METR/ENVS 113

Lecture 10: Introduction to Epidemiology

SJSU Fall Semester 2020

Module 4: Air Pollution Epidemiology

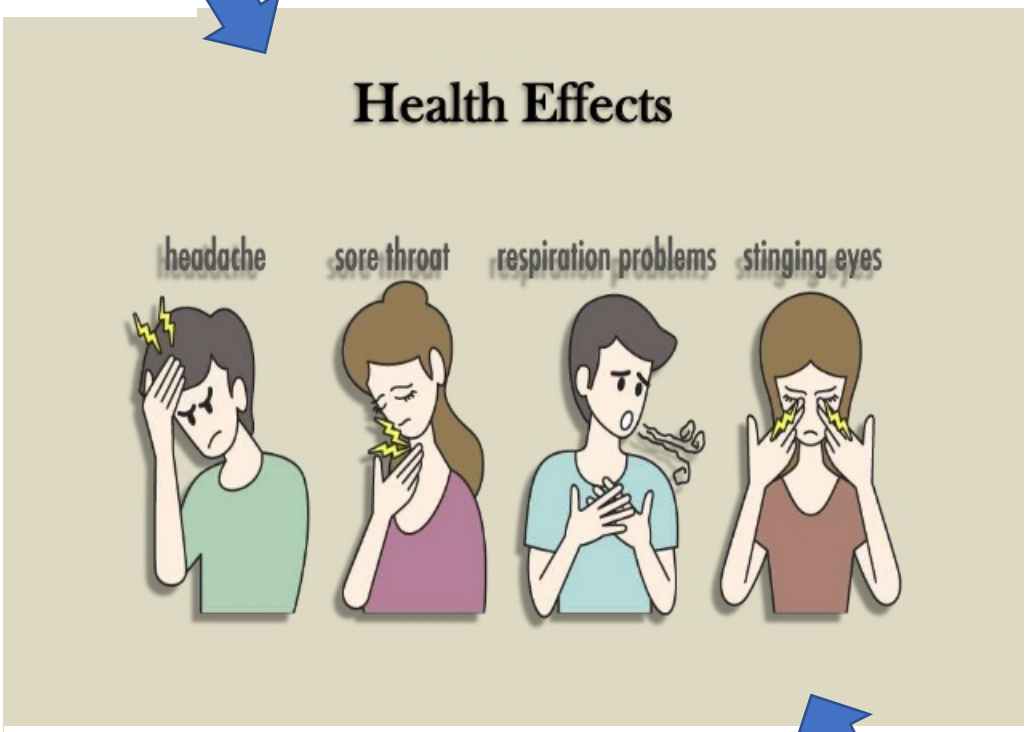
Frank R. Freedman (Course Instructor)

Environmental Factors

Air Pollution?

Other pollution sources (water, soil, indoor environment)?

?



?

Population

Hereditary, Gender, Age?

Diet, Smoking, Exercise, Body Weight?

Pre-Existing Health Condition?

Life-Style, Stress, Anxiety?

Module 4: Outline

- **Introduction to Epidemiology** (Lecture 10)
- **Video Exercise:** *Salmonella* in the Caribbean
- **Video Exercise:** Long-term health effects of PM2.5, Epidemiological Studies, Arden Pope talk

Lecture 10: Introduction to Epidemiology (Outline)

- **Background**
  - Exposure vs Outcome
  - Experimental Study Design & Randomized Control Trials
  - Presenting Results of Epidemiological Studies
  - Absolute Risk & Risk Ratio
- **Epidemiological Study Design**
  - Prevalence vs. Incidence
  - Cross-Sectional
  - Case-Control
  - Cohort
- **Causal Inference**
  - Chance, Statistical Significance, Confidence Intervals)
  - Bias & Confounders

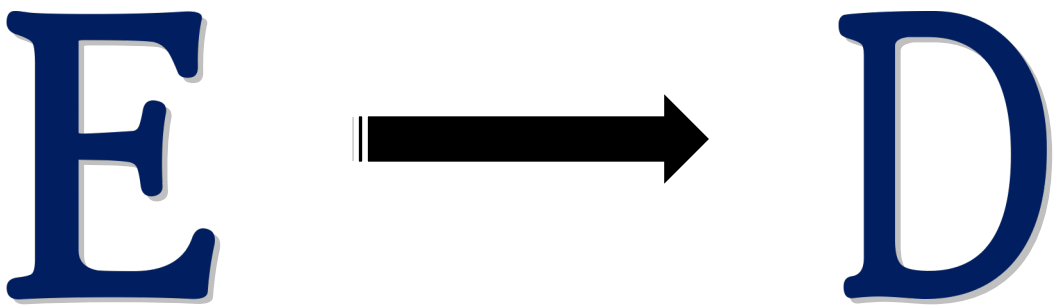
This lecture is adapted from the original PPT lecture ***"Introduction to Epidemiology and Study Designs"***
by Thomas Songer, Ph.D., University of Pittsburgh and the Supercourse Team

Background

Basic Question

(Epidemiological Studies)

Are exposure and disease linked?



"Exposure"

"Disease"
"Outcome"

