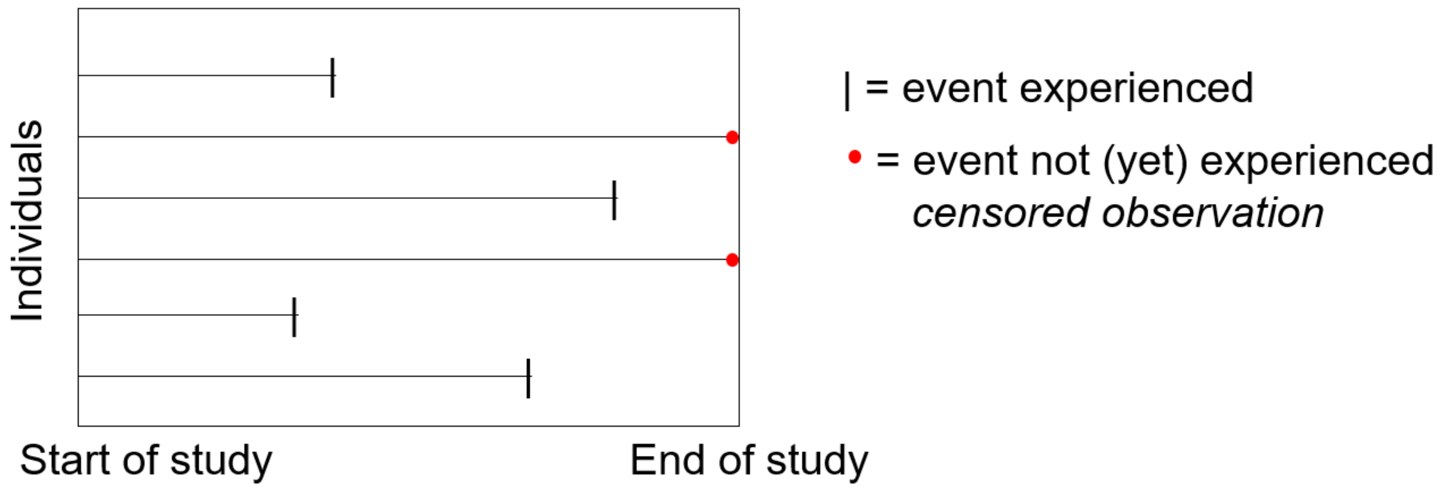
UPPSALA UNIVERSITET

Analysis of Survival Data

Lecture 2: Censoring and truncation



Inger Persson

Program L2

- **Censoring and truncation**
  - Right, left and interval censoring
  - Right and left truncation
  - Likelihood for censored and truncated data

Censored observations

Typical for survival data:
not all individuals experience the event of interest.

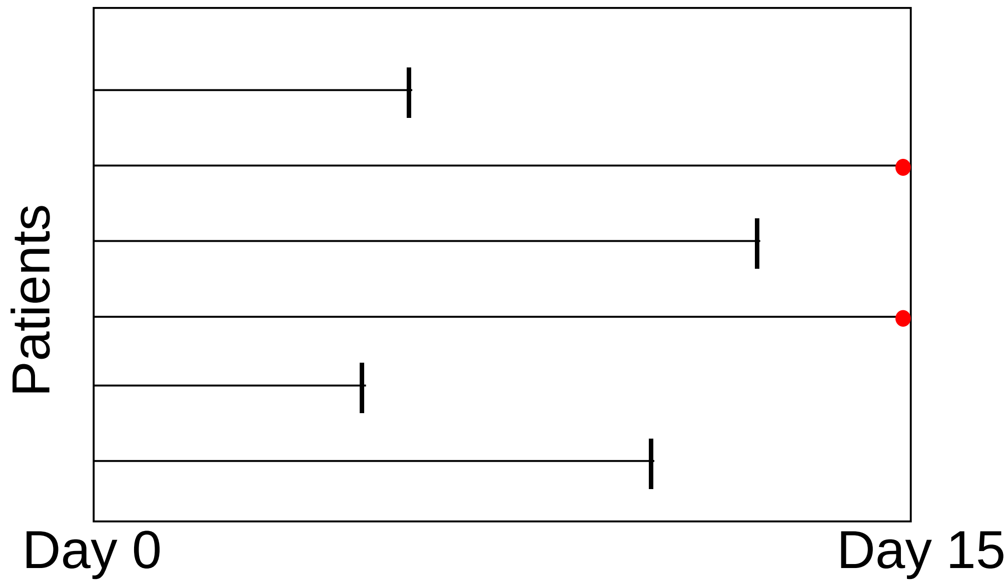
When the study is closed, some patients have experienced the event and others have not.

Or, some patients may have experienced the event but the exact time point is unknown.

