Public Health Sciences 310
Epidemiologic Methods

Lecture 5
Bias

January 18, 2024

Brian C.-H. Chiu

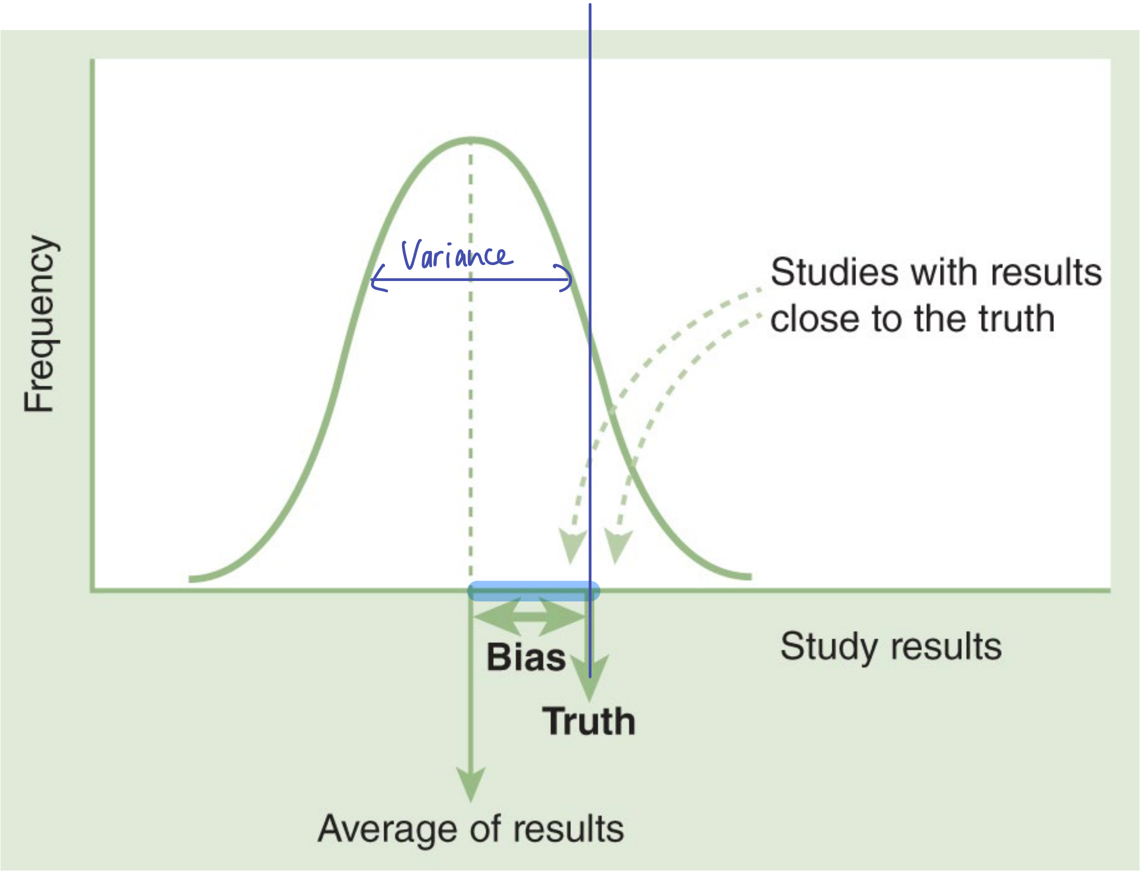
Possible Explanations for an Association Between a Risk Factor and a Disease

Explanation	**Assessment Strategy**
Random Variability (Sampling Error)	Estimation of precision (95% confidence interval)
Confounding	Experimental design; adjustment/matching
Bias (systematic error)	Quality assurance and quality control
Causal Relationship	Eliminate alternative explanations; causality guidelines

Epidemiological Definition of Bias

Deviation from the mean

- "Deviation of results or inferences from the truth, or processes leading to such deviation. Any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth."
  - Last J: A Dictionary of Epidemiology, ed. by J. Last, 3rd Edition, IEA



