بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Epidemiological and Research Studies

Part 2 & 3

Prof DR. Waqar Al – Kubaisy

10-Decemb. -2023

Analytical studies

1. Cross-sectional
2. Case-control
3. Cohort

Case-control Study

Part 2

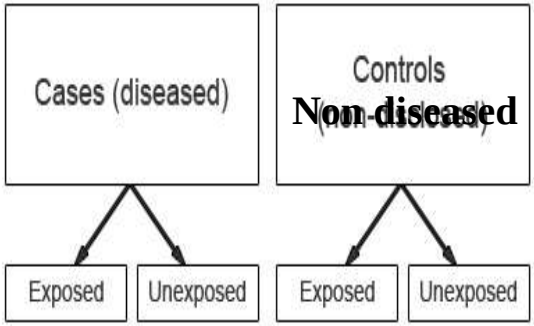
1. Basic concepts, application of CCS
2. Issues in the design CCS
3. Common sources of bias in a CCS
4. Analysis of CCS
5. Strengths and weaknesses of CCS

Analytical studies

case-control studies

basic concepts, application of CCS
Issues in the design CCS
Common sources of bias in a CCS
Analysis of CCS
Strengths and weaknesses of CCS

Case-control studies are one of the frequently used study designs **due to the relative ease of its** application in comparison with other study designs



Case-control Studies (CCS)
start with the identification of:

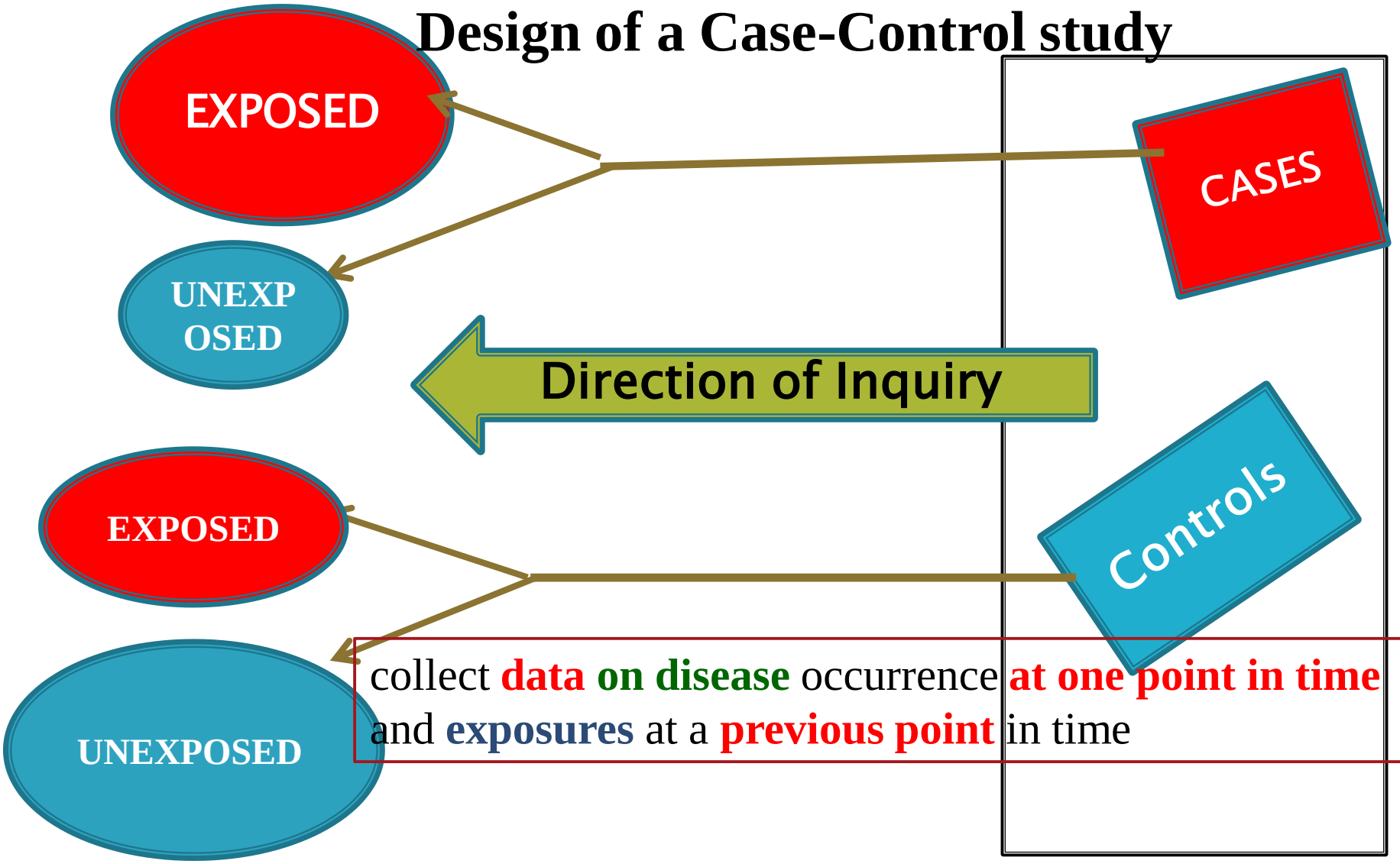
- a **group of cases** (individuals **with** a
- particular health outcome) in a given population
- **and a group of controls** (individuals **Without** the health outcome) to **be included** in the study.
- In **CCS** the **prevalence of exposure** to a potential **risk factor(s) is**