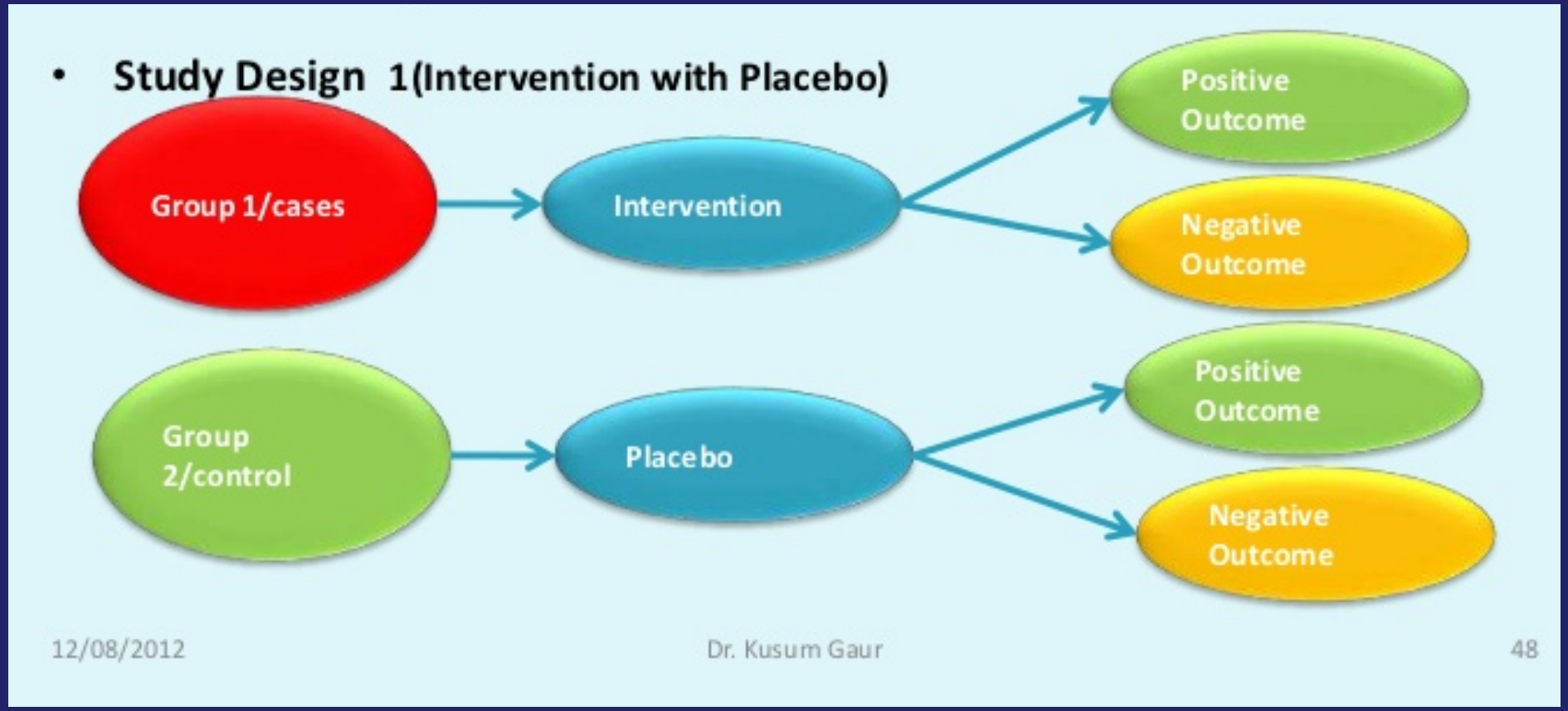
Experimental Study

Classic "placebo-controlled" experimental laboratory study

Four possible outcomes

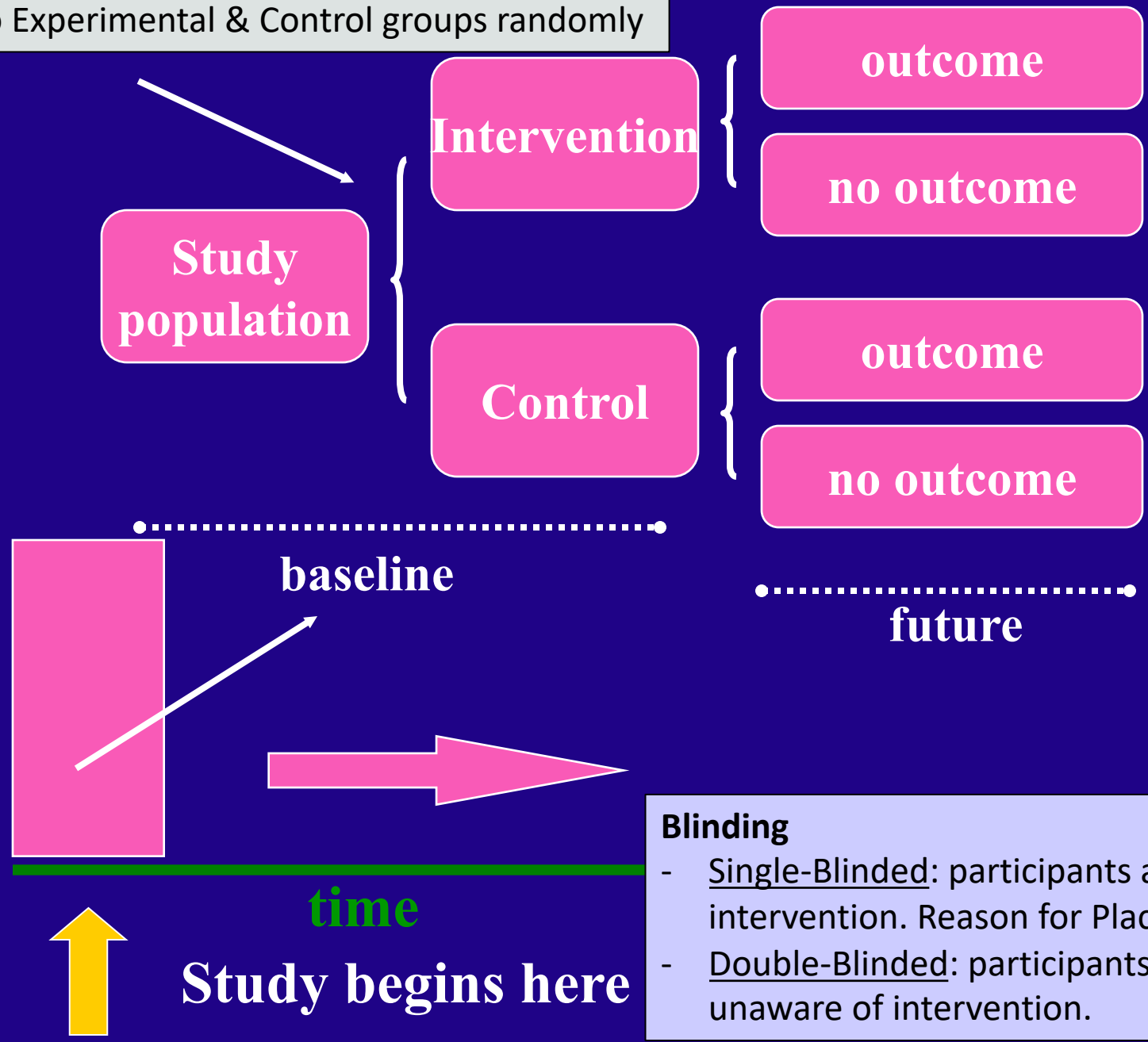
Experimental

Control



Experimental Design

Randomization
- Participants assigned to Experimental & Control groups randomly



Blinding
- Single-Blinded: participants are unaware of intervention. Reason for Placebo.
- Double-Blinded: participants & researchers are unaware of intervention.