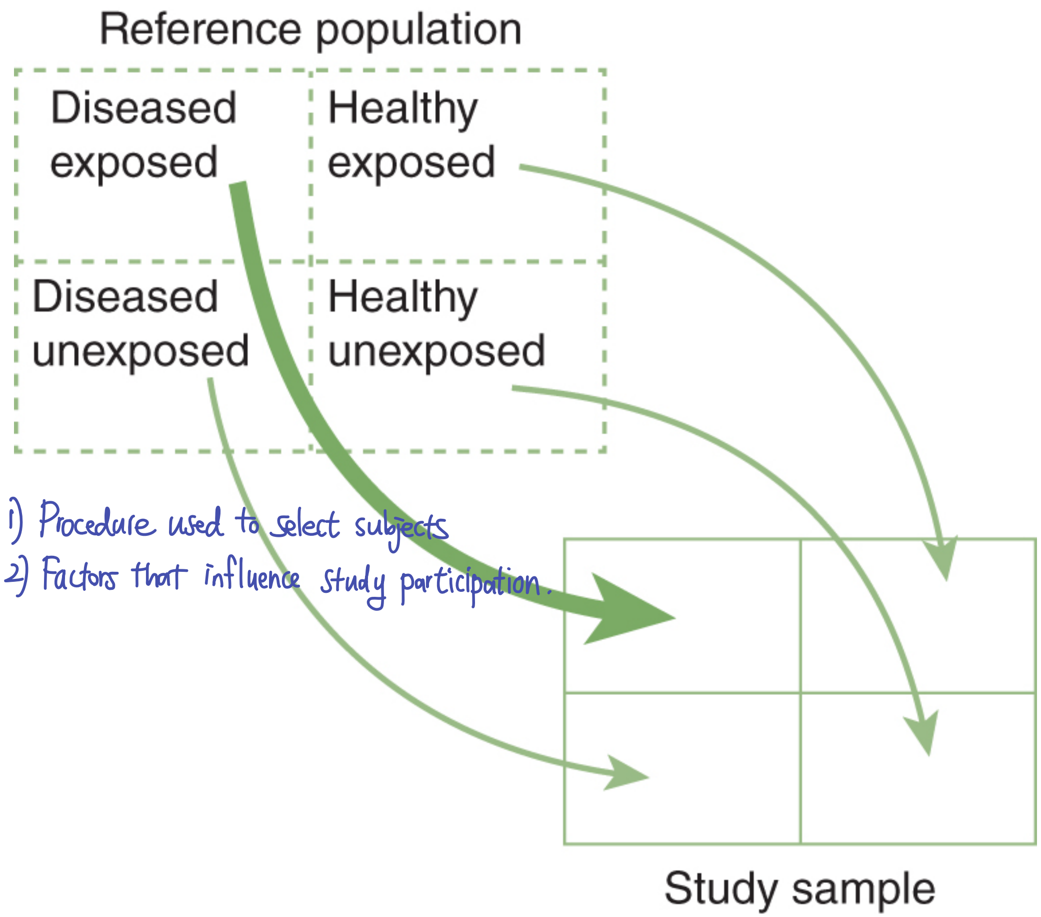
Main Types of Bias

- Selection Bias
- Information Bias (and misclassification)
- Mixed Biases

Selection Bias in Association Studies

A. Selection bias:
Occurs when the probability of ascertaining study subjects depends on both exposure and disease status (i.e., one group in the population has a greater likelihood of inclusion in the study)

More inclined to include a certain group in the Study



1) Procedure used to select subjects
2) Factors that influence study participation.

B. Selection bias can be a result of:

a. Differential response (participation) based on both exposure status and disease status. Respone / Participation Bias

b. Differential referral based on both exposure and disease status. Referral Bias

Response/Participation Bias

% of people participating in each group.

- Differential rates of participation between cases and controls does not itself introduce bias? No.
- EXAMPLE: true association:

No difference in exposure

	Case	Control
Exp	80 a	20 b
Unexp	20 c	80 d
	100	100

$$OR=\frac{\frac{{}^{a}80}{{}^{c}20}}{\frac{{}^{b}20}{{}^{d}80}}=16$$

True Association

· Positive association between exposure and cases (disease+).

WHAT IF: 90% of cases respond and 70% of controls respond

Difference in cases/controls does not produce bias

	Case	Control
Exp	80 (0.9)	20 (0.7)
Unexp	20 (0.9)	80 (0.7)

$$OR=\frac{\frac{80\,(\cancel{0}.9)}{20\,(\cancel{0}.9)}}{\frac{20\,(\cancel{0}.7)}{80\,(\cancel{0}.7)}}=16$$

(No Introduce of Bias)

Response/Participation Bias (cont'd)

- If the participation rates are not independent of both case status and exposure status, the OR will be biased. The direction of that bias depends on the response rates for each exposure-disease group.
- EAMPLE: 90% of exposed cases respond, 70% of unexposed cases respond, but 100% of controls respond.

Participation differs across cases within different exposures.

