Cohort Study

Ram Rangsin, MD MPH DrPH
Department of Military and Community Medicine
Phramongkutklao College of Medicine, THAILAND

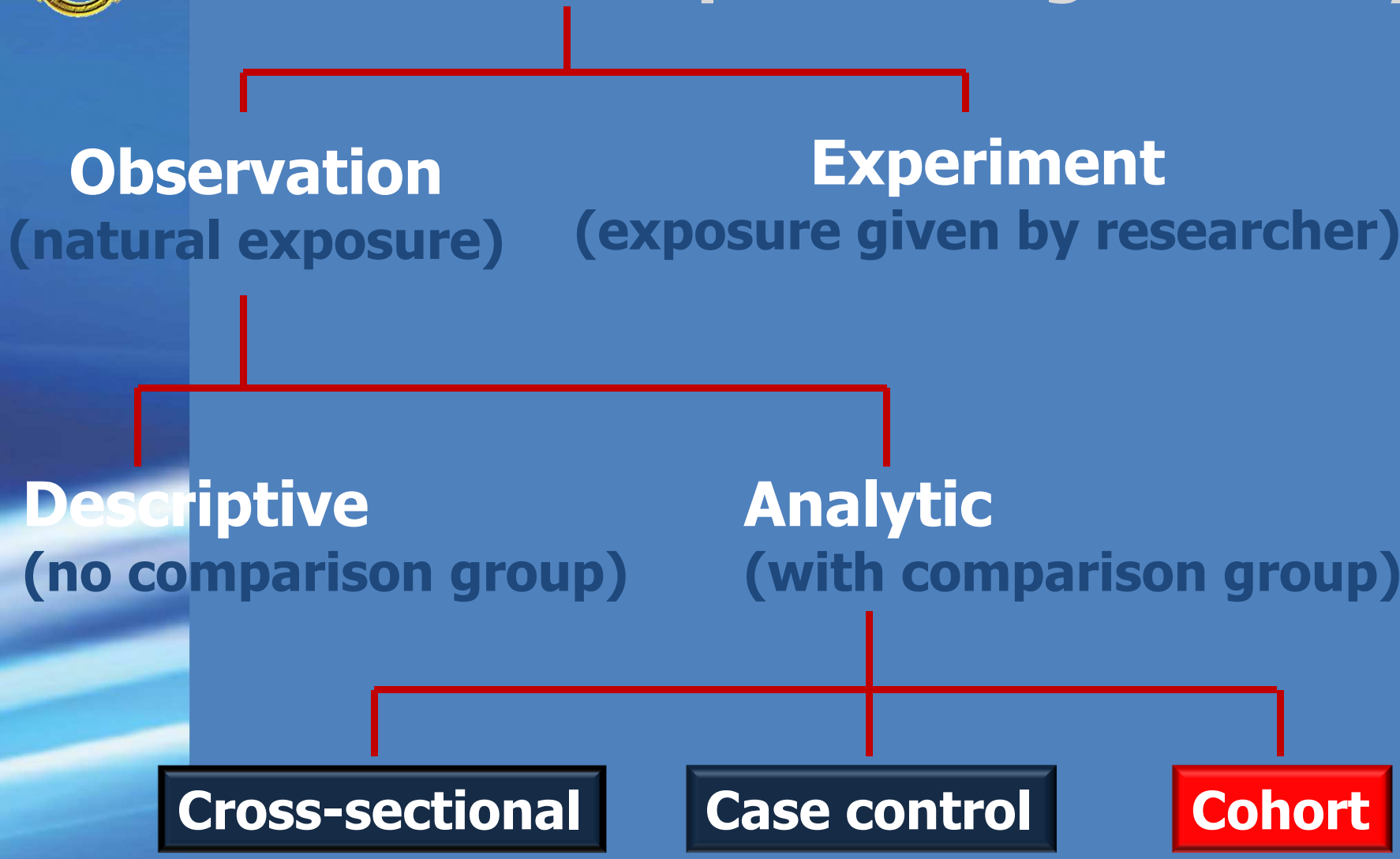
Learning objectives

- Students are able to identify the fundamental concepts of cohort study design.
- Students are able to identify types of cohort studies
- Students are able to design a simple cohort study
- Student are able to understand the concept of Poisson Regression Model using in data analysis of the Cohort Study

Epidemiological Studies

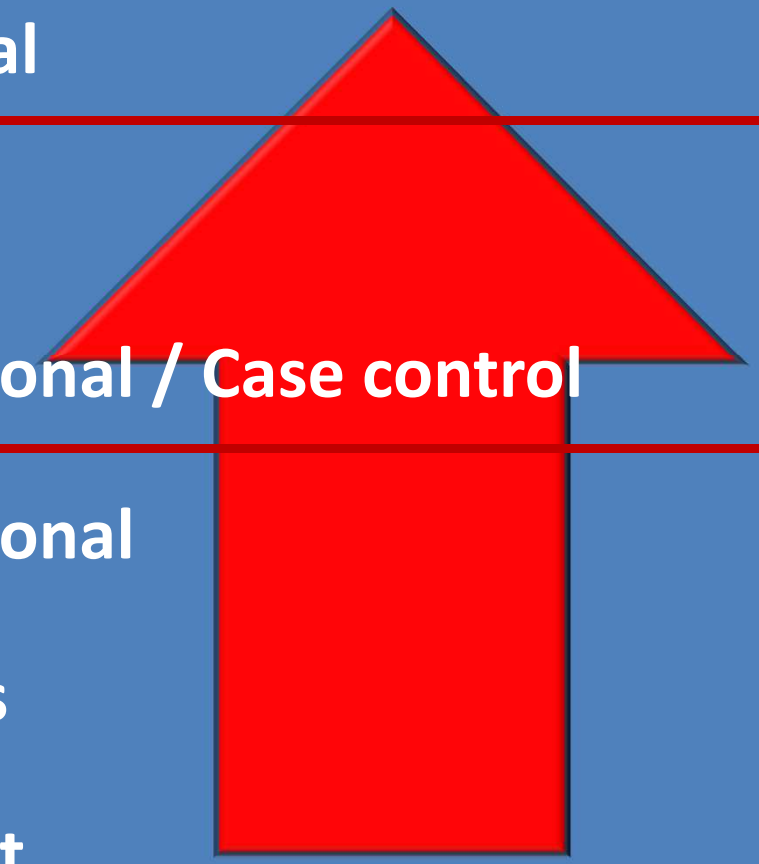
- Describe the problem
  - How large is the problem?
  - How dose the problem distribute in the population?
  - Descriptive Study
    - Time
    - Place
    - Person

Classification epidemiological study



Hierarchy of Epidemiological studies

- Clinical Trial
- Cohort
- Cross-sectional / Case control
- Cross-sectional
- Case series
- Case report