compared between cases and controls.

Cont. ..case-control studies

- If the prevalence of exposure is more common among cases than controls, it may be a risk factor for the outcome under investigation.
- The investigators collect data on disease occurrence
  - at one point in time and
  - exposures at a previous point in time.
- CCS provide a relatively simple way to
  - investigate causes of diseases, especially rare diseases
- A major characteristic of CCS is that data on
  - potential risk factors are collected retrospectively
    - and as a result may give rise to bias.
- This is a particular problem associated with CCS and therefore needs to be carefully considered during the design and conduct of the study.

Design of a Case-Control study



collect **data on disease** occurrence **at one point in time** and **exposures** at a **previous point** in time

Retrospective studies

investigator is looking backward to a possible cause.

Cont. ..case-control studies

basic concepts, application of CCS
Issues in the design CCS
Common sources of bias in a CCS
Analysis of CCS
Strengths and weaknesses of CCS

case-control studies

- **CCSs** are **longitudinal**, in contrast to cross-sectional studies.

- **CCSs** have been called **Retrospective studies since the investigator is looking backward from the disease to a possible cause.**

  - The **terms retrospective** and **prospective** are also **used to describe the timing of data collection in relation to the current date**.

Hepatitis C Virus Prevalence and Genotyping among Hepatocellular Carcinoma Patients in Baghdad

Waqar Abd Al Qahar Al-Kubaisy,Kadhim Jawad Obaid,Nor Aini Mohd Noor,Nik Shamsidah Binti Nik Ibrahim,Ahmed Albu-Kareem Al-Azawi

Hepatocellular carcinoma (HCC) is the third most common cause for cancer death in the world, now being especially linked to chronic hepatitis C virus (HCV) infection. This **case-control study consisting of 65 HCC patients and 82 patients with other malignant tumours as controls** was **conducted to determine the association of HCV markers with HCC**. Serum of each participant was obtained for detection of HCV Ab and RNA by DNA enzyme immunoassay (DEIA). Twenty six per cent (26.0%) of HCC patients had positive anti-HCV which was significantly greater than the control group (p=0.001). **HCC patients significantly have a risk** of **exposure to HCV infection almost 3 times** than the control group (**OR=2.87, 95**% C.I=1.1-7). Anti-HCV seropositive rate was significantly (p=0.03) higher among old age HCC patients and increases with age. Males with HCC significantly showed to have more than 9 times risk of exposure to HCV infection (OR=9.375, 95 % CI=1.299-67.647) than females. HCV-RNA seropositive rate was (70.8%) significantly higher among HCC patients compared to (22.2%) the control group (p=0.019). The most prevalent genotype (as a single or mixed pattern of infection) was HCV1b. This study detected a significantly higher HCV seropositive rate of antibodies and RNA in HCC patients.