	Case	Control
Exp	80 (0.9)	20
Unexp	20 (0.7)	80

$$OR = \frac{\frac{80\ (0.9)}{20\ (0.7)}}{\frac{20}{80}} = 20.6$$

Overestimate:

Biased

Reason:

More paticipation for Cases of exposure, mistakenly ↑ OR by 1.285.

- Conclusions:
  - Biased exposure odds in cases
  - Unbiased exposure odds in controls
  - Biased odds ratio

Referral Bias

* Hospital-Based Case Control

· Less common in population-based C.C.

- Differential referral patterns of hospitalization for exposed and unexposed cases vs. controls can distort the estimates of association from a hospital-based case-control study

Unbiased population

	Stroke	No stroke
Diabetes	20	20
No diabetes	10	40

$$OR = \frac{\frac{20}{10}}{\frac{20}{40}} = 4.0$$

Well-known hospital

Hospital A

	Stroke	No stroke
Diabetes	18	10
No diabetes	5	20

$$OR = \frac{\frac{18}{5}}{\frac{10}{20}} = 7.2$$

Overestimate { Reason:
More likely to get referred,
receive more patients.

Country-side hospital

Hospital B

	Stroke	No stroke
Diabetes	2	10
No diabetes	5	20

$$OR = \frac{\frac{2}{5}}{\frac{10}{20}} = 0.8$$

Underestimate { Reason:
Less likely to
be referred to
receive patients.

Referral Bias (cont.)

SOLUTION

- Establish a hospital network that captures all cases within a defined geographic area. ex) Cook County: Need more than 2 hospitals from the area to establish network.
- Control (hospital or population) should be taken from the same geographic area as the cases. ↑ Hospital Network: ↓ Bias ☺
  ↑ Cost ☹
  ↓ Feasability ☹
  } "Trade-Off"

Berkson Bias (Admission Rate Bias)

* Hospital-Based Case Control

· Less common in population-based c.c.

Multiple comorbidities

- Patients with more than one disease or condition are more likely to be hospitalized than patients with only one disease or condition

General Population

	Stroke	No stroke
Cancer	10	30
No Cancer	30	90

True OR

$$OR = \frac{\frac{10}{30}}{\frac{30}{90}} = 1.0$$

Associated Hospitalization Rates

	Stroke	No stroke
Cancer	50% More likely for risk	10%
No Cancer	10%	5%

Hospitalized Population

	Stroke	No stroke
Cancer	5	3
No Cancer	3	5

Overestimate

OR=2.8

OR appears larger, appears "more sick".

Sutton-Tyrrell K. Stroke. 1991

Main Types of Bias

- Selection Bias
- Information Bias (and misclassification)
- Mixed Biases

Information Bias

- Some degree of misclassification of exposure or outcome information

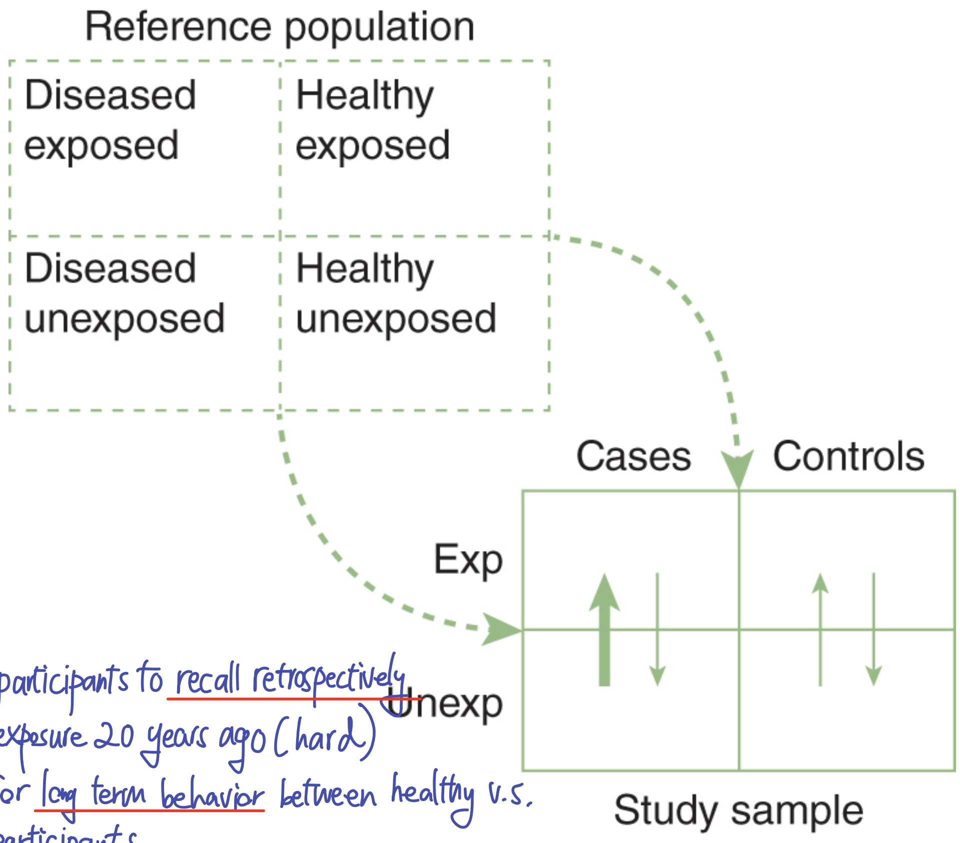
- In case-control studies: usually results from misclassification of exposure. Some degree of misclassification of the exposure information exists in both cases and controls, but unexposed cases in this example tend to mistakenly report past exposure to a greater extent than do controls.

