3.1 Introducing time
(inclusive, exclusive, concurrent sampling)

Time-to-event data

Until now, we have dealt with binary (Yes/No) outcomes:

Disease, non-disease

Case, control

However, in a cohort study, where individuals are followed up, it is usual to know not only if the disease occurred but:

- For cases, **when** it was diagnosed
- For non-cases, **how long** they were known to be disease-free

Features of time-to-event data

1. The "outcome" of each individual has two components:
   - the time for which they were followed
   - whether or not they had the event of interest.

2. "censoring": subjects may:
   - enter the study at different times (left censoring)
   - not be observed for the whole period (right censoring)

Individuals who were followed for the entire study period and remained event-free are also defined as censored (this is called administrative censoring).

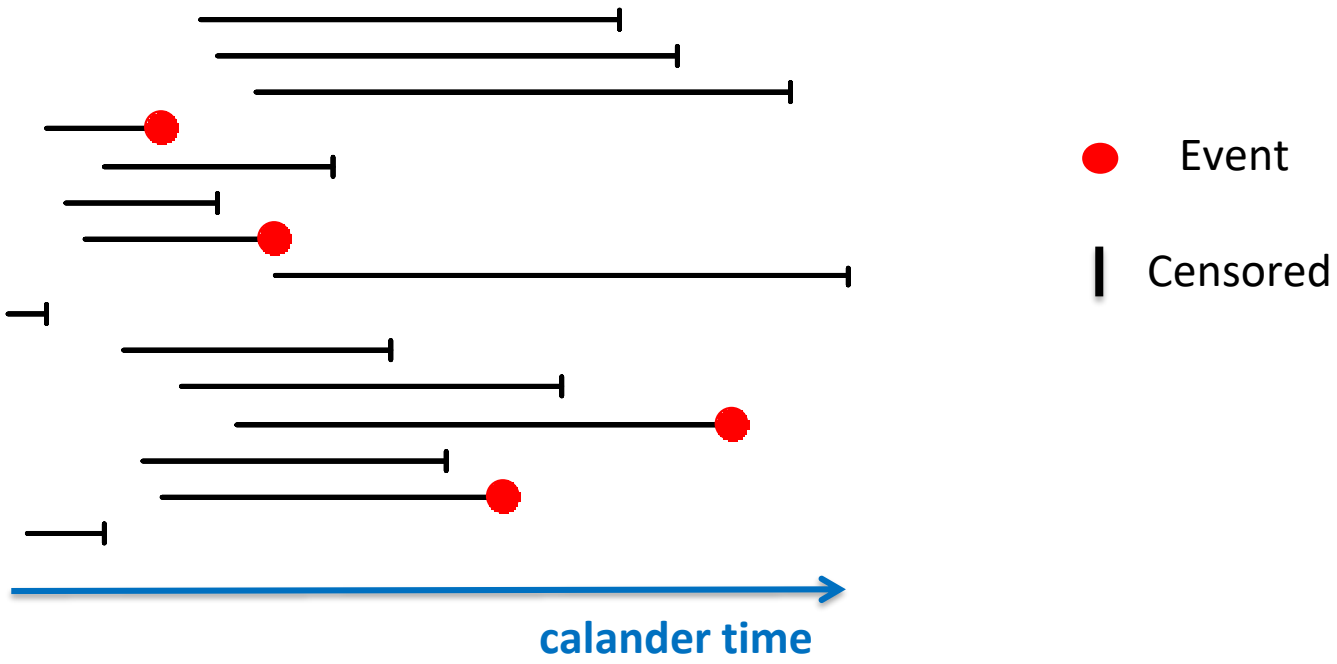
Illustration of censoring

The study subjects are enrolled at different calendar times.

Example below of 15 study subjects, 4 of whom had the event.

Others are censored, sometimes called “lost to follow-up” e.g. due to death, migration, or other reasons

Example:



Time is information!

Each individual´s period of observation starts when they join the study and stops when they:

- experience the outcome
- the follow-up period ends
- they are lost to follow-up (censored).

