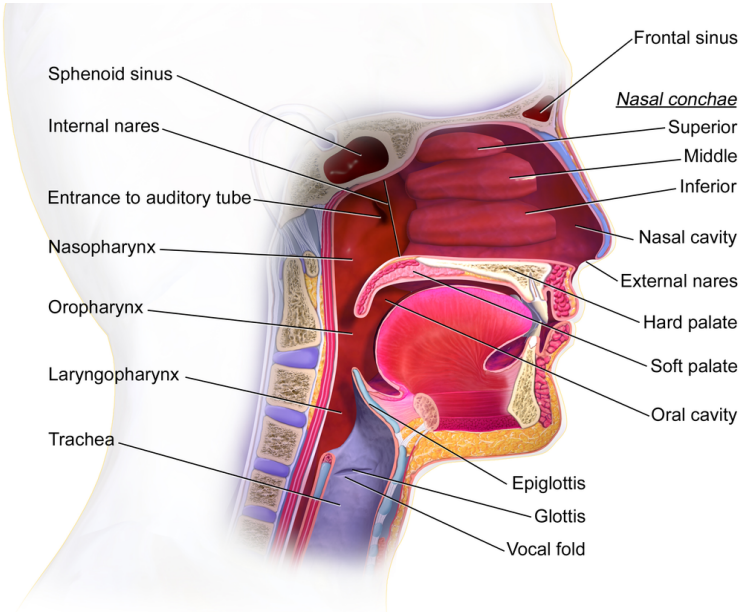


- The median survival time is a parameter often meaningful to estimate; its interpretation is that half of the time to events are shorter.
- The median survival time is estimated to be 6 months (95%-CI= [3;10]).
- Here, we estimate that half of the women become pregnant within 6 months after laparoscopy (95%-CI= [3;10]).

Case: carcinoma randomized clinical trial data

Data, $n = 100 + 95$ (experimental + standard):

	time	status	trt
1	631	1	1
2	270	1	0
3	327	1	1
4	243	1	1
5	916	1	0
6	1823	0	0



The Upper Respiratory System

Research question: Does an experimental treatment of carcinoma of the oropharynx, which combines radiotherapy and chemotherapy, improve survival chances, as compared to standard radiotherapy treatment?

Reference/source: Kalbfleisch and Prentice, *The statistical analysis of failure time data*, 2002, Appendix II.

Which null hypothesis?

We need to be specific about what we mean, when we aim to compare "survival chances". Two simple and common approaches and their corresponding null hypothesis ($\mathcal{H}_0$) are:

- Comparison at a specific time point t, e.g., t=2 years.

$$\mathcal{H}_0: \quad S_1(t) = S_2(t) \quad ,$$

  meaning that the survival chances at t years are the same. For the interpretation, this is similar to the 2x2 table case of Lecture 5.
  $\rightarrow$ the choice of t should be prespecified and justified.

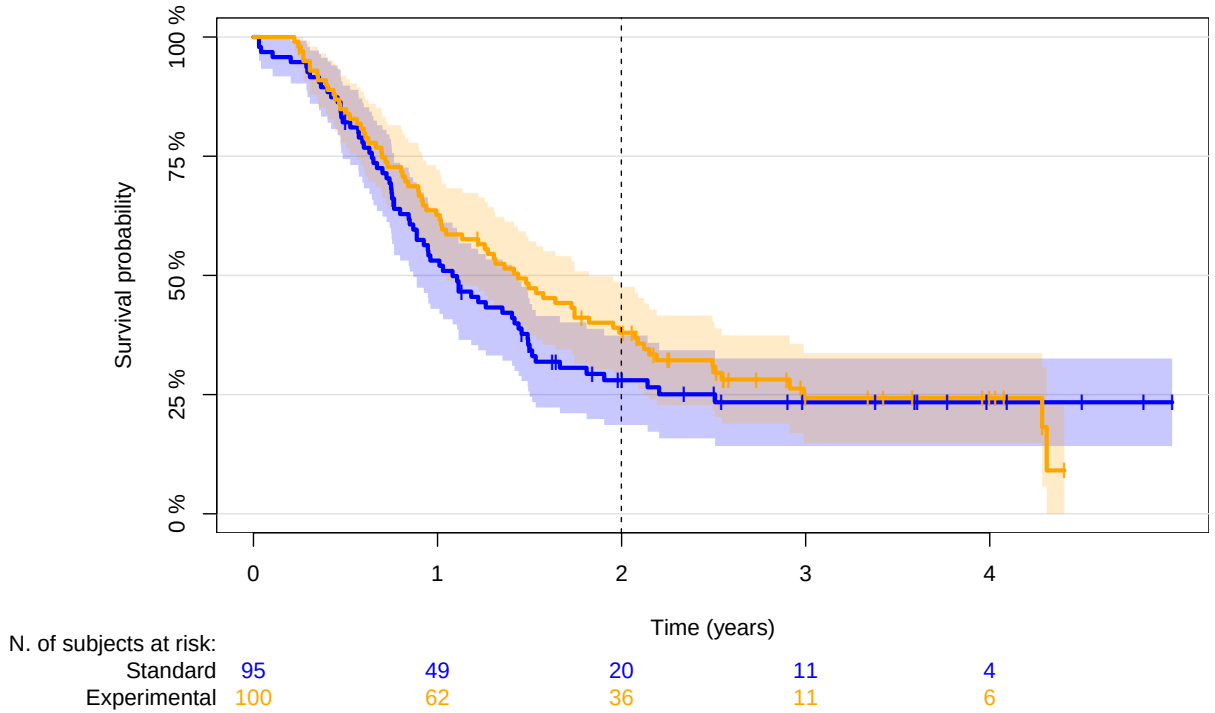
- Comparison of the survival curves "overall".

$$\mathcal{H}_0: \quad S_1(t) = S_2(t) \quad \text{for all time } t,$$

  meaning that the survival curves are the same (everywhere). Often useful when we expect than one curve is above the other, i.e. $S_1(t) > S_2(t)$ for all time t.
  $\rightarrow$ most common choice, for which a log-rank test[6] is useful.

[6] See e.g. Bland & Altman. "The logrank test." (2004) BMJ 328(7447), 1073.

Pointwise comparison at $t = 2$ years

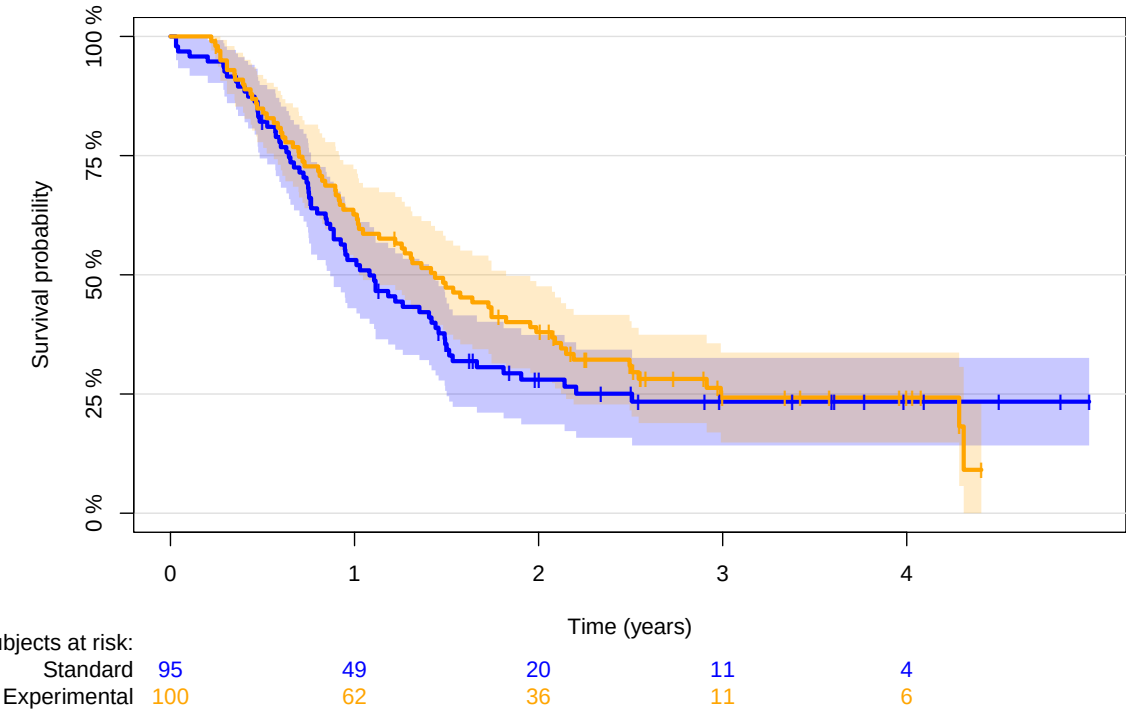


- In each group:

  38.0% (95%-CI=[28.3, 47.6]) versus 28.0% (95%-CI=[18.6, 37.4])

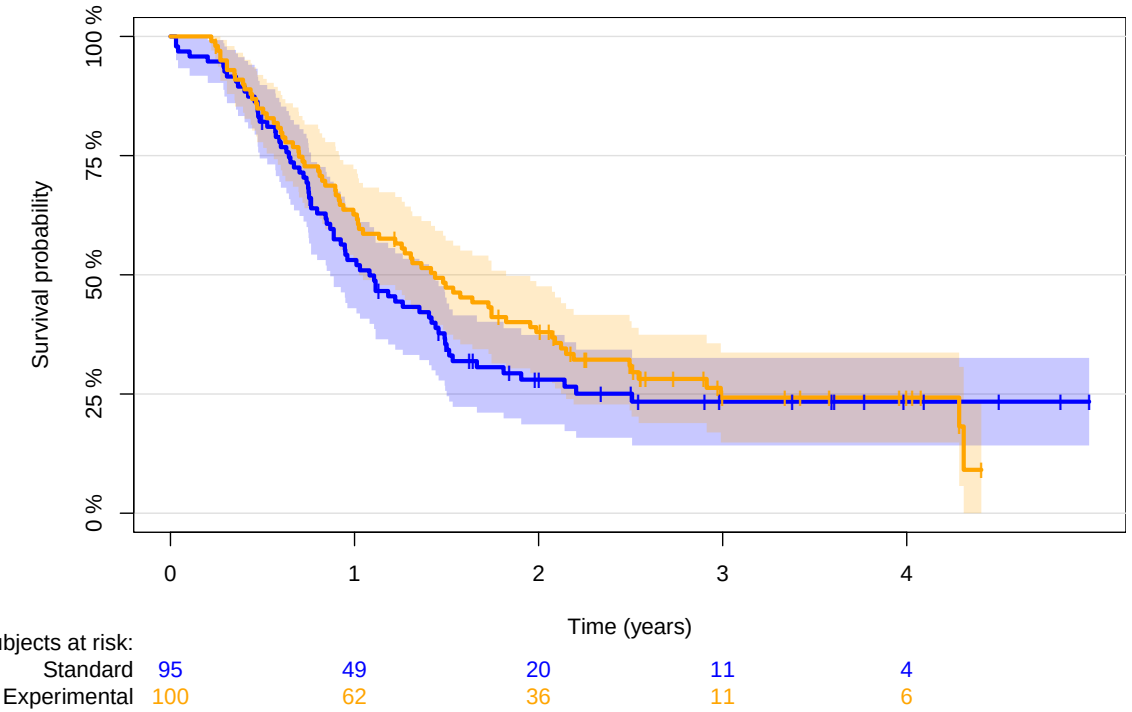
- The survival difference is 10.0% (95%-CI=[-3.5, 23.4], p-value=0.146)
- See R-demo for computation.