The observation for an individual who has not experienced the event, or for whom the event time point is unknown, is said to be a **censored observation.**

Example: Time to treatment response

Patients with a certain disease are being followed for 15 days after starting a new treatment to see how long it takes before they respond to the treatment.



| = response

• = no response (yet)
censored observation

UPPSALA UNIVERSITET

Survival analysis

Survival analysis takes censoring into account, and calculates likelihoods based on the combination of the probability to experience the event at a certain time with the probability to experience the event at a certain time *or later*.

Censoring categories

Different categories of censoring:

- Right censoring
- Left censoring
- Interval censoring

Each type of censoring will lead to a different likelihood function.

UPPSALA UNIVERSITET

Right censoring

Right censoring occurs when the event is observed only if it occurs before a certain time, e.g. the predetermined end of a study.

Starting times and censoring times may be fixed or vary from individual to individual.

- Type I censoring
- Progressive type I censoring
- Generalized type I censoring
- Type II censoring
- Random censoring

UPPSALA UNIVERSITET

Type I censoring (right censoring)

The starting point is the same for all individuals.

Censoring times may vary from individual to individual.

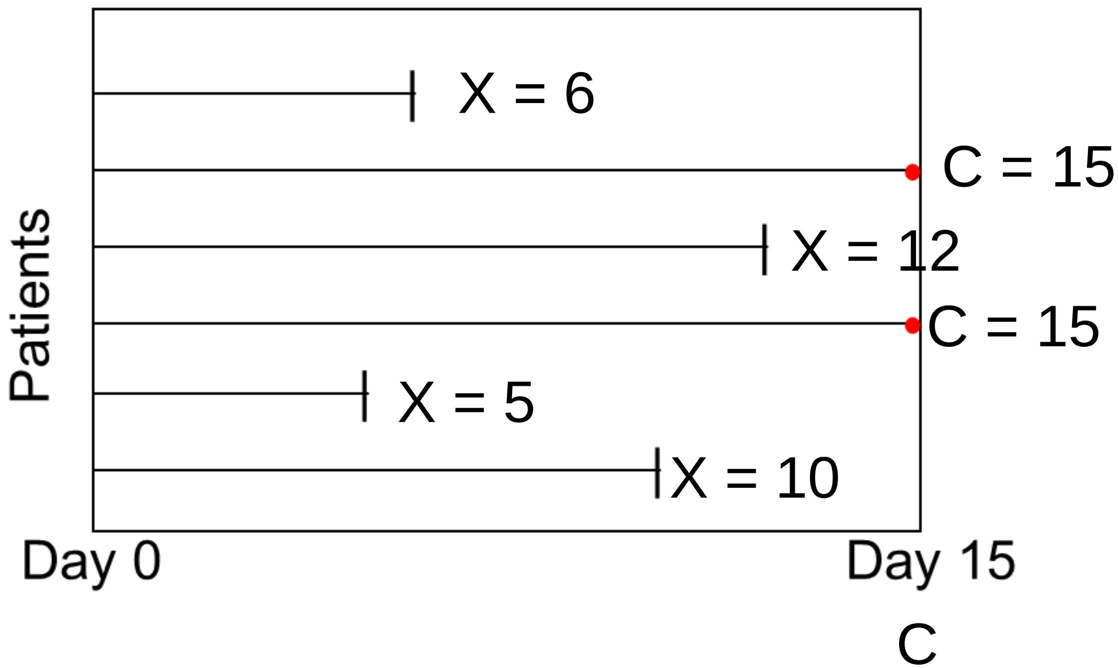
Example: Time to treatment response

X = time to response

C = censoring time

T = min (X, C)

$$\delta = \begin{cases} 1 & \text{If event occurred} \\ 0 & \text{If censored} \end{cases}$$



T	δ
6	1
15	0
12	1
15	0
5	1
10	1

UPPSALA UNIVERSITET

Progressive Type I censoring

Progressive type I censoring is when some individuals are censored at a certain time point, and others are further observed until a second time point.

