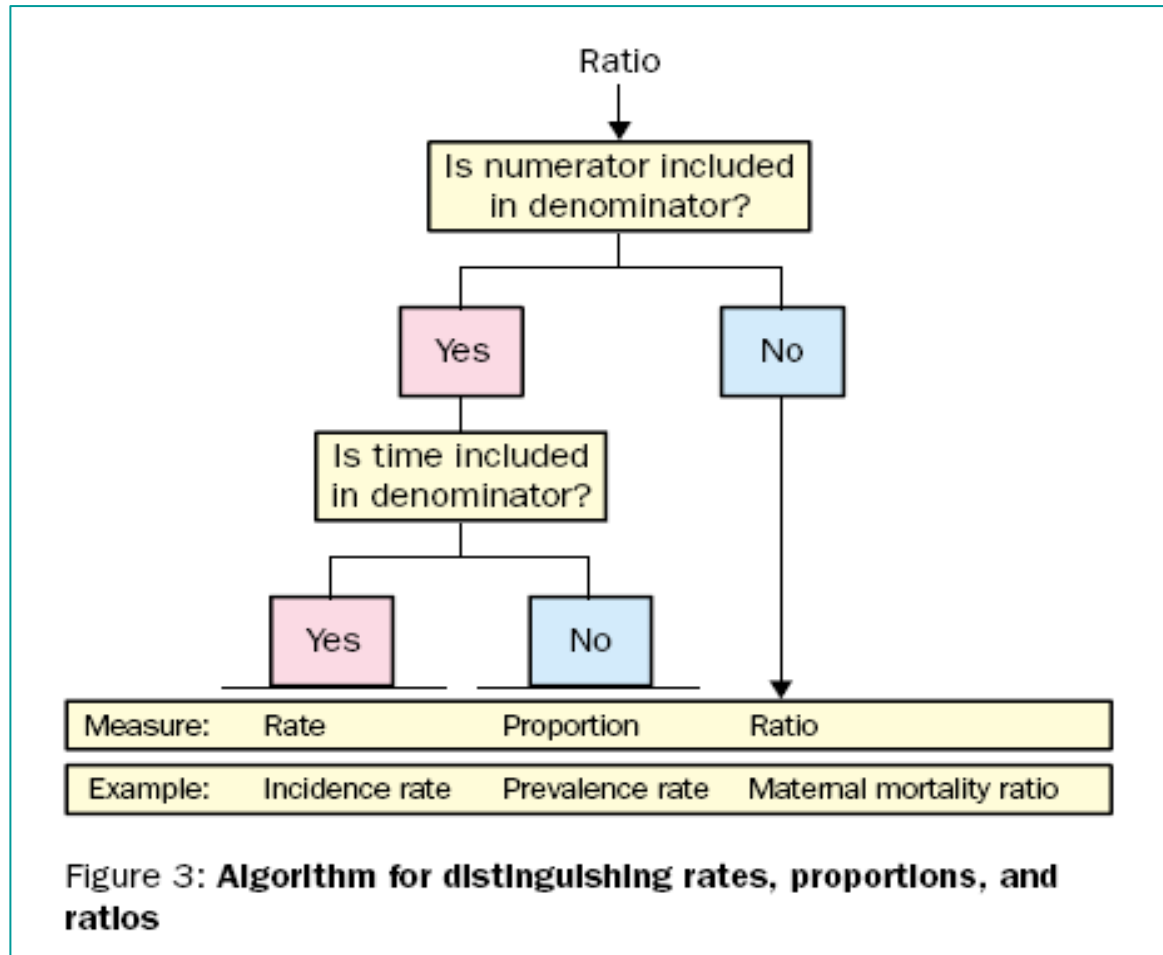


Introduction to Study Design

Thomas P. Erlinger, MD, MPH
Director, Clinical Research
Seton Family of Hospitals