ex) Recall Bias

ex) Questionnare:
- Ask participants to recall retrospectively their exposure 20 years ago (hard)
- Ask for long term behavior between healthy v.s. sick participants.

↓ sick ppl ↑ likely to care more therefore provide better information than healthy ppl.

- In cohort studies: usual results from misclassification of outcome, although there could also be misclassification of exposure



Examples of Information Bias

- Exposure Identification Bias Case Control Study → Retrospective
  - Recall Bias
  - Interviewer/Observer Bias
- Outcome Identification Bias Cohort Study → Prospective
  - Observer Bias
  - Respondent Bias

Recall Bias

- Inaccurate or differential recall of past exposure between cases and controls
- Result in misclassification of exposure status, thus biasing the results of the study
- Most often cited exposure bias
- Common concern in case-control studies : Study method, Data Collection.

Interviewer Bias

- When data collection in a case-control study is not masked with regard to the disease status of study participants, observer bias in ascertaining exposure, such as interviewer bias, may occur.
  - Interviewer bias may be a consequence of trying to “clarify” questions when such clarifications are not part of the study protocol, failing to follow either the protocol-determined probing or skipping rules of questionnaires.

Expectation or opinions of the interviewer interferes with judgement of participant.

Respondent Bias

- Outcome ascertainment bias may occur during follow-up of a cohort when information on the outcome is obtained by participant response.
  - Whenever possible, information given by a participant on the possible occurrence of the outcome of interest should be confirmed by more objective means.

When participant complete rating scales in ways that don't accurately reflect their true responses.

Two types of Misclassification Bias

- **Nondifferential (random) misclassification:**
  - Errors in assignment of group happens in more than one direction
  - Tends to bias the association toward the null OR=1
  - More Predictable Can also bias the association away from the null OR≠1
- **Differential misclassification:**
  - The degree of misclassification differs between the groups being compared
  - Less Predictable May bias the association either toward or away from the null hypothesis

Nondifferential Misclassification Bias

True Classification

	Case	Control	Total
Exp	100	50	150
Unexp	50	50	100

OR=2.0 RR=1.3

Nondifferential misclassification: overestimate exposure in 10 cases, 10 controls – Bias towards null

	Case	Control	Total
Exp	110 +10	60 +10	170
Unexp	40 -10	40 -10	80

Underestimate of true strength

↓OR=1.8 =RR=1.3

Misclassify across all participants with equal probability
(Ratio relatively =)

Differential Misclassification Bias (Worst)

True Classification

	Case	Control	Total
Exp	100	50	150
Unexp	50	50	100

OR=2.0 RR=1.3

Differential misclassification: overestimate exposure in 10 cases – Inflate rates

	Case	Control	Total
Exp	110 +10	50	160
Unexp	40 −10	50	90

↑OR=2.8 ↑ RR=1.6

Proportion of misclassify subjects differs between study groups.

(Ratio relatively ≠)

Differential Misclassification Bias (cont.)

True Classification

	Case	Control	Total
Exp	100	50	150
Unexp	50	50	100

OR=2.0 RR=1.3

Differential misclassification: overestimate exposure in 10 controls – deflate rates

	Case	Control	Total
Exp	100	60 +10	160
Unexp	50	40 -10	90

↓OR=1.3 ↓RR=1.1

Differential Misclassification Bias (cont.)

