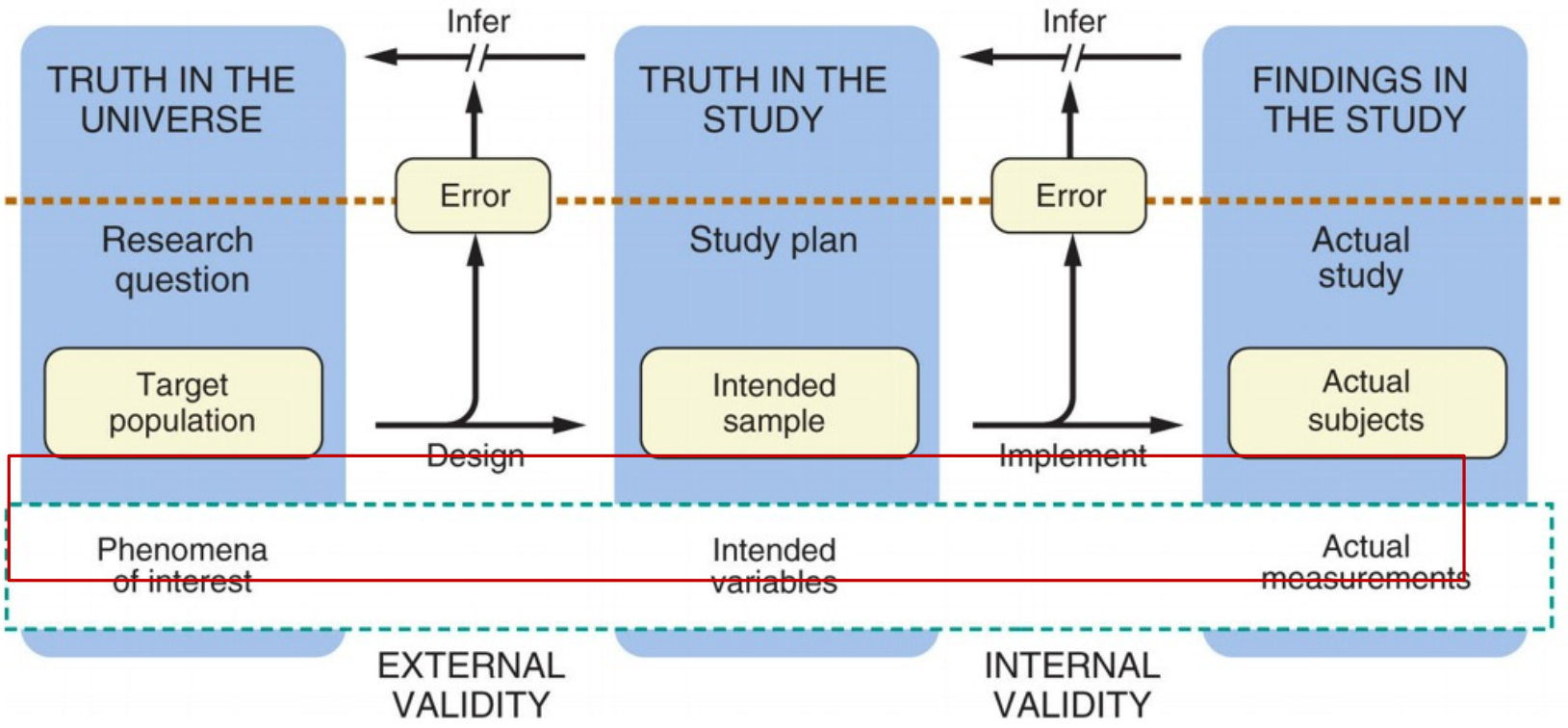
Grant Writing Seminar Series
SESSION 4:

Observational Studies

Feng Liu-Smith, PhD
Slides modified from Simonne Nouer, MD, PhD

General Idea: know your goals and weakness of studies



Source: Hulley, Stephen B.. Designing Clinical Research. LWW. Kindle Edition

Outline

- **Overview of Epidemiological Study Designs**

- **Descriptive Studies**
- Cross-Sectional
  - Design; Analytical approach; Strengths; Weakness
  - Random error, Systematic error, and Confounding
- **Observational Studies**
  - Cohort Study
    - Design; Analytical

  - Case-Control Study (Dr. Zhao)

Two Types of Epidemiology

Descriptive

Describe disease patterns

1. To monitor public's health
2. To evaluate success of intervention programs
3. To generate hypotheses about causes of disease

Identify and count cases of disease in populations and conduct simple studies

- *Case Report*
- *Case Series*
- *Cross-Sectional Study*
- *Ecologic Study*

Analytic/ Scientific

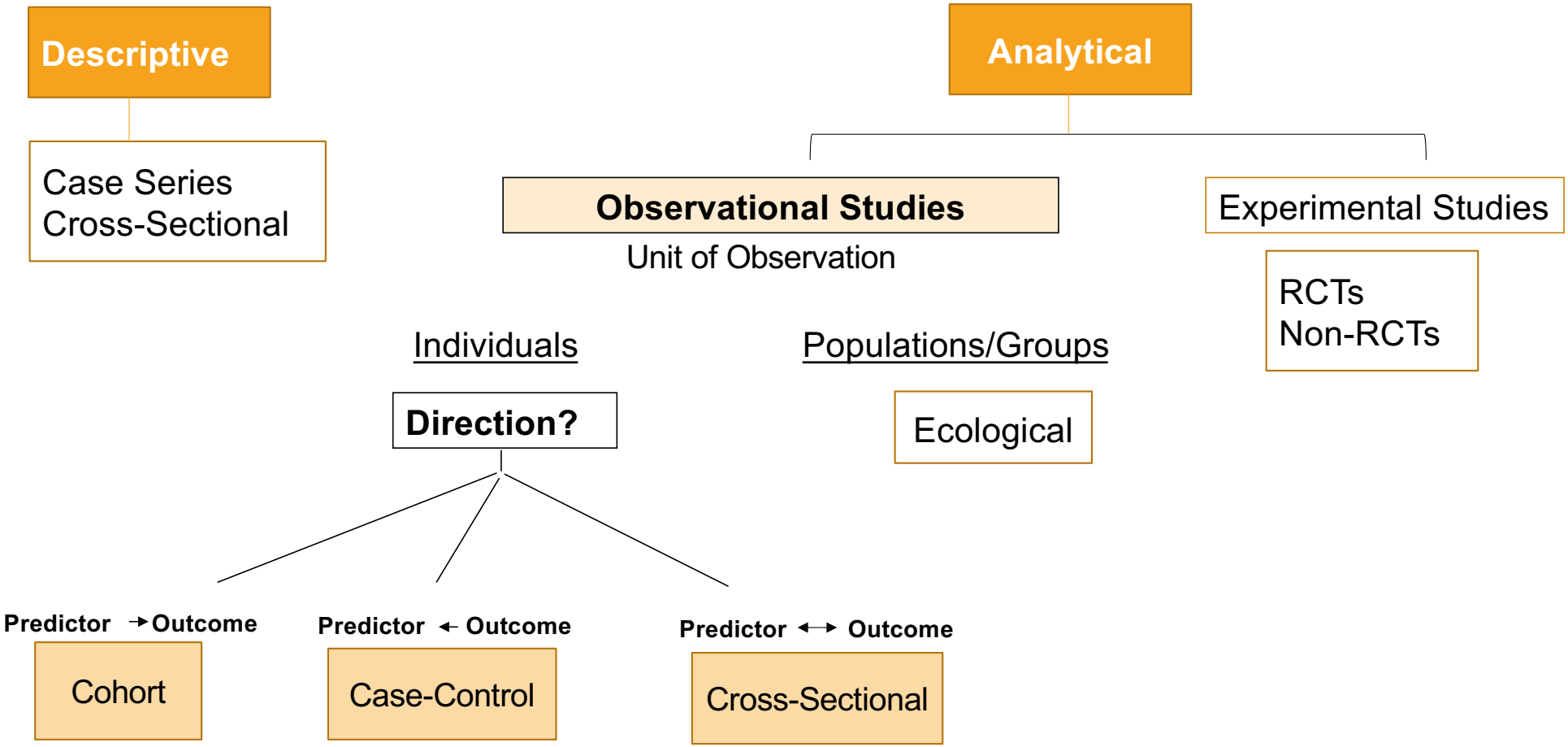
Search for disease causes and preventions

