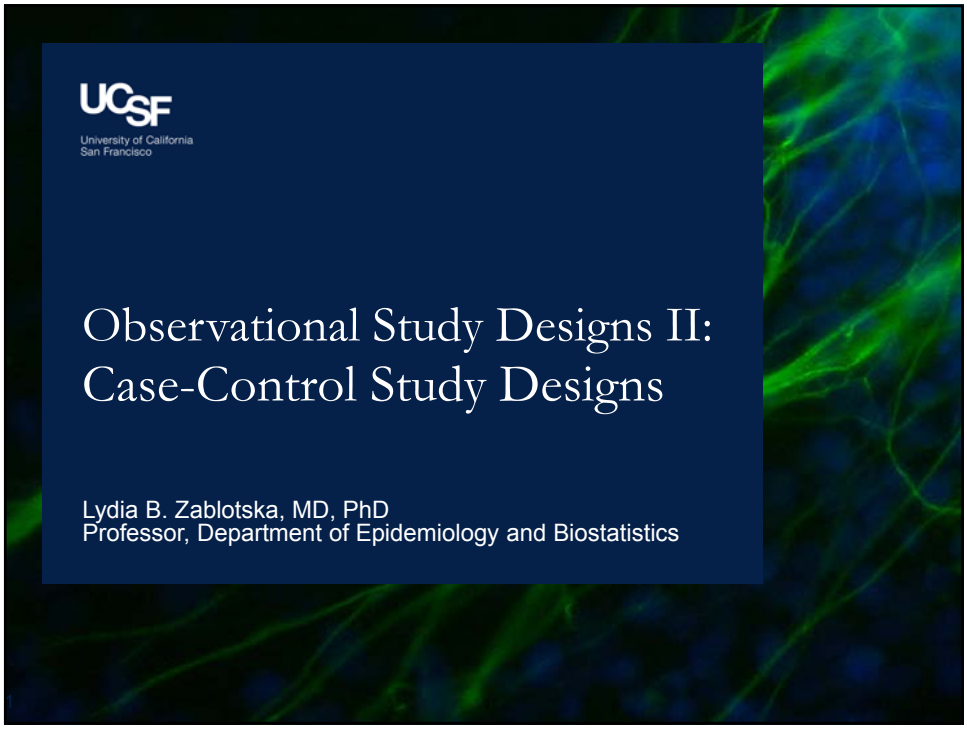


Highlights from Week #3

- **Introduced new ways to classify epidemiological study designs:**
  - Goal
  - Timing and directionality
  - Unit of analysis
  - Outcome
- **Discussed design features of cohort studies**
  1. Cohort specification (closed vs. open cohorts)
  2. Selection of exposed
  3. Selection of unexposed
  4. Sources of exposure information
  5. Classifying person-time (induction, latency, lag time, immortal PYs)
  6. Sources of outcome information
  7. Approaches to follow-up
- **Stressed the importance of field methods (research methods, research administration methods)**

2

UCSF

UCSF

Highlights from Week #3

- **Discussed how principles of experimental studies are applied to observational cohort studies**
  1. Randomization of treatment so groups are comparable on known and unknown confounders.
  2. Use placebo in order to reduce bias.
  3. Blinding to avoid bias in outcome ascertainment.
- **Answered a question: When does observational study emulate an RCT design?**
  1. The values of treatments under comparison correspond to well-defined interventions
  2. The conditional probability of receiving every value of treatment, though not decidedly by the investigators, depends only on the measured covariates (exchangeability)
  3. The conditional probability of receiving every value of treatment is greater than zero, i.e., positive (positivity)

UCSF

Learning Objectives

- Design features of case-control studies
  1. Selection of cases and controls
  2. Advantages and disadvantages of methods used to increase study accuracy:
     - Restriction
     - Matching
  3. Exposure classification
- Special types of case-control studies
  - Nested case-control
  - Case-cohort
  - Two-stage case-control study
  - Counter-matched studies

