Validity, accuracy, precision

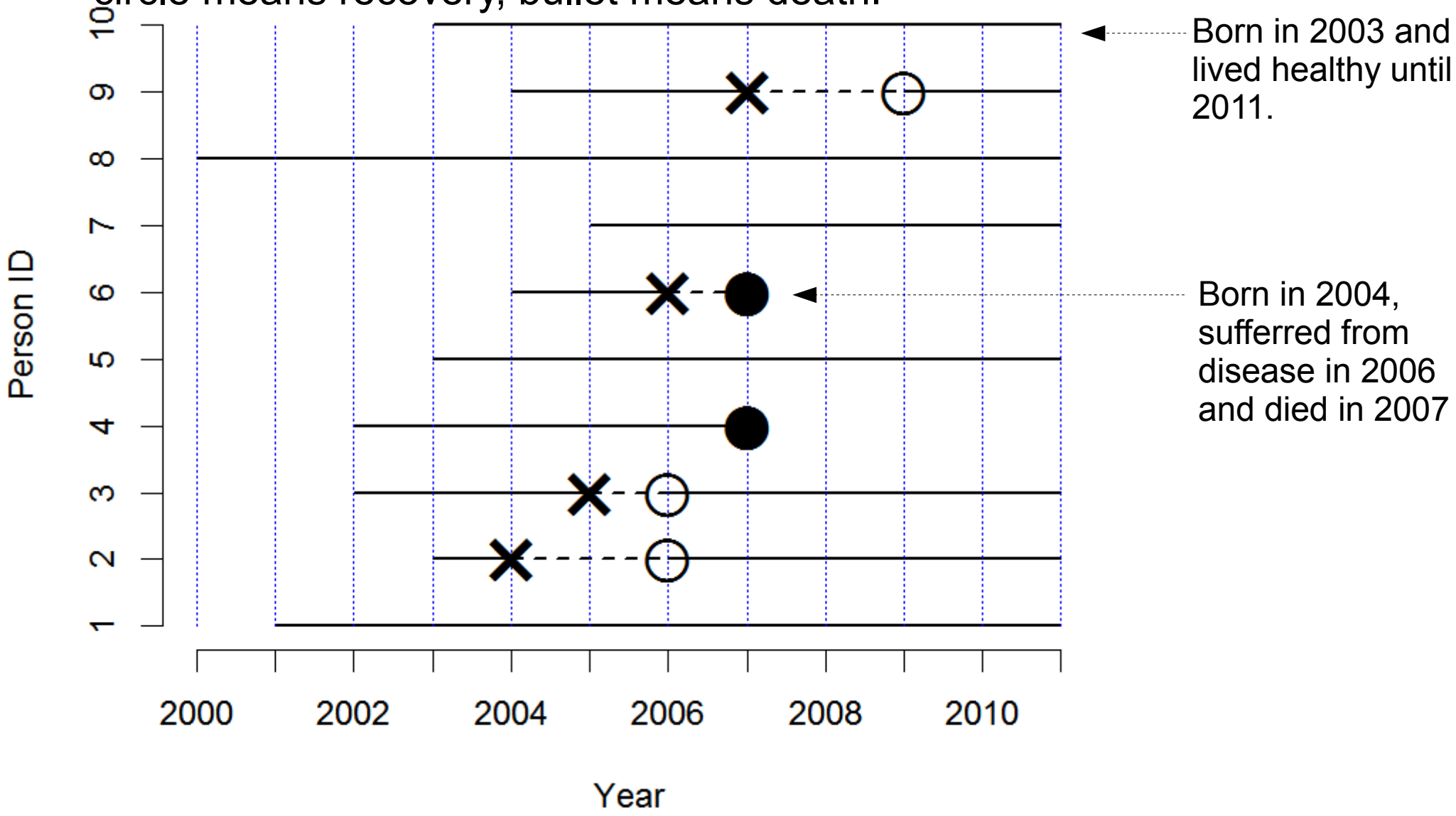
- Validity (妥当性)
  - Correctly measure what you want to measure
  - Less specificity in ELISA antibody, the measurement is invalid.
- Accuracy (正しさ，正確性)
  - Small bias (small systematic error)
  - Mistake in zero-adjustment cause inaccuracy
- Precision (精度)
  - Small random error
  - Less sensitive measurement has low precision

Cross tabulation

- Evaluating the extent of relationships between two categorical variables: the relationship between disease/not disease and exposed/unexposed.
  - How to evaluate the disease amount
    - prevalence（有病割合） / risk（リスク） / incidence rate（罹患率）
  - How to evaluate the extent of relationship between disease and exposure: the effect（効果）
    - difference（差） / ratio（比）
  - How to test and estimate / usage of p-value function
    - p-value function is recommended, but test is convenient.

