Study Designs in Epidemiology

Siamak Sabour, MD, MSc, DSc, PhD, Postdoc
Clinical Epidemiologist

Persian International Epidemiology Network
(PIEPNET)

Comparison

Qualitative

- Understanding
- Interview/observation
- Discovering frameworks
- Textual (words)
- Theory generating
- Quality of informant more important than sample size
- Subjective
- Embedded knowledge
- Models of analysis: fidelity to text or words of interviewees

Quantitative

- Prediction
- Survey/questionnaires
- Existing frameworks
- Numerical
- Theory testing (experimental)
- Sample size core issue in reliability of data
- Objective
- Public
- Model of analysis: parametric, non-parametric

Descriptive Studies: Uses

- Hypothesis Generating
- Suggesting Associations

Descriptive Studies

- Case reports
- Case series
- Population studies

Analytical Studies

- Observational
- Experimental

Observational Studies

- Cross-sectional
- Case-control
- Cohort

Cross-sectional study; Definition

- A **cross-sectional** studies
  - a type of observational or descriptive study
  - the research has no control over the exposure of interest .
- It involves
  - identifying a defined population at a particular point in time
  - measuring a range of variables on an individual basis
  - include past and current exposure

Cross-sectional Study

- Data collected at a single point in time
- Describes associations
- Prevalence

A "Snapshot"

Prevalence vs. Incidence

- Prevalence (شیوع)
  - The total number of cases at a point in time
  - Includes both new and old cases
- Incidence (بروز)
  - The number of new cases over time

Uses of cross-sectional studies

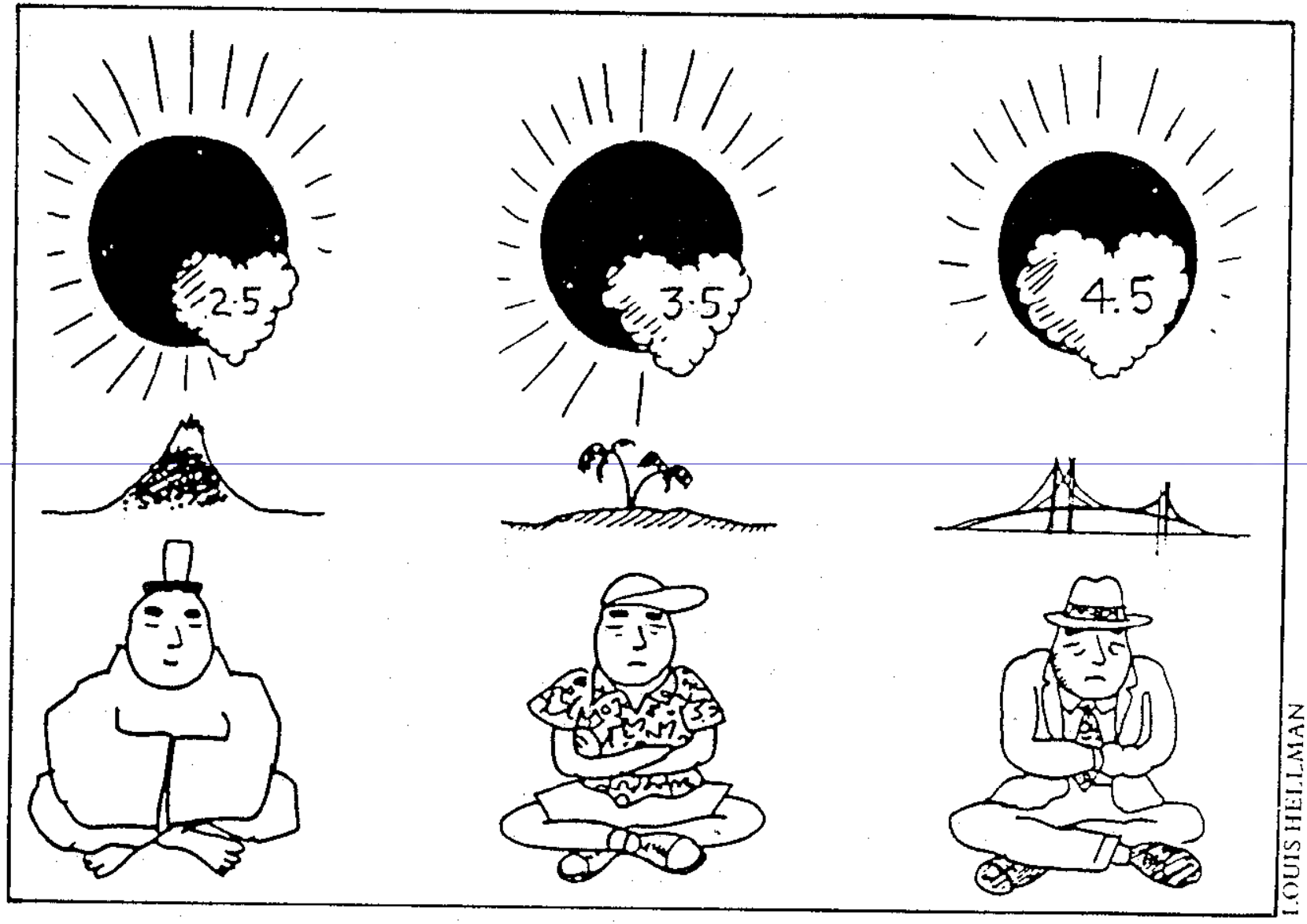
- **Prevalence survey:**

  The studies are commonly used to describe the burden of disease in the community and its distribution.

- **Describe population characteristics:**

  In terms of person (who?) and place (where?)

  - The British National Diet and Nutrition Survey
  - Nutrition and Health Survey in Taiwan
  - To describe various age groups in the population in terms of food and nutrient intake and range of other personal and lifestyle characteristics.



g. 13.3 Prevalence (percentage) of coronary heart disease (as indicated by Q waves in electrordiogram) among men of Japanese ancestry living in Japan (left), Honolulu (centre), and San Francsico Bay area (right). (Source: Marmot and Davey-Smith.[9])

Design of cross-sectional survey

- The problem to be studied must be **clearly described (PICO format)** and a thorough literature review undertaken **before starting the data collection.**
- **Specific objectives** need to be formulated.
- Data collection **techniques** need to be decided.
- **Sampling** is a particularly important issue.

Limitation of cross-sectional study

- It is **not possible** to talk about **causality. (علیت)**

Causality by cross-sectional design means(不能判定因果關係) !!!

- **Confounded** results may lead to misinterpretation.

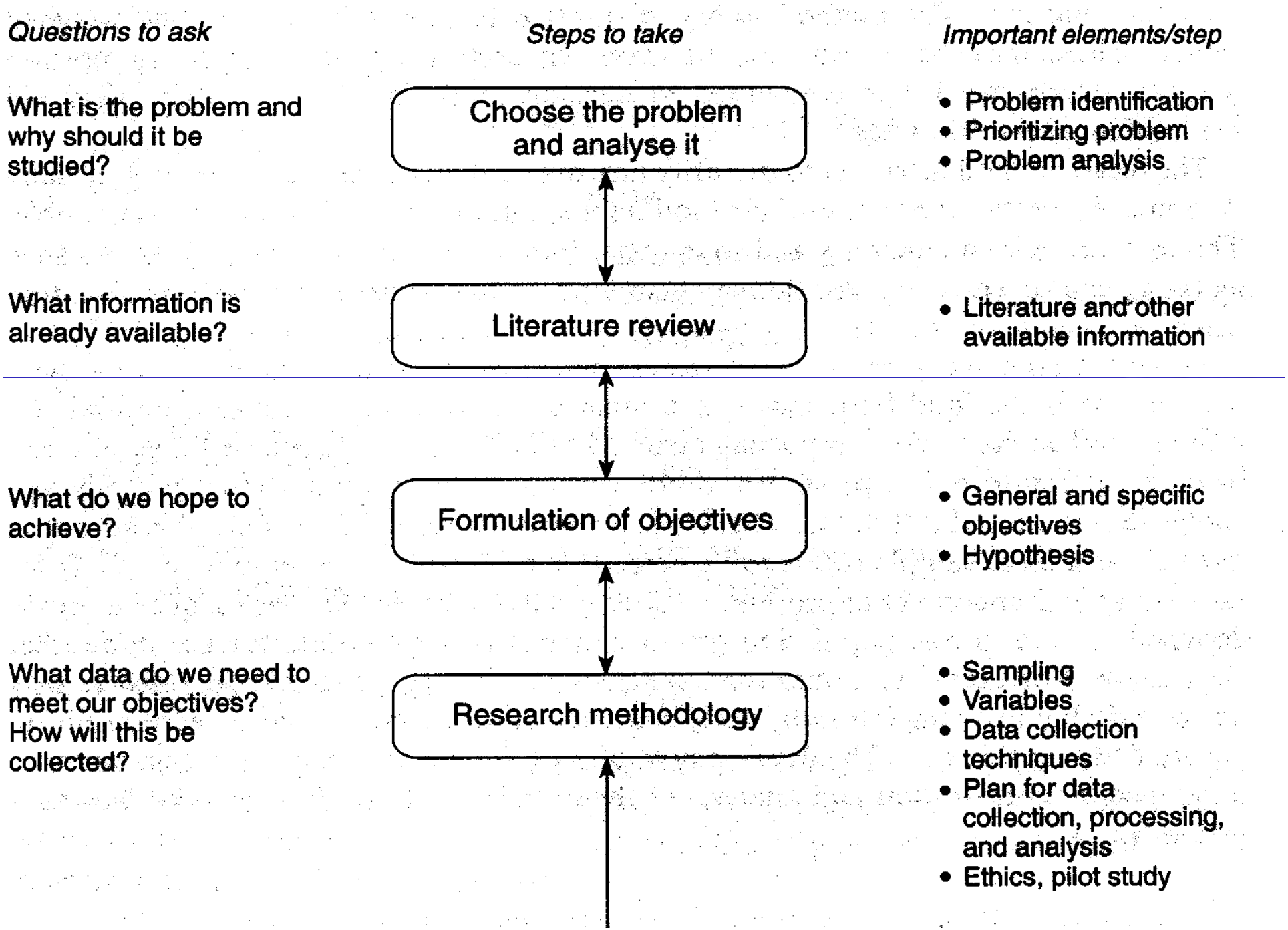
*e.g.: Association of **Boldness & Heart Diseases (p<0.05)***

- etc

Cross-sectional study

- Fieldwork needs planning:
  - **Who** is available to collect the data ?
  - Do they need **training** ?
  - If more than one is to collect the data then it is necessary to assess between-observer variation.
- The collection, coding and entry of data need planning.
- A pilot study is essential to test the proposed methods and make any alternations as necessary.

The steps are summarized below


Questions to ask
Steps to take
Important elements/step
What is the problem and why should it be studied?
Choose the problem and analyse it
• Problem identification
• Prioritizing problem
• Problem analysis
What information is already available?
Literature review
• Literature and other available information
What do we hope to achieve?
Formulation of objectives
• General and specific objectives
• Hypothesis
What data do we need to meet our objectives? How will this be collected?
Research methodology
• Sampling
• Variables
• Data collection techniques
• Plan for data collection, processing, and analysis
• Ethics, pilot study

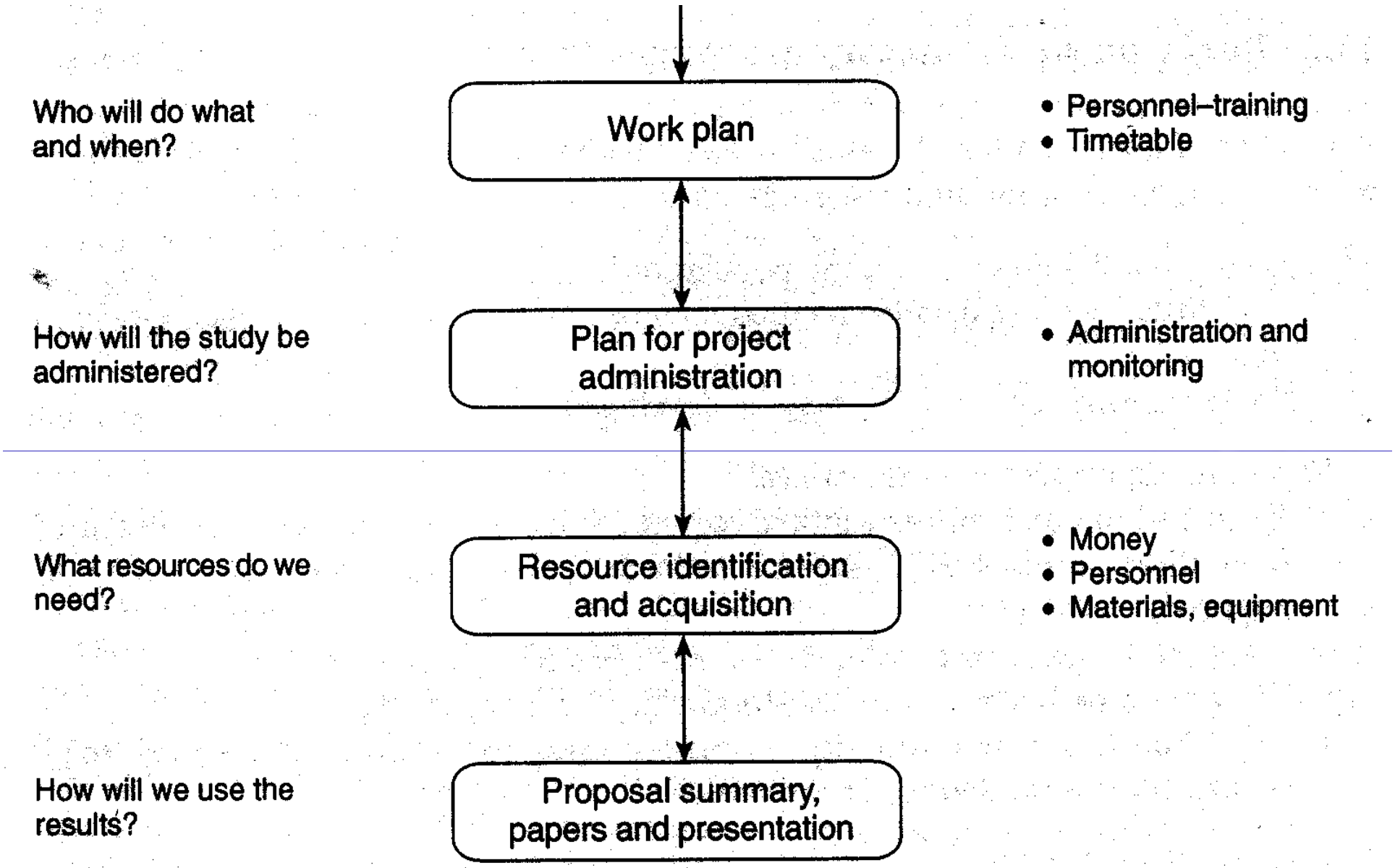


Fig. 13.5 Steps in the design of a cross-sectional study. (Modified from Varkevisser *et al.*[23])

Analysis of cross-sectional study

- Before starting any analysis, the data should be checked for any errors.
  - Obvious error must be corrected.
  - Checking normality of data distribution.
    The Kolmogorov-Smirnov or Histogram for distribution of data
  - Outliers
  - etc

Cross-sectional studies

- People are studied at a "point" in time, without follow-up.
- Can combine a cross-sectional study with follow-up to create a **cohort (longitudinal) study.**
- Can conduct repeated cross-sectional studies to measure **change in a population**.

Cross-sectional studies

- 43.0 million people in the U.S., under age 65 years old, were uninsured (16.4%)

(National Health Interview Survey, 2007)

- 66.3% of no institutionalized U.S. adults age 20+ years were overweight or obese.

(National Health and Nutrition Examination Survey, 2003-2004)

- 35% (~7.4 million) of births to U.S. women during the preceding 5 years were mistimed or unwanted.

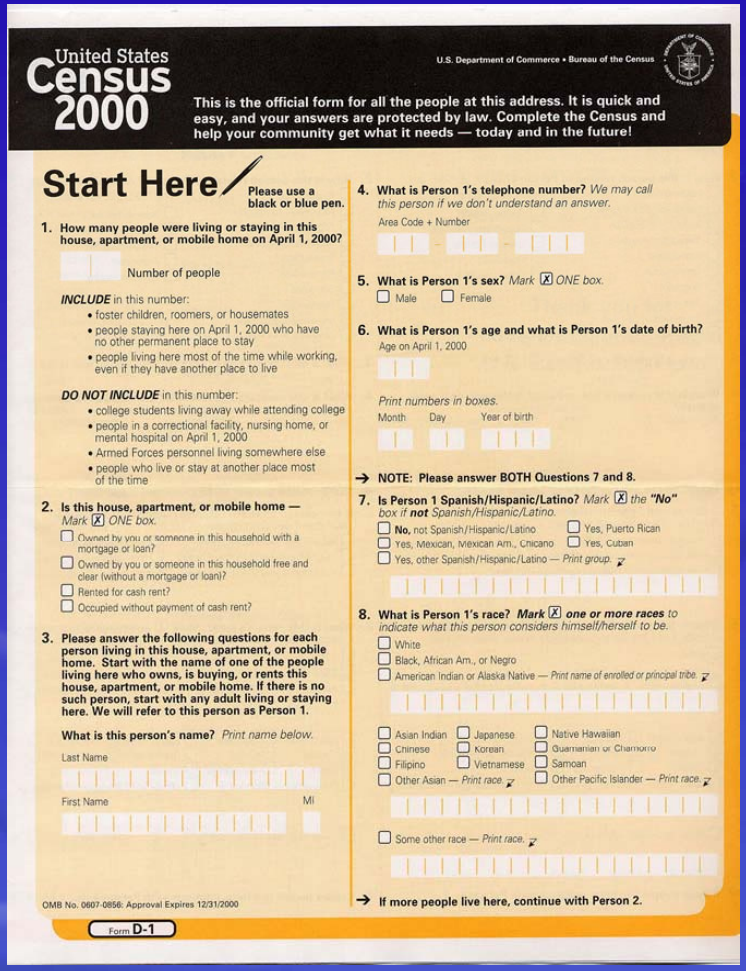
(2002 National Survey of Family Growth, Series 23, No. 25, Table 21)

Cross-sectional studies

- Incidence information is not available from a typical cross-sectional study.
- The source of most of what we know about the population

Population census

- A cross-sectional study of an entire population
- Provides the denominator data for many purposes (e.g., estimation of rates, …)
- A huge effort – people can be difficult to find and to count; may not want to provide data and …



United States Census 2000

U.S. Department of Commerce • Bureau of the Census

This is the official form for all the people at this address. It is quick and easy, and your answers are protected by law. Complete the Census and help your community get what it needs — today and in the future!

Start Here Please use a black or blue pen.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?
   Number of people

   INCLUDE in this number:
   - foster children, roomers, or housemates
   - people staying here on April 1, 2000 who have no other permanent place to stay
   - people living here most of the time while working, even if they have another place to live

   DO NOT INCLUDE in this number:
   - college students living away while attending college
   - people in a correctional facility, nursing home, or mental hospital on April 1, 2000
   - Armed Forces personnel living somewhere else
   - people who live or stay at another place most of the time

2. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*
   - ☐ Owned by you or someone in this household with a mortgage or loan?
   - ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
   - ☐ Rented for cash rent?
   - ☐ Occupied without payment of cash rent?

3. Please answer the following questions for each person living in this house, apartment, or mobile home. Start with the name of one of the people living here who owns, is buying, or rents this house, apartment, or mobile home. If there is no such person, start with any adult living or staying here. We will refer to this person as Person 1.

   What is this person's name? *Print name below.*
   Last Name
   First Name MI

4. What is Person 1's telephone number? *We may call this person if we don't understand an answer.*
   Area Code + Number

5. What is Person 1's sex? *Mark ☒ ONE box.*
   ☐ Male ☐ Female

6. What is Person 1's age and what is Person 1's date of birth?
   Age on April 1, 2000
   Print numbers in boxes.
   Month Day Year of birth

→ NOTE: Please answer BOTH Questions 7 and 8.

7. Is Person 1 Spanish/Hispanic/Latino? *Mark ☒ the "No" box if not Spanish/Hispanic/Latino.*
   ☐ No, not Spanish/Hispanic/Latino ☐ Yes, Puerto Rican
   ☐ Yes, Mexican, Mexican Am., Chicano ☐ Yes, Cuban
   ☐ Yes, other Spanish/Hispanic/Latino — *Print group.*

8. What is Person 1's race? *Mark ☒ one or more races to indicate what this person considers himself/herself to be.*
   ☐ White
   ☐ Black, African Am., or Negro
   ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
   ☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
   ☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
   ☐ Filipino ☐ Vietnamese ☐ Samoan
   ☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
   ☐ Some other race — *Print race.*

→ If more people live here, continue with Person 2.