Epidemiologic Study Designs

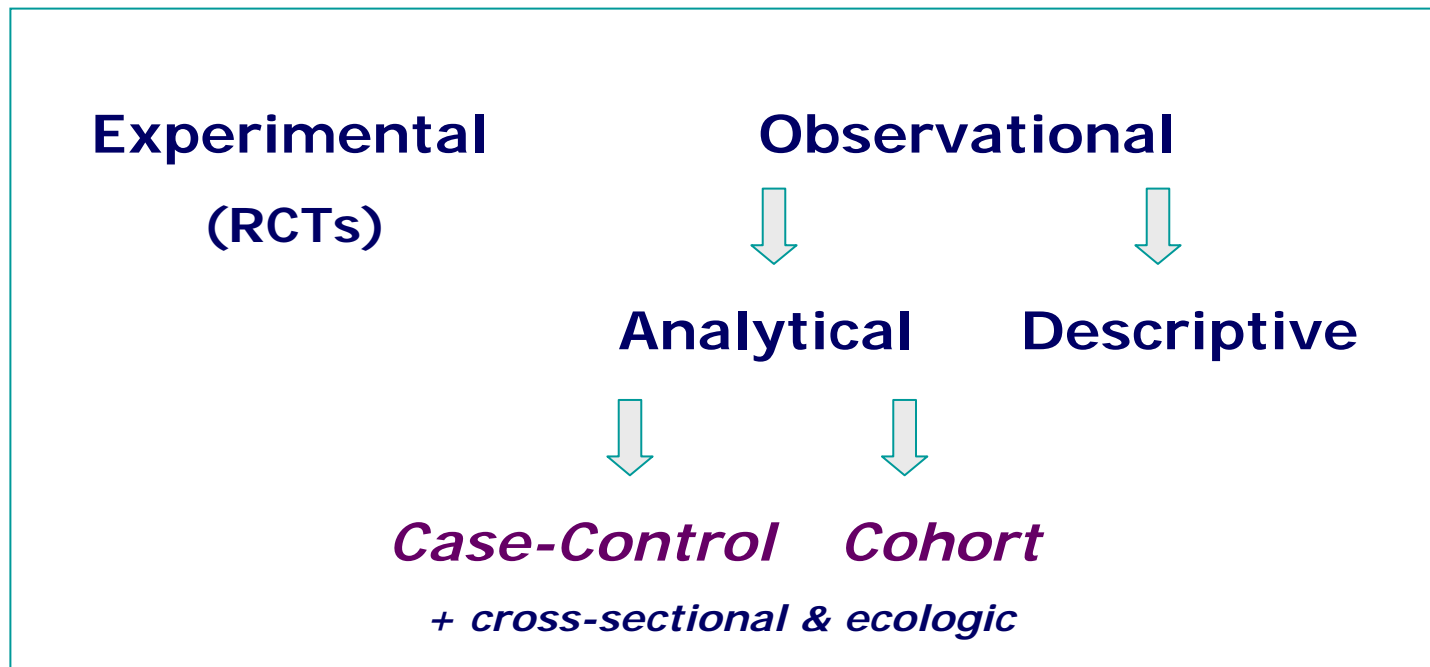


Types of Study Designs

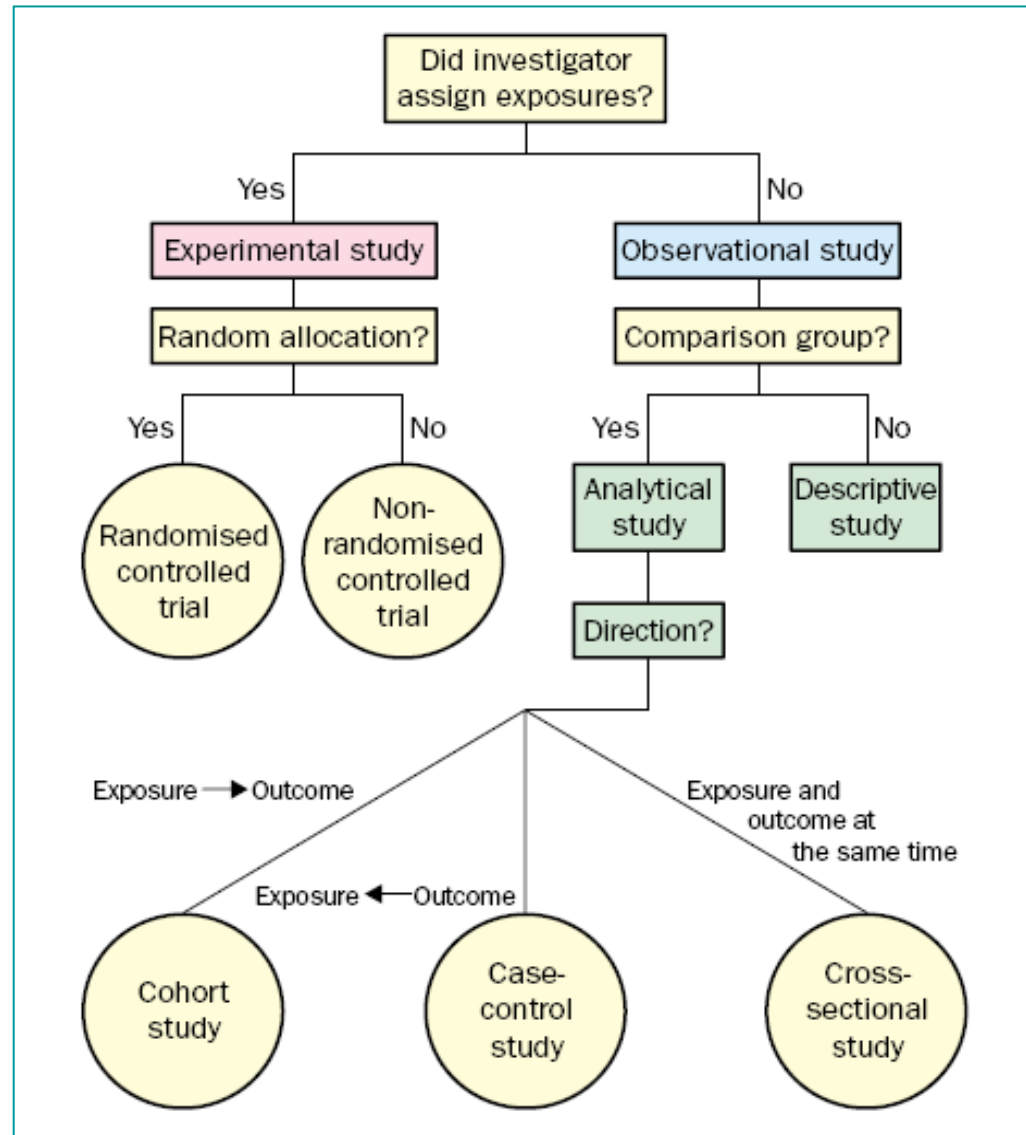
- Cross-sectional
- Case-Control
- Cohort
- Randomized Controlled Trials



