Case-Control Studies

Vincent Lo Re, MD, MSCE, FISPE
Department of Medicine (Infectious Diseases)
Center for Pharmacoepidemiology Research and Training
Perelman School of Medicine
University of Pennsylvania

4th MURIA – June 18, 2018

CPeRT Center for Pharmacoepidemiology Research and Training

ISPE

Learning Objectives

- Understand conceptual framework underlying case-control design
- Learn principles underlying selection of controls for case-control studies
- Consider sources of bias in case-control studies and approaches to limit their influence

Outline

- Overview of case-control study design
- Defining source population
- Selection of cases, controls
- Measure of association (odds ratios)

Outline

- Overview of case-control study design
- Defining source population
- Selection of cases, controls
- Measure of association (odds ratios)

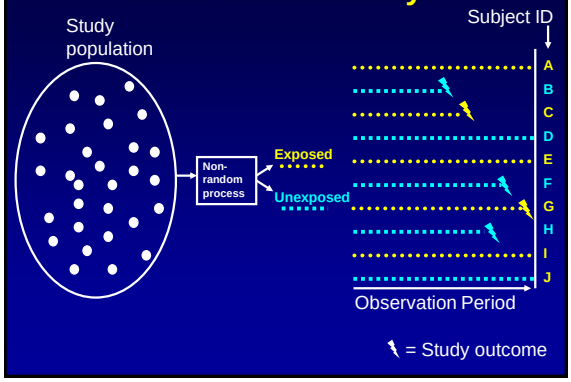
Options in Research Design

Descriptive Studies	Analytic Studies
• Case reports	• Case-control
• Case series	• Retrospective cohort
• Analysis of secular trends	• Prospective cohort
	• Experimental

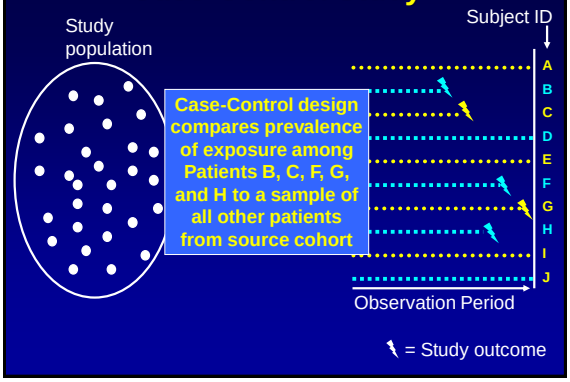
Case-Control Study

- Definition
  - Compares diseased to non-diseased patients for how frequently risk factor is present
- Use
  - Study risk factors for disease (esp. rare)
- Main limitation
  - Biases must be avoided (e.g., historically obtained data must be complete, accurate)

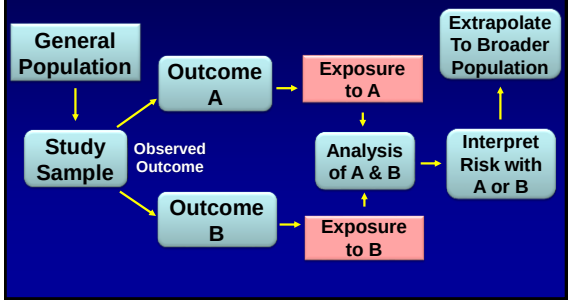
Case-Control Study



Case-Control Study



Case-Control Study



Conceptual Framework of Case-Control Design

- Goal is to capture all (or sample) of cases as if conducting a cohort study
- Select controls such that exposure distribution among controls is same as among population that is source of cases

Case-Control Concepts - 1

- Efficient alternative when cohort study is impractical (e.g., too large)
- Require fewer patients than cohort study
- Relative risks can be estimated from case-control studies

Case-Control Concepts - 2

- Requires extra step in sampling according to outcome/disease (vs. cohort design)
  - Sampling creates more efficiency, but introduces potential biases

