# Cohort Studies

Nguyen Thi Tu Van[1] Nguyen Quang Vinh[2]

[1] University of Medicine and Pharmacy of Ho Chi Minh City

[2] Nguyen Tri Phuong Hospital

JICA project - August 5, 2014

# Outline

# Introduction

Cohort $\equiv$ group of people sharing a common condition.

A cohort study: cohort(s) being free from the disease under investigation:

- followed-up to assess the incidence of the outcome(s) of interest.
  The required length to follow up is about the length of the latent period of the outcome of interest.
- group(s) are selected by different levels of selected exposure (>2 groups).
  More often, the presence or absence of a suspected risk factor for a disease (2 groups).

# Introduction

Cohort $\equiv$ group of people sharing a common condition.
A cohort study: cohort(s) being free from the disease under investigation:

- followed-up to assess the incidence of the outcome(s) of interest.
  The required length to follow up is about the length of the latent period of the outcome of interest.
- group(s) are selected by different levels of selected exposure (>2 groups).
  More often, the presence or absence of a suspected risk factor for a disease (2 groups).

# Goal

- to measure (always)
- to compare (usually)

the incidence of disease in one or more study cohorts

# Goal

- to measure (always)
- to compare (usually)

the incidence of disease in one or more study cohorts

# Descriptive - Analytic

- Descriptive study: looking at individual variables, one at a time.
- Analytic study: analyzing the relationships among $\geq 2$ variables $\longrightarrow$ to predict outcomes, to draw inferences about cause-effect.

# Descriptive - Analytic

- Descriptive study: looking at individual variables, one at a time.
- Analytic study: analyzing the relationships among $\geq 2$ variables $\longrightarrow$ to predict outcomes, to draw inferences about cause-effect.

# Descriptive - Analytic

- Descriptive study: looking at individual variables, one at a time.
- Analytic study: analyzing the relationships among $\geq 2$ variables $\longrightarrow$ to predict outcomes, to draw inferences about cause-effect.

# Prospective - Retrospective

Classified as either Prospective or Retrospective $\in$ the initiation time of the study, $t_{initiation} \sim$ the occurence of *outcome(s)*:

$$
\begin{array}{rlll}
Retrospective: & outcome(s) \rightarrow & t_{initiation} & \\
Prospective: & & t_{initiation} \rightarrow & outcome(s) \\
Retro-Prospective: & outcome(s) \rightarrow & t_{initiation} \rightarrow & outcome(s)
\end{array}
$$

# The choice of Descriptive or Analytic

## $\in$ The study objective

The study can be designed:

Descriptive

*or*

Descriptive and Analytic

# The choice of Descriptive or Analytic

$\in$ The study objective

The study can be designed:

Descriptive

*or*

Descriptive and Analytic

# The choice of Retrospective or Prospective

$\in$ Scientific $\sim$ Logistic considerations

- Prospective:
  - minimizing bias in the ascertainment of exposure.
  - time consuming, expensive.
- Retrospective:
  - more quickly & cheaply.
  - the availability of adequate records.

# The choice of Retrospective or Prospective

$\in$ Scientific $\sim$ Logistic considerations

- Prospective:
  - minimizing bias in the ascertainment of exposure.
  - time consuming, expensive.
- Retrospective:
  - more quickly & cheaply.
  - the availability of adequate records.

# The choice of Retrospective or Prospective

$\in$ Scientific $\sim$ Logistic considerations

- Prospective:
  - minimizing bias in the ascertainment of exposure.
  - time consuming, expensive.
- Retrospective:
  - more quickly & cheaply.
  - the availability of adequate records.

# Example 1 - An descriptive prospective cohort study

| **A Cohort of** | | **Outcome** |
|---|---|---|
| *Nulliparous women* | $\rightarrow$ | *Incidence of Pregnancy Induced Hypertension* |
| *Pregnant women* | $\rightarrow$ | *Incidence of Pregnancy Induced Diabetes* |

# Example 2 - An analytic prospective cohort study

| **Predictor** | | **Outcome** |
|---|---|---|
| *Pregnancy Intention* | $\rightarrow$ | *Maternal Confidence in child rearing* |
| *Unintended* | | *No or not sure* |
| *Intended* | | *Yes* |

# Example 3 - An analytic cohort study

## Prospective or Retrospective

| **Predictors** | | **Outcome** |
|---|---|---|
| *Medicine University Entrance Scores* | $\rightarrow$ | *Doctor Graduation* |
| *Mathematics* | | *Passed* |
| *Chemistry* | | *Failed* |
| *Biology* | | |

- *If* prospective:
  - can control for other factors...
  - costly in terms of time, money, personels,...
- *If* retrospective:
  - quickly & no cost added.
  - completely depend on available records.