Log-rank test to compare the survival curves



- The p-value of the log-rank test is p=0.336. We do not reject the null hypothesis that the survival curves are the same in the two treatment groups.

Log-rank test to compare the survival curves



- The p-value of the log-rank test is p=0.336. We do not reject the null hypothesis that the survival curves are the same in the two treatment groups.
- See R-demo for computation.
- But can we provide a matching 95%-CI of an effect size? (see next slides)

Hazard function (aka hazard rate function)

The log-rank test actually compares the survival curves in each group via the comparison of the hazard functions.

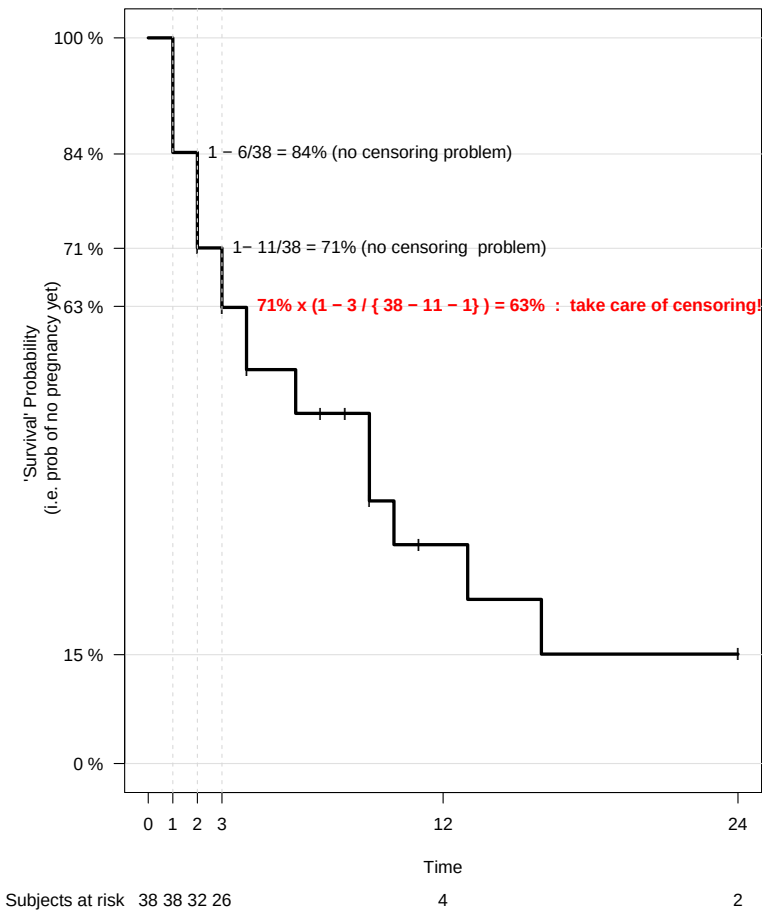
The hazard function is: $\lambda(t) = \lim_{dt \to 0} P(t \leq T < t + dt \,|\, T \geq t)/dt$.

Informally, for any (very) short time duration dt, the hazard $\lambda(t)$ multiplied by this small duration dt is the probability of death before time $t + dt$, given that a subject has survived until time t.

Knowing the hazard function is equivalent to knowing the survival function, as $S(t) = \exp\big(-\int_0^t \lambda(u)du\big)$, which explains why comparing the hazard functions is equivalent to comparing the survival curves.

Why using hazards?

Numerous statistical methods and concepts in survival analysis rely on the hazard function. This is because censoring is "easily" accounted for when estimating the hazard function, e.g. as in Kaplan-Meier computation.



Remember from previous slide

$$\widehat{P}(T > 3) = \widehat{P}(T > 3|T > 2) \times \widehat{P}(T > 2)$$

and

$$\begin{aligned}\widehat{P}(T > 3|T > 2) &= 1 - \widehat{P}(T = 3|T > 2) \\ &= 1 - \widehat{P}(3 \leq T < 4|T \geq 3) \\ &= 1 - \widehat{\lambda}(t) \times dt\end{aligned}$$

where $t = 3$ and $dt = 1$ and

$$\widehat{\lambda}(t) = \frac{\text{N. of events at time } t}{\text{N. of subjects at risk at time } t}$$

which gives $\widehat{\lambda}(3) = 3/26$.

Cox model and hazard ratios (univariate model)

The (univariate) Cox model is a popular model for the hazard function.

$$\lambda(t|X) = \lambda_0(t)\exp(X\beta)$$

- $\lambda_0(t)$ is called the baseline hazard function and nothing specific is assumed about this component of the model.
- $\exp(\beta)$ is interpreted as a hazard ratio (HR), since

$$\frac{\lambda(t|X=x+1)}{\lambda(t|X=x\quad)} = \frac{\lambda_0(t)\exp(\ (x+1)\beta\ )}{\lambda_0(t)\exp(\ x\beta\ )} = \exp(\beta)$$

- $\beta = 0 \Leftrightarrow \exp(\beta) = 1$ means 'no effect', i.e., that the survival curves are the same whatever the covariate value $X = x$.
- $\left\{ \begin{array}{c} \beta < 0 \\ \beta > 0 \end{array} \right\}$ means $\left\{ \begin{array}{c} \text{higher} \\ \text{lower} \end{array} \right\}$ survival when x increases.
- An important assumption is the hazard ratio (HR) does not depend on t. This is the so-called he proportional hazards assumption.

Example: Cox regression, one binary variable (1/3)

Research question: Does an experimental treatment of carcinoma of the oropharynx, which combines radiotherapy and chemotherapy, improve survival chances, as compared to standard radiotherapy treatment?

Statistical model: we model the hazard function as

$$\lambda(t|X = x) = \lambda_0(t) \exp(\beta x) \ ,$$

where

$$x = \begin{cases} 1 & \text{for experimental treatment} \\ 0 & \text{for standard treatment} \end{cases} \ .$$

Cox regression with one binary variable (2/3)

R code:

```
library(survival)
cox1 <- coxph(Surv(time,status)~trt,carcinoma)
summary(cox1)
```

Output:

```
Call:
coxph(formula = Surv(time, status) ~ trt, data = carcinoma)

  n= 195, number of events= 142

       coef exp(coef) se(coef)      z Pr(>|z|)
trt1 -0.1622    0.8503   0.1685 -0.963    0.336

     exp(coef) exp(-coef) lower .95 upper .95
trt1    0.8503      1.176    0.6111     1.183

Concordance= 0.53  (se = 0.023 )
Likelihood ratio test= 0.92  on 1 df,   p=0.3
Wald test            = 0.93  on 1 df,   p=0.3
Score (logrank) test = 0.93  on 1 df,   p=0.3
```

Example: Cox regression, one binary variable (3/3)

- `coef`: the log of the estimated hazard ratio, $\hat{\beta} = -0.1622$.
- `exp(coef)`: the hazard ratio, $\exp(\hat{\beta}) = 0.8503$.
- `lower .95` and `upper .95`: 95% confidence interval for the hazard ratio, 95-CI=[0.61,1.18].
- `Pr(>|z|)`: p-value for the null hypothesis $\boxed{H_0 : \beta = 0}$ or equivalently $\boxed{H_0 : \exp(\beta) = 1}$, i.e., no treatment effect. Here p=0.336.
- the p-value is non significant and (equivalently) the confidence interval of the hazard ratio does not include 1.
- `Score (logrank) test`: provide a p-value for the log-rank test, in this specific case where there is only one categorical variable X in the model.[7]

[7] In theory, it should match that of the `survdiff` function. However, it can be slightly different, due to minor differencies in the computation. Usually it does not matter when we round the results as appropriate, e.g. at the third digit. Note, however, that the number of digits presented by default in this output can be too small (here only one digit!).

How to interpret $HR = 0.85$?

When comparing two subjects alive after any t days since treatment initiation (e.g. after 1, 10, 100 or 1000 days[8]), we estimate that the risk of death **within the next day**[9] is 0.85 times lower for patients treated with the experimental treatment than for those treated with the standard treatment. That is, we estimate that the instantaneous risk of death is reduced by 15% $(= 1 - 0.85)$.

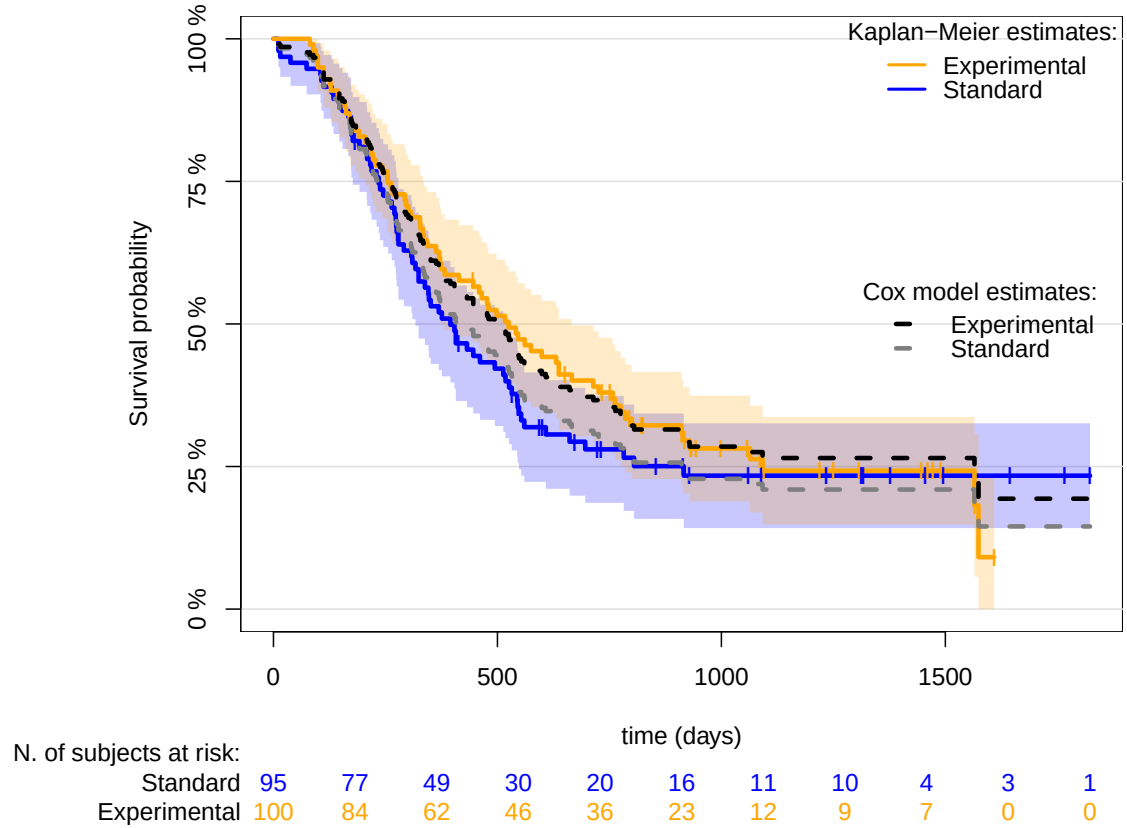
However, the 95% confidence interval tells us that we cannot rule out that the instantaneous risk is reduced by as much as 39% or, on the contrary, increased by up to 18%.