UCSF

UCSF

Case-Control Studies (from Epi 203)

- Imagine a population in which a cohort study could be conducted
  - Identify cases as you would in the cohort study
  - Sample from the study base (**person-time**) to determine exposure distribution in the population that gave rise to the cases = Controls
  - Controls must be sampled independent of exposure!
- + More efficient version of cohort study
- - Sampling creates new opportunities for bias

5

UCSF

Case-Control Key Concepts (from Epi 203)

1. Think of the selection of cases and controls as occurring from an underlying cohort (aka "study base")
2. Distinguish if underlying study base is fixed (closed) or dynamic (open) cohort
3. A inappropriate control group is usually the result of the inability to identify a well defined secondary study base (or the result of ignoring the study base concept entirely)
4. Incident sampling of cases is preferable to prevalent sampling
5. Plans for measurement are part of study design

6

UCSF

UCSF

Case-Control Key Concepts (from Epi 203)

➢1. Think of the selection of cases and controls as occurring from an underlying cohort (aka "study base")

2. Distinguish if underlying study base is fixed (closed) or dynamic (open) cohort

3. A inappropriate control group is usually the result of the inability to identify a well defined secondary study base (or the result of ignoring the study base concept entirely)

4. Incident sampling of cases is preferable to prevalent sampling

➢5. Plans for measurement are part of study design

7

UCSF

"TROHOC" STUDIES

- This disparaging term was given to case-control studies because their logic seemed backwards (trohoc is ?? spelled backwards) and they seemed more prone to bias than other designs.
- No basis for this derogation.
- Case-control studies are a logical extension of cohort studies and an efficient way to learn about associations.

8

UCSF

When is it desirable to conduct a case-control study?

- When exposure data are expensive or difficult to obtain
  - Ex: Pesticide and breast cancer study
- When disease has long induction and latent period
  - Ex: Cancer, cardiovascular disease
- When the disease is rare
  - Ex: Studying risk factors for birth defects
- When little is known about the disease
  - Ex. Early studies of AIDS
- When underlying population is dynamic
  - Ex: Studying breast cancer on Cape Cod

9

UCSF

Case-Control Studies as Approximation of Cohort Studies

OR is assumed to approximate RR

Q1. Where do you get information on the numerators for OR?

Q2. Where do you get the information for the denominators of OR?

10

UCSF

UCSF

Main Epidemiologic Study Designs for Testing Hypotheses

- Randomized controlled trial = Investigator assigns exposure and follows participants overtime to ascertain the outcome
- Cohort study = Analogous to RCT, but investigator does not assign exposure
- Case-control study = Analogous to cohort study, but with more efficient sampling

11

UCSF

Case-Control Studies as Approximation of Cohort Studies

Q1. Where do you get information on the numerators?

- Cases = numerators of the rates of disease being compared you would have measured in exposed and unexposed groups in the corresponding cohort study

Q2. Where do you get the information for the denominators?

- If this were a cohort study, you would have the total population (if you were calculating cumulative incidence) or total person-years (if you were calculating incidence rates) for both the exposed and non exposed groups, which would provide the denominators for the compared rates.
- Controls = relative size of the exposed and unexposed person-time in the study base ≈ person-time denominators you would have measured in the corresponding cohort study

12

UCSF

UCSF

Definitions of Cases and Controls (from Epi 203)

- Cases are the same people that would be cases in the underlying cohort study
  - Can randomly sample cases if sampling is independent of exposure
- Controls are a random, or conditionally random within strata, sample of study base
- Exposure distribution in the controls is the same as in the population that gave rise to the cases, conditional on matching factors

13

UCSF

Sources of Controls (from Epi 203)