# Example 3 - An analytic cohort study

## Prospective or Retrospective

| **Predictors** | | **Outcome** |
|---|---|---|
| *Medicine University Entrance Scores* | $\rightarrow$ | *Doctor Graduation* |
| *Mathematics* | | *Passed* |
| *Chemistry* | | *Failed* |
| *Biology* | | |

- *If* prospective:
  - can control for other factors...
  - costly in terms of time, money, personels,...
- *If* retrospective:
  - quickly & no cost added.
  - completely depend on available records.

# Example 3 - An analytic cohort study

## Prospective or Retrospective

| **Predictors** | | **Outcome** |
|---|---|---|
| *Medicine University Entrance Scores* | $\rightarrow$ | *Doctor Graduation* |
| *Mathematics* | | *Passed* |
| *Chemistry* | | *Failed* |
| *Biology* | | |

- *If* prospective:
  - can control for other factors...
  - costly in terms of time, money, personels,...
- *If* retrospective:
  - quickly & no cost added.
  - completely depend on available records.

# Advantages of Cohort studies

- Measure incidence of disease in group(s)
- $E \sim O$ temporal sequence is more clearly established
- Well suited for:
  - Assessing multiple effects of a single exposure
  - Assessing the effects of rare exposure

# Disadvantages of Cohort studies

- Issues of failure to follow-up.
- Inefficient for studying rare diseases, *unless the %AR is high.*

# Efficient designs for rare outcome(s) and/or expensive predictor(s)

- *Nested Case-Control studies*
  - sample is from the cohort when cases occur (incidence density sampling)
- *Case-Cohort studies*
  - sample is from the cohort at baseline (a subcohort)

# Selection of the exposed

- For common exposure: smoking, coffee drinking, alcohol, ...
  $\rightarrow$ general population cohort(s).
- For rare exposure: occupational, environmental/geographical locations: dioxin, ionizing radiation, stress (after earthquakes, terrorism), ...
  $\rightarrow$ special exposure cohort $\equiv$ **more efficient**
  *(with a caution that's special exposure people are, on average, healthier than usual ones).*

# How to select good comparison group(s)

- The major principle: comparable $\equiv$ as similar as possible, *except* the determinant under investigation.
- Confounder(s): social-economic, demographic, geographic... characteristics.
- Other co-risk factors: alcohol, foods...
- ($\pm$) Multiple comparison groups.

# Variety of sources of data

- Interviews, or questionnaires
- Medical examination, or testing
- Measuring the environment where cohort members have lived, or worked
- Records: medical charts, computerized database
- ...

# The importance of follow-up rate

- Failure to follow-up: the major issue leading to bias in cohort studies $\rightarrow$ un-interpretable results.
- The longer the required observation period, the more difficult to achieve satisfactory follow-up rate.