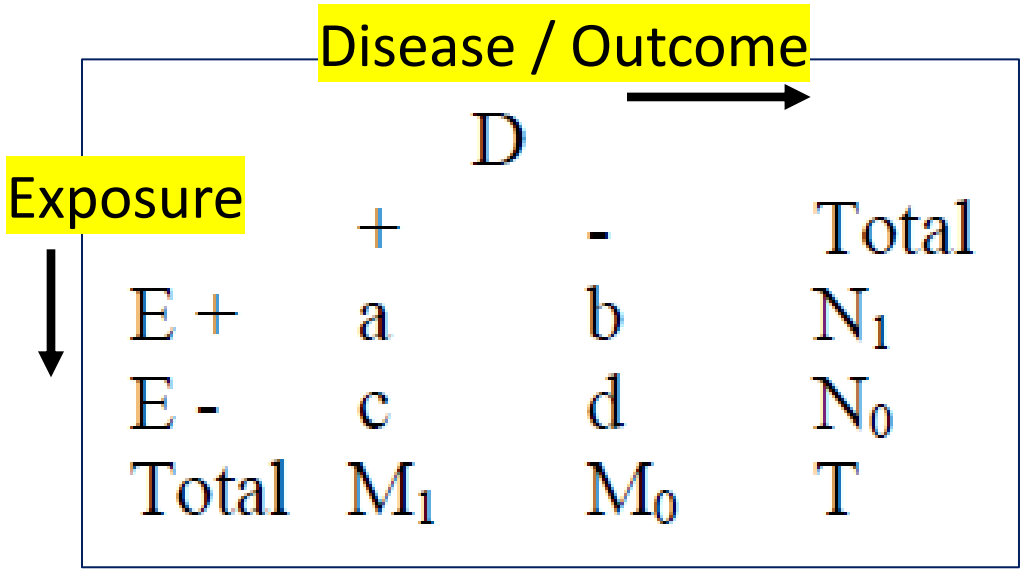
Experimental Studies: Summary

- Investigator specifies exposure and examines outcomes.
- “Control” vs. “Experimental” groups
- Usually includes some level of randomization & blinding
- Clinical drug trials are the most common example
- **Double-Blinded, Randomized Controlled Trials (RCTs)**
  - The “gold standard” of research design. Provides most convincing evidence of relationship between exposure and effect

Experimental Studies: Disadvantages

- Very expensive.
- It may be unethical to assign persons to certain treatment or comparison groups. Not possible to use RCTs to test effects of exposures that are expected to be harmful to humans, for ethical reasons.
- Not appropriate to study human exposures and diseases in natural environmental setting … like ambient air pollution exposure.

Presenting Epidemiological Study Results



2 x 2 table to display results of experiment

Symbol Definition

E+ (exposed), **E-** (not exposed), **D+** (got disease), **D-** (did not get disease)

a: exposed that got disease, **b**: exposed that did not get disease
c: not exposed that got disease, **d**: not exposed that did not get disease

N_1: Total number exposed (= a+b), **N_0**: Total number not exposed (=c+d)
M_1: Total number disease (= a+c), **M_0**: Total number not disease (=b+d)

Example

- 180 Participants split into 80 Exposed and 100 Not-Exposed
- Of the 80 Exposed, 10 got Disease
- Of the 100 Not Exposed, 5 got Disease

	D+	**D-**	**Total**
E+	10 (= a)	70 (= b)	80 (= N_1)
E-	5 (= c)	95 (= d)	100 (= N_0)
Total	15 (= M_1)	165 (= M_0)	180

Absolute Risk

<u>Definition</u>: Fraction of Total Number in a Group that Got the Disease (Outcome)

2x2 table

	D		
	+	-	Total
E +	a	b	N_1
E -	c	d	N_0
Total	M_1	M_0	T

E+ (exposed), E- (not exposed),
D+ (got disease), D- (did not get disease)
N_1: Total number exposed, N_0: Total number not exposed
M_1: Total number disease, M_0: Total number not disease

Exposed Subjects (E+) $R_1 = a/N_1$ R_1: Absolute Risk (Exposed)

Non-exposed Subjects (E-) $R_0 = c/N_0$ R_0: Absolute Risk (Not Exposed)

Relative Risk or "Risk-Ratio" (RR)

<u>Definition</u>: Ratio of Absolute Risks of Exposed to Non-Exposed Groups.

- Measures how much more/less likely Exposed Group got disease compared to Non-Exposed Group.
- RR > 1, exposed group more likely to have gotten disease
- RR < 1, exposed group less likely to have gotten disease

Estimated Risk of Disease

Exposed Subjects (E+)	$R_1 = a/N_1$
Non-exposed Subjects (E-)	$R_0 = c/N_0$

Measure of Effect	Formula
Risk Ratio (Relative Risk)	$RR = R_1 / R_0$

Returning to Our Example

- 180 Participants split into 80 Exposed and 100 Not-Exposed
- Of the 80 Exposed, 10 got Disease
- Of the 100 Not Exposed, 5 got Disease

	D+	**D-**	**Total**
E+	10 (= a)	70 (= b)	80 (= N_1)
E-	5 (= c)	95 (= d)	100 (= N_0)
Total	15 (= M_1)	165 (= M_0)	180

Absolute Risk:
R_1 = 10 diseases per 80 exposed = 10/80 = 0.125
R_0 = 5 outcomes per 100 not-exposed = 5/100 = 0.05

Relative Risk (Risk Ratio): = R_1/R_0 = (10/80) / (5/100) = 0.125/0.05 = 2.5

Interpretation: Exposed group was 2.5 times more likely to get disease than unexposed group. Exposed group was 2.5 times more “at risk” of getting disease.

Epidemiological Study Design

Epidemiology

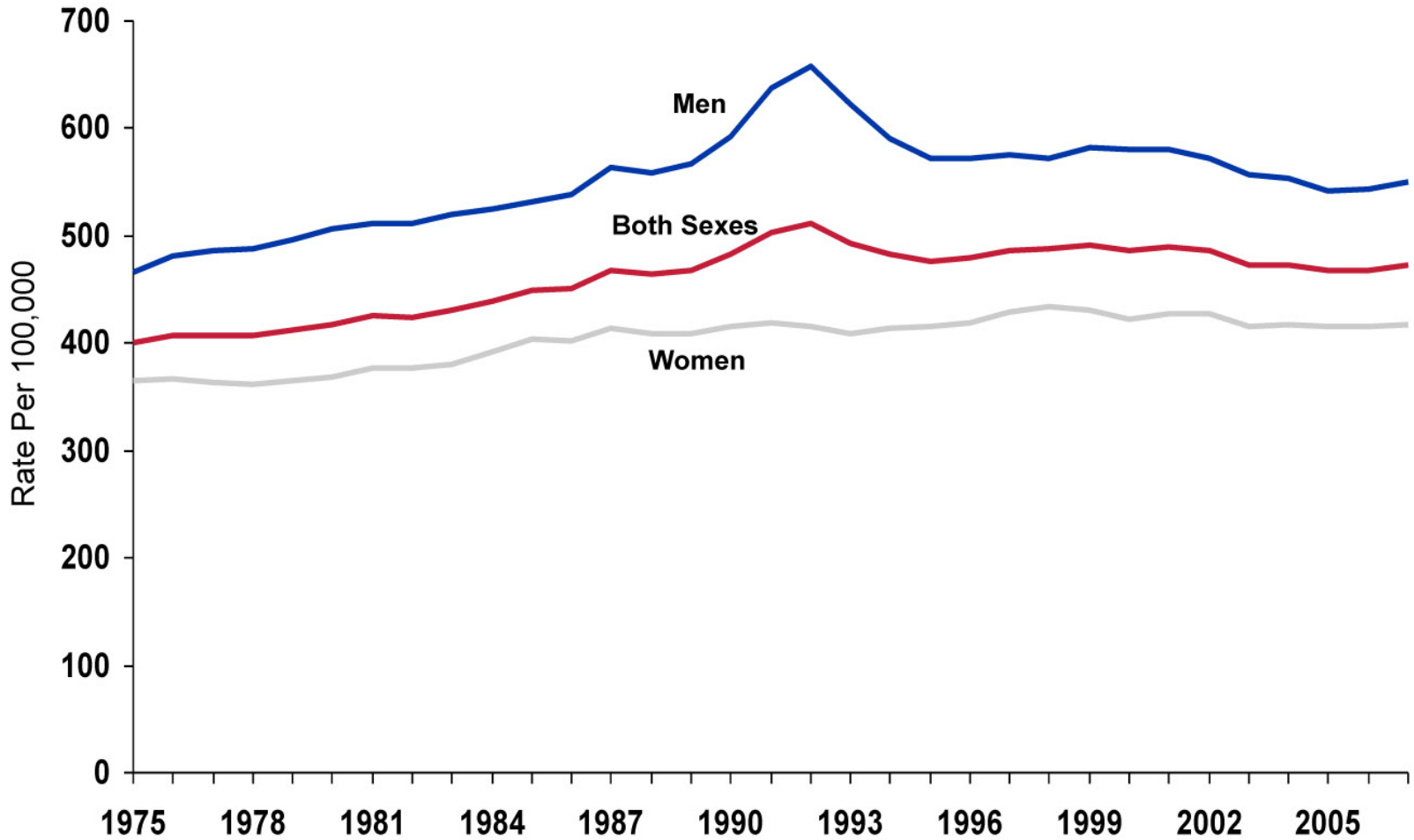
***Definition**: the study of the distribution and determinants of disease frequency as it exists across human populations.*

