Case-Control Studies
Heather J Baer, ScD
7/20/2016

Case-Control Studies

Heather J. Baer
EPI 208
July 20, 2016

Recall: Cohort Studies

- Main features of cohort studies:
  - Selection or classification of subjects based on exposure status
  - "Follow" subjects for outcome (in past or future)
  - Compare incidence of outcome in two or more exposure groups
- Problem: in some situations, cohort studies may not be the best approach

Limitations of Cohort Studies

- To obtain sufficient number of outcomes for rare outcomes (i.e., low risk), will need:
  - Large number of subjects and/or
  - Long follow-up period
- Can be difficult and expensive to obtain information on exposure and other variables from everyone in cohort
- Cohort study can be inefficient
- Alternative: case-control study

Case-Control Studies: Big Picture

- Same goal as other epidemiologic studies
  - To examine (causal) relation between an exposure and an outcome
- Compare exposure histories for two groups of subjects
  - Subjects who develop the outcome of interest (cases)
  - Subjects who reflect the exposure distribution in the population that produced the cases (controls)

Key Features

- Enrollment of subjects based on disease status
- Cases are same as would be in cohort study: subjects who develop the outcome over some time period
- Controls are a sample from population that produced the cases (source population, population at risk)
  - Purpose: to provide estimate of exposure distribution in the source population
- If done properly, can provide same information as cohort study at much less cost and time
  - More "efficient"

How would you do a case-control study of statins and hip fracture?

Case-Control Studies
Heather J Baer, ScD
7/20/2016

Case-Control Study of Statins and Hip Fracture

- Cases: people who had a hip fracture over some time period
- Controls: sample from people who did not have hip fracture (but were at risk)
- Compare statin use in cases and controls

Role of Controls

- Case-control studies compare exposure history of cases and controls
- Controls should represent exposure distribution in source population that produced the cases
  - Tell you about the relative size of exposed and unexposed groups in the source population
  - Controls must be sampled independently of exposure status
- Hardest part of case-control study: selecting an appropriate control group

Nested Case-Control Study

- "Nested" refers to a case-control study that is conducted within a well-defined cohort
  - All cohort members can be easily identified (i.e., there is an actual list or roster available)
  - May need additional information on exposure or other factors that would be difficult or expensive to collect for everyone in original cohort
- But almost any case-control study can be thought of as being nested within some underlying cohort, although it may not be well-defined

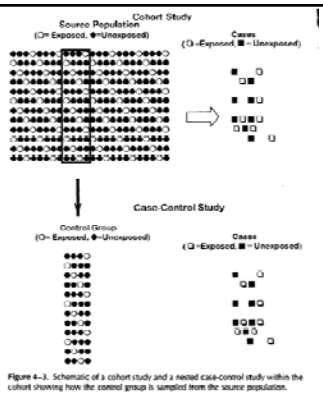


Figure 4–3. Schematic of a cohort study and a nested case-control study within the cohort showing how the control group is sampled from the source population.

Rothman. *Epidemiology: An Introduction (2nd Edition)*, 2012.

Nested Case-Control Study: Example

The Influence of Family History on Breast Cancer Risk in Women With Biopsy-Confirmed Benign Breast Disease

Results from the Nurses' Health Study

Laura C. Collins, MD[1]
Heather J. Baer, ScD[2,3]
Rulla M. Tamimi, ScD[2,3]
James L. Connolly, MD[1]
Graham A. Colditz, MD, DrPH[2,4]
Stuart J. Schnitt, MD[1]

[1] Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts.
[2] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.

BACKGROUND. An association between histologic category of benign breast disease (BBD) and breast cancer risk has been well documented. However, the influence of a positive family history (FH) on breast cancer risk among women with biopsy-confirmed BBD is less certain.