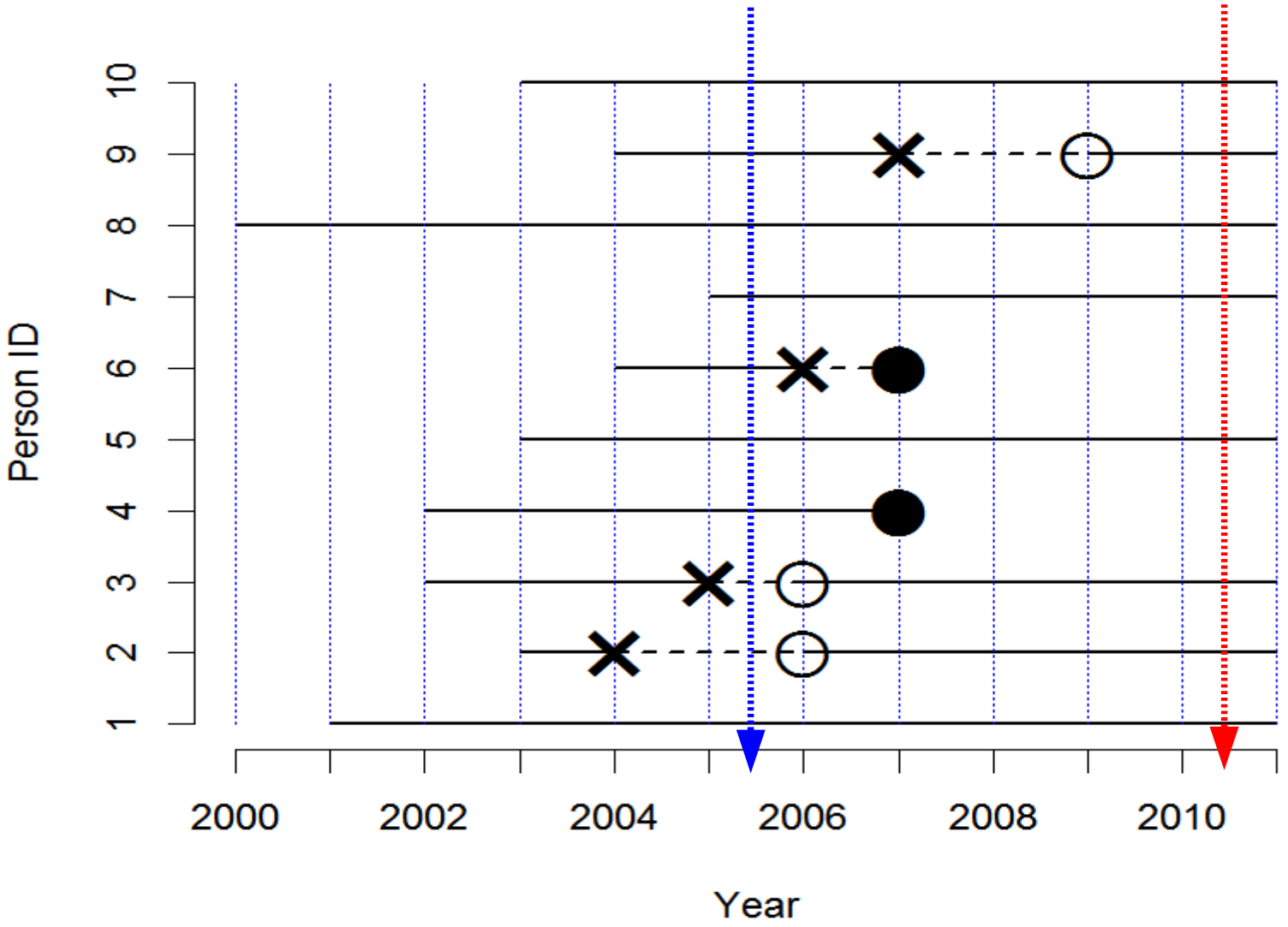
Evaluating disease amount

- Consider the actual situation of disease occurence. In the figure below, each line means 1 individual; x axis is years of observation; solid line means healthy and dashed line means sick, "x" is disease occurence, circle means recovery, bullet means death.