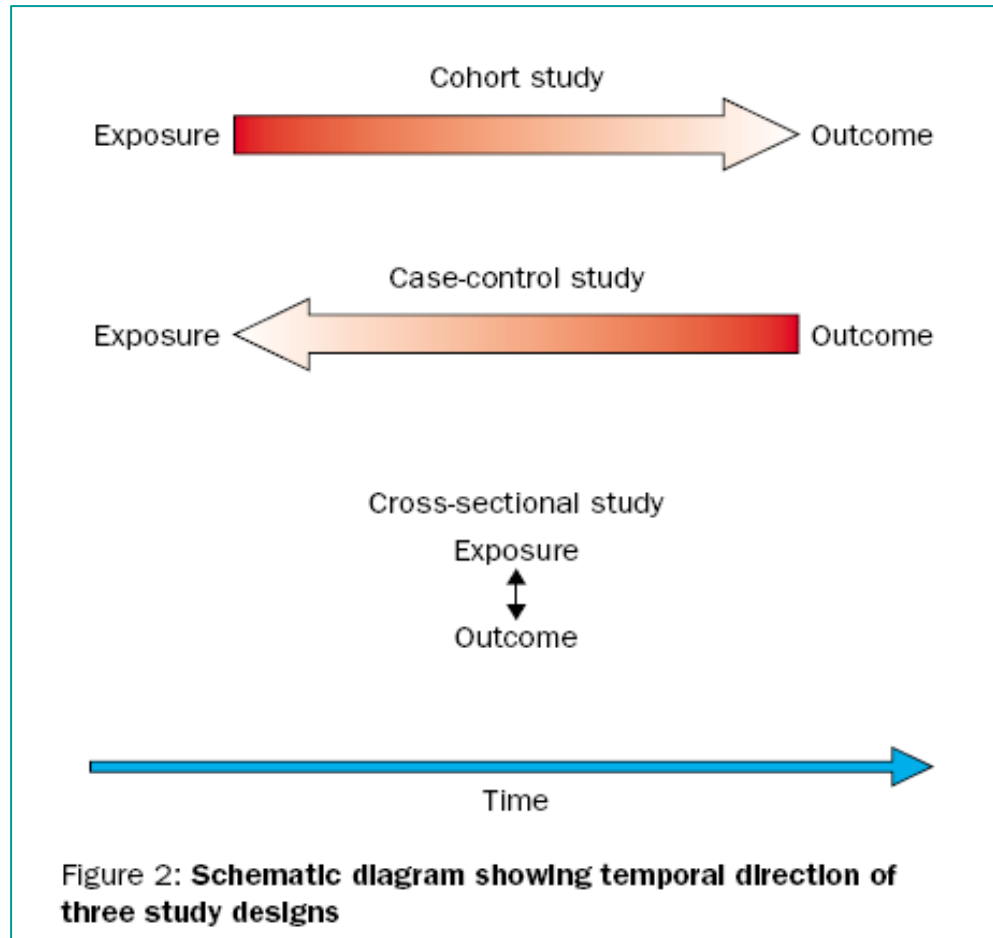
Epidemiologic Study Designs



Epidemiologic Study Designs



Epidemiologic Study Designs



Question

Sir Norman Gregg, an Australian Ophthalmologist, observed a number of infants and young children in his practice that had an unusual form of cataract (eye disease).

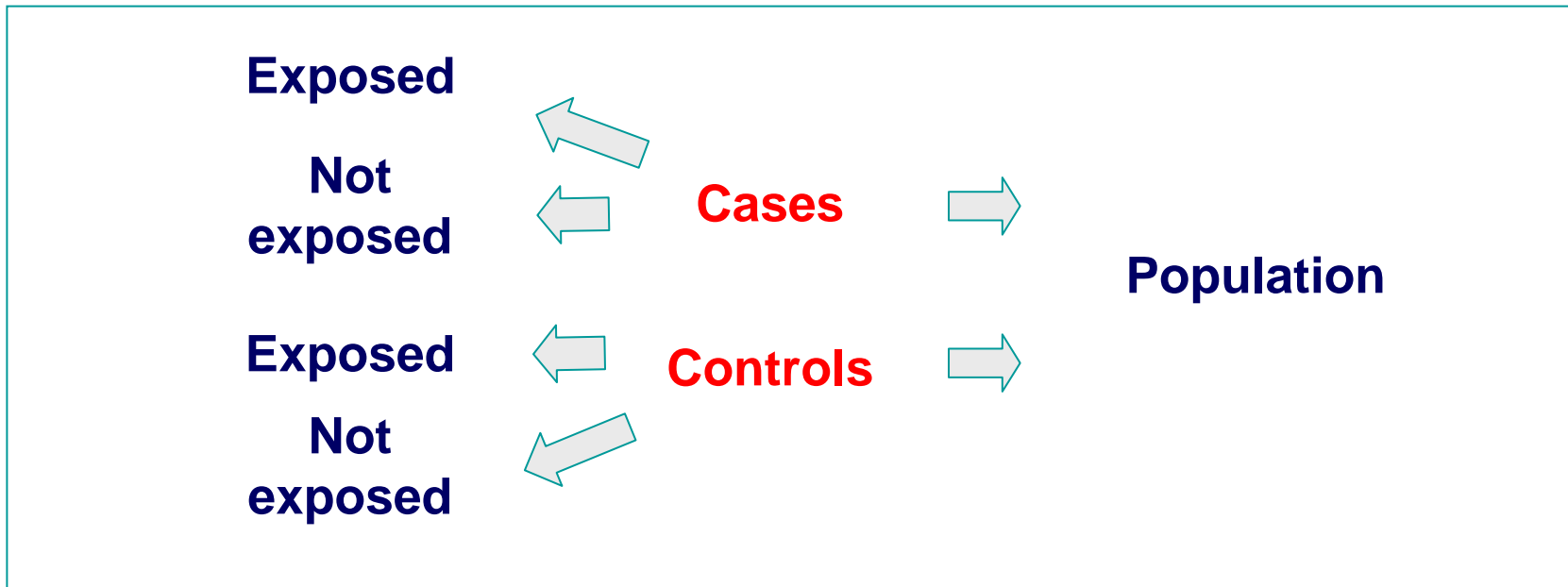
Observed: children in utero during rubella outbreak.

Hypothesis: Prenatal rubella exposure leads to the development of cataracts.