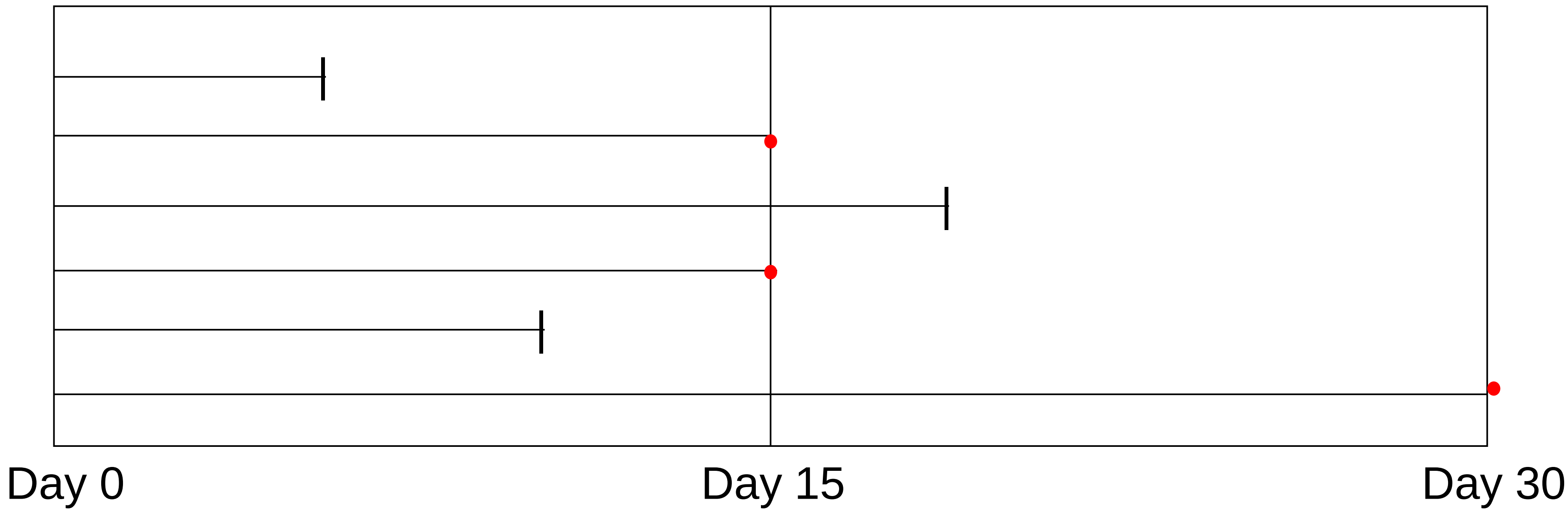
A type of study usually performed by practical and economical reasons.

Example: Time to treatment response

X= time to response of the new treatment.
Patients are followed for 15 days after treatment start.

In addition, lagged response effects are of interest.

Some patients are followed for another 15 days, after stopping the treatment.



UPPSALA UNIVERSITET

Generalized Type I censoring

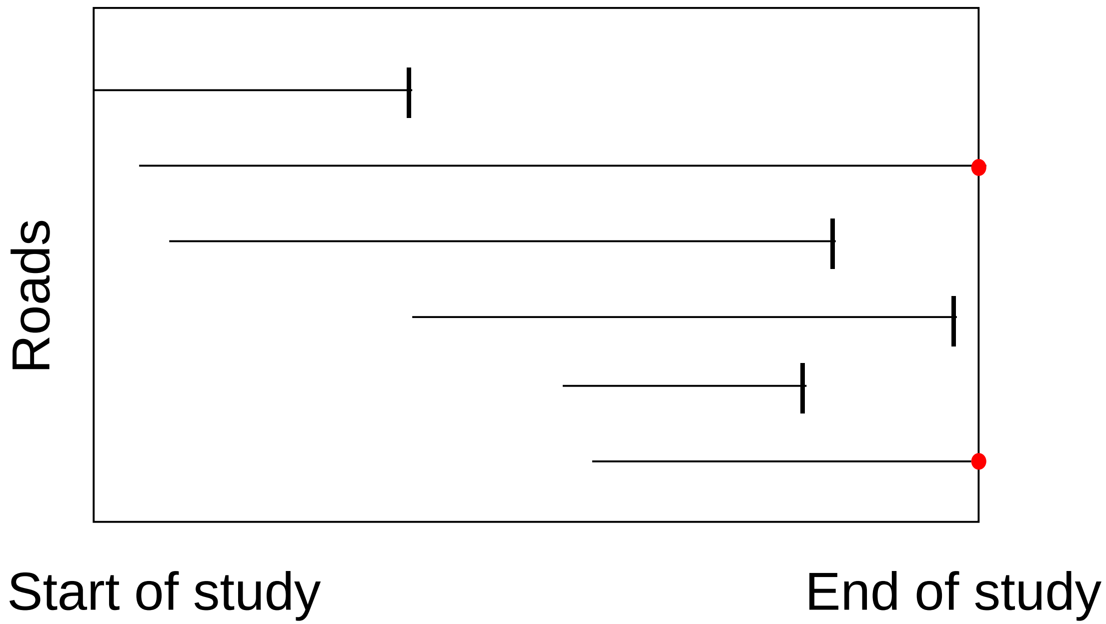
Generalized type I censoring is when individuals enter the study at different times, and the end of the study is predetermined.

Censoring times are known when individuals enter the study.

A very common type of censoring in patient studies, where patients are included e.g. at the time of diagnosis.

