Biostatistics
First Year: Summer Course
Study Design

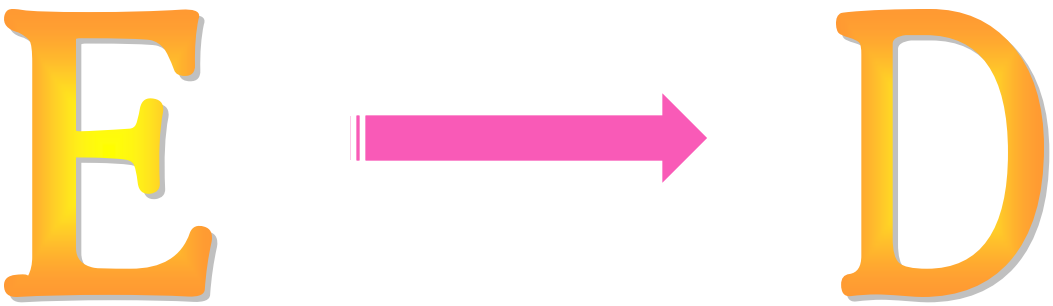
Associated Professor Dr. Nedal Alnawaiseh
M. B. Ch. B (MD), Baghdad, Iraq.
MSc, JUST, Jordan.
MSPH, Tulane University, USA.
PhD, UKM, Malaysia.
PhD, UNU, IIGH.
Public Health & Community Medicine Department, Medical School, Mutah University, Jordan
Mobile:+962795****. nawayseh@gmail.com**

Types of Primary Studies

- **Descriptive Studies**
  - **describe occurrence of an outcome**
    without analysis or association. It can't correlate outcome to the exposure
- **Analytic Studies**
  - **describe the potential *association* between exposure and outcome,**
    usually for chronic diseases which have latency period
    - **Incidence is the proportion of new cases.**
    - **Prevalence is the proportion of old and new cases.**

Basic Question in Analytic Epidemiology

- Are exposure and disease linked?

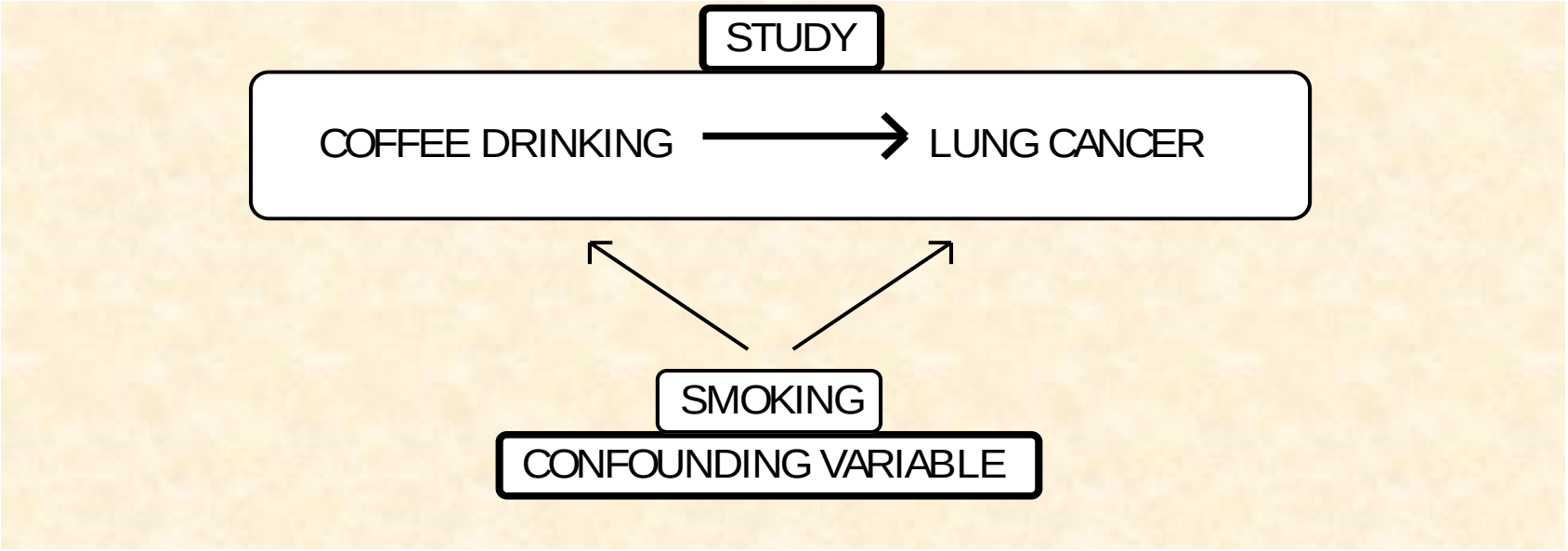


Exposure **Disease**

- STATISTICS

- **CHANCE - p = 1 in 20 (0.05).**
- **> 1 in 20 (0.051) = not significant**
- **< 1 in 20 (0.049) = statistically significant**
- **CONFIDENCE INTERVALS**
- **what is the range of values between which we could be 95% certain that this result would lie if this intervention was applied to the general population**