OMB No. 0607-0856: Approval Expires 12/31/2000

Form D-1

Case - Control Studies

- **Characteristics**: two source populations; (*assumption that non-cases are representative of the source population of cases.)*
- **Merits**: least expensive; least time-consuming; suitable for study of rare diseases
- **Limitations**: not suitable for rare exposures; liable to selection bias and recall bias; not suitable for calculation of frequency measures.
- **Effect measure: Odds Ratio (نسبت شانس)**

Design of Case-Control Studies

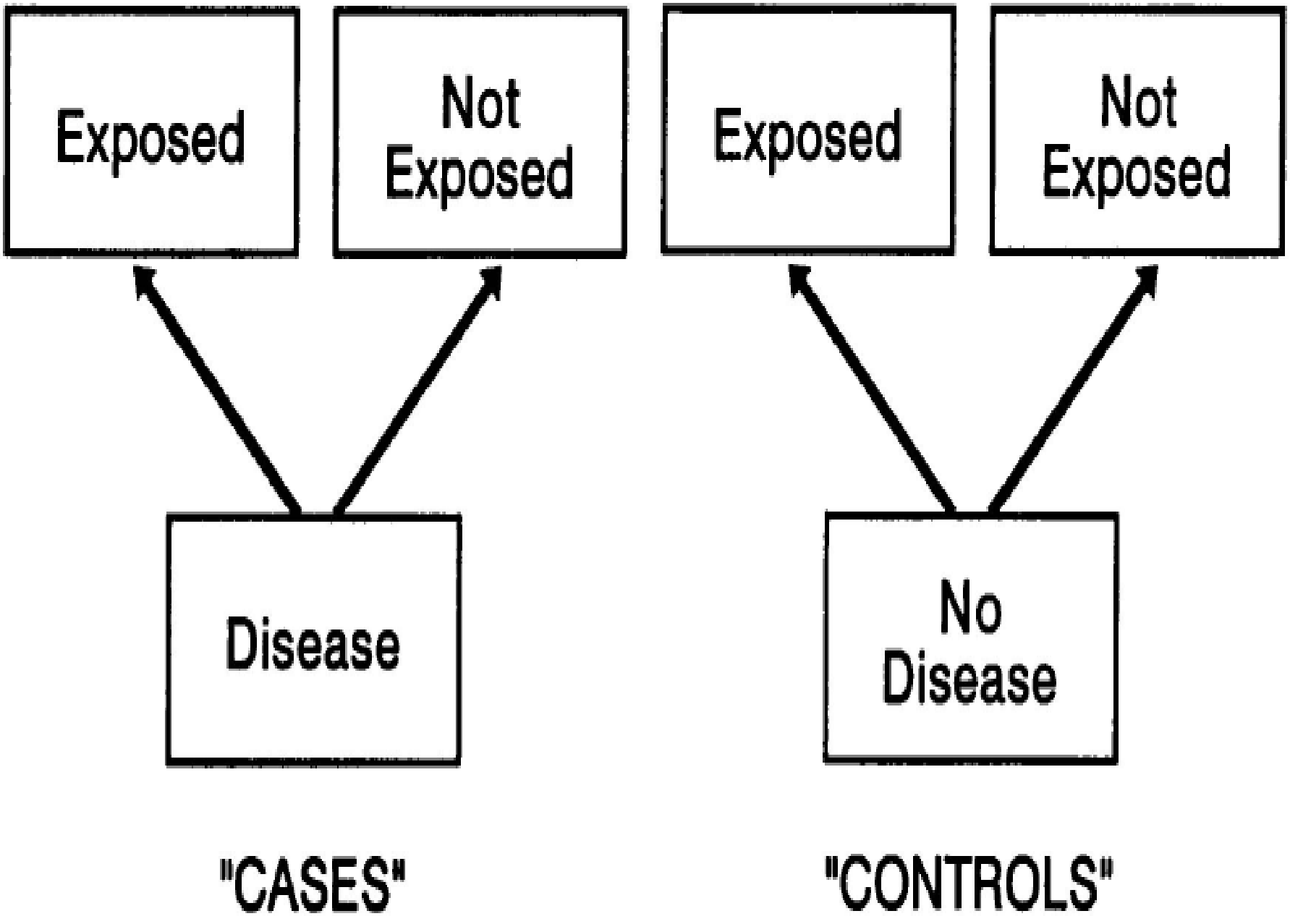
- **Cases:**

  persons/group with a given disease

- **Controls:**

  persons/group without the given disease

- Ascertain exposure or background of the two groups and compare the odds
- Best suited for study of diseases where medical care usually sought, (hip fracture, cancer) because this makes it easier to identify cases


Exposed
Not Exposed
Exposed
Not Exposed
Disease
No Disease
"CASES"
"CONTROLS"

Selection of Cases

- **Ideally**, investigator identifies & enrolls all **incident cases** in a defined population in a specified time period.
- Select cases from registries or hospitals, clinics
- *When all incident cases in a population are included, the study is representative; otherwise there is potential for bias (e.g. referral bias)*
- Use of **prevalent** vs. incident cases (rare diseases)

Selection of controls

- Critical that the exposure in the controls is representative of the exposure in the population
- Ideal controls would have same/similar characteristics as the cases
- Matching cases to controls *(avoid overmatching)*

Population-Based Controls

- The **best control group** is a random sample of individuals from same source population (as the cases) who have not developed the disease

- Population-based controls are the best way to ensure that the **distribution of exposure among the controls is representative**

Hospital Controls

- Hospital controls are the most frequently used source
- Hospital controls may not be representative of exposure rates in the target population
- The use of other ill persons as controls will provide a valid result **only if** their illness is **unrelated** to the exposure in question.

Hospital Controls

- Convenient
- Cheap
- Numerous
- Avoids non-response

When a population-based case registry is not available, hospital controls better represent the subpopulation from which the cases arose

Other Controls

- **Neighborhood** controls are somewhat matched on SES & environmental exposures but may "overmatch" & be expensive

- **Friends & relatives** also cause problems with "overmatching" with habits, environment and occupation & are generally a **poor choice for controls**

Multiple Controls

- Control to Case ratio used is usually 1:1; if large number and cost is the same for both groups

- If a study has a small number of cases, increasing the number of controls increases power of study

Control to Case **ratio**	**power** of study
1:1	65-70%
2:1	75-80%
3:1	90-95%
4:1	**95-97%**

Advantages of Case Control Design

- Relatively inexpensive
- **Good** for diseases with long latency
- **Optimal** for rare diseases
- Multiple exposure evaluated for single disease
- Shorter time
- Smaller sample

Limitations of Case Control Design

- Identifying controls may be difficult
- Temporal relationship between exposure & disease difficult to establish
- Prone to bias (Recall) compared with other study designs

Limitations of Case Control Design

- Difficult to determine representativeness of cases & controls

- Can't measure incidence of disease

  Nested case control design can measure incidence of disease

- Bad for rare exposures

Case Control Design

- At baseline:
  - Selection of cases (disease) and controls (no disease) based on disease status
  - Exposure status is unknown
- Retrospective design – lacks temporality !

Case Control Design

Sources of cases and controls

CASES	CONTROLS
All cases diagnosed in the community	Sample of general population
All cases diagnosed in a sample of the population	Non-cases in a sample of the population
All cases diagnosed in all hospitals	Sample of patients in all hospitals who do not have the disease
All cases diagnosed in a single hospital	Sample of patients in the same hospital who do not have the disease
Any of the above methods	Spouses, siblings or associates of cases

Assessing Exposure in Case-Control Design

- Exposure estimates are subject to

  recall bias and *interviewer bias*

  – Some protection may be afforded by blinding interviewers and carefully phrasing interview questions

- **Potential confounders** need to be **accurately** assessed in order to be controlled in the analysis

Odds Ratio (OR)

- A ratio that measures the odds of exposure for cases compared to controls
- Odds of exposure = number exposed ÷ number unexposed
- OR Numerator: Odds of exposure for cases
- OR Denominator: Odds of exposure for controls

Odds Ratio (OR)

Exposure Status	Disease Status: CHD cases (Cases)	Disease Status: No CHD (Controls)
Smoker	112	176
Non-smoker	88	224
Total	200	400

$$\text{Odds Ratio} \approx \frac{AD}{BC} = \frac{112 \times 224}{176 \times 88} \approx 1.62$$

Interpreting the Odds Ratio

- The odds of exposure for cases are 1.62 times the odds of exposure for controls.
- Those with CHD are 1.62 times more likely to be smokers than those without CHD.
- Those with CHD are 62% more likely to be smokers than those without CHD.

	OR<1	OR=1	OR>1
Odds comparison between cases and controls	Odds of exposure for cases are less than the odds of exposure for controls	Odds of exposure are equal among cases and controls	Odds of exposure for cases are greater than the odds of exposure for controls
Exposure as a risk factor for the disease?	Exposure reduces disease risk (Protective factor)	Particular exposure is not a risk factor	Exposure increases disease risk (Risk factor)

Possible Sources of Bias and Error

- **Information on the potential risk factor (exposure) may not be available either from records or the study subjects' memories**
- **Information on potentially important confounding variables may not be available either from records or the study subjects' memories**

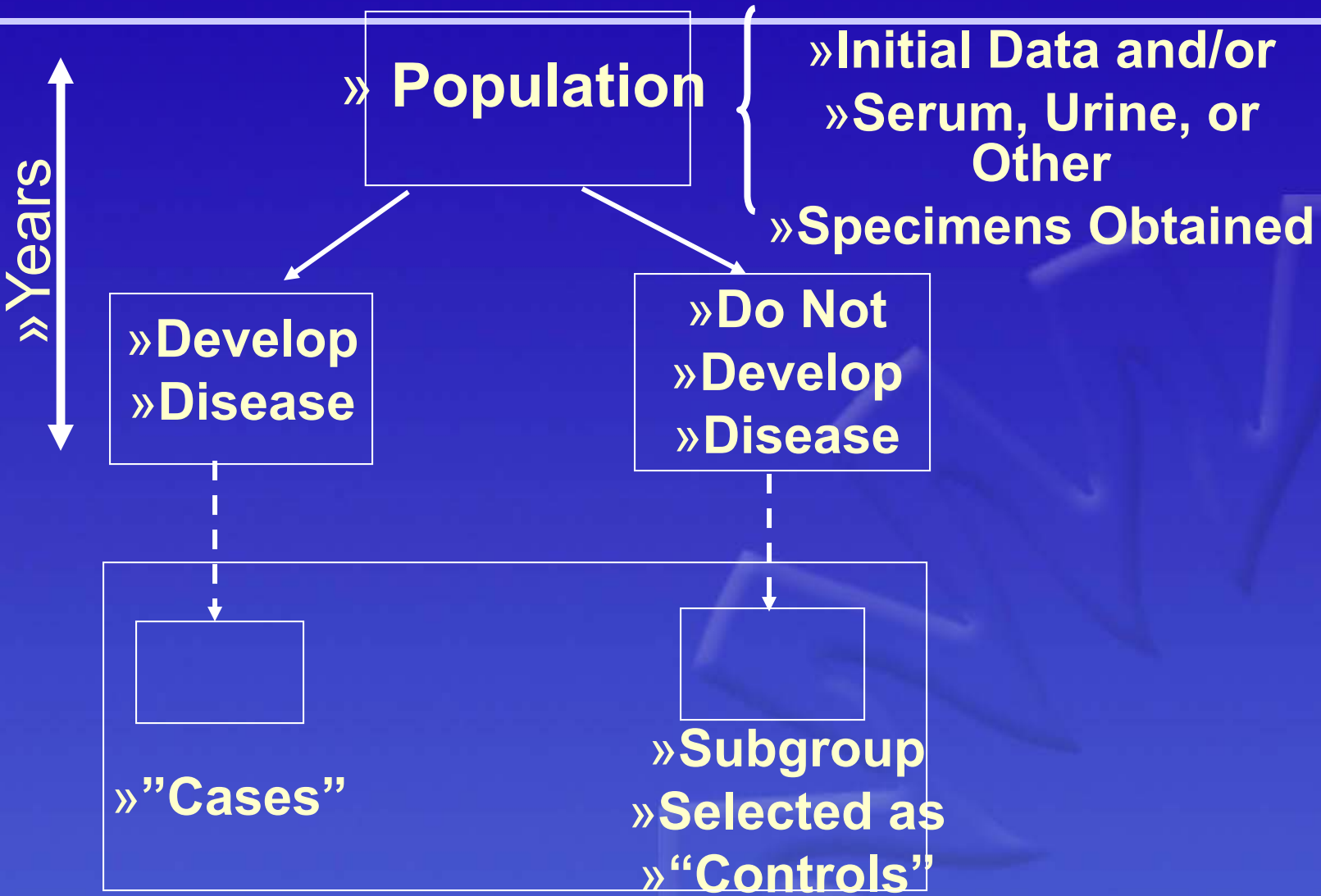
Possible Sources of Bias and Error (cont.)

- **Cases may search for a cause for their disease and thereby be more likely to report an exposure than controls (recall bias)**
- **The investigator may be unable to determine with certainty whether the suspected agent caused the disease or whether the occurrence of the disease caused the person to be exposed to the agent**

Possible Sources of Bias and Error (cont.)

- **Identifying and assembling a case group representative of all cases may be unduly difficult**
- **Identifying and assembling an appropriate control group may be unduly difficult**

Nested Case-Control Study



»CASE-CONTROL STUDY

ORs, P-Values and 95% CIs for Case-Control Study with 3 Different Sample Sizes

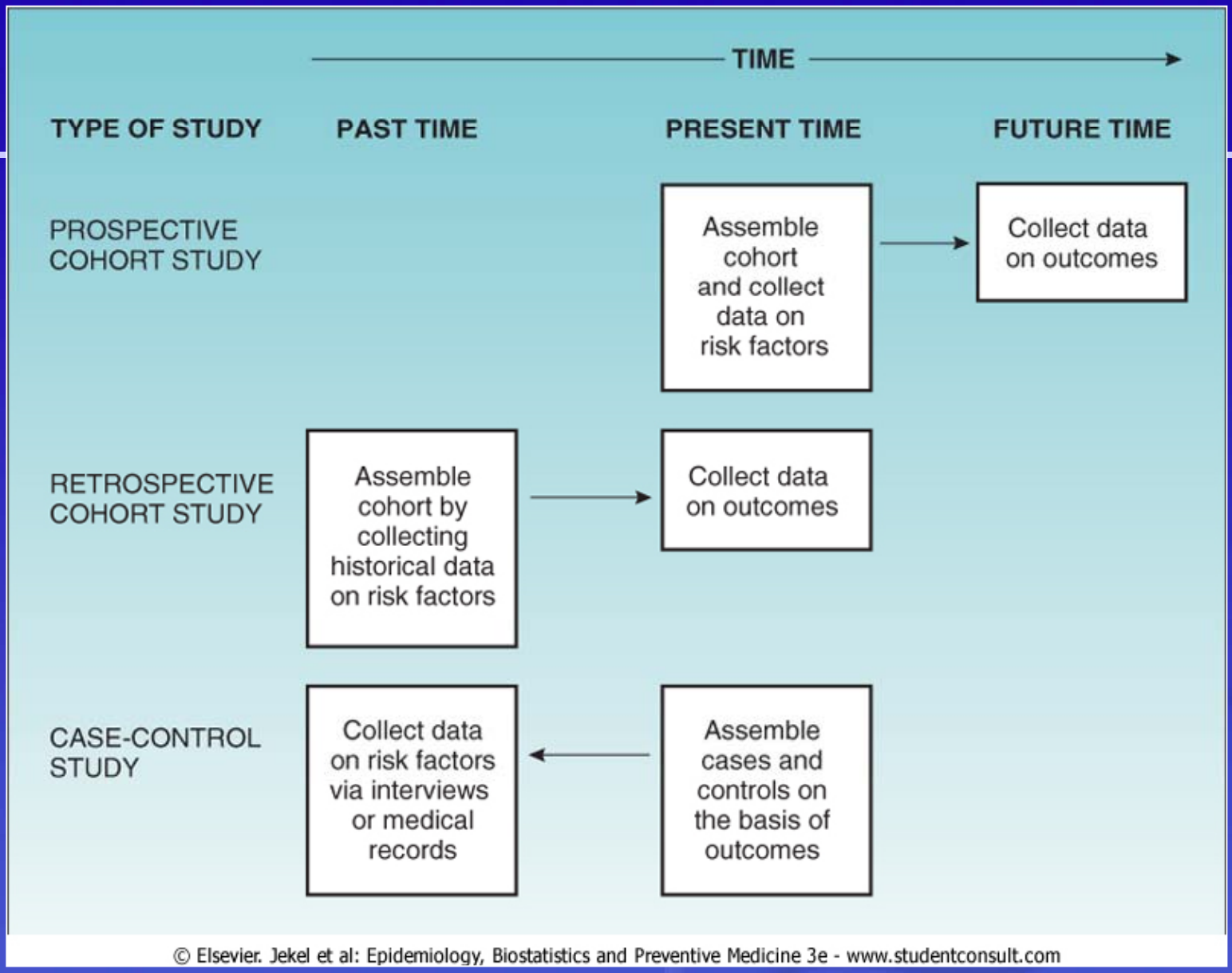
	Sample Size		
Parameter Computed	**n=20**	**n=50**	**n=500**
OR	2.0	2.0	2.0
p-value	0.500	0.200	0.001
95% CIs	0.5, 7.7	0.9, 4.7	1.5, 2.6

Advantages of Case-Control Studies

- Quick and easy to complete, cost effective
- Most efficient design for rare diseases
- Usually requires a smaller study population than a cohort study

Disadvantages of Case-Control Studies

- Uncertainty of exposure-disease time relationship
- Inability to provide a direct estimate of risk
- Not efficient for studying rare exposures
- Subject to biases (recall & selection bias)

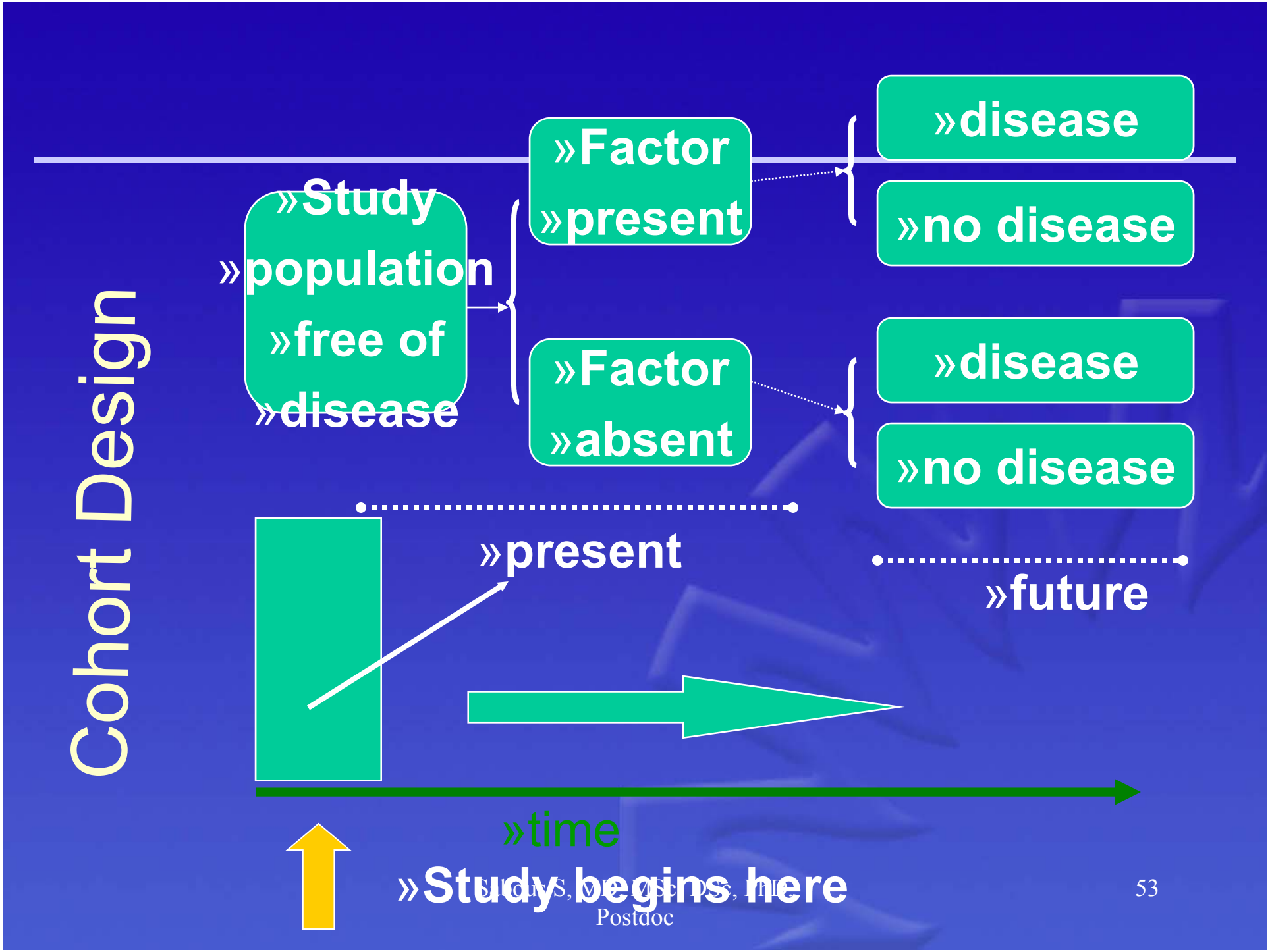

TIME
TYPE OF STUDY
PAST TIME
PRESENT TIME
FUTURE TIME
PROSPECTIVE COHORT STUDY
Assemble cohort and collect data on risk factors
Collect data on outcomes
RETROSPECTIVE COHORT STUDY
Assemble cohort by collecting historical data on risk factors
Collect data on outcomes
CASE-CONTROL STUDY
Collect data on risk factors via interviews or medical records
Assemble cases and controls on the basis of outcomes
© Elsevier. Jekel et al: Epidemiology, Biostatistics and Preventive Medicine 3e - www.studentconsult.com