METHODS. The authors conducted a nested case-control study of BBD and breast cancer risk among 2005 women who were enrolled in the Nurses' Health Study. Cases were women with breast cancer who had a previous benign breast biopsy (n = 395 women). Controls were women who also had previous biopsy-confirmed BBD but were free from breast cancer at the time the corresponding case was diagnosed (n = 1610 women). BBD slides were reviewed and categorized as either nonproliferative lesions, proliferative lesions without atypia, or atypical hyperplasia (AH).

RESULTS. Compared with women who had nonproliferative lesions and no FH,

Collins et al. *Cancer* 2006; 107: 1240-7.

Measures of Association

- Odds ratio (OR) is only measure of association directly calculated from case-control studies
  - Exposure OR= $\frac{\text{(odds of exposure among cases)}}{\text{(odds of exposure among controls)}}$
- Depending on how controls are sampled, OR may be an estimate of the Risk Ratio or Incidence Rate Ratio that would have been obtained from a cohort study
- Cannot directly measure risks or rates for outcome, because investigator usually determines relative sizes of case and control groups
  - What is "risk" if 50 cases and 50 controls?
  - What is "risk" if 50 cases and 100 controls?

absolute risk ratio/difference can not be used

Sampling of Controls

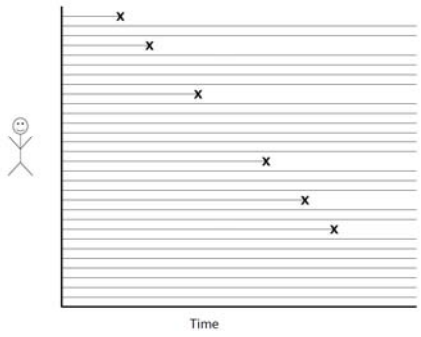
- Controls can be sampled in different ways
- Common options:
  - Cumulative sampling → cumulative case-control study
  - Density-based sampling → density case-control study
    - Risk set sampling
  - Case-cohort study
  - Case-crossover study

Other variants, but less common

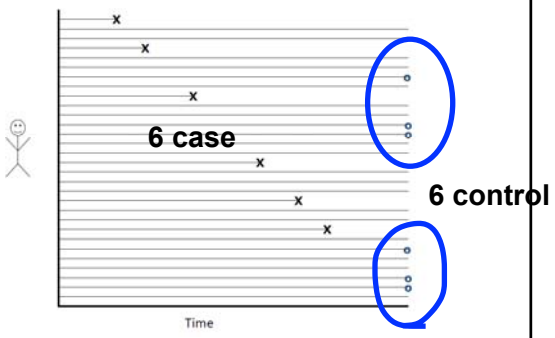
Cumulative Case-Control Study

- Conducted within an underlying closed cohort
  - Cases that occurred during some period of risk
  - Controls sampled from those who did not become cases during period of risk ("survivor sampling")
  - Predominant method used in traditional case-control studies

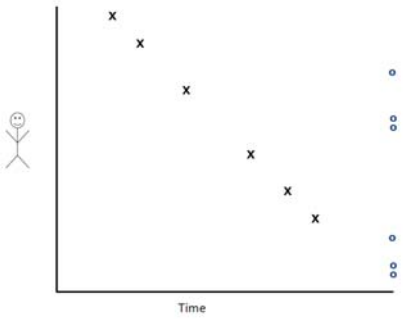
Underlying Cohort (Closed)



Cumulative Case-Control Study



Cumulative Case-Control Study



Cumulative Case-Control Study

- Identify cases of disease that occurred during some time period
- Select sample of controls from subjects who did not develop disease during that time
- Assess exposure in cases and controls
- Calculate odds ratio
  - Exposure OR = Disease OR
  - Odds ratio approximates risk ratio from closed cohort study (if outcome is rare, i.e., low risk in the population)

2 x 2 Table from Closed Cohort

	Disease +	Disease -	Total
Exposure +	A	B	N_1
Exposure -	C	D	N_0

Risk Ratio = RR = $(A/N_1) / (C/N_0)$
= $(A/C) / (N_1/N_0)$
= (odds of exposure among cases)
(odds of exposure in source population)