[8] Can be any day up to the maximum follow-up time, here 1823 days.

[9] Here we consider that dt =1 day is a 'small enough' duration for interpretation.

Simple check of the model assumption

Reminder: an important assumption is the hazard ratio (HR) does not depend on t (proportional hazards assumption).



For a univariate Cox model with a binary covariate, we can graphically check the 'model fit' (and so the single modeling assumption) by comparing the survival curves estimated by the Cox model to those obtained via Kaplan-Meier (which makes no modeling assumptions). If the model fit is good, the curves estimated by both approaches should be "close" (approximately, as there is some sampling variability, as shown by 95%-CIs).

Hazard ratio vs Risk ratio (1/3)

We have just seen that we can correctly interpret:

- the hazard ratio (HR) as the ratio of two instantaneous risks.
- 1-HR as a relative reduction in instantaneous risk (e.g. 15%).

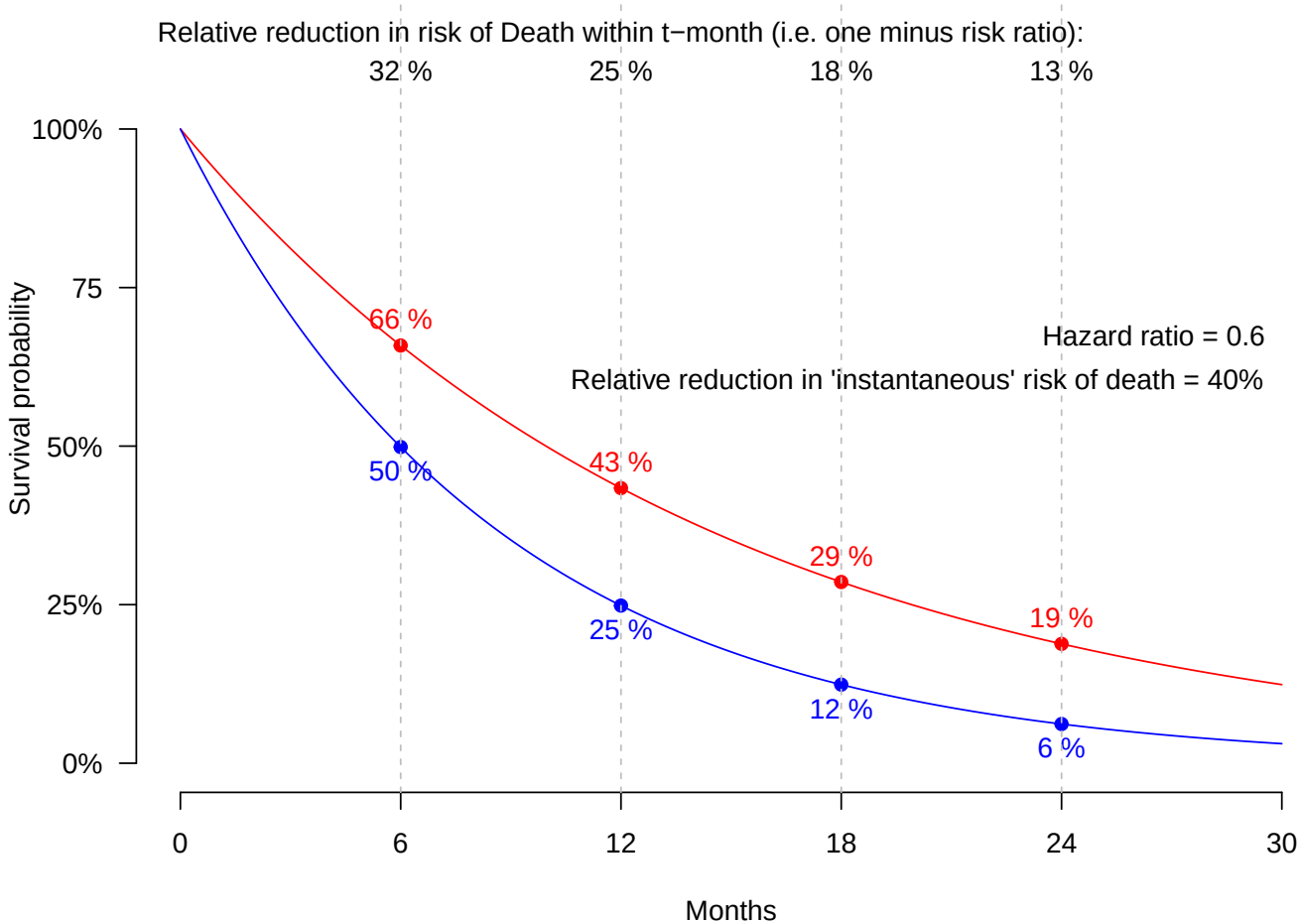
However, **we CANNOT safely interpret**:

- the hazard ratio (HR) as the ratio of two "long term" risks.
- 1-HR as a relative reduction in "long term" risk.

This is, however, a common misunderstanding. [10] The misunderstanding might come from the fact that some researchers talk about "risks" without clarifying whether they mean 1 day, 1 year or 10 years risks and forget that this is not the same! Also, for historical reasons, hazard ratios are often referred to as a "relative risks", which is a bit confusing.

[10] See e.g. Sutradhar & Austin, (2018). Relative rates not relative risks: addressing a widespread misinterpretation of hazard ra... Annals of epidemiology, 28(1), 54-57; or Sashegyi and Ferry. "On the interpretation of the hazard ratio and communication of survival benefit." The oncologist 22.4 (2017): 484-486.

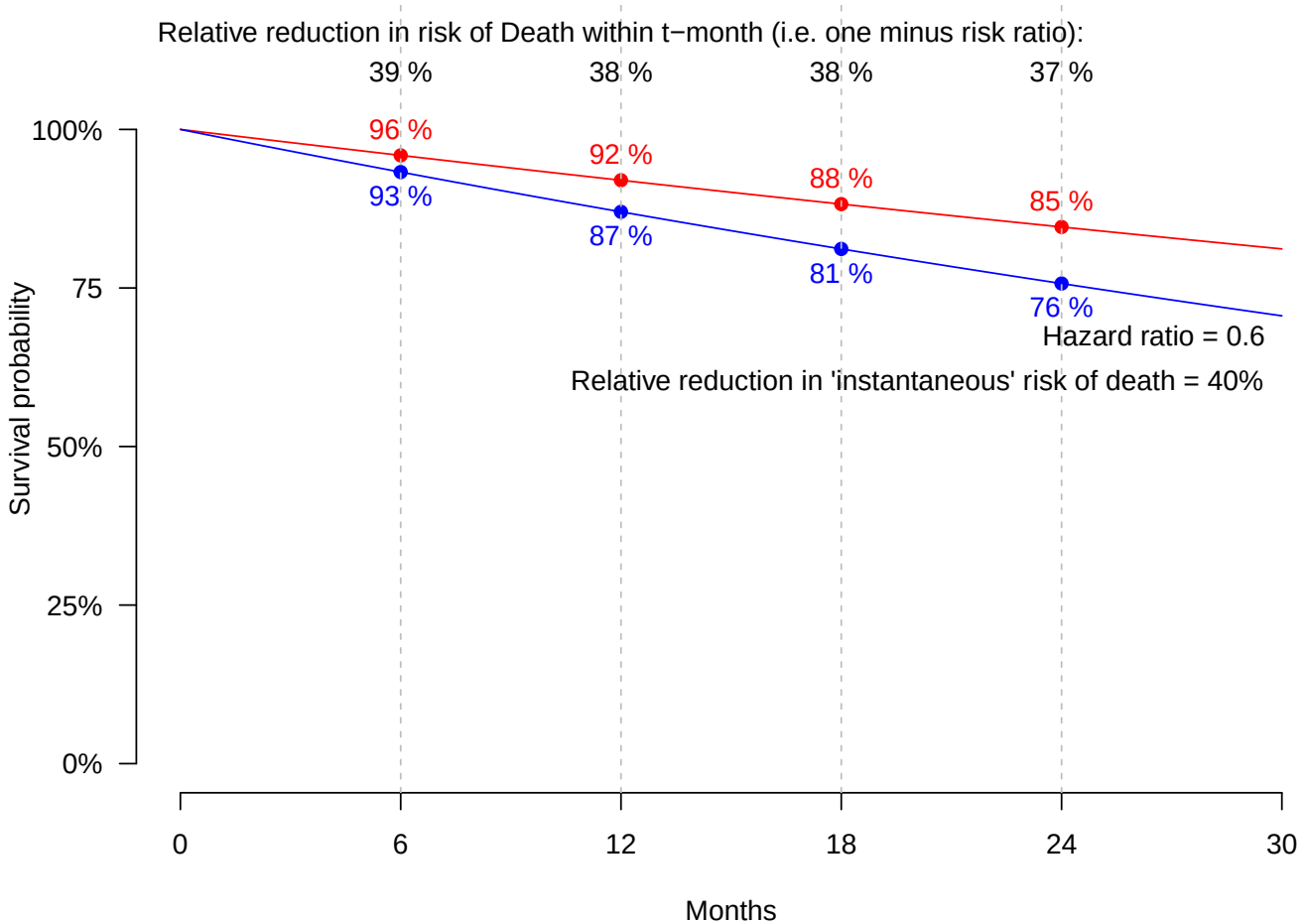
Hazard ratio vs Risk ratio (2/3)



The relative reduction in "long term" risk can be very different from one minus the hazard ratio (e.g. 13% vs 40%), especially when the risks are large! [11] Here, $13\% = 1 - \frac{1-0.19}{1-0.06}$.