»Comparison of Case/Control & Cohort Studies

	»Case/Control	»Cohort
Rare Exposure	Inefficient	Efficient
Rare Disease	Efficient	Inefficient
Disease with long latency	Efficient	Inefficient
Cost	Cheap	More Expensive
Time	Shorter	Longer
Size	Smaller	Larger
	Difficult to assess	Good to determine (RR)

Cohort Studies

- Characteristics: follow-up period (prospective; retrospective)
- Merits: no temporal ambiguity; several outcomes could be studied at the same time; suitable for incidence estimation
- Limitations (of prospective type): expensive; time-consuming; inefficient for rare diseases; may not be feasible
- **Effect measure:** Risk Ratio (Relative Risk)


Cohort Design
»Study
»population
»free of
»disease
»Factor
»present
»Factor
»absent
»disease
»no disease
»disease
»no disease
»present
»future
»time
»Study begins here

Study Design

- Objectives
- ☞ *1. Introduce concepts of "counterfactual argument" and "study base"*
- ☞*2. Review the three fundamental study designs*
- *-Cohort (including clinical trials)*
- *-Case-Control*
- *-Cross-Sectional survey*
- ☞ *3. Discuss Cohort Studies*
- *-Uses*
- *-Strengths/weaknesses*
- *-Measure of effect (Relative Risk)*

Cohort Studies

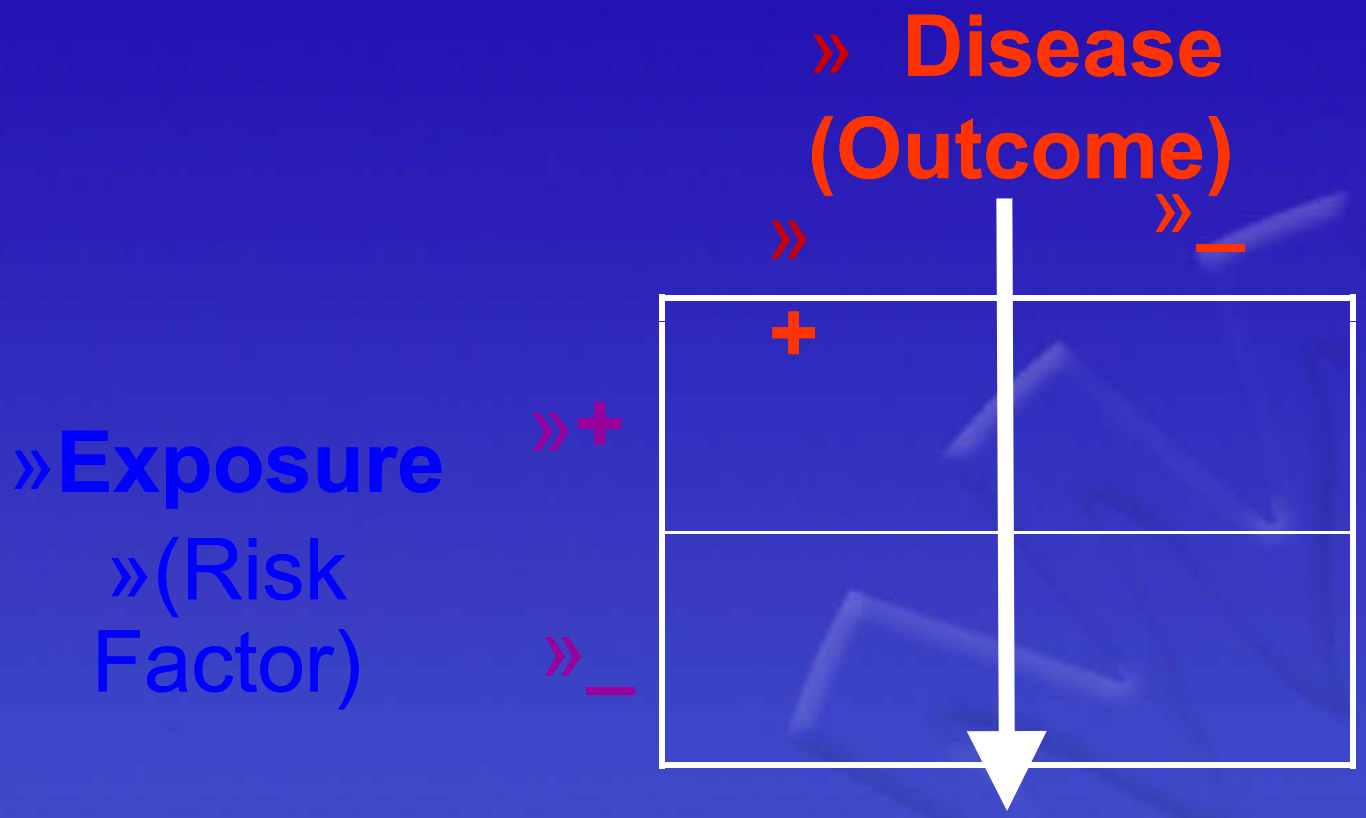
- ➢Begin with sample → "Healthy Cohort" (i.e., subjects without the outcome *yet*)
- ➢Start with Exposure status, then compare subsequent disease experience in exposed vs. unexposed.

Study Design

Case-Control Studies

- ➢ Begin with sample of "Cases and Controls"
- ➢ Start with Disease status, then assess and compare Exposures in cases vs. controls.

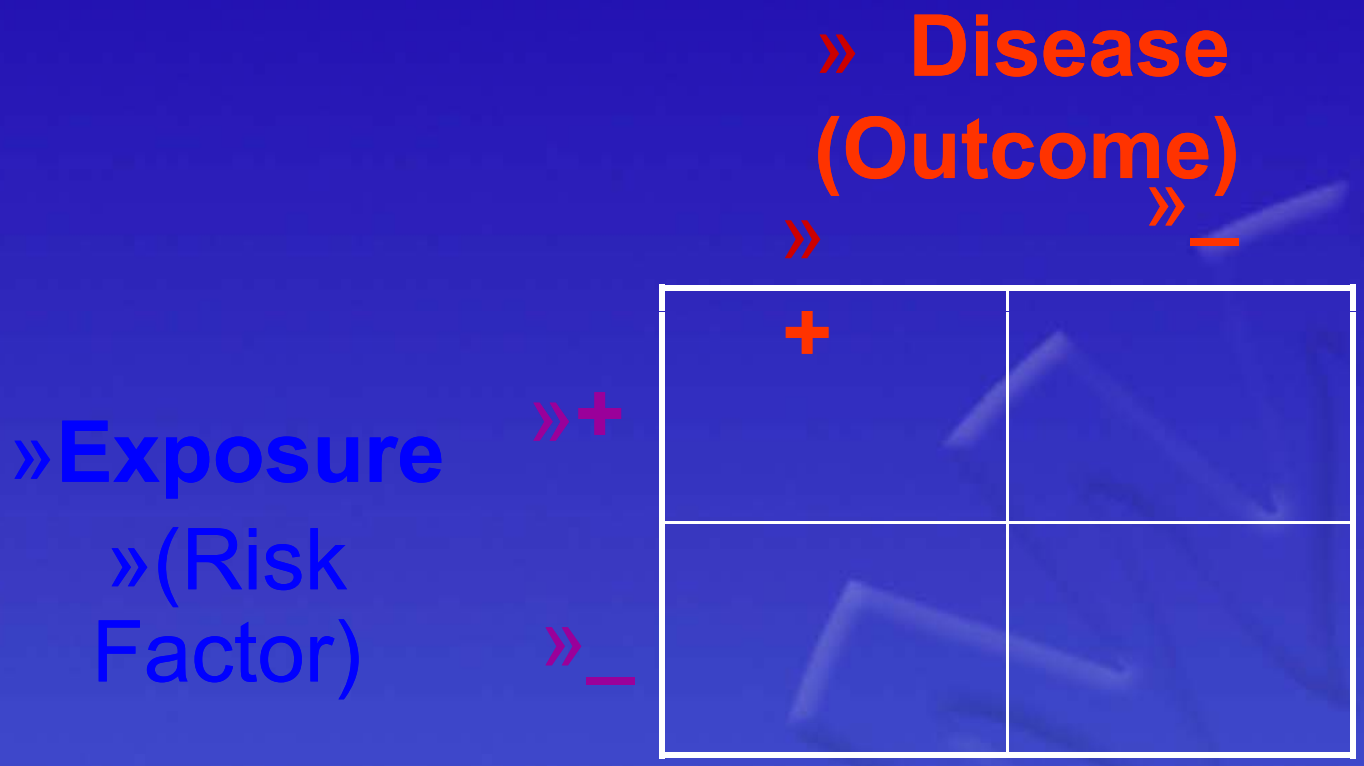
Study Design



Cross-Sectional Studies

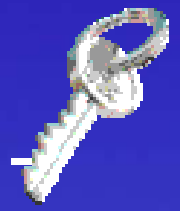
- ➢ Begin with "Cross-sectional" sample
- ➢ Determine Exposure and Disease at same time

Study Design



COHORT STUDIES

- Cohort Study



Key Point:

– Presence or absence of risk factor is determined before outcome occurs.

COHORT STUDIES

		»Disease »(+)	»(-)
»R.F	»(+)	50%	50%
.	»(-)	10%	90%

»**Basic Idea**:

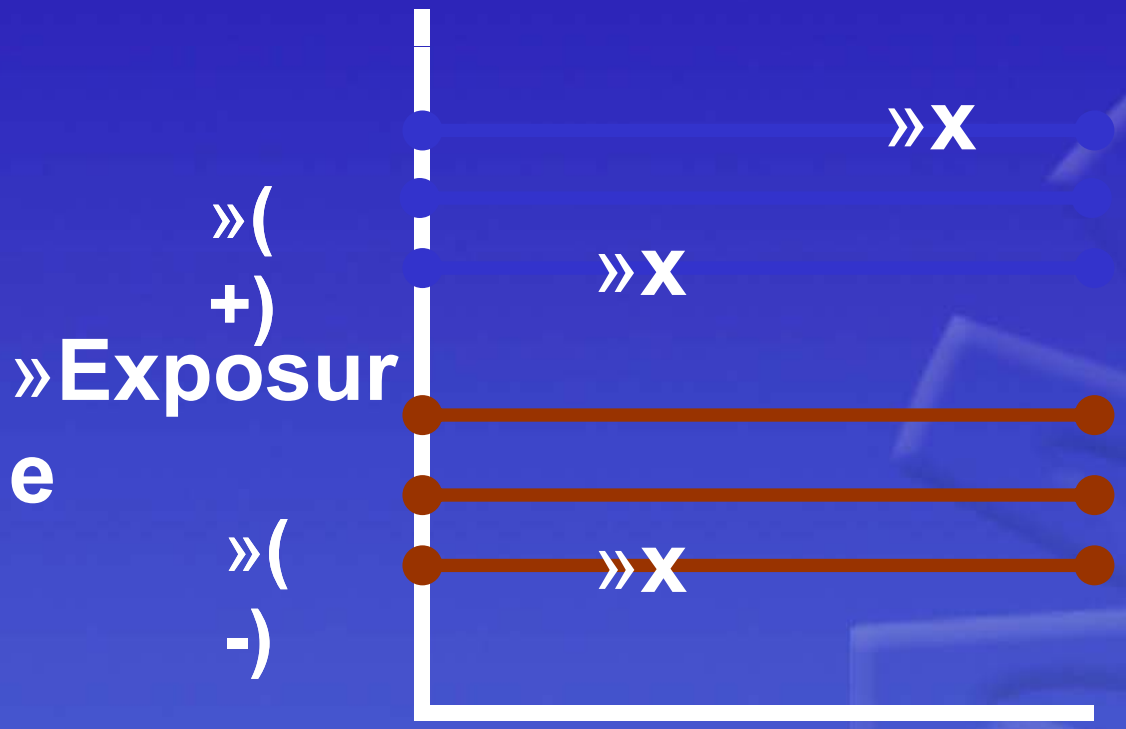
»See if those with the risk factor develop more disease than those without the risk factor

COHORT STUDIES

- Basic Approach: Cohort Study
  - Identify Cohort (s)
  - Measure exposure and outcome variables
  - Follow for development of outcomes

COHORT STUDIES

- Fixed Cohort



COHORT STUDIES

»Fixed cohort

»Disease = Hepatitis A

»S a l a d	»(+)	»(-)
»(+)	30 a	70 b
»(-)	3 c	57 d

»a + b » 100

»c + d 60

»Risk =a/(a+b) =0.3

»Risk =c/(c+d) =0.05

»Rel. risk

$$\frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

»= 0.3/0.05 »= 6

COHORT STUDIES

»Disease = Hep A

»Rel. risk= $\frac{\frac{a}{a+b}}{\frac{c}{c+d}}$ »= 0.3/0.05»= 6

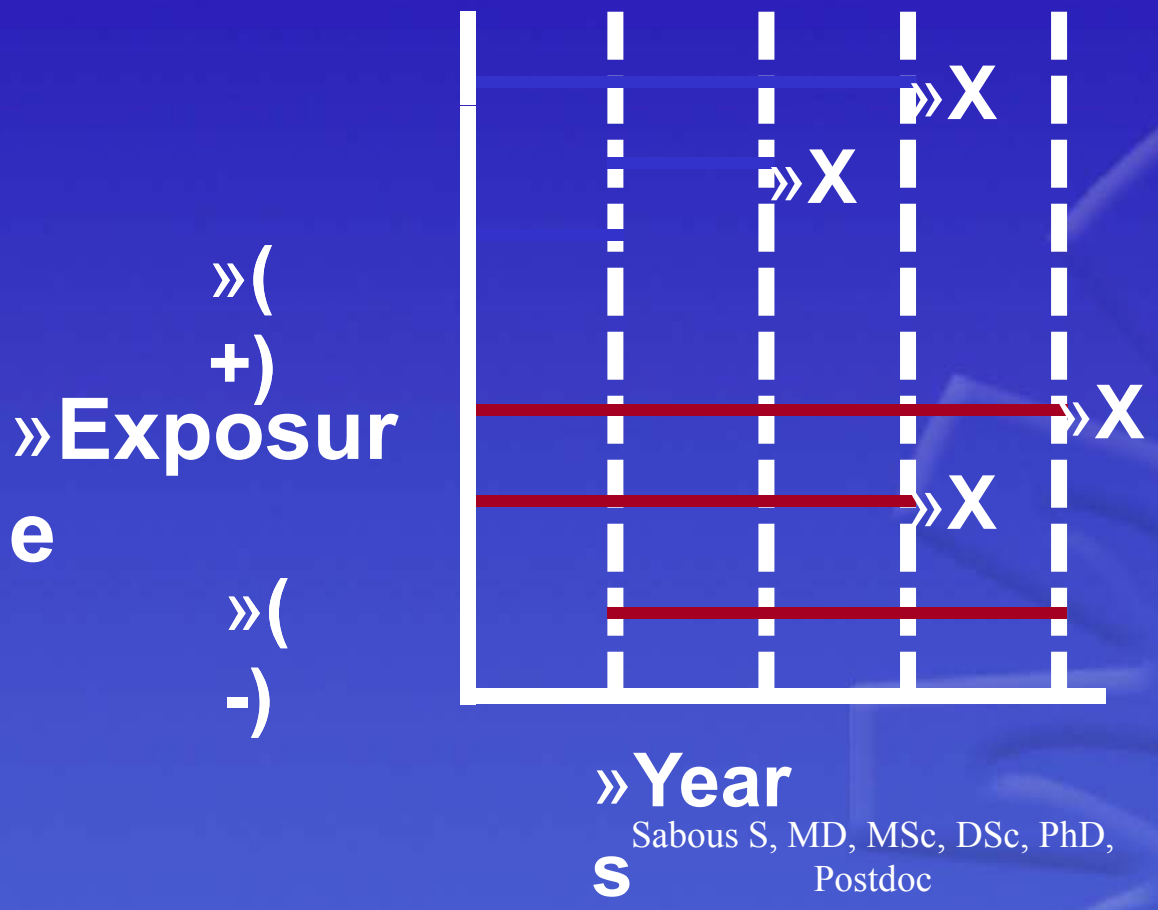
»Salad	»(+)	»(-)	
»(+)	30 a	70 b	»a + b »= 100
»(-)	3 c	57 d	»c + d »= 60

»Odds Ratio: (a/c)/(b/d)=(a/b)/(c/d)

»(30/3)/(70/57)= 8.14

COHORT STUDIES

- Dynamic Cohort



»Rel. Risk

»= 2/3 /2/3 =1

»or

»$2/5_{py}$ /$2/10_{py}$

»= 2.0

COHORT STUDIES

- Cohort : 16, 936 Harvard grads
- Measure: Question re: activity level
- Follow: “Sedentary”: 24 CHD deaths per 10,000 person-years
- vs. “Active”: 16 CHD deaths per 10,000 person-years
- Relative risk = 24/16 = 1.5

COHORT STUDIES

- Questions:
- Findings due to confounding?
- Could subclinical disease have affected the risk factor (activity)?

COHORT STUDIES

- Take-Home Message:
- The best measure of effect is the “relative risk.” For a fixed cohort, this will be the ratio of the cumulative incidences. For a dynamic cohort, this will be the ratio of the incidence rates.
- The odds ratio can be used for fixed cohorts comparing cumulative incidences. It will be close to the relative risk for rare diseases.

COHORT STUDIES

- Variations on a theme:

- Retrospective (Historical) Cohort

COHORT STUDIES

- Prospective: Outcomes have not yet occurred as study begins. Example: Women's Health Study.
- Retrospective: Outcomes have already occurred as the study begins. Example: finding a trove of medical records allowing you to follow a cohort born in 1880 to death.