Experiment

Analytic

Descriptive

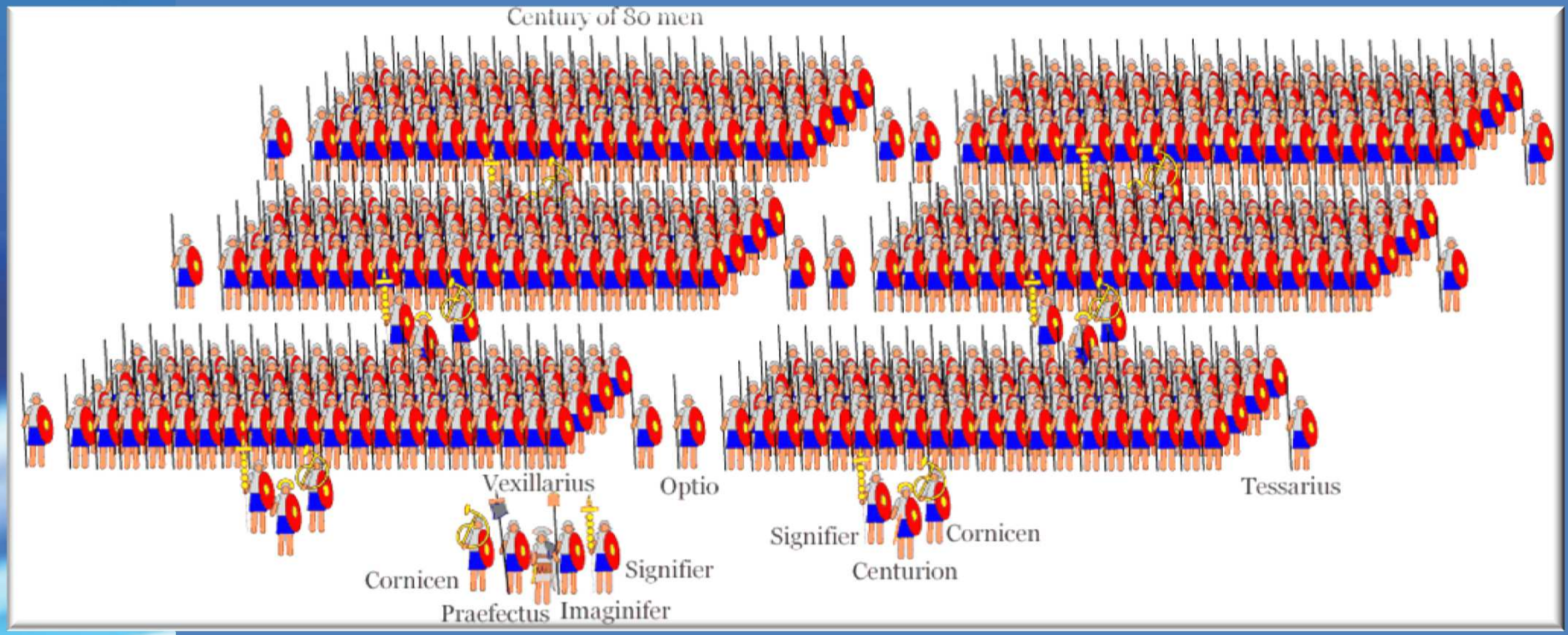
Cohort Study

- **The most powerful observational study for identifying an association between risk factors and a disease**
- **The most time consuming**
- **The most expensive**

“COHORT”

A unit of 300-600 men
in the ancient Roman army

A Roman Cohort



Two centuries made one maniple and three maniples made up one cohort.