The case group consisted of 65 patients, **histologically confirmed with HCC** and **a serum level of alpha-fetoprotein exceeding 400ng/ ml**, **while 82 patients with other malignant tumours (not related to gastro intestinal system)** were considered **as a control group**. **Four hospitals** namely Baghdad Teaching Hospital-Baghdad Medical City, Al Kadhmiya Teaching Hospital, Radiology and Nuclear Medicine Institute and Al Yarmuk General Teaching Hospital **were chosen for** data collection. **Only respondents with informed consent were interviewed** using a structured questionnaire and serum samples were taken for HCV markers analysis. **Patients with positive serum HBV were excluded** from this study.

case-control studies

1. Issues in the design of case-control studies

- ❑ **Formulation of a clearly defined hypothesis**
- ❖ **Selection of cases**
- ❖ **Selection of controls**
- ❖ **Measuring exposure status**

basic concepts, application &of CCS
Issues in the design CCS
Common sources of bias in a CCS
Analysis of CCS
Strengths and weaknesses of CCS

❑ **Formulation of a clearly defined hypothesis**

the **beginning** of a **CCS should begin** with the formulation of **a clearly defined hypothesis.**

❖ **Case definition** It is essential that the **case definition is clearly defined** at the onset of the investigation **to ensure that all cases included in the study are based on the same diagnostic criteria**

- **Source of cases** The **source of** cases needs to be clearly defined.

Cases may be recruited from

- **Cases may be recruited from a number of sources**; they may be recruited from a **hospital, clinic**, or may be **population bases**.

Population based case control studies are generally **more expensive and difficult to conduct**

Selection of cases

- CCS may use **incident** or **prevalent** cases.
- **Incident cases**: comprise **cases newly diagnosed during a defined time period**.
- **The use of incident** cases is considered **as preferential**, as the **recall of past exposure(s) may be more accurate** among newly diagnosed cases. **In addition,**
- **the temporal sequence of exposure and disease is easier to assess among incident cases**

Prevalent cases

comprise **individuals who have had the outcome** under investigation **for some time**.

The use of **prevalent cases** may **give rise to recall bias as** prevalent cases may **be less likely to accurately report past exposures(s).**

As a result, the **interpretation of results** based on prevalent cases may **show more problem** as it may be more difficult to ensure that reported events **relate to a time before** the development of disease rather **than to the consequence of the disease process itself**. For example, **individuals may modify their exposure following the onset of disease**.

Selection of controls

- **A particular problem inherent in CCS is the selection of a comparable control group.**
- **Controls are used to estimate the prevalence of exposure in the population which gave rise to the cases.**

Therefore, the ideal control

Formulation of a clearly defined hypothesis
Selection of cases
Selection of controls
Measuring exposure status

- **Therefore, the ideal control group would comprise**
- **a random sample from the general population that gave rise to the cases.**

❖ However, this is not always possible in practice.

- **The goal is to select individuals** in whom the **distribution of exposure** status would **be the same as that of the cases** in the absence of an exposure disease association. That is, **if there is no true association between exposure and disease**, the cases and controls should have the same distribution of exposure.

❑ **The source of controls is dependent on the source of cases. In order to minimize bias, controls** should be selected **to be a representative sample of the population which produced the cases.**

case-control studies

Formulation of a clearly defined hypothesis
Selection of cases
Selection of controls
Measuring exposure status

- In **CCS where cases are hospital based**, it is common to recruit **controls from the hospital population**.
- **However**, the choice of controls from a hospital setting **should not include individuals with an outcome related to the exposure(s) being studied**.

For example, in a case-control study of the association between smoking and lung cancer the inclusion of controls being treated for a condition related to smoking (e.g. chronic bronchitis) may result in an underestimate of the strength of the association between exposure (smoking) and outcome.

- **Recruiting more than one control per case**
- **may improve the statistical power of the study**,
- **though** including more than 4 controls per case is generally considered to be no more efficient. **Also,**

- **the exposures** of controls should be **measurable with similar accuracy** to those of the cases

❑ Measuring exposure status

- ❑ Exposure status is measured **to assess** the
- ❑ **presence** and/or **level of exposure** for each individual
- ❑ **for the period of time prior to the onset of the disease** or condition under investigation **when the exposure**
- ❑ would have **acted as a causal factor**.
  - ➢ Note **that in CCS** the **measurement** of **exposure is**
  - ➢ **established after the development of disease** and as a result **is prone to both recall and observer bias.**
- ❑ **Various methods can be used to ascertain exposure status. These include:**

Standardized questionnaires
Biological samples
Interviews with the subject
Interviews with spouse or other family members
Medical records
Employment records
Pharmacy records

The procedures used for the collection of exposure data should be the same for cases and controls.