- Descriptive Epidemiology: Measuring disease distribution across population.
- Analytical Epidemiology: Measuring the association between a disease and its risk factors (exposure).

Prevalence vs. Incidence

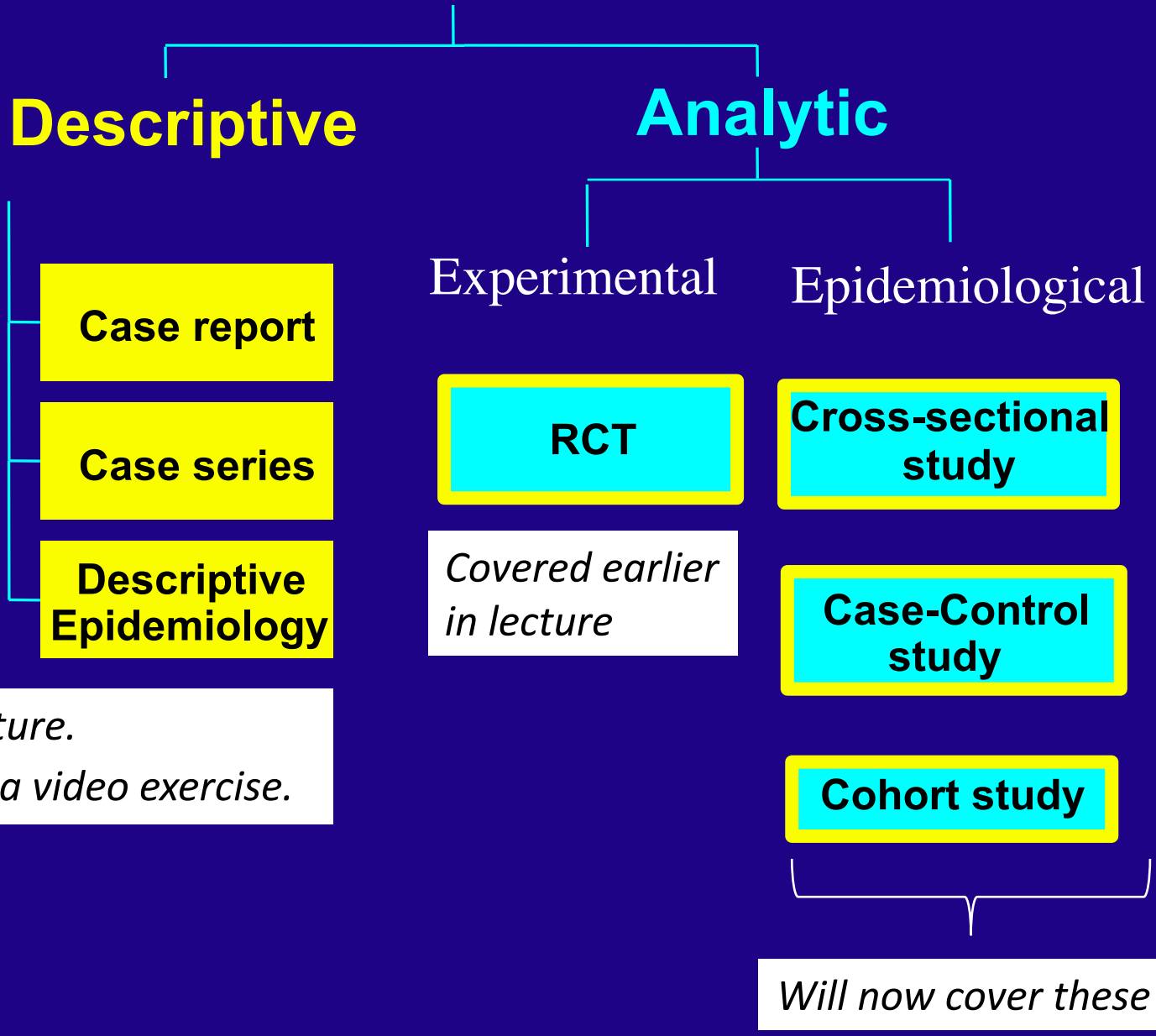
- **Prevalence:** measures the existence of a disease in a population at a particular time.
- **Incidence:** tracks the occurrence of new cases of a disease in a population over a period of time.
- Both are commonly expressed as # of disease cases per population (e.g. 100 cancers of all types per 100,000 people)

Cancer Incidence Rates* by Sex, US, 1975-2007



*Age-adjusted to the 2000 US standard population and adjusted for delays in reporting.
Source: Surveillance, Epidemiology, and End Results Program, Delay-adjusted Incidence database: SEER Incidence Delay-adjusted Rates, 9 Registries, 1975-2007, National Cancer Institute, 2010.

Study Designs



Covered earlier in lecture

Not covered in this Lecture.
Some in CDC Salmonella video exercise.