CHANCE, BIAS, CONFOUNDING VARIABLES

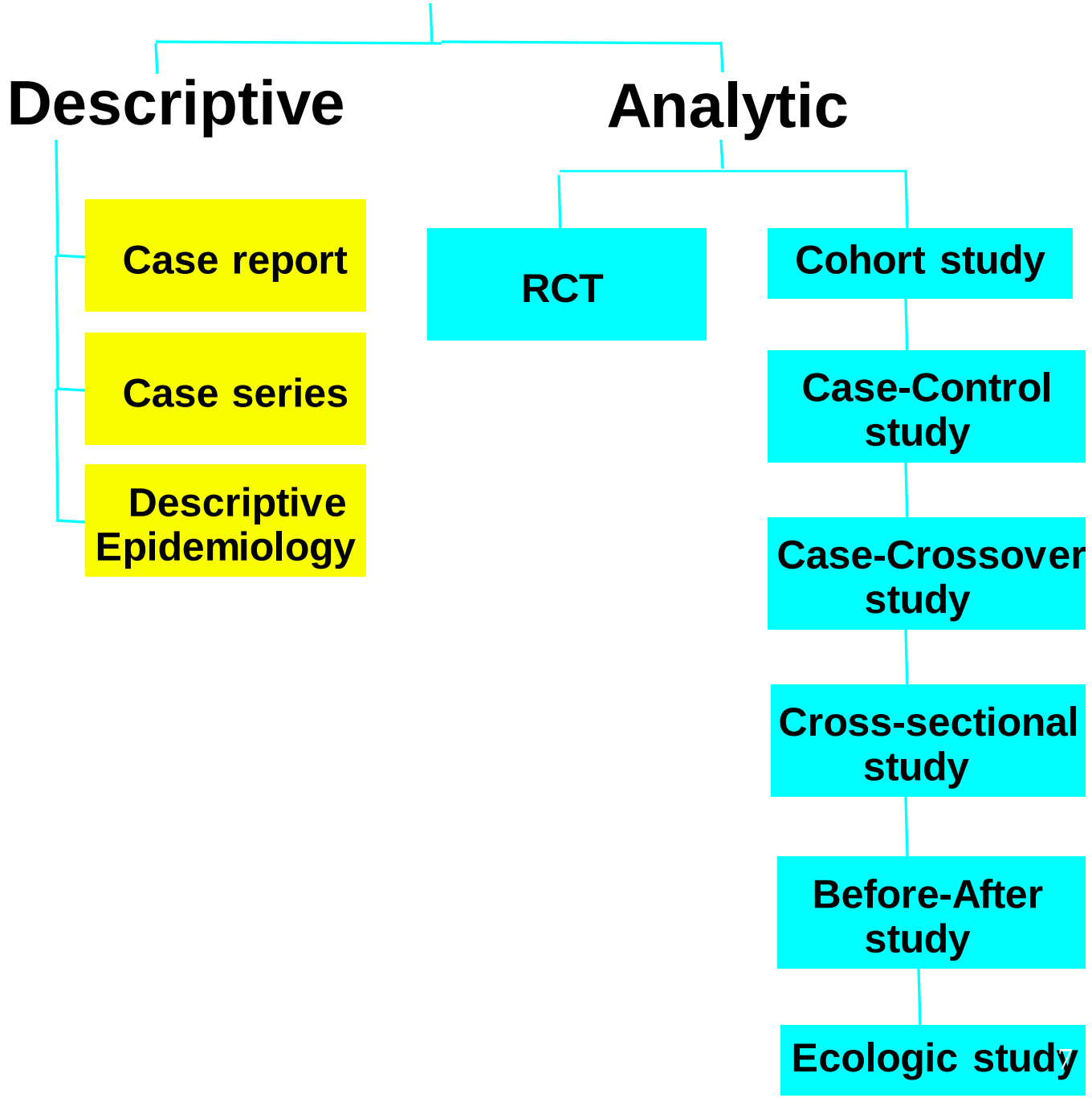


(Lung cancer Is more common between people, who drink a lot of coffee. However, coffee is not the truth cause, instead there is a confounding factor which is smoking among coffee-drinking people)

Basic Questions in Analytic Epidemiology

- Look to link exposure and disease
  - **What is the exposure?**
  - **Who are the exposed?**
  - **What are the potential health effects** **(the environmental factors upon the sample, such as chemical effect on people who deal with insecticide)?**
  - **What approach will you take to study the relationship between exposure and effect?**

Study Designs

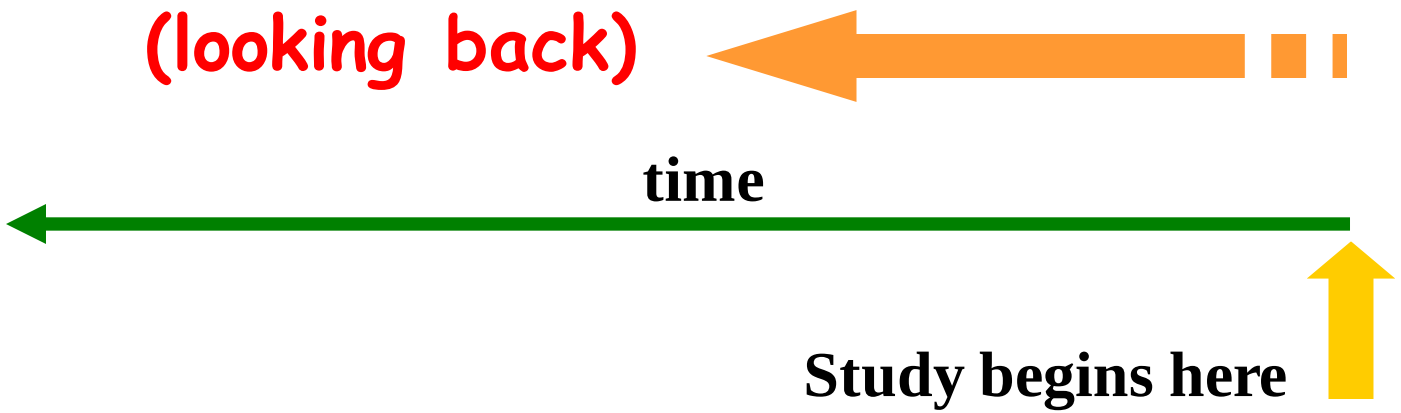


Timeframe of Studies

- **Prospective Study - looks forward, looks to the future, examines future events, follows a condition, concern or disease into the future**