This leads to differences in the follow-up time of individuals, that needs to be considered in rates

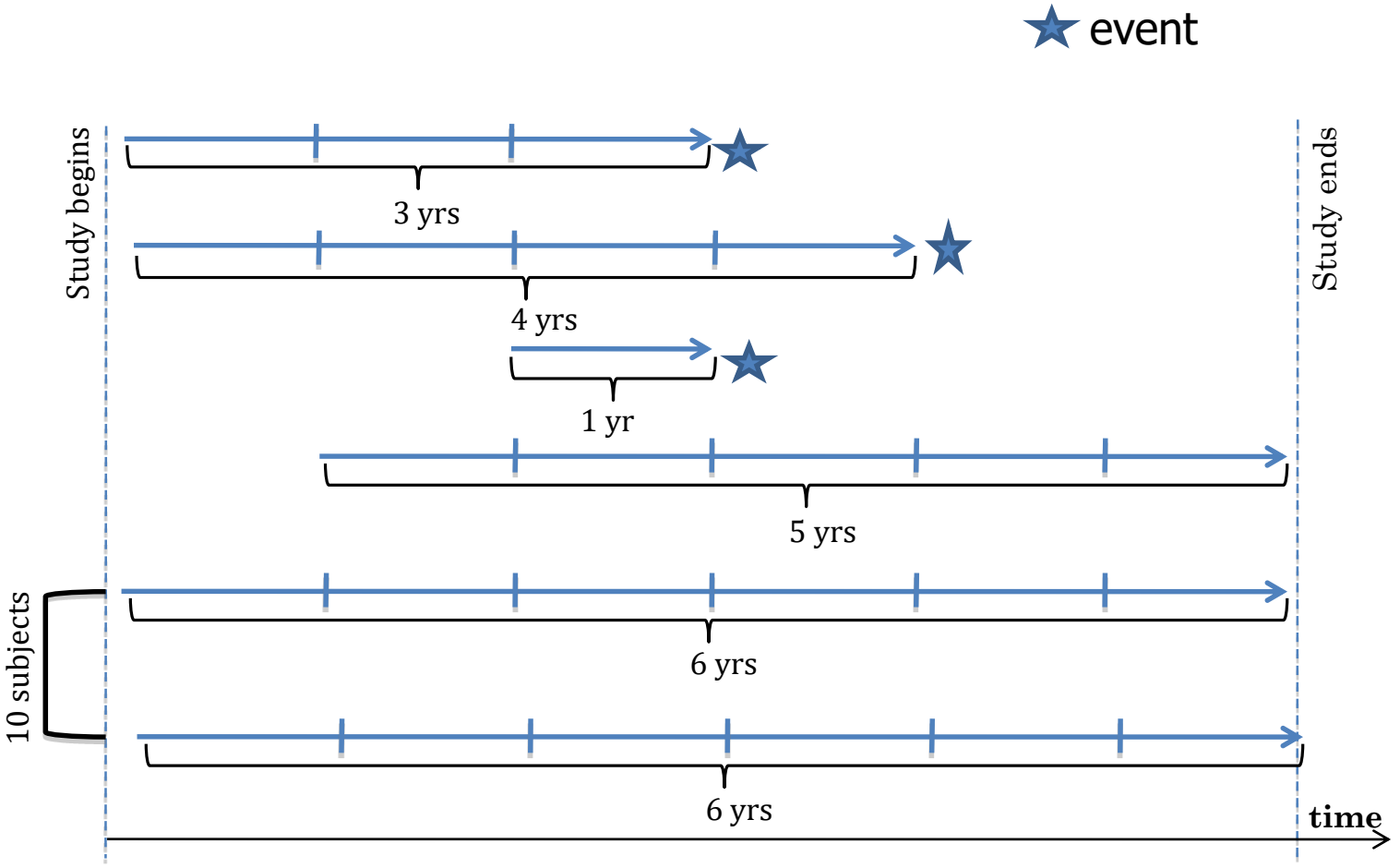
Recall the definition of a rate

The rate is defined as the events per **person-time of observation**.

$$\textbf{Incidence rate} = \frac{\text{Number of new cases of disease}}{\text{Total person-time at risk}}$$

$$\textbf{Mortality rate} = \frac{\text{Number of deaths}}{\text{Total person-time at risk}}$$