Will now cover these

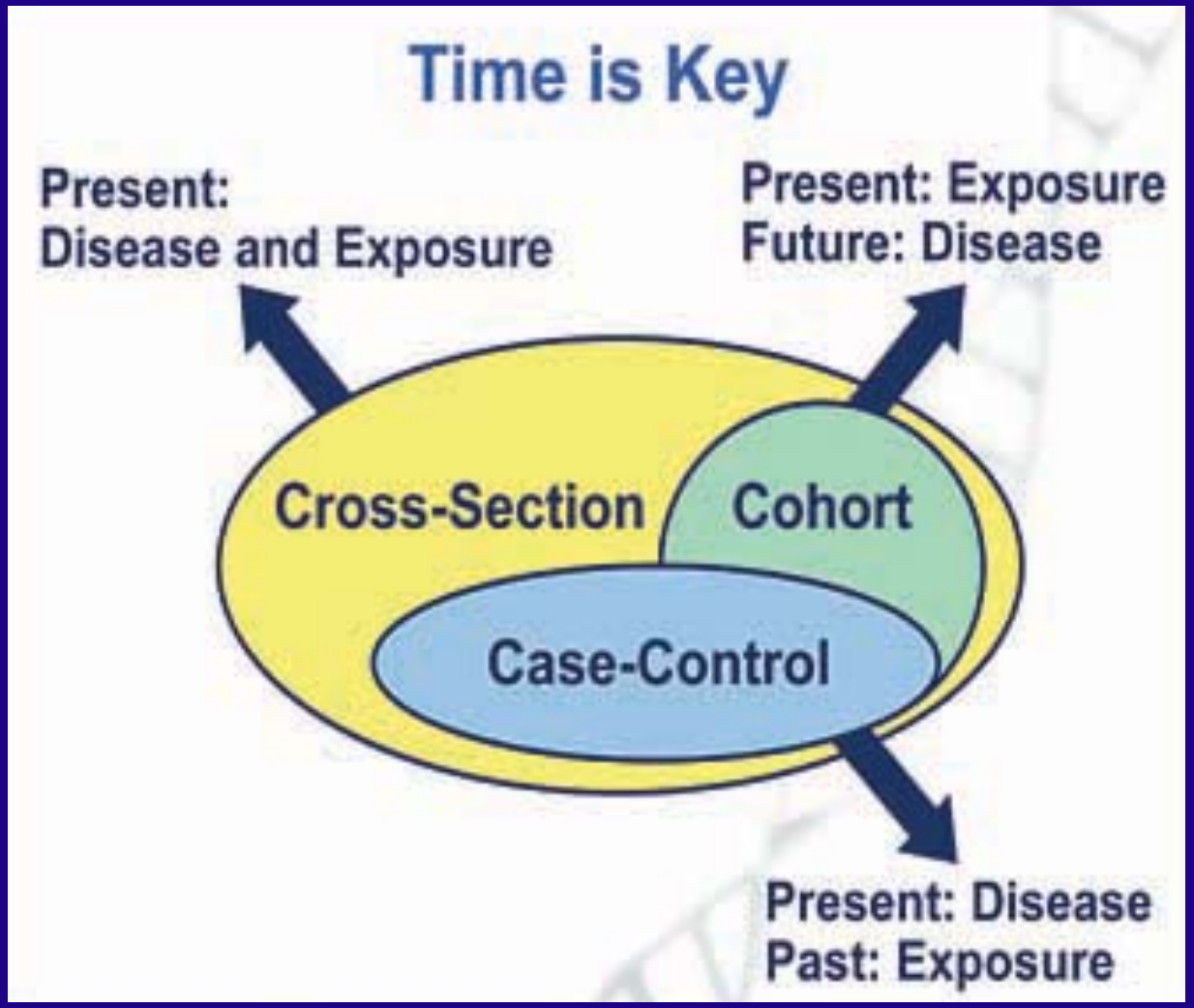
Analytical Epidemiological: Study Design

- **Cross-Sectional:** Measures prevalence of exposure(s) and outcome(s) over a population at a given time.

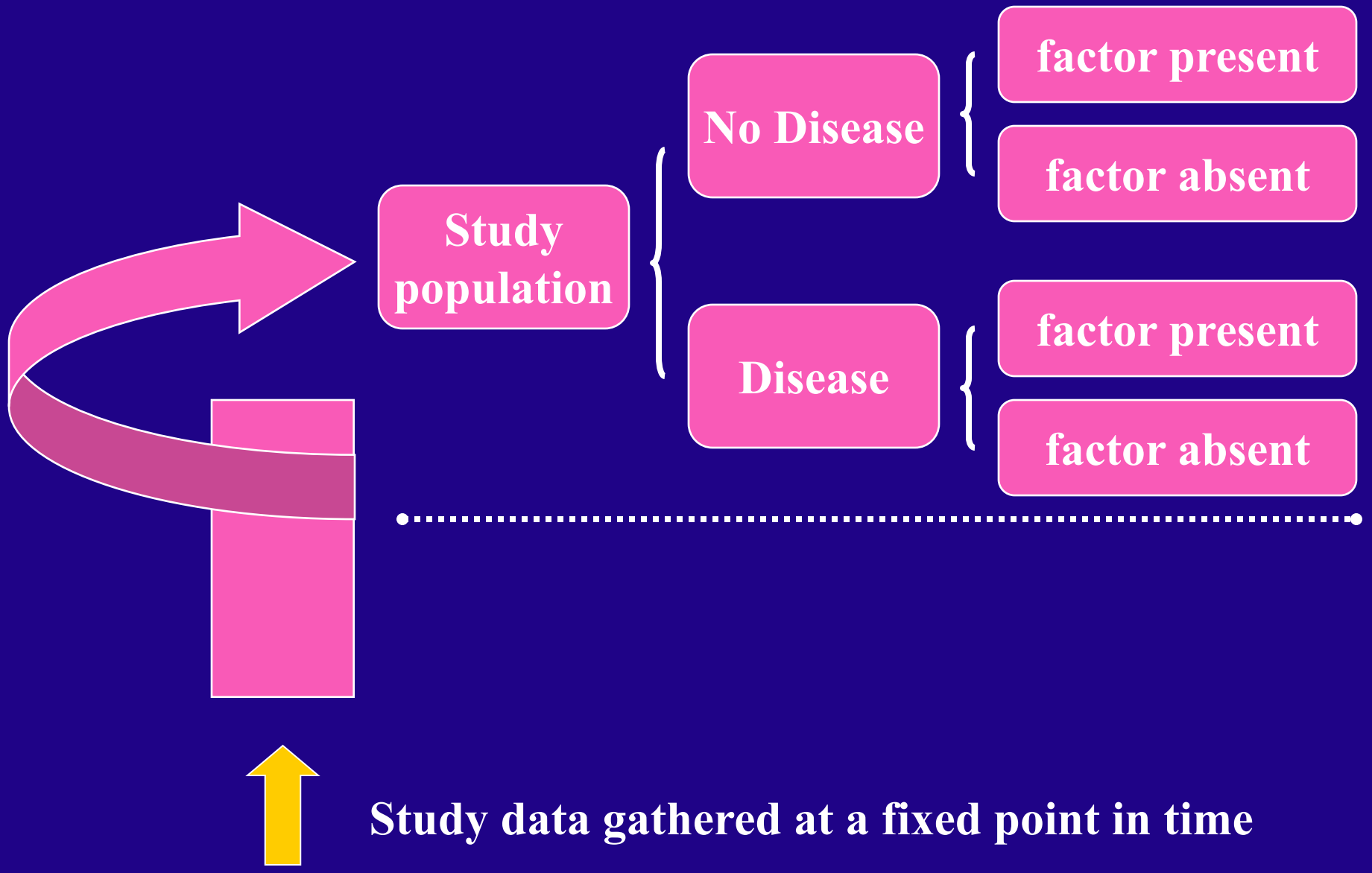
- **Case-Control:** Tracks previous history of exposure in disease ("case") versus non-disease ("control") cases. The analysis tracks exposure backwards in time. This is therefore a retrospective study.