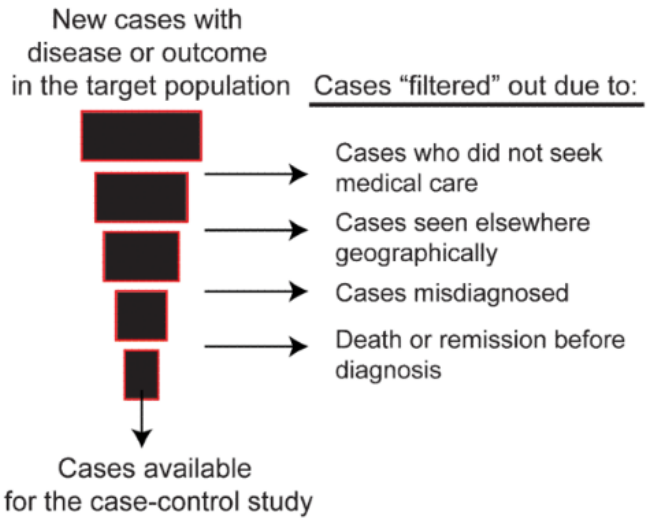
- Primary study base
  - Base population known
    - Study of Kaiser Permanente members
    - Study conducted within existing cohort, “Nested case-control”
- Secondary study base
  - Identify cases first, base population not known
    - Hospital-based case-control study
  - Must identify person-time contributed by persons who would have become a case in your study had they developed the disease

14

UCSF

UCSF

Selection of cases



15

UCSF

1. Design features of case-control studies

1. Selection of Controls

- General population controls

- Advantages

  Because of selection process, investigator is usually assured that they come from the same base population as the cases.

- Disadvantages

  Time consuming, expensive, hard to contact and get cooperation; may remember exposures differently than cases

16 RGL2008 Ch 8

UCSF

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Neighborhood Controls

- Sample residences
- May individually match cases to one or more controls residing in the same neighborhood
- If neighborhood is associated with exposure, must control for matching in the analysis
- Neighbors may not be the source population of the cases
  - Cases at a VA hospital

17

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Friend/Family Controls

- Being named as a friend control may be related to exposure
  - Reclusive people are less likely to be named
- Investigators depend on cases for identifying controls
- Friend groups often overlap, so persons with more friends are more likely to be selected as a control

18

UCSF

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Random Digit Dialing

- Case eligibility should include residence in a house with a telephone
- Probability of calling a number ≠ probability of contacting an eligible control
  - Households vary in the number of people, amount of time a person is at home, and the number of operating phones
- Method requires a great deal of time and labor

19

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Random Digit Dialing

- Answering machines, voicemail, and caller ID reduce response rates
- Cell phones reduce validity of assuming source population can be randomly sampled using this method
  - Recent CDC survey showed 2% increase in binge drinking compared to 2009 data – more cell phone numbers included, and average age of respondents decreased
- May not be able to distinguish business and residential numbers - difficult to estimate proportion of non-responders

20

UCSF

- Random Digit Dialing

American Journal of Epidemiology

Vol. 178, No. 1
DOI: 10.1093/aje/kwt133
Advance Access publication:
September 5, 2013

Practice of Epidemiology

Comparison of Address-based Sampling and Random-digit Dialing Methods for Recruiting Young Men as Controls in a Case-Control Study of Testicular Cancer Susceptibility

Bartholt Clagett, Katherine L. Nathanson, Stephanie L. Ciosek, Monique McDermoth, David J. Vaughn, Nandita Mitra, Andrew Weiss, Rachel Martonik, and Peter A. Kanetsky*

Table 2. Response Rates for Landline Random-digit Dialing, Cell-Phone Random-digit Dialing, and Address-based Sampling, Philadelphia, Pennsylvania, 2009–2012

	RDD, %		ABS, %
	Landlines	Cell Phones	
Screening contact rate[a]	59.6	64.3	
Screening cooperation rate[b]	54.2	63.5	
Screening response rate[c]	32.3	40.8	3.2
Field response rate[d]	35.3	10.0	53.2
Field response rate, ages 18–39 years[e]	32.7	10.0	51.8
Overall response rate[f]	11.4	4.1	1.7