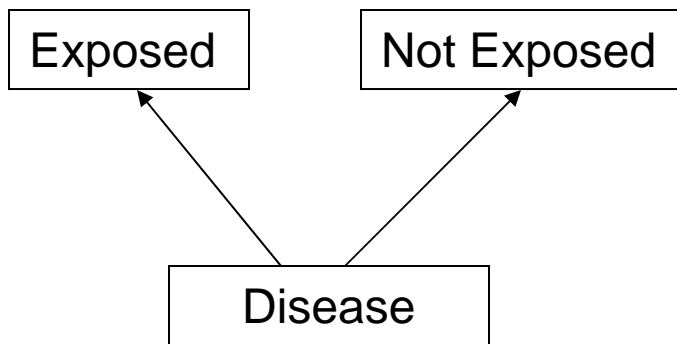
What happened?

- Gregg observed that 90% of infants had been in utero during the rubella outbreak.
- Is this sufficient to justify the conclusion that rubella exposure caused cataracts?
- What is needed?
 - A control group
- Why?
 - Could be that 90% of ALL mothers had been pregnant during the outbreak. If that were true, the exposure history would not be relevant.

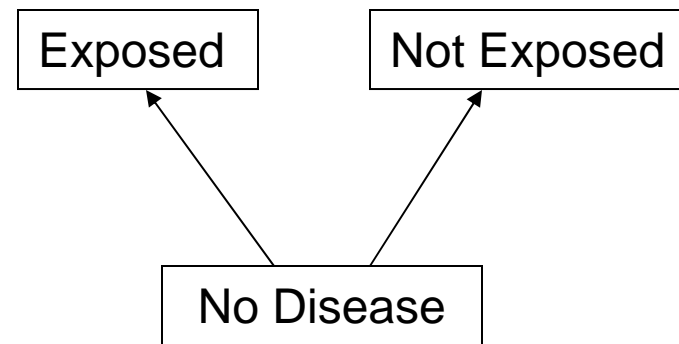
Case-Control Studies



Design of a Case-Control Study



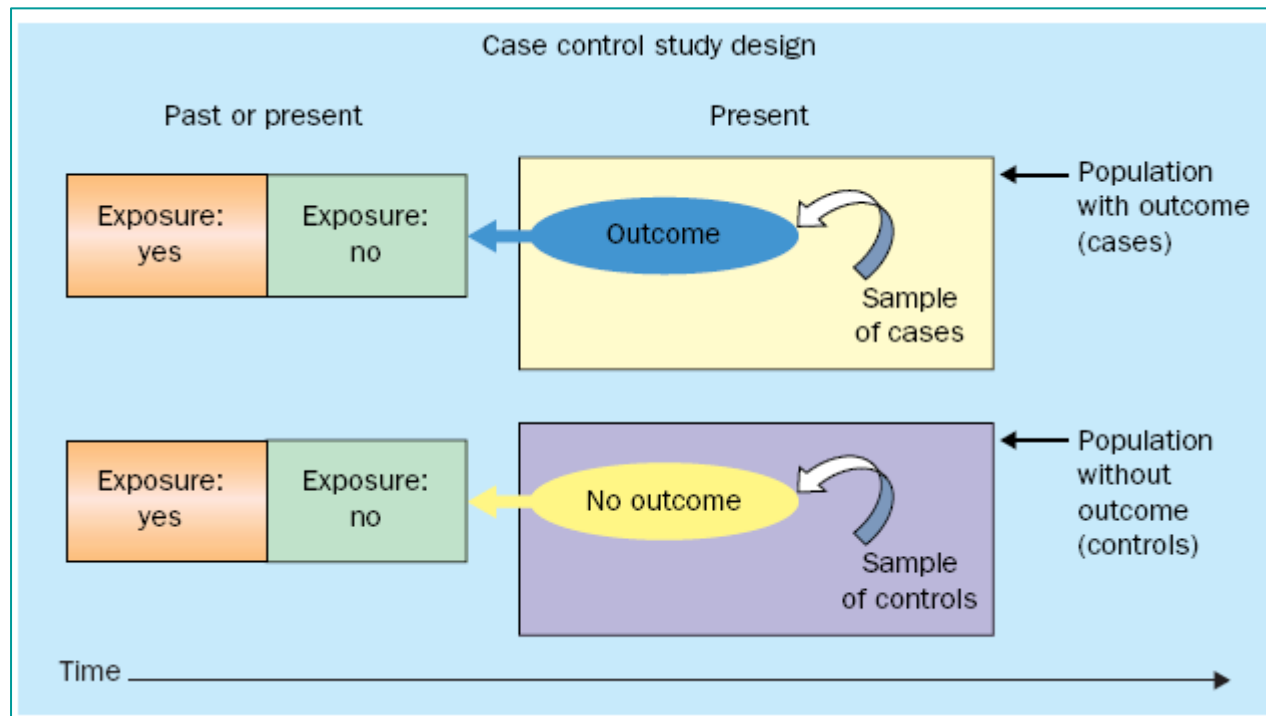
“CASES”



“CONTROLS”

Case control studies START with persons who already have the disease.

Case-Control Studies



Schulz & Grimes, 2002 ([www](#)) ([PDF](#))

Case-Control Study

First, Select:

		Cases (with Dz)	Controls (without Dz)
<u>Then</u> , measure past Exposure	Were Exposed	a	b
	Were NOT Exposed	c	d
	Total	a + c	b + d
	Proportions Exposed	$\frac{a}{a + c}$	$\frac{b}{b + d}$

Case-Control Study

- Thus, in a case-control study, if there is an association of an exposure with a disease, the prevalence of history of exposure should be higher in persons who have the disease (cases) than in those who do not (controls).

Case-Control Study

	CHD	Controls
Smoke	112	176
Do not smoke	88	224
Total	200	400
% Smoking Cigarettes	56.0	44.0

Notes on Case-Control Studies

- IF we use only data from a case-control study, we cannot estimate the prevalence of the disease.
 - The number of controls selected is up to the investigator, and is arbitrary.
 - Could choose case 1 control per case, 2 controls per case, 4 controls per case, etc.