1. To evaluate hypothesis about causes of disease
2. To evaluate success of intervention programs

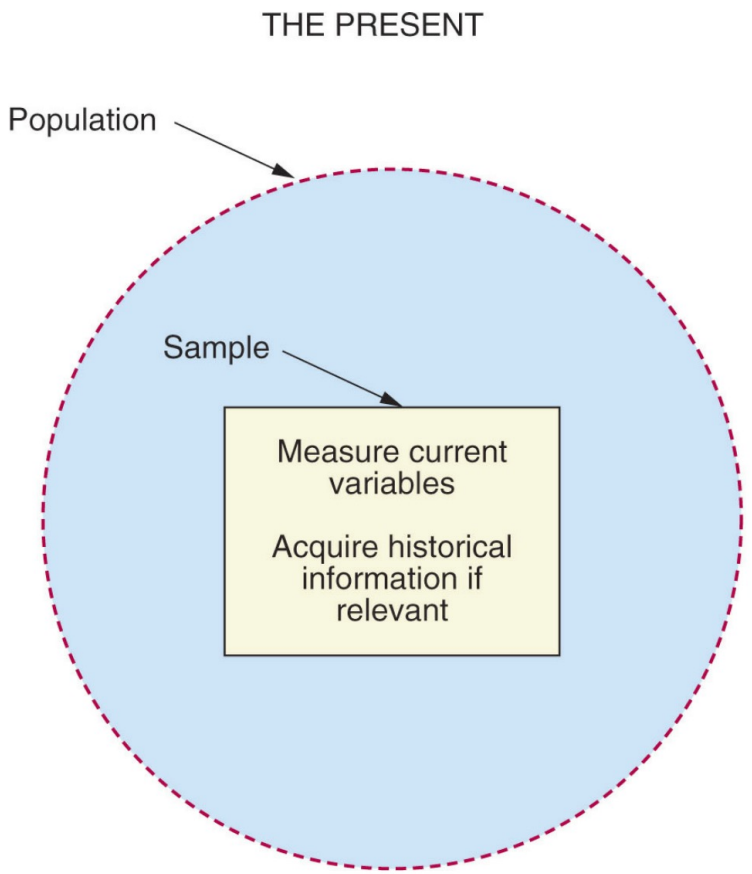
Compare groups & systematically determine: is there an association?

- *Clinical Trial*
- *Experimental Study*
- *Case-Control Study*
- *Cohort Study*

Epidemiological Study Designs



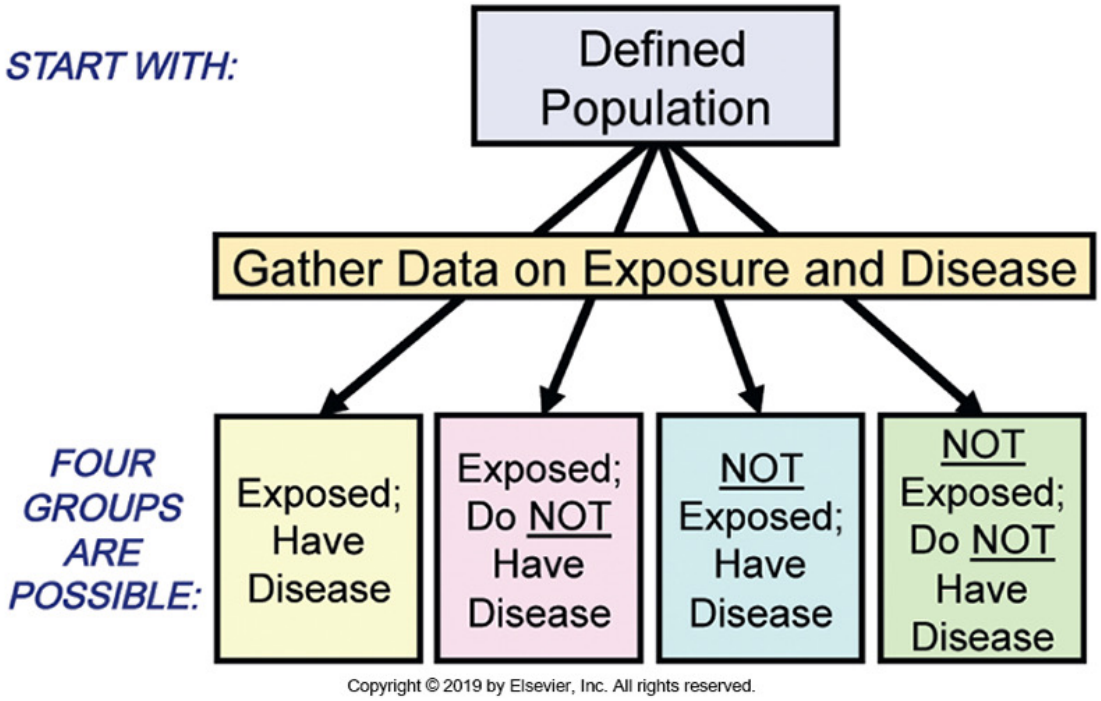
CROSS-SECTIONAL STUDY

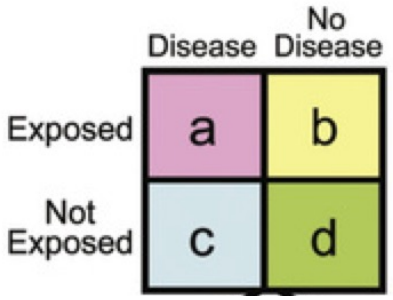


■ FIGURE 7.1 In a cross-sectional study, the steps are to:
• Define selection criteria and recruit a sample from the population.
• Measure current values of predictor and outcome variables, often supplemented by historical information.