Example: Person-time



In total, these 14 subjects contribute 3+4+1+5+(6*10) = 73 **person-years**

Incidence rate = 3/73 events per person year = 4.1 per 100 py

Person-time, rates and risks

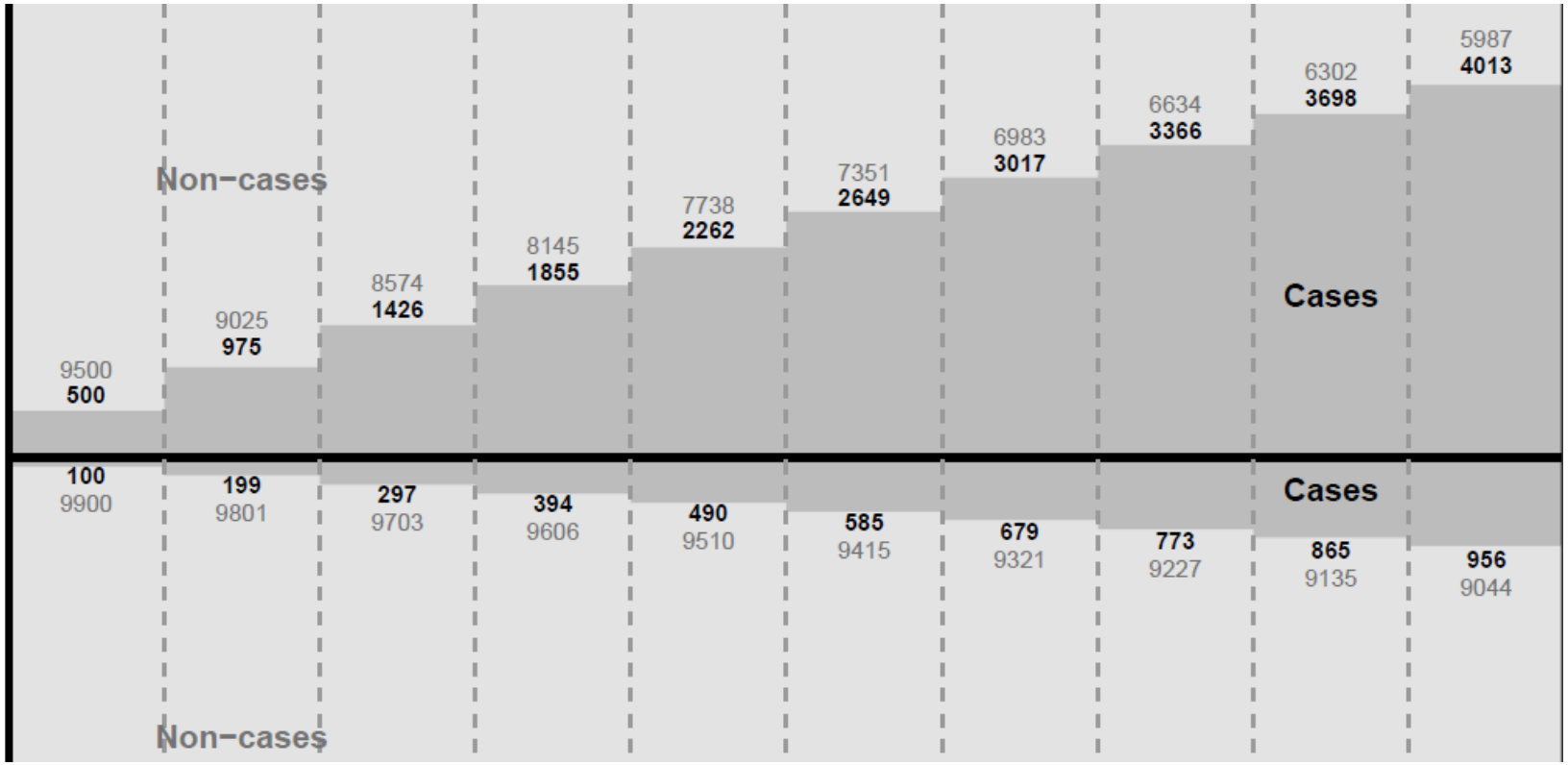
- An individual´s follow-up time (or time at risk) is the time during which, were they to experience the event, the event would be recorded in the study.

- The rate expresses the number of new events to the total observation *time* (in contrast to risk, in which the denominator is the number of *individuals* at risk at the beginning of the observation period.

- The longer the period of observation the greater the risk will be, since there will be more time for the event to occur.