Example: Road maintenance

X= time to when a road in the Swedish road network needs to be maintained (e.g. with new pavement)

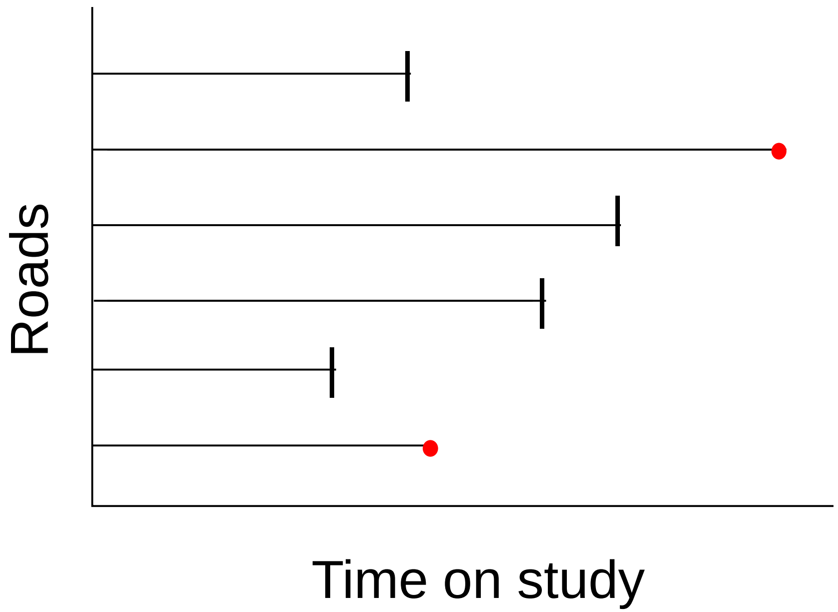


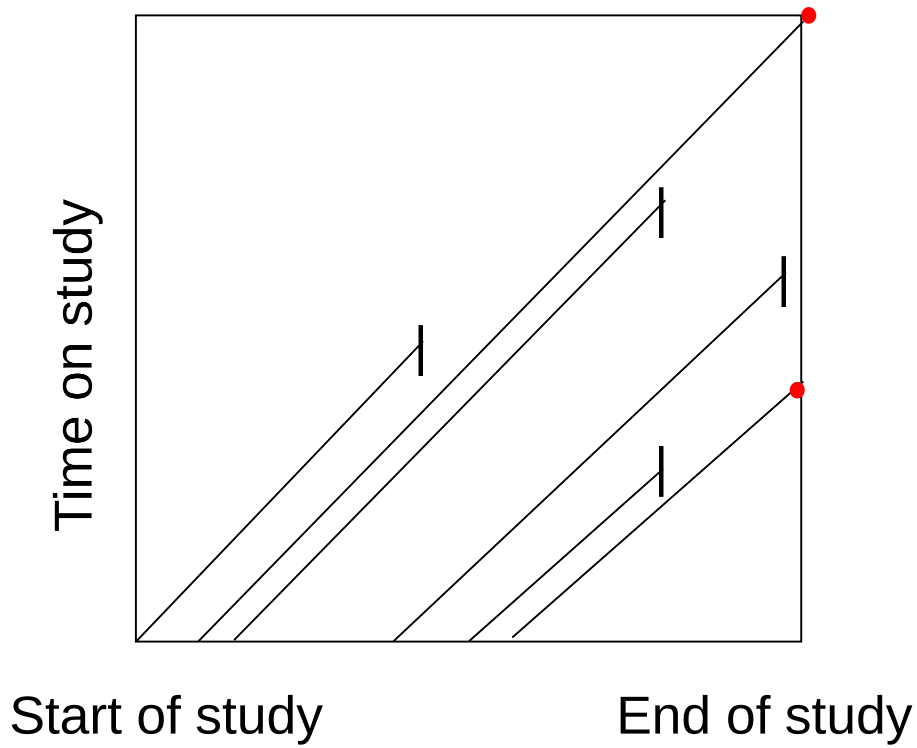
Generalized Type I censoring

Two common ways of representing data with generalized type 1 censoring:

1) Shifting each individual's starting time to 0
2) Lexis diagram

Example: Road maintenance





Each starting time shifted to 0

Lexis diagram

Type II censoring (right censoring)

Type II censoring is also a type of right censoring.

The study is terminated when r individuals have experienced the event ($r < n$, predetermined numbers).

Advantages:

- You can be sure to include a certain number of events in your data (e.g. sample size calculations are based on the number of events).
- Data consist of the r shortest survival times, thus the theory of order statistics can be used (simpler).