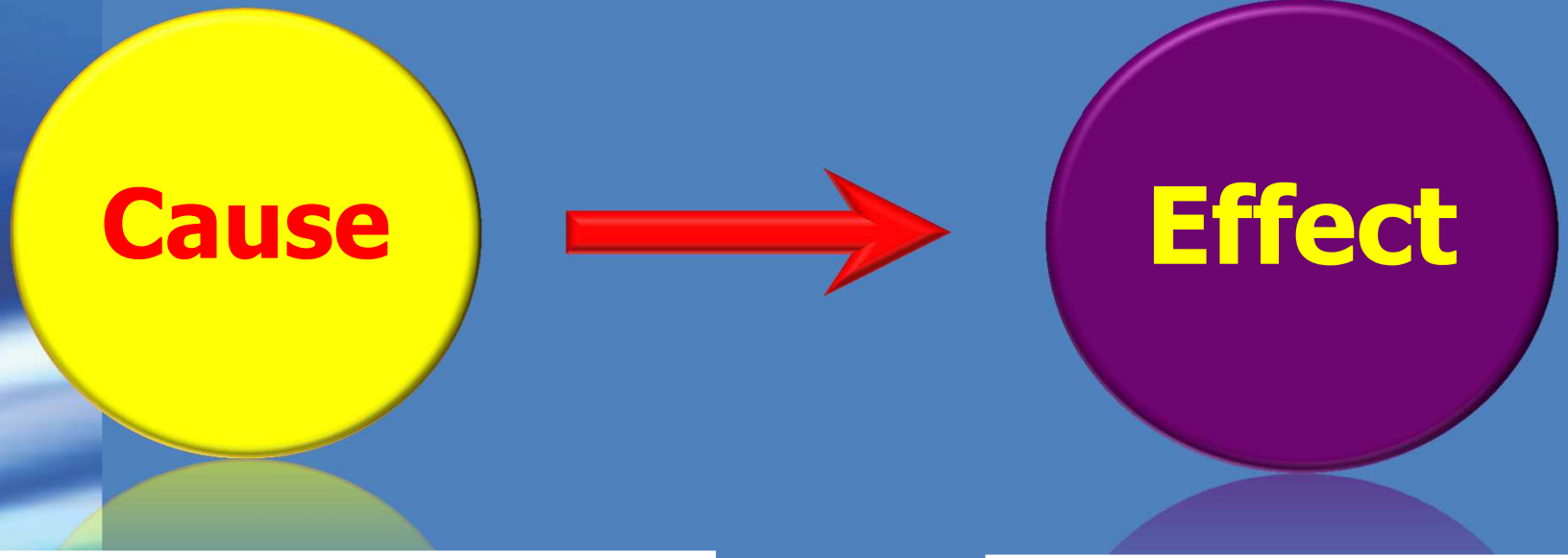
"COHORT"
in Epidemiology

A group of persons who are followed over time

Cohort Study

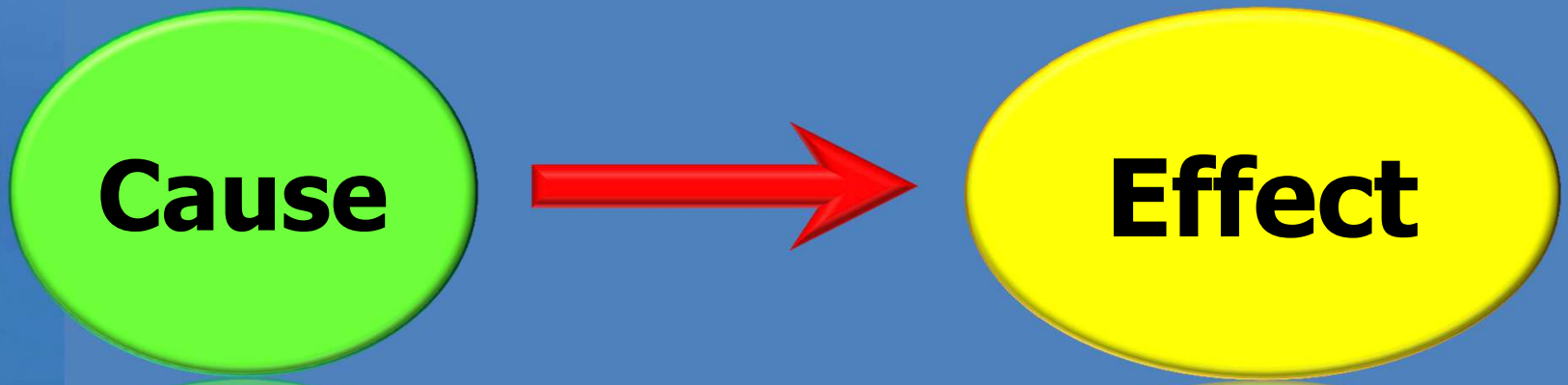
- Start with a group of people without the disease
- Then divide people based on the basis of the exposure to a suspected risk factor
- Follow the "whole group" for a period of time
- Then asses the disease occurrence outcome

