Public Health Sciences 310
Epidemiologic Methods

Case-Control Studies

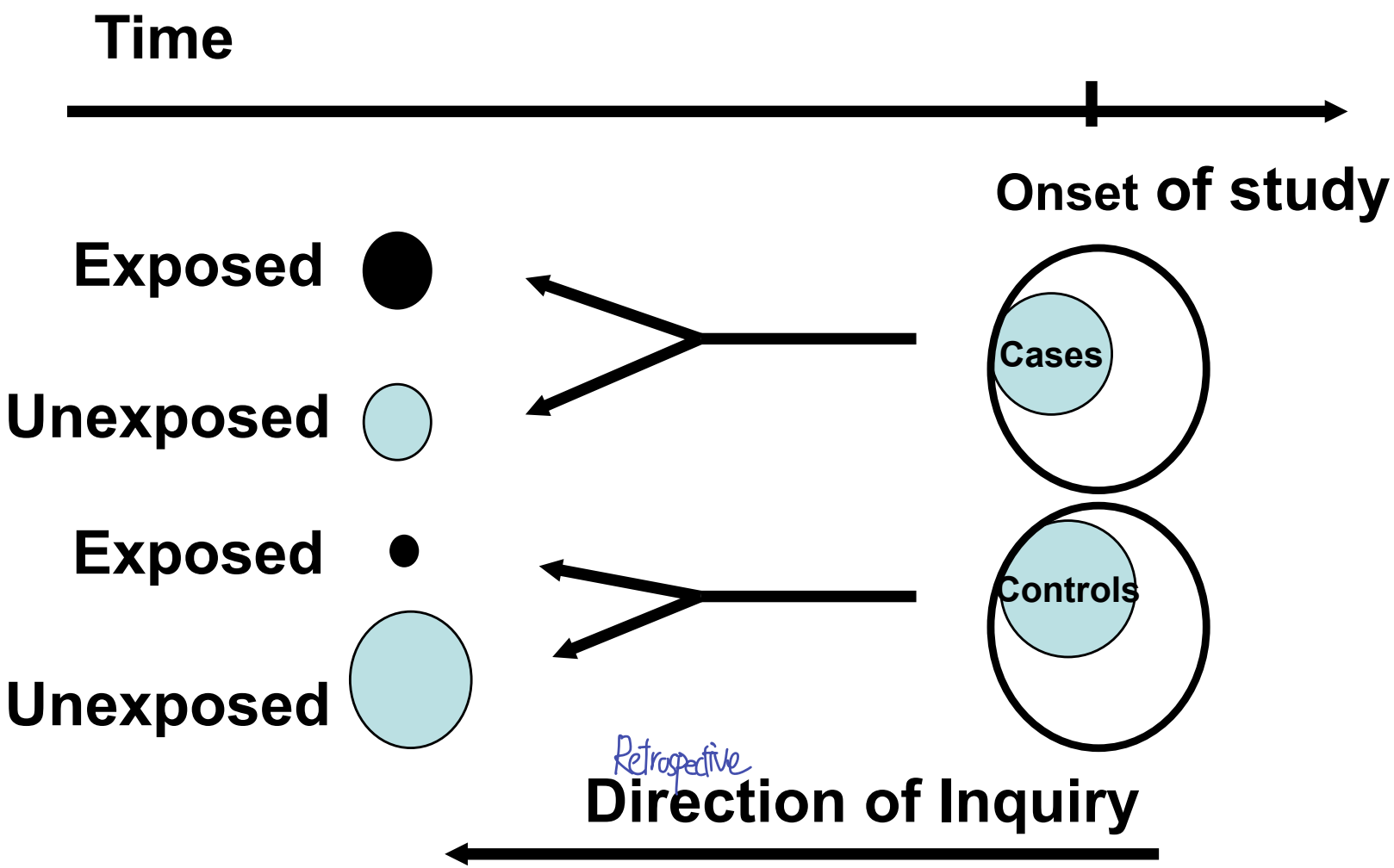
February 13, 2024

Brian C.-H. Chiu

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
- Design Options
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Case-Control Study Design



Outline

- Basic Study Design
- Selection of Cases
  - Types of cases
  - Study base principle
- Selection of Controls
- Types of Controls
- Design Options
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Types of Cases

- Incident cases
- Prevalent cases

Incident Cases

- Prefer new (incident) cases
- Generally defined as first contact with medical system
- Note that past biologic event (at least for chronic disease) will have occurred in the distant past
  - Limitation of case-control studies is that they have difficulty elucidating etiologic factors with long latent periods

Prevalent Cases

- Prevalence = Incidence x Duration
- Duration determined by rate at which the patient leaves the disease state (recovers or dies) or persists in the disease state (slow course or successful palliation)
- Thus, risk factors for prevalent cases are risk factors for both incidence and duration. You cannot distinguish the relative contributions of the two parts.
- Worst case scenario: exposure X causes a extremely lethal form of a disease; this lowers the proportion of prevalent cases that are exposed to X; a case-control study of prevalent disease would show a decrease in the odds ratio (compared to an incident study) as a lower proportion of prevalent cases would be exposed.