Competing risks

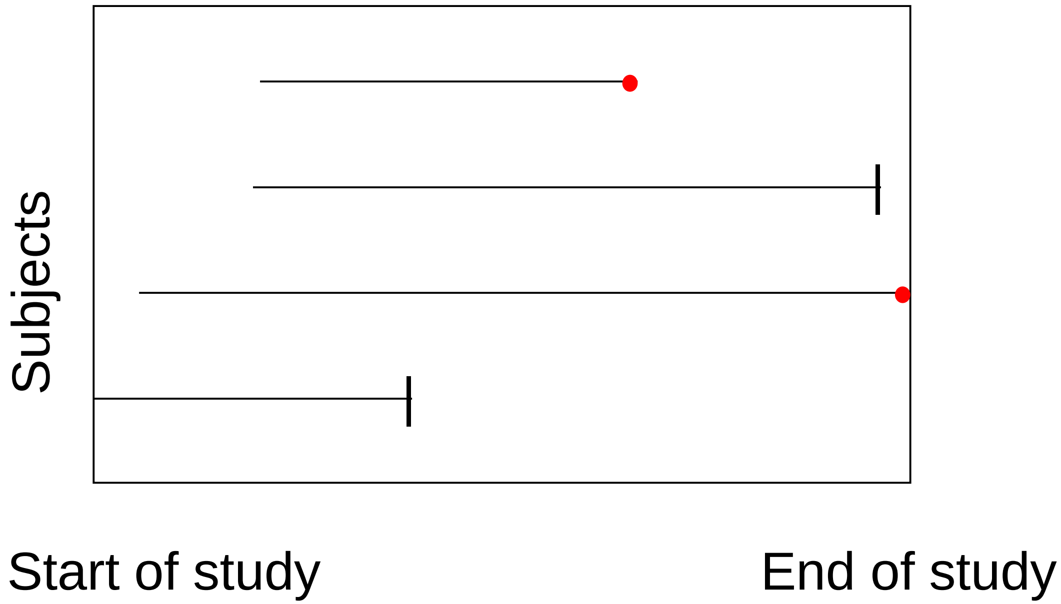
Sometimes the individuals in the study may fail due to **competing risks**.

A competing risk is something that if it occurs, it prevents the event of interest from being observed.

E.g. death in the study of time to response of a medical treatment, or death from heart-failure when the lifetime of cancer patients is being investigated, etc.

Special cases of competing risks can be handled by random censoring.

Random censoring (special case of competing risks censoring)



E.g. a study of how long marriages last. If one of the spouses dies, that couple’s observation is censored.

Another example is when people in the study move abroad, or no longer want to participate.

Left censoring

Left censoring occurs when the event has happened prior to the study start.

The event is known to have happened but the exact time is unknown.

Example: Children learning to read

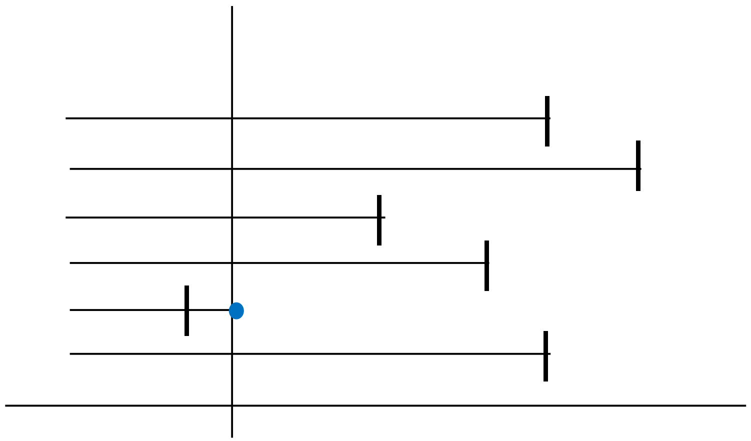
X = The time it takes for children in school to learn to read

C = School start

T = max (X, C)

$$\varepsilon = \begin{cases} 1 & \text{If event is observed} \\ 0 & \text{If left censored} \end{cases}$$

• = left censored observation



School start
= study start
(C=0)

Double censoring

Double censoring occurs when both left and right censoring are present in the study.