[11] Same example in Sashegyi and Ferry. "On the interpretation of the hazard ratio and communication of survival benefit." The oncologist 22.4 (2017): 484-486.

Hazard ratio vs Risk ratio (3/3)



The relative reduction in "long term" risk can be less different when the risks are smalls.

A limitation of hazard ratios and an alternative

When using a Cox model, we (heavily) rely on the proportional hazards assumption.

Sometimes it does not make sense and sometimes we simply wish to use an alternative "non-parametric" method wich does not rely on any assumption about how the two survival curves might differ.

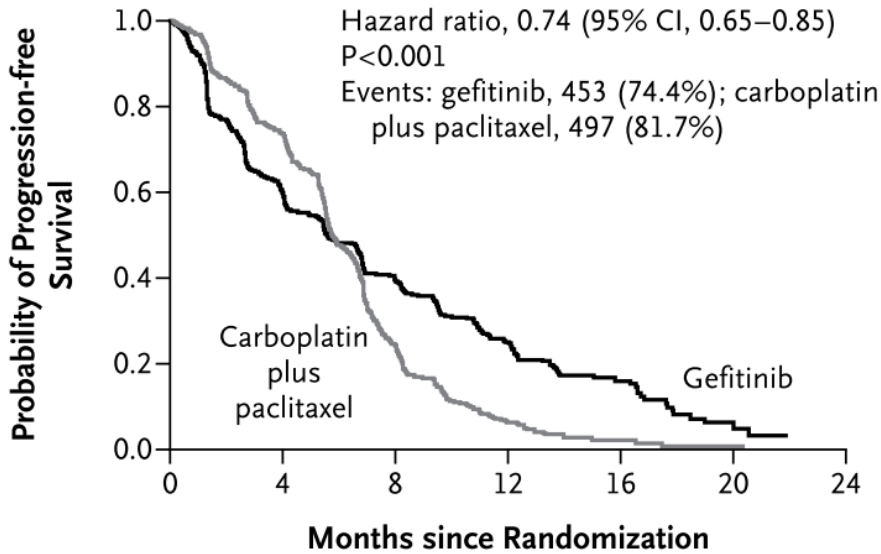
For instance, the proportional hazards assumption implies that the (true) survival curves cannot cross, which might be thought as a "strong" arbitrary assumption in some contexts.

Instead of systematically planning to report hazard ratios when comparing two survival curves, it is now increasingly recommended to also consider reporting differences in restricted mean survival times.[12]

[12] See e.g. Uno et al. "Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis." Journal of clinical Oncology 32.22 (2014): 2380. See also, e.g., the statistical reporting guidelines of the New England Journal of Medicine, available from the website of the journal.

Crossing survival curves (expected or observed)

- Is the hazard ratio meaningful to summarize the difference in survival chances over time shown below? Very questionable...
- The difference in restricted mean survival times is an attractive alternative



No. at Risk							
Gefitinib	609	363	212	76	24	5	0
Carboplatin plus paclitaxel	608	412	118	22	3	1	0

Figure 2.A in Mok et al. "Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma." N. Engl. J. Med. 361.10 (2009): 947-957.

Restricted Mean Survival Time (RMST)

Why a "restricted" mean and not a usual mean? The mean survival time could be a good summary of the survival time distribution, but it typically cannot be estimated well because of limited follow-up (censoring).

The RMST up to time τ, e.g. expressed in years, is simply the population average of the amount of event-free time experienced during the initial τ years of follow-up.

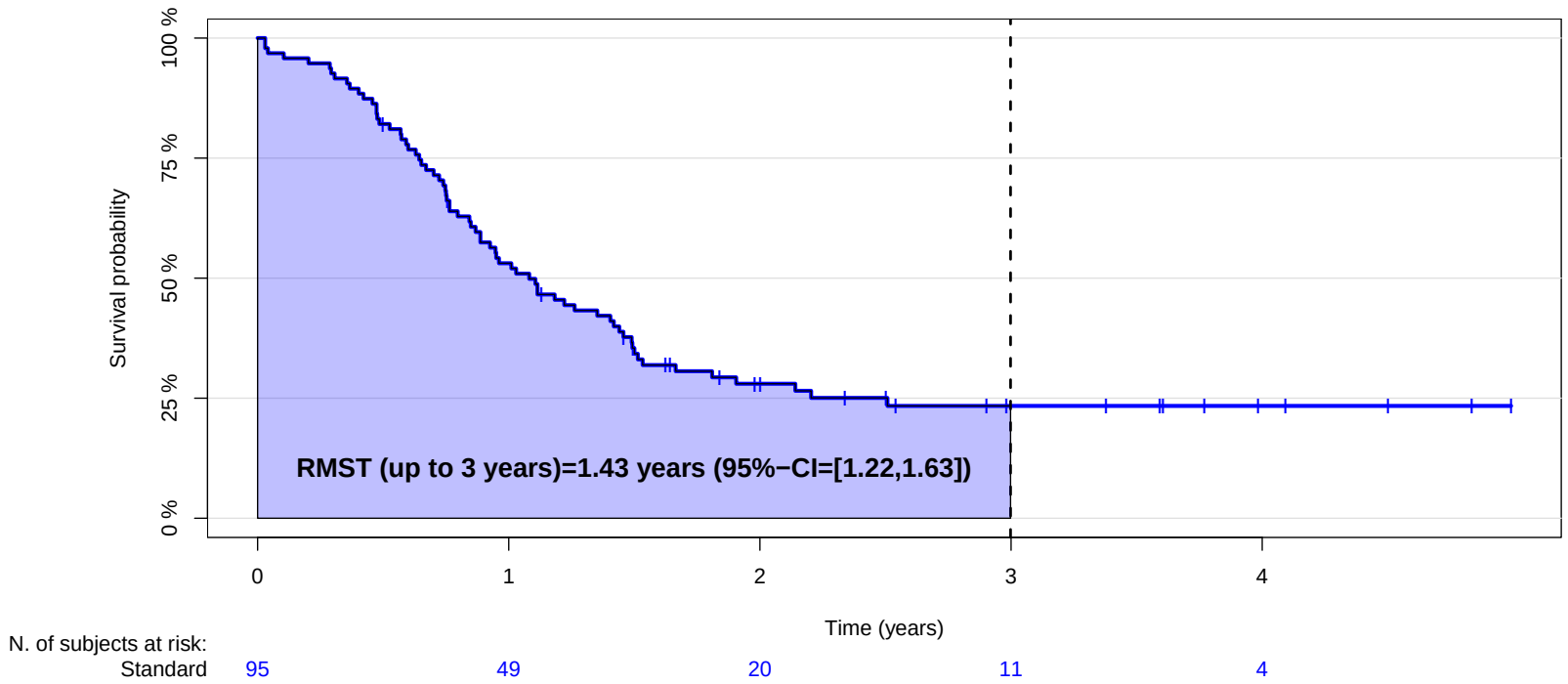
It can be estimated for any relevant time τ no larger than the largest possible follow-up time.

If we do not observed censored observations, we can estimate the RMST by a simple average. With censored data, we can instead estimate it by the area under the Kaplan-Meier curve up to time τ.

The choice of time τ should be prespecified! [13]

[13] We usually choose it close to the maximal follow-up time, but slightly smaller, to observe 'enough' subjects at risk at time τ.

RMST: area under the (Kaplan-Meier) survival curve



The interpretation is that if we treat future patients from the study population similarly and follow them for 3 years, the average time spent alive would be approximately 1.43 years (95%-CI=[1.22,1.63]).

Comparing RMSTs: example

R code:

```
library(survRM2)
RMSTfit <- rmst2(time=d$time/365, status=d$status, arm=d$trt, tau=tau/365)
RMSTfit
```

Output:

```
The truncation time: tau = 3  was specified.

Restricted Mean Survival Time (RMST) by arm
               Est.    se lower .95 upper .95
RMST (arm=1) 1.641 0.103     1.440     1.842
RMST (arm=0) 1.428 0.105     1.223     1.633


Restricted Mean Time Lost (RMTL) by arm
               Est.    se lower .95 upper .95
RMTL (arm=1) 1.359 0.103     1.158     1.560
RMTL (arm=0) 1.572 0.105     1.367     1.777


Between-group contrast
                       Est. lower .95 upper .95     p
RMST (arm=1)-(arm=0) 0.213    -0.074     0.500 0.146
RMST (arm=1)/(arm=0) 1.149     0.952     1.388 0.149
RMTL (arm=1)/(arm=0) 0.865     0.710     1.053 0.148
```

Interpretation

- `RMST (arm=0)`: we estimate that if we treat future patients from the study population with the standard treatment (`trt`=0) and follow them for 3 years, the average time spent alive would be approximately 1.43 years (95%-CI=[1.22,1.63]).
- `RMST (arm=1)`: we estimate that if we treat future patients from the study population with the experimental treatment (`trt`=1) and follow them for 3 years, the average time spent alive would be approximately 1.64 years (95%-CI=[1.44,1.84]).
- `RMST (arm=1)-(arm=0)`: on average, patients treated with the experimental treatment are estimated to be alive 0.21 years (i.e. 77 days) longer than patients treated with the standard treatment, within the 3 years following treatment initiation (95%-CI=[-0.074,0.500] years, i.e., [-27,183] days, p-value=0.146).
- `RMTL` means "restricted mean time lost" and it is computed as τ-RMST (where here $\tau = 3$ years).

Example: Cox regression, one continuous variable (1/2)

Research question: Do young patients have better survival chances than old patients (in the standard treatment group)?

Statistical model: we can model the hazard function via as

$$\lambda(t|X = x) = \lambda_0(t) \exp(\beta x) \ ,$$

where x represents the age of the patient at treatment initiation.

Example: Cox regression, one continuous variable (2/2)

R code:

```
d0 <- d[d$trt==0,]
coxAge <- coxph(Surv(time,status)~age,data=d0)
summary(coxAge)
```

Output:

```
Call:
coxph(formula = Surv(time, status) ~ age, data = d0)

  n= 95, number of events= 69

         coef exp(coef)  se(coef)      z Pr(>|z|)
age -0.003678  0.996329  0.011640 -0.316    0.752

    exp(coef) exp(-coef) lower .95 upper .95
age    0.9963      1.004    0.9739     1.019

Concordance= 0.516  (se = 0.038 )
Likelihood ratio test= 0.1  on 1 df,   p=0.8
Wald test            = 0.1  on 1 df,   p=0.8
Score (logrank) test = 0.1  on 1 df,   p=0.8
```

How to interpret $HR = 0.9963$?