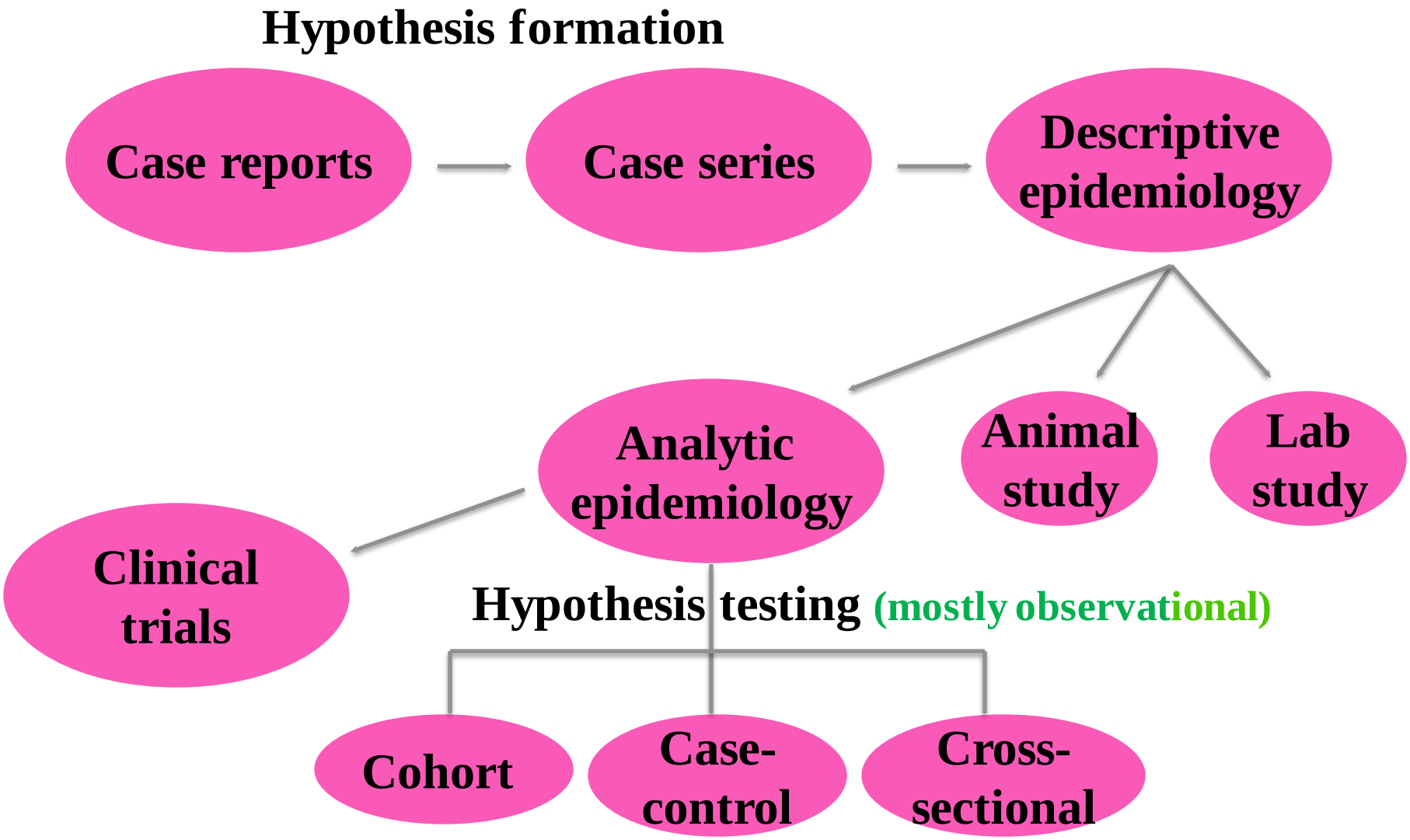
Timeframe of Studies

- **Retrospective Study - "to look back", looks back in time to study events that have already occurred**



e.g. Case-Control Study

Study Design Sequence



Increasing Knowledge of Disease/Exposure

Descriptive Studies — **Develop hypothesis**

Case-control Studies — **Investigate it's relationship to outcomes**, studied by a special ratio called odds ratio

Cohort Studies — **Define it's meaning with exposures,** and can determine incidence

Clinical trials — **Test link experimentally**

Descriptive Studies

Case Reports

- **Detailed presentation of a single case or handful of cases**
- **Generally report a new or unique finding**
    - **e.g. previous undescribed disease**
    - **e.g. unexpected link between diseases**
    - **e.g. unexpected new therapeutic effect**
    - **e.g. adverse events**

Case Series

- **Experience of a group of patients with a similar diagnosis**
- **Assesses prevalent disease**
- **Cases may be identified from a single or multiple sources**
- **Generally report on new/unique condition**
- **May be only realistic design for rare disorders**

Case Series

Case series
From Wikipedia, the free encyclopedia
http://en.wikipedia.org/wiki/Case_series

- **Advantages**
  - **Useful for hypothesis generation. However, for testing the hypothesis, we have to apply analytical studies.**
  - **Informative for very rare disease with few established risk factors**
  - **Characterizes averages for disorder**

- **Disadvantages**
  - **Cannot study cause and effect relationships**
  - **Cannot assess disease frequency**

Descriptive Studies

Case Report → **One case of unusual findings**

Case Series → **Multiple cases of findings**

Descriptive Epidemiology Study → **Population-based cases with denominator**

prevalence=number of cases/ population

Analytical Studies

Study Designs - Analytic Epidemiology

- **Experimental Studies**
  - **Randomized Controlled Clinical Trials (RCT)**

*considered as the gold standard of studies.

*under full control and intervention of researcher. *assigned blindly.