Advantages of Case-Control Studies

- Cheap, easy and quick studies
- Multiple exposures can be examined
- Rare diseases and diseases with long latency can be studied
- Suitable when randomization is unethical (alcohol and pregnancy outcome)

Disadvantages of Case-Control Studies

- Case and control selection troublesome
- Subject to bias (selection, recall, misclassification)
- Direct incidence estimation is not possible
- Temporal relationship is not clear
- Multiple outcomes cannot be studied
- If the incidence of exposure is high, it is difficult to show the difference between cases and controls
- Not easy to estimate attributable fraction
- Reverse causation is a problem in interpretation - especially in molecular epidemiology studies

Case-Control Studies: Potential Bias

Panel 2: Introduction of bias through poor choice of controls

Cases	Control selection	Non-representativeness	Selection bias
Colorectal cancer patients admitted to hospital	Patients admitted to hospital with arthritis	Controls probably have high degrees of exposure to NSAIDs	Would spuriously reduce the estimate of effect (odds ratio)
Colorectal cancer patients admitted to hospital	Patients admitted to hospital with peptic ulcers	Controls probably have low degrees of exposure to NSAIDs	Would spuriously increase the estimate of effect (odds ratio)

NSAIDs=non-steroidal anti-inflammatory drugs.

Cause-and-Effect Relationship

Temporal sequence

Did exposure precede outcome?

Strength of association

How strong is the effect, measured as relative risk or odds ratio?

Consistency of association

Has effect been seen by others?

Biological gradient (dose-response relation)

Does increased exposure result in more of the outcome?

Specificity of association

Does exposure lead only to outcome?

Biological plausibility

Does the association make sense?

Coherence with existing knowledge

Is the association consistent with available evidence?

Experimental evidence

Has a randomised controlled trial been done?

Analogy

Is the association similar to others?

Cause-and-Effect Relationship

Panel 1: What to look for in observational studies

Is selection bias present?

In a cohort study, are participants in the exposed and unexposed groups similar in all important respects except for the exposure?

In a case-control study, are cases and controls similar in all important respects except for the disease in question?

Is information bias present?

In a cohort study, is information about outcome obtained in the same way for those exposed and unexposed?

In a case-control study, is information about exposure gathered in the same way for cases and controls?

Is confounding present?

Could the results be accounted for by the presence of a factor—eg, age, smoking, sexual behaviour, diet—associated with both the exposure and the outcome but not directly involved in the causal pathway?

If the results cannot be explained by these three biases, could they be the result of chance?

What are the relative risk or odds ratio and 95% CI?^{11,12}

Is the difference statistically significant, and, if not, did the study have adequate power to find a clinically important difference?^{13,14}

If the results still cannot be explained away, then (and only then) might the findings be real and worthy of note.

Selection of Cases

****need to specify selection criteria****

- Cases should be “representative”, i.e. registries vs. hospitals
 - Cases selected from hospitals may track with other features associated with admission to that hospital
 - Disease registries are valuable for non-biased case selection.

Selection of Cases

- Incident vs. Prevalent Cases
 - In studies of etiology, preferable to use incident cases.
 - Why?
 - Factors associated with prevalent disease may be associated with survival rather than to the development (incidence) of the disease.
 - E.g. cases may die soon after disease onset and thus, not be represented in a study of prevalent cases.

Selection of Controls (An example)

In 1929, Dr. Raymond Pearl of JHU conducted a case-control study to test the hypothesis that tuberculosis protected against cancer. The following are his data.

	Cases (with cancer)	Controls (without cancer)
Total no. of autopsies	816	816
No. (%) with tuberculosis	54 (6.6)	133 (16.3)

Selection of Cases

(An example, cont.)

- Pearl concluded that TB did, in fact, protect against cancer.
- Was he correct?
 - Depends on the validity of the control selection.
 - If the prevalence of TB in the non-cancer patients (controls), was similar to that in the general population, his conclusion would be valid.
 - However, in 1929, TB was a major reason for admission to the hospital.
 - Therefore, he inadvertently selected a group of controls with a higher than normal prevalence of TB.

What are we talking about?

- Selection bias
 - A systematic under or over – representation of the exposure of interest.
 - Could impact cases or controls
 - E.g. of faulty control selection
 - Lung cancer / smoking ----- controls were persons with emphysema rather than lung cancer. Both groups have higher prevalence of smoking than general population, so would obscure associations
 - Pancreatic cancer / coffee ----- controls were persons without PC seeing the same GASTROENTEROLOGIST. Coffee often stopped for other GI problems, so prevalence was abnormally low in control group.

Final Thoughts

- Controls should be similar to cases in all other respects other than having the disease in question.

VS.

- Controls should be representative of all persons without the disease in the population from which the cases were selected.

Matching

- Group Matching –
 - Selecting controls in such a manner that the proportion of controls with a certain characteristic is identical to the proportion of cases with the same characteristic.
- Individual Matching –
 - For each case selected in the study, a control is selected to who is similar to the case in terms of the specific variable or variables of concern, e.g. age and gender.