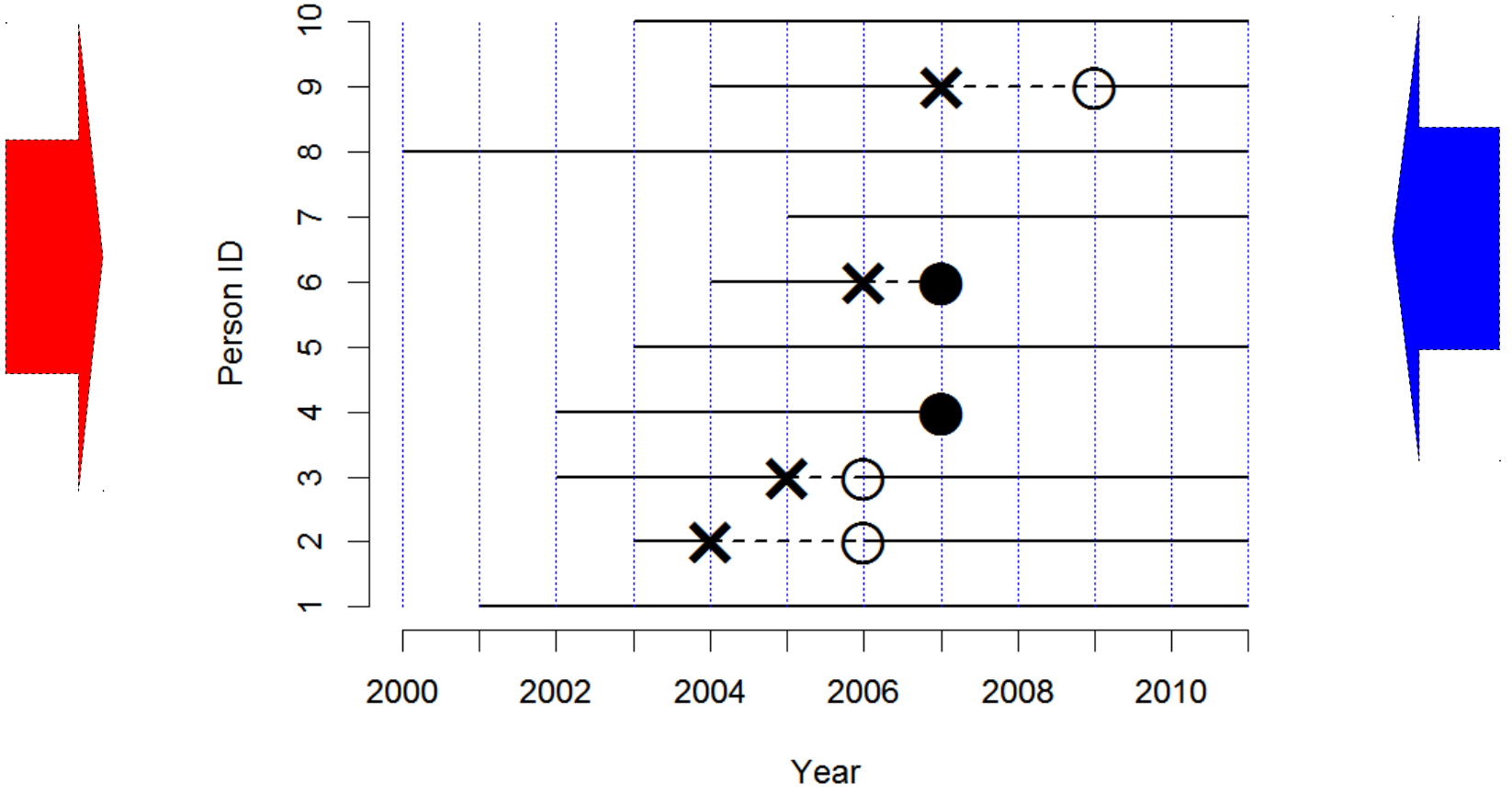
Prevalence by cross-sectional study

- The cross-sectional study in 2005 gives 0.2 (2 patients among 10) as the prevalence (2/8=0.25 as disease odds)←Easy, first-hand
- However, study in 2010 gives 0 as prevalence←Low representative