Using epidemiology to identify the cause of a disease

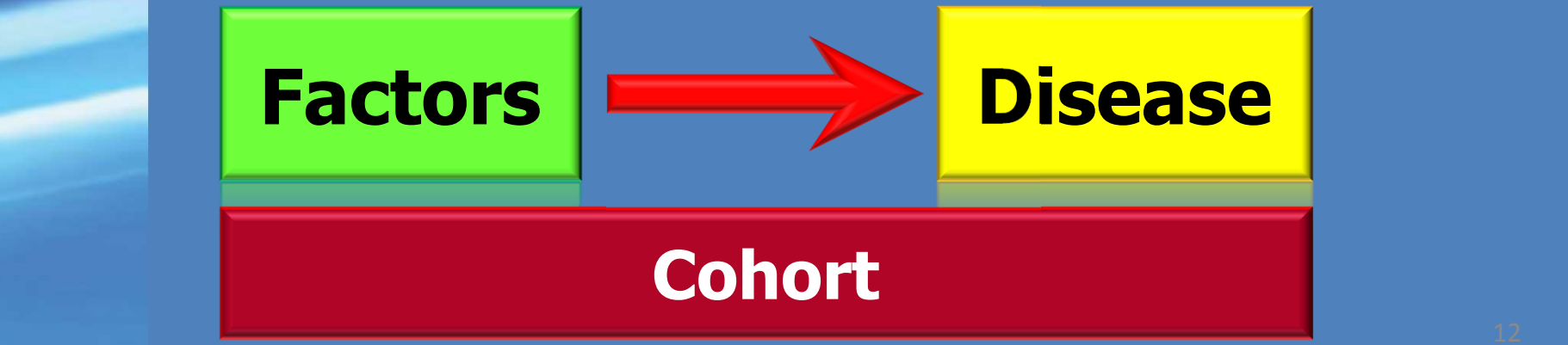


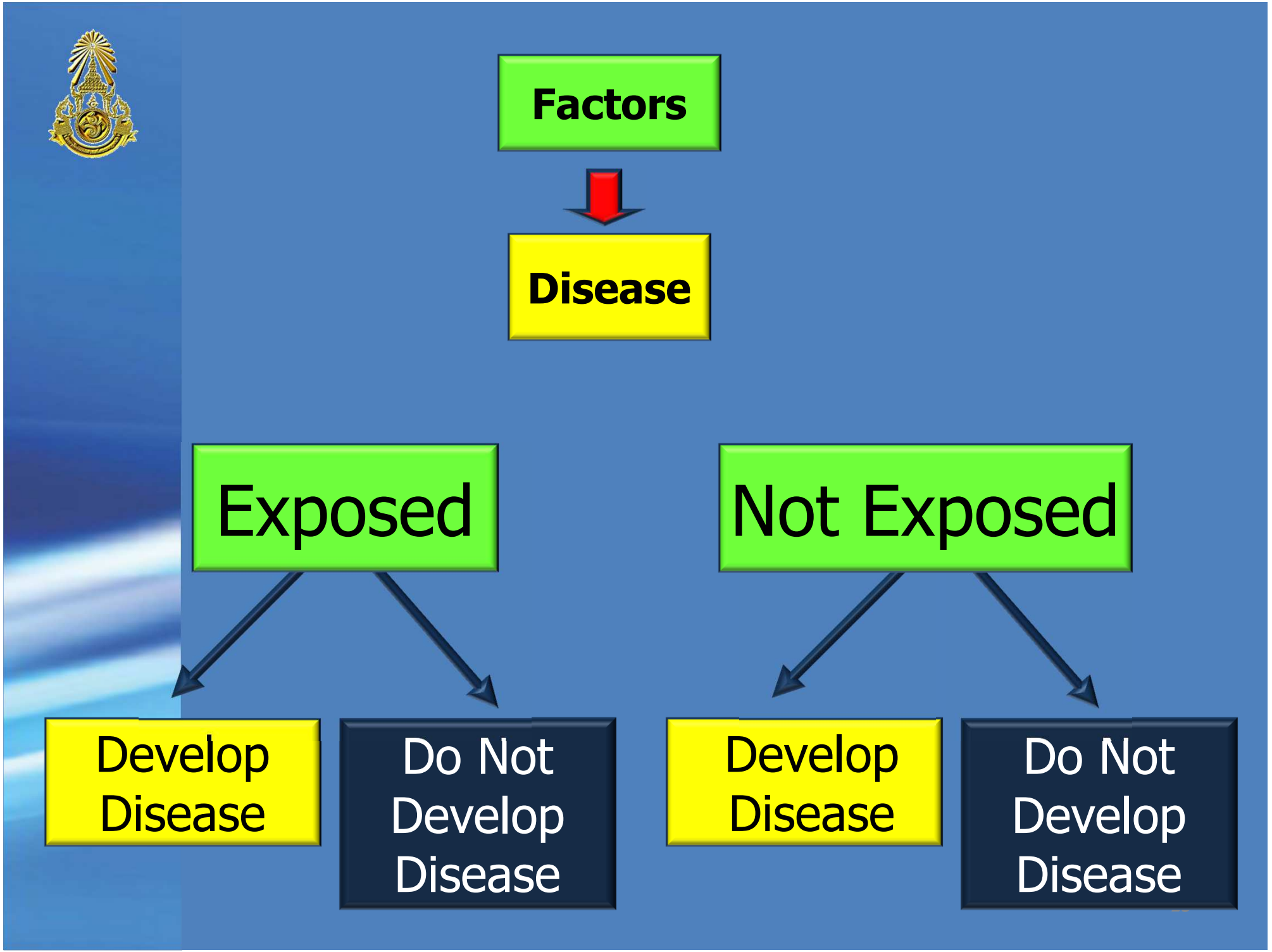
- **RISK FACTOR**
  - **Cigarette**

- **DISEASE**
  - **Lung Cancer**


Cause
Effect


Factors
Disease
Case-Control

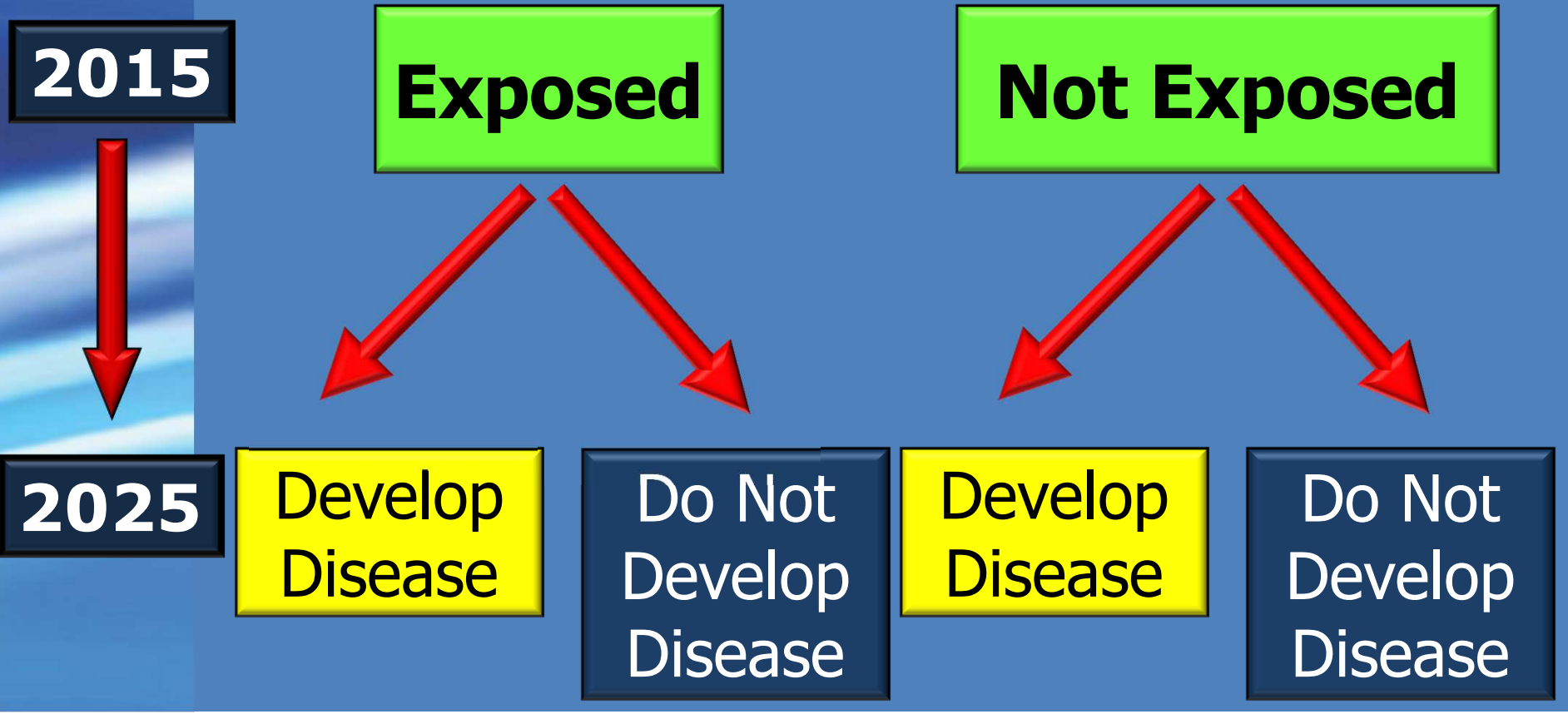

Factors
Disease
Cohort


Factors
Disease
Exposed
Not Exposed
Develop Disease
Do Not Develop Disease
Develop Disease
Do Not Develop Disease

Type of Cohort Study

I. Concurrent Cohort Study

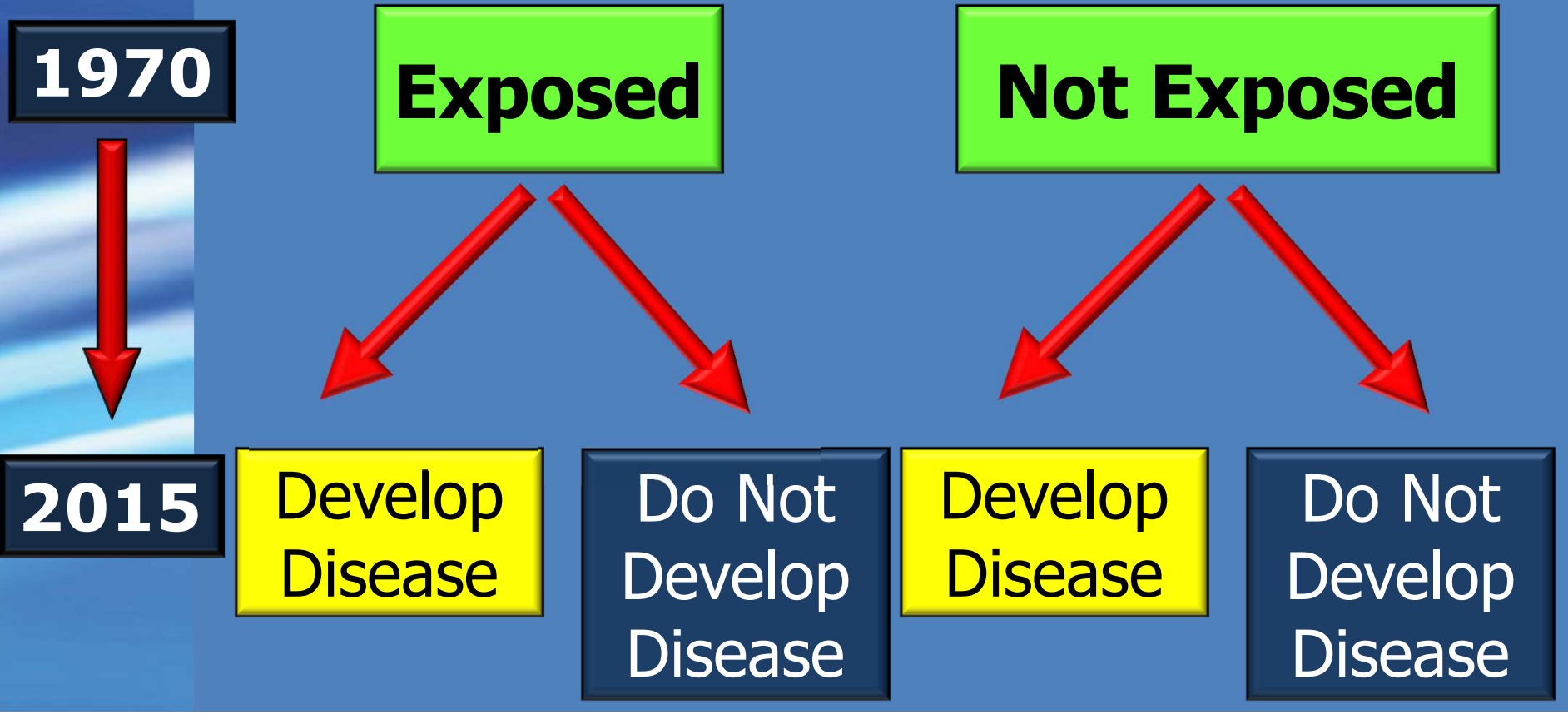
(Prospective Cohort Study)