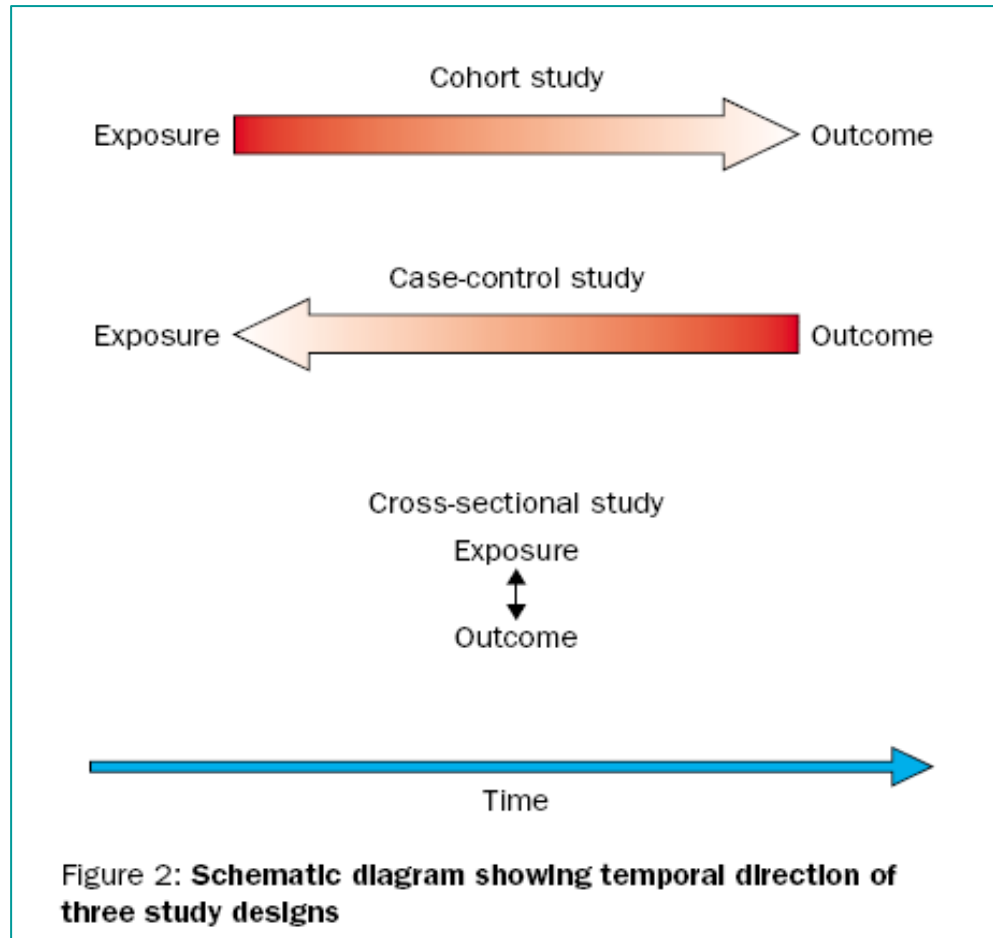
Problems with Matching

1. Too many matching variables might render the study logistically impossible.
2. Once we've matched on a variable, we cannot study that particular characteristic...i.e. we've artificially equalized the groups on that factor so it can't differ.

Main Biases in Case-Control Studies

1. Selection bias
2. Recall bias
 - Cases may intuit the reason for the question, i.e. patients with brain cancer asked, “Do you use a cell phone?”

