

# Case-control studies

Dr. Lale Say  
WHO/RHR

Training Course in Sexual and Reproductive Health Research  
Geneva 2016

# Introduction

- ▶ Helps determining if an exposure is associated with an outcome (disease, condition of interest)
  - Cases (a group known to have the outcome)
  - Controls (a group known to be free of the outcome)
  - Comparison of the two groups in terms of frequency of exposure
- ▶ E.g., whether ovarian cancer (outcome) is related to oral contraceptive use (exposure)

# Introduction

A case control study looks backwards

**Comparison between  
study groups**

► Exposure  
(oral contraceptive)

20%  
40%



**Defines study group**

► Outcome  
(ovarian cancer)

Yes (case)  
No (control)

# Advantages

- ▶ Quick, inexpensive, and easy.
- ▶ Particularly appropriate for
  - investigating outbreaks – quick information is needed
  - studying rare diseases or outcomes – because starts with people known to have the outcome (rather than starting with a population free of disease and waiting to see who develops it) it is possible to enroll a sufficient number of patients with a rare disease.

# Disadvantages

- ▶ Cannot generate incidence data
  - We do not know the population at risk
- ▶ Subject to bias
- ▶ Difficult if record keeping is either inadequate or unreliable
  - If data collection relies on existing data sets
- ▶ Selection of controls can be difficult

# Caution

- ▶ All studies which contain 'cases' and 'controls' are not case-control studies.
  - One may start with a group of people with a known exposure and a comparison group ('control group') without the exposure and follow them through time to see what outcomes result, but this does not constitute a case-control study

# Cases

- ▶ The definition of a case needs to be very specific
  - Strict criteria
- ▶ Sources to select the cases
  - Hospital based
  - Population-based
    - ▶ Locating all cases (or a random sample) from a defined population

# Controls

- ▶ Controls should be chosen who are similar in many ways to the cases – most critical issue when designing a case-control study
  - There is a need to obtain comparable information from cases and controls
  - The selected control group must be **at similar risk** of developing the outcome
  - Any restrictions made in identification of the cases should apply to the controls



# Controls – sources to identify

- ▶ Hospital controls
  - Easy to find
  - More likely to recall earlier events
  - Might be similar if cases are also from hospitals
  - More likely to accept participation
- ▶ General population
  - Costly and time-consuming
  - Less likely to accept participation – might be different from those who participated
- ▶ Friends/relatives
  - Are similar to cases in important factors (ethnicity, social class)
  - Cases and controls might be similar in terms of risk factor in question (e.g., eating or smoking habits)
    - ▶ Could underestimate the effect of exposure in question on the outcome

# Data collection

- ▶ After clearly defining cases and controls, decide on data to be collected; the same data must be collected in the same way from both groups.
- ▶ Care must be taken to be objective in the search for past risk factors, especially since the outcome is already known.
- ▶ Sometimes it will be necessary to interview patients about potential factors (such as history of smoking, diet, use of medicines, sexual behaviour, etc.) in their past. It may be difficult for some people to recall/report all these details accurately.
- ▶ Participants who have the outcome (cases) are likely to scrutinize the past, remembering details of exposures more clearly than controls.

# Measure of effect

- ▶ Cases are selected according to disease status
  - Do not know the denominator/cannot calculate the incidence of the disease/outcome
  - We can think of cases and controls as proportions of all possible cases and controls – but do not know what that proportion is
  - The ODDS OF EXPOSURE among cases and controls is calculated
  - **ODDS RATIO** is used to estimate the effects of exposure on risk of disease

# Calculating the odds ratio

	Disease		Total
	Yes	No	
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	N

For the cases: the ODDS of EXPOSURE =  $a/c$

For the controls: the ODDS of EXPOSURE =  $b/d$

**ODDS RATIO** = ratio of odds of exposure =  $(a/c)/(b/d) = ad/bc$

# Example = is IUD use associated with PID?

Cases: women 18-44 yrs admitted for PID

controls: women 18-44 yrs without PID

exposure main contraceptive method used within 30 days of hospitalisation

	Disease		Total
	Yes	No	
Exposed	841	518	1359
Unexposed	724	967	1691
Total	1565	1485	3050

$$\text{ODDS RATIO} = ad/bc = 841 \times 967 / 724 \times 518 = 2.2$$

Interpretation: PID users were 2.2 times more likely to be current IUD users

# Odds ratio interpretation

- ▶ If odds ratio (and the the lower and upper bound of its 95% confidence interval)
  - Less than 1                      protective effect of the exposure
  - 1                                      no effect
  - Greater than 1                  harmful effect

# Sources of bias 1 – selection bias

- ▶ Ideally, control group has to be exactly the same as the group of cases – except for the presence of disease
  - Selection process should ensure that both cases and controls are likely to come from a similar population
- ▶ Selection bias occurs when the persons in one group are different on some factor (other than disease)
- ▶ E.g., women admitted to hospital with DVT (clinical diagnosis) are 6 times more likely to have used OC in the month prior to admission than those admitted for surgery
  - OC users are more likely to be referred to hospital with a diagnosis of DVT than non-OC users
    - ▶ However, the clinical diagnosis is not usually confirmed – e.g., only 16.7% of suspected cases are found to have confirmed DVT
    - ▶ The same study when considers confirmed cases finds an OR of 1.5 (95% CI 0.5 - 4.4)

# Sources of bias 2 – ascertainment bias

- ▶ Equal ascertainment of exposure in both the cases and the controls has to be ensured
- ▶ May arise because;
  - Cases may recall exposure better
  - Investigators may search for exposure more thoroughly in cases
  - Different data collection instrument may be used for the controls
- ▶ E.g., is OC use associated with endometrial cancer?
  - Cases: all women with endometrial cancer in region X
  - Controls: a random sample of resident women in region X age:20-60
  - Data collection: cases in hospital, controls telephone interview
  - Cases are more likely to report previous OC use



# Sources of bias 3 – confounding

- ▶ The two groups differ in some characteristic which is associated with both the outcome and exposure being studied
  - A confounding variable is one that can influence both the exposure and the outcome
  - E.g., in relation between OC use and myocardial infarction, cigarette smoking is a likely confounder
    - ▶ Women who use OC are more likely to smoke, and smoking is strongly associated with myocardial infarction. Age could be another confounder:
- ▶ E.g., is OC use associated with myocardial infarction?
  - Cases: all women with admitted to hospital with MI aged 20-49 in region X
  - Controls: a random sample of resident women in region X; age: 20-49 who have not had MI
  - Exposure: OC use during 3 months prior to interview
  - Data collection: personal interview of cases and controls – cases in hospital, controls telephone interview
  - Potential bias: cases are older than controls. Age is related to both exposure (OC use) and outcome (MI)

# Dealing with confounding

- ▶ Design of the study
  - Matching cases and controls on the relevant confounding variables
    - ▶ E.g., matching in age
- ▶ Analysis phase
  - Restrict the analysis to a limited age group – discards much of the collected information
  - Stratification /Multivariate analysis
- ▶ Confounding bias is the only type of bias that can be controlled in the analysis (neither selection, nor attrition bias can be handled at the analysis stage)

# Summary

- ▶ Case-control studies begin at the end – outcome is the basis to select the comparison groups
- ▶ They test hypotheses concerning the association and magnitudes of a relationship between outcomes and exposures
- ▶ The strength of association is measured by an odds ratio (OR) which is a good proxy for a relative risk when disease is rare
  - when the prevalence is  $<5\%$ , OR approximates RR
- ▶ Case-control studies are easy to do..... badly
  - Inherent problems with bias
- ▶ They are limited to evaluate an intervention