COHORT STUDIES

- Utility and Strengths
- Incidence and natural history
- Temporal sequence
- Avoid survivor bias
- Avoid reporting bias
- Look at multiple outcomes

COHORT STUDIES

- Limitations:
-
- Inefficient for rare diseases
- Confounding may occur
- Sub-clinical disease may affect risk factor levels
- Loss to follow-up

»Epidemiologic Study Designs

»M. Tevfik DORAK

»HUMIGEN LLC
»Genomic Immunoepidemiology Laboratory
»Hamilton, NJ
»USA

»Clinical Studies & Objective Medicine

»Bodrum, 15-16 April 2006

»Epidemiologic Study Designs

»Experiment
»(RCTs)

»Observational

»Analytical
»*Case-Control*
»*Cohort*

»Descriptive
»+ *cross-sectional & ecologic*

»Epidemiologic Study Designs

»*Descriptive studies*

»Examine patterns of disease

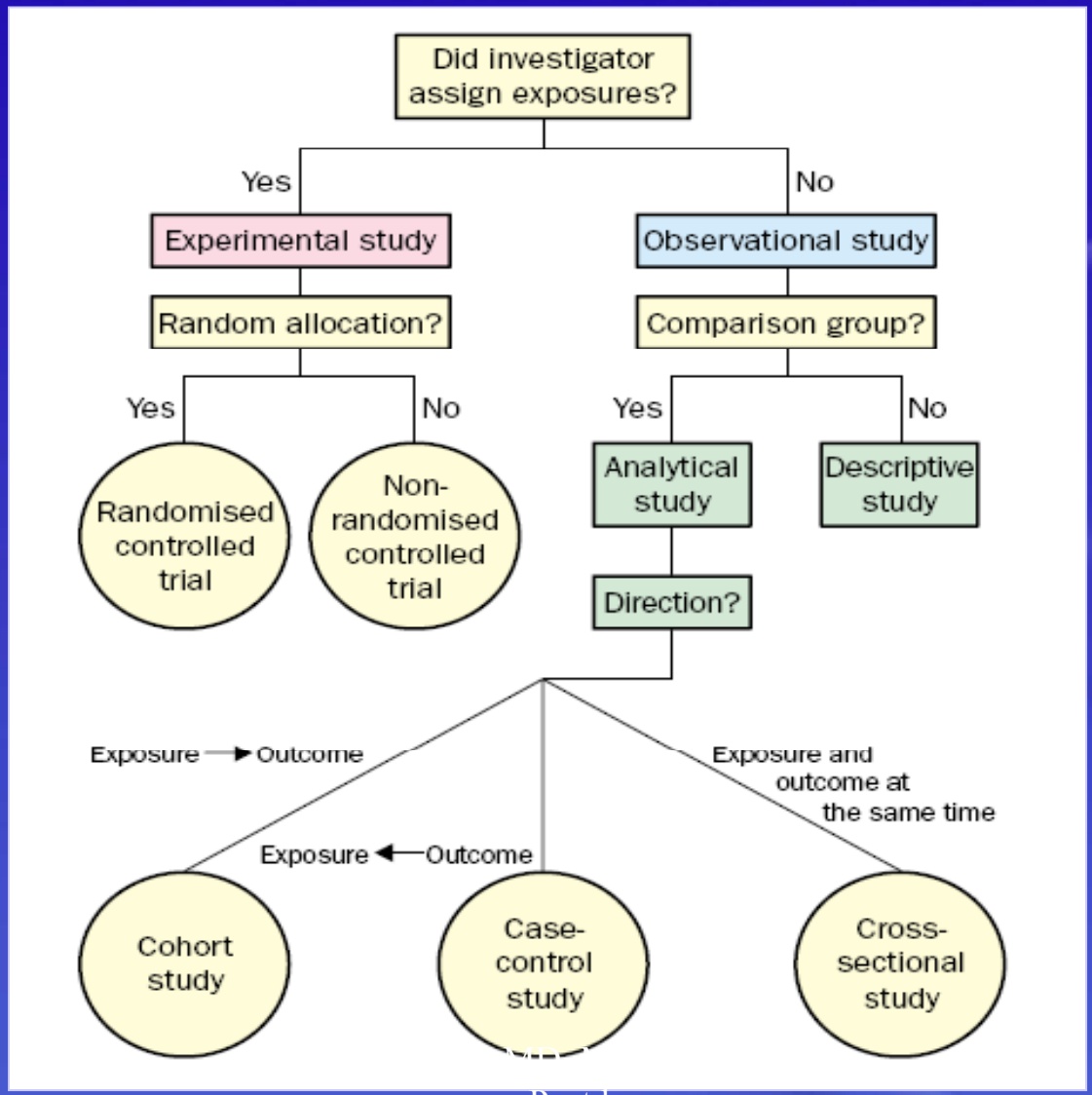
»*Analytical studies*

»Studies of suspected causes of diseases

»*Experimental studies*

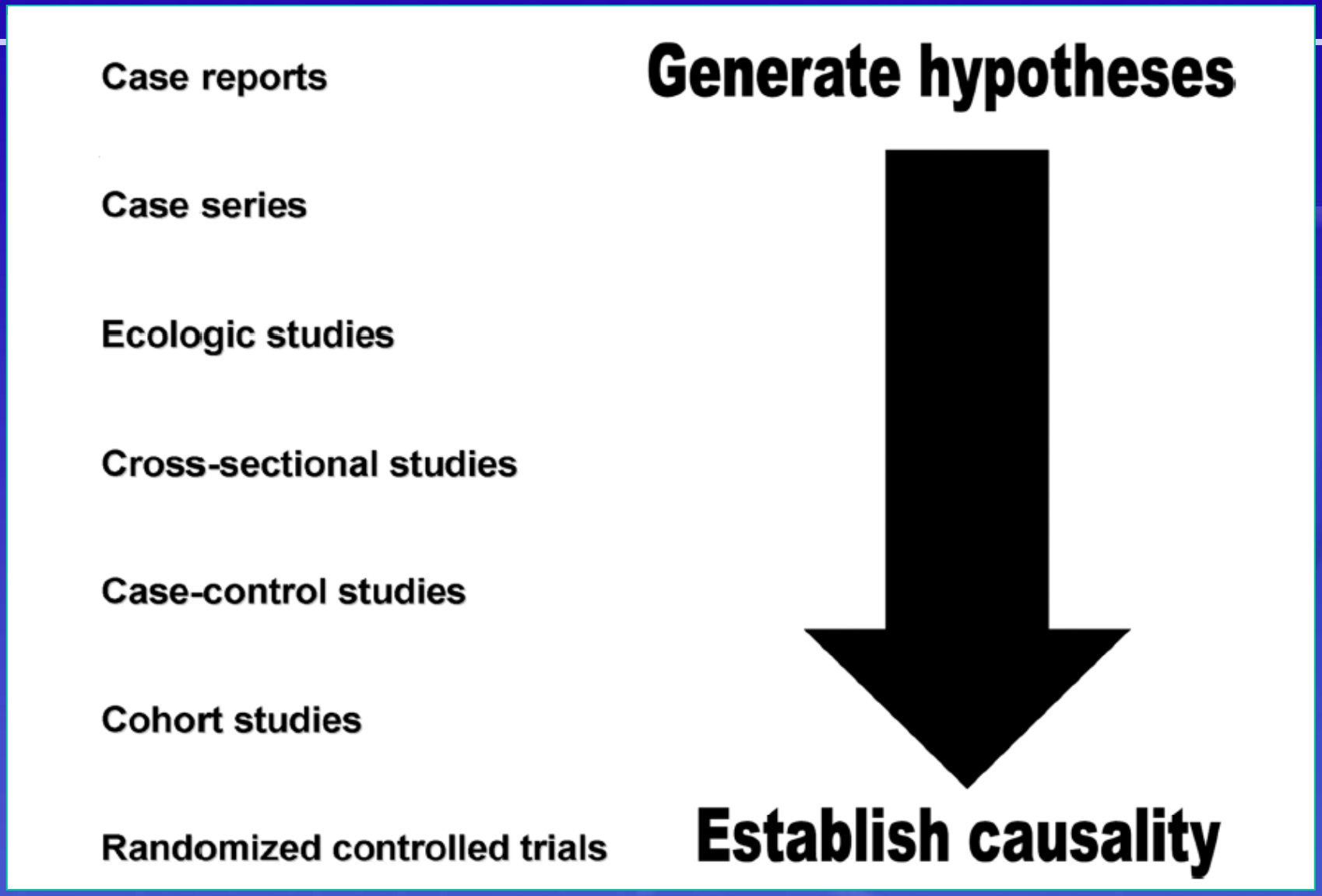
»Compare treatment modalities

»Epidemiologic Study Designs



»Grimes & Schulz, 2002 (www)

»Hierarchy of Epidemiologic Study Design



»Tower & Spector, 2007 (www)

»Observational Studies

»(no control over the circumstances)

» - Descriptive: Most basic demographic studies

» - Analytical: Comparative studies testing an hypothesis

» * cross-sectional

» (a snapshot; no idea on cause-and-effect relationship)

» * cohort

» (prospective; cause-and-effect relationship can be inferred)

» * case-control

»Epidemiologic Study Designs

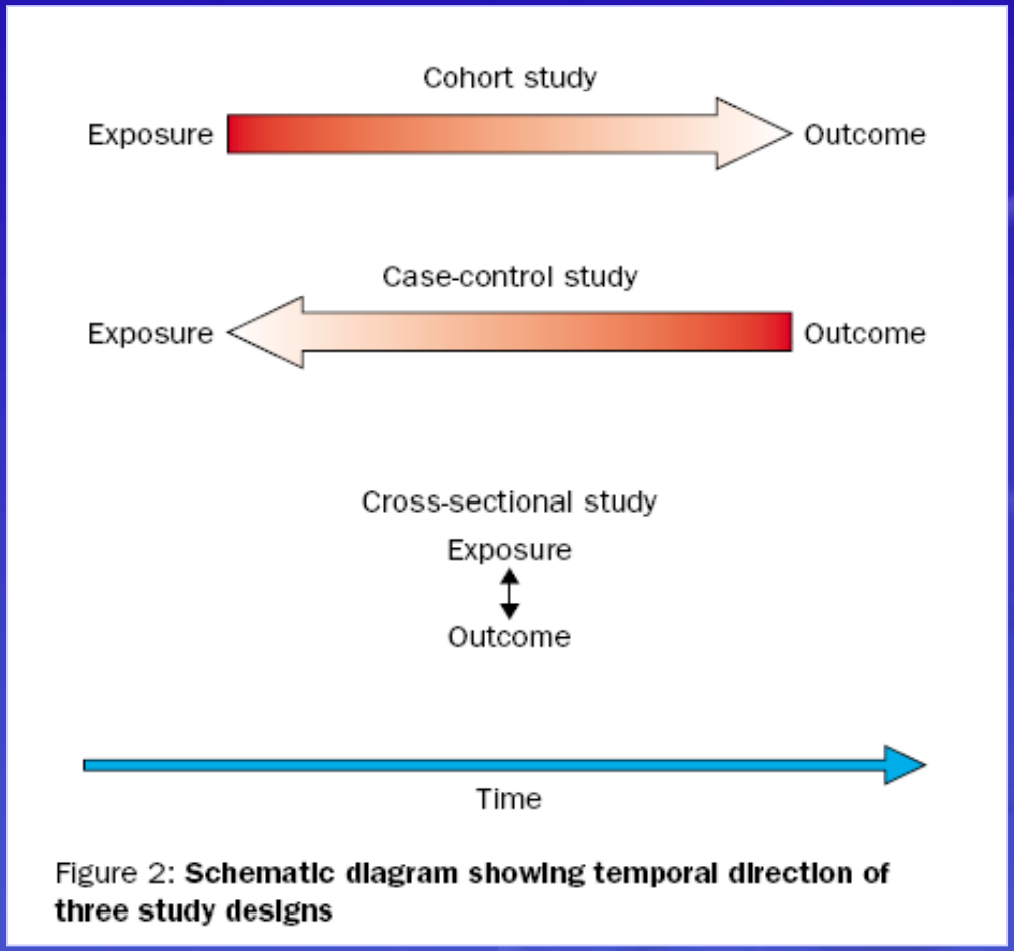


Figure 2: **Schematic diagram showing temporal direction of three study designs**

»Grimes & Schulz, 2002 (www)

»Analytical Studies

»(comparative studies testing an hypothesis)

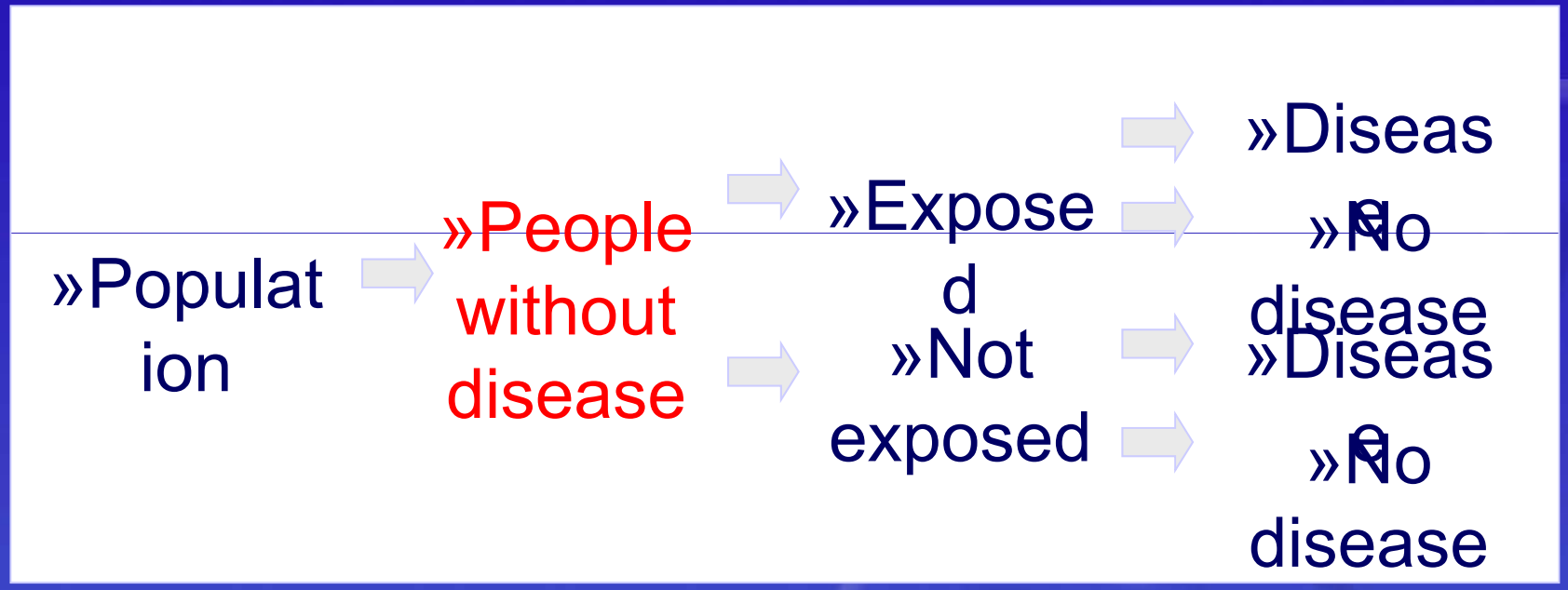
» * cohort (prospective)

» *Begins with an exposure (smokers and non-smokers)*

» * case-control (retrospective - trohoc)

» *Begins with outcome (cancer cases and healthy controls)*

»Cohort Studies



»Examples of Cohort Studies

- *»* Framingham Heart Study (www)*
- *»* NHANES Studies (www)*
- *»* MACS (www)*
- *»* Physicians' Health Study (www)*
- *»* Nurses' Health Study (www)*
- *»* ALSPAC (www)*

»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

»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, 195

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

»Source: Doll & Hill

»Presentation of data: Various exposure levels

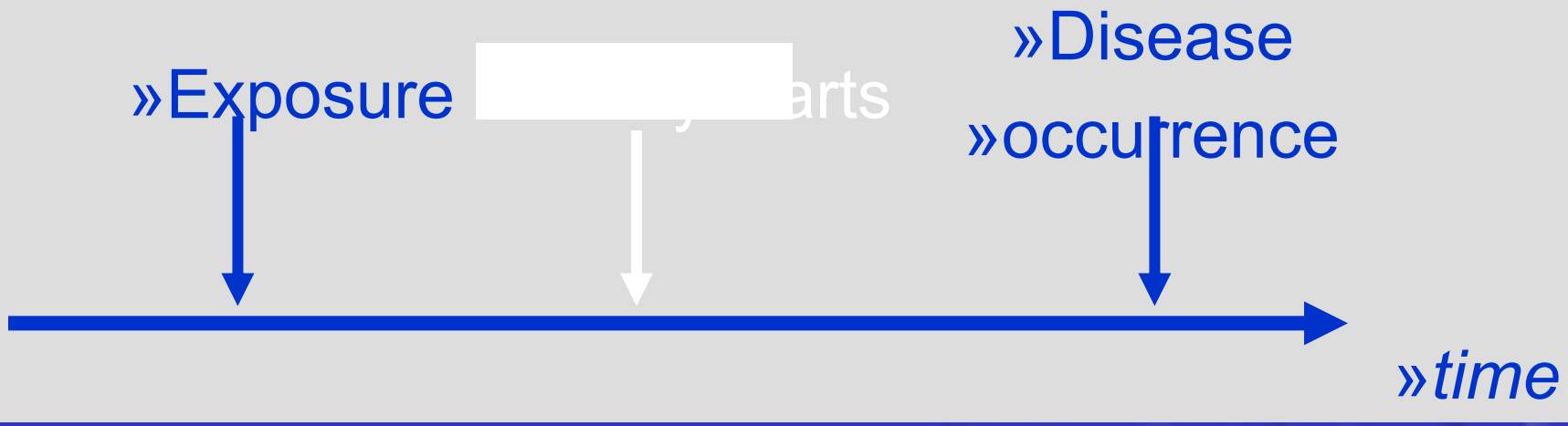
Daily number of cigarettes smoked	Person-years at risk	Lung cancer cases
> 25	25,100	57
15 - 24	38,900	54
1 - 14	38,600	22
none	42,800	3

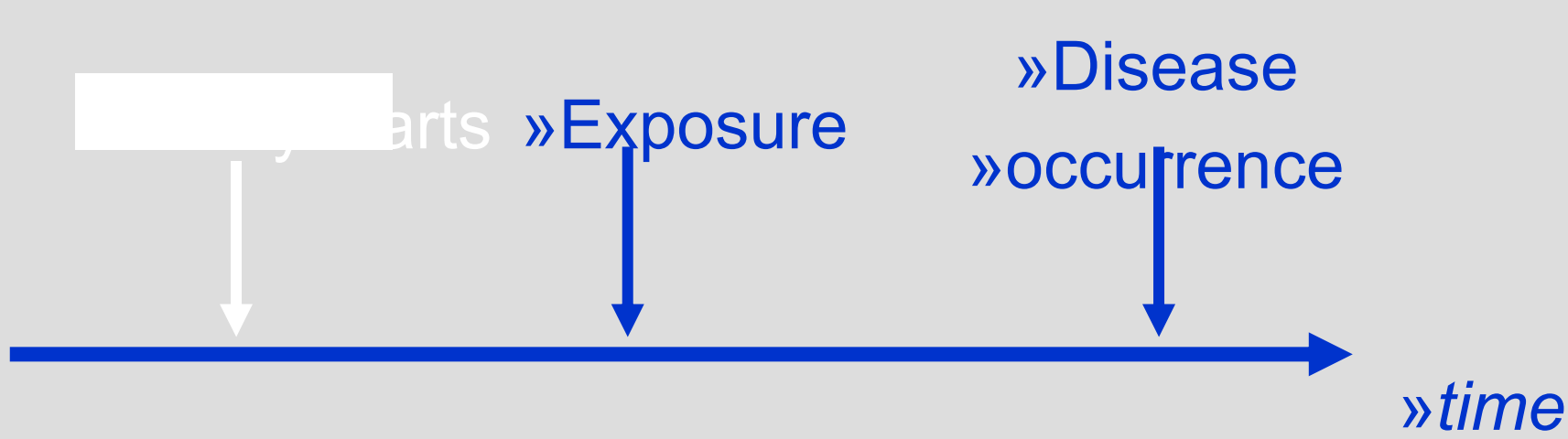
Cohort study: Tobacco smoking and lung cancer, England & Wales, 1951

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.