Type of Cohort Study

II. Retrospective Cohort Study

(Take advantage of records collected)




1970
Persons without the disease!!!!
Exposed
Not Exposed
Smoke
#500 persons
Not Smoke
#500 persons
2015
Disease
No Disease
Disease
No Disease
Lung Cancer
45
No Lung Cancer
455
Lung Cancer
1
No Lung Cancer
499

	CA Lung	No CA	
Smoke	**45**	**455**	**500**
Not smoke	**1**	**499**	**500**

- Incidence of Smoker who develop Lung Cancer = 45 /500
- Incidence of Non -Smoker whodevelop Lung Cancer = 1 /500
- Relative Risk of smoking for Lung Cancer = $\frac{45/500}{1/500}$ = 45
- **Those who smoked were *45 times more likely* to get lung cancer.**

Relative Risk

	CA Lung	No CA	
Smoke	A	B	A+B
Not smoke	C	D	C+D

$$\text{Relative Risk} = \frac{A/A+B}{C/C+D}$$

Interpretation of Relative Risk (RR)

- Relative Risk of smoking for CA Lung = 45
- **Those who smoked were *45 times more likely* to develop lung cancer than those who did not smoke.**

Cohort Study

Comparison between

"a group of persons *with*

a factor -- Exposed"

vs

"a group of persons *without*

the factor -- Non-exposed"

Measurement of Associations

- Cross-sectional → **Prevalence Rate Ratio**
- Case-Control → **Odds Ratio**
- Cohort → **Relative Risk**

Advantages of a cohort study

- Temporal sequence (exposure occur prior the disease) can be more clearly established
- Well suited for assessing the effect of RARE EXPOSURE (e.g. Radiation,)
  - Persons are enrolled on the basis of exposure

Advantages of a cohort study

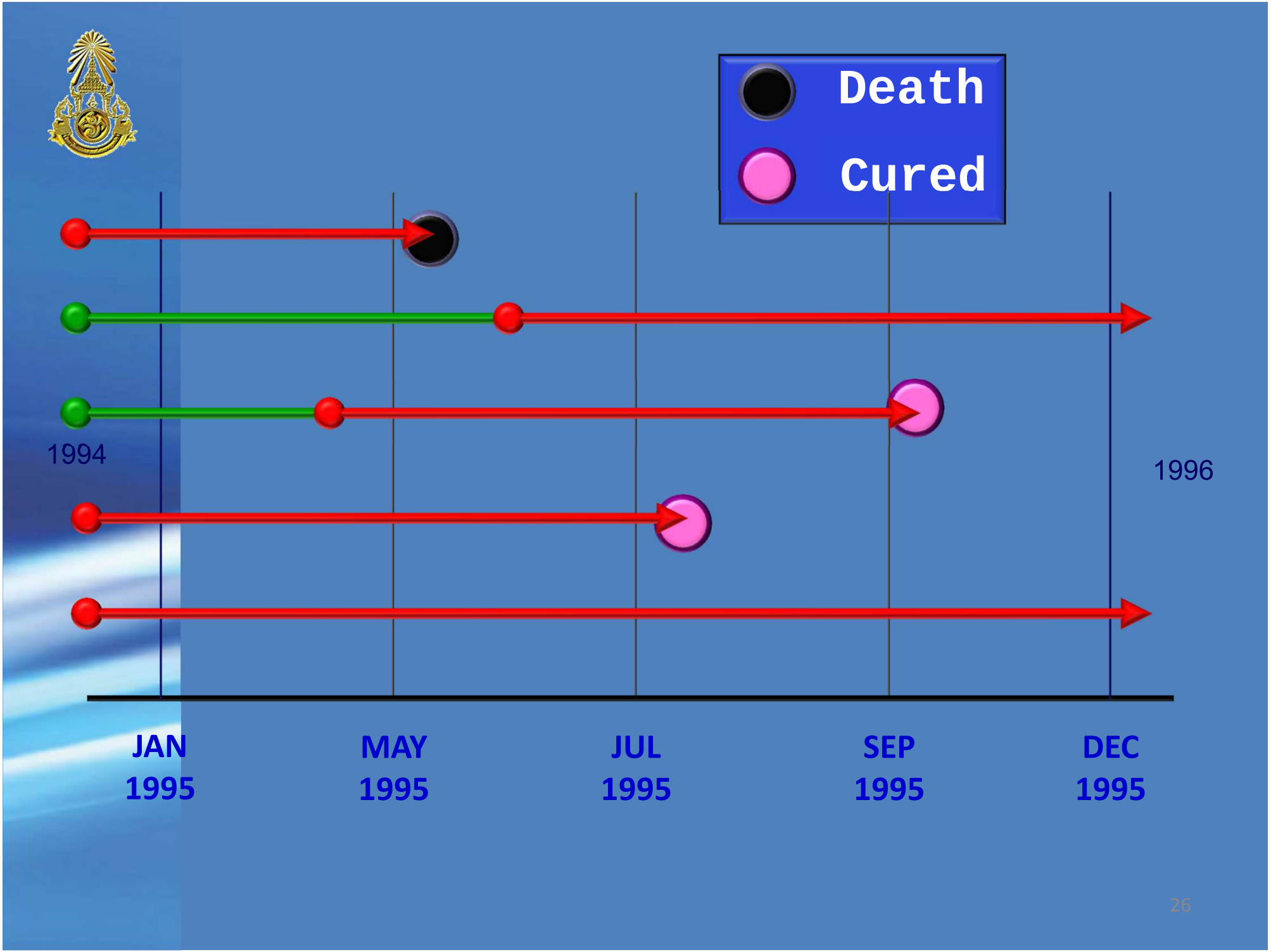
- **Able to examine multiple diseases outcome of a single exposure**
  - **The Nurse Health Study, USA**
  - **120,000 female nurses**
  - **Exposure:** Oral Contraceptive Pill
  - **Outcomes:**
    - Breast cancer
    - Ovarian Cancer
    - Malignant melanoma