  - **Community trials** (applied to the whole community)
- **Observational Studies**
  - **Group data** (i.e. we don't have subject level info)
    - **Ecologic** (correlation study of non medical subjects, such as the rate of using umbrella and rain-fall rates. It is statistically significant but not always logical)
  - **Individual data**
    - **Cross-sectional**
    - **Cohort**
    - **Case-control**
    - **Case-crossover**

An Introduction to Epidemiology (CDC)
http://www.cdc.gov/excite/classroom/intro_epi.htm

Experimental Studies

- **Treatment and/or exposures occur in a "controlled" environment**
- **Planned research designs**
- **Clinical trials are the most well known experimental design. Clinical trials use randomly assigned data.**

Observational Studies

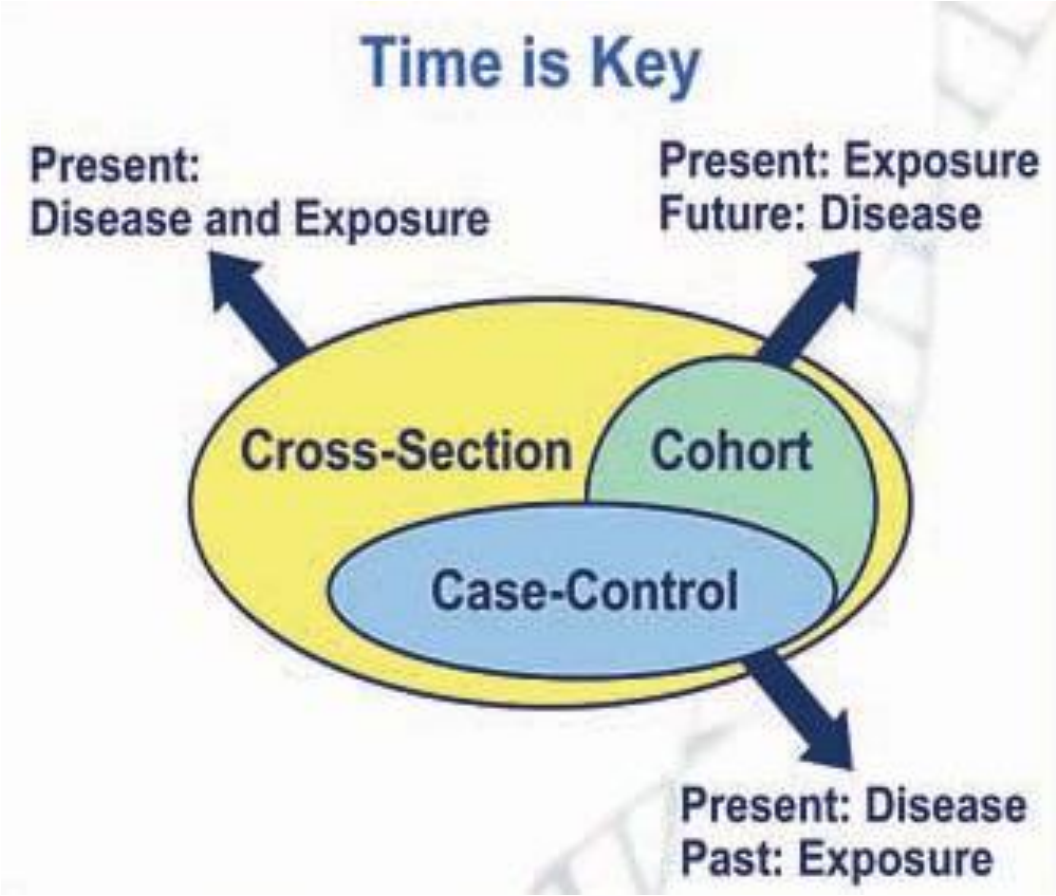
1. Non-experimental
2. Observational because there is no individual intervention
3. Treatment and/or exposures occur in a "non-controlled" environment
4. Individuals can be observed prospectively, retrospectively, or currently (i.e. cross-sectional)

Cross-sectional studies

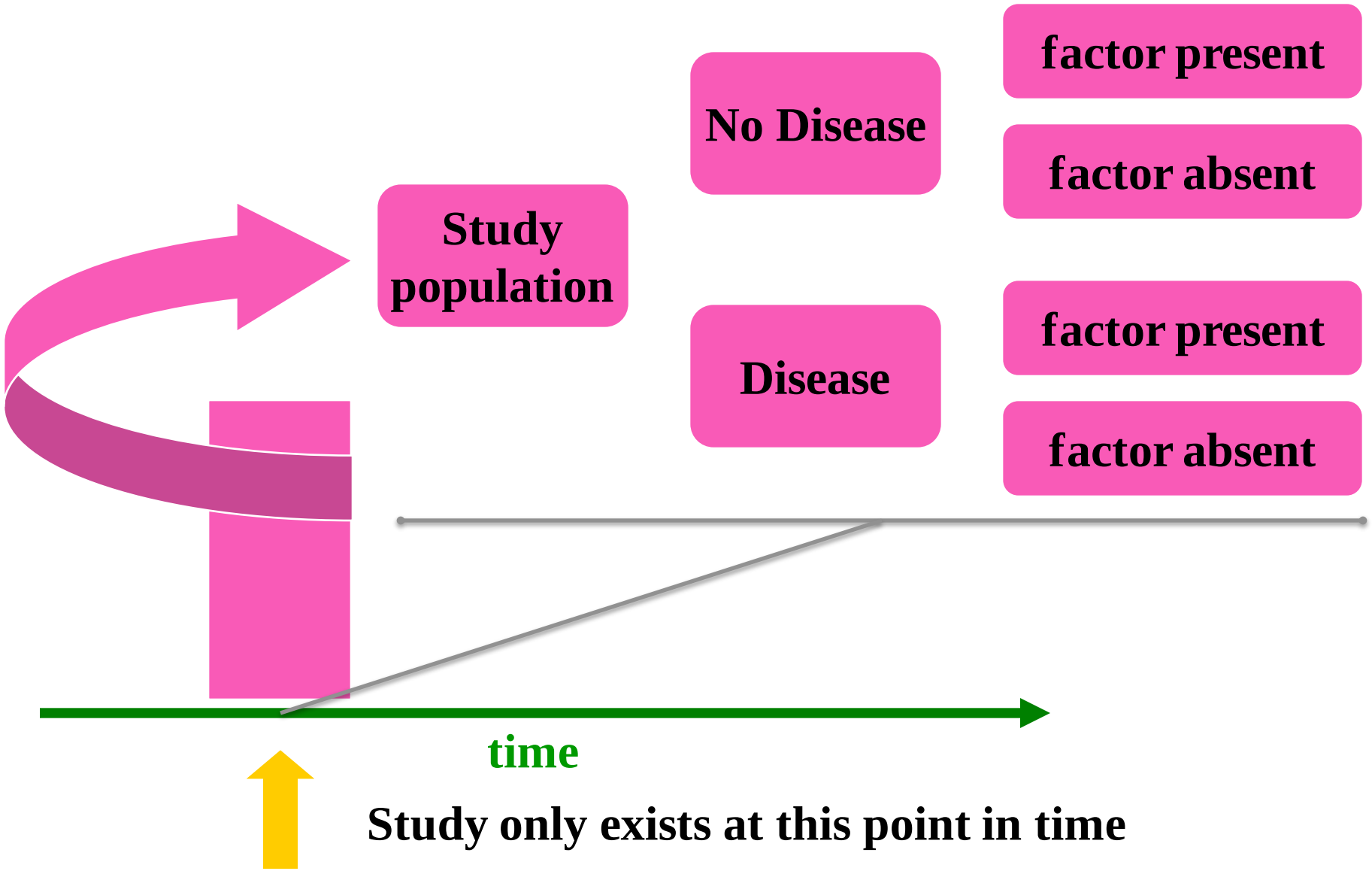
- **An "observational" design that surveys exposures and disease status at a single point in time (a cross-section of the population)**