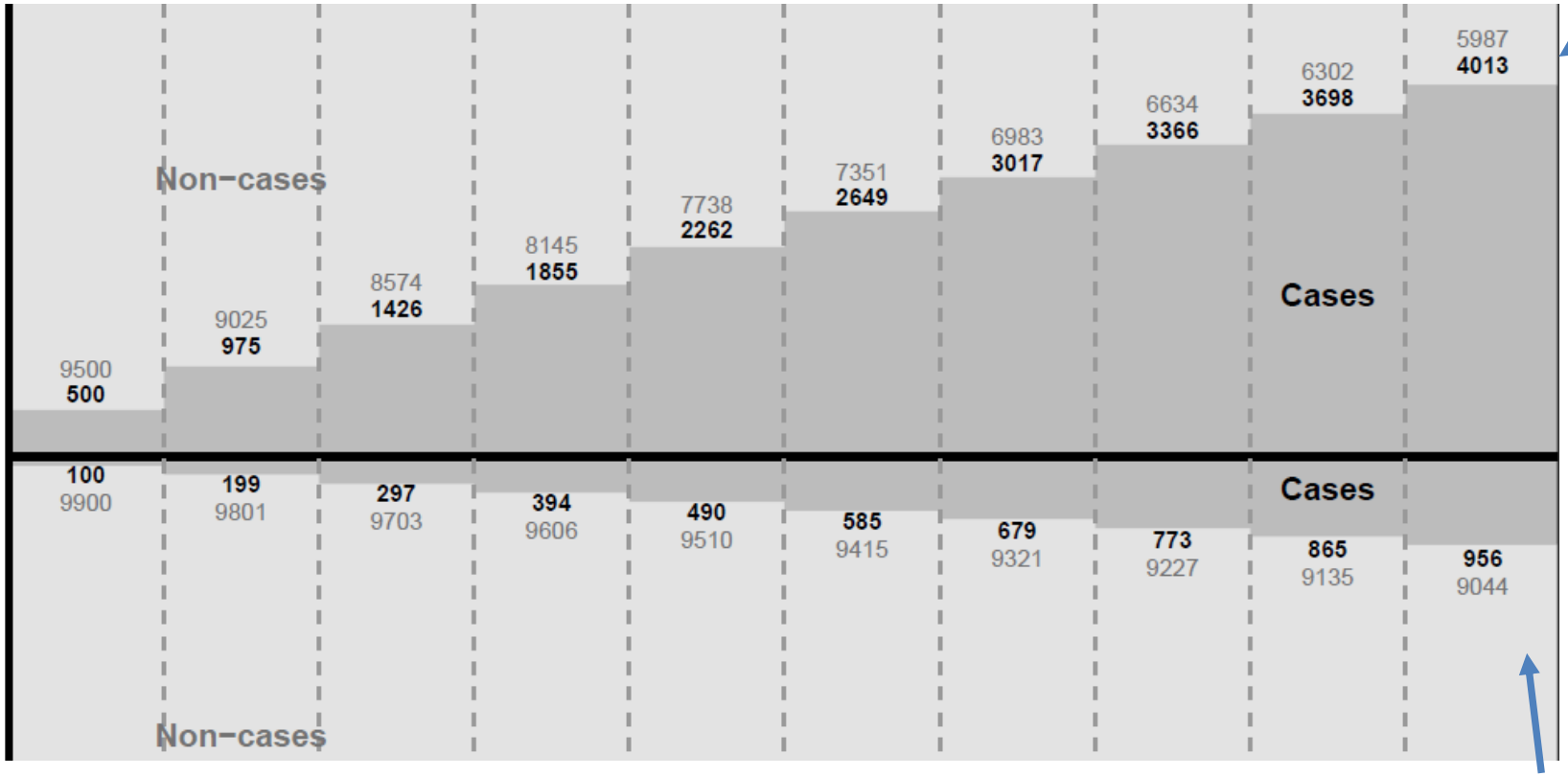
- Measures of risk therefore contain an implicit but not explicit time element.

Cohort of 10,000 exposed and 10,000 unexposed (incidence = 5/100 and 1/100 person years)



Cohort of 10,000 exposed and 10,000 unexposed (incidence = 5/100 and 1/100 person years)

Risk in exposed =4013/10,000



RR=4013/956 = 4.2

Risk in unexposed =956/10,000

To estimate RR from a sample

Select all cases (4969 at end of follow-up)