21 Clagett et al 2013

UCSF

1.Design features of case-control studies

1. Selection of Controls

- Neighborhood controls

- crucial for studies of vector-borne disease, such as Lyme disease, for which risk is intrinsically linked to geographical location
- can help control for variation in ecological risk factors that are tied to geographical location, like vector and host habitat in the peridomestic environment
- Use RDD or residential telephone numbers from a commercial marketing database

Table 1. Final Disposition of Each Telephone Number Selected From the InfoUSA.com[a] Database for Enrollment in a Lyme Disease Case-Control Study, Connecticut, 2005–2007

Disposition	No.	%
Control interview completed	353	3.7
Answering machine/voicemail	4,704	48.8
Not eligible	1,849	19.2
Refused	1,141	11.8
No answer (after ≤3 attempts)	800	8.3
Nonworking number	465	4.8
Business number	196	2.0
Fax number	115	1.2
Busy signal	8	0.1
Other	5	<0.1

22 Clagett et al 2013

UCSF

UCSF

http://epiville.ccnmtl.columbia.edu/interactive/bias_interactive_exercise.html

1. Design features of case-control studies

1. Selection of Controls

- Hospital-Based Controls

- If not randomly selecting controls, must be cautious that control selection is independent of exposure
- Limit diagnoses for controls to conditions with no association with the exposure
  - May exclude most potential controls
  - Exclusion criteria only applies to the cause of the current hospitalization
- Reasonable to exclude categories of potential controls on the suspicion that a given category might be related to exposure
- Imprudent to use only a single diagnostic category as a source of controls

23

UCSF

http://epiville.ccnmtl.columbia.edu/interactive/bias_interactive_exercise.html

1. Design features of case-control studies

1. Selection of Controls

- Hospital-Based Controls, cont.

- May not represent exposure distribution in source population if exposure is associated with hospitalization, other diseases, or both
  - Example: Hospital-based study examining smoking and pancreatic cancer where controls are selected from persons admitted to the hospital for other conditions.
  - Special case: Berkson's bias (Berksonian bias, Berkson's fallacy)

24

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Hospital-Based Controls, cont.

- Berkson's Bias: a form of selection bias will occur if the exposure directly affects risk of being hospitalized, even if exposure is unrelated to the study disease or control diseases

UCSF

Berkson's bias

- Joseph Berkson, 1940: drawing cases and controls from hospitalized patients could lead to biased results
  - A type of selection bias
  - Occurs when **both** exposure and disease affect selection and specifically because they affect selection (e.g., rare and acute conditions)
  - Generates downward bias when both exposure and disease increase the chance of selection
  - Downward bias can induce a negative association if the association in the source population is positive but not as large as the bias
  - **Example:** c-c studies showed 10-fold increase in endometrial cancer odds for women taking estrogen regularly (estrogens →bleading→detection of cancer)
  - **Remedy:** use more than one control group, separately and aggregately

UCSF

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Hospital-Based Controls, cont.

Advantages

- Same selection factors that led cases to hospital also led controls to hospital
- Easily identifiable and accessible (so less expensive than population-based controls)
- Accuracy of exposure recall comparable to that of cases since controls are also sick
- More willing to participate than population-based controls

27

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Hospital-Based Controls, cont.