2 x 2 Table from Cumulative Case-Control Study

	Case	Control
Exposure +	a	b
Exposure -	c	d
Total	M_1	M_0

Odds Ratio = OR = (a/c) / (b/d) = ad/bc
≈ $(A/C) / (N_1/N_0)$
= RR from cohort study (if outcome is rare)

when a or d increase OR increase, when b or C decreases OR decrease

Cumulative Case-Control Study: Example

130

PREDIAGNOSTIC SERUM SELENIUM AND RISK OF CANCER

WALTER C. WILLETT, B. FRANK POLK, J. STEVEN MORRIS, MEIR J. STAMPFER, SARA PRESSEL, BERNARD ROSNER, JAMES O. TAYLOR, KENNETH SCHNEIDER, CURTIS G. HAMES

ON BEHALF OF THE HYPERTENSION DETECTION AND FOLLOW-UP PROGRAM COOPERATIVE GROUP*

Department of Epidemiology, Harvard School of Public Health, Channing Laboratory and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland; Research Reactor Facility, University of Missouri, Columbia, Missouri; University of Texas Health Science Center at Houston, University of Texas School of Public Health; Medical Center Hospital, Duke University, Durham, North Carolina; and Cardiovascular Epidemiology Unit, Evans County, Georgia, USA

Summary Selenium levels in serum samples collected in 1973 from 111 subjects in whom cancer developed during the subsequent 5 years were compared with those in serum samples from 210 cancer-free subjects

Willett et al. *Lancet* 1983; 2: 130-4.

Cumulative Case-Control Study: Example

- Nested case-control study of serum selenium and cancer within the Hypertension Detection Follow-up Programme (HDFP)
  - Randomized clinical trial
  - Frozen blood samples for 4480 participants
- 111 new cases of cancer during 5-year follow-up
- 210 controls selected, with no cancer before or during the study
- Selenium measured on 321 blood samples

2 for each case

Willett et al. *Lancet* 1983; 2: 130-4.

To complete cohort study, we need to fill 2 by 2 table with total sample data, meaning calculate all Pt's selenium level (B,D)

How would you have done this as a cohort study within the HDFP?

2 x 2 Table from Closed Cohort unknown

	Cancer Yes	Cancer No	Total
Low selenium	57	B	N_1
High selenium	54	D	N_0
Total	111	4369	4480

Risk Ratio = RR = $(A/N_1) / (C/N_0)$
= $(57/N_1) / (54/N_0)$
= $(57/54) / (N_1/N_0)$
= ?

In cohort study, would have had to measure selenium levels in all 4480 participants to figure out N_1 and N_0

selection of control make it possible to fill 2 by 2 table

2 x 2 Table from Cumulative Case-Control Study

	Case	Control
Low selenium	57	84
High selenium	54	126
Total	111	210

by selection

Odds Ratio = OR = (57/54) / (84/126)
= [(57)(126)] / [(54)(84)]
= 1.6

In case-control study, controls are telling you about ratio of N_1/N_0

OR = 1.6 ≈ RR from closed cohort (if outcome is rare)

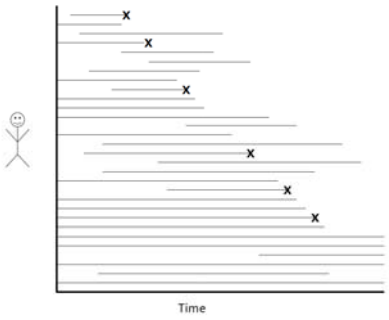
Density Case-Control Study

- Conducted within an underlying open cohort (dynamic population)
  - Cases that occur during some time period of risk
  - Controls sampled over time from person-time contributed by individuals at risk
- Controls represent person-time distribution of exposure in the source population
  - Probability of being sampled as a control is proportional to person-time contribution (more time in cohort = greater chance of being selected as a control)
  - Odds ratio from density case-control study is an estimate of incidence rate ratio from open cohort study