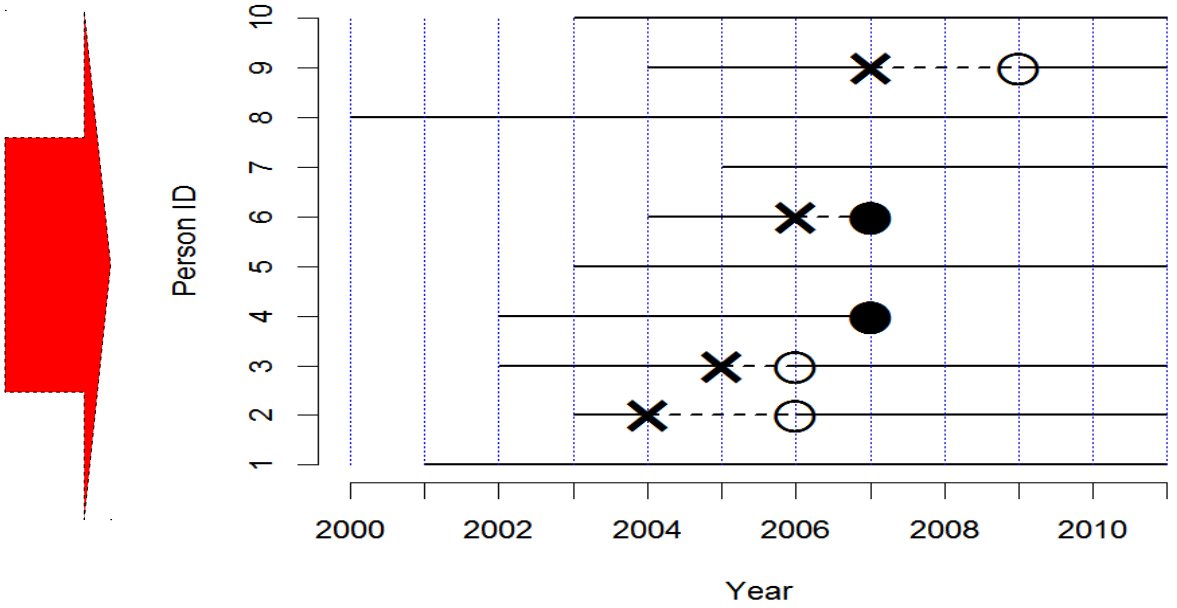
Risk (= cumulative incidence rate)

- Asking to parents of surviving 8 children in 2011, 3 disease experiences will be reported. So the risk for recent 11 years is 3/8=0.375←Easy, cheap study. But already-died cases are lost.
- Cohort study for 11 years since 2000 gives 4 disease incidence among 10 children. So the risk for 11 years is 0.4. However, the risk for 1 year after birth is 1/10=0.1←Risk depends on obs. period.