2. Common sources of bias in CCS

- Due to the **retrospective nature** of CCS, they **are particularly susceptible to the effects of bias, which may be introduced as a result of:**

❖ **poor study design** or

❖ **during the collection of exposure and outcome data.**

- Because the disease and exposure have already occurred at the onset of a **CCS**, **there may be differential reporting of exposure information between cases and controls based on their disease status.** Eg **cases and controls may recall past exposure differently (recall bias).**

- Similarly, **the recording of exposure information may vary depending on the investigator's knowledge of an individual's disease status (interviewer/observer bias).**

Therefore, the **design & conduct**

- **Therefore,** the **design and conduct of the study must be**
- **carefully considered,** as there are limited options for the control of bias during the analysis.
- **Selection bias in CCS Selection bias is** a particular problem inherent in case-control studies, where **it**
- **gives rise to non-comparability** between cases and controls.
- **Selection bias in CCS may occur when:** '

➢ **cases or controls** are **included in or excluded from** a study because of **some characteristic they exhibit which is related to exposure to the risk factor under evaluation**' .

- **Selection bias** may occur when those individuals selected as **controls are unrepresentative** of the population that produced the cases.

Cont.....Common sources of bias in CCS

The aim of a case-control study is to **select study controls** who are representative of the population which produced the cases. Controls are used to provide an estimate of the exposure rate in the population.

The potential for **selection bias** in CCS is a **particular problem** when **cases and controls** are **recruited exclusively from hospital or clinics.**
Then, estimates of the exposure among controls may be different from that in the reference population, which may result in a biased estimate of the association between exposure and disease.

- As the potential for **selection bias** is likely to **be less of a problem in population based CCS**
- **Neighborhood** controls may be a preferable choice when using cases from a hospital or clinic setting.
- **Alternatively**, the potential for **selection bias may be minimized** by
- Selecting controls from **more than one source**, such as by using both hospital and neighborhood controls.

➢ **Selection bias** may also be introduced in CCS **when exposed cases are more likely to be selected than unexposed cases.**

3. Analysis of case-control studies

RISK ESTIMATES (Odds ratio)

- **Odds ratio (OR)**
- **Used in cross sectional, case-control**

Results of a case-control study can be presented in a **2x2 table** as follow

	Case (diseases)	control	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	N

- **OR= a/(a+c) ÷ b/(b+d)= a/c ÷ b/d =ad/bc**
- **c/(a+c) d/(b+d)**

Example:

To study to identify whether **smoking is a risk factor for cancer (CA) of the pancreas.** A case-control study **was conducted** among **100 cases of CA pancreas** and **400 free of CA pancreas**. they found that the number of smokers were 60 and 100 individuals, among the CA pancreas and no CA pancreas groups respectively

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

	Cases	Controls	Total
Smokers	60 (a)	100 (b)	160
Non-smokers	40 (d)	300 (d)	340
Total	100	400	500

$$OR = \frac{60 \times 300}{100 \times 40} \qquad OR = 4.5$$

The OR estimates that **cancer of the pancreas occurrence is 4.5 times more among smokers than non-smokers.**

NB: The odds ratio of smoking and cancer of the pancreas has been performed without

adjusting for potential confounders.

the confounding variable(s) to be calculated

and a statistical test of significance can also be calculated.

In addition, **confidence intervals for the odds ratio would also be presented. (2.84-7.13)**

Cont. „Analysis of case-control studies

- **The odds ratio (OR) is used in CCS** to **estimate the strength of the association between exposure & outcome**.

it is **not possible to estimate the incidence** of disease from a CCS :

The **OR is a measure of the odds of exposed in the disease compared to the odds of exposed in the controls** and is calculated as:

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

which is the ratio of the odds of exposure among the cases to the odds of exposure among the controls.

2x2 table

	Cases	Controls	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

4. Strengths and weaknesses of CCS

Strengths

- ✓ **Cost effective** relative to other analytical studies such as cohort studies.
- ✓ CCS are retrospective, and cases are identified at the beginning of the study; therefore **there is no long follow up** period (as compared to cohort studies)
- ✓ **Efficient** for the study of **diseases with long latency periods.**
- ✓ **Efficient** for the study of **rare diseases**
- ✓ Good for examining **multiple exposures.**

Weaknesses

- ➢ Particularly **prone to bias** especially **recall ,selection**, and **observer bias**.
- ➢ CCS limited to **examining one outcome.**
- ➢ **Unable to estimate incidence rates** of disease
- ➢ **Poor** choice for the study of **rare exposures.**
- ➢ **The temporal sequence** between exposure and disease may be **difficult to determine.**

Epidemiological and Research Studies

Part 3

Prof DR. Waqar Al – Kubaisy

Analytical studies
~~**Cross-sectional**~~
~~**Case-control**~~
Cohort

begins with group of people free of disease & classified into subgroups
a group of individuals **exposed to a risk factor**
a group who are **unexposed to the risk** factor
are followed over time (often years)

Cohort Study

Issues in the design of cohort studies understand the differences from a CCS,
*Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (RR,AR
*appreciate its strengths and weaknesses.