True Classification

	Case	Control	Total
Exp	100	50	150
Unexp	50	50	100

OR=2.0 RR=1.3

Differential misclassification: underestimate exposure in 10 cases – deflate rates

	Case	Control	Total
Exp	90 −10	50	140
Unexp	60 +10	50	110

↓OR=1.5 ↓RR=1.2

Differential Misclassification Bias (cont.)

True Classification

	Case	Control	Total
Exp	100	50	150
Unexp	50	50	100

OR=2.0 RR=1.3

Differential misclassification: underestimate exposure in 10 controls – inflate rates

	Case	Control	Total
Exp	100	40 −10	140
Unexp	50	60 +10	110

↑ OR=3.0 ↑ RR=1.6

Main Types of Bias

- Selection Bias
- Information Bias (and misclassification)
- Mixed Biases

Mixed Biases

- Cross-Sectional Biases
  - Incidence–Prevalence Bias Neyman Bias

    Measure simoutaneously
- Biases Related to the Evaluation of Screening Interventions

Incidence-Prevalence Bias

ex) Etiology

portion of ppl†, portion of new ppl†

Prevalence ↓, Incidence ↑.

·Actually have more cases but are not able to be included in the study.

·Did not include true risk.

- Those persons who develop the disease but die prior to the time of study obviously cannot be included in the study population. If exposure status happens to be over- or underrepresented in the survivors, then the use of prevalence data to estimate incidence-based risk of odds ratios can lead to biased results.

ex) Disease detected for a long time

"Make disease appear less severe."

- Persons diagnosed with a disease may modify their risk factors upon diagnosis (esp. biological measures, diet)

Incident vs Prevalent Cases (cont'd)

- EXMAPLE: Framingham Cohort in which the association between CHD risk and hypercholesterolemia was investigated

Incident cases

	Dev CHD by exam 6	Did not dev CHD by exam 6
High chol at exam 1	85	462
Lower chol at exam 1	116	1511

↑

$$OR = \frac{85 \times 1511}{462 \times 116} = 2.4$$

when individuals who + cases but died are excluded:
Biased in estimating association between risk and exposure

Prevalent cases

	CHD at exam 6	No CHD at exam 6
High chol at exam 6	38	34
Lower chol at exam 6	113	117

$$OR = \frac{38 \times 117}{34 \times 113} = 1.16$$

"There's actually more risk than you think!"

Length and Lead-Time Bias

Overestimation of Survival

- Length Bias:

  Length Time Bias

  ex) slow growing tumor

  - Disease detected by screening is less aggressive than the disease detected without screening or between screening exams (interval). So, cases detected on screening appear to live longer.

    ex) aggressive tumor

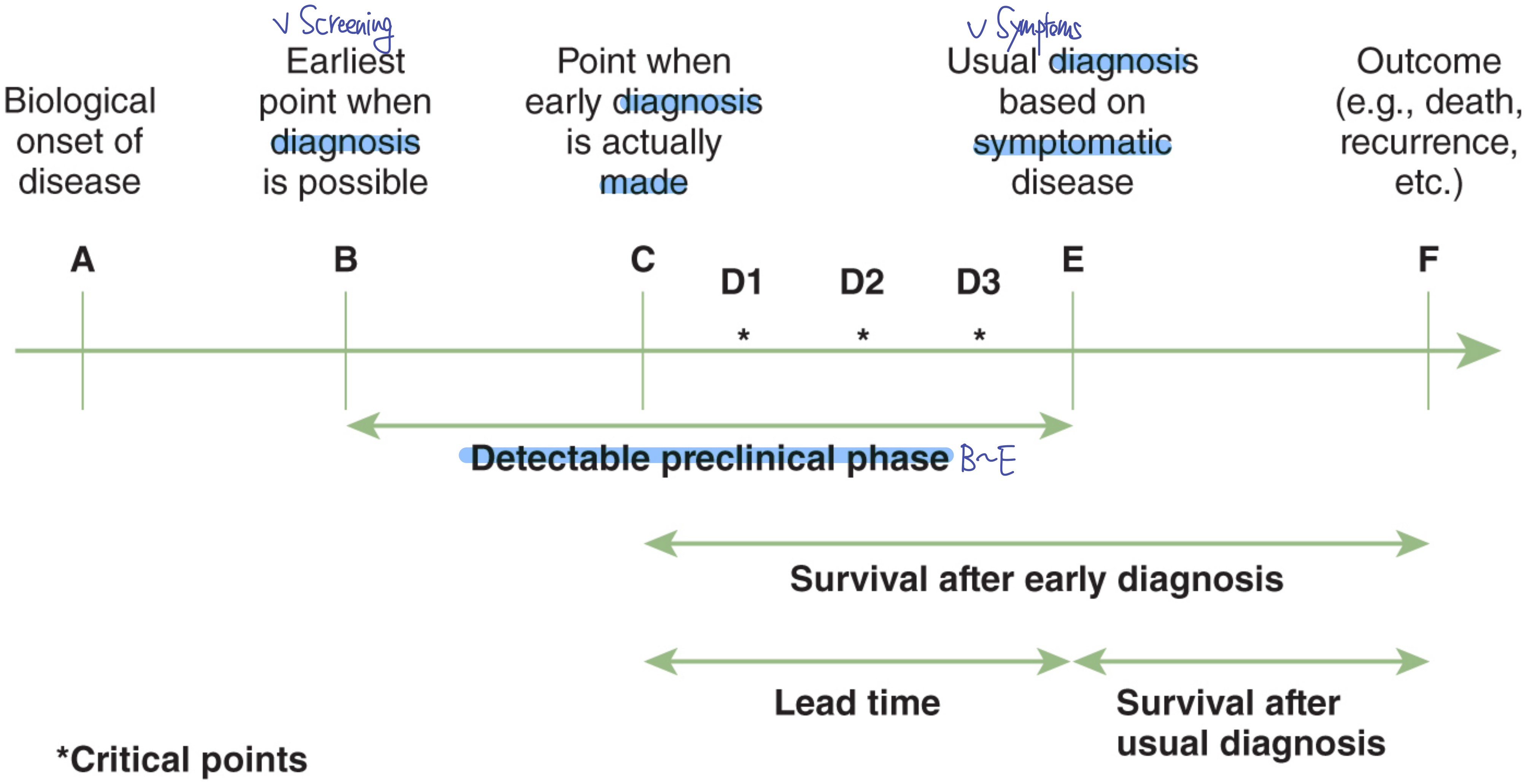
  · Slower progressing diseases have better prognosis → lead to longer survival.