Select *sample of cohort at start* of follow-up

Odds of exposure among cases = 4013/956 =4.2

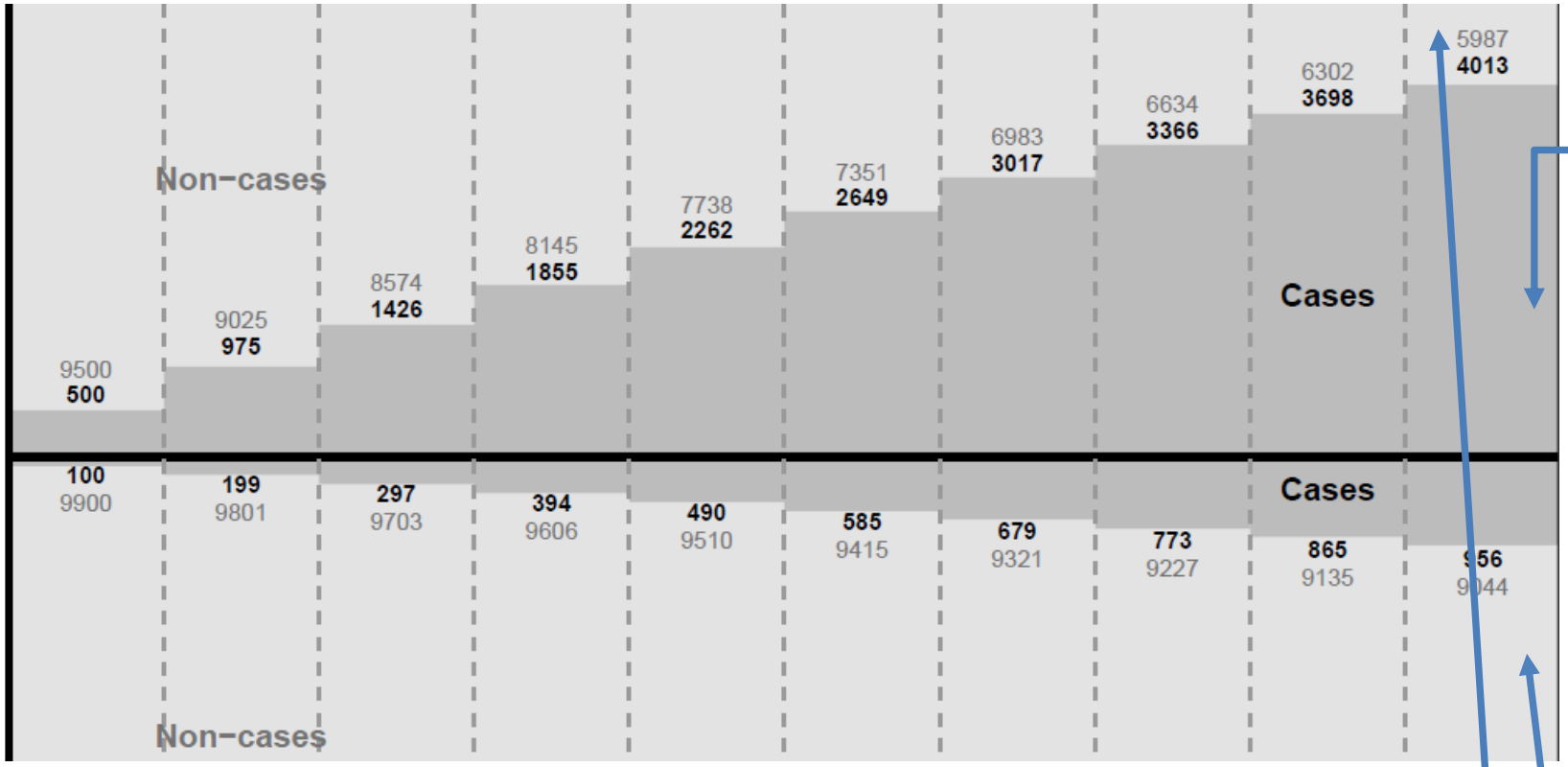
Odds of exposure among sample at start ≈ 1

OR= 4.2 =**RR** in population

This is ”inclusive sampling” / case-cohort sampling

Cohort of 10,000 exposed and 10,000 unexposed (incidence = 5/100 and 1/100 person years)