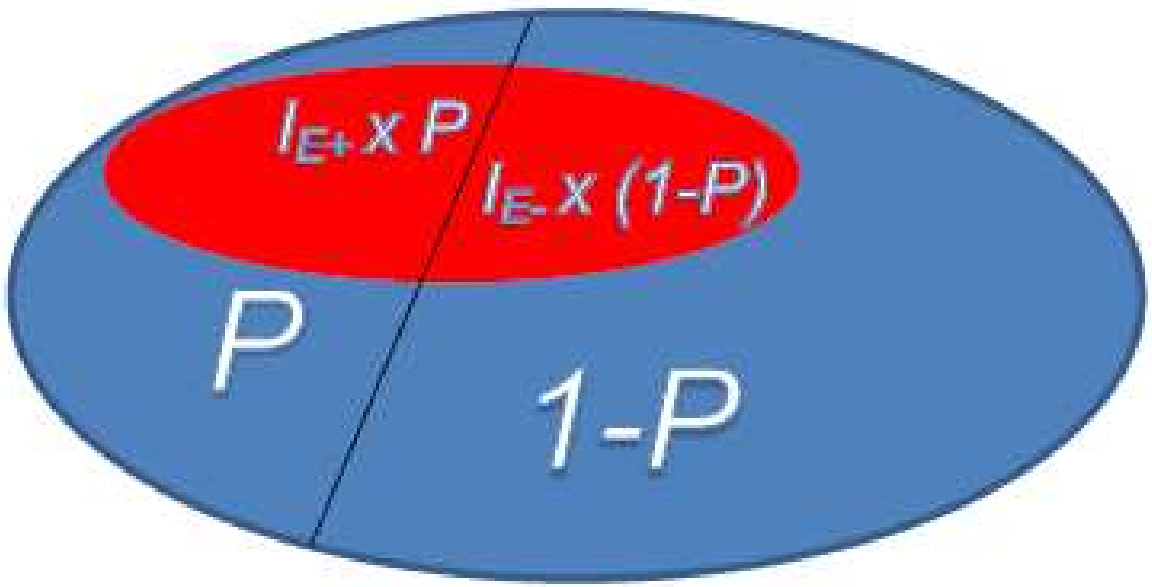
# Analysis

| | $D^+$ | $D^-$ |
|---|---|---|
| $E^+$ | $a$ | $b$ |
| $E^-$ | $c$ | $d$ |

- $Incidence\ E^+ = \frac{a}{a+b}$
- $Incidence\ E^- = \frac{c}{c+d}$
- $Incidence\ T = \frac{a+c}{a+b+c+d}$
- $Relative\ Risk\ (RR) = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{I_{E^+}}{I_{E^-}}$
- $Excess\ Risk\ (ER) = \frac{a}{a+b} - \frac{c}{c+d} = I_{E^+} - I_{E^-}$
- $Attributable\ Risk\ (AR) = \frac{\frac{a}{a+b} - \frac{c}{c+d}}{\frac{a}{a+b}} = \frac{I_{E^+} - I_{E^-}}{I_{E^+}} = \frac{RR-1}{RR}$

# Analysis

- Levin's Population Attributable Risk - the risk of the disease in a population:
  $Levin's\ ARp = [I_{E+} \times P] + [I_{E-} \times (1 - P)]$
  ($P$ = *prevalence of exposure in a population*)




**Levin's ARp**

# Analysis

- Population Attributable Risk - the difference risk (between $E^+$ and $E^-$) of the disease in a population:
  $ARp = Excess\ Risk \times P = (I_{E^+} - I_{E^-}) \times P$[1]

- Population Attributable Fraction - the fraction of the disease occurrence associated with the risk factor in a population:
  $AFp = \frac{ARp}{I_T}$
  ($I_T = Total\ incidence\ of\ disease\ in\ a\ population$)

[1] $ARp = Levin's\ ARp - I_{E^-}$

# Analysis

- Population Attributable Risk - the difference risk (between $E^+$ and $E^-$) of the disease in a population:
  $ARp = Excess\ Risk \times P = (I_{E^+} - I_{E^-}) \times P$[1]

- Population Attributable Fraction - the fraction of the disease occurrence associated with the risk factor in a population:
  $AFp = \frac{ARp}{I_T}$
  ($I_T = Total\ incidence\ of\ disease\ in\ a\ population$)