Cohort Study

Analytical studies
~~Cross-sectional~~
~~Case-control~~
Cohort

Issues in the design of cohort studies
understand the differences from a CCS
*Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (RR,AR
*appreciate its strengths and weakness

Also called:follow up study, or incidence studies,

Definition:

Study in which persons,

- **based on their exposure to a determinant** and
- **free of the disease outcome at the start** of the study
- **are followed in time** to **assess the occurrence of the disease outcome**

- It begins with a group of people who **are free of disease**
- and who are **classified into subgroups according to exposure** to a **potential cause of disease or outcome**
- **Cases** are excluded at the beginning
- **Variables of interest are specified and measured** and the
- **whole cohort is followed up to see how the subsequent**
- **development of new cases of the disease** (or other outcome) **differs** between the groups **with** and **without exposure**

Cohort studies

*Issues in the design of cohort studies understand the differences from a CCS, *Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (relative risk, attributable risk
*appreciate its strengths &weaknesses.

Cohort studies are a **form of longitudinal study** design that **flows from** the **exposure** **to** **outcome**.

In a cohort study,

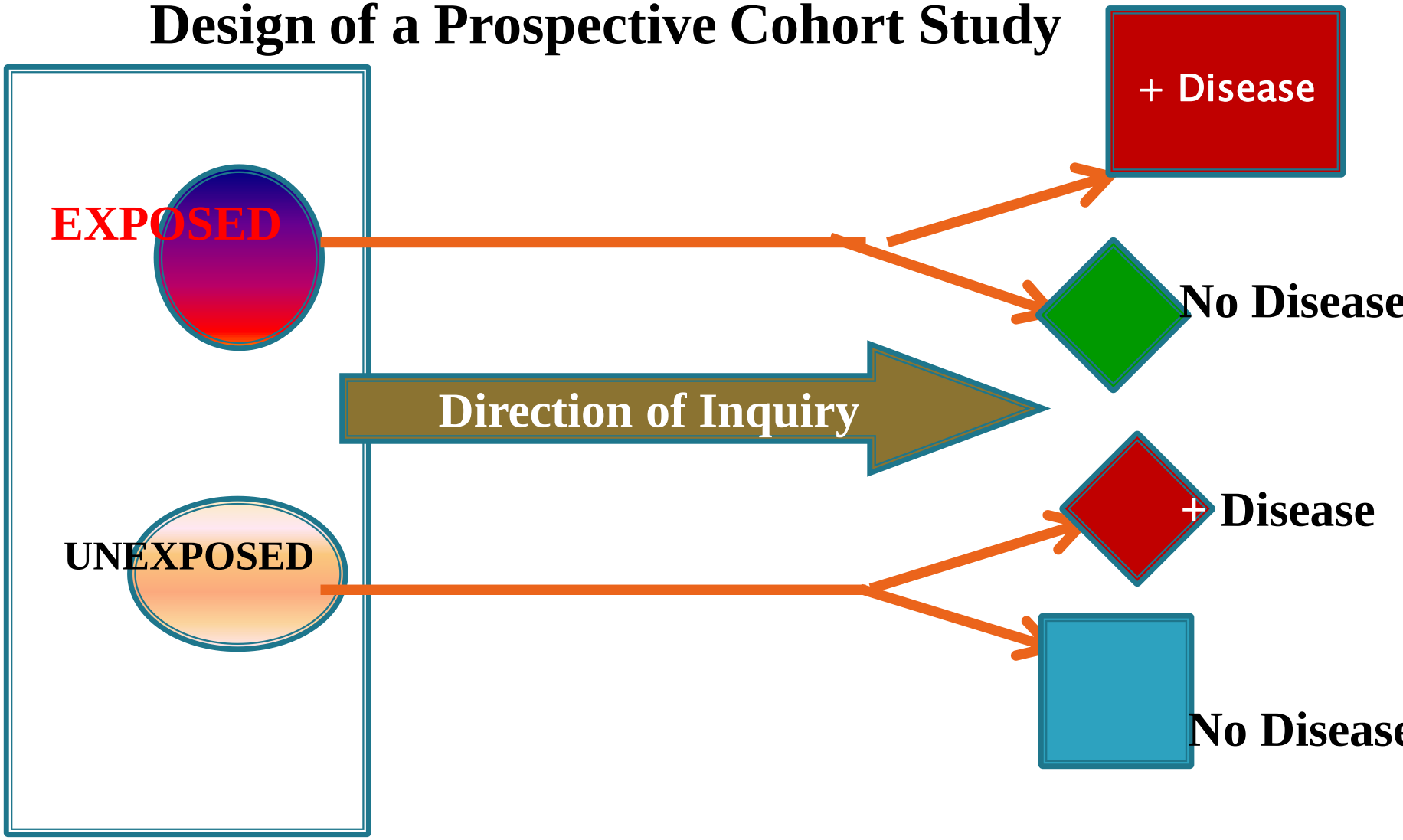
- **Group** of individuals **exposed to a putative risk factor**
- **Group** who are **unexposed to the risk** factor
- **are followed over time** (often years)
- **to determine the occurrence of disease**.
- The **incidence of disease**
- in the **exposed group** is **compared with** the
- **incidence** of disease in the **unexposed group**.