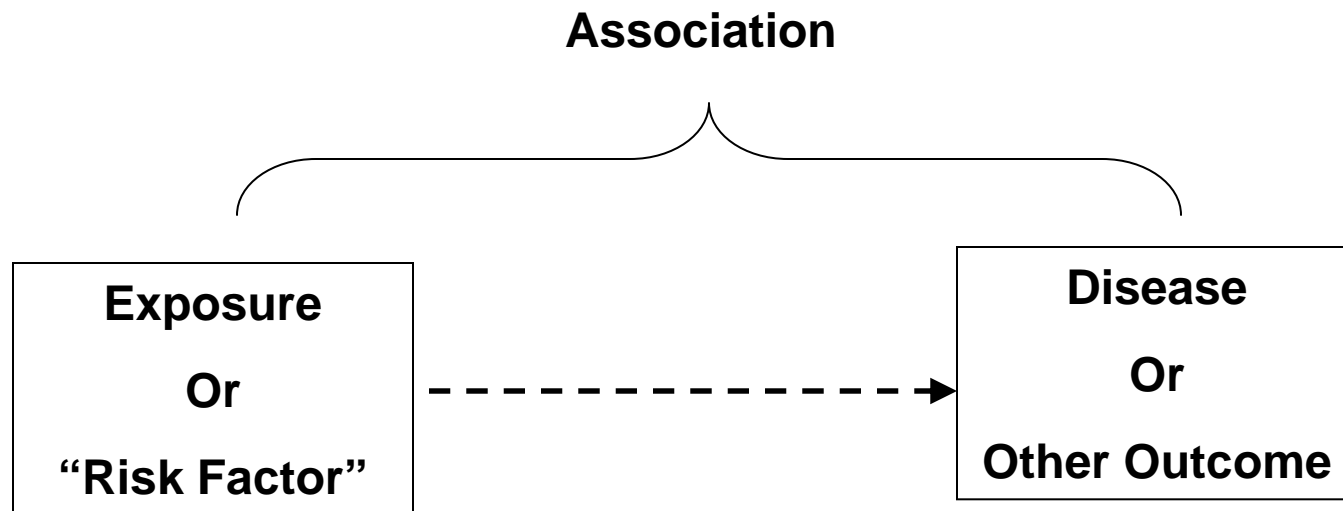
Epidemiologic Study Designs



Cross-Sectional Studies

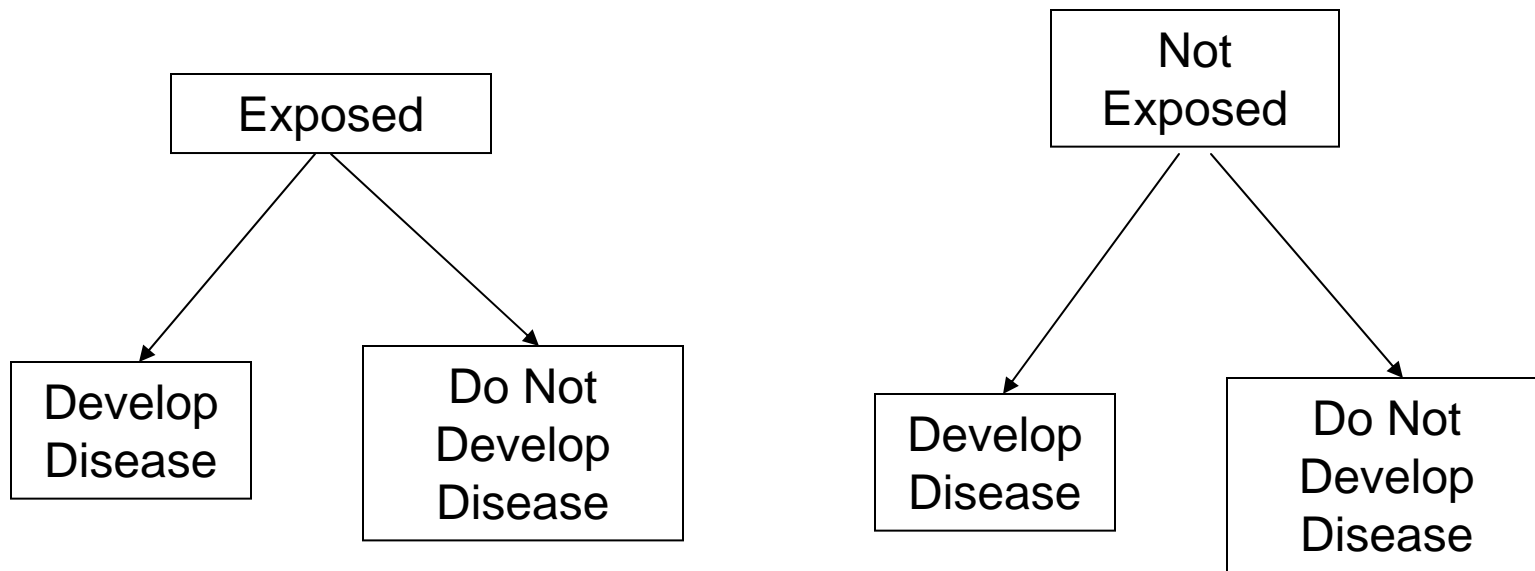
- Both exposure and disease are determined at the “same time”
- “cases” are prevalent cases rather than incident cases (excludes those who have already died)
- Cannot determine which came first, “exposure (risk factor)” or “Disease”

Cohort Studies



If an association is observed, "Is it causal?"

Design of a Cohort Study



Cohort studies START with an observation period or an exposure.

Cohort Study

Then follow to see if:

	Disease Develops	Disease Does Not Develop	Totals	<i>Incidence Rates</i> of Disease
<u>First, Select</u>				
Exposed	a	b	a + b	$\frac{a}{a + b}$
NOT Exposed	c	d	c + d	$\frac{c}{c + d}$

NOTE: Cohort studies allow for precise measurement of exposure.

Cohort Study (smoking and CHD)

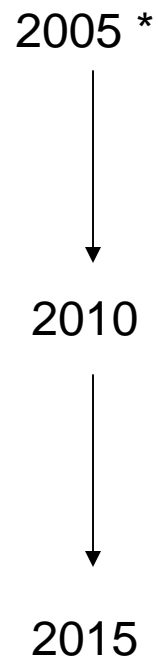
Then follow to see if:

	Disease Develops	Disease Does Not Develop	Totals	<i>Incidence per 1,000 per Year</i>
<u>First, Select</u>				
Smoke	84	2,916	3,000	28.0
Do Not smoke	87	4,913	5,000	17.4

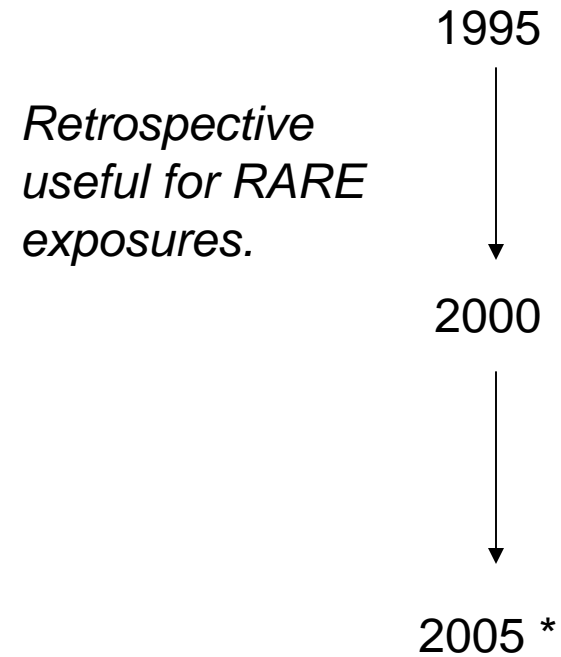
Because these are NEW cases, we can establish a temporal relationship between the exposure (smoking) and CHD.