Source: Hulley et. All,. Designing Clinical Research. LWW. Kindle Edition.

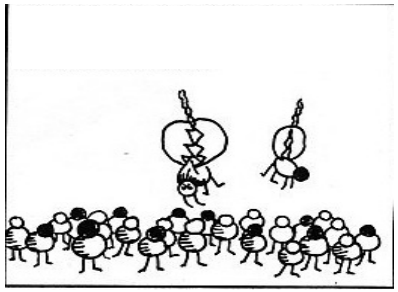
Cross-Sectional Studies





Source: Celentano & Szklo. Gordis Epidemiology. Elsevier Health Sciences.

Cross-Sectional Studies



Sample Size – needs to be calculated

Sampling Methods

- Random sampling: purest form of probability sampling. Each member of the population has an equal chance of being selected.
- Systematic sampling: use of pre-established sequences to select from a source of participants (e.g. medical records)
- Stratified sampling: sample based on certain demographic characteristics, (systematic or random sampling)
- Convenience sampling: the sample is selected because they are convenient (college students, patients, person on the street)

Cross-Sectional Studies – When to use

- **Goal is to describe variables and their distribution pattern**
  - **Example: National Health and Nutrition Examination Survey (NHANES study)**
    - Sample designed to represent the US population -- interviewed and examined
    - Each cross-sectional study -- major source of information on health and habits of the US population (e.g., prevalence of smoking in various demographic groups)
- **Can be used to examine associations**
    - Which variables to label as predictors and outcome depends on the investigator hypothesis
    - Temporal relationship usually cannot be established

Cross-Sectional Studies

Analytical Approach

Exposure	Outcome		Total
	Present	Absent	
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b + c + d

Prevalence $_{total}$ = ((a+c) / (a+b+c+d)) x 10^n

Prevalence $_{exposed}$ = (a / (a+b)) x 10^n

Prevalence $_{non\text{-}exposed}$ = (c / (c+d)) x 10^n

Measure of association

Prevalence Ratio = $P_{exposed}$ /$P_{non\text{-}exposed}$

Cross-Sectional Studies: Example 7.1

Analytical Approach

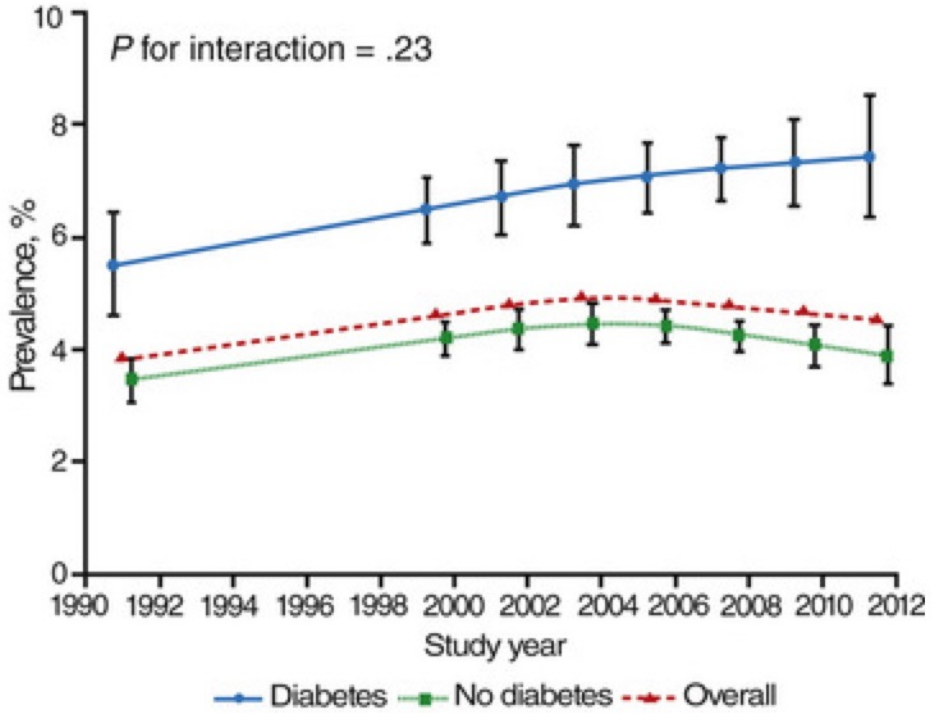
Sargent et al. (2) sought to determine whether exposure to movies in which the actors smoke is associated with smoking initiation. The steps in performing the study were to:

1. Define selection criteria and recruit the population sample. The investigators did a random-digit-dial survey of 6,522 U.S. children aged 10 to 14 years.
2. Measure the predictor and outcome variables. They quantified smoking in 532 popular movies and for each subject asked which of a randomly selected subset of 50 movies they had seen (predictor variable). Subjects were also asked about a variety of covariates such as age, race, gender, parent smoking and education, sensation-seeking (e.g., "I like to do dangerous things"), and self-esteem (e.g., "I wish I were someone else"). The outcome variable was whether the child had ever tried smoking a cigarette.

3. Results and conclusion: 1) Overall, 10% of the population had tried smoking. Quartile (Q) of movie smoking exposure was significantly associated with the prevalence of smoking initiation; 2) This association did not differ significantly by race/ethnicity or census region. 3) After controlling for sociodemographics, friend/sibling/parent smoking, school performance, personality characteristics, and parenting style, the adjusted odds ratio for having tried smoking were 1.7 (95% confidence interval [CI]: 1.1, 2.7) for Q2, 1.8 (95% CI: 1.2, 2.9) for Q3, and 2.6 (95% CI: 1.7, 4.1) for Q4 compared with adolescents in Q1. 4) The covariate-adjusted attributable fraction was 0.38 (95% CI: 0.20, 0.56), suggesting that exposure to movie smoking is the primary independent risk factor for smoking initiation in US adolescents in this age group.

Serial Surveys

A cross-sectional following time



Adjusted prevalence of chronic kidney disease in US adults – NHANES – 1988-1994 thorough 2011-2012 .