Therefore **Relative risk (incidence risk or incidence rate)** is used

to assess whether the exposure and disease are causally linked.

Cohort studies be **prospective**

Design of a Prospective Cohort Study



It begins with group of people **free of disease** and **classified into subgroups**
a group of individuals **exposed to a risk factor**
a group who are **unexposed to the risk** factor
are followed over time (often years)

*Issues in the design of cohort studies understand the differences from a CCS, *Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (relative risk, attributable risk etc
*appreciate its strengths &weaknesses.

- **Cohort studies** be **prospective**
- **A prospective cohort** study is also called a
- **concurrent cohort study**,

where the **subjects have been followed up for a period and the outcomes of interest are recorded.**

1. Issues in the design of cohort studies:

- **Selection of study groups**
- **Measuring exposure**
- **Measuring outcome**
- **Methods of follow-up**

*Issues in the design of cohort studies understand the differences from a CCS, *Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (relative risk, attributable risk etc
*appreciate its strengths &weaknesses.

Cont…..cohort studies

Issues in the design of cohort studies understand the differences from a CCS,
Potential bias in cohort studies
Analysis of cohort studies
calculate the basic measures (relative risk, attributable risk etc
appreciate its strengths and weaknesses.

CONT. ..Issues in the design of cohort studies

Selection of study groups

- The aim of a cohort study is to **select study participants who are identical with the exception of their exposure status**.

- **All study participants must be**
- **Free of the outcome under investigation and**
- **have the potential to develop the outcome** under investigation.

❑ Measuring exposure

❖ **Levels of exposure** (e.g. packs of cigarettes smoked per year)are

- **measured for each individual at baseline at the**
- **beginning of** the study **and**
- **Assessed at intervals during the period of follow-up**.

❑ **When several** exposures are being considered simultaneously, the non-exposed group should comprise all those with none of the risk factors under investigation.

❖ **A particular problem** occurring in cohort studies is **whether individuals in the control group are truly unexposed. For example**, study participants may start smoking or they may fail to correctly recall past exposure.

❖ **Similarly**, those in the **exposed group may change** their **behavior in relation to** the exposure such as diet, smoking or alcohol consumption.

Exposure data may be obtained

Issues in the design of cohort Selection of study groups
Measuring exposure
Measuring outcome
Methods of follow-up

- **Exposure data may be obtained** from a number of sources including:
- **Medical** or **Employment records**,
- **standardized** questionnaires, interviews and by
  - **physical examination**

Issues in the design of cohort Selection of study groups
Measuring exposure
Measuring outcome
Methods of follow-up

Measuring outcome

- **Outcome measures may be obtained from**
  - **various sources, including**
  - **Directly from the participant**
  - **Routine surveillance of cancer** registry data,
  - **Death Certificates**,
  - **Medical records**
  - **Method used to ascertain outcome**
    - **must be identical for both exposed and unexposed groups**

Cont…..cohort studies

❑ Methods of follow-up

The follow-up of study participants in a cohort study is a major challenge.