*** The main disadvantage is loss of temporal sequence, for instance it is not possible to determine whether hypertension causes stroke, or hypertension is caused by angina.**

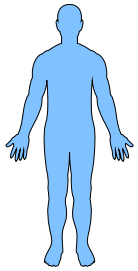
Observational Studies and Timeframe



Cross-sectional Design

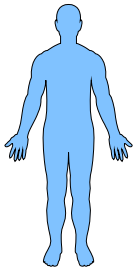


Cross-sectional Studies



- **Often used to study conditions that <u>are relatively frequent with long duration of expression</u> (nonfatal, chronic conditions)**
- **It measures *prevalence*, not *incidence* of disease**
- **Example: community surveys**
- **Not suitable for studying rare (the most suitable study is case-control) or highly fatal diseases or a disease with short duration of expression**

Cross-sectional studies



- **Disadvantages**
  - **Weakest observational design, although it is most common, (it measures prevalence, not incidence of disease). Prevalent cases are survivors**
  - **The Temporal Sequence of exposure and effect may be difficult or impossible to determine**
  - **Usually don't know when disease occurred**
  - **Rare events a problem. Quickly emerging diseases are also problem.**

Analysis of cross-sectional studies

✓**In a cross-sectional study, to calculate prevalence, multiple parameters are measured simultaneously – questions, observations, and answers.**

✓**Prevalence = No of cases at a given time / No of people at the same given time**

✓**Prevalence is a proportion, not integral number**

✓***For continuous variables, they fall along a scale within a given range. To calculate prevalence, the values have to be below or above predetermined level or else median levels may be calculated.***

Cross-Sectional Studies Advantages and Disadvantages

Advantages of cross-sectional studies

1. **Relatively quick to conduct**
2. **All variables are collected at one go**
3. **Multiple outcomes can be researched at once**
4. **Prevalence for all factors can be measured**
5. **Good for descriptive analysis**
6. **Can be used as a springboard for further research**

Disadvantages of cross-sectional studies

1. **Cannot be used to get timeline based research**
2. **Tough to find people that fall under the exact same variables**
3. **Associations are tough to interpret, association doesn't mean causation.**
4. **When strong feelings are involved, there could be a case of a bias**
5. **Does not help to determine cause**

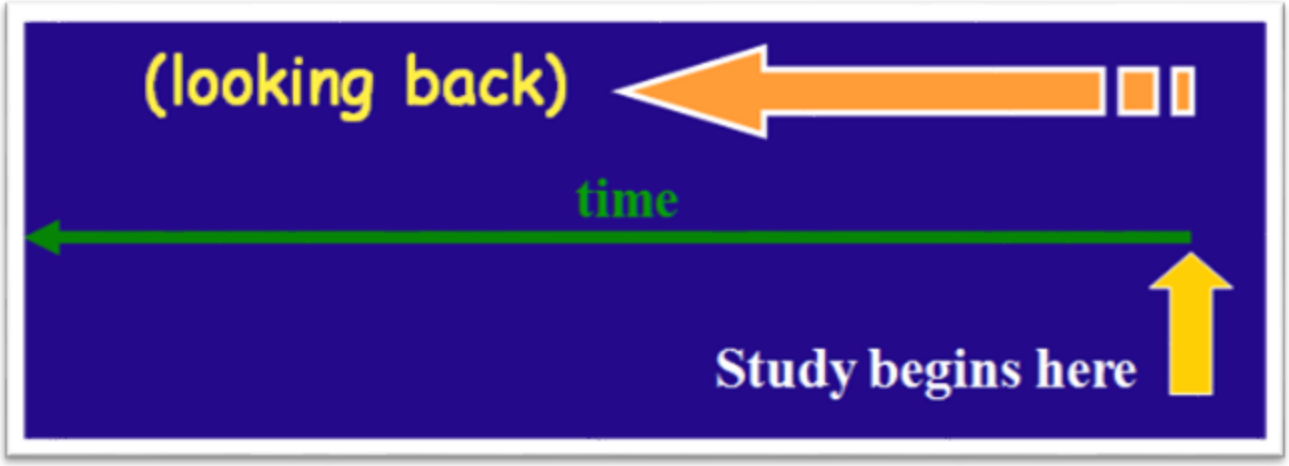
Epidemiologic Study Designs

- **Case-Control Studies**
  - **an "observational" design comparing exposures in disease cases vs. healthy controls from same population**
  - **exposure data collected retrospectively**
  - **most feasible design <u>where disease outcomes are rare</u>**