Source: Murphy et All., Ann Intern Med. 2016;165:473–481. *in: Celentano & Szklo. Gordis Epidemiology . Elsevier Health Sciences. Kindle Edition.*

Cross-Sectional Studies – Random and Systematic Error

Random error – by chance – may affect precision in both outcome and exposure measures (frequencies or relationship) – solution: increase the sample size

Systematic error (bias) -- can happen in design, conduct, analysis or reporting of a study

Selection bias:

Sampling Bias – Not using representative sample of the source population

Incidence-Prevalence Bias – Inclusion of prevalent cases in a study (overrepresentation of those who have lived the longest)

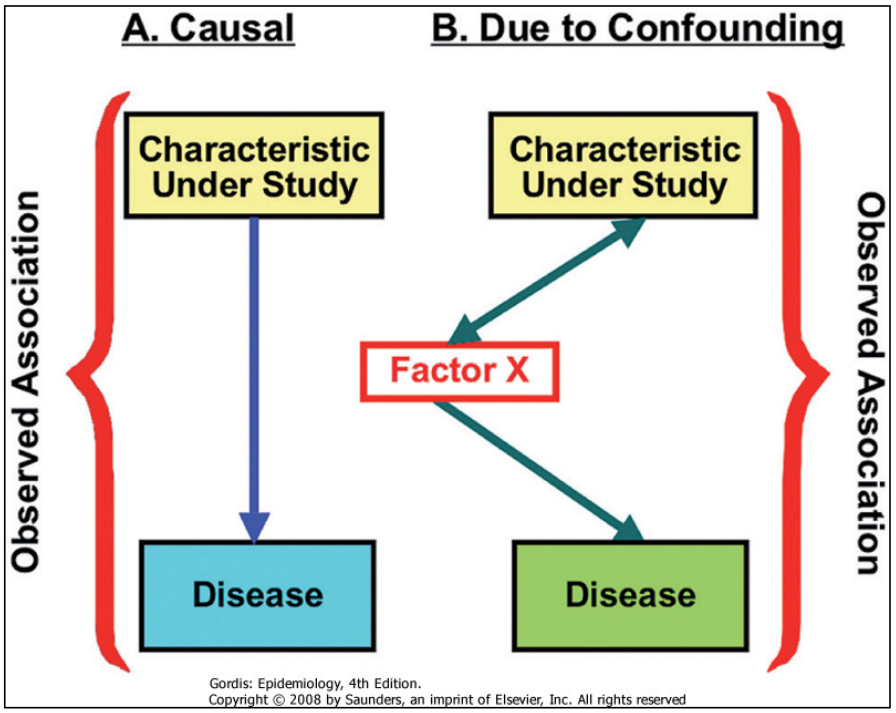
Information bias:

Recall bias – use of self-reporting – differences in accuracy or completeness of recall of past events/experiences

More error details refer to :https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7318122/

Cross-Sectional Studies - Confounding

A distortion in the association between an exposure and disease brought about by extraneous factors (confounders)



Cross-Sectional Studies

Strengths

- No waiting for the outcome to occur
  - Fast; Inexpensive; No loss of follow-up
- Can be a first step in a cohort or a clinical trial

Weakness

- Impractical for studies of rare diseases (if collecting data from the general population)
- Not suited for diseases of short-duration
- Difficult to establish causal relationship

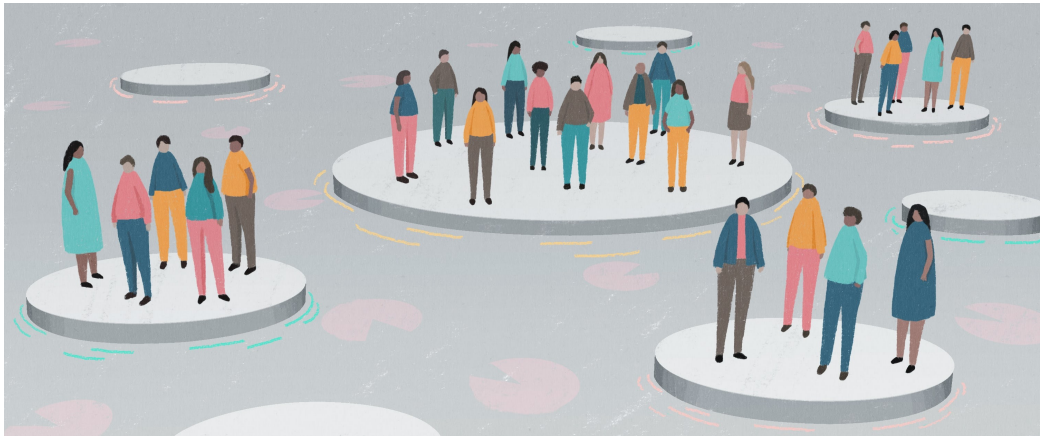
Cohort Studies

Cohort

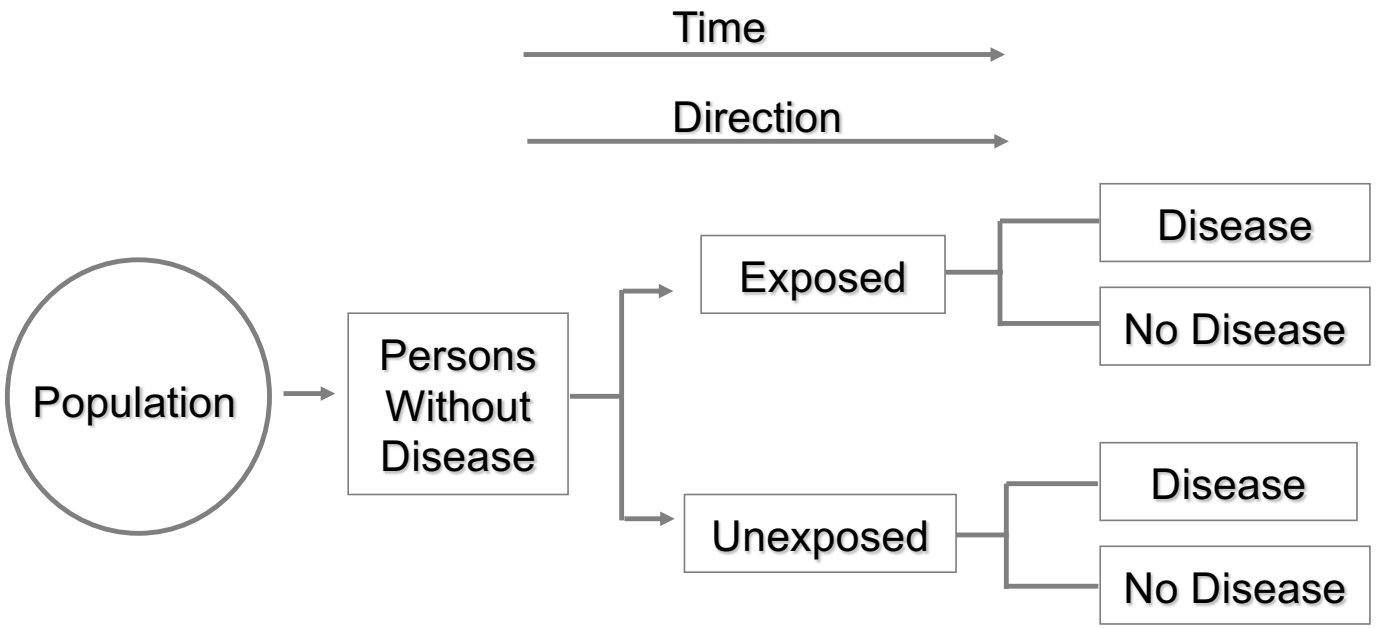
an epidemiological term to identify a group of persons that share a given experience