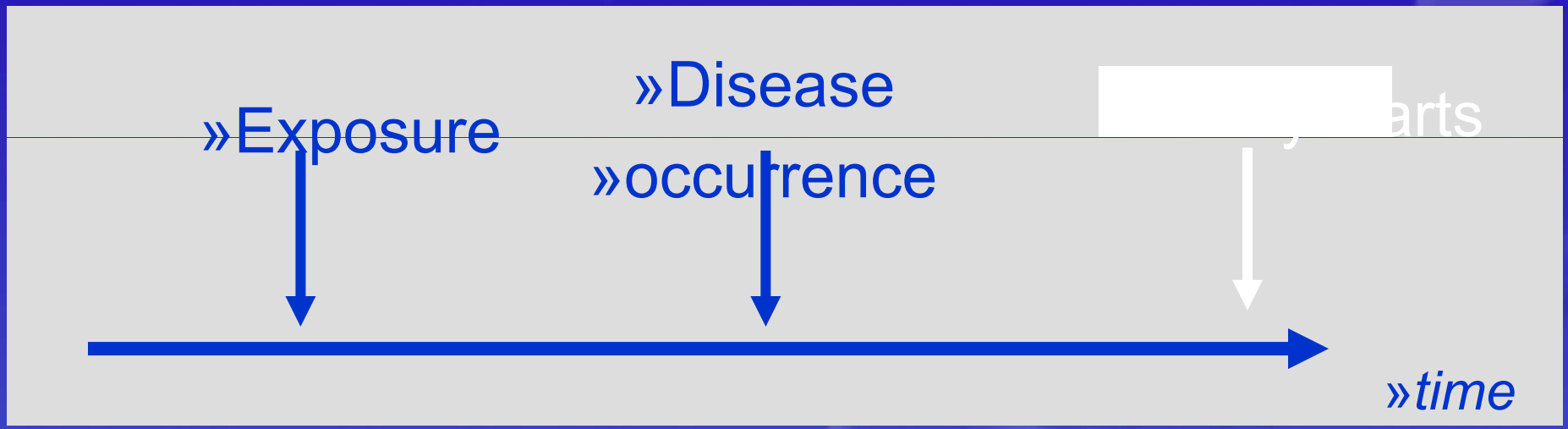
»Source: Doll & Hill

Prospective cohort study





»Retrospective cohort studies



»Cohort Studies

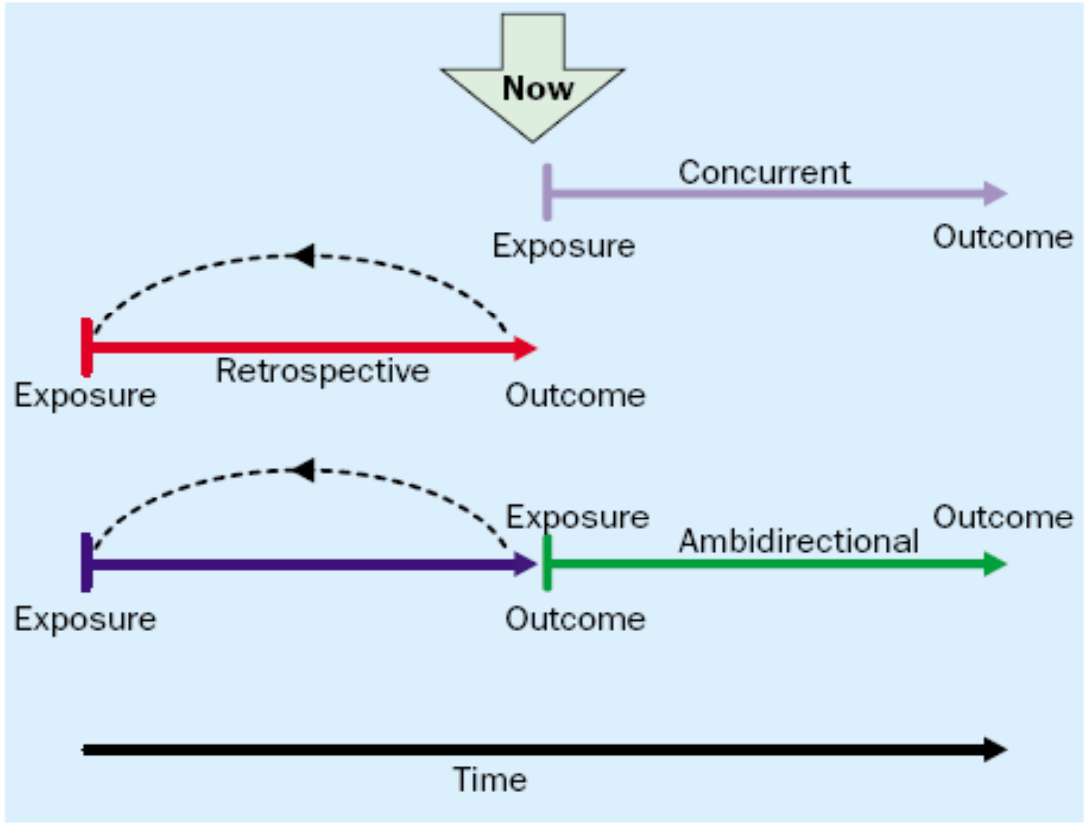


Figure 2: **Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies**

»Grimes & Schulz, 2002 (www) (PDF)

»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?

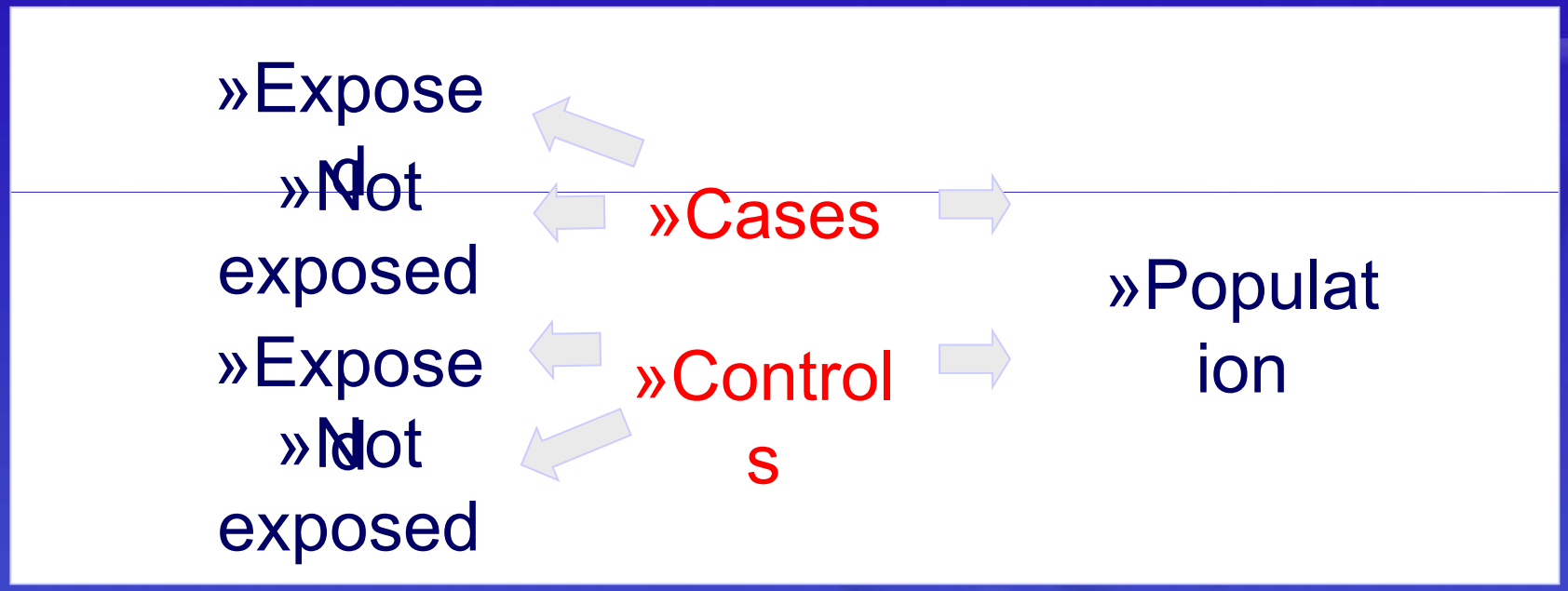
1 What efforts were made to limit loss to follow-up?
1 Was loss to follow-up similar in both groups?

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

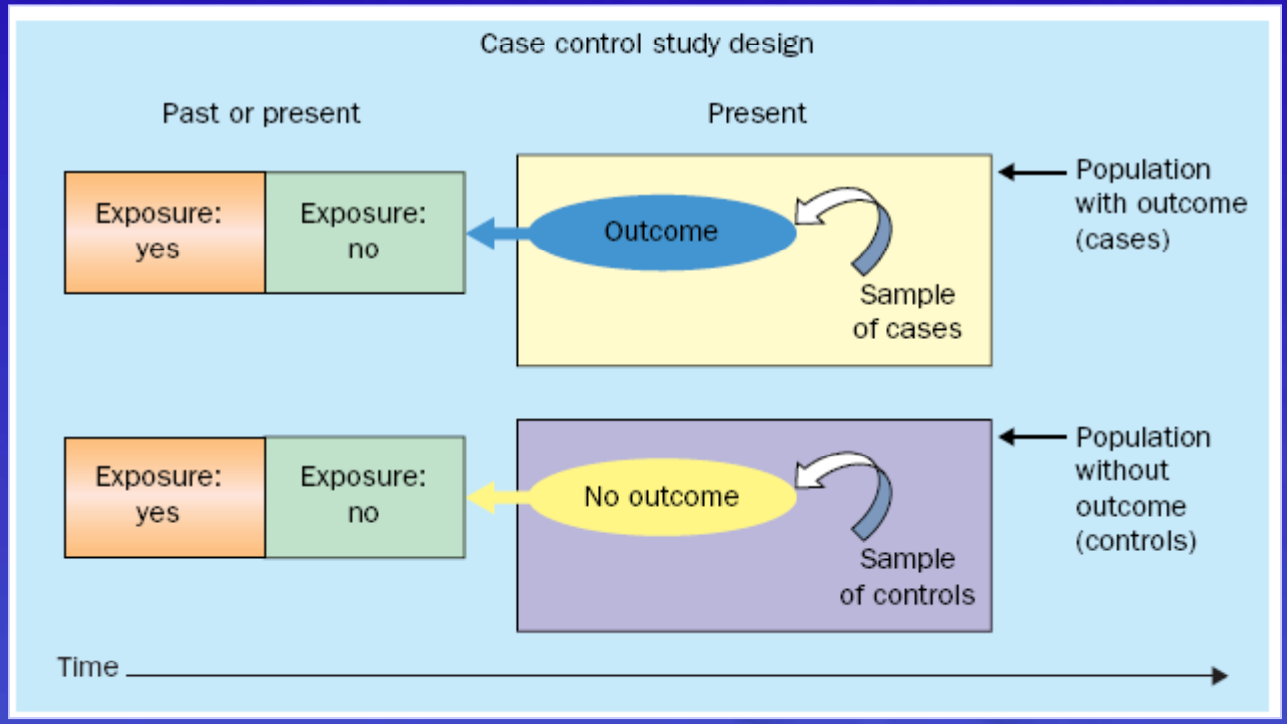
1 Did the investigators anticipate and gather information on potential confounding factors?
1 What method(s) were used to assess and control for confounding?

»Grimes & Schulz, 2002 (www) (PDF)

»Case-Control Studies



»Case-Control Studies



»Schulz & Grimes, 2002 (www) (PDF)

»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

»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.

»Schulz & Grimes, 2002 (www) (PDF)

»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?

»Grimes & Schulz, 2002 (www) (PDF)

»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.

»Grimes & Schulz, 2002 (www) (PDF)

»Epidemiologic Association / Impact Measures

»(Absolute Risk) (AR)

»Relative Risk (Risk Ratio) (RR)

»Odds Ratio (OR)

»Phi coefficient / Cramer's V / Contingency coefficient

»Attributable Fraction (AF)

»Attributable Risk (AR)

»Relative Risk Reduction (RRR)

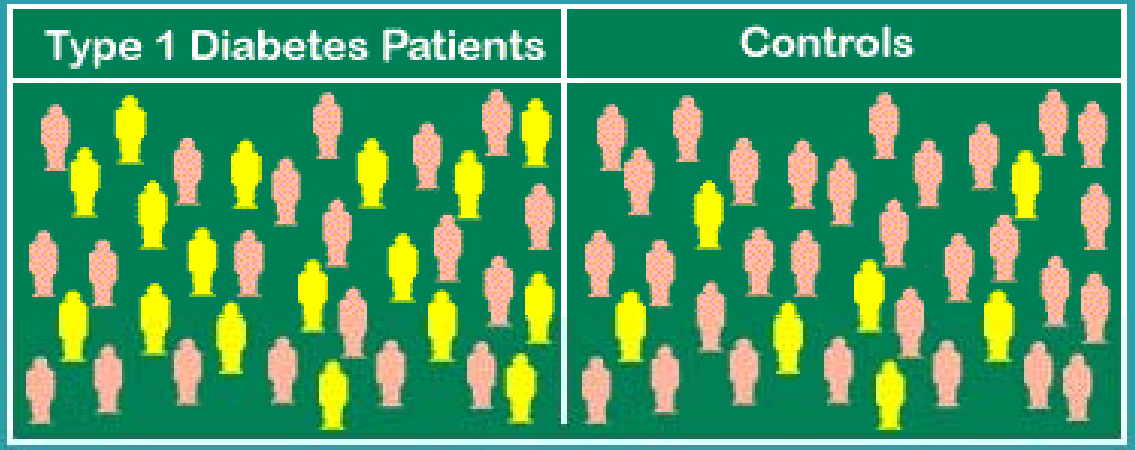
»Absolute Risk Reduction (ARR)

»Number Needed to Treat (NNT)

»Measures of test accuracy:

»Sensitivity, specificity, positive and negative predictive value (PPV, NPV)

Association Studies



Genotype	Type 1	Controls	Total
HLA DR4	17	7	24
NON-HLA DR4	20	30	50
	37	37	

$X^2_{.05} = 5.377$

$p < 0.025$

= HLA DR4

= non-HLA DR4

»Odds Ratio: 3.6
»95% CI = 1.3 to 10.4

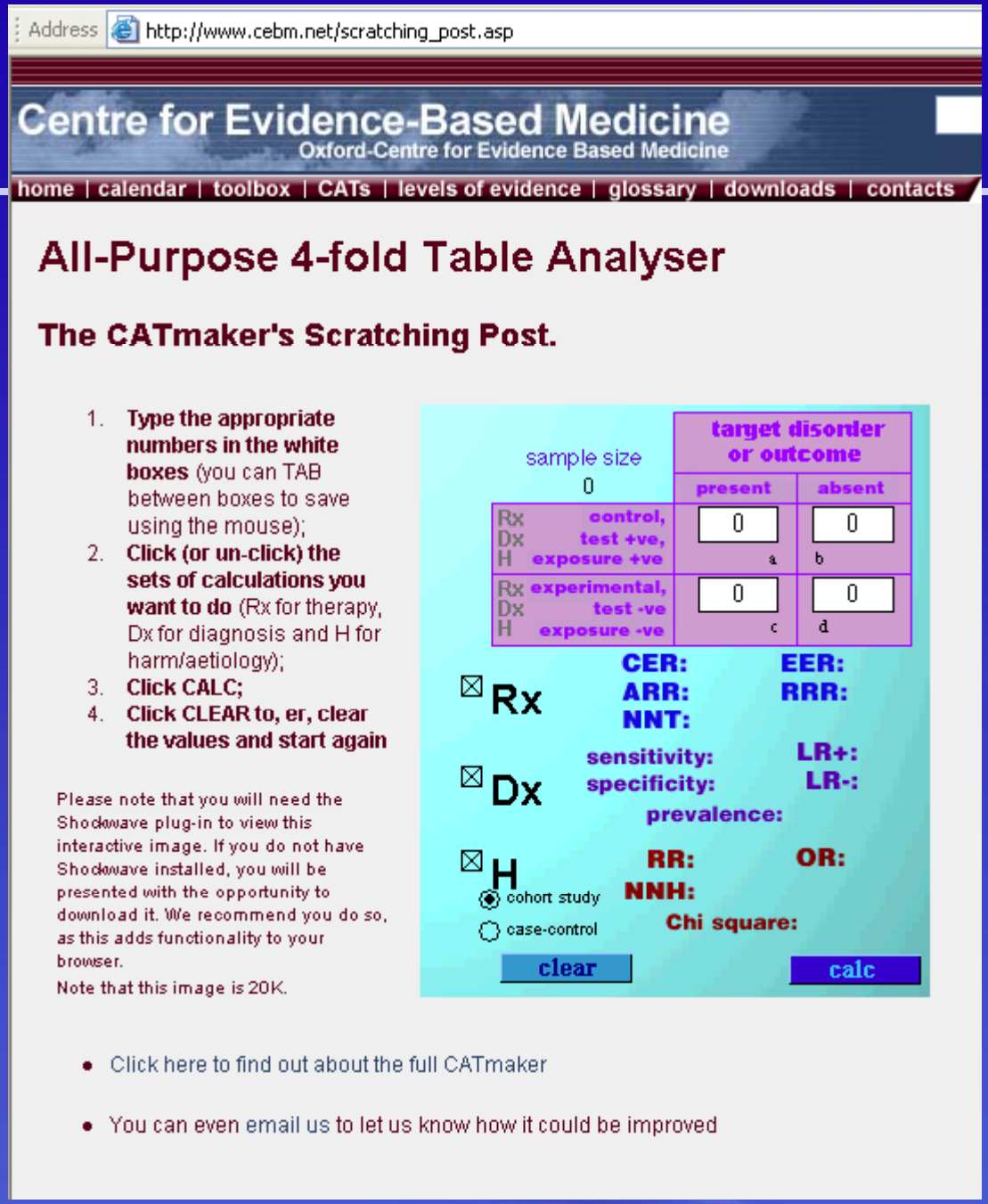
»ROCHE Genetic Education (www)

Genotype	Type 1	Controls	Total
HLA DR4	17	7	24
NON-HLA DR4	20	30	50
	37	37	

»a = 17
»b = 20
»c = 7
»d = 30

»OR = ad / bc = 17*30 / 20*7 = 3.6
»RR = (a/(a+c)) / (b/(b+d)) = (17/24)/(20/50) = 1.8

»EBM toolbox (www)
»EpiMax Table Calculator (www)



»EBM toolbox (www)

EpiMax Table Calculator

Epidemiology & Lab Statistics from Study Counts
With Chi Square, NNT & "Cost to Treat" Estimates

[For Demonstration Only-Not for Official Use]

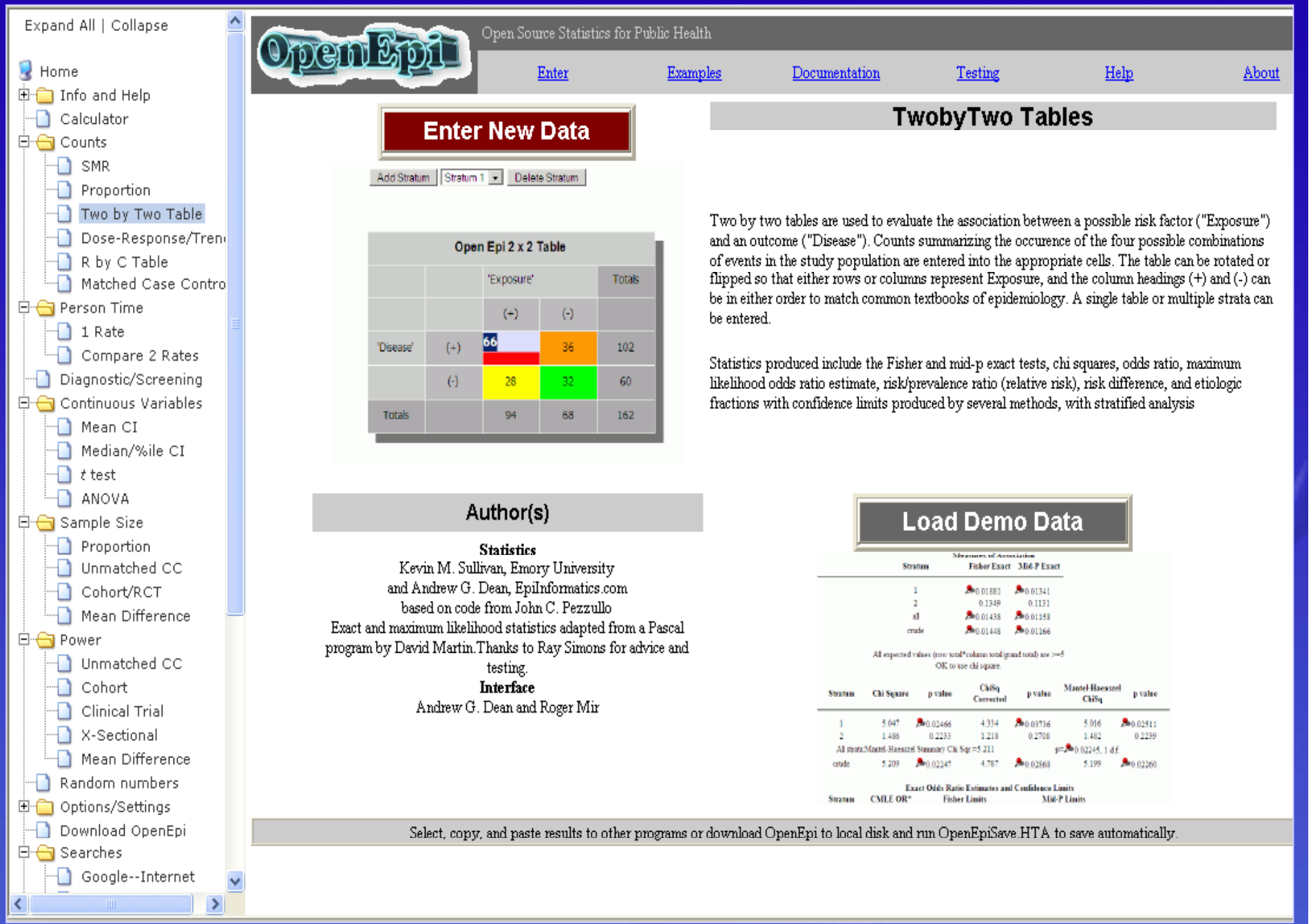
Clinical & Economic Software Solutions	
Health Decision Strategies, LLC	Princeton, New Jersey USA

Data Entry Page

Instructions: Using 2x2 study data, you can change the "Title" and fill in the four center cells in the table below (the cells in blue) and if you wish, enter a "Cost Per Person" value. Hit the "Calculate" button to see the estimated results.