When comparing two subjects alive after any t days since treatment initiation (e.g. after 1, 10, 100 or 1000 days[14]), one being one year older than the other, we estimate that the risk of death **within the next day**[15] is 0.9963 times lower for the older patient. That is, we estimate that the instantaneous risk of death is reduced by 0.37%.

However, the 95% confidence interval tells us that we cannot rule out that the instantaneous risk is reduced by as much as 2.61% or, on the contrary, increased by up to 1.93%.

It might be clinically more relevant to report the results for e.g. a 10 years difference.

[14] Can be any day up to the maximum follow-up time, here 1823 days.

[15] Here we consider that dt =1 day is a 'small enough' duration.

The usual computational trick

R code:

```
d0$age10 <- d0$age/10
coxAge10 <- coxph(Surv(time,status)~age10,data=d0)
summary(coxAge10)
```

Output (partial):

```
          coef exp(coef) se(coef)      z Pr(>|z|)
age10 -0.03678   0.96389  0.11640 -0.316    0.752

      exp(coef) exp(-coef) lower .95 upper .95
age10    0.9639      1.037    0.7673     1.211
```

Interpretation: we estimate that the instantaneous risk of death of a patient 10 years older than another (e.g. 60 versus 50) is reduced by 3.61%[16] (95%-CI=[-21%,23%], p-value=0.752).

16 Note: 3.61%=1-0.9639

Important property/assumption: log-linearity

The model assumes that the hazard ratio (HR) is **log-linear**, that is,

$$\frac{\lambda(t|X = x + \Delta_x)}{\lambda(t|X = x \quad)} = \exp(\Delta_x \beta) = \{\exp(\beta)\}^{\Delta_x}$$

for all values of x and Δ_x.

This means that we assume that the HR is the same when comparing e.g.:

- age 50 and age 40 (x=40, Δ_x=10)
- age 70 and age 60 (x=60, Δ_x=10)

and the square root of that when comparing e.g.:

- age 45 and age 40 (x=40, Δ_x=5)
- age 65 and age 60 (x=60, Δ_x=5)

since

$$\exp(0.5\Delta_x \beta) = \sqrt{\exp(\Delta_x \beta)} \ .$$

Remark: one should carefully think whether this assumption makes sense, in each specific context it is used.[17] When this does not make sense, categorizing a continuous variable is a simple solution (i.e. making age groups).

[17] How to best check this assumption using the data is not 'easy' and it is beyond the scope of this introductory course.

Example: multiple Cox regression, without interaction

Research question: Does a patient receiving the experimental treatment have better survival chances than a patient receiving the standard treatment, when both patients have the same age, tumor size and disability at baseline? [18]

Statistical model: we model the hazard function as

$$\boxed{\lambda(t|X=x) = \lambda_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)}$$

where $X = (X_1, \ldots, X_4)$, $x = (x_1, \ldots, x_4)$, with $x_2 =$ age ,

$$x_1 = \begin{cases} 1 & \text{experimental trt} \\ 0 & \text{standard trt} \end{cases}, \quad x_3 = \begin{cases} 1 & \text{tumor size } >4\text{cm} \\ 0 & \text{tumor size } \leq 4\text{cm} \end{cases}, \quad x_4 = \begin{cases} 1 & \text{disability} \\ 0 & \text{no disability} \end{cases}.$$

[18] Here, because of randomization, this question is more or less the same as the simpler question "Does a patient receiving the experimental treatment have better survival chances than another receiving the standard treatment?". Indeed, due to randomization the patients of the two groups are similar (on average). With observational data, the simpler question could be much less interesting.

R code:

```
cox2 <- coxph(Surv(time,status)~trt+age+Tsize+disability, data=carcinoma)
summary(cox2)
```

Output (partial):

```
                    coef  exp(coef)   se(coef)      z Pr(>|z|)
trt1          -0.0994906  0.9052984  0.1699927 -0.585    0.558
age            0.0001749  1.0001749  0.0080015  0.022    0.983
Tsize>4cm      0.1871734  1.2058364  0.2407842  0.777    0.437
disabilityYes  1.0862726  2.9632083  0.1953929  5.559 2.71e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
trt1              0.9053     1.1046    0.6488     1.263
age               1.0002     0.9998    0.9846     1.016
Tsize>4cm         1.2058     0.8293    0.7522     1.933
disabilityYes     2.9632     0.3375    2.0204     4.346
```

Interpretation

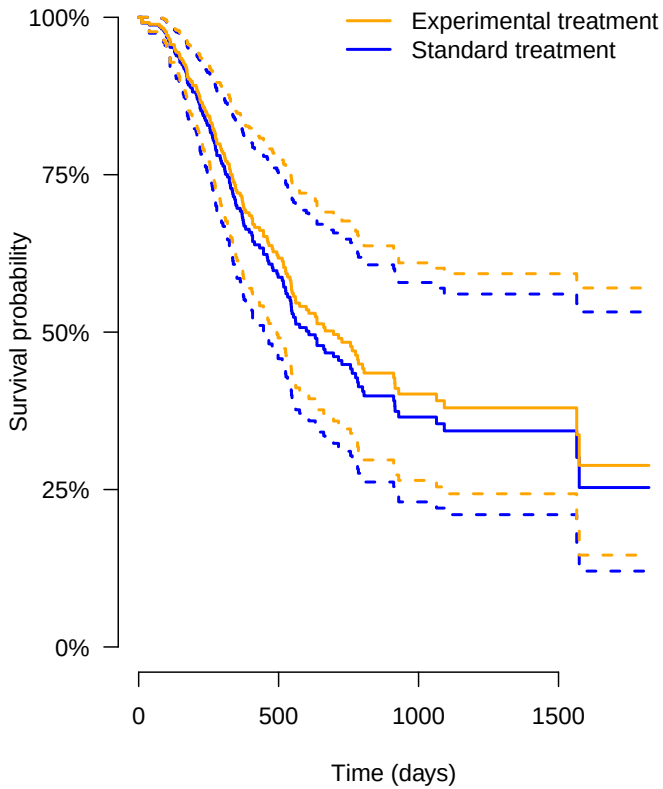
- `trt1`: When comparing two subjects alive after any t days since treatment initiation (e.g. after 1, 10, 100 or 1000 days[19]), we estimate that the risk of death **within the next day**[20] is 0.9053 times lower for patients treated with the experimental treatment than for those treated with the standard treatment, when both patients have the same age (whatever it is), disability (either both are disable or both are not disable) and tumor size (either both with tumor >4cm or both with $\leq$ 4cm) . That is, we estimate that the "adjusted" instantaneous risk of death is reduced by 9.5% (Hazard Ratio= 0.95, 95%-CI=[1.10,0.65], p-value=0.558).

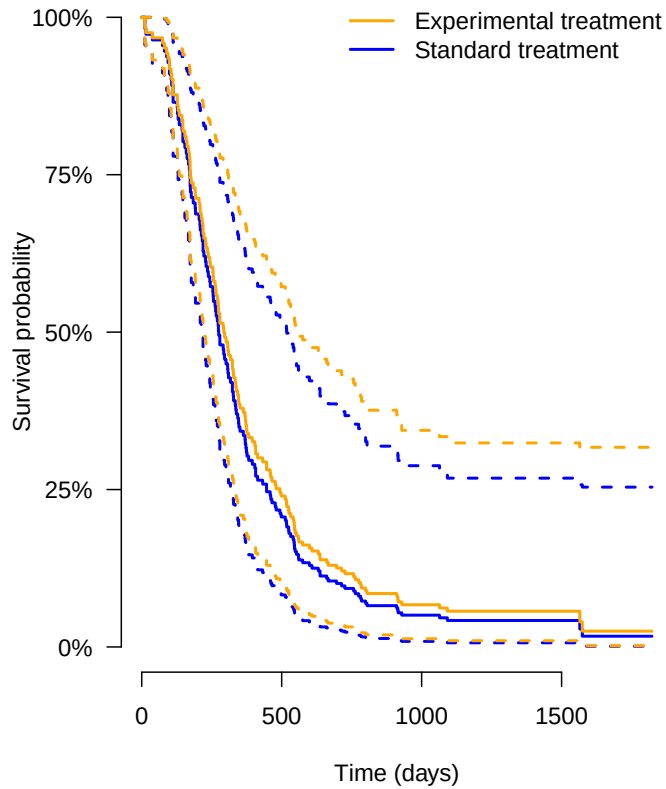
- Other lines: not related to the research question. One can still have a similar interpretation, although it might not relate to a relevant research question.

[19] Can be any day up to the maximum follow-up time, here 1823 days.

[20] Here we consider that dt =1 day is a 'small enough' duration for interpretation.

Appendix: Estimated or "predicted" curves (with pointwise 95%-CI)





Example: two examples of two patient's profiles, when one receives the standard treatment, the other the experimental treatment, both aged 60, with tumor size $\leq$ 4cm and either both are not disabled (left plot) or both are disabled (right plot).

Independent data and "large sample" assumptions

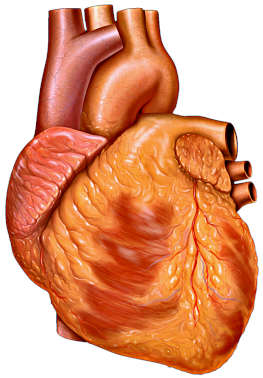
Note that for all methods discussed today:

- We assume that the individual observations are independent. This would probability not be realistic if e.g. we were studying the survival times of dental sealants from patients contributing with several observations because of several treated teeth. E.g. level of oral hygiene or eating habits might create strong correlations...
- 95%-CI and p-value computation are based on "large sample" approximations. They might be moderately accurate with smallish sample sizes... Some alternative "exact" methods exist and can be used in specific contexts (but they are not widely known/used, yet).