- **A great deal of cost and time is required to ensure follow-up** of cohort members
- and to **update measures of exposures and confounders**,
- in addition to **monitoring participants' health outcomes**. .

The failure to collect outcome data for all members of the cohort will affect the validity of study results

2. Potential sources of bias Cont…..cohort studies

1 Issues in the design of cohort studies understand the differences from a CCS,
2 Potential bias in cohort studies
3 Analysis of cohort studies
4 calculate the basic measures (relative risk, attributable risk etc
5 appreciate its strengths and weaknesses.

A major source of potential bias in cohort studies is due to:

- **losses to follow-up.**
- **Cohort members may; die, Migrate, Change jobs or**
- **Refuse to continue to participate** in the study.

In addition, losses to follow-up **may be related to the**

- **exposure, outcome or both**. For example, individuals who develop the outcome may be less likely to continue to participate in the study.
- The degree to which losses to follow-up are correlated with exposure and outcome will lead to serious bias in the measures of effect of exposure and outcome

- **A major source of potential bias in cohort studies arises from**
- **The degree of accuracy with which subjects have been classified with respect to their exposure or disease status.**
- **Differential misclassification** can lead **to an over or underestimate** of the effect between **exposure** and **outcome**

1 Issues in the design of cohort studies understand the differences from a CCS,
2 Potential bias in cohort studies
3 Analysis of cohort studies
4 calculate the basic measures (relative risk, attributable risk etc
5 appreciate its strengths and weaknesses.

Analysis of cohort studies

- Analysis of a cohort study uses either
- the risk or the rate ratio of disease in the exposed cohort
- compared with the rate or risk in the unexposed cohort

Risk estimates:

To estimate risk of event to occur when exposed to a risk factor.

Relative risk (RR)

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

Exposure	+Ve Out com	no outcome	Total
positive	a	b	
Negative	c	d	
Total			

RR Used in cohort study

The risk is the relative incidence in the exposed and non exposed group

Example

Risk estimates

RR = proportion of disease in exposed group / proportion of disease in unexposed group

Example: A cohort study of smoking and cancer of the pancreas was conducted for 90,049 individuals and followed for 1 year. Consist of 27,042 smokers of them 42 were developed CA pancreas, while only 7 developed the CA pancreas among non smokers group . Is smoking is a risk factor for CA pancreas

Exposure	CA pancreas	no CA pancreas	Total
positive	a 42	b 27,000	27,042
Negative	C 7	d 63,000	63,007
Total			90,049

Analysis of cohort studies

Example: A cohort study of smoking and cancer of the pancreas was conducted for 90,049 individuals and followed for 1 year. Consist of 27,042 smokers of them 42 were developed CA pancreas, while only 7 developed the CA pancreas among non smokers group . Is smoking is a risk factor for CA pancreas

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

the data, taken from a hypothetical **cohort** study **to investigate the association between smoking and cancer of the pancreas,** the **relative and attributable risk** can be calculated as follows:

Example

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

$$\text{Rate Ratio} = \frac{\text{Incidence rate in exposed group (r1)}}{\text{Incidence rate in unexposed group (r0)}}$$

RR = 1.5/0.1 = 15

The RR of 15 indicates that
the risk of cancer of the pancreas is 15 times higher among smokers than non-smokers.

Attributable risk can be calculated ???

Population attributable Risk PAR

Cont…..cohort studies

1 Issues in the design of cohort studies understand the differences from a CCS,
2 Potential bias in cohort studies
3 Analysis of cohort studies
4 calculate the basic measures (relative risk, attributable risk etc
5 appreciate its strengths &weaknesses.

4. Strengths and weaknesses of cohort studies

Weaknesses

- **Costly and time consuming.**
- **Prone to bias due to loss to follow-up.**
- **Prone to confounding.**
- **Participants may move between one exposure category**
- **Knowledge of exposure status may bias classification of the outcome.**
- **Being in the study may alter participant's behaviour.**
- **Poor choice for the study of a rare disease.**
- **Classification of individuals (exposure or outcome status) can be affected by changes in diagnostic procedures.**

Strengths

- ✓ **Multiple outcomes** can be measured for any one exposure.
- ✓ **Can look at multiple exposures.**
- ✓ **Exposure is measured before the onset of disease**
- ✓ **Good for measuring rare exposures,** for example among different occupations.
- ✓ **Demonstrate direction of causality.**
- ✓ **Can measure incidence**

Thank you for attention

Qs ????