(Results generated will appear in the boxes outside and below the center cells.)

		Target Disorder or Outcome		
	Analysis Title:	Present	Absent	
	TPA vs Strepto: Fiction	Case	Control	
		True Positive(a)	False Positive(b)	a + b
Rx Dx H	(•) Control Group ( ) Diag. Test positive ( ) Exposed to Risk Factor	100	900	
		False Negative(c)	True Negative(d)	c + d
Rx Dx H	-- Experimental Group -- Diag. Test negative -- Not Exposed to Risk	90	910	
	Incremental Cost Per Person (CPP) Per Duration	a + c	b + d	a+b+c+d
	$ 2000			



»Open-Epi Calculator (www)

»Epidemiologic Study Designs

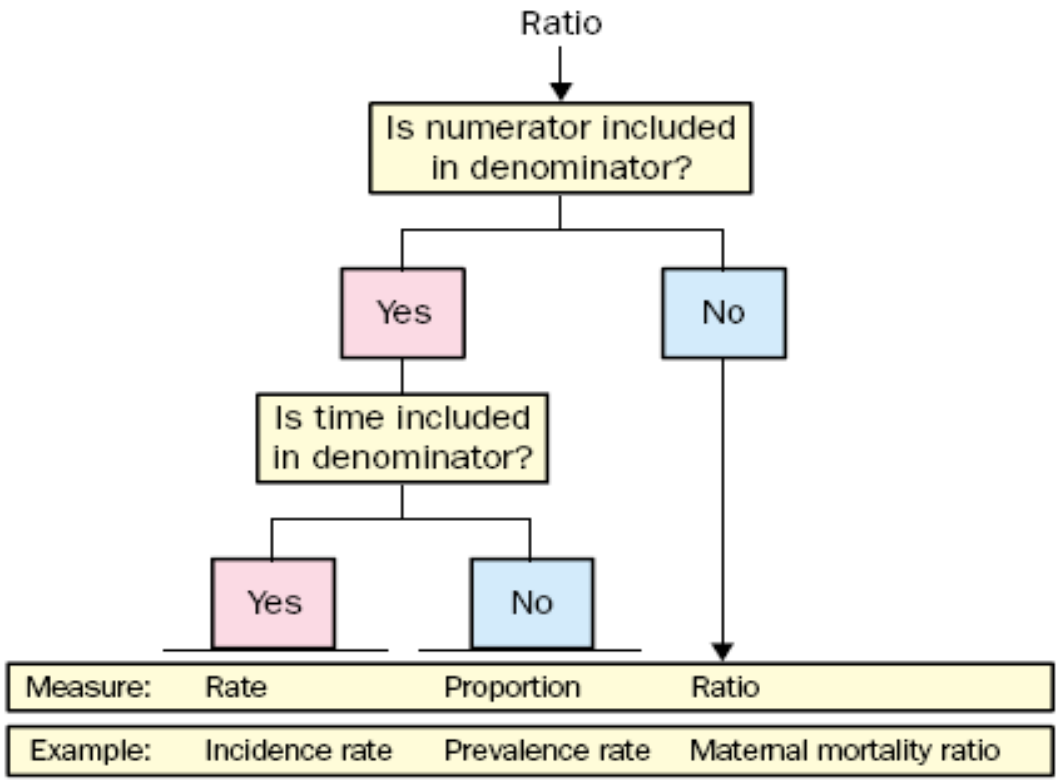
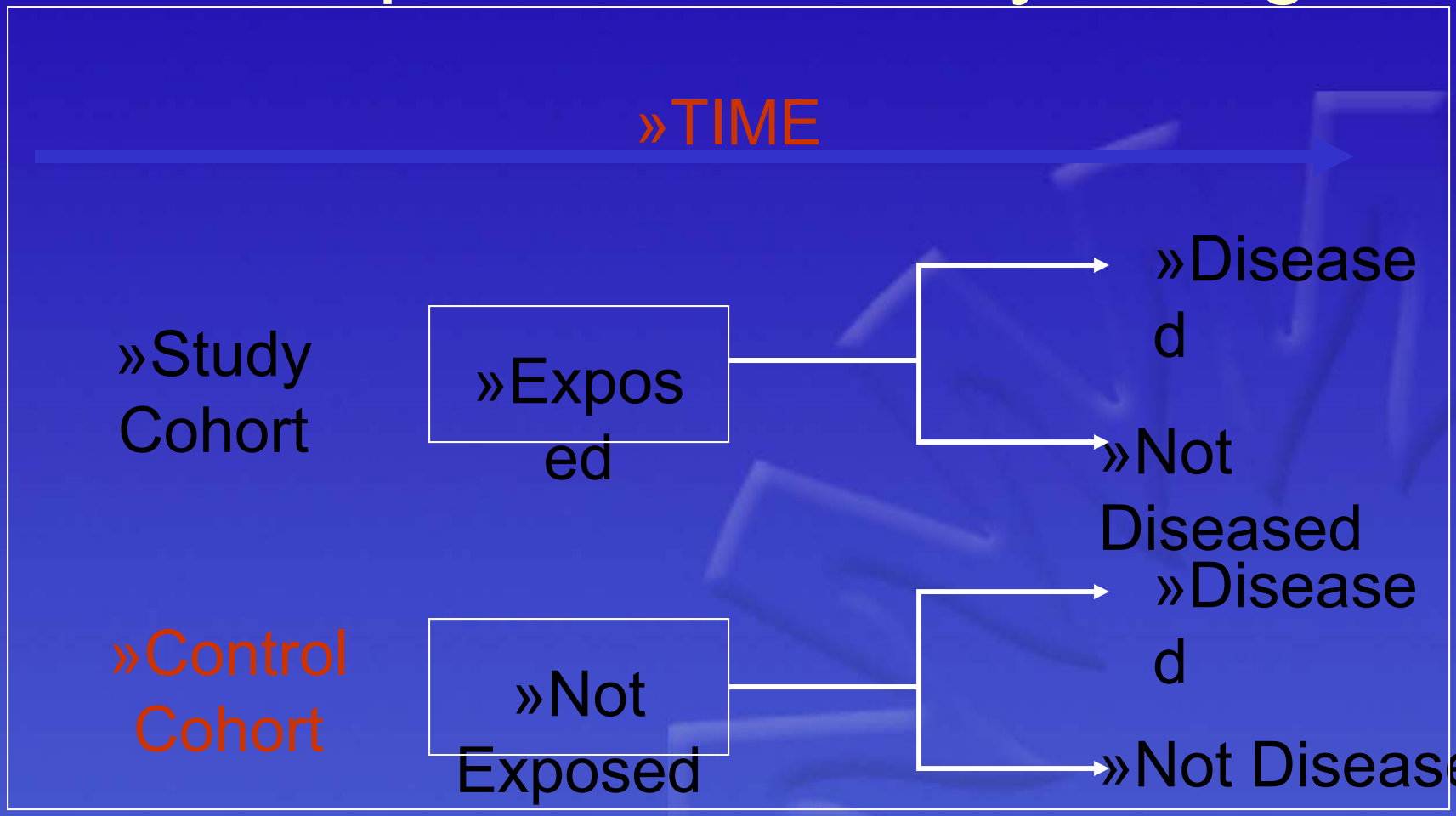


Figure 3: **Algorithm for distinguishing rates, proportions, and ratios**

»Grimes & Schulz, 2002 (www)

Multi-Sample Cohort Study Design



Selecting Comparison (Control) Groups

- **If the cohort is the general population, subjects are selected based on exposure and the comparison group is internal - from the same sample - who do not have the exposure**
- **If the cohort is based on a high risk population selected on the basis of a given exposure (e.g., Chernobyl residents, asbestos workers), external controls must be sought**
- **Sometimes both comparison groups are**

Selecting Comparison Groups (cont.)

- **If a comparison group cannot be assembled, known population rates for outcomes may be acceptable but only if they are adjusted for the exposure**
  - **Lung cancer rates are based on the population but should not be used for comparison to compare to populations with high smoking rates, such as miners. WHY?**
  - **Leukemia rates from the general population can be used to compare rates to Three Mile Island residents. WHY?**

Determining Exposure

- **Valid means of determining exposure include:**
  - **Questionnaires**
  - **Laboratory tests**
  - **Physical measurements**
  - **Special procedures**
  - **Medical records**

- **What if the exposure is chronic, such as radon or smoking?**

Measuring Disease

- **You must determine endpoints in a similar manner for both the exposed and the non-exposed**
  - **That is, procedures for disease identification must be the same for the exposed and the non-exposed**
- **Define the outcomes of interest (set diagnostic criteria)**
  - **If you are looking for multiple outcomes, each must be defined**

Measuring Disease (cont.)

- **Mortality may be ascertained from medical records, autopsy records, death certificates, physician records, or next-of-kin**
  - **Using mortality records does not allow for multiple outcomes**
- **Hospital records can be scanned for specific types of admissions**
  - **Health records of employers and schools can be monitored**
  - **Reportable diseases may be ascertained from state registries**
- **Absenteeism may be monitored with work records, self reporting, school records or household surveys**
- **Common ailments that do not usually require medical care may be monitored through self-reports, telephone surveys or calendar sheets**

Relative Risk (RR)

- **A ratio that measures the risk of disease among the exposed to the risk among the unexposed**
- **RR Numerator: Incidence rate in the exposed**
- **RR Denominator: Incidence rate in the unexposed**

Example: Calculating the Relative Risk

Exposure Status	Disease Status: CHD cases (Cases)	Disease Status: No CHD (Controls)	TOTAL
Smoker	112	176	288
Non-smoker	88	224	312

»Relative Risk »= »A/(A+B) / »C/(C+D) »= »112 / 288 / »88 / 312 »= »1.38

Example: Interpreting the Relative Risk

»Relative Risk »= »1.38

»The risk of developing CHD is 1.38 times higher for a smoker than for a nonsmoker.

»o

r

»The risk of developing CHD is 38% higher for a smoker than for a nonsmoker.

Types of Cohort Studies

- **Prospective**
  - **Exposure baseline in the present**
  - **Follow-up period: present to future**
- **Retrospective:**
  - **Exposure baseline in the past**
  - **Follow-up period: past to present**
- **Historical prospective or ambispective:**
  - **Exposure baseline in the past**

Types of Cohort Studies (cont.)

- **You may also NEST a case-control study within a cohort study**

Example:

  - **Begin with a cohort of 10,000 individuals without rheumatoid arthritis**
  - **Test for the presence of RA antigen**
  - **Assume those with RA antigen are the exposed and those without the controls**
  - **Follow for 10 years and determine the incidence of disease among both cohorts**
  - **This reduces the cost of testing**

Outcome Measures

- **Incidence in the exposed**
- **Incidence in the unexposed**
- **Relative risk**
- **Attributable risk (risk difference)**
- **Population attributable risk**
- **Attributable risk percent**
- **Population attributable risk percent**
- **Standardized mortality ratio**

Advantages of Cohort Studies

- **Temporality: Exposure precedes outcome because the cohort is disease free at baseline**
- **Efficient for studying rare exposures**
- **May be used to study multiple outcomes**
- **Allows for calculation of incidence of diseases in exposed and unexposed individuals**
- **Minimizes recall bias**

Disadvantages of Cohort Studies

- **Tend to be expensive (large sample size) and time consuming (long follow-up period)**
- **Loss to follow-up**
  - **When multiple outcomes or specific disease incidence is the outcome of interest, bias can be a serious problem**
- **Inefficient to study rare diseases**

Disadvantages of Cohort Studies (cont.)

- **Nonparticipation (selection bias) – it cannot be assumed that those who chose to participate had the same prevalence of exposures nor incidence of disease as those who did not participate**
  - **A difference in prevalence of exposure in nonparticipants will not bias the results**
  - **A difference in rate of disease among nonparticipants will bias the results**

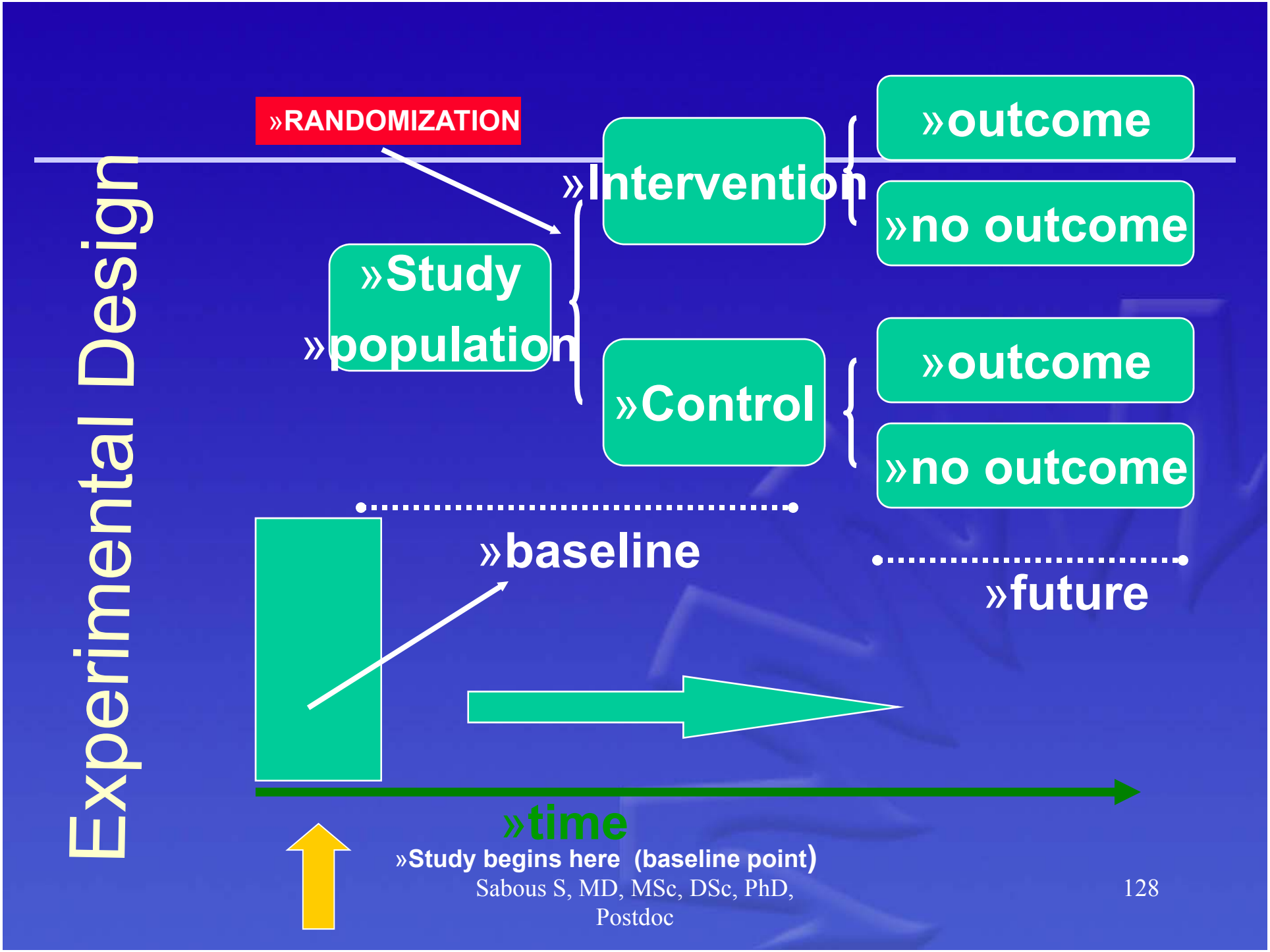
Experimental Designs

Experimental Study Design

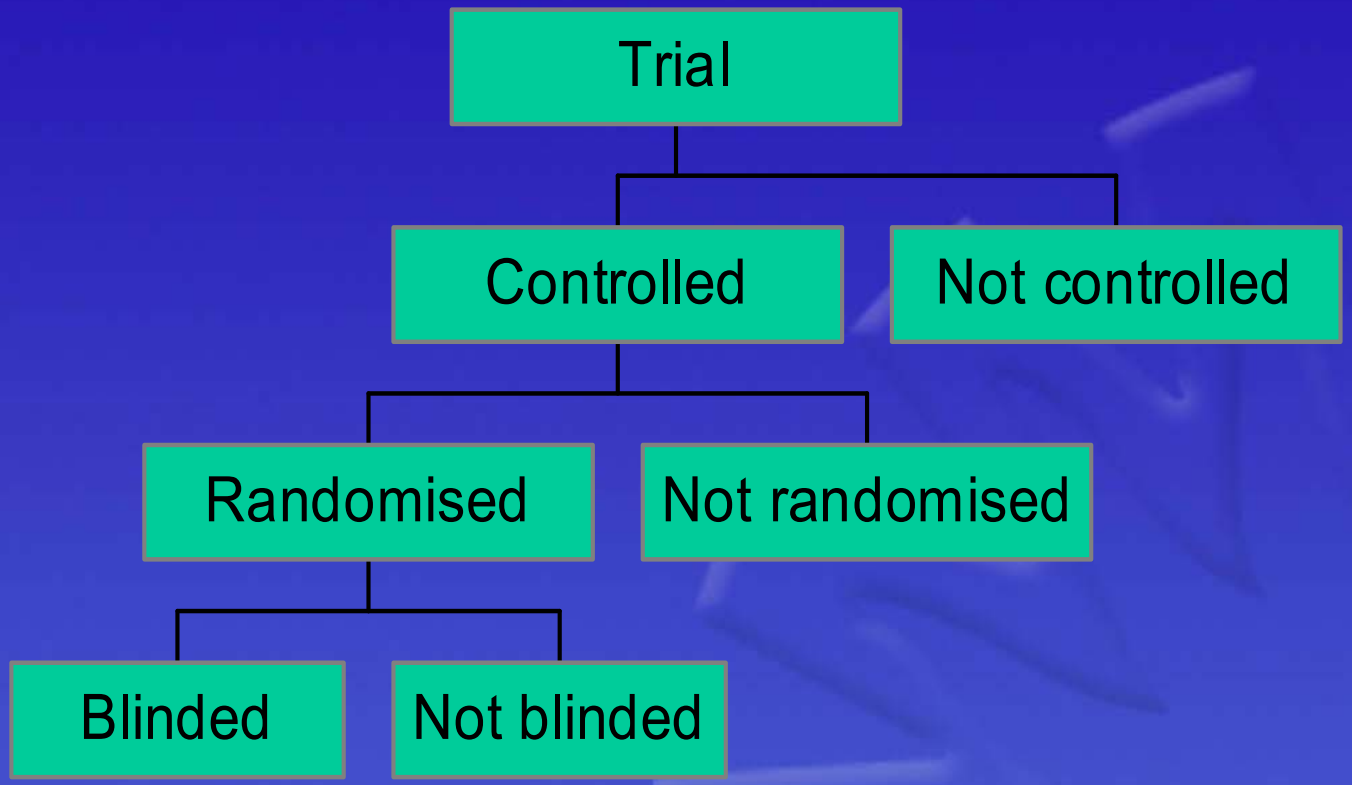
A study in which a population is selected for a planned trial of a regimen, whose effects are measured by comparing the outcome of the regimen in the experimental group versus the outcome of another regimen in the control group. Such designs are differentiated from observational designs by the fact that there is manipulation of the study factor (exposure), and randomization (random allocation) of subjects to treatment (exposure) groups.

Why Performed ?