Case: Stanford heart transplant program[21]

Data, n=103:

	futime	fustat	transplant
1	49	1	0
2	5	1	0
3	15	1	1
4	38	1	1
5	17	1	0
6	2	1	0



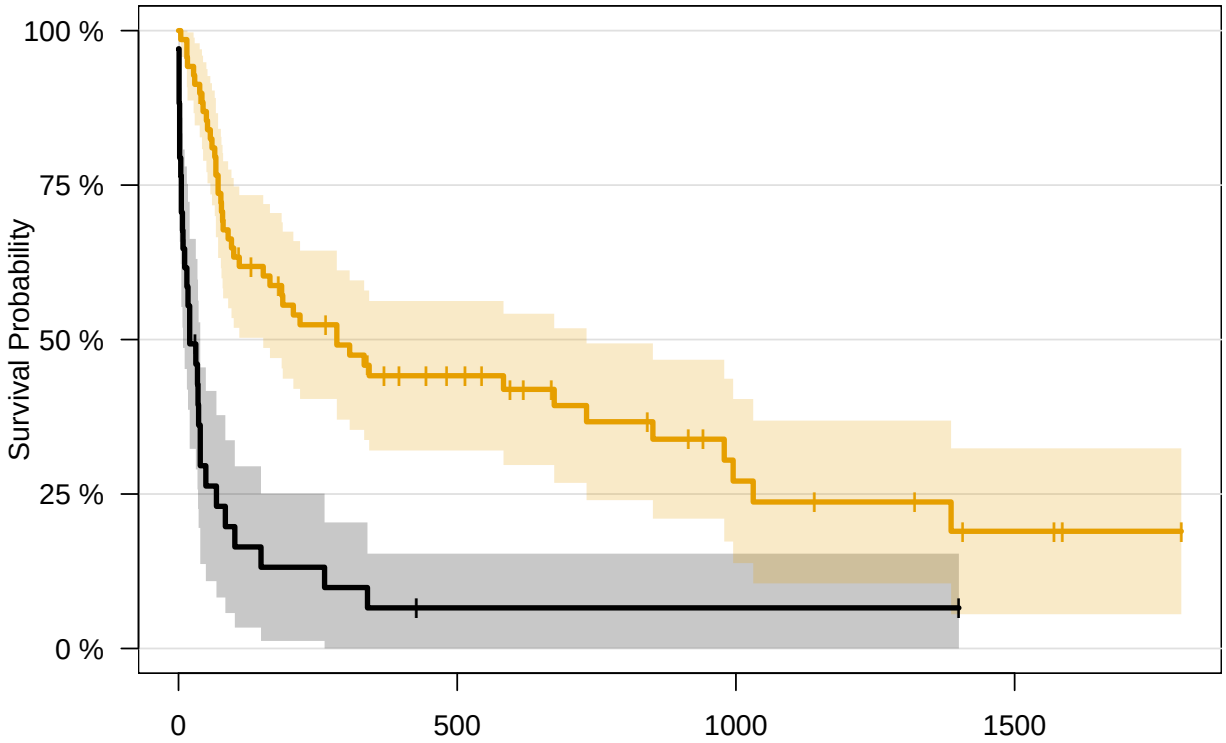
Variables (many others actually available...):

- `futime`: time in days since acceptance into the transplantation program
- `fustat`: dead (1) or alive (0)
- `transplant`: received transplantation (1) or not (0)

Research question: Does cardiac transplantation prolong life?

[21] **Reference:** Gail. "Does cardiac transplantation prolong life? A reassessment." Annals of Internal Medicine 76.5 (1972): 815-817; see also e.g., Moore. Applied survival analysis using R. Springer, 2016.

Common, but **WRONG** analysis!



N. of subjects at risk:											
Not Transplanted	34	4	2	1	1	1	1	1	0	0	0
Transplanted	69	37	26	21	15	12	7	6	3	1	1

What is the interpretation? Is it meaningful? What's wrong?

What is wrong?

We compared the survival chances of those who will be transplanted one day (in the future), to those who will not, and showed that those who will be transplanted one day (in the future) survive longer.

This is completely meaningless !

[22] It kind of answers the questions *"If I am transplanted one day what are my survival chances? And what if I am never transplanted?"*

[23] See e.g. Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. American journal of epidemiology, 167(4), 492-499.

What is wrong?

We compared the survival chances of those who will be transplanted one day (in the future), to those who will not, and showed that those who will be transplanted one day (in the future) survive longer.

This is completely meaningless !

1. At time zero, i.e. at time of acceptance in the program, neither the doctors nor the patient knows whether the patient will be transplanted. So, who could ever benefit from knowing these "survival chances"? This analysis answers an irrelevant research question![22]

[22] It kind of answers the questions *"If I am transplanted one day what are my survival chances? And what if I am never transplanted?"*

[23] See e.g. Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. American journal of epidemiology, 167(4), 492-499.

What is wrong?

We compared the survival chances of those who will be transplanted one day (in the future), to those who will not, and showed that those who will be transplanted one day (in the future) survive longer.

This is completely meaningless !

1. At time zero, i.e. at time of acceptance in the program, neither the doctors nor the patient knows whether the patient will be transplanted. So, who could ever benefit from knowing these "survival chances"? This analysis answers an irrelevant research question![22]
2. Even if the "intervention" (here transplantation) has no effect on survival, the analysis will show that those receiving the intervention will survive longer. Why?

[22] It kind of answers the questions *"If I am transplanted one day what are my survival chances? And what if I am never transplanted?"*

[23] See e.g. Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. American journal of epidemiology, 167(4), 492-499.

What is wrong?

We compared the survival chances of those who will be transplanted one day (in the future), to those who will not, and showed that those who will be transplanted one day (in the future) survive longer.

This is completely meaningless !

1. At time zero, i.e. at time of acceptance in the program, neither the doctors nor the patient knows whether the patient will be transplanted. So, who could ever benefit from knowing these "survival chances"? This analysis answers an irrelevant research question![22]
2. Even if the "intervention" (here transplantation) has no effect on survival, the analysis will show that those receiving the intervention will survive longer. Why? Just because the patients need to survive "long enough" to receive the intervention. These patients can never be observed dead before they receive the intervention. They are "immortal" until they receive the intervention. This introduces a so-called " **immortal time bias**"[23]

[22] It kind of answers the questions *"If I am transplanted one day what are my survival chances? And what if I am never transplanted?"*

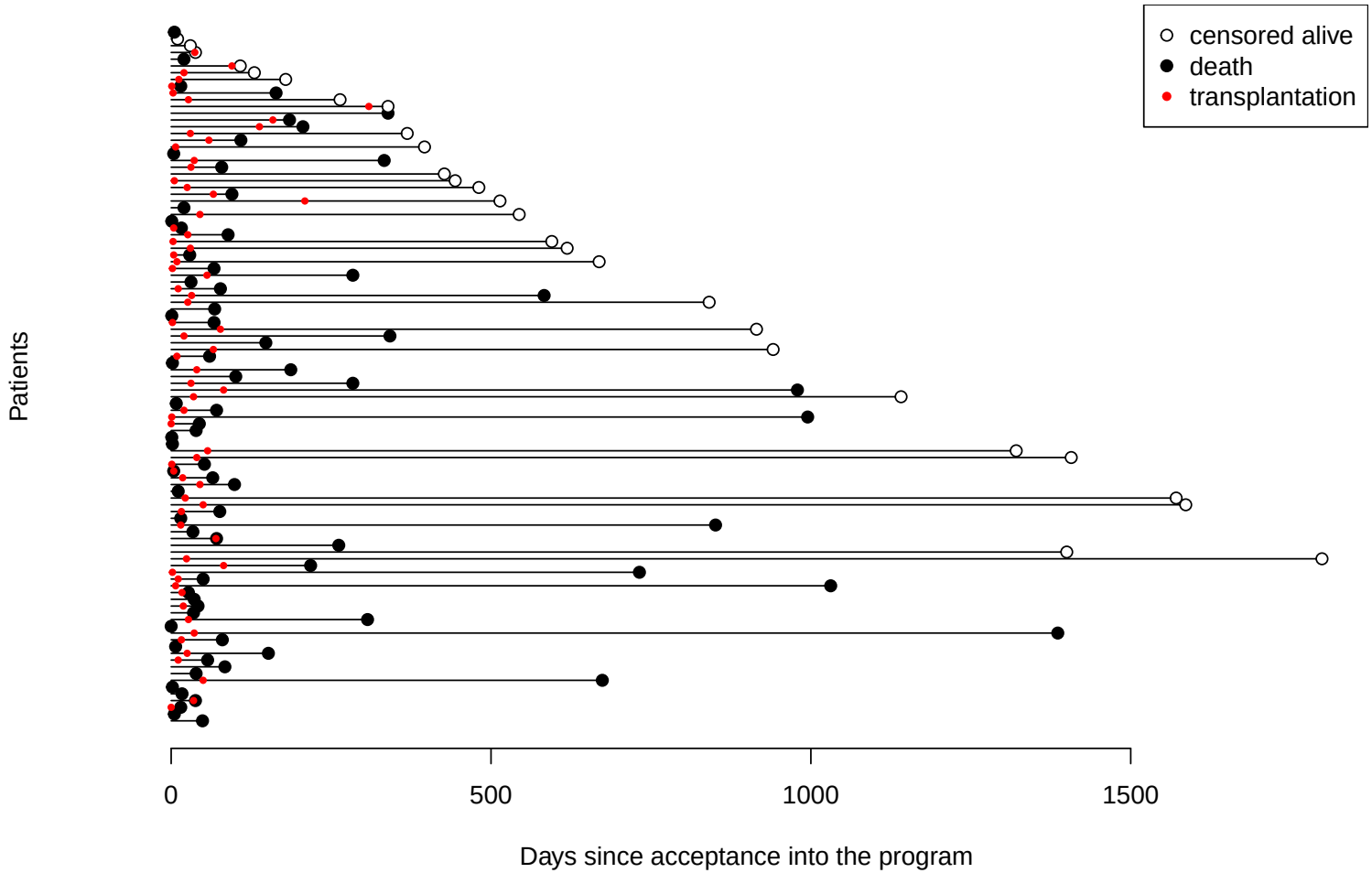
[23] See e.g. Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. American journal of epidemiology, 167(4), 492-499.

Take home message

Do not condition on the future!