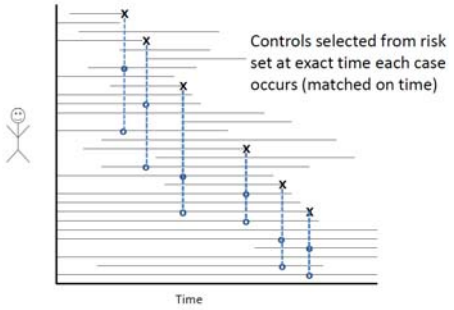
Risk Set Sampling

- Most common way to implement density sampling
  - Controls are matched to cases on time
  - Controls selected from unique set of people in the source population who are at risk at the time that each case is diagnosed (risk set)
  - Risk set changes from one case to the next
  - Theoretically, a control can later become a case, and same control may be selected (by chance) more than once

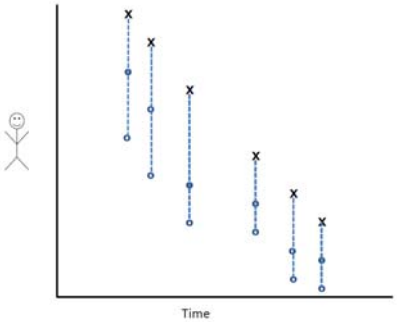
Underlying Cohort (Open)



Case-Control Study with Risk Set Sampling



Case-Control Study with Risk Set Sampling



2 x 2 Table from Open Cohort

	Cases	Person-time
Exposure +	A	K_1
Exposure -	C	K_0

Incidence Rate Ratio = IRR
= $(A/K_1) / (C/K_0)$
= $(A/C) / (K_1/K_0)$
= (odds of exposure among cases) / [(person-time among exposed)/(person-time among unexposed)]

Density Case-Control Study (Risk Set Sampling)

- Assess exposure in cases and in controls
- Calculate odds ratio
- Don't know K_1 and K_0 (person-time among exposed and unexposed)
- Purpose of controls is to provide estimate of K_1/K_0 in the source population

2 x 2 Table from Density Case-Control Study

	Case	Control
Exposure +	a	b
Exposure -	c	d
Total	M_1	M_0

Odds ratio = OR = (a/c)/(b/d) = ad/bc
= $(A/C) / (K_1/K_0)$
= IRR from open cohort

In case-control study, controls are telling you about ratio of K_1/K_0

Risk Set Sampling: Example

Antibiotic use and risk of gynecological cancer

Hani M. Tamim [a,b,*], Khaled M. Musallam [c], Hanan MF Al Kadri [b], Jean-François Boivin [d,e], Jean-Paul Collet [f]

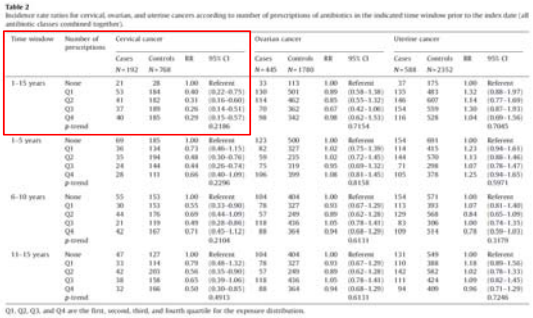
Tamim et al. *Eur J Obstet Gynecol Reprod Biol* 2011; 159: 388-93.

Risk Set Sampling: Example

- Source population: "Dynamic cohort defined by registration with Saskatchewan Health from January 1, 1981-December 31, 2000"
- Cases: 1225 women diagnosed with cervical, ovarian, or uterine cancer from 1981-2000
- 4 controls selected for each case
  - Alive and free of cancer in month case was diagnosed
  - Index date for each case = date of cancer diagnosis
  - Matched controls assigned same index date as case
- Assessed antibiotic use between 1976 and index date for cases and controls, using prescription data

Tamim et al. *Eur J Obstet Gynecol Reprod Biol* 2011; 159: 388-93.