Incidence rate by cohort study

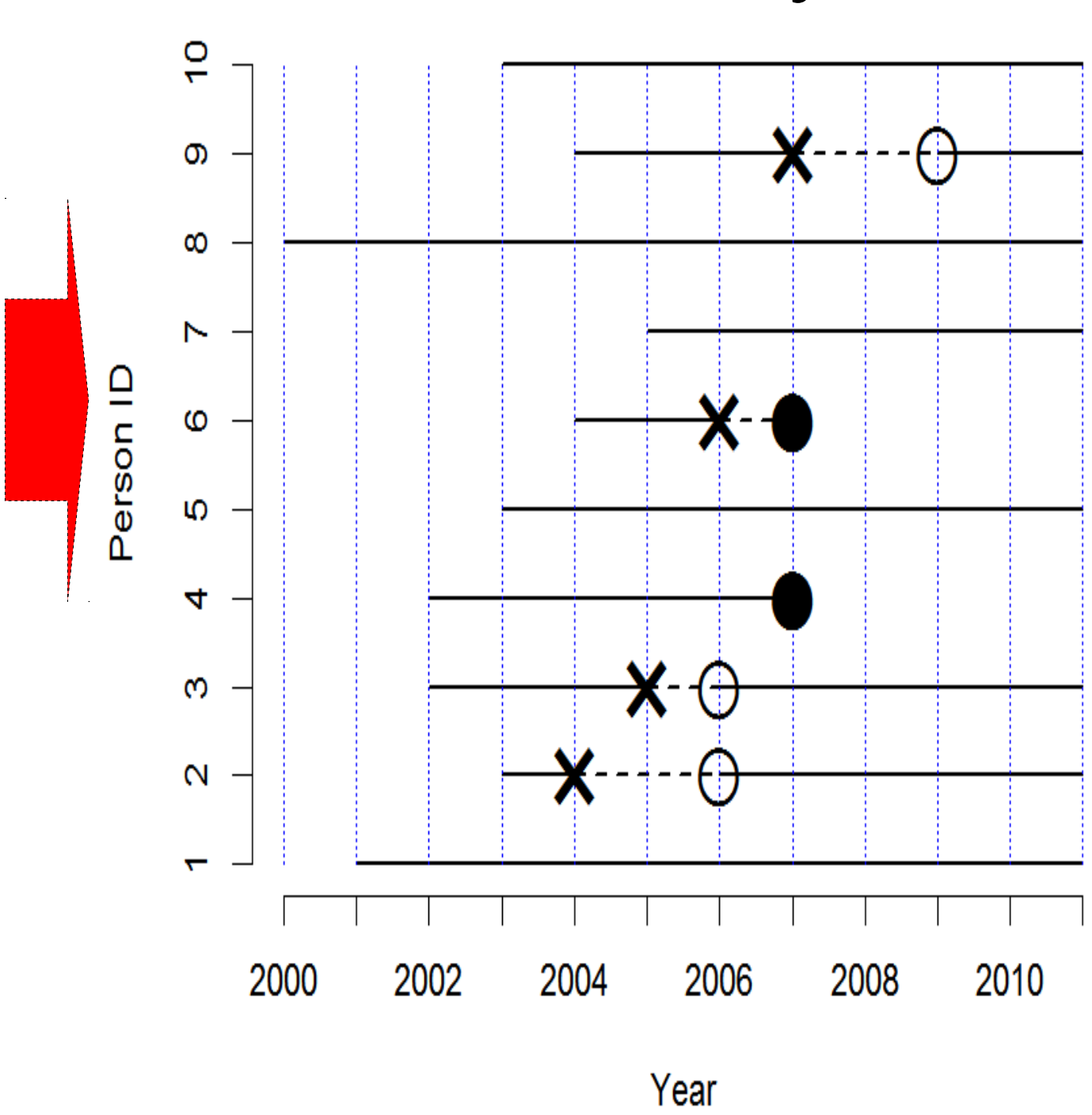
- The cohort study since 2000 can provide whole observation person-times.
- If the disease attacks only once in life, the individual looses susceptibility to the disease after the incidence of the disease. Patients are removed from the population at risk. Let the total of observation periods with susceptibility (the individual belongs to the population at risk) as denominator, and the total incidence of the disease as numerator, it becomes the "incidence rate". Dimension is 1/year.
- As summary measure for the diseases with multiple attacks, let the denominator mid-year population at risk, the numerator the incidence of the year. It gives incidence rate (usually per 100,000).



Let's calculate the incidence rate of "X".

The answer is in the next page.