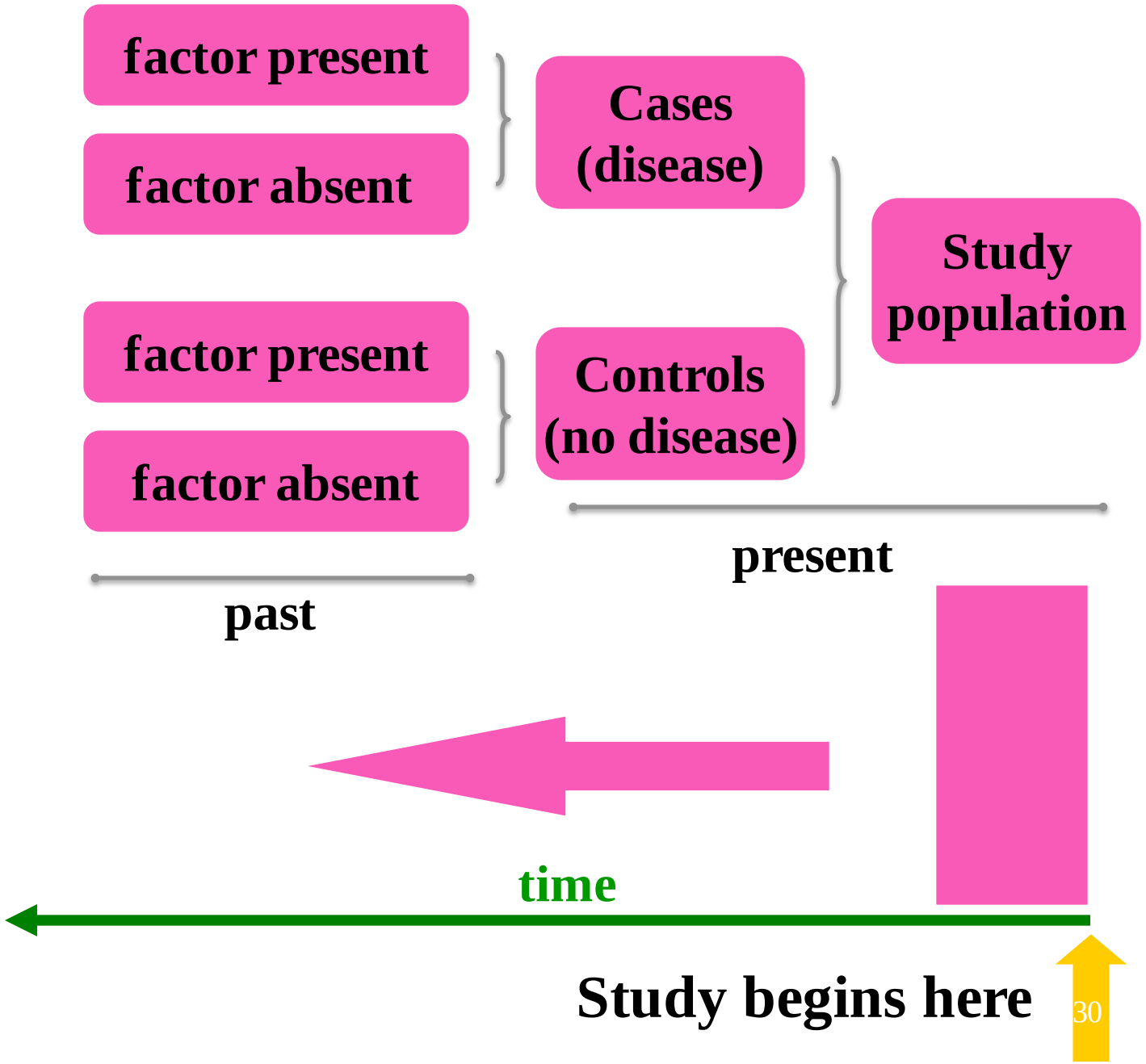
Case-Control Studies

Cases: Disease

Controls: No Disease



Case-Control Design



Case-Control Study

- **Strengths**
  - **Less expensive** compared with cohort **and time consuming** because the disease is already exists.
  - **Efficient for studying rare diseases**
- **Limitations**
  - **Exposure measurements taken after disease occurrence** (recall bias: patients are more likely to remember events associated with their condition than non-diseased people)
  - **Disease status can influence selection of subjects**

Calculating the Odds Ratio

It is better that cases equal to control, and control must not be less than cases.

Exposure Status	**Disease Status: CHD cases (Cases)**	**Disease Status: No CHD (Controls)**
Smoker	**112**	**176**
Non-smoker	**88**	**224**
Total	**200**	**400**

$$\text{Odds Ratio} = \frac{AD}{BC} = \frac{112 \times 224}{176 \times 88} = 1.62$$

	OR<1	OR=1	OR>1
Odds comparison between cases and controls	Odds of exposure for cases are less than the odds of exposure for controls	Odds of exposure are equal among cases and controls	Odds of exposure for cases are greater than the odds of exposure for controls
Exposure as a risk factor for the disease?	Exposure reduces disease risk (Protective factor)	Particular exposure is not a risk factor	Exposure increases disease risk (Risk factor)

Interpreting the Odds Ratio

Those with CHD are 1.62 times more likely to be smokers than those without CHD

Those with CHD are 62% more likely to be smokers than those without CHD

	Cases	Controls
Exposed (ate)	13 a	32 b
Not Exposed (did not eat)	17 c	23 d

Odds Ratio (OR) = (a/c) / (b/d)
= (13/17) / (32/23)
= 0.55

ORs, P-Values and 95% CIs for Case-Control Study with 3 Different Sample Sizes

Parameter Computed	Sample Size n=20	n=50	n=500
OR	2.0	2.0	2.0
p-value	0.500	0.200	0.001
95% Cis	0.5, 7.7	0.9, 4.7	1.5, 2.6

When confidence interval contain (1), it is certainly not significant

The interval doesn't include (1), so it is significant.