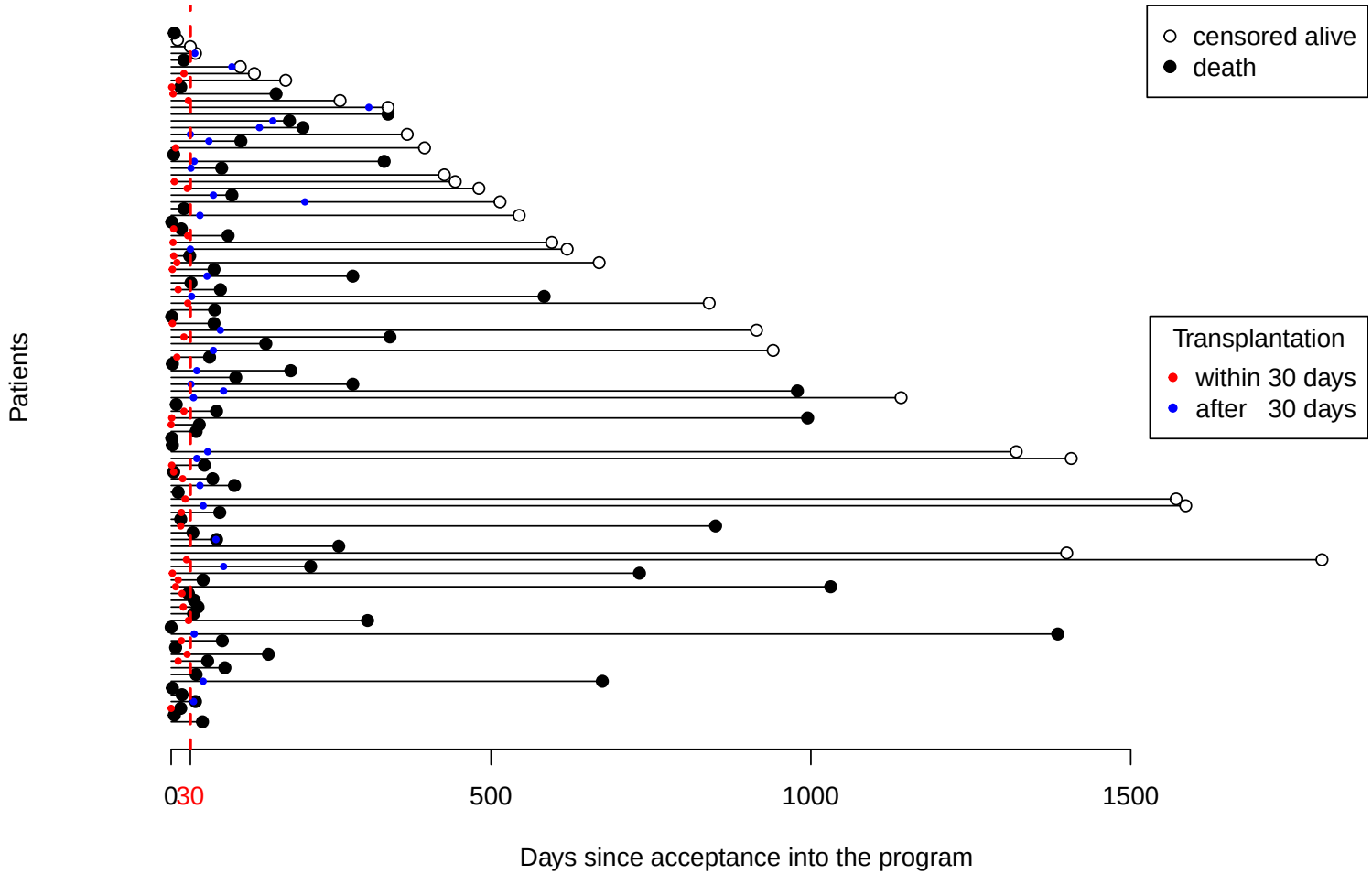
- In survival analysis, define groups (and "adjust") based on what is known at "time zero" only.
- This might seem "obvious" today, but keep in mind that once the data are recorded in your excel sheet or csv file or database, it is often no longer obvious to know when e.g. a blood sample, diagnosis or claimed prescriptions has been observed, and whether it was before or after the start of the follow-up.
- Advanced statistical methods can sometimes help, although not always, to analyze "time-dependent" covariates measured after the start of follow-up. Seek help from a statistician!

A simple, better, analysis: landmarking (sometimes meaningful, not always, 1/4)



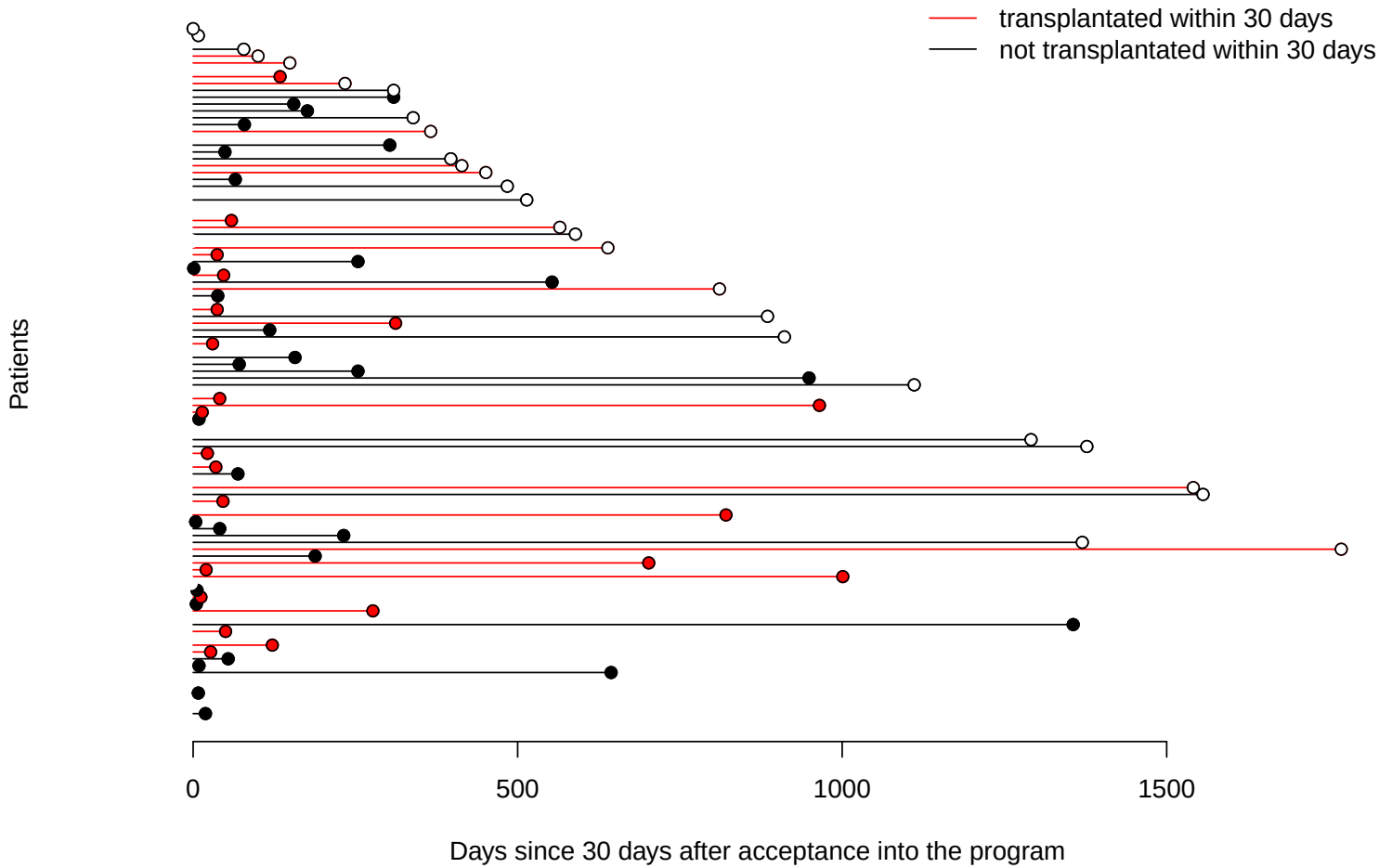
- The full data...

A simple, better, analysis: landmarking (sometimes meaningful, not always, 2/4)



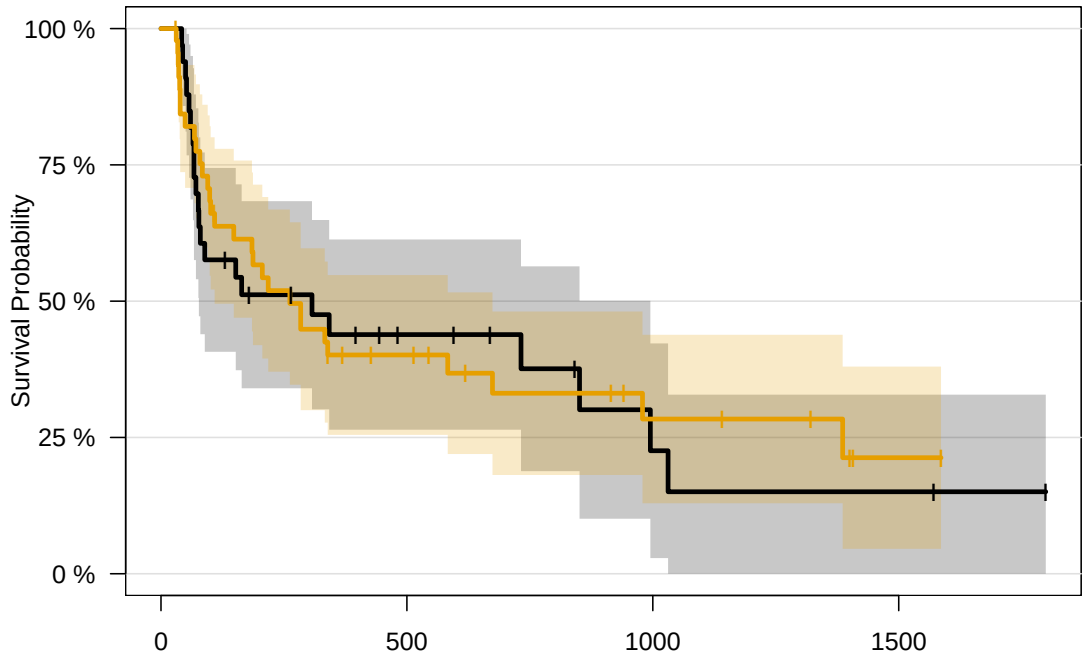
- Let's look at what has happened within the first, say, 30 days.

A simple, better, analysis: landmarking (sometimes meaningful, not always, 3/4)



- We move "time zero" from date of acceptance into the program to the same date plus 30 days. Now the two groups that we compare are "well defined" at new "time zero". Note: we study those alive at 30 days only!

A simple, better, analysis: landmarking (sometimes meaningful, not always, 4/4)



N. of subjects at risk:											
Not Transplanted	33	15	12	9	7	4	2	2	2	1	1
Transplanted	46	26	16	13	9	9	6	5	1	0	0

Interpretation: a doctor meets two patients accepted in the program exactly 30 days they have been accepted into the program. One has been transplanted within the 30 days, the other has not. The curves show the estimated survival curves ("prognosis") for these two patients.

Digression: Does the statistical method matter? Yes !

Note that maybe the wrong analysis gave the right answer. But that is not the point!

1. In practice, you do not know what is right or wrong, and that is why analyzing the data at hand should be interesting. Hence an obviously wrong analysis cannot be useful.

2. Even if the conclusion is correct, the claim that the data support the conclusion cannot be correct, if the statistical analysis is incorrect.