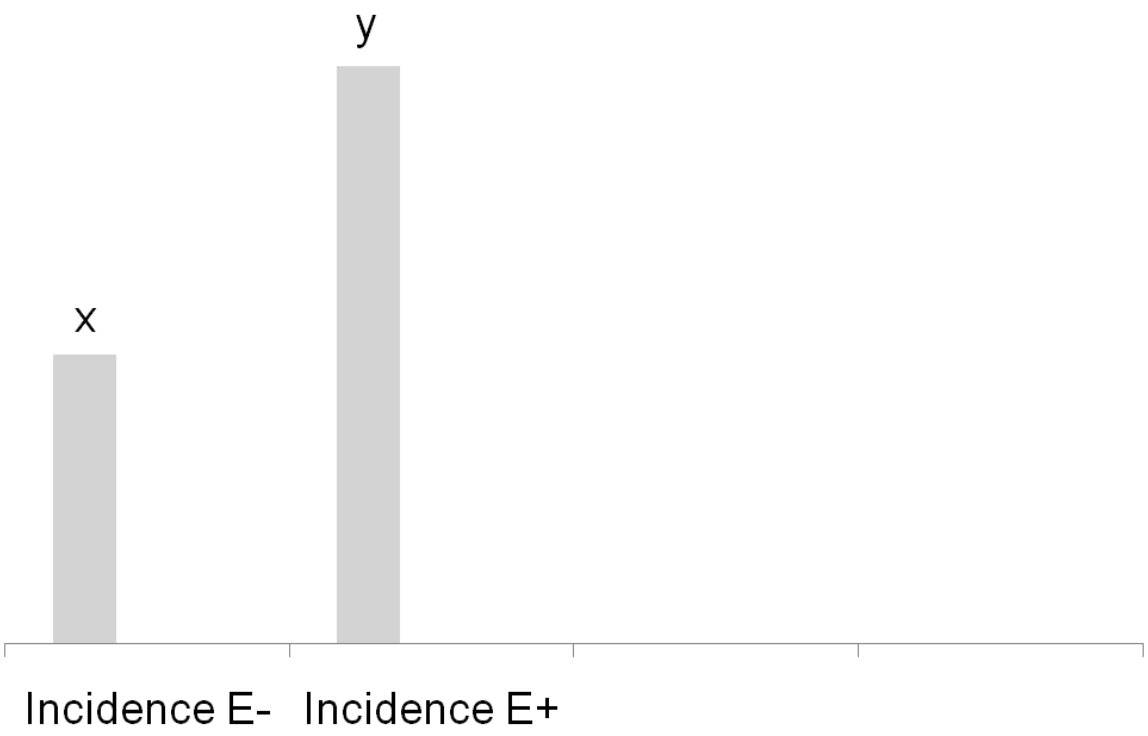
[1] $ARp = Levin's\ ARp - I_{E^-}$

# Other terms

- $Relative\,Risk = Risk\,Ratio\,(RR) = \frac{I_{E^+}}{I_{E^-}}$
- $Excess\,Risk\,(ER) = Risk\,Difference\,(RD) = Attributable\,Risk\,(AR) = I_{E^+} - I_{E^-}$
- $Attributable\,Risk\,(AR) = Attributable\,Risk\,Percent\,(\%AR) = \frac{I_{E^+} - I_{E^-}}{I_{E^+}}$

# $RR \sim ER$

- $RR = \frac{y}{x}$
- $ER = y - x$



# The use of Odds Ratio in Cohort studies

- *Odds Ratio* can approximate *Relative Risk*[2]
- *Odds Ratio* can be used to control simultaneously potential confounders

[2] *OR* is a good estimation for *RR*

- only when the probability of the event of interest $< 10\%$.
- above this boundary, *OR* will overestimate *RR*.

# The use of Odds Ratio in Cohort studies

- *Odds Ratio* can approximate *Relative Risk*[2]
- *Odds Ratio* can be used to control simultaneously potential confounders

[2] *OR* is a good estimation for *RR*

- only when the probability of the event of interest $< 10\%$.
- above this boundary, *OR* will overestimate *RR*.

# Issues in interpretation

## Chance

## Biases

- assessment of the outcome
- information, due to:
  - the extent and quality of obtaining $E^+$ or $E^-$, is different in $D^+$ and $D^-$ groups *(especially, in retrospective)*
  - non-differential misclassification
- nonrespondense & losses to follow-up
- analytic, due to "strong preconception" investigator

## Confounder(s) ($\neq$ *effect modification*)

# Issues in interpretation

Chance

Biases

- assessment of the outcome
- information, due to:
  - the extent and quality of obtaining $E^+$ or $E^-$, is different in $D^+$ and $D^-$ groups *(especially, in retrospective)*
  - non-differential misclassification
- nonrespondense & losses to follow-up
- analytic, due to "strong preconception" investigator

Confounder(s) ($\neq$ *effect modification*)

# Issues in interpretation

Chance

Biases

- assessment of the outcome
- information, due to:
  - the extent and quality of obtaining $E^+$ or $E^-$, is different in $D^+$ and $D^-$ groups *(especially, in retrospective)*
  - non-differential misclassification
- nonrespondense & losses to follow-up
- analytic, due to "strong preconception" investigator

Confounder(s) ($\neq$ *effect modification*)

# Issues in interpretation

Chance

Biases

- assessment of the outcome
- information, due to:
  - the extent and quality of obtaining $E^+$ or $E^-$, is different in $D^+$ and $D^-$ groups *(especially, in retrospective)*
  - non-differential misclassification
- nonrespondense & losses to follow-up
- analytic, due to "strong preconception" investigator

Confounder(s) ($\neq$ *effect modification*)

# Summary

- Cohort is the only design to establish incidence of disease in group(s)
- The $E \sim O$ temporal precedence is clear
- Issues of failure to follow-up

# For Further Reading

1. Stephen B. Hulley, MD, MPH; Steven R. Cummings, MD; Warren S. Browner, MD, MPH; Deborah Grady MD, MPH; Norman Hearst, MD, MPH; Thomas B. Newman, MD, MPH. *Designing Clinical Research - An Epidemiologic Approach,* Second Edition. Lippincott Williams & Wilkins, 2001.
2. Robert H. Fletcher, MD, MSc; Suzane W. Fletcher, MD, MSc. *Clinical Epidemiology - The Essentials,* Fourth Edition. Lippincott Williams & Wilkins, 2005.
3. Leon Gordis, MD, MPH, DrPH. *Epidemiology,* Fourth Edition. Saunders, Elsevier, 2009.

# Exercise

Do a brainstorm for topic(s) in your field, which can be designed as cohort studies.