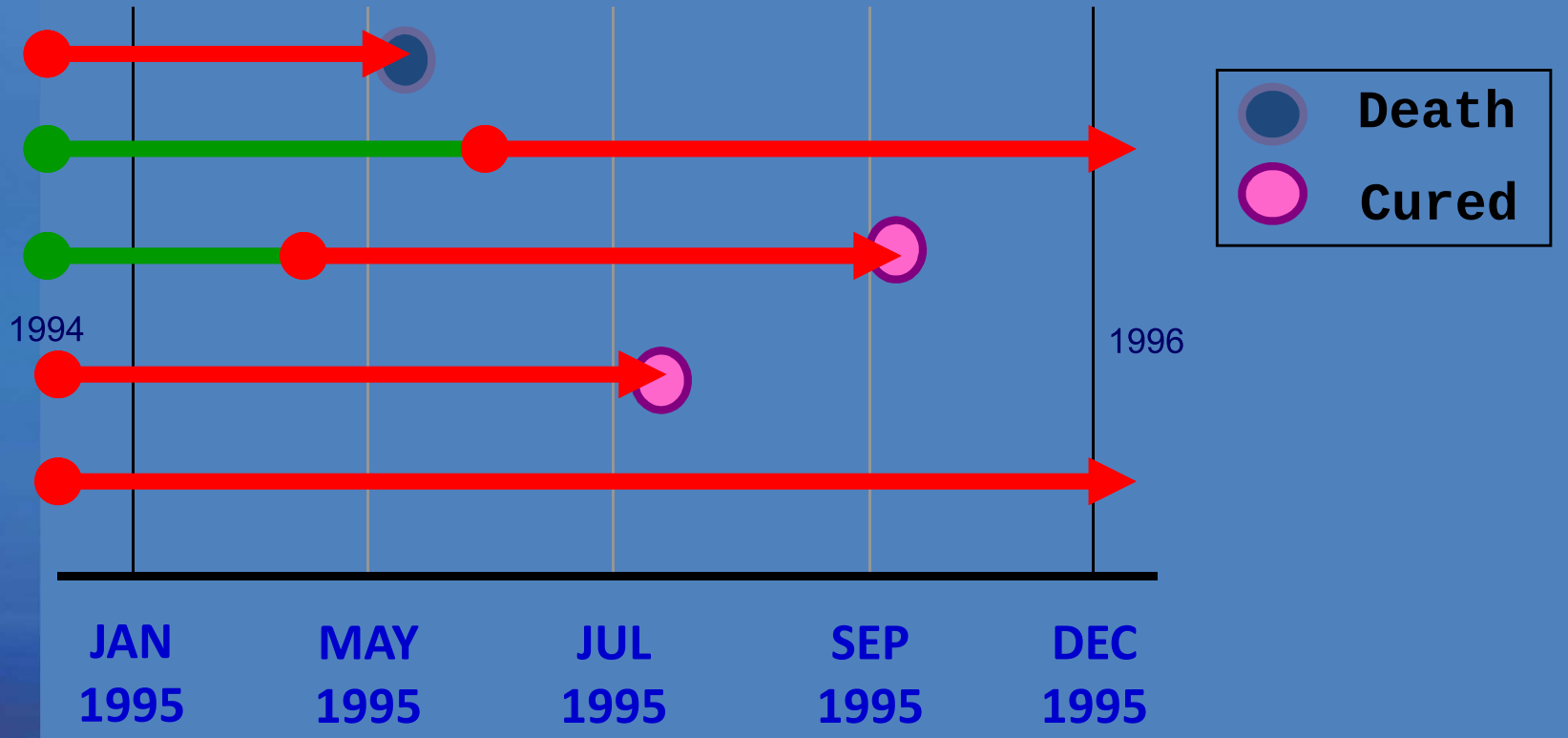
Disadvantages

- Insufficient for the evaluation of rare diseases
- Extremely expensive and time consuming (Prospective)
- Required the availability of adequate records (Retrospective)
- Loss to follow-up

When we are conducting a cohort study,

we are dealing with “INCIDENCE”.