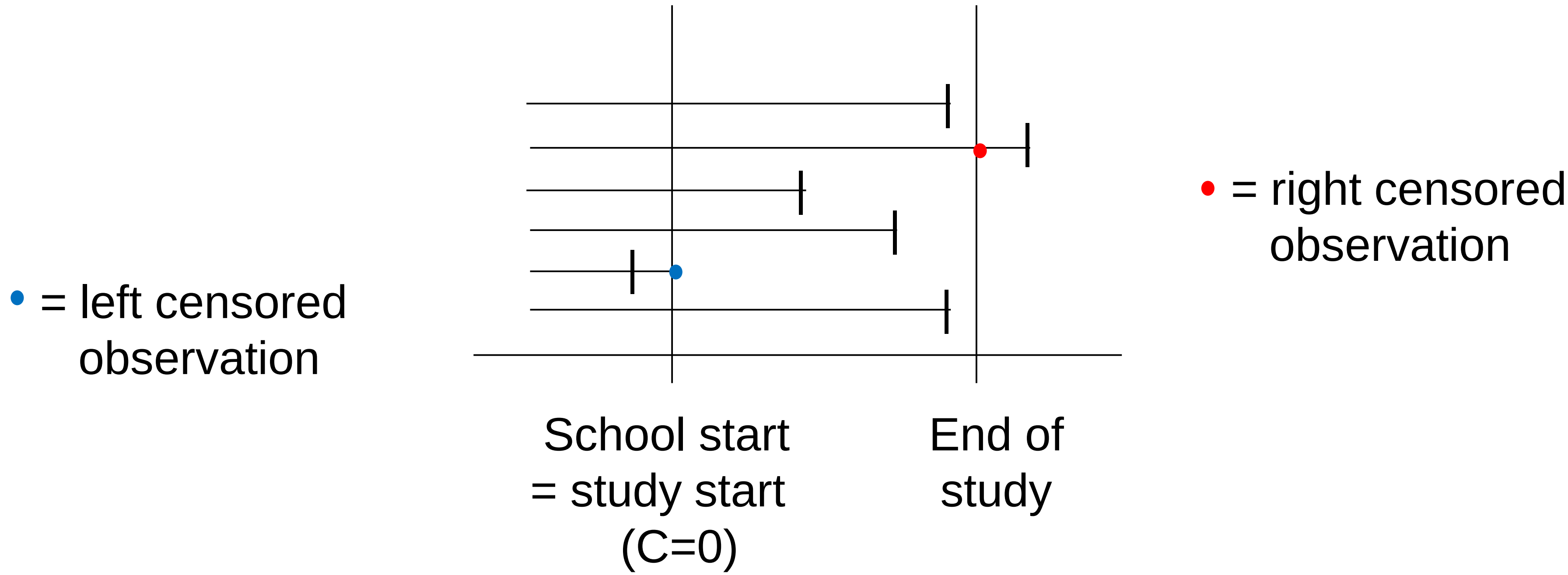
Example: Children learning to read

X = The time it takes for children in school to learn to read

C = School start

T = max (X, C)

$$\delta = \begin{cases} 1 & \text{If event is observed} \\ 0 & \text{If right censored} \\ -1 & \text{If left censored} \end{cases}$$



UPPSALA UNIVERSITET

Interval censoring

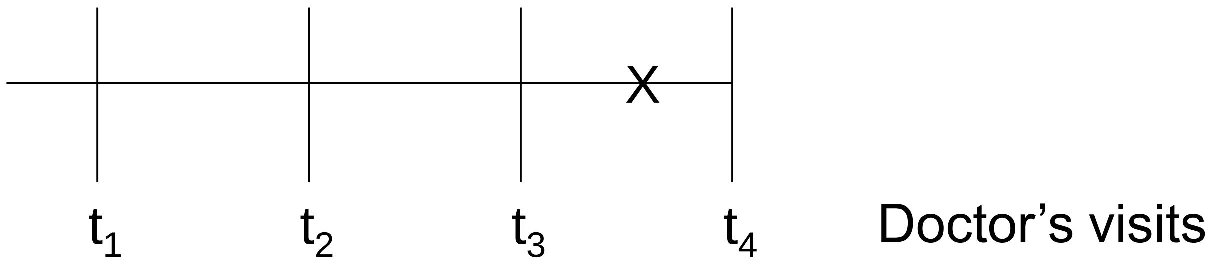
Interval censoring occurs when the event is known to happen only within an interval of time.

Happens in e.g. clinical trials or longitudinal studies when patients have periodic follow-up and the exact event date is not recorded between follow-up visits.

You know that the event occurred sometime between time points A and B.

Example: Doctor's visits

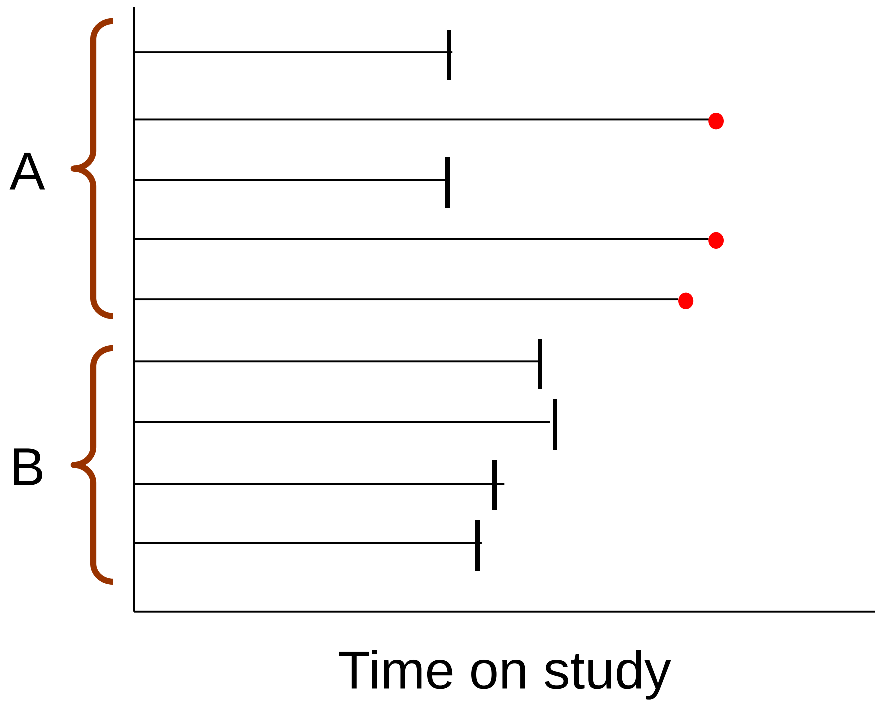
X = Age at first symptom of hearing loss



At a certain doctor's visit it is discovered that the patient has some hearing loss, but the exact time when it started is unknown.