- **Cohort** – Tracks incidence of disease (or other outcome) of enrolled subjects over time that have different measures of exposure to a risk factor for developing the outcome. Exposure status and outcomes of study participants are observed forward in time. This is therefore a prospective study.

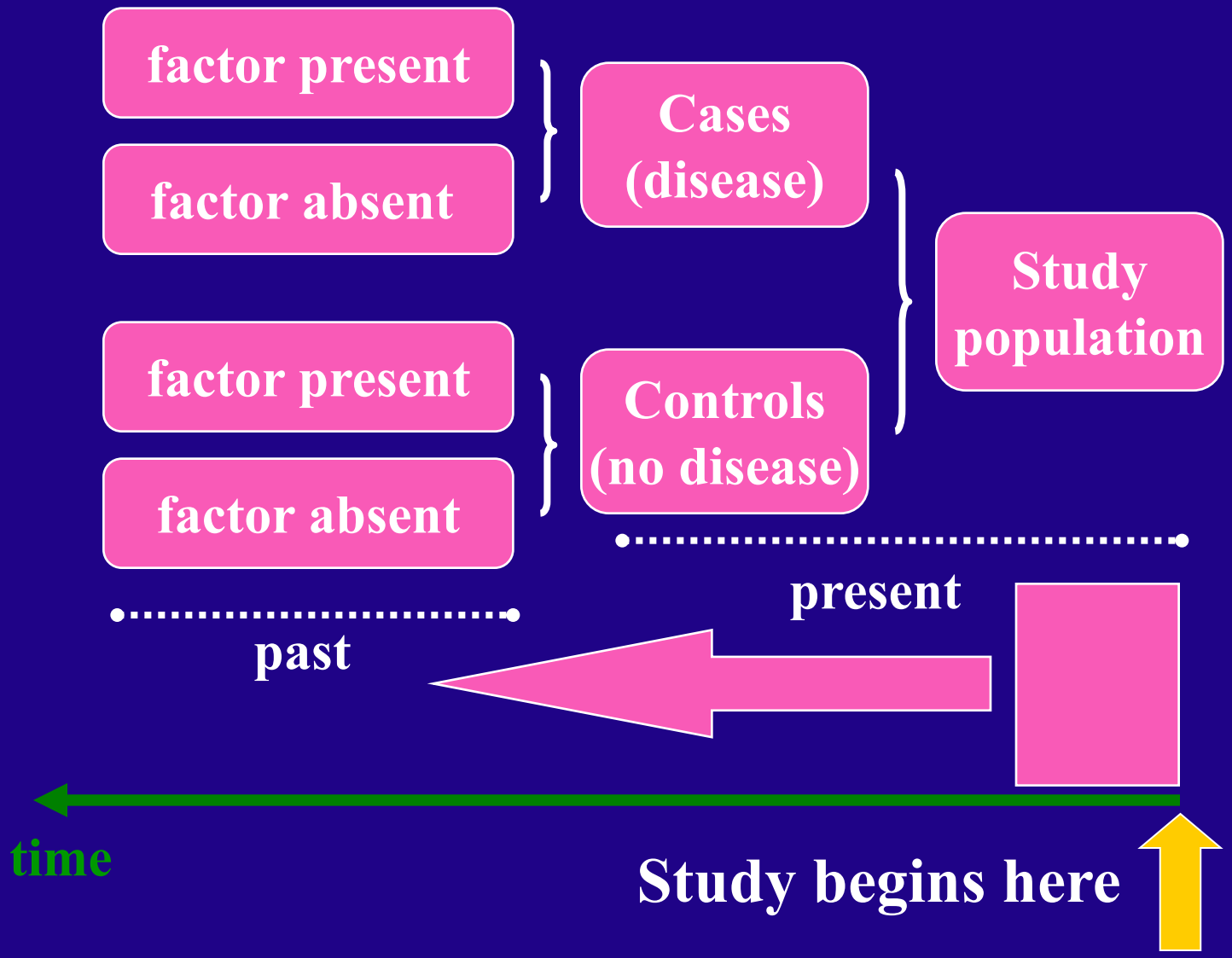
Epidemiological Studies: Analytical



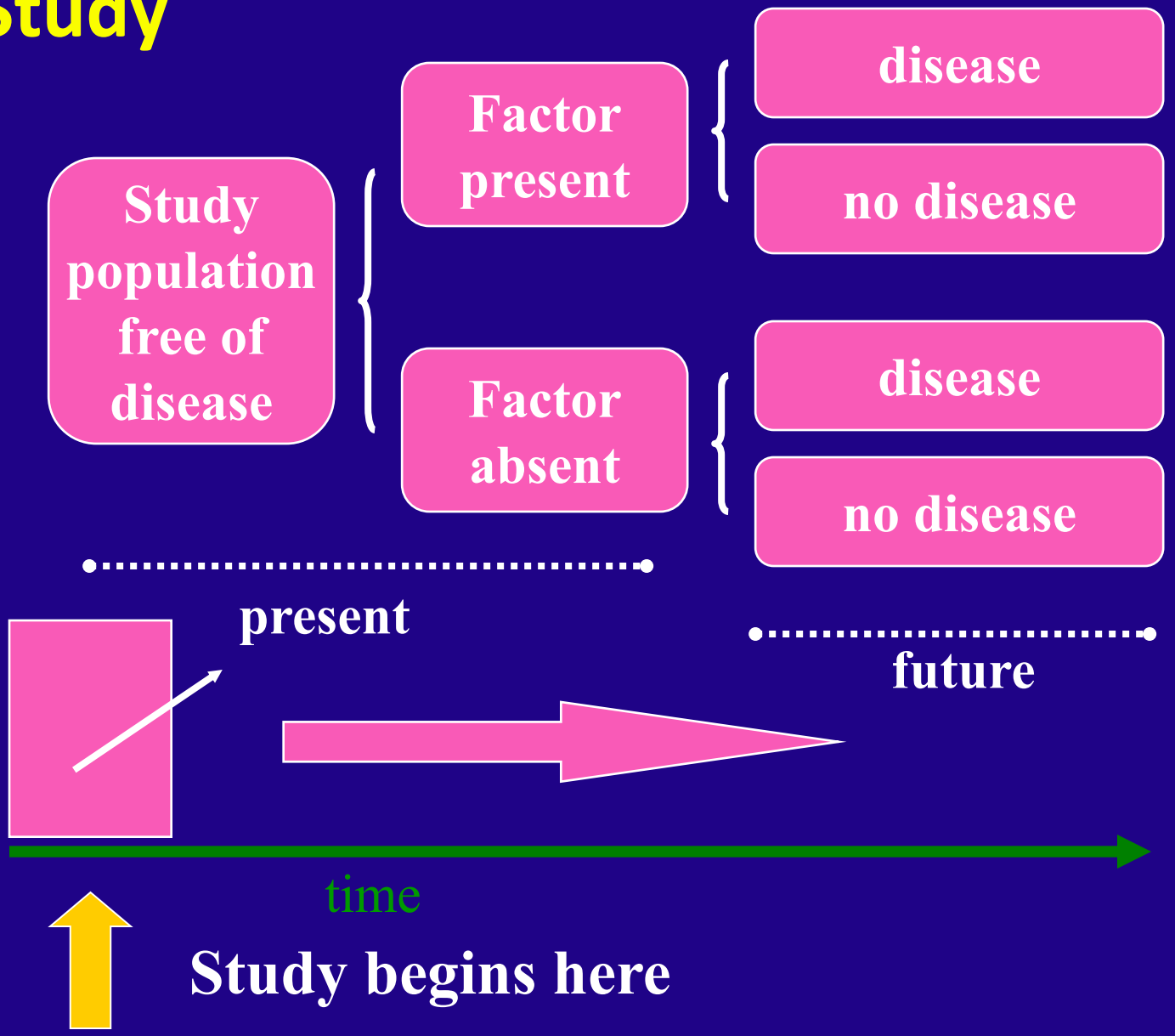
Cross-Sectional Study



Case-Control Study



Cohort Study



Causal Inference

Interpreting Study Results
("Causal Inference")

How to interpret the causes for study results pertaining to different outcomes for exposed vs. unexposed groups?

Example:

- Research study is designed to investigate link between smoking and higher incidence of coronary heart disease (CHD).
- Researchers conduct a cohort study. Results are a relative risk (RR) of CHD of smokers to non-smokers equal to 1.25, meaning smokers experienced a 25% higher incidence of CHD than non-smokers in the study.
- Can researchers therefore conclude smoking is associated a higher rate of CHD incidence based on these study results?
- What other possible explanations must be considered?

Example (continued …)

Study Results: RR = 1.25 for smokers getting CHD

Interpretation?

- **Possibility 1**: Smoking causes coronary heart disease (CHD).
- **Possibility 2**: There are other reasons for getting different CHD outcomes for smokers vs. non-smokers independent of the exposure differences (smoking vs. non-smoking). Among the main alternative possible explanations are …
  - Chance
  - Bias / Confounding

Chance