EXAMPLES:

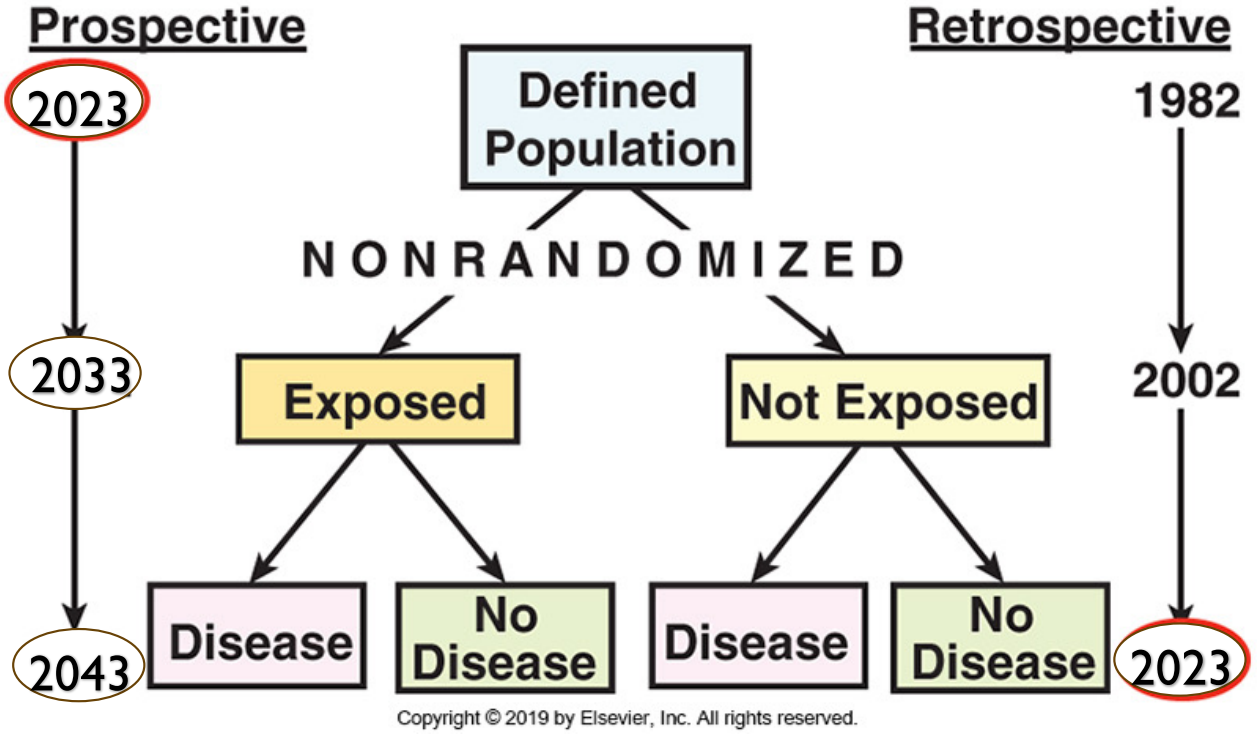
Students
Patients
Employees
Migrants
Pregnant women
Infants
…etc.



Cohort Studies

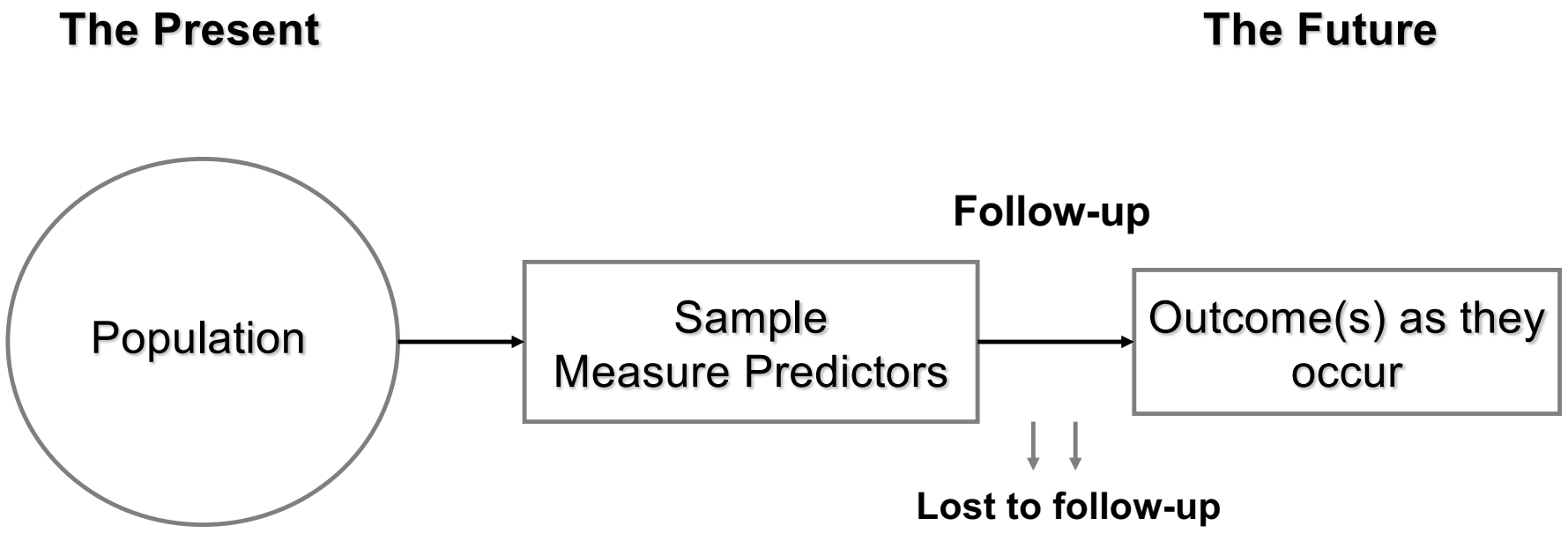


Types of Cohort Studies



Source: Celentano & Szklo. Gordis Epidemiology. Elsevier Health Sciences.