Incidence rate by cohort study



Observation

8

3

11

6

2

8

5

3

1

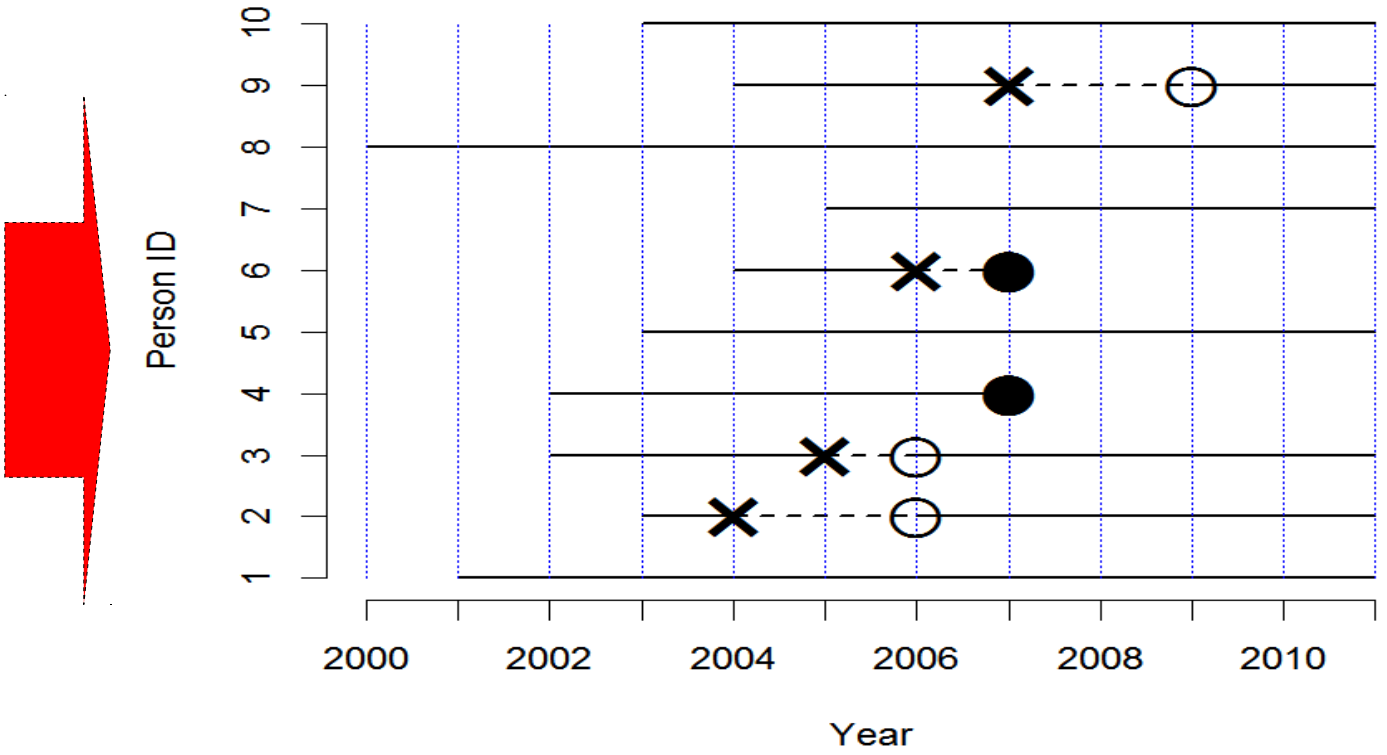
10

57 person-years

4/57=0.07(/yr)

Mortality rate

- Changing the end-point of observation from disease incidence to death gives mortality rate instead of incidence rate.
- Death is similarly treated as the disease attacking only once in life.
- As summary measure for large scale population, dividing the number of annual death by mid-year population gives annual mortality rate (usually per 1,000 or per 100,000).
  →in this case, 0.2/yr in 2007, 0 in other years.



Observed person-yrs
8→8
3→7
11→11
6→6
2→3
8→8
5→5
3→9
1→8
10→10
57→75 person-yrs
2/75=0.027

Relationship between exposure and disease = comparison of disease amount by exposure status

- Typical comparison
  - difference (absolute comparison)
  - ratio (relative comparison)
- Both meaningful
- Varies by disease amount
  - risk -> risk difference or risk ratio
  - incidence rate -> rate difference or rate ratio
  - mortality rate -> mortality rate difference or mortality rate ratio
  - prevalence -> "odds ratio"!!

Absolute comparison ~ risk difference or rate difference = attributable risk (=excess risk)