The **two types of case-control studies** are:

1. **Non-matched case-control study:** this is the simplest form. Find a person with the disease and enroll them in the study. Then enroll a control and determine their exposure status.
2. **Matched case-control:** Find a person with the disease and enroll them in the study. Match the person for some characteristic (e.g. sex, age **(the most confounding variable),** weight) with a control. **This can eliminate or minimize confounding variables** . However, it generally results in a longer study; the more characteristics being "matched", the longer the study takes. **(all factors should be the same except the factor under study)**

Advantages and Disadvantages

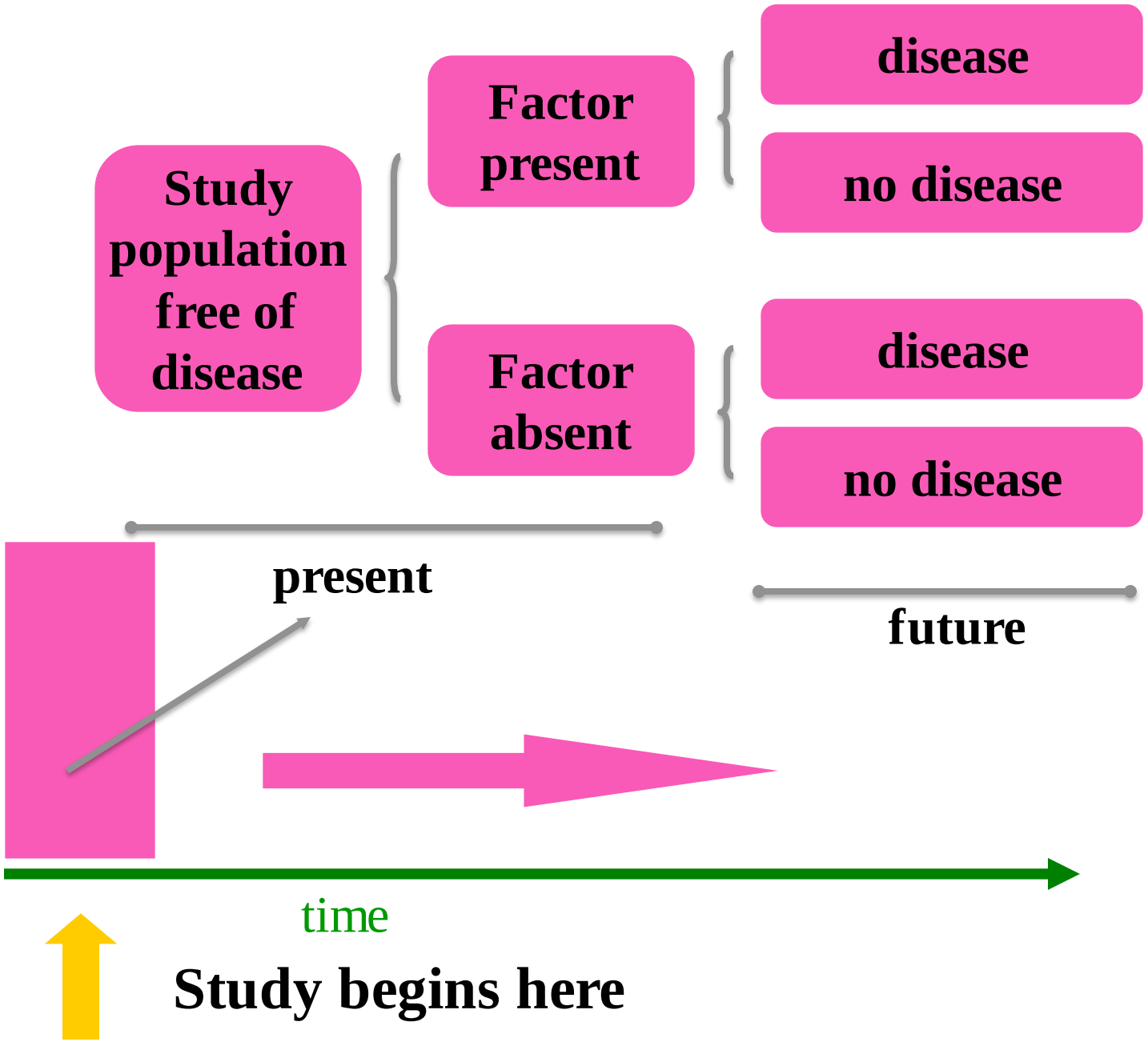
Other Advantages:

1. Short term study that doesn't require waiting for events to happen, as they have already occurred.
2. Inexpensive.
3. Multiple risk factors can be studied at the same time.
4. Quickly establishes associations between risk factors and disease. This can be especially useful with disease outbreaks, as causes can be identified with small sample sizes.
5. Stronger than cross-sectional studies for establishing causation.

Disadvantages:

1. Control groups can be difficult to find.
2. Results can easily be tainted by **Recall Bias**, where people with the disease or condition are more likely to remember past details compared to people who don't have the disease or condition.
3. Is weaker than a cohort study for establishing causation.
4. Usually not generalizable.