Prospective Cohort Study



Prospective Cohort Studies

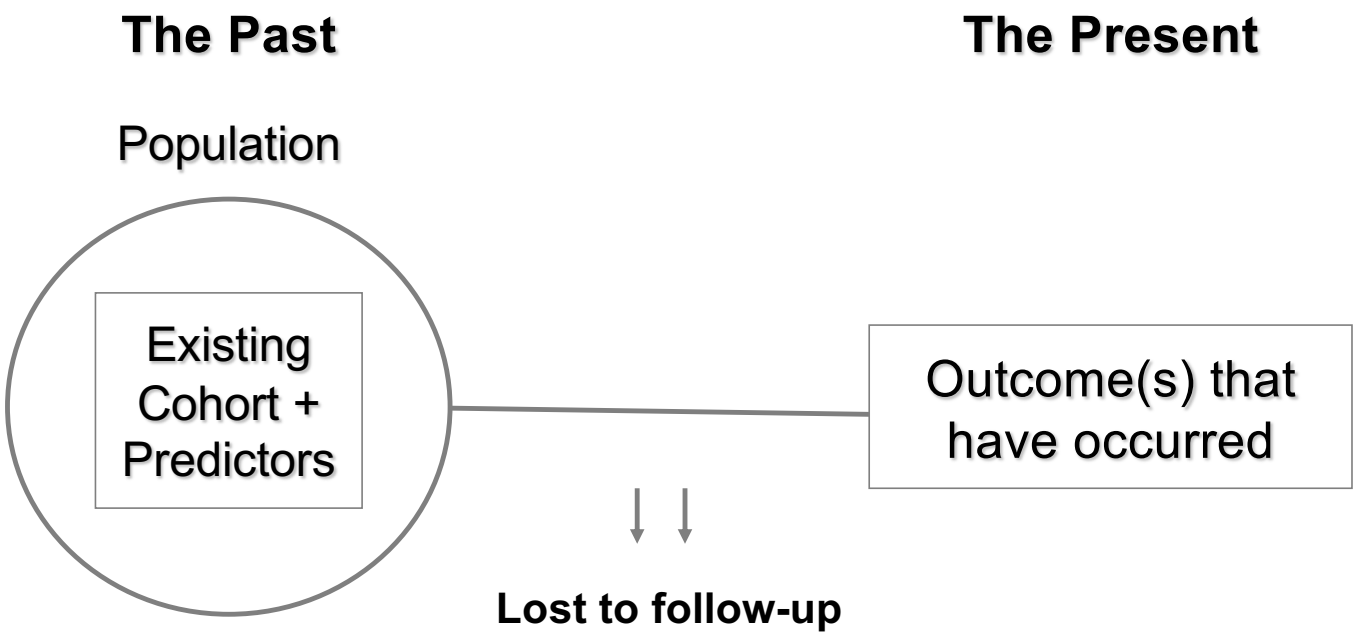
Strengths

- Allows calculation of incidence, hence estimation of risk
- Temporal relationship between predictors and outcome can be established
- Less possibilities of introducing bias if good criteria and procedures for conducting the study are established in advance
- Information can be obtained on participants whose exposure to risk factors have changed

Weakness

- Potential for influences of confounding variables
- High cost and long duration
- Inefficient for rare outcomes

Retrospective Cohort Study



Retrospective Cohort Studies

Strengths

- Same as Prospective Cohort
- And…
  - Less costly
  - Less time consuming

Weakness

- Investigator has limited control over sampling, follow-up of population, quality of baseline measurements

Cohort Studies – Analytical Approach

Exposure or characteristic	Developed disease		Total
	Yes	No	
Present (exposed)	a	b	a + b
Absent (not exposed)	c	d	c + d

Incidence $_{total}$ = ((a+c) / (a+b+c+d)) x 10^n

Incidence $_{exposed}$ = (a / (a+b)) x 10^n

Incidence $_{non\text{-}exposed}$ = (c / (c+d)) x 10^n

When denominator is total time of follow-up for each participant – Rate Ratio

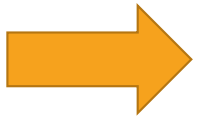
Cox Proportional Hazards -- Hazard Ratio

Measure of association

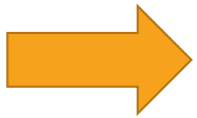
Relative Risk= I $_{exposed}$ /I $_{non\text{-}exposed}$

Cohort Studies – Issues Reviewers Evaluate and Why

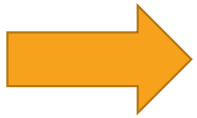
- Is there a well-characterized cohort defined at the beginning of follow-up?
- Will the sample size be large enough?
- Are cohort members readily available to follow-up?
- Do the measures of predictors/outcomes have good reliability and validity?



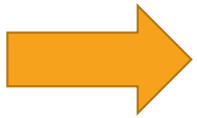
Selection bias
(inclusion and exclusion criteria)



Random error
(a must-have component)



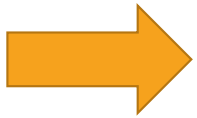
Selection bias
(your proposal's feasibility)



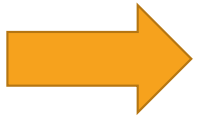
Random error & bias
(quality of your study)

Cohort Studies – Issues Reviewers Evaluate and Why

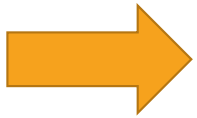
- Does the protocol include standardized assessment criteria? (e.g., blinding)
- Have potential confounders and effect modifiers been included?
- What steps will be taken to maximize retention?
- How will the longitudinal data be analyzed appropriately?