Types of Cohort Studies

Concurrent



Retrospective



*** Current time**

Cohort Studies

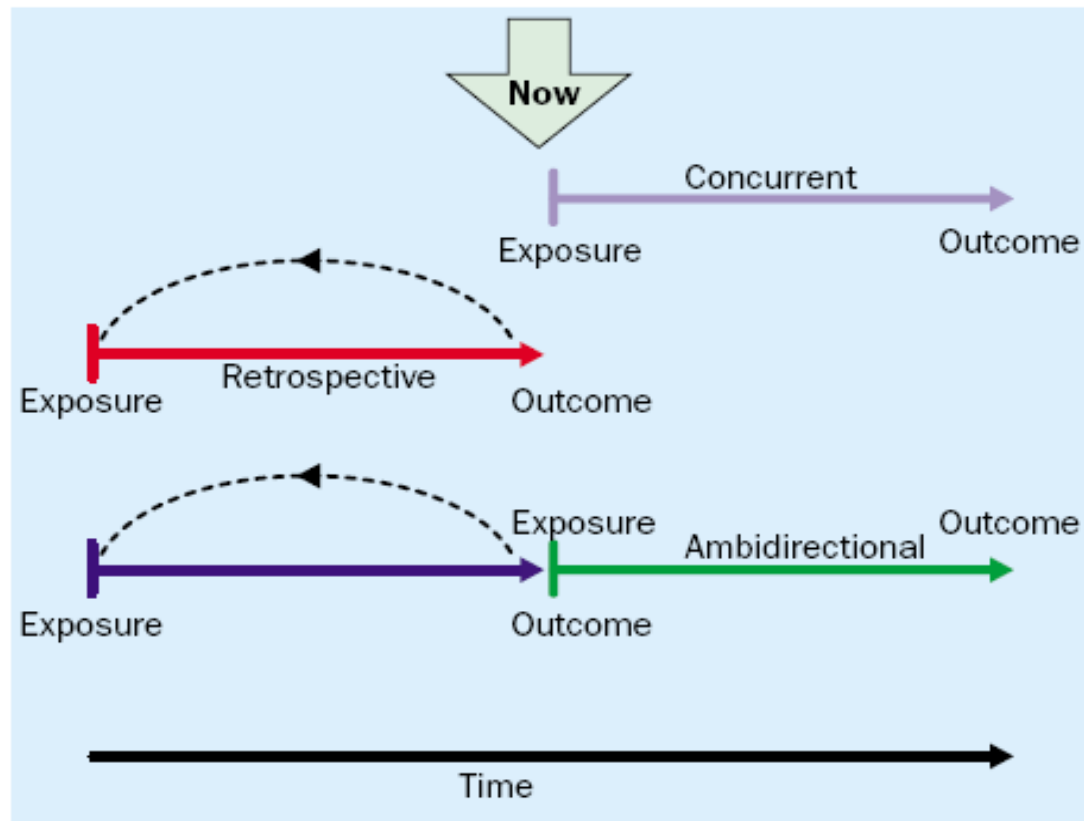


Figure 2: **Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies**

Potential Biases in Cohort Studies

1. Bias in Assessment of Outcome

(Person assessing outcome aware of exposure)

- Solution – blinding (masking) of process or personnel involved in outcome assessment

2. Information Bias

(quality and extent of outcome is different for exposed and non-exposed individuals)

3. Non-response or loss to follow-up

(e.g. persons with disease are differentially lost to follow-up compared to persons without disease)

Advantages of Cohort Studies

- **Can establish population-based incidence**
- **Accurate relative risk (risk ratio) estimation**
- **Can examine rare exposures (asbestos > lung cancer)**
- **Temporal relationship can be inferred (prospective design)**
- **Time-to-event analysis is possible**
- **Can be used where randomization is not possible**
- **Magnitude of a risk factor's effect can be quantified**
- **Selection and information biases are decreased**
- **Multiple outcomes can be studied**
(smoking > lung cancer, COPD, larynx cancer)

Disadvantages of Cohort Studies

- Lengthy and expensive**
- May require very large samples**
- Not suitable for rare diseases**
- Not suitable for diseases with long-latency**
- Unexpected environmental changes may influence the association**
- Nonresponse, migration and loss-to-follow-up biases**
- Sampling, ascertainment and observer biases are still possible**

Presentation of cohort data: Population at risk

Does HIV infection increase risk of developing TB
among a population of drug users?

	Population (follow up 2 years)	Cases
HIV +	215	8
HIV -	289	1

Source: Selwyn et al., New York, 1989

Does HIV infection increase risk of developing TB among drug users?

Exposure	Population (f/u 2 years)	Cases	Incidence (%)	Relative Risk
HIV +	215	8	3.7	11
HIV -	298	1	0.3	

Presentation of cohort data: Person-years at risk

Tobacco smoking and lung cancer, England & Wales, 1951

	Person-years	Cases
Smoke	102,600	133
Do not smoke	42,800	3

Source: Doll & Hill

EPIET ([www](http://www.epiet.com))

Presentation of data: Various exposure levels

Daily number of cigarettes smoked	Person-years at risk	Lung cancer cases
> 25	25,100	57
15 - 24	38,900	54
1 - 14	38,600	22
none	42,800	3

Cohort study: Tobacco smoking and lung cancer, England & Wales, 1951

Cigarettes smoked/d	Person-years at risk	Cases	Rate per 1000 p-y	Rate ratio
> 25	25,100	57	2.27	32.4
15 - 24	38,900	54	1.39	19.8
1 - 14	38,600	22	0.57	8.1
none	42,800	3	0.07	Ref.

Source: Doll & Hill

EPIET ([www](#))

Cohort Studies

Panel 2: Features to look for in a cohort study

How much selection bias was present?

- 1 Were only people at risk of the outcome included?
- 1 Was the exposure clear, specific, and measurable?
- 1 Were the exposed and unexposed groups similar in all important respects except for the exposure?

What steps were taken to minimise information bias?

- 1 Was the outcome clear, specific, and measurable?
- 1 Was the outcome identified in the same way for both groups?
- 1 Was determination of outcome made by an observer blinded as to treatment?

How complete was the follow-up of both groups?

- 1 What efforts were made to limit loss to follow-up?
- 1 Was loss to follow-up similar in both groups?

Were potential confounding factors sought and controlled for in the analysis?

- 1 Did the investigators anticipate and gather information on potential confounding factors?
- 1 What method(s) were used to assess and control for confounding?

Randomized Controlled Trials (RCT)

- Type of cohort study (prospective) except “exposure” is randomly assigned.
 - E.g. drug vs. placebo
- Randomization removes bias and confounding as long as proper study procedures are followed.
 - Blinding of treatment assignment (double)
 - Blinding of outcome assessment