Choosing Case-Control Design

- **Efficiency** is main reason for choosing case-control design
  - Rare outcomes
  - Long latency
  - Multiple exposures
  - Time-varying exposures (incidence density sampling)

Outline

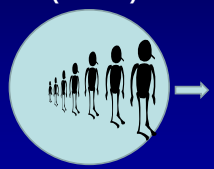
- Overview of case-control study design
- **Defining source population**
- Selection of cases, controls
- Measure of association (odds ratios)

Source Cohort

- Population (cohort) that gave rise to cases included in study
  - Rare outcomes
  - Long latency
  - Multiple exposures
  - Time-varying exposures (incidence density sampling)

Primary Base

Source Population (Cohort)



- Defined as population of interest, or a defined cohort
- Population-based case-control study uses primary base, where population is defined geographically and temporally
- Difficulty → ascertain cases

Secondary Base

Source Population (Cohort)

- Cases are defined before base is identified
- Base is source of cases, and controls would have become cases if they had developed disease
- Challenge → definition of study base

Requires thoughtful identification of referral patterns and other factors (e.g., reputation of hospital in a certain specialty)

Outline

- Overview of case-control study design
- Defining source population
- **Selection of cases, controls**
- Measure of association (odds ratios)

Case Selection

- **Cases can come from hospitals, clinical practices, registries, or cohorts**
- **Must choose to use incident or prevalent cases**
  - **Incident cases → associations more clearly reflect associations with development of disease**
- **Cases must be chosen independently of exposure**

Selection of Controls - 1

- Controls should be selected to have the same exposure distribution as the study base
- Controls should be selected independently of their exposure status
- Controls are NOT selected because they have similar characteristics to the cases

Selection of Controls - 2

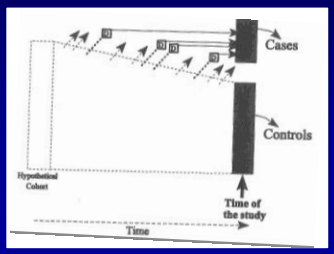
- The time during which a subject is eligible to be a control should be the time in which that individual is also eligible to become a case

Principles of Control Selection

- Controls are people who do not have the disease but otherwise meet the same inclusion and exclusion criteria as the cases
- Need to pick subjects who would have become cases in the study had they developed the disease: i.e. they are representative of the underlying population
- Must be selected independent of exposure status

Strategies for Selecting Controls

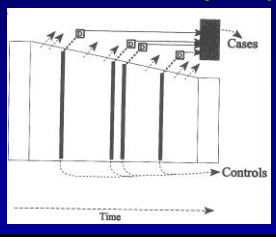
Select controls from set of non-cases once you have defined the set of cases



Szklo, Nieto. 2000.

Strategies for Selecting Controls

- **Select controls from set of individuals in source population who are *at risk* of becoming a case at the time the case is diagnosed (*risk-set* or *incidence density sampling*)**



Nested Case-Control Study

Szklo, Nieto. 2000.

Controls as Later Cases

- With risk-set (incidence density) sampling:
  - Individual selected as control who later develops disease and is selected as a case should be included in study both as a control and case

Outline

- Overview of case-control study design
- Defining source population
- Selection of cases, controls
- Measure of association (odds ratios)

Odds Ratio Approximates Risk Ratio from Underlying Cohort If Disease is Rare

	Disease	No Disease
Exposed	A	B
Unexposed	C	D

If this were a cohort:

$Risk_{Exposed} = A / (A + B)$

$Risk_{Unexposed} = C / (C + D)$

Risk Ratio = [A / (A + B)] / [C / (C + D)]

Odds Ratio Approximates Risk Ratio from Underlying Cohort If Disease is Rare

	Disease	No Disease
Exposed	A	B
Unexposed	C	D

If disease is uncommon in exposed, B>>A → A+B≈B

If disease is uncommon in unexposed, D>>C → C+D≈D

Recall: Risk Ratio = [A / (A + B)] / [C / (C + D)]

Substituting B for A+B and D for C+D, we get:

Risk Ratio=(A/B) / (C/D) = AD/BC = Odds Ratio

ORs and RRs

- Rare diseases: OR ≈ RR
- As prevalence of disease increases, OR departs from RR

Pr (D|E)	Pr (D|No E)	RR	OR
0.002	0.001	2	2.002
0.01	0.005	2	2.01
0.06	0.03	2	2.06
0.10	0.05	2	2.11
0.16	0.08	2	2.19
0.20	0.10	2	2.25

Rare Disease Assumption

- If controls are selected from set of non-cases, OR ≈ RR only when disease is rare
- If use risk-set (incidence density) sampling, rare disease assumption is not necessary
  - OR ≈ RR even if disease is not rare

Odds Ratio Approximates Risk Ratio from Underlying Cohort If Disease is Rare

	Hip Fracture	No Hip Fracture	
Thiazide	46	70	116
No Thiazide	340	316	656
	386	386	772

Odds Ratio = AD/BC = (46 x 316) ÷ (70 x 340) = 0.6
= Unbiased estimate of risk ratio

Interpretation: Risk of femur fracture is 40% lower in thiazide users than non-users.

Herings RMC. *J Clin Epidemiol* 1996;49:115-19.

Summary

- Case-control studies are way to improve efficiency of cohort study
- Odds ratio from cumulative incidence sampling approximates relative risk if disease is rare
- Odds ratio from risk-set (incidence density) sampling is equivalent to relative risk (**no rare disease assumption is needed**)

Descriptive Studies

Vincent Lo Re, MD, MSCE, FISPE
Department of Medicine (Infectious Diseases)
Center for Pharmacoepidemiology Research and Training
Perelman School of Medicine
University of Pennsylvania

4th MURIA – June 18, 2018

CPeRT Center for Pharmacoepidemiology Research and Training

ISPE

Outline

- Review cohort, case-control designs
- Descriptive studies
  - Cross-sectional
  - Correlational
  - Case reports, series

Outline

- Review cohort, case-control designs
- Descriptive studies
  - Cross-sectional
  - Correlational
  - Case reports, series

Overview of the Scientific Method

Study Sample

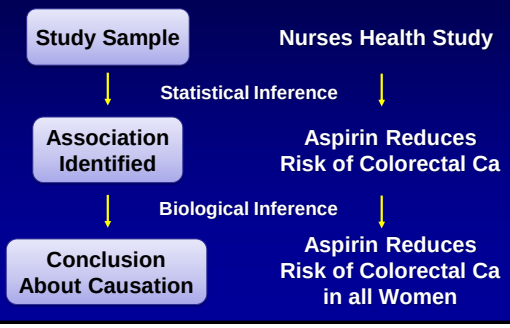
↓ Statistical Inference

Conclusion About a Population (Association)

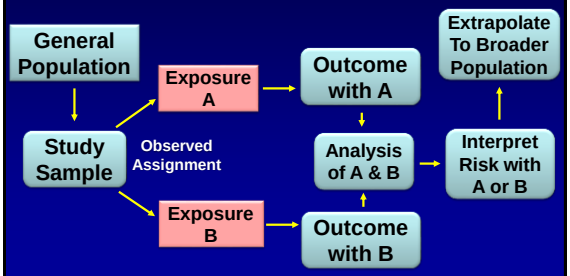
↓ Biological Inference

Conclusion About Scientific Theory (Causation)

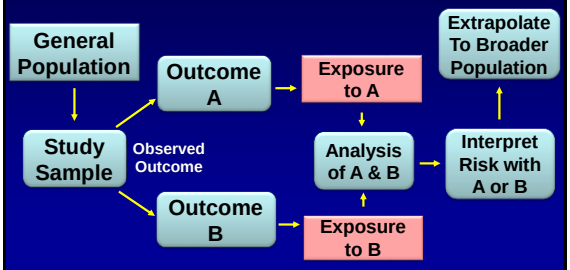
Overview of the Scientific Method



Cohort Study



Case-Control Study



Analytic Studies

- Explicit comparisons of individuals with respect to:
  - Exposure (cohort, experimental studies)
  - Disease status (case-control study)
- Allow testing of epidemiologic hypotheses