How to deal with censored observations?

Example: a study of two different treatments, A and B.



B patients have the longest survival times, if you ignore the censored observations.

If you include the censored observations the A patients have the longest survival times.

How to deal with the censored times?

UPPSALA UNIVERSITET

Program L2

- **Censoring and truncation**
  - Right, left and interval censoring
  - Right and left truncation
  - Likelihood for censored and truncated data

Truncation

Sometimes only individuals with a certain characteristic are included in the study.

There is a risk of missing individuals whom you have an interest of including in the study.

Right truncation

Right truncation is when only subjects who experience the event within a certain observational window (e.g. the study period) are included in the study.

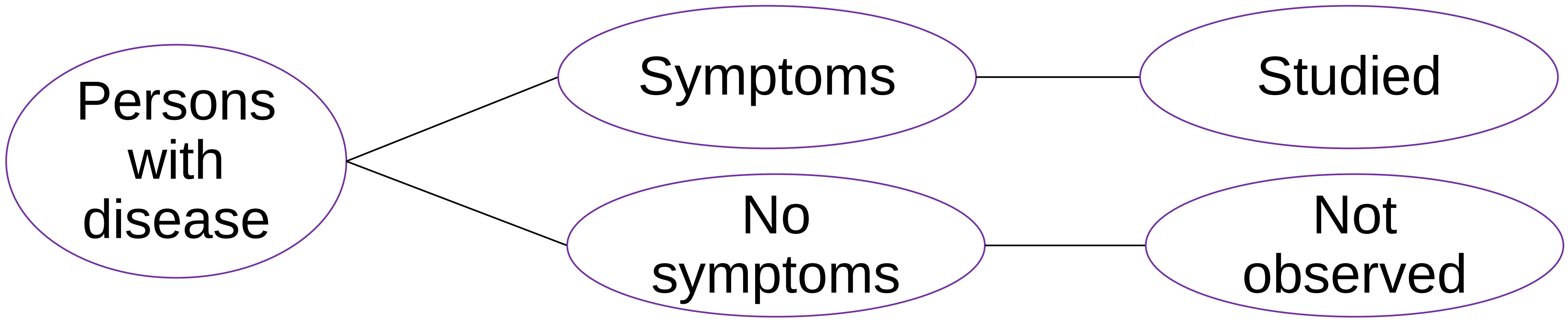
E.g. failure time of electronic devices. If they don't break down within the limited guarantee time they won't be handed in for service and will thus not be observed.

Or, any study based on death records.

Left truncation

Left truncation is when there is a selection before the event, you must have experienced something earlier.

Example: survival for individuals with a certain disease:



Individuals who die before the truncation time (e.g. study start) are not observed, and the survival is overestimated.

UPPSALA UNIVERSITET

Program L2

- **Censoring and truncation**
  - Right, left and interval censoring
  - Right and left truncation
  - Likelihood for censored and truncated data

Likelihood construction

No censored observations:

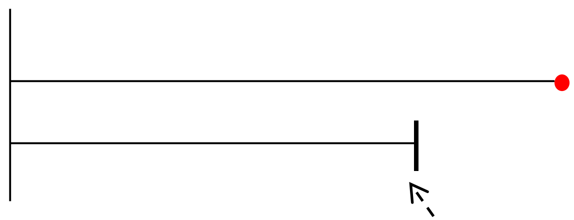
$$L = \prod_{i=1}^{n} f(x_i) \qquad \text{(Sample } (x_1, \ldots x_n))$$

Probability to get the observed data.

UPPSALA UNIVERSITET

Likelihood construction for censored data

For right censored data:



Information provided by the survival function evaluated at the time on study.

Provides information on the probability of the event occurring at this time

The different categories of information are to be combined.

UPPSALA UNIVERSITET

Likelihood construction: right censored data

$$L = \prod_{i \in E} f(x_i) \prod_{i \in R} S(C_i)$$

E = event R = right censored

$$L = \prod_{i=1}^{n} f(t_i)^{\delta} S(t_i)^{1-\delta}$$

$$\delta = \begin{cases} 1 & \text{If event occurred} \\ 0 & \text{If censored} \end{cases}$$

Critical assumption:

Lifetimes and censoring times are independent.

Example: Exponential distribution

Four observed survival times:

2, 3, 4+, 4+ (+ denotes right censoring)

Assume that the survival times follow an exponential distribution with hazard rate λ.

Find the maximum likelihood estimator of λ.

UPPSALA UNIVERSITET

Example: Exponential distribution

$$f(x) = \lambda e^{-\lambda x}$$

$$S(x) = e^{-\lambda x}$$

$$L = \lambda e^{-\lambda 2} \lambda e^{-\lambda 3} e^{-\lambda 4} e^{-\lambda 4}$$

$$= \lambda^2 e^{-\lambda(2+3+4\cdot 2)} = \lambda^2 e^{-13\lambda}$$

To be maximized
How do we do that?