Study Base Principle

- Case and controls should be representative of the same base experience
- The study base can be thought of as the members of an underlying source population for the cases during the time periods when they are eligible to become cases
- Primary study base vs secondary study base

Primary Study Base (e.g., population-based)

- Define the study base ***a priori,*** and then ascertain all of the cases (or a representative sample) in that study base
- Most common example is the population-based case-control study; the population is generally defined geographically (e.g., all cases of leukemia diagnosed in the State of Illinois between 1/1/2024 and 12/31/2026)
- Mechanisms to get cases: disease registries, survey all hospitals and pathology laboratories, birth or vital records, etc.
- Advantage: easy to get controls (study base well defined)
- Major challenge: complete case identification in the base, esp. for diseases with mild symptoms or do not require medical attention

Secondary Study Base (often hospital-based)

- Cases are defined ***a priori*** before the study base is defined; the challenge is the proper identification of the secondary study base in order to ensure proper sampling for controls
- Most common example is the hospital-based case-control study (e.g., all cases of leukemia diagnosed at the University of Chicago Hospital between 1/1/2024 and 12/31/2026)
- Advantages: easy to get cases
- Major challenge: defining the study base in order to identify controls
  - **Controls should be individuals who would have become study cases if they had developed the disease of interest during the study period**
  - Particularly difficult for tertiary care hospitals

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
- Design Options
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Principle

- Controls are individuals who would have become cases if they had developed the disease during the time of the investigation. There is no control group that is optimal for all situations.
  - **Primary base: (e.g.)** If cases are all women diagnosed with Parkinson's disease in the State of Illinois between 1/1/24 and 12/31/25, then appropriate control pool would be all women without Parkinson's disease living in Illinois during that time period
  - **Secondary base: (e.g.)** If cases are all women diagnosed with Parkinson's disease at University of Chicago Hospital between 1/1/24 and 12/31/25, then appropriate control pool would be all women who would have been diagnosed at University of Chicago Hospital had they developed Parkinson's disease
  - Study base principle requires that controls be representative of the base but not necessarily the general population

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
  - Population controls
  - Hospital or disease registry controls
  - Other types
- Design Options
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Population Controls

- Studies using a primary base
- Advantage
  - Controls are drawn from the same source as the cases
  - Increase generalizability of study results
- Disadvantages
  - Inconvenient
  - Cost (vs secondary base)
  - Less motivation
    - If 10% to 20% of invited eligible controls participate, is that really a population-based control group?

Population Controls (cont)

- Selection with a roster
  - e.g., birth certificate records, health plan enrollees, voting lists, etc.
- Selection without a roster
  - Remember: the goal is ensure that every eligible subject in the study base has the same chance of selection
  - Random-digit dialing (RDD)
    - Why not use a telephone directory?
    - Sample unit: telephones
    - Must specify how multiple eligible controls will be handled
    - Efficiency can be low
  - Neighborhood controls
    - Sampling unit: residences
    - Advantages:
      - Confounding factors associated with the neighborhood should be balanced between cases and control (esp. SES and environmental factors)
    - Disadvantages:
      - high cost; multiunit dwellings; potential for overmatching; adapts less well for rural areas

RDD Screening Outcomes

Outcomes	Calls
Total numbers in sample	9,911
Non-working numbers	2,411
Non-households (business, teen, etc)	1,369
Undetermined (ring-no-answer)	507
Unscreened household (hang-up, impairment hearing, maximum attempts)	1,180
Screened households	4,444
- eligible, refused contact by study manager	55
- Ineligible (exclusion criteria)	3,714
- Eligible and agreed to be contacted by SM	675

Hospital Controls

- Advantages
  - Easily identified and readily available
  - Often more willing to participate than health individual (e.g., blood draws, physical examinations, etc)
- Disadvantages
  - Different catchments for different diseases
  - More likely to share similar exposures with cases (overmatching)
  - Berkson's bias
- Do not use patients w/ dx known to be related to any risk factor of interest
  - Example: coffee and pancreatic cancer
- Select from different diagnostic groups so that no particular risk factors will be overrepresented