Risk Set Sampling: Example



Tamim et al. *Eur J Obstet Gynecol Reprod Biol* 2011; 159: 388-93.

Sources for Controls

- Want controls to represent exposure distribution in source population that produced the cases
  - Nested case-control study: the cohort is the source population
  - Non-nested case-control study: source population may be harder to define, but need to think about it
- Population controls: appropriate when cases are identified from a well-defined population (e.g., residents of a geographic area)
- Hospital- or clinic-based controls: appropriate when cases are identified from hospitals or clinics

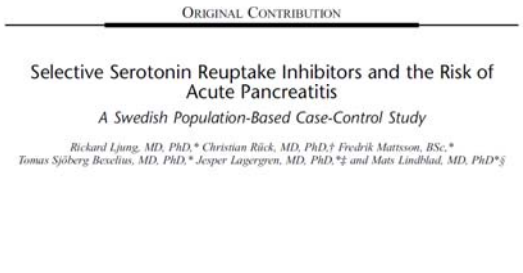
The "Would" Criterion

- If this condition is met, member of the control group "would" have been included as a case in the study if they had developed the outcome
- Guiding principle: cases and controls should come from the same source population
  - Where are the cases coming from? What is the source population that produced the cases?

Population Controls

- Appropriate when cases are identified from a well-defined population (e.g., residents of a geographic area)
- When a population roster or registry is available, random selection of population controls is relatively simple
  - Census lists
  - Birth certificates
  - Electoral rolls
- If no population roster is available, other approaches are:
  - Random digit dialing
  - Neighborhood or friend controls

Population Controls: Example 1



ORIGINAL CONTRIBUTION

Selective Serotonin Reuptake Inhibitors and the Risk of Acute Pancreatitis

A Swedish Population-Based Case-Control Study

Rickard Ljung, MD, PhD, Christian Rück, MD, PhD,† Fredrik Mattsson, BSc,* Tomas Sjöberg Bexelius, MD, PhD,* Jesper Lagergren, MD, PhD,*‡ and Mats Lindblad, MD, PhD*§*

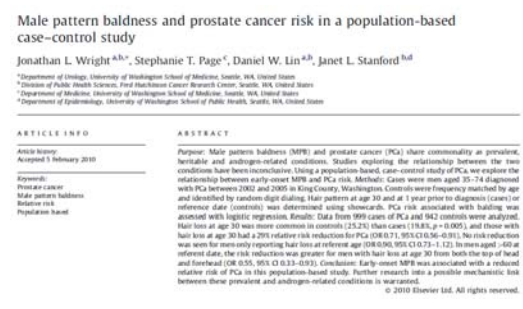
Ljung et al. *J Clin Psychopharmacol* 2012; 32: 336-40.

Population Controls: Example 1

- Source population: all Swedish residents ages 40-84
- *The Patient Register* was used to identify all cases of first episode of acute pancreatitis in this population (Jan. 1, 2006-Dec. 31, 2008)
- *The Register of the Total Population* was used to randomly select control subjects from the general population

Ljung et al. *J Clin Psychopharmacol* 2012; 32: 336-40.

Population Controls: Example 2



Male pattern baldness and prostate cancer risk in a population-based case–control study

Jonathan L. Wright [a,b,*], Stephanie T. Page [c], Daniel W. Lin [a,b], Janet L. Stanford [b,d]

ARTICLE INFO

ABSTRACT

Wright et al. *Cancer Epidemiol* 2010: 34: 131-5.

populationの地理情報を必ず記載する

ケースもコントロールも Population Controls: Example 2

- Source population: male residents of King County, Washington
- Cases: residents of King County, Washington with histologically-confirmed prostate cancer ascertained from Seattle-Puget Sound SEER cancer registry (Jan. 1, 2002-Dec. 31, 2005)
- Controls: male residents of King County with no history of prostate cancer, identified using random digit dialing

Wright et al. *Cancer Epidemiol* 2010: 34: 131-5.