The derivative of the likelihood (or the logarithm of the likelihood) gives the maximum

UPPSALA UNIVERSITET

Example: Exponential distribution

$$\ln L = \ln \lambda^2 e^{-13\lambda} = 2\ln\lambda - 13\lambda$$

$$\frac{\partial \ln L}{\delta\lambda} = \frac{2}{\lambda} - 13$$

$$\frac{2}{\hat{\lambda}} - 13 = 0$$

$$\frac{2}{\hat{\lambda}} = 13 \qquad \hat{\lambda} = \frac{2}{13} \approx 0.15$$

Likelihoods for different types of censoring

$$L = \prod_{i \in E} f(x_i) \prod_{i \in R} S(C_i) \prod_{i \in L} [1 - S(C_i)] \prod_{i \in I} [S(L_i) - S(R_i)]$$

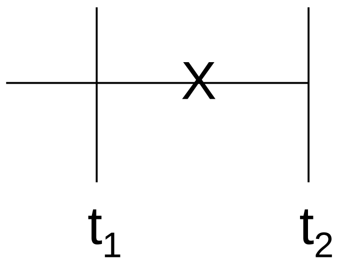
E = event

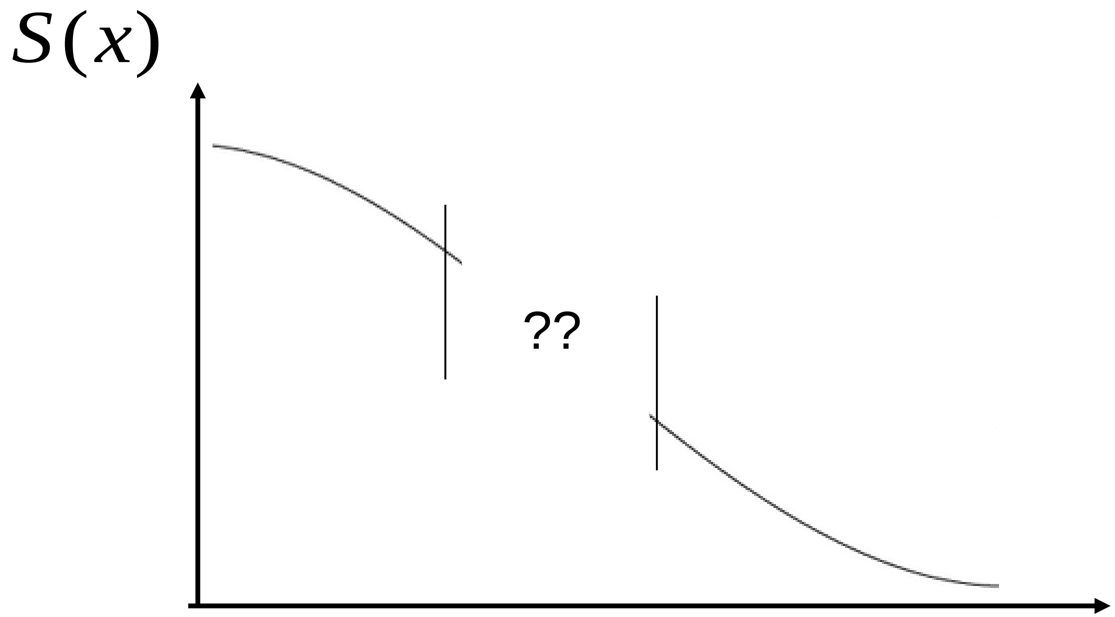
R = right censored

L = left censored

I = interval censored
L=Left interval time
R=Right interval time

Example: Interval censoring vs survival





$$S(t_1) > S(t_2)$$

Likelihood construction: right truncated data

Only events are observed (all times, T_i, are event times, X_i).

The probability to observe an event at T_i given that $T_i \leq Y_i$:

$$f(T_i | T_i \leq Y_i) = \frac{f(T_i)}{P(T_i \leq Y_i)} = \frac{f(T_i)}{1 - S(Y_i)} = \frac{f(X_i)}{1 - S(Y_i)}$$

$$L_i = \prod_i \frac{f(x_i)}{1 - S(Y_i)}$$

Likelihood construction: left truncated data

The observed time T_i is left truncated at Y_i.

Then we have to consider the conditional distribution of T_i given that $T_i \geq Y_i$:

$$f(T_i | T_i \geq Y_i) = \frac{f(t_i)}{P(T_i \geq Y_i)} = \frac{f(t_i)}{S(Y_i)}$$

Likelihood construction: left truncated data

Replace $f\,(x_i)$ by $\dfrac{f\left(x_i\right)}{S\left(Y_i\right)}$ Y = truncation time

and replace $S\,(C_i)$ by $\dfrac{S\left(C_i\right)}{S\left(Y_i\right)}$