Disadvantages

- Since hospital based controls are ill, they may not accurately represent the exposure history in the population that produced the cases
- Hospital catchment areas may be different for different diseases

28

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Deceased Controls

- Not members of the source population for the cases
- If exposure is associated with mortality, dead controls will misrepresent exposure distribution in source population
- Even if cases are dead, generally better to choose living controls
- Do not need a proxy interview for living controls of dead cases

29

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Comparability of Information

- Comparability of information is often used to guide control selection and data collection
- BUT
  - Non-differential exposure measurement error does not guarantee that bias will be toward the null
  - Efforts to ensure equal accuracy of exposure data tend to produce equal accuracy of data on other variables
  - Overall bias due to non-differential error in confounders and effect modifiers can be larger than error produced by unequal accuracy of exposure data from cases and controls

30

UCSF

1. Design features of case-control studies

1. Selection of Controls

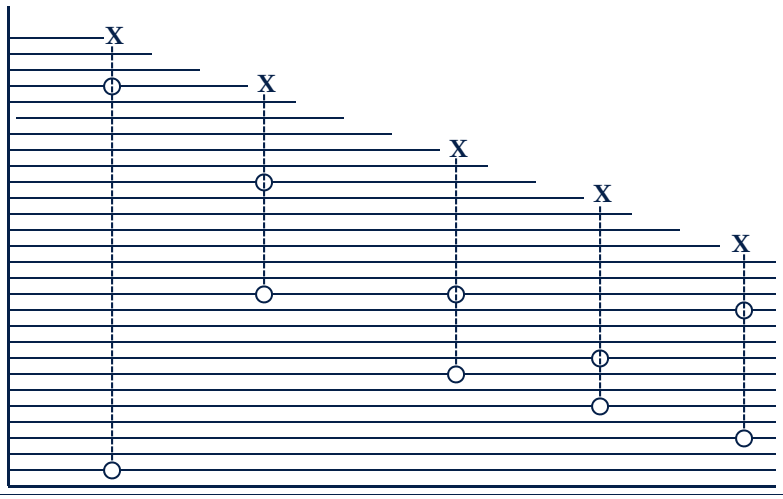
- Incidence Density Sampling

- Sample controls at a steady rate per unit time over period in which cases are sampled
- Probability of being selected as a control is proportional to amount of time person spends at risk of disease in source population
- Individual may be selected as a control while they are at risk for disease, and subsequently become a case
- Incidence density sampling or "risk set sampling" is a form of density sampling in which you match cases and controls on time

31

UCSF

Incidence Density Sampling (full cohort)