- When results of study are within expected random variation among samples, independent of exposure differences.
- Differences in results between exposed and unexposed groups must generally be beyond the "95% confidence interval (CI)" to confidently rule out random chance as an explanation.
- Differences beyond 95% CI: Rule out chance, differences are "statistically significant" (p-value < 0.05)
- Differences within 95% CI: Can be due to chance, differences are not statistically significant (p-value > 0.05)
- Larger the sample sizes ("large n") lead to greater likelihood of statistically significant results.

Apply to our example …

Study Results: RR = 1.25 for smokers getting CHD

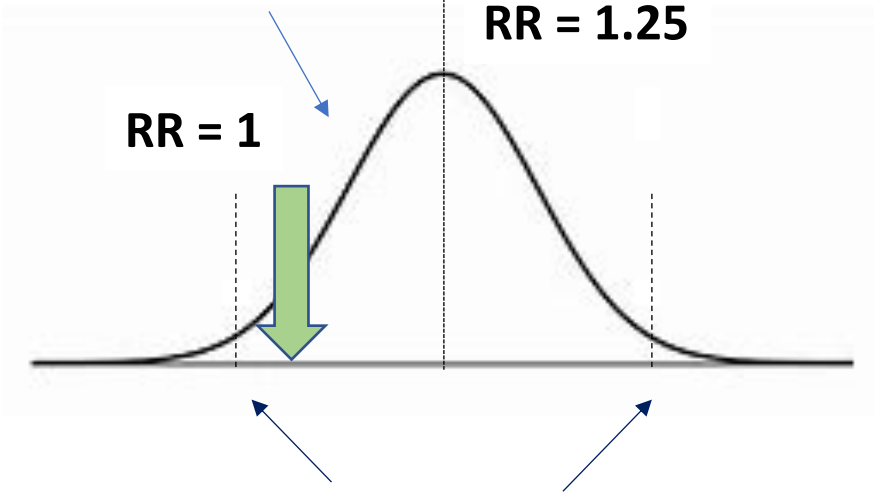
Interpretation?

- **Possibility 1**: Smoking causes coronary heart disease (CHD).
- **Possibility 2**: There are other reasons for getting different CHD outcomes for smokers vs. non-smokers independent of the exposure differences (smoking vs. non-smoking). Among the main alternative possible explanations are …
  - Chance?
  - Bias / Confounding

Chance: Applied to Our Example

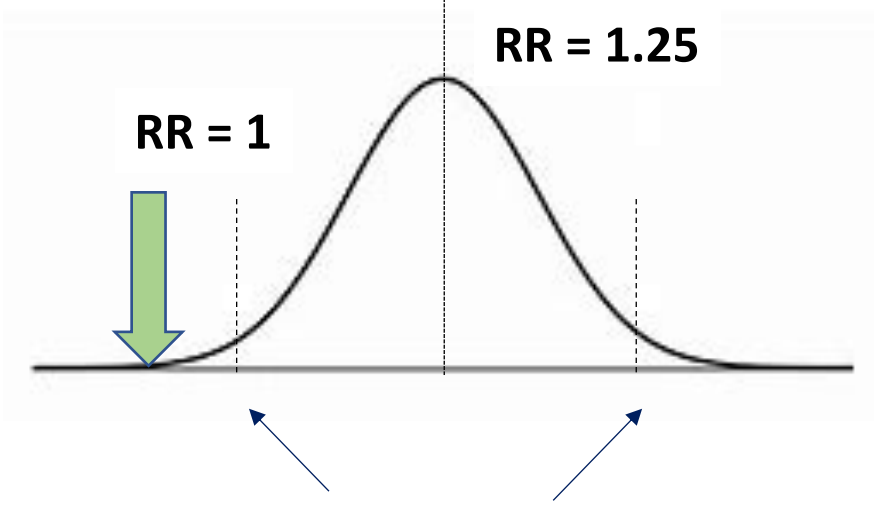
Expected variation of Risk Ratio due to random chance ("bell curve")

RR = 1 within 95%-CI



RR = 1 within 95% CI ($p > 0.05$)
*(RR = 1.25 **is not** statistically significant)*
*(**Cannot** rule out chance with sufficient confidence)*

RR = 1 outside 95%-CI



RR = 1 outside 95% CI ($p < 0.05$)
*(RR = 1.25 **is** statistically significant)*
*(**Can** rule out chance with sufficient confidence)*

Example (continued …)

Study Results: RR = 1.25 for smokers getting CHD

Interpretation?