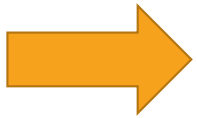
Random error & bias
(quality of your study)



Confounding
(ensuring correct conclusion)



Selection bias from loss to follow-up
(feasibility)



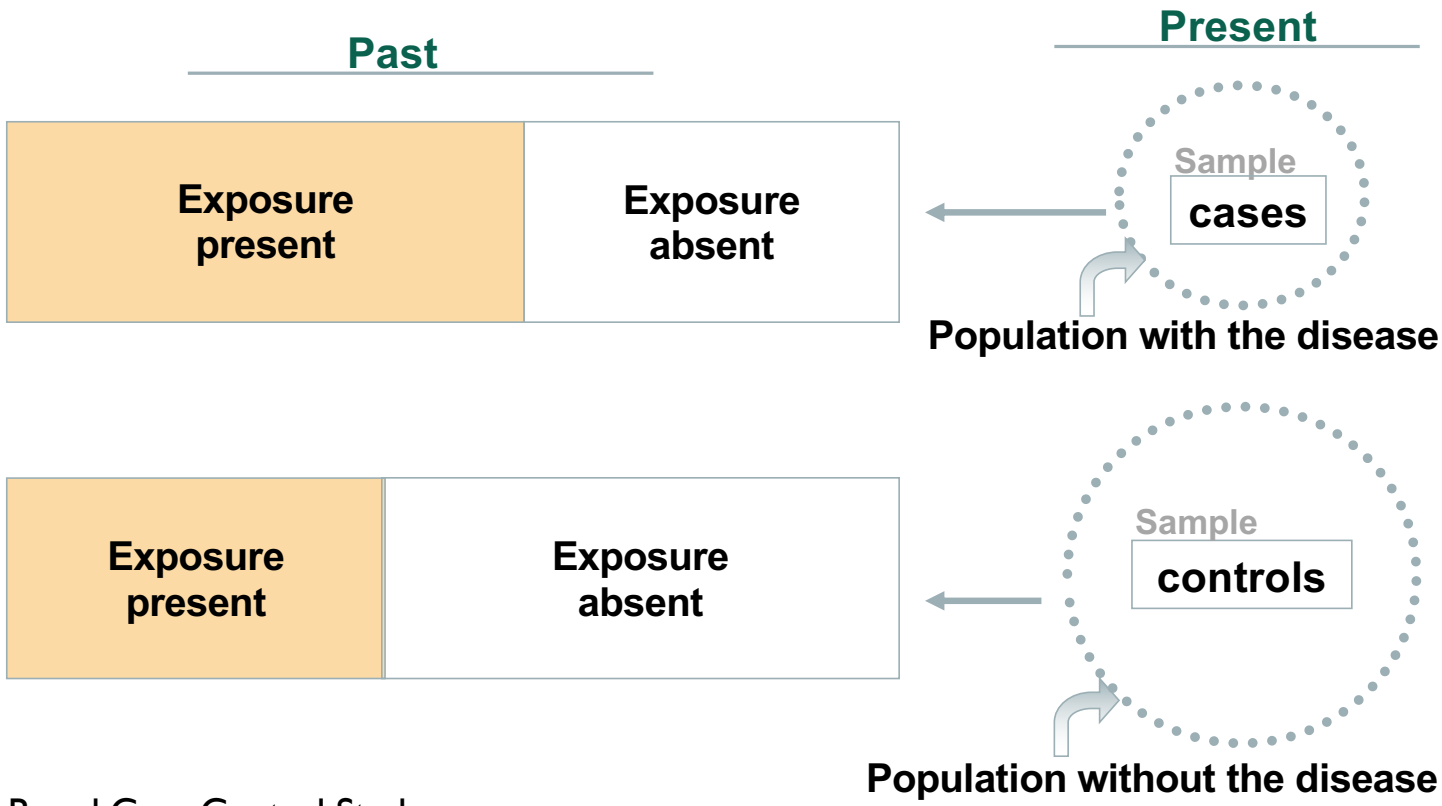
Statistical inference bias
(quality of study)

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Case-Control Study

Qi Zhao, MD, PhD
Associate Professor of Epidemiology
Department of Preventive Medicine
10.27.2023

Case-Control Studies



Case-Based Case-Control Study

Case-Control Studies

Selection of Cases

The source of cases depends on the disease of interest

Hypertension, stroke ----- hospital, clinics
HIV infected individuals ----- STD clinics, community
Cancer ----- Cancer registration

Incident (new case/newly diagnosed) or Prevalent (old case/previously diagnosed) Cases?

Case-Control Studies

Selection of Controls

- One of the major challenges in a case-control studies
- Controls should be similar to the cases in all respects other than having the disease (event) in question
- Controls should be representative of all persons without the disease in the population from which the cases are selected

Case-Control Studies

Multiple Controls

- Controls from the same source -- two or three controls for each case are used to increase the statistical power of the study

- Controls from different sources – e.g., hospital controls and neighbourhood controls.

Case-Control Studies – Analytical Approach

Exposure or characteristic	Disease/Event	
	Cases	Controls
Present (exposed)	a	b
Absent (not exposed)	c	d

Odds Ratio = (a/b) / (c/d)
= (a*d) / (b*c)

Logistic Regression -- Multivariable approach

Case-Control Studies – Strengths

- Efficient for rare outcomes
- Require fewer participants than cohort studies, which means that more expensive and rigorous tests can be used
- There is no problem with losses to follow-up

Case-Control Studies – Weakness

- Cannot estimate the incidence or prevalence of the diseases
- Information on the exposure or risk factor is obtained after the occurrence of disease, so there is not a clear way to estimate the time between exposure and start of disease
- Only one outcome can be studied
- Susceptibility to bias

Case-Control Studies – Weakness

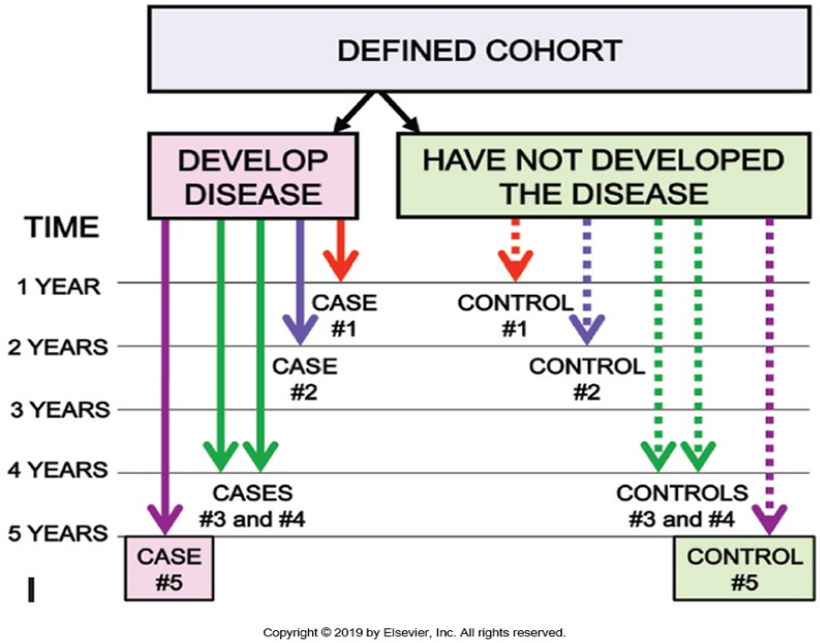
- **Bias sources**
  - **Selection bias**
    - **Control selection**
  - **Information bias**
    - **Recall bias:** e.g., patients with disease may overreport a certain exposure
    - **Interviewer bias: e.g.,** observer may tend to ask cases and controls differently about their exposure