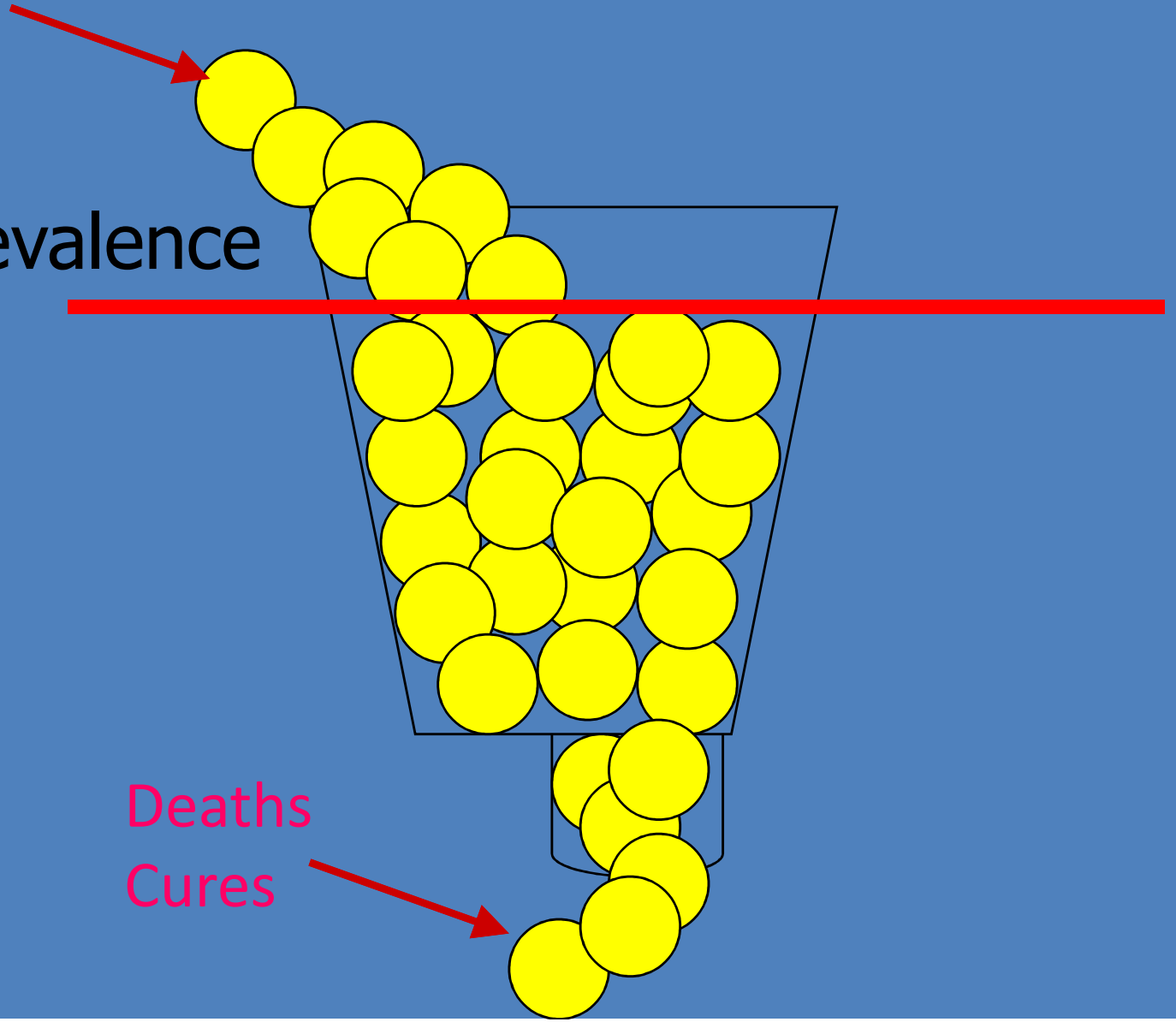

Death
Cured
1994
1996
JAN
1995
MAY
1995
JUL
1995
SEP
1995
DEC
1995



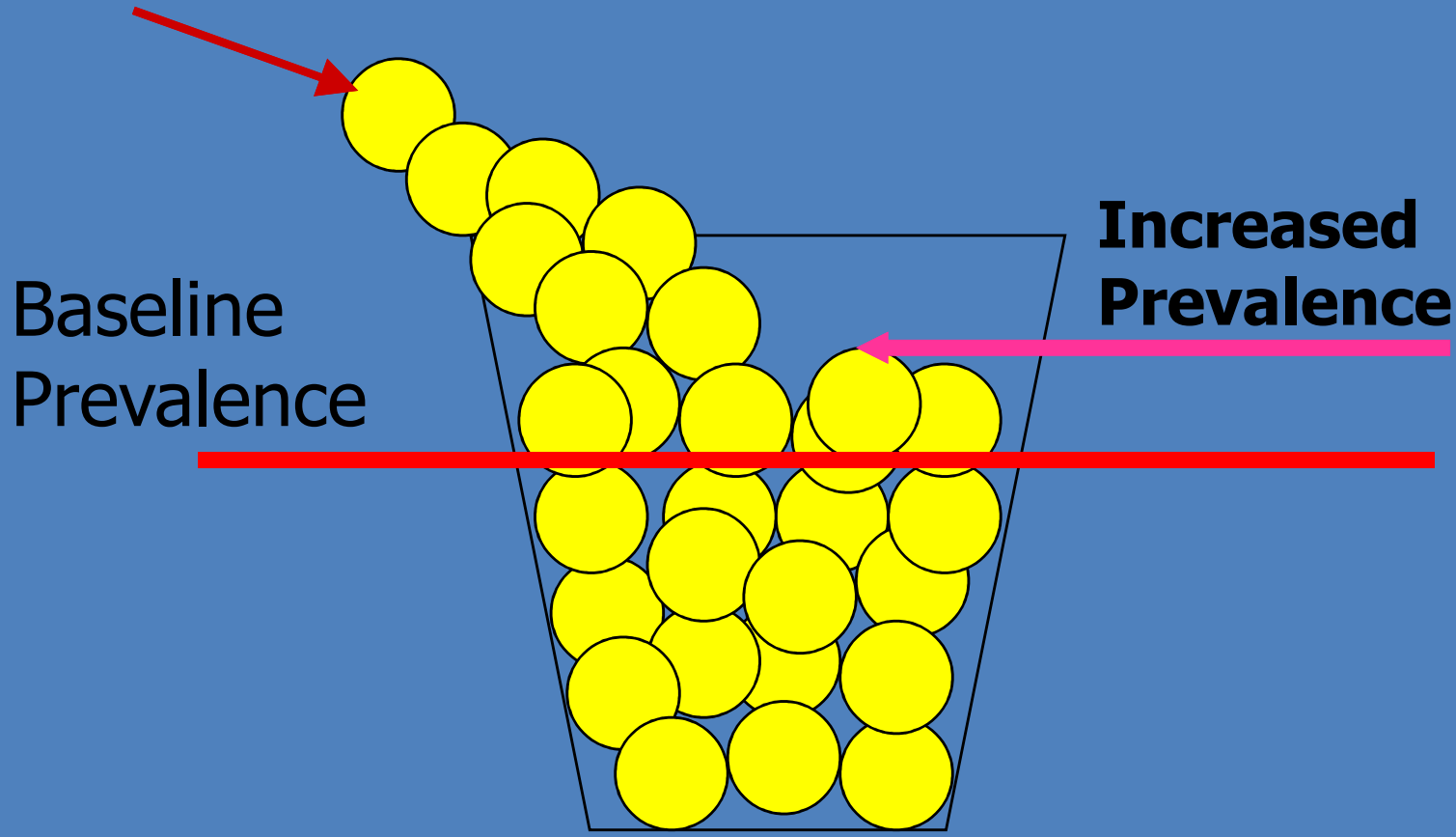
Incidence in 1995 = ?

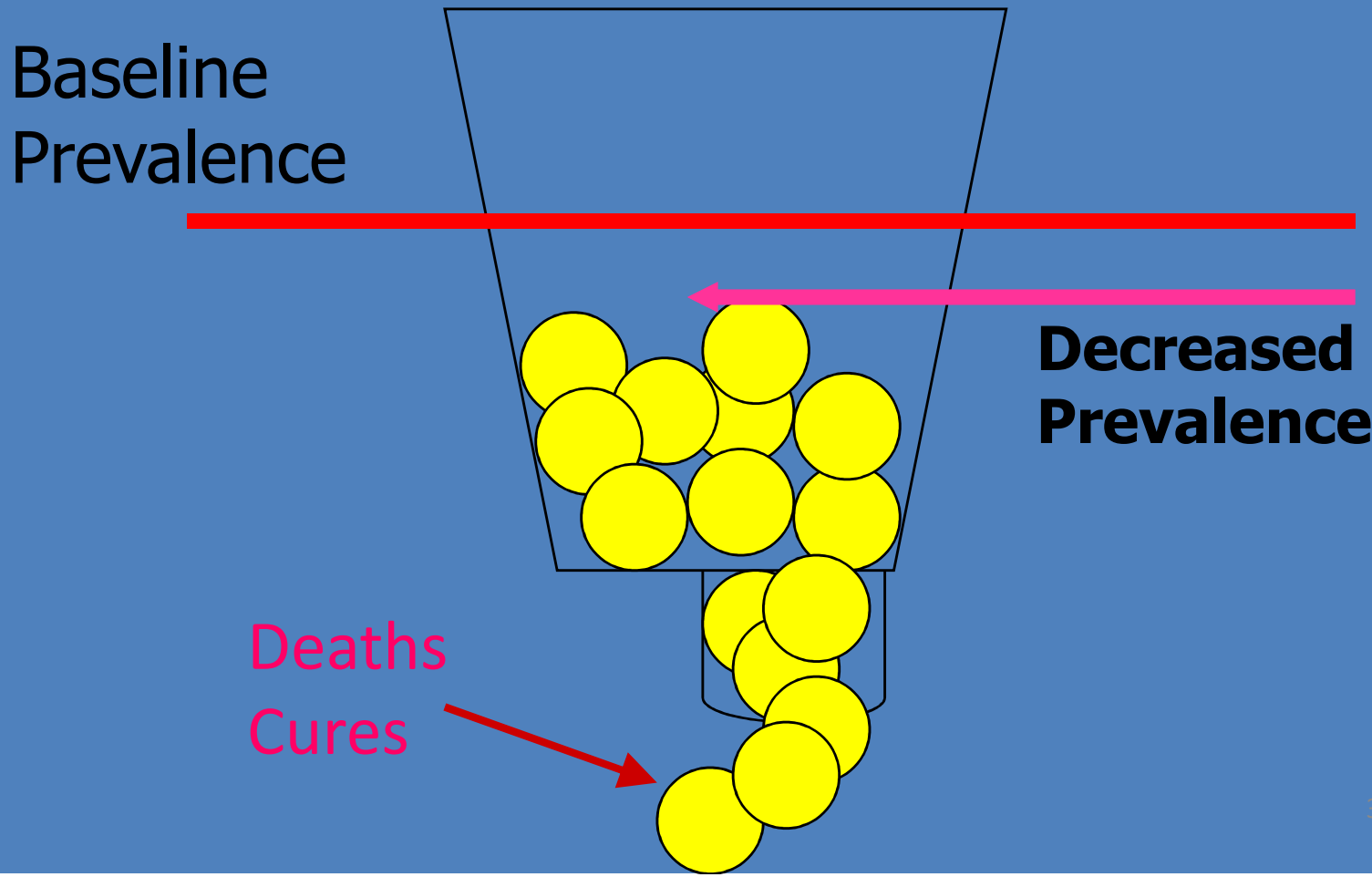
Point Prevalence at July 1995 = ?