- **Possibility 1**: Smoking causes coronary heart disease (CHD).
- **Possibility 2**: There are other reasons for getting different CHD outcomes for smokers vs. non-smokers independent of the exposure differences (smoking vs. non-smoking). Among the main alternative possible explanations are …
  - Chance
  - Bias / Confounding?

Bias

- **Selection Bias** – errors in selecting participants that would systematically affect exposed and non-exposed groups differently.
- **Measurement Bias** – errors in measuring or classifying exposures or outcomes among participants that would systematically affect exposed and non-exposed groups differently.
- If these are an important driver explaining the cause of findings, one would say the study results are “biased”.

Confounding

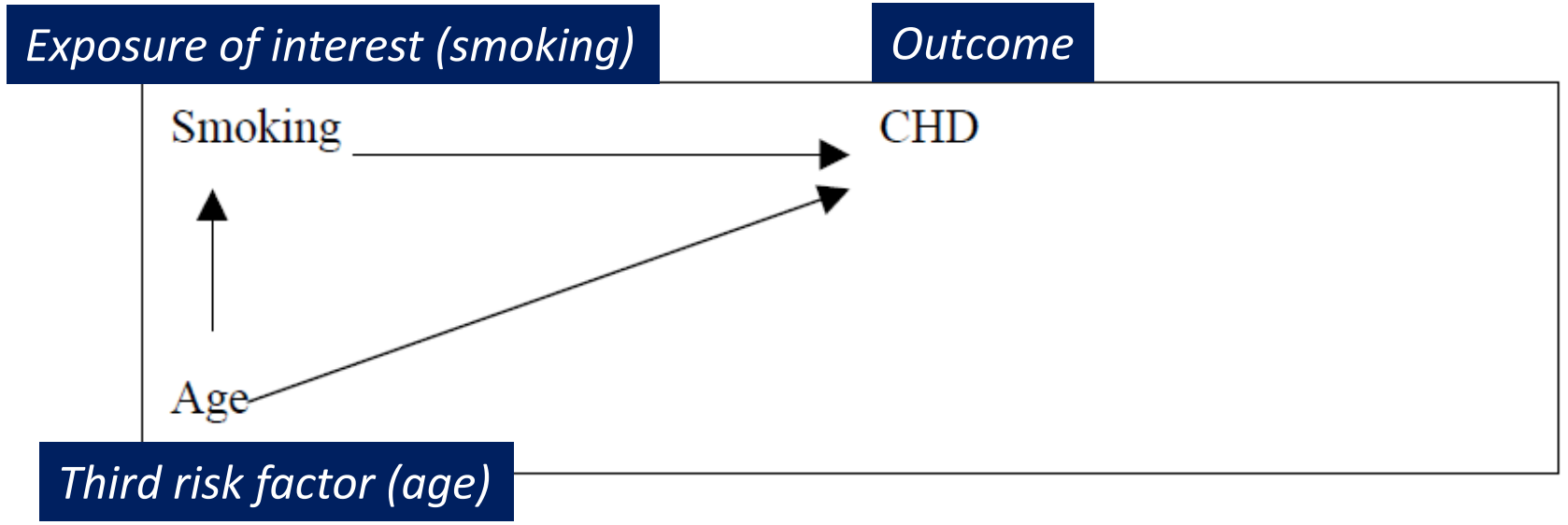
- Outcome differences are due to a “third risk factor”, aside from the exposure or intervention studied. This third risk factor is the main driver of the outcome.
- Two requirements for something to be a “confounder”

1. Presence of individuals exposed to third risk factor is different in exposed versus non-exposed groups (for example, a selection or measurement bias was made).
2. The third risk factor is associated with the outcome, independent of the one studied.

Confounding

Requirements for "third" risk factor to be a confounder

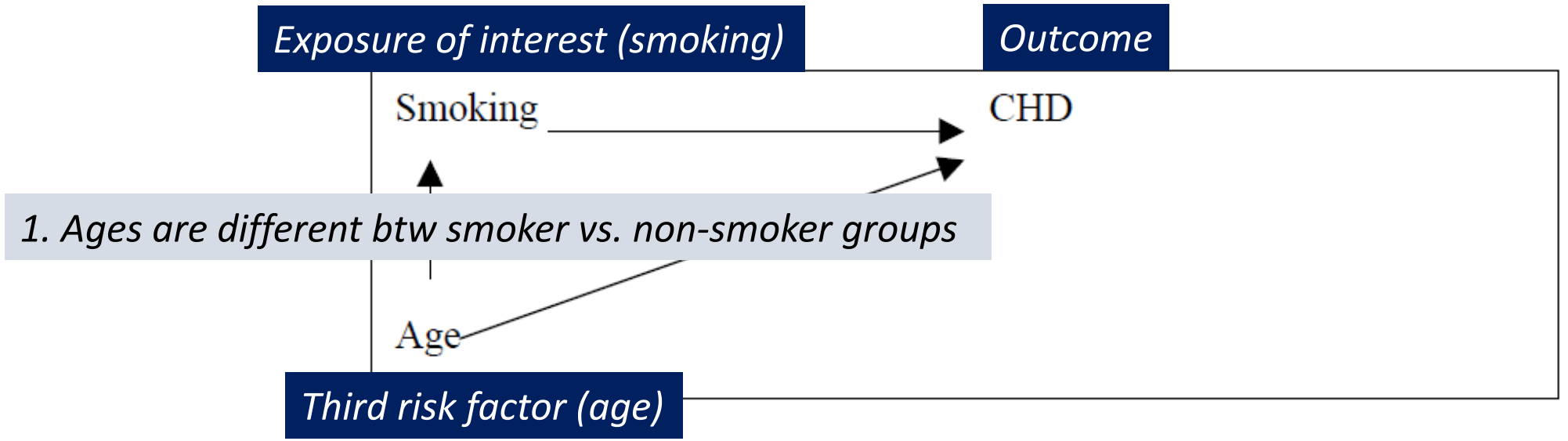
Example: Study on the link between smoking and CHD



Confounding

Requirements for "third" risk factor to be a confounder

Example: Study on the link between smoking and CHD



Confounding

Requirements for "third" risk factor to be a confounder

Example: Study on the link between smoking and CHD