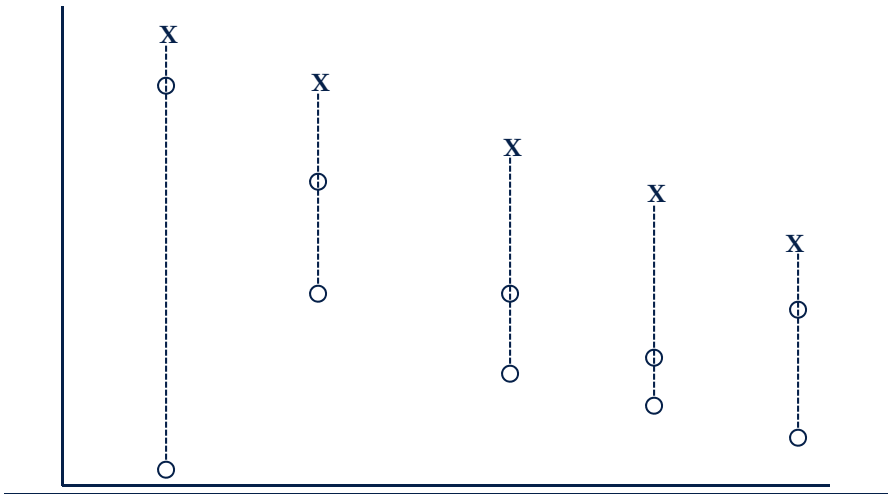
32

Time

UCSF

UCSF

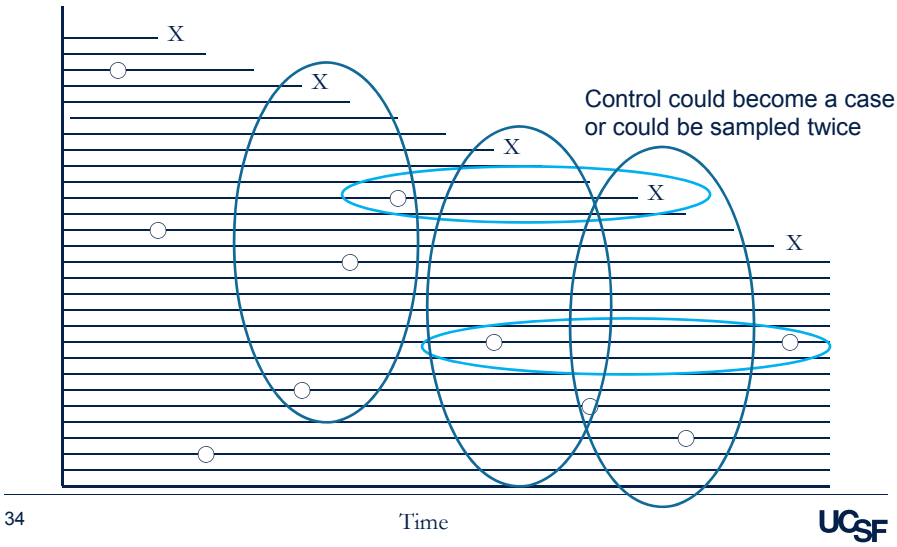
Incidence Density Sampling



33

UCSF

Incidence Density Sampling

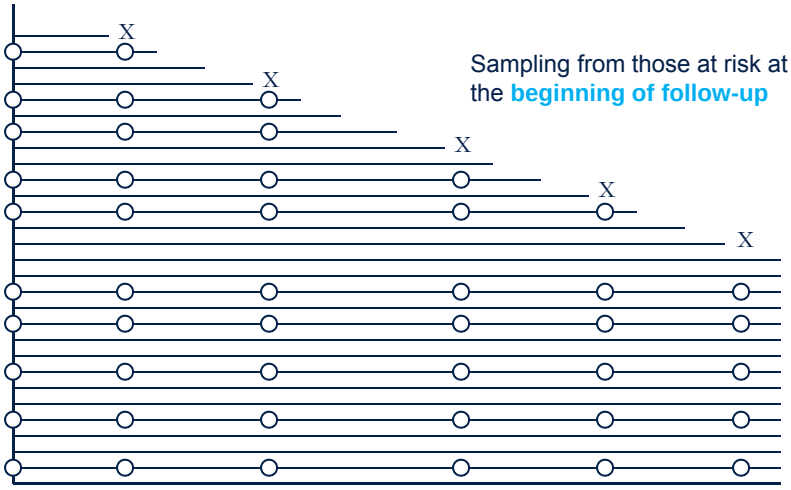


34

UCSF

UCSF

Case-Cohort Sampling



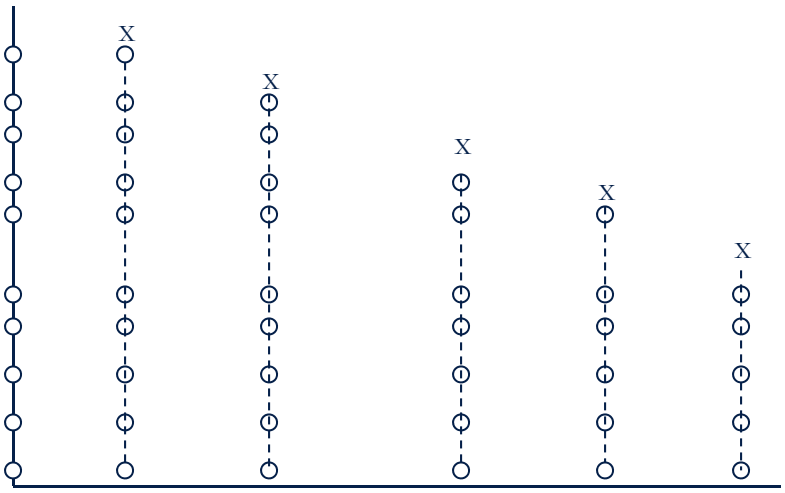
Sampling from those at risk at the **beginning of follow-up**