- Lead Time Bias:

  - An increase in survival as measured from the disease detection until death, without lengthening life.

  · Early diagnosis of disease fasely make it appear patient is living longer.

Natural History of Disease



√ Screening

√ Symptoms

A: Biological onset of disease

B: Earliest point when diagnosis is possible

C: Point when early diagnosis is actually made

D1 * D2 * D3 *

E: Usual diagnosis based on symptomatic disease

F: Outcome (e.g., death, recurrence, etc.)

Detectable preclinical phase B~E

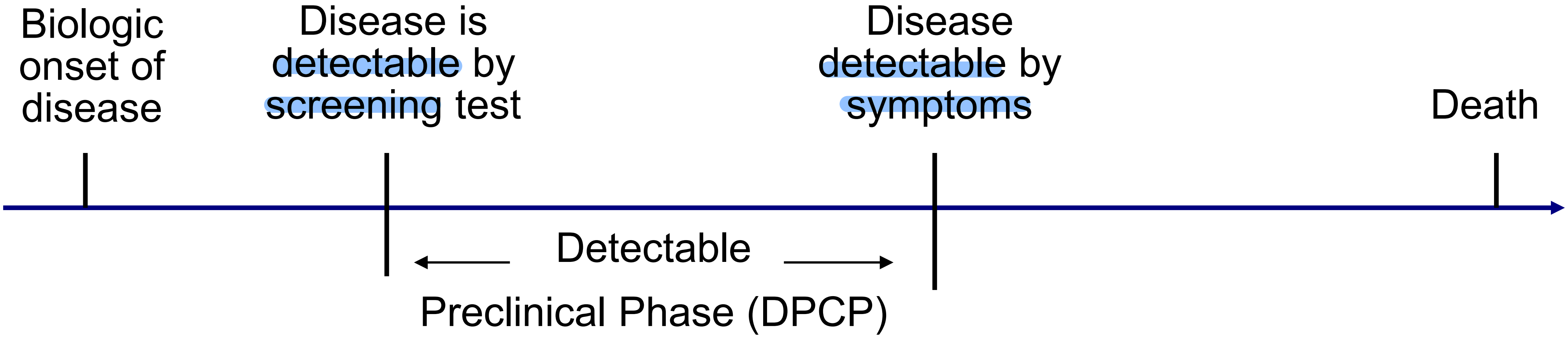
Survival after early diagnosis

Lead time