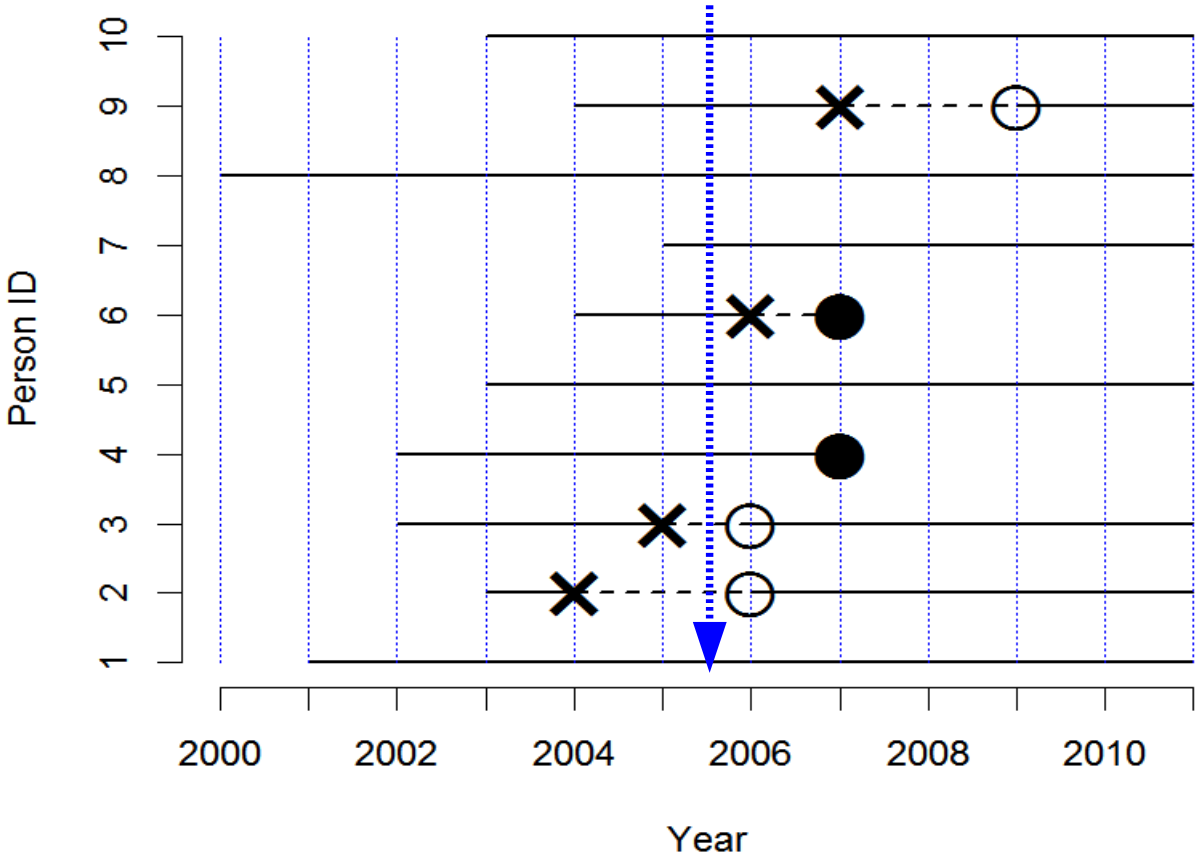
- (example) Followup 100000 individuals who live near power lines (Exposed Group) for 5 years. 2 leukemia patients are found every year. Followup 100000 individuals who live apart from power line (Control Group). 1 leukemia patient is found every year. There is no difference but residence.
- Risk difference = 10/100000-5/100000=5/100000 (=5e-5)
- Incidence rate difference = 10/(100000+99998+99996+99994+99992) -5/(100000+99999+99998+99997+99996) ≒0.0000100006（/year）
- The difference looks small because the incidence rate itself is very small.

Relative comparison (1) = risk ratio and rate ratio

- Same example
- Risk ratio = (10/100000)/(5/100000)=2
- Incidence rate ratio = $\frac{10/(100000+99998+99996+99994+99992)}{5/(100000+99999+99998+99997+99996)} \fallingdotseq 2$
- Both means "Living near the power line increases the leukemia risk about 2 times."
- Is it statistically significant?
  - Testing null-hypothesis that the ratio is 1.
  - It can be done using software like R (EZR), SAS, JMP

Relative comparison (2) Odds ratio

- In cross-sectional study, the prevalence proportion is the indicator of disease amount. However, difference or ratio of prevalence is not evaluated. Instead, odds ratio is calculated.
- If the subjects 2, 4, 6, 8, 10 are smokers in 2005, disease odds are 1/4 for smokers (exposed group) and also 1/4 for non-smokers (non-exposed group=control group). **The odds ratio is (1/4)/(1/4) = 1.**



* In case-control study, the odds ratio is the exposure odds of cases divided by the exposure odds of controls. If the subjects 2 and 3 are found as cases in 2005 and others are sampled as controls (and smokers are 2, 4, 6, 8, 10), the odds ratio is (1/2)/(4/8)=1.