Odds of exposure in cases =4013/956



OR=4013/956÷ 5987/9044 =6.36

Odds exposure in controls= 5987/9044

To estimate OR from a sample

Select all cases

Select *sample of survivors at end* of follow-up

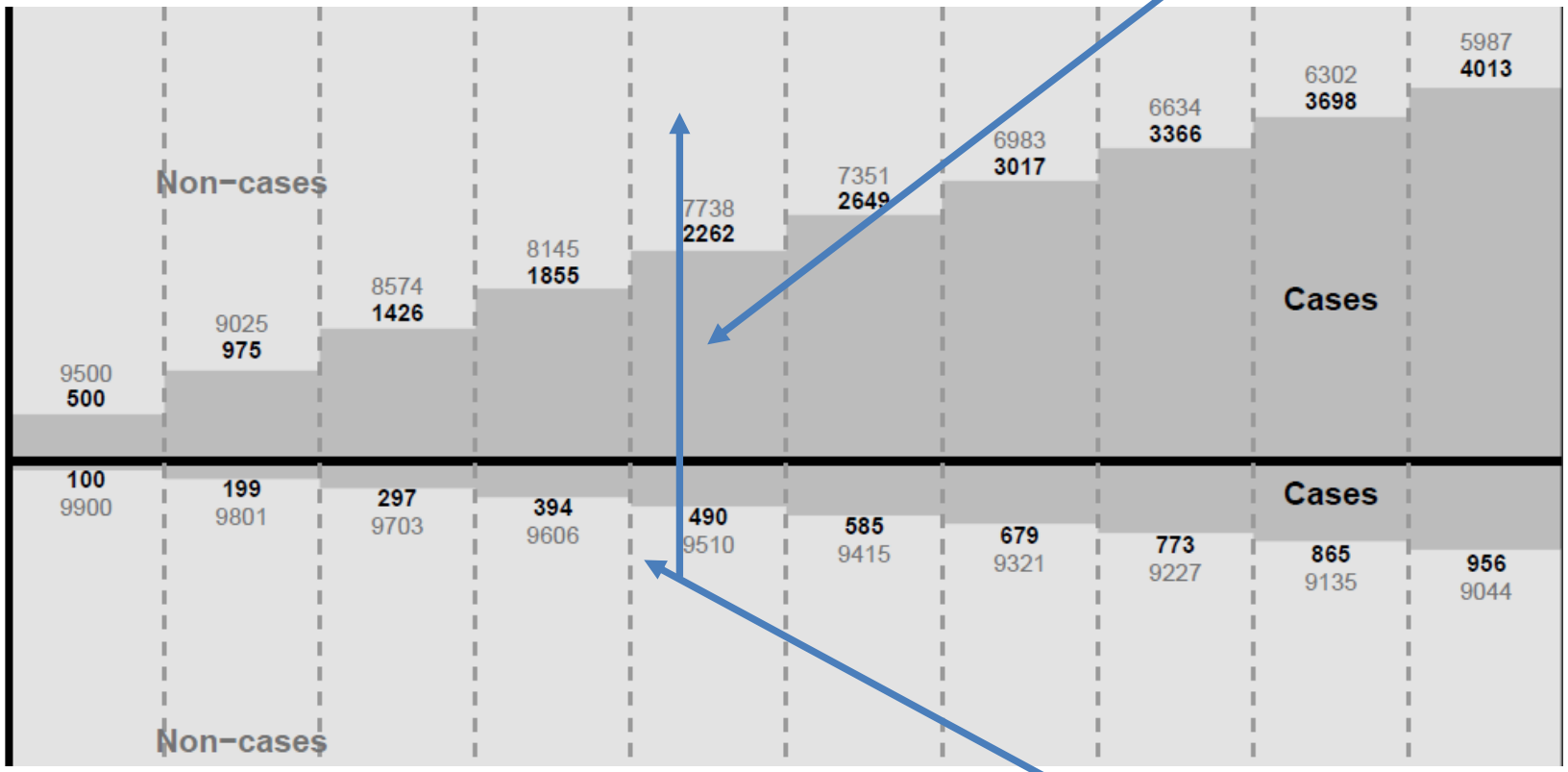
OR in sample = **OR** in population

e.g. odds exposure in cases = 4013/956= 4.2
odds of exposure in survivors ≈ 5987/ 9044 = .66
OR = 6.4

This is ”exclusive sampling” / classical case-control study

Cohort of 10,000 exposed and 10,000 unexposed (incidence = 5/100 and 1/100 person years)

Odds of exposure in cases (year 5)



OR= ■ ÷ ■ = ■

Odds exposure in controls (year 5)

Can we estimate IRR from a sample?

Yes!

Select all cases

Select sample of non-cases from ”concurrent” members of cohort
(”risk-set sampling”)

OR in sample = **IRR** in population
(we will demostrate this in the exercise)

Sampling ”controls” from a cohort

Design	Which controls
Classic CC	Sample from survivors at end (”exclusive” sampling)
Case-cohort	sample from start (”inclusive” sampling)
Nested CC	Incidence density or risk set sampling (”concurrent” sampling ”)

What OR estimates

OR from Design	estimates these parameters in study base
Classic CC	Incidence OR (≈ RR, IRR for ”rare disease”)
Case-cohort	RR
Nested CC	IRR