Survival after usual diagnosis

*Critical points

Length Bias



Biologic onset of disease

Disease is detectable by screening test

Disease detectable by symptoms

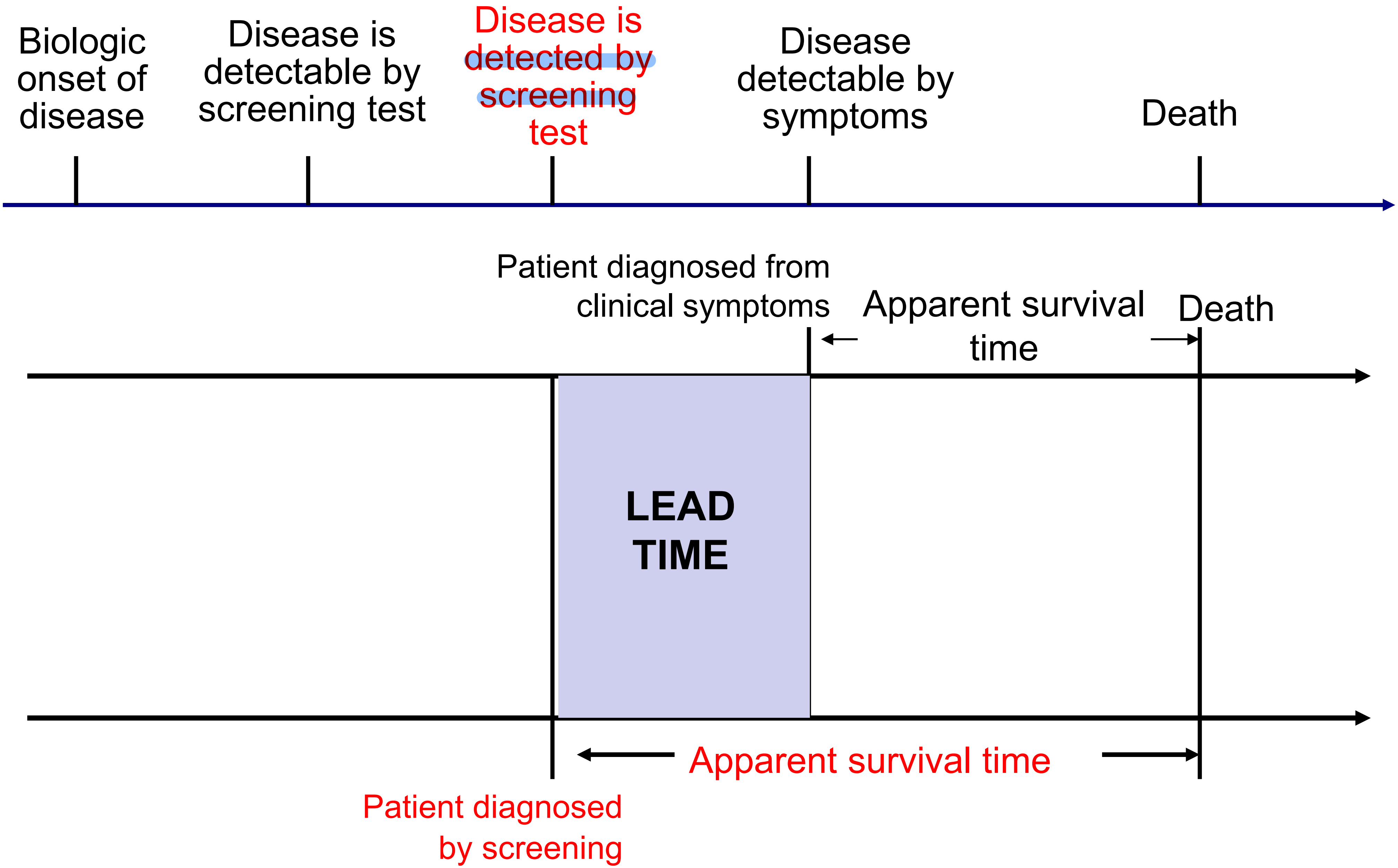
Death

Detectable Preclinical Phase (DPCP)

- DPCP is directly related to the rate of disease progression.
- Effectiveness of screening is positively related to length of the DPCP.
- Length bias occurs when disease detected by screening is simply catching the diseases with a better prognosis, i.e. where the length of the DPCP is long → Prolong survival

ex) slow progressive, less aggressive disease.