VS　地域ベースコントロール Hospital-based Controls

- Appropriate when cases are identified from hospitals or clinics
- Source population = hospital's catchment area
  - Often poorly defined
  - Who would come to this particular hospital if they had this particular condition?
  - Could be different for different conditions
- Controls selected from other diagnostic groups being treated at the same hospital
  - Share same catchment area as cases
  - Control condition(s) must be unrelated to exposure under study
  - May include more than one diagnostic group

Hospital-based Controls: Example

630 THE NEW ENGLAND JOURNAL OF MEDICINE March 12, 1981

COFFEE AND CANCER OF THE PANCREAS

BRIAN MACMAHON, M.D., STELLA YEN, M.D., DIMITRIOS TRICHOPOULOS, M.D., KENNETH WARREN, M.D., AND GEORGE NARDI, M.D.

Abstract We questioned 369 patients with histologically proved cancer of the pancreas and 644 control patients about their use of tobacco, alcohol, tea, and coffee. There was a weak positive association between pancreatic cancer and cigarette smoking, but we found no association with use of cigars, pipe tobacco, alcoholic beverages, or tea. A strong association between coffee consumption and pancreatic cancer was evident in both sexes. The association was not affected by controlling for cigarette use. For the sexes combined, there was a significant dose-response relation ($P \sim 0.001$); after adjustment for cigarette smoking, the relative risk associated with drinking up to two cups of coffee per day was 1.8 (95 per cent confidence limits, 1.0 to 3.0), and that with three or more cups per day was 2.7 (1.6 to 4.7). This association should be evaluated with other data; if it reflects a causal relation between coffee drinking and pancreatic cancer, coffee use might account for a substantial proportion of the cases of this disease in the United States. (N Engl J Med. 1981; 304:630-3.)

MacMahon et al. *New Engl J Med* 1981; 304: 630-3.

Hospital-based Controls: Example

- Coffee and cancer of the pancreas
- Cases: patients with histologic diagnoses of cancer of the pancreas who were in any of 11 large hospitals in the Boston metro area and Rhode Island (Oct. 1974-Aug. 1979)
- Controls: other patients under care of same physician in same hospital at same time
  - Excluded patients with other diseases of the pancreas or hepatobiliary tract – why?

MacMahon et al. *New Engl J Med* 1981; 304: 630-3.

Threats to Validity in Case-Control Studies

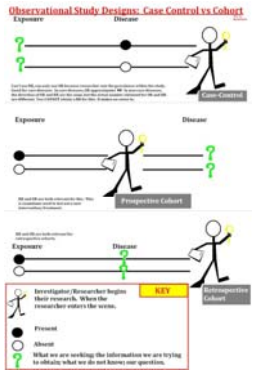
- Selection bias
  - Bias due to how subjects are selected into study
  - Often because controls not selected independently of exposure (preferential selection of exposed or unexposed controls)
- Information (or measurement) bias
  - Bias due to how information is collected on subjects in study (e.g., recall bias)
- Confounding
  - Other factors that predict risk of outcome are distributed unequally in exposure groups

Strengths of Case-Control Studies

- Often more efficient than cohort studies, especially in setting of:
  - Rare outcomes, or
  - Exposures that are difficult and/or costly to assess
- Usually much less expensive and time-consuming than cohort studies
- Convenient for studying multiple exposures

Limitations of Case-Control Studies

- Cannot directly measure risks or rates
- Potential for selection bias due to sampling of controls
- Reduced precision due to sampling, although this loss can be kept small if number of controls selected per case is large
  - Loss is offset by cost savings of not having to obtain exposure information on everyone



https://en.wikipedia.org/wiki/File:ExplainingCaseControlSJW.jpg

Coming Up

- Major epidemiologic study designs
  - Randomized controlled trials, cohort studies, case-control studies
  - Measures of association from these studies
- We usually are interested in whether observed association is causal
  - Not all associations are causal
- What are alternative explanations for results?
  - Confounding, bias, chance