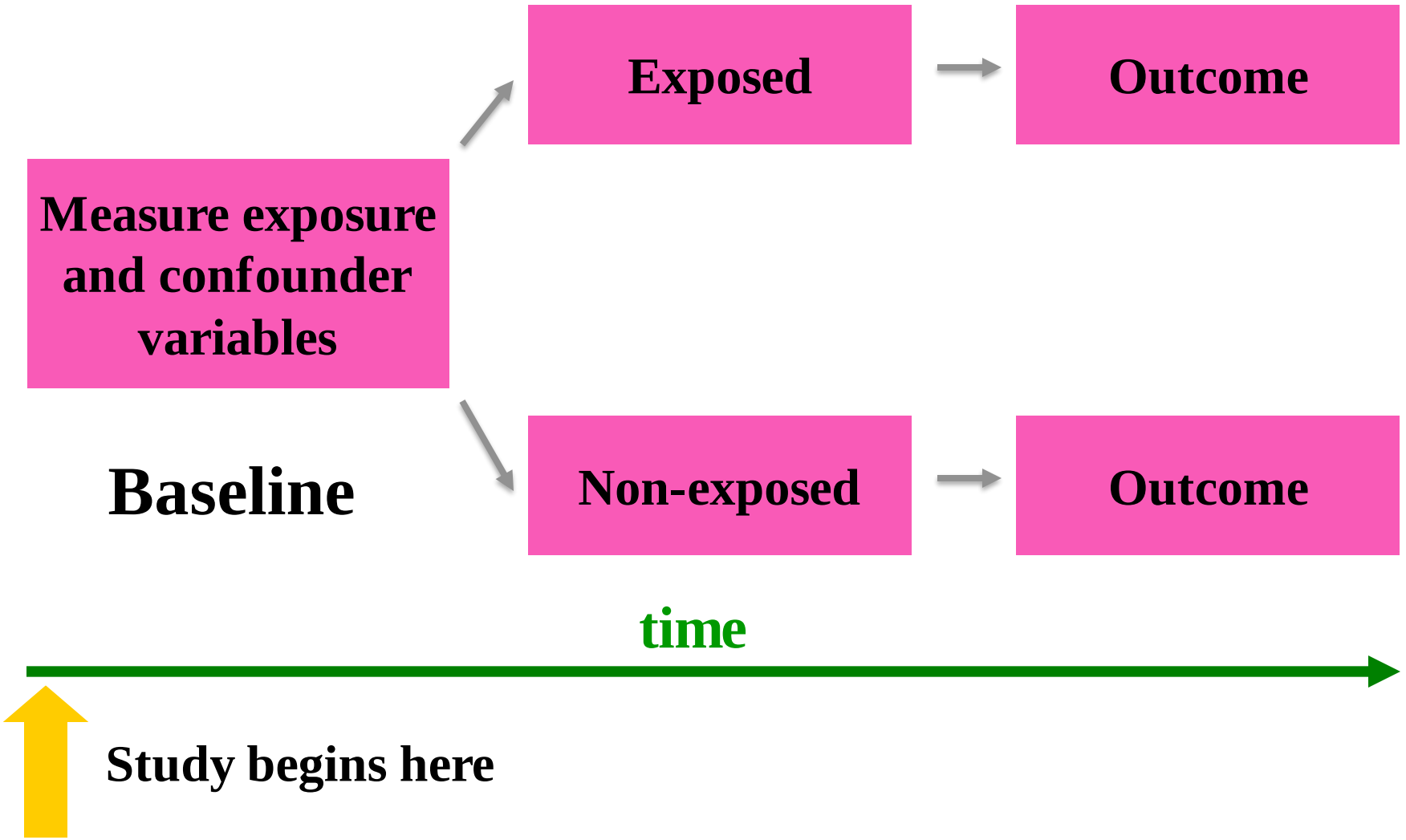
Cohort Design



Timeframe of Studies

- **Prospective Study - looks forward, looks to the future, examines future events, follows a condition, concern or disease into the future**

Prospective Cohort study



Cohort Study

- **Strengths**
  - **Exposure status determined before disease detection**
  - **Subjects selected before disease detection**
  - **Can study several outcomes for each exposure**
- **Limitations**
  - **Expensive (due to follow up) and time-consuming**
  - **Inefficient for rare diseases or diseases with long latency**
  - **Loss to follow-up (attrition, according to inclusion and exclusion criteria, beside of that patients may migrate from the country or die suddenly)**

Incidence rates of outcome: Cohort Study

Exposure Status	Disease Status: Yes	Disease Status: No	Total	
Yes	a	b	a+b	Study cohort
No	c	d	c+d	Comparison cohort
	a+c	b+d	N	

Incidence rates of outcome: Cohort Study

- Incidence among exposed = $\frac{a}{a+b}$
- Incidence among non-exposed = $\frac{c}{c+d}$

Estimation of risk

- **Relative Risk**

$$RR = \frac{\text{incidence of disease among exposed}}{\text{Incidence of disease among non-exposed}}$$

$$= \frac{a/a+b}{c/c+d}$$

Estimation of Risk

- **Attributable Risk**

  Incidence of disease among exposed – incidence of disease among non exposed

AR = ——————————————————————

Incidence of disease among exposed

a/a+b – c/c+d

AR = ——————————

a/a+b

• Smoking	• Lung cancer		• Total
	• YES	• NO	
• YES	• 70	• 6930	• 7000
• NO	• 3	• 2997	• 3000
	• 73	• 9927	• 10000

Find out RR and AR for above data

- Incidence of lung cancer among smokers

  70/7000 = 10 per 1000

- Incidence of lung cancer among non-smokers

  3/3000 = 1 per thousand

RR = 10 / 1 = 10

(lung cancer is 10 times more common among smokers than non smokers)

AR = 10 – 1 / 10 X 100

= 90 %

(90% of the cases of lung cancer among smokers are attributed to their habit of smoking)

1. **Meta-analysis**
2. **Systematic Review**
3. **Randomized Controlled Trial**
4. **Cohort Study**
5. **Case Control Study**
6. **Cross Section Study**
7. **Case Reports, Series**

arranged from the most reliable and valid to the less

Cohort Study

Definition

A study design where one or more samples (called cohorts) are followed prospectively and subsequent status evaluations with respect to a disease or outcome are conducted to determine which initial participants exposure characteristics (risk factors) are associated with it. As the study is conducted, outcome from participants in each cohort is measured and relationships with specific characteristics determined

Advantages

1. **Subjects in cohorts can be matched, which limits the influence of confounding variables**
2. **Standardization of criteria/outcome is possible**
3. **Easier and cheaper than a randomized controlled trial (RCT)**

Disadvantages

1. **Cohorts can be difficult to identify due to confounding variables**
2. **No randomization, which means that imbalances in patient characteristics could exist**
3. **Blinding/masking is difficult**
4. **Outcome of interest could take time to occur**