Outline

- Review cohort, case-control designs
- Descriptive studies
  - Cross-sectional
  - Correlational
  - Case reports, series

Descriptive Studies

- Describe patterns of disease occurrence with respect to person, place, or time
- Generate etiologic hypotheses
- Types of descriptive studies:
  - Cross-sectional
  - Correlational
  - Case reports, series

Cross-Sectional Study

- Survey of a sample of a population
- Presence/absence of exposure and disease are assessed at the same time
- Can assess prevalence (disease burden)
  - Setting priorities
  - Allocating resources
  - Plan prevention, education services

Time and Prevalence Measures in Cross-Sectional Studies

- Point prevalence: at single time point
  - Prevalence of antiretroviral use in HIV+
- Period prevalence: over specified time
  - Often used for conditions with short duration
  - Prevalence of steroid use among patients with Crohn's disease during one-year period

Cross-Sectional Studies to Estimate Performance of Diagnostic Tests

- Test and gold standard applied at same time
- Prevalence of Test+ among diseased
- Prevalence of Test- among non-diseased

	Gold Standard: Disease	Gold Standard: No Disease
Test+	A	B
Test-	C	D

Cross-Sectional Studies Can Estimate Associations

- Odds ratios are frequently used to assess results
- OR = AD/BC
  - Estimates relative risk if disease is rare

	Gold Standard: Disease	Gold Standard: No Disease
Test+	A	B
Test-	C	D

Is There a Relationship Between Hepatitis C and Diabetes?

1,000 Patients With Hepatitis C	1,000 Patients Without Hepatitis C
25% have Diabetes	12.5% have Diabetes

Odds Ratio = 2.33, p<0.001

There is an association between hepatitis C & diabetes, but is the relationship causal?

Limitations of Cross-Sectional Studies

- Do not capture concept of elapsed time
- No information about transitions from states of health → disease
- Do not distinguish between outcomes that developed recently versus long ago
- Uncertainty as to whether exposure or outcome occurred first

Outline

- Review cohort, case-control designs
- **Descriptive studies**
  - Cross-sectional
  - **Correlational**
  - Case reports, series

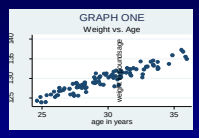
Correlational Studies

- Also referred to as:
  - Ecological studies
  - Analyses of secular trends
- Use aggregated data
- Evaluate correlations, trends over time

Features of Correlational Studies

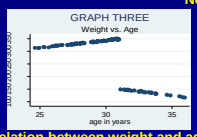
- Measured with correlation coefficient

Correlation as a Measure



Positive Correlation

Negative Correlation



Relation between weight and age is different for younger vs. older

Features of Correlational Studies

- Measured with correlation coefficient
- Popular for initial hypothesis generation
- Relatively inexpensive
- Can rapidly perform with existing data

Limitations of Correlational Studies

- Lack of patient-level data
  - Unable to link exposure and outcome in individual patient
- Inability to control for confounding factors
- Small attributable risks difficult to detect
- Represent average levels of exposures rather than actual levels

Outline

- Review cohort, case-control designs
- Descriptive studies
  - Cross-sectional
  - Correlational
  - Case reports, series

Limitations of Premarketing Clinical Trials

Too…

Few

Ideal

Narrow

Brief

Spontaneous Reports of Adverse Effects

- Clinical description of single patient or series of patients
- Large size → detection of rare events
- Can assist in regulatory decisions
- Vital for hypothesis generation

Limitations of Spontaneous Reports - 1

- Cannot calculate true incidence of event
- Under-reporting in numerator
  - Recognition of event
  - Know how to report, take effort to report
- Lack of denominator

Limitations of Spontaneous Reports - 2

- Report quality
  - Often important data missing
- Bias
  - Reported cases different from unreported
- Lack of comparator group
  - Event rate in unexposed rarely known