35

Time

UCSF

Case-Cohort Sampling



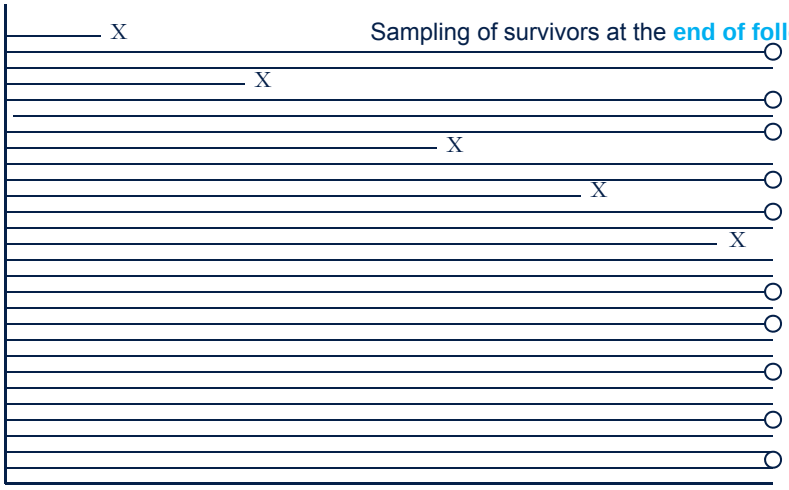
36

Time

UCSF

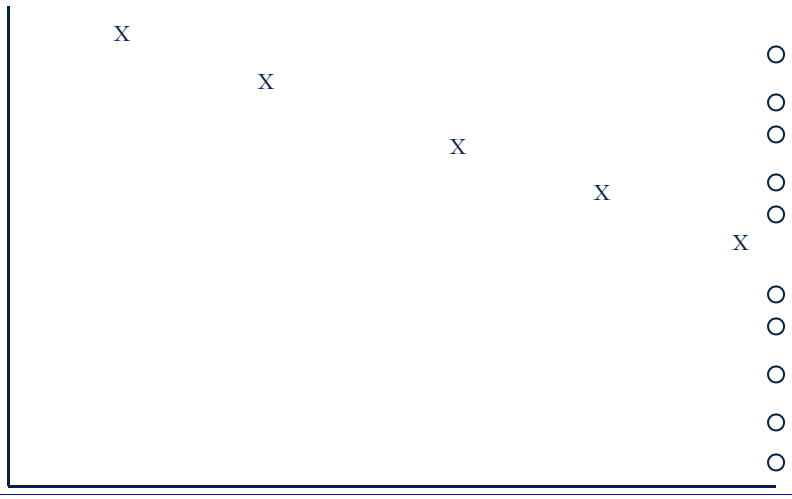
UCSF

Cumulative Sampling



37

Cumulative Sampling



38

1. Design features of case-control studies

1. Selection of Controls

- Measures of Association

Person-time Data			Case-control Data		
	Exposed	Unexposed		Exposed	Unexposed
Cases	x_1	x_0	Cases	x_1	x_0
Person-time	T_1	T_0	Controls	y_1	y_0

$$OR = \frac{x_1/y_1}{x_0/y_0} = \frac{x_1/[(y_1/T_1)T_1]}{x_0/[(y_0/T_0)T_0]} = \frac{x_1/(f \cdot T_1)}{x_0/(f \cdot T_0)} = \frac{x_1/T_1}{x_0/T_0} = IRR$$