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
- Design Options
  - Matching
  - Ratio of controls to cases
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Matching - Principles

- One or more controls matched to each case with respect to one or more confounders (e.g., age, sex)
- Criteria on selecting matching variables
  - Number should be kept to a minimum
  - Should be potential confounding variables
  - Should be readily available at little or no cost
  - Note: modern statistical methods allow more flexible confounder control than matching. Match when you cannot adjust.
- Types of matching
  - Individual matching
  - Frequency-matching or category-matching: match on crude categories (e.g., 20% of cases are over age 65, make sure ~20% of the controls are too)

Advantages and Disadvantages

- Advantages
  - Control for confounding variables that are difficult to measure directly
  - Control of unmeasured confounders
  - Potential gain in the precision of estimating ORs
- Disadvantages
  - Once you match on a factor, you cannot examine its association with the outcome
  - Longer study duration
  - Potential for overmatching

Ratio of Controls to Cases

- Little increase in efficiency beyond 4:1; even at 2:1 there is little increase in efficiency (so use >1:1 only if controls are cheap)
- Better strategy is to get more cases
- Generally 1:1 best approach

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
- Design Options
- Assessment of Exposure
- Estimates Made from Case-Control Studies
- Summary

Ascertainment of Exposure

- For cases, important to make sure that exposures preceded the disease
- Interviewers should remain unaware of the specific hypotheses being tested (reduce observation bias)
- Sources of data
  - Pre-existing records (e.g., medical records, employer records)
  - Interviews and questionnaires
  - Direct measurement (e.g., physical examination, etc.)

Selection Bias

- When the inclusion of cases OR controls into the study is somehow (inadvertently!) associated with exposure or disease status
- Some ways this can happen:
  - Inappropriate control group (e.g., coffee and pancreatic cancer)
  - Case identification related to exposure or media attention
    - e.g., toxic shock syndrome and tampon use
  - Low response rate in controls

Information Bias: Recall Bias

- Cases (or families of affected individuals) recall past exposures differently than controls
  - e.g., maternal x-ray exposure and childhood cancer
  - What would you do to address this issue?
- Disease itself, diagnosis, or treatment may change actual habits as well as the perception of current and past habits

Outline

- Basic Study Design
- Selection of Cases
- Selection of Controls
- Types of Controls
- Design Options
- Estimates Made from Case-Control Studies
- Summary

Estimates Made from Case-Control Studies

- The odds ratio of disease
- Need to account for any known potential confounding variables
- Need to describe effect modifiers
- Need to evaluate the role of chance

Summary – Case-Control Studies

- Strengths
  - Optimal for rare disease or latency period is long
  - Can examine multiple etiologic factors for a single disease
  - Relatively quick and inexpensive compared with other analytic designs
  - Generally require a smaller number of subjects
- Limitations
  - Inefficient for rare exposures
  - Temporal relationship may be difficult to establish
  - Cannot directly compute incidence rates
  - Particularly prone to bias (Selection bias and recall bias)

Consequences When Bias Goes Unrecognized

- Pearl R. *Cancer and Tuberculosis*. Am J Hyg 1929;9:97-159
  - Cases: pts autopsied in the Johns Hopkins Hospital and found to have a malignant tumor
  - Controls: pts without tumors on autopsy were age-, sex-, and race-matched to the cases
  - Active tuberculous lesions: found in 16% of controls and in only 7% of cases
  - Conclusion: “there is a definite and marked incompatibility or antagonism between the two diseases, cancer and tuberculosis”
- Researchers at Johns Hopkins began treating terminal cancer patients with tuberculin (Pearl R et al., Experimental treatment of cancer with tuberculin. Lancet 1929;216:1078-80)
- Bias invalidated the results in the C-C study: Study included an overrepresentation of exposed controls (many control subjects had died from tuberculosis), causing a spurious inverse association between cancer and tuberculosis
- Another study using a representative control group (patients who died from heart disease) found that the prevalence of tuberculous lesions was the same in both cancer and control groups (Carlson HA, J Cancer Res 1929;13:126-35)
- Pearl R: Discussion: A note on the association of diseases. Science 1929;70:191-2.

- If properly designed and carried out, the case-control study….
  - is a valuable research tool that can provide reliable tests of particular epidemiologic hypotheses
  - provides information that mirrors what could be learned from a cohort study, usually at considerably less cost and time