1. Provide stronger evidence of the effect (outcome) compared to observational designs, with maximum confidence and assurance
2. Yield more valid results, as variation is minimized and bias controlled
3. Determine whether experimental treatments are safe and effective under "controlled environments" (as opposed to "natural settings" in


Experimental Design
»RANDOMIZATION
»Study
»population
»Intervention
»Control
»outcome
»no outcome
»outcome
»no outcome
»baseline
»future
»time
»Study begins here (baseline point)

Types of trials



RCT Advantages (I)

- **the "gold standard" of research designs. They thus provide the most convincing evidence of relationship between exposure and effect. Example:**
  - ***trials of hormone replacement therapy in menopausal women found no protection for heart disease, contradicting findings of prior observational studies***

RCT Advantages (II)

- **Best evidence study design**
- **No inclusion bias (using blinding)**
- **Controlling for possible confounders**
- **Comparable Groups (using randomization)**

RCT Disadvantages

- **Large trials (may affect statistical power)**
- **Long term follow-up (possible losses)**
- **Compliance**
- **Expensive**
- **Public health perspective ?**
- **Possible ethical questions**

Choice of Design (I)

Depends on:

- Research Questions
- Research Goals
- Researcher Beliefs and Values
- Researcher Skills
- Time and Funds

Choice of design (II)

It is also related to:

- **Status of existent knowledge**
- **Occurrence of disease**
- **Duration of latent period**
- **Nature and availability of information**
- **Available resources**

Comparing study designs

- **Theme**
- **Ease**
- **Timing**
- **Maintenance and continuity**
- **Costs**
- **Ethics**
- **Data utilisation**
- **Main contribution**
- **Observer bias**
- **Selection bias**
- **Analytic output**

Overlap in the conceptual basis of quantitative study designs

- The cross-sectional study can be repeated
- If the same sample is studied for a second time i.e. it is followed up, the original cross-sectional study now becomes a cohort study.
- If, during a cohort study, possibly in a subgroup, the investigator imposes an intervention, a trial begins.
- Cohort study also gives birth to case-control studies, using incident cases (nested case control study).
- Cases in a case-series, particularly a population based one, may be the starting point of a case-control study or a trial.
- Not every epidemiological study fits neatly into one of the basic designs.

Conclusion (I)

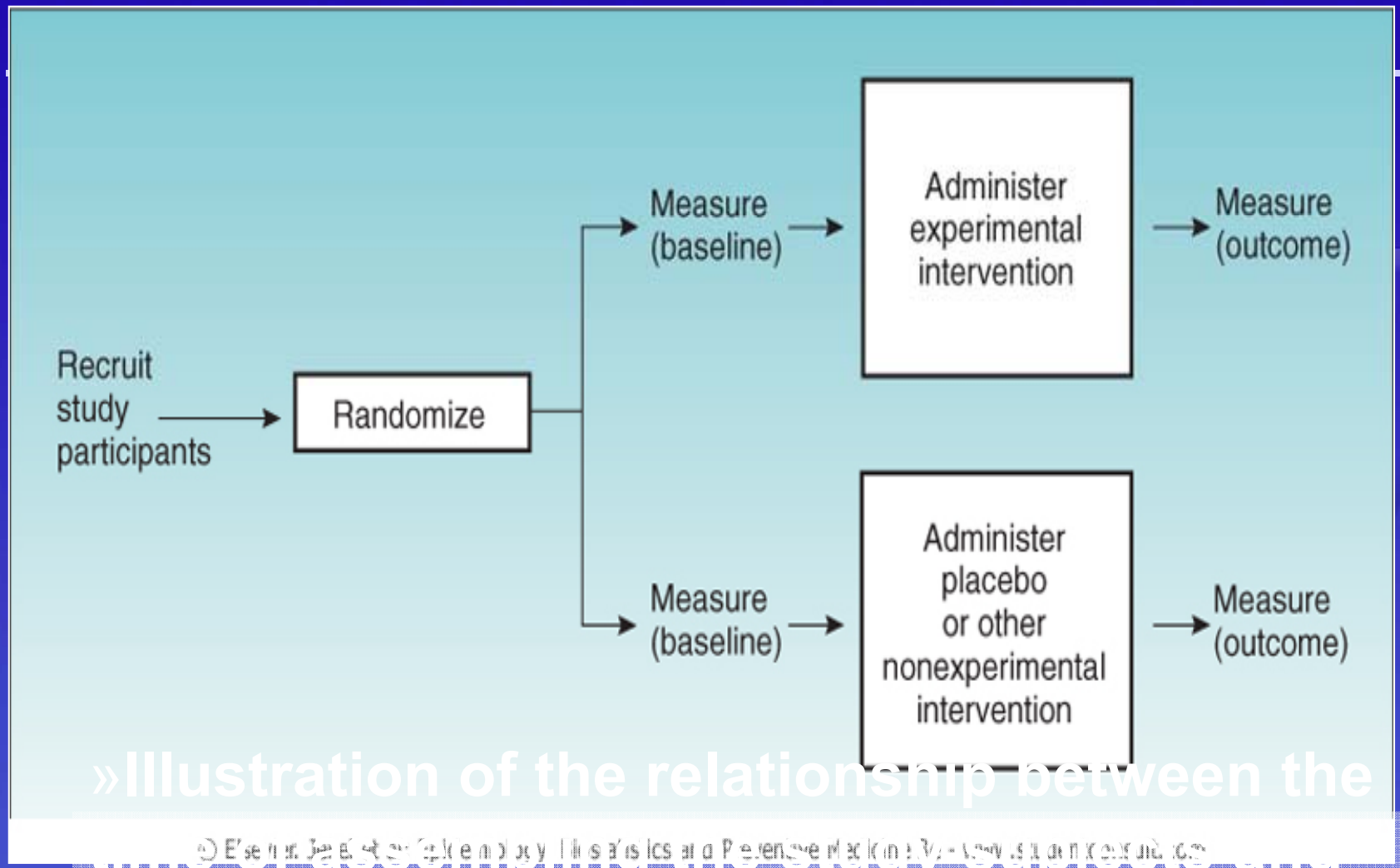
- **Qualitative designs are complementary to quantitative designs, are important in study of social determinants of health problems**
- **Quantitative designs have a common goal to understand the frequency and causes of health-related phenomena**
- **Seeking causes starts by describing associations between exposures (causes) and outcomes**

Conclusion (II)

- Case-series is a coherent set of cases of a disease (or similar problem).
- Cases are compared with reference group, we have a case control study
- In a population studied at a specific time and place (a cross-section) the primary output is prevalence data, though association between risk factors and disease can be generated.
- In cross-sectional studies, we are looking for both exposure and outcome
- In case-control studies, we know the outcome, looking for the exposure
- In cohort studies, we know the outcome, following up looking for the outcome in question

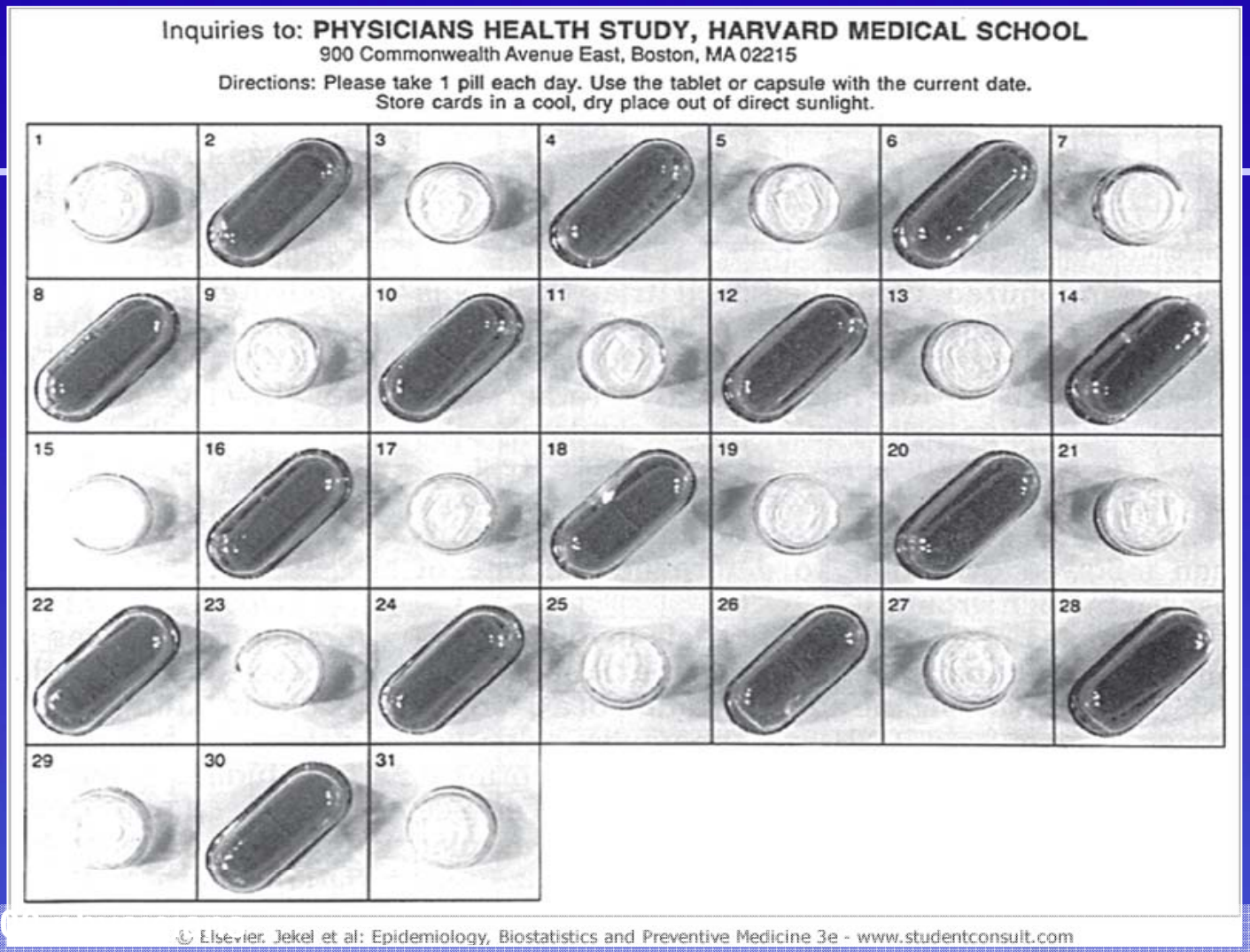
Conclusion (III)

- **If the population in a cross-sectional survey is followed up to measure health outcomes, this study design is a cohort study.**
- **If the population of such a study are, at baseline, divided into two groups, and the investigators impose a health intervention upon one of the groups the design is that of a trial.**
- **Studies based on aggregated data are commonly referred to as ecological studies.**
- **Mostly, ecological studies are mode of analysis, rather than a design.**
- **Interpretation and application of data are easier when the relationship between the population observed and the target population is understood**
- **RCTs represent the "gold standard" of research designs. They thus provide the most convincing evidence of relationship between exposure and effect..**



»Illustration of the relationship between the time of assembling the study subjects and the time of data collection

»in an RCCT and an RCFT.



»22,0

»Aspirin to reduce cardiovascular disease and beta carotene to prevent cancer.

»To have true blinding, the nonexperimental treatment must appear identical (e.g., in size, shape, color, taste) to the experimental treatment

Randomized Controlled Clinical Trials

- Blinding is impossible and unethical:

1. **Surgical intervention**
2. **Intervention were the best available**
3. **Prenatal care**

Problems of RCCT

- **Lost to follow-up** (for various reasons)
- **Therapy changes** (due to side effects)
- **Publication bias** (only positive results are publishing)

Randomized Controlled Field Trials

- An RCFT is similar to an RCCT except that the intervention in an RCFT is **preventive** rather than therapeutic, and usually it is done in the **community**.
- Appropriate subjects are randomly allocated to receive the preventive measure (e.g., **a vaccine or an oral drug**) or to receive the placebo (e.g., **an injection of sterile saline or an inert pill**).
- They are followed over time to determine the rate of disease in each group.

RCCT & RCFT

- **Disadvantages:**

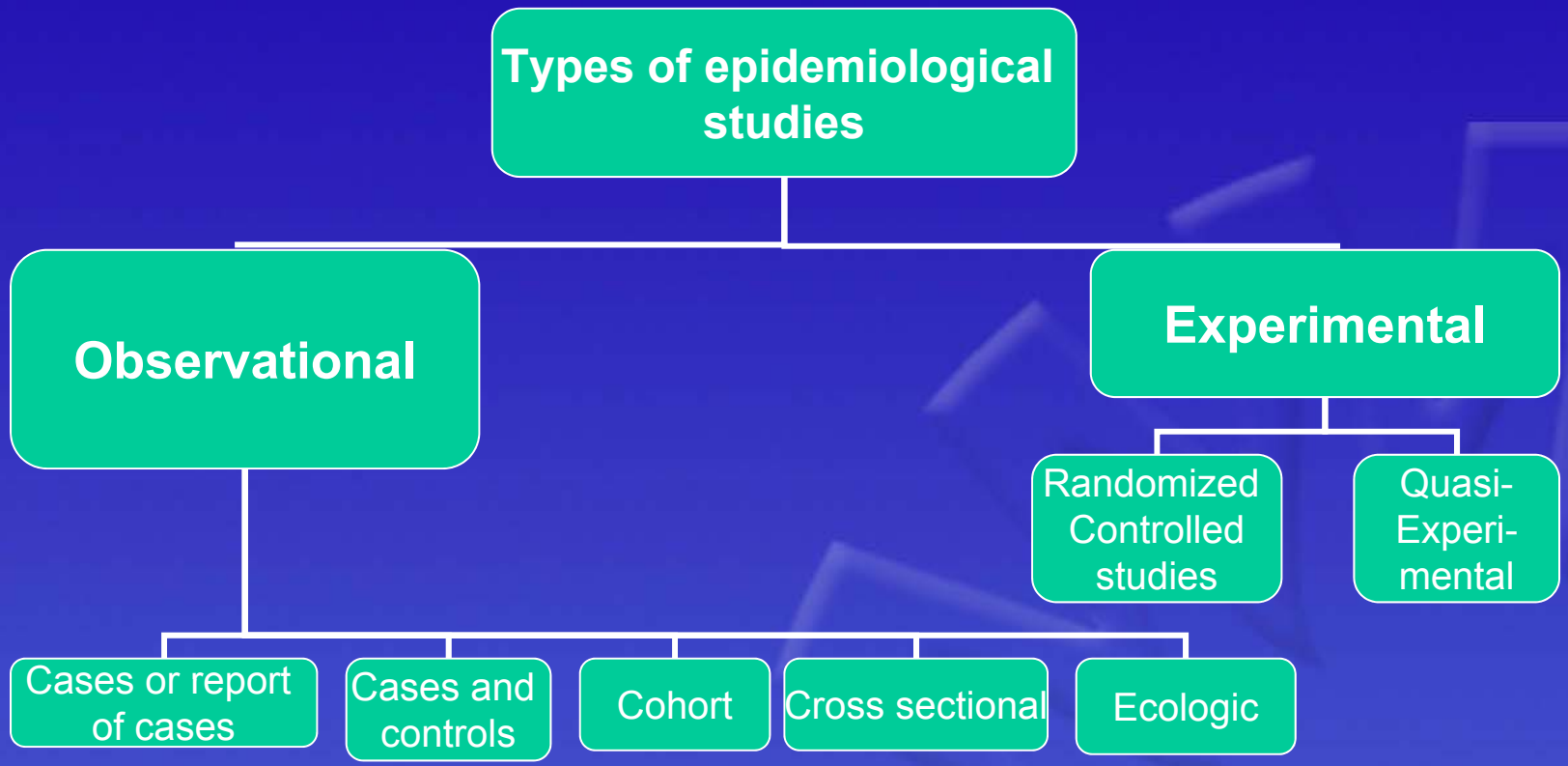
1. The results may take a **long time** to obtain
2. Has to do with **external validity**

(which is the ability to generalize the findings to other groups in the population as opposed to **internal validity**, which concerns the validity of results for the persons in the study)

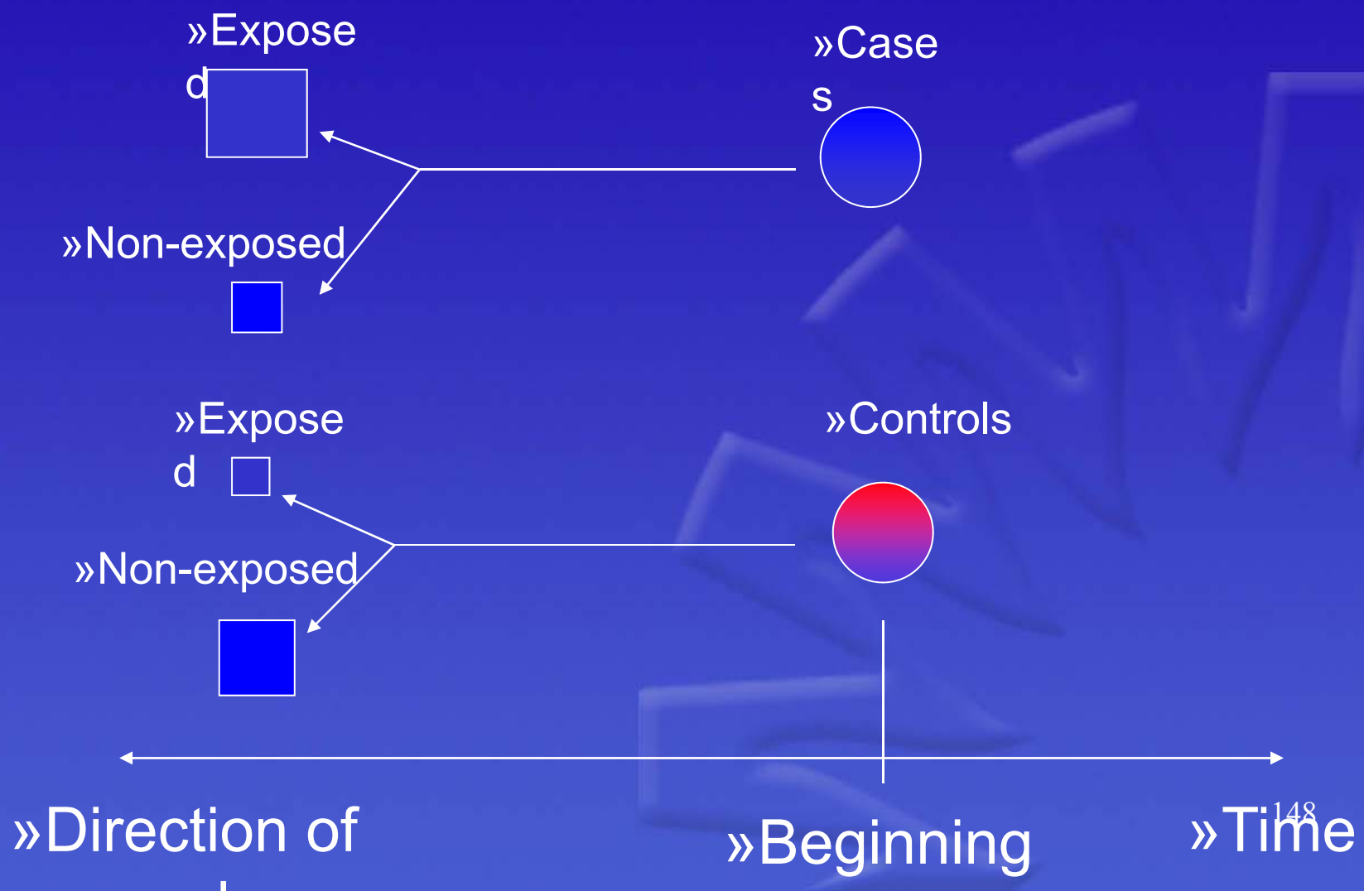
Headlines

- **Epidemiological research**
- **Classification of designs**
- **Qualitative methods**
- **Quantitative methods**
- **Choice of design**