Outline

Survival Data

- ILO: to recognize survival data and list contexts in which we meet them
- ILO: to define censoring and explain the challenges it creates
- ILO: to distinguish censoring from a competing risk

Simple & common analyses: possibilities and pitfalls

- ILO: to perform a Kaplan-Meier analysis and a log-rank test
- ILO: to fit and interpret a Cox model
- ILO: to list the main limitations of the Cox model
- ILO: to perform a Restricted Mean Survival Time (RMST) analysis
- ILO: to exemplify the difference between a risk ratio and a hazard ratio
- ILO: to recognize and avoid immortal time bias

Competing risks

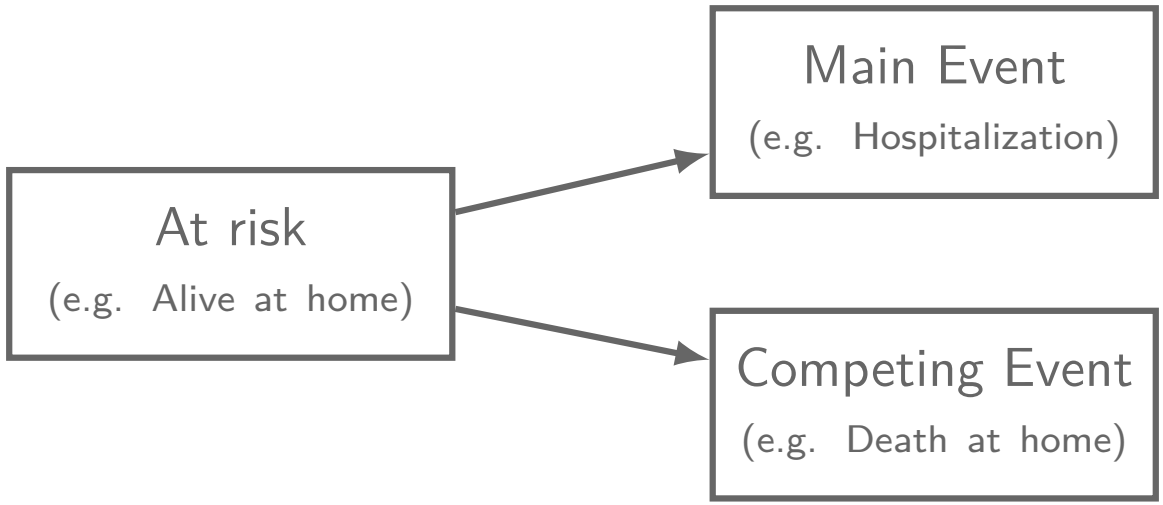
- ILO: to exemplify competing risks data
- ILO: to describe a very common mistake
- ILO: to employ a basic (but appropriate!) method for competing risks data

What is a competing risks situation?

A competing risks' situation, which is frequent in epidemiological follow-up studies, is the situation in which the event of interest (e.g., stroke) does not happen for all subjects, as one or several other types of event, called "competing events", prevent the main event from happening (e.g, non cardiovascular death prevents strokes, as we cannot suffer a stroke once we are dead).

Note that censoring is different, as it only prevents the event from being observed; it does not prevent the event from happening.

Competing risks examples



Other examples:

Main event	Competing event
Stroke	Death (without stroke before)
Cancer	Death (without cancer before)
Leaving ICU[24]	Death (in ICU)
Weaning[25]	Death (before weaning)
Healing complication [26]	"Normal" loss of primary teeth
Death at work	Retiring
Pregnancy	Stop trying or menopause

[24] ICU: intensice care unit

[25] Among premature babies who require parenteral (intravenous) nutrition

[26] in primary teeth

Case: HF-ACTION randomized trial data [27]

Data, n=377 + 364 :

	time	status	trt
1	3.4771	0	1
2	0.7639	2	0
3	1.0897	0	0
4	0.7009	1	0
5	0.3012	1	0
6	0.3778	1	1

The usual competing risks data format!

Variables:

- `time`: time in years to an event or end of follow-up transplantation program
- `status`: type of event: (first) hospitalization (1), death (2) or censoring (0)
- `trt`: exercise training (1) or usual care (0)

Research question: Does exercise training, in top of usual care, lower the risk of hospitalization among heart failure patients?

[27] **Reference:** O'Connor et al. Jama 301.14 (2009): 1439-1450. Data (subgroup) from the R package WA.

Common, but **WRONG** Kaplan-Meier analysis

Unfortunately, many researchers use status=0 for both status=0 and status=2, to be back to the usual survival data format, and then run a Kaplan-Meier analysis. That is, they treat death as censoring. This is fundamentally wrong.

The risks of hospitalization which are computed in that way are not meaningful. Here they would actually estimate the risk of hospitalization in the hypothetical world in which nobody dies without being hospitalized first. Alternatively, we can conclude that they overestimate the true risks.[28]

In short, treating a competing risk as censoring corresponds to estimate risks in the hypothetical world in which the competing risk does not exist. It is usually better to *"stick to this world"*.[29]

[28] Huebner et al. "Competing risks need to be considered in survival analysis models for cardiovascular outcomes." The Journal of thoracic and cardiovascular surgery 153.6 (2017): 1427-1431.

[29] See e.g. Andersen, P. K., & Keiding, N. (2012). Statistics in Medicine, 31(11-12), 1074-1088.

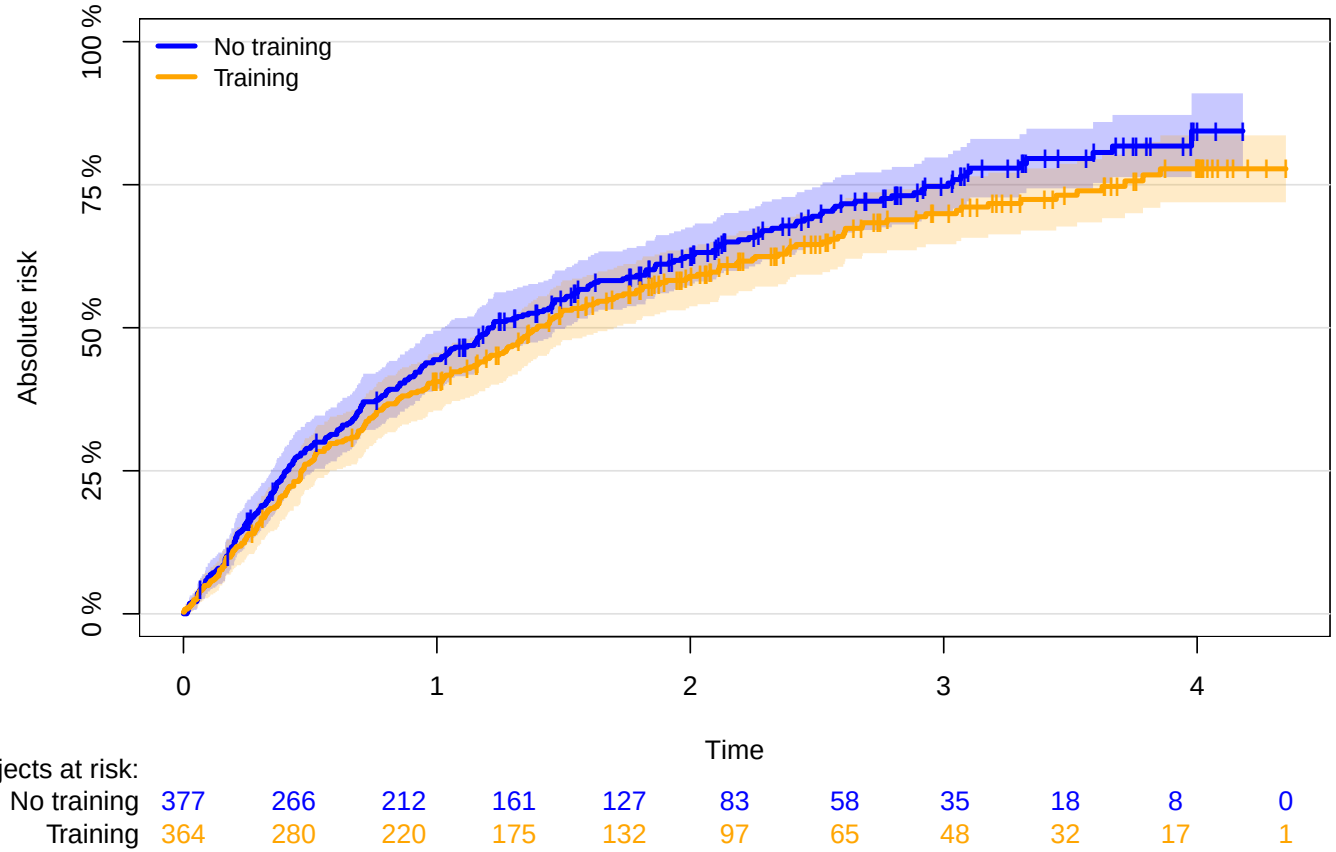
Correct analysis (1/2)

To estimate risks in a competing risks situation, we used specific methods. We often say that we estimate the "cumulative distribution function" or the "absolute risk" in that case. Using this "jargon" emphasizes that we aim to properly account for competing risks.

- Instead of Kaplan-Meier, we can use the Aalen-Johansen estimator.[30]
- Regression models also exist. Seek help from a statistician.

[30] Huebner et al. "Competing risks need to be considered in survival analysis models for cardiovascular outcomes." The Journal of thoracic and cardiovascular surgery 153.6 (2017): 1427-1431.

Correct analysis (2/2)



- Estimates of the absolute risks of hospitalization obtained with the Aalen-Johansen estimator.[31]
- We see that, for e.g., the 3-year absolute risks are estimated as 69.9% (95%-CI [64.6, 75.3]) and 74.7% (95%-CI [69.7, 79.8]) with and without "training". This corresponds to a difference -4.8% (95%-CI [-12.1, 2.6], p=0.204).

[31] See R-demo for the (simple) R code.

Reminder of today's topics and ILOs

Survival Data

ILO: to recognize survival data and list contexts in which we meet them
ILO: to define censoring and explain the challenges it creates
ILO: to distinguish censoring from a competing risk

Simple & common analyses: possibilities and pitfalls

ILO: to perform a Kaplan-Meier analysis and a log-rank test
ILO: to fit and interpret a Cox model
ILO: to list the main limitations of the Cox model
ILO: to perform a Restricted Mean Survival Time (RMST) analysis
ILO: to exemplify the difference between a risk ratio and a hazard ratio
ILO: to recognize and avoid immortal time bias

Competing risks

ILO: to exemplify competing risks data
ILO: to describe a very common mistake
ILO: to employ a basic (but appropriate!) method for competing risks data