Case-Control Studies

Confounding

- **Matching**
  - To increase the comparability of cases and controls by controlling a confounding variable in the study design: controls are matched to cases based on having the same value of the confounder (e.g. age)
  - More than one control may be matched to each case

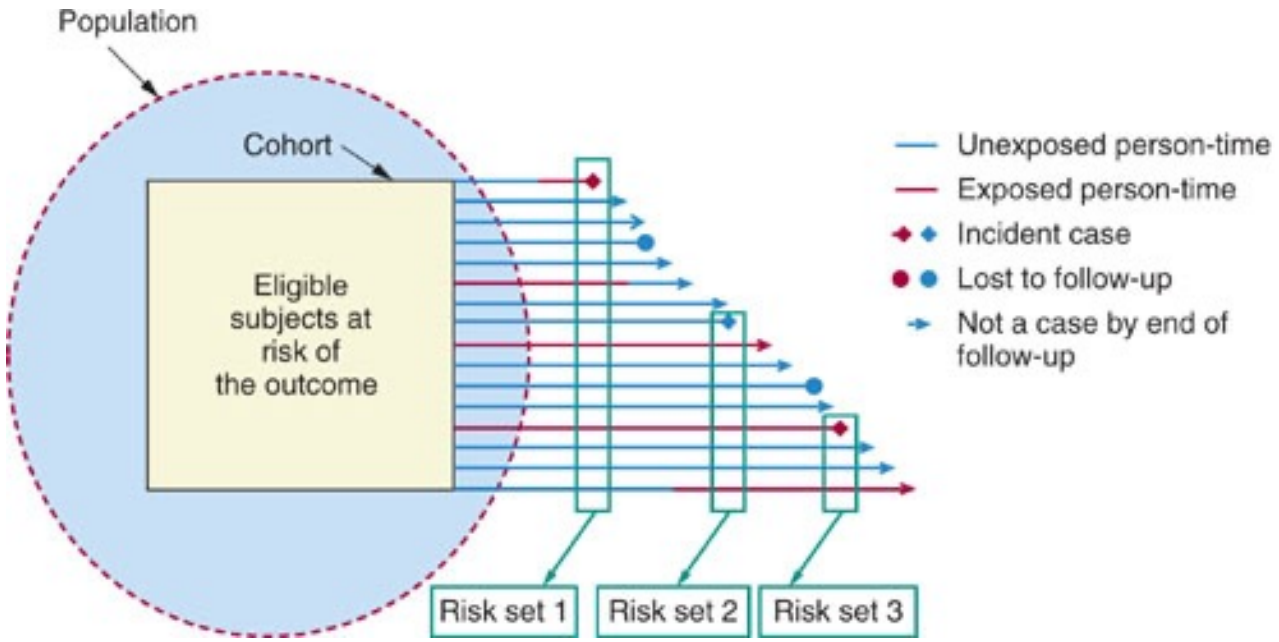
Nested Case-Control Studies



Nested Case-Control Study

Source: Celentano & Szklo. Gordis Epidemiology. Elsevier Health Sciences.

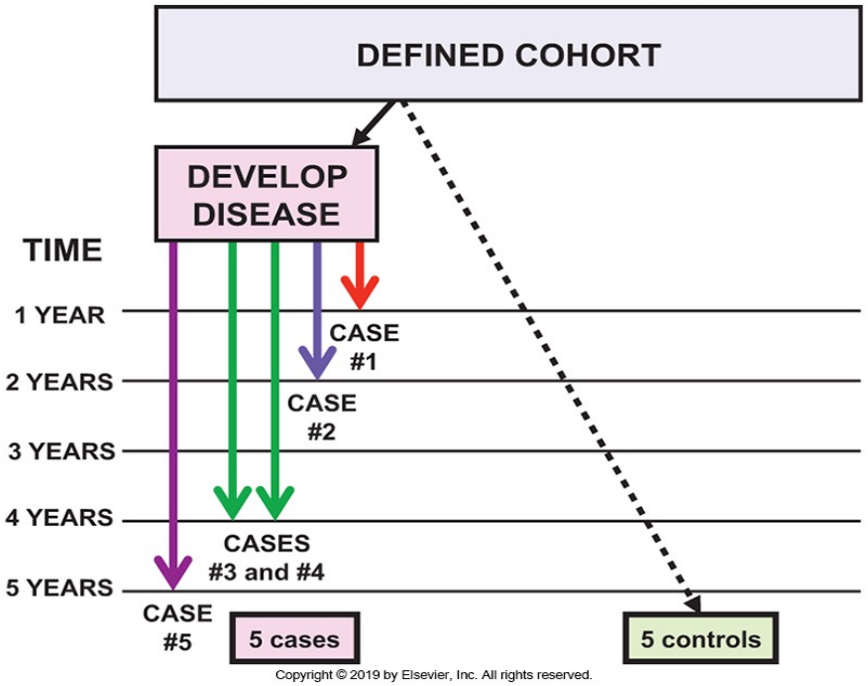
Cohort-based Case-Control Studies



Incidence-Density Nested Case-Control Study

Source: Hulley SB, et al. Designing clinical research. 4th edition.

Cohort-based Case-Control Studies



Nested Case-Cohort Study

Source: Celentano & Szklo. Gordis Epidemiology. Elsevier Health Sciences.

Nested Case-Control Studies

- **Strengths**
  - Useful for costly measurements on specimens that have been archived at the beginning of the study
  - Avoids the potential biases of conventional case–control studies that cannot make measurements on fatal cases and that draw cases and controls from different populations
  - Retains the advantages of cohort studies -- collect predictor variables before the outcomes have happened
- **Weakness**
  - Same as other observational studies including potential for confounding

Considerations in Grant Application

Bias

1) Study design: e.g., nested case-control study; case or control selection; inclusion and exclusion criteria; multiple control groups
2) Data collection: e.g., staff training, blinded to case and control status; additional data collection for evaluating potential bias
3) Data analysis plan: e.g., analyze additional data

Confounding

1) Study design (study population): e.g., matched study design
2) Data collection: e.g., collect potential confounding factors
3) Data analysis plan: e.g., stratification analysis; multivariable modeling

Questions?