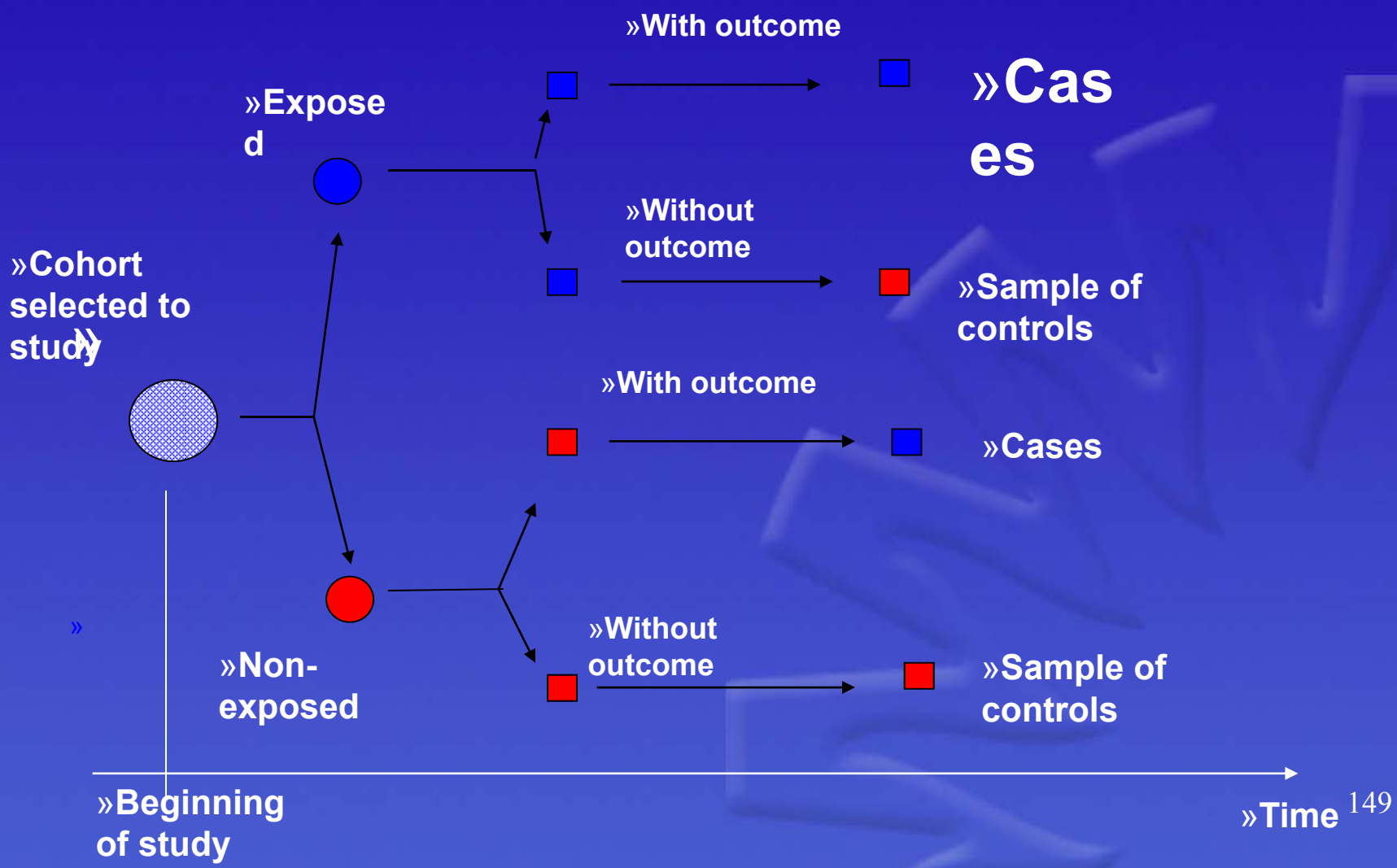
Types of studies



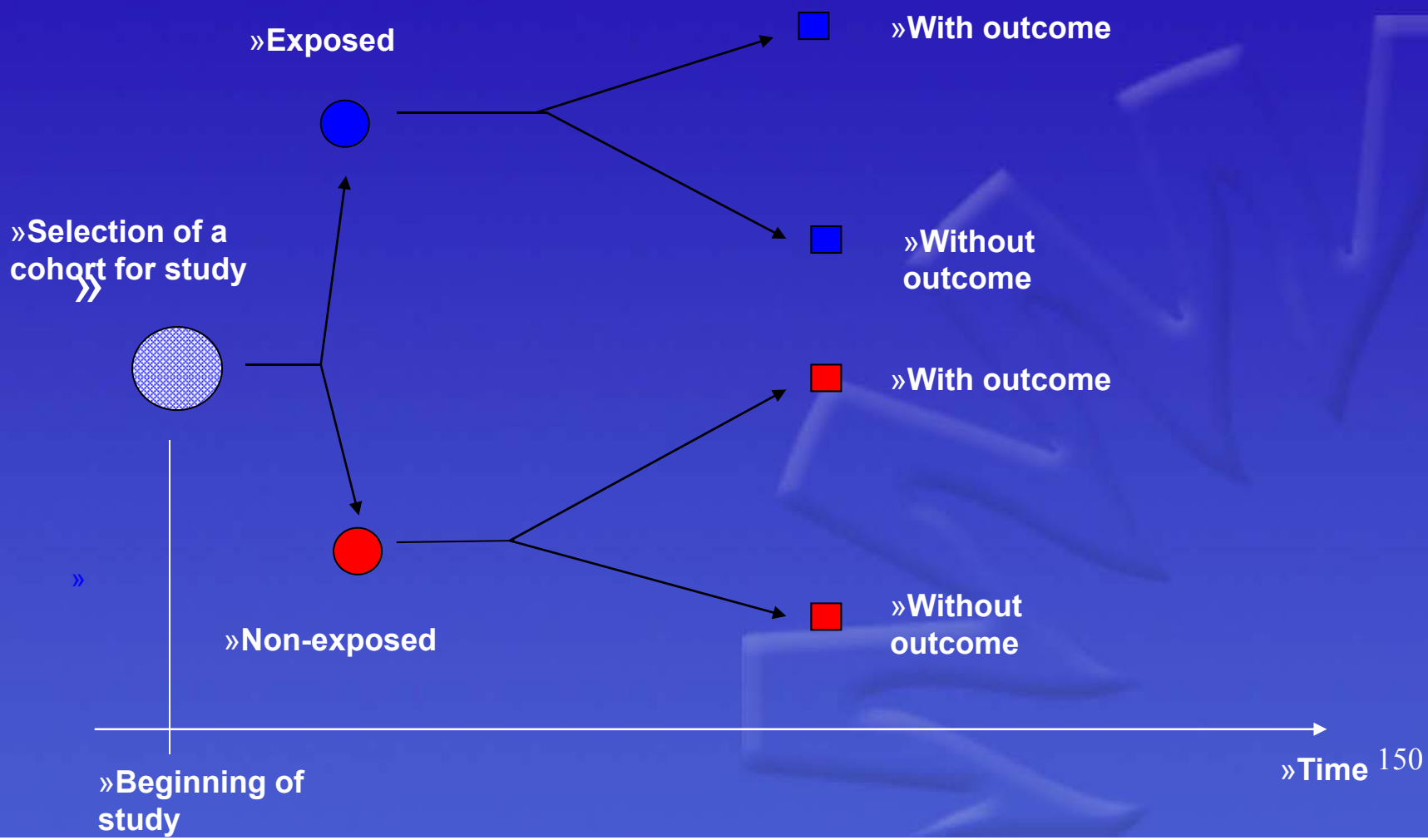
Cases and controls studies



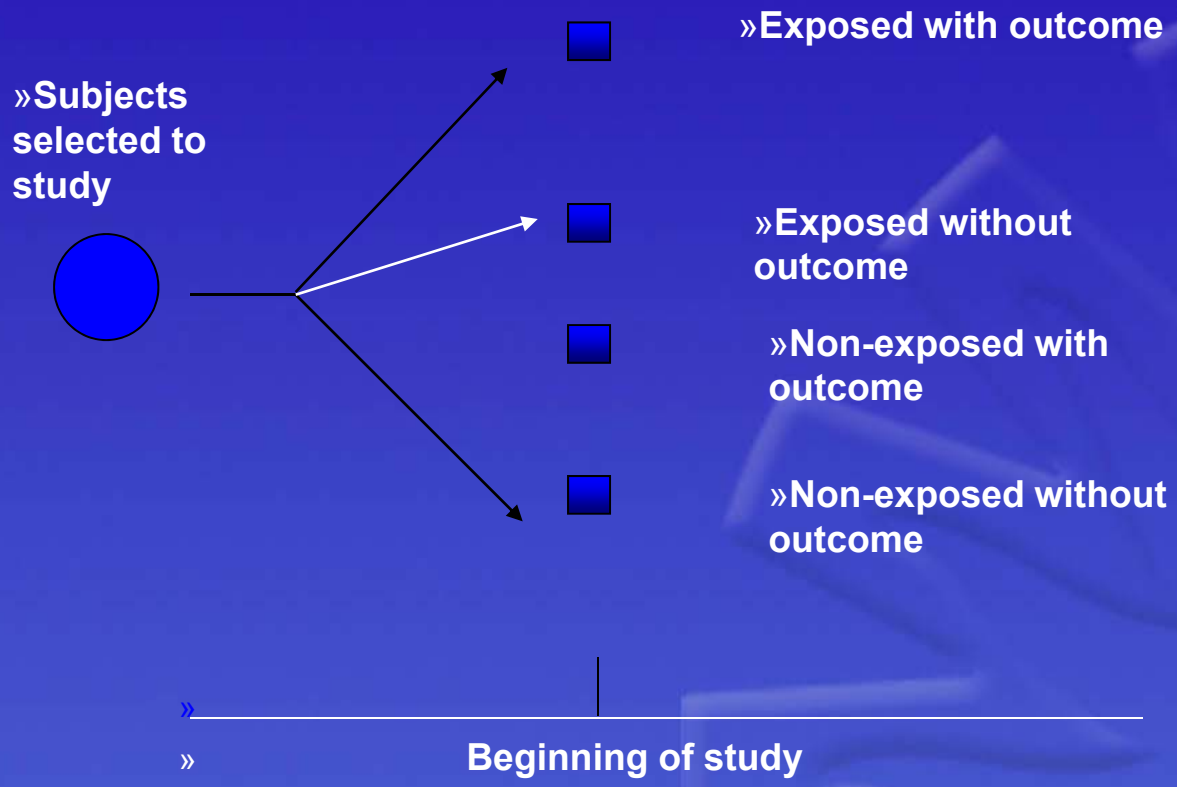
»Nested cases and controls studies



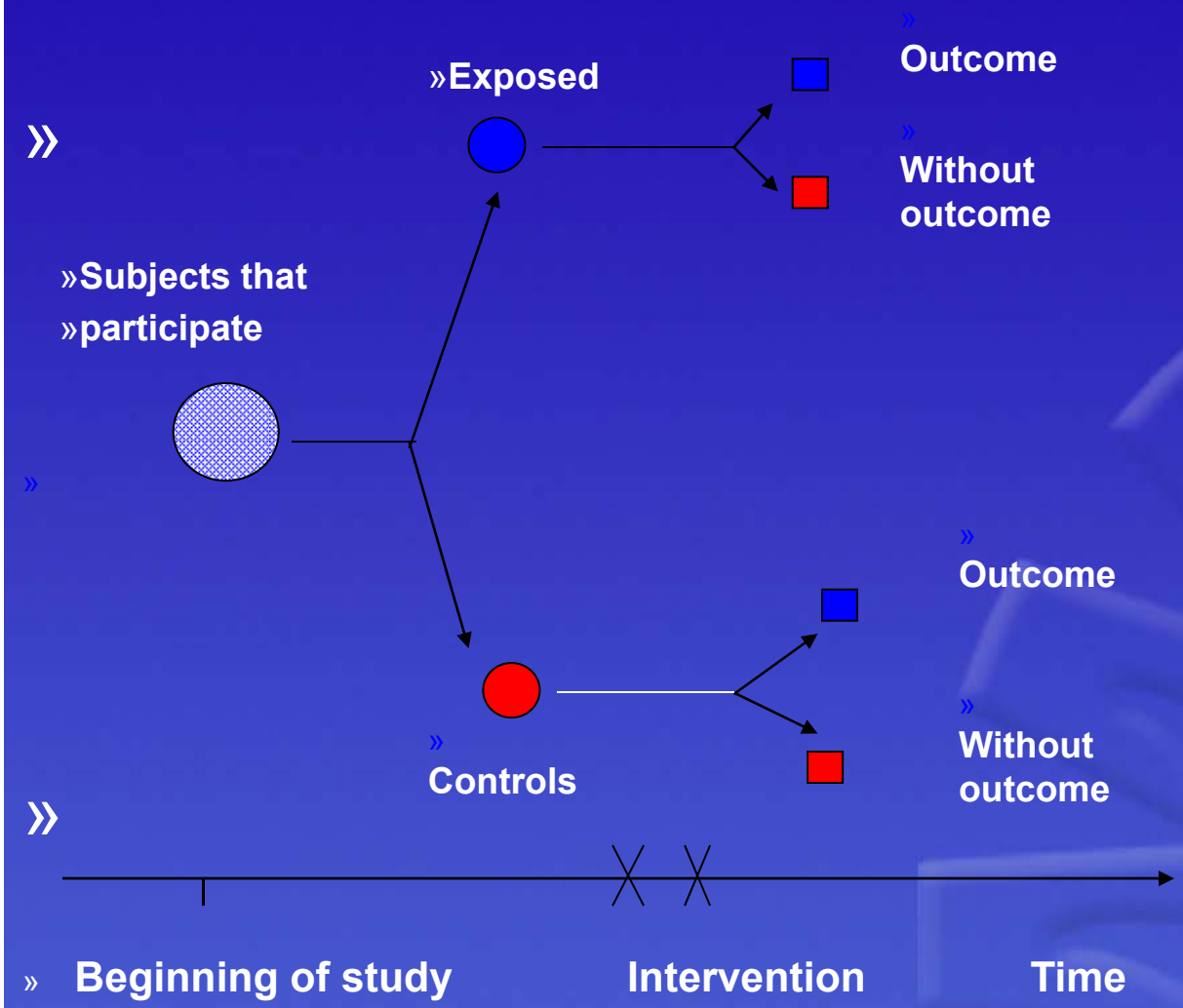
Cohort studies



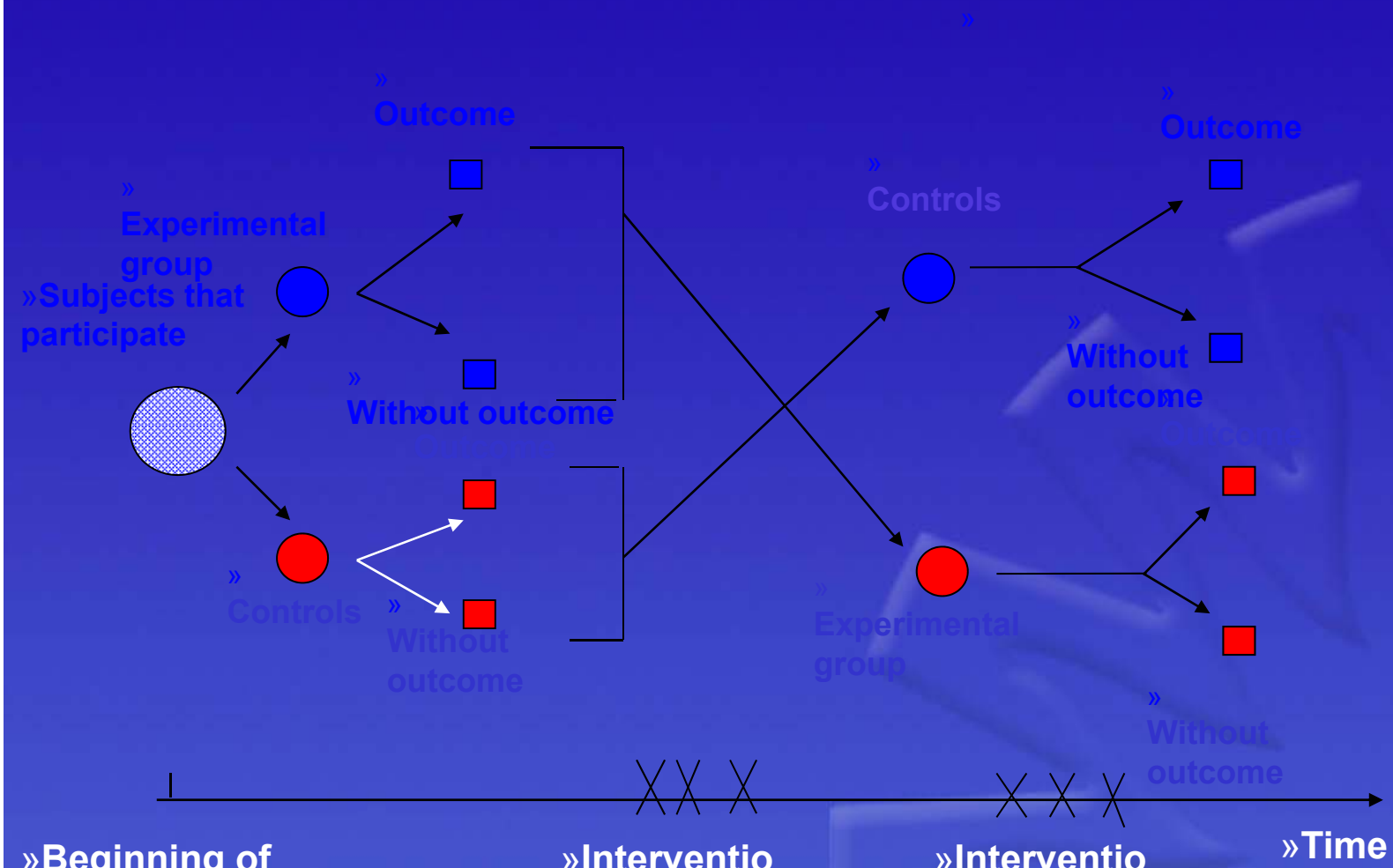
Cross sectional studies



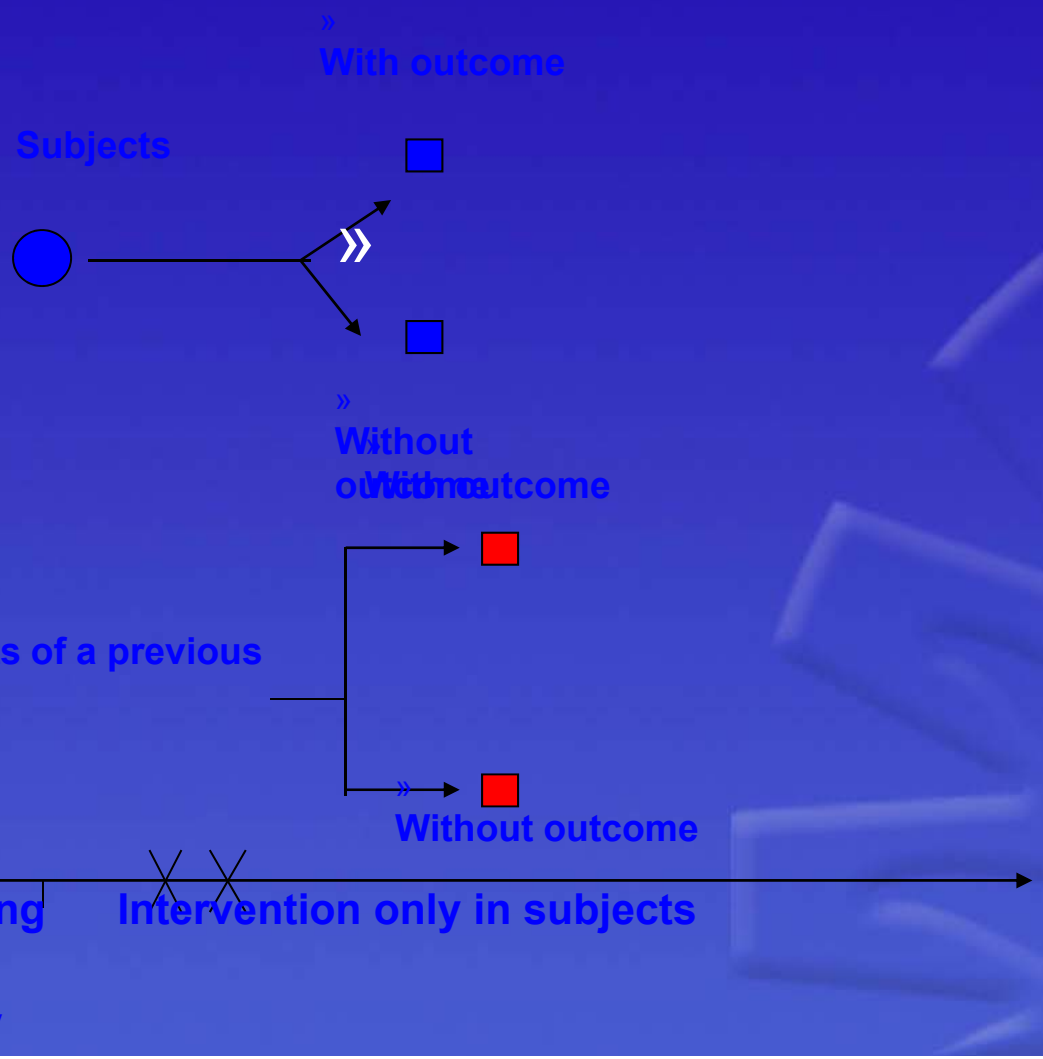
Experimental studies



Experimental studies



Experimental studies



Thank You