Incidence



Incidence




Baseline
Prevalence
Decreased
Prevalence
Deaths
Cures

Rate

- The central tool of Epidemiology is the comparison of RATES
- RATE = $\frac{\text{Numerator}}{\text{Denominator}}$
  - Mortality Rate
  - Prevalence
  - Incidence

Measuring the incidence

There are two ways of measuring

1) Cumulative incidence

$$= \frac{\text{number of new case in specified time}}{\text{population at risk in specified time}} \times 10^{(n)}$$

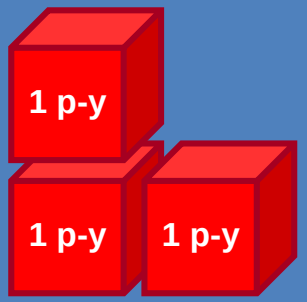
$$= \frac{40}{32{,}000} = 1.25 \text{ /1,000}$$

Measuring the incidence

2) Incidence density or Incidence rate

- Adding "TIME Dimension" into the denominator

"Person-time"



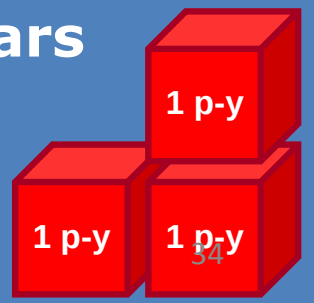
- person-month, person-year
- 1 person-year = Following 1 person for 1 year period
- 10 person-year = Following 1 person for 10 year period
  = Following 10 persons for 1 year period

Measuring the incidence

- 2) Incidence density or Incidence rate

$$= \frac{\text{Number of new case in specified time} \times 10^{(n)}}{\text{Person-years of observation which is disease free}}$$

- **If <u>100 subjects are followed for 1 year</u> and <u>20 develop disease</u>, the incidence density is**
- **20 cases/100 person-years of observation**
- $\frac{\textbf{20}}{\textbf{100 person-years}} = \textbf{20 / 100 person-years}$



Relationship between cigarette smoking and incidence rate of stroke in a study of 118,539 population in over 8 years period

Smoking	No. of stroke	Person-years of observation	Incidence rate /100,000 person-years
Never	70	395,594	17.7
Ex-smoker	65	232,712	27.9
Smoker	139	280,141	49.6
Total	274	908,477	30.2

Database of 118,539 subjects

ID	Age	smoking	Stroke	Enter	Last Contact	Person-Year
1	18	No	No	1990	1998	8
2	36	No	No	1990	1992	2
3	50	Yes	Yes	1991	1998	7
4	42	Ex	No	1993	1995	2
.	.	.	.	.	.	.
118,539	24	Yes	No	1993	1998	5
Total						**908,477**

Relationship between cigarette smoking and incidence rate of stroke in a study of 118,539 population in over 8 years period

Smoking	No. of stroke	Person-years of observation	Incidence rate /100,000 person-years
Never	70	395,594	17.7
Ex-smoker	65	232,712	27.9
Smoker	139	280,141	49.6
Total	274	908,477	30.2

- **RR: Smoke VS Never =**
- **RR: Ex-Smoke VS Never =**

Data Analysis in Cohort Study

- Incidence
  - Cumulative incidence = … / 100 persons
  - Incidence Rate (Density) = … / 100 person-years
- Relative Risk: Univariate Analysis
  - Cumulative incidence = Relative Risk (… / 100 persons)
  - Incidence Rate (Density) = Relative Risk (… / 100 person-years)
- Relative Risk: Multivariate Analysis
  - Cumulative incidence = Adjusted Relative Risk
    → Multiple Logistic Regression Model
  - Incidence Rate (Density) = Adjusted Relative Risk
    → Poisson Regression Model

Exposure assessment

- Exposed VS Non-exposed
- Fixed Exposure
- Time-dependent Exposure(Exposure level changes through time)

Fixed Exposure

- Exposure do not change over time
  - Sex (Male / Female)
  - Blood group (A / B / O / AB)
  - Race (White / Black / Asian)
  - Expose to radiation from the power plant explosion
  - Adult height

Time-dependent

- **Exposure level changes over time**
  - **Body weight**
  - **Alcohol consumption**
  - **Blood pressure level**

Outcome assessment

- **Disease**
- **Specify clearly what is your final outcome**
  - **Disease**
  - **Death**
  - **Intermediate outcomes**
    - **CD4+ count**
    - **Increased Creatinine**

Conducting a Cohort Study

- **Selecting a group of people without the disease**
- **Defining the Exposed group**
- **Defining the Non-exposed group**
- **Evaluate the disease outcome among both Exposed and Non-exposed**
- **Calculating Relative Risk**

Conducting a Cohort Study

- **You are interested in the association between blood cholesterol level and coronary artery heart disease**
- **Please conduct a cohort study to verify the association**

Conducting a Cohort Study

- **What population would you like to start with?**

Conducting a Cohort Study

- **How can you identify those who will be the “study population”?**

Conducting a Cohort Study

- **How can you identify exposed and non-exposed groups?**

Conducting a Cohort Study

- **What is you follow-up plan?**
  - **What is you outcome of interest?**
  - **How often you would like to asses the outcome?**
  - **How long will you follow the population?**

Conducting a Cohort Study

- What is you plan for the analysis?
  - What will be the measurement of association from your study?
  - What would you like to compare?

Framingham study

- **Framingham study of cardiovascular disease**
  - **Individuals 30 – 62 years old in community at risk for disease**
  - **Framingham, MA, 1948 to present**

Framingham study

	No. Men	No. Women	Total
Random Sample	3,074	3,433	6,507
Respondents	2,024	2,445	4,469
Volunteers	312	428	740
Respondents free of CHD	1975	2,418	4,393
Volunteers free of CHD	307	427	734
Total free of CHD	2,282	2,845	5,127

Framingham study

Cholesterol level	CHD	No CHD	Total
`High'	57	305	362
`Low' <250	71	1098	1169

References

- Gordis, L. 2000. Epidemiology, 2nd ed., Philadelphia: W.B. Saunders.
- Last, M. et. al. 2001. Dictionary of Epidemiology, 4th ed., New York; Oxford University Press.
- Hennekens, H. H., et. al. 1987, Epidemiology in Medicine, Boston: Little, Brown.

- rrangsin@pcm.ac.th