Lead Time Bias



Biologic onset of disease

Disease is detectable by screening test

Disease is detected by screening test

Disease detectable by symptoms

Death

Patient diagnosed from clinical symptoms

Apparent survival time

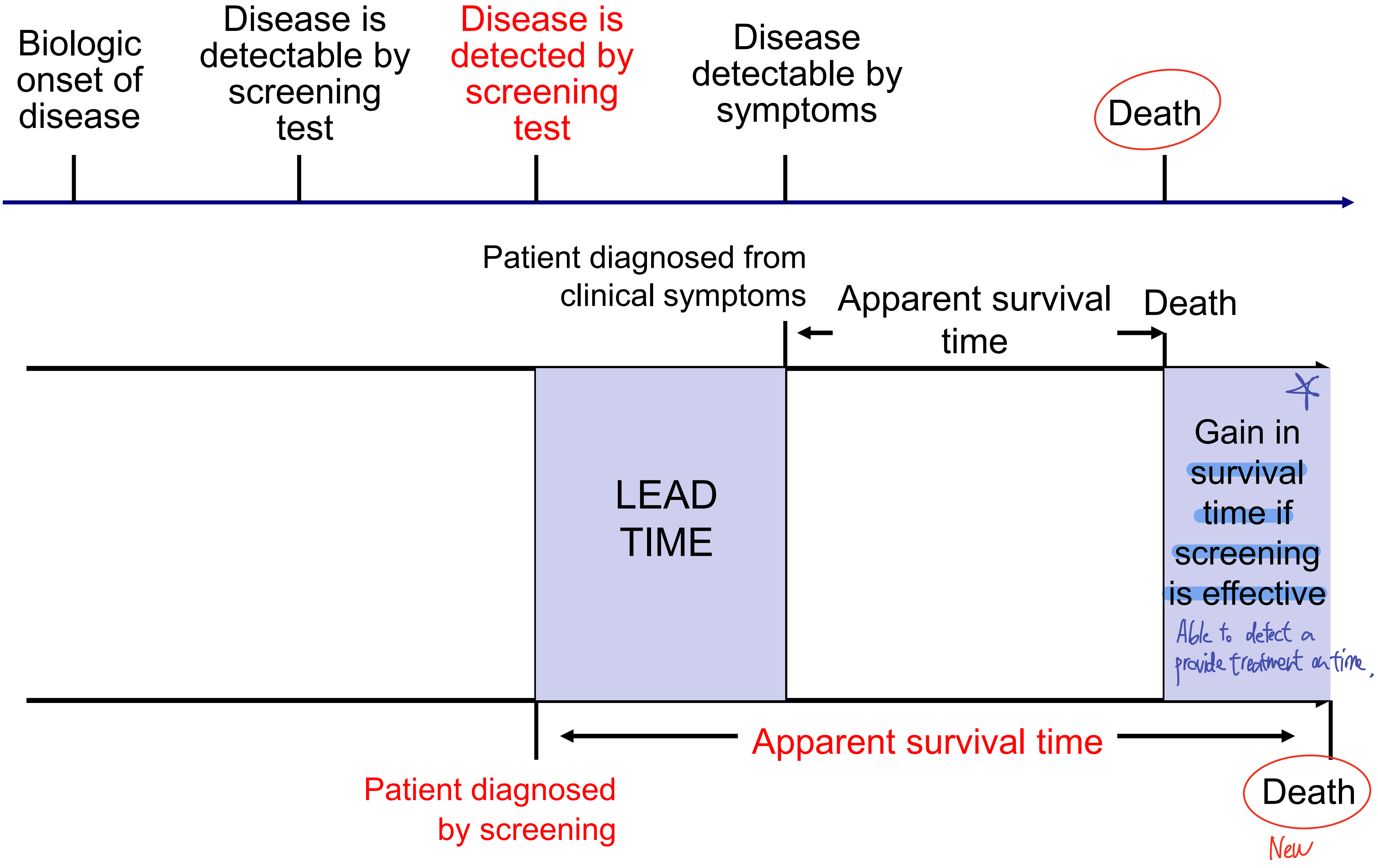
Death

LEAD TIME

Apparent survival time

Patient diagnosed by screening

Lead Time Bias (cont.)



Biologic onset of disease

Disease is detectable by screening test

Disease is detected by screening test

Disease detectable by symptoms

Death

Patient diagnosed from clinical symptoms

Apparent survival time

Death

LEAD TIME

Gain in survival time if screening is effective

Able to detect a provide treatment on time.

Apparent survival time

Patient diagnosed by screening

Death

New

Control of Bias

It is important to minimize any biases that may occur at any stage of the study, particularly when the association is weak.

Prevent or control bias on three phases

- Ensure study design is appropriate for hypothesis
- Establish careful data collection procedures
- Use appropriate analytic techniques

Today: Common in observational study from study participant.

Next: Medical Record related Bias?