- Can also estimate the risk ratio or incidence odds ratio from case-control studies
- The measure of association estimated by the OR depends on the control sampling scheme

39

UCSF

1. Design features of case-control studies

1. Selection of Controls

- Measures of association based on control sampling scheme

Control Sampling Method	Description	Measure of effect estimated by the OR
Case-cohort	Persons at risk of disease at baseline	Average risk ratio* (incidence proportion) Rate ratio
Density sampling	Proportional to person-time accumulated by persons at risk of disease during follow-up	Rate Ratio
Cumulative case-control	Persons at risk of disease who are non-cases at the end of follow-up	Incidence Odds Ratio Risk Ratio*

* Only need rare disease assumption when estimating the risk ratio from the odds ratio.

40

UCSF

UCSF

Findings from a hypothetical cohort study of 20,000 persons followed for 10 yrs

	Exposed	Non-exposed	Ratio
Cases	1813 (a)	952 (b)	
Non-cases	8187 (c)	9048 (d)	
Initial population size	10 000 (N_1)	10 000 (N_0)	
Person-years	90 635 (Y_1)	95 163 (Y_0)	
Incidence rate	0.0200 (I_1)	0.0100 (I_0)	2.00
Incidence proportion (average risk)	0.1813 (R_1)	0.0952 (R_0)	1.90
Incidence odds	0.2214 (O_1)	0.1052 (O_0)	2.11

1 < RR < IRR < OR

	Exposed	Non-exposed	OR
Cases	1813 (a)	952 (b)	
Controls			
From survivors (cumulative sampling)	1313 (c)	1452 (d)	2.11
From source population (case–cohort sampling)	1383 (c)	1383 (d)	1.90
From person-years (density sampling)	1349 (c)	1416 (d)	2.00

41

UCSF

Published by Oxford University Press on behalf of the International Epidemiological Association International Journal of Epidemiology 2012;41:1480–1489
doi:10.1093/ije/dys147

Case–control studies: basic concepts

Jan P Vandenbroucke[1]* and Neil Pearce[2,3]

New developments
(why we don't need a rare disease assumption)

- Of all case–control studies involving incident cases, 82% sampled from a dynamic population; only 18% of studies sampled from a fixed cohort
- Case–control studies can be conducted in a dynamic population, and the resulting OR directly estimates the IRR from this dynamic population, *provided that* either the exposure distribution is in steady state in the dynamic population, or sampling of control subjects is matched on time in a dynamic population

42 Vandenbroucke and Pearce 2012

UCSF

UCSF

Exercise

Q: What effect estimate could be obtained from cumulative case-control studies?

A) OR estimating IRR

B) OR estimating RR

C) Both

43

UCSF

Exercise Answer

- Question: What effect estimate could be obtained from cumulative case-control studies?
- Answer: C

The odds ratio in a cumulative study estimates either rate ratio (IRR) or risk ratio (RR), but only if disease is rare. The approximation to the rate ratio is closer than the approximation to the risk ratio. Each control represents a sample from a fixed number of people, rather than person-time.

In the case-cohort study, the people who serve as the source of the control sample constitute the denominator for the risk, but in the cumulative case-control study, controls are drawn just from non-cases, which is not the full denominator. With rare disease, almost everyone is a non-case, so the odds ratio in a cumulative case-control study is a good approximation to the risk ratio, but will be even closer to the rate ratio.

44

UCSF

1. Design features of case-control studies

2. Methods used to increase study accuracy

- Restriction

- Limit who can be included in a study to prevent confounding
- Restricts pool of potential participants
- While this may decrease generalizability, it may help with validity
- If a population is too heterogeneous might not be able to answer any questions.

45

UCSF

1. Design features of case-control studies

2. Methods used to increase study accuracy

- Apportionment of study subjects

- Try to improve study efficiency by selecting certain proportion of subjects into groups.
- Can be based on exposures and disease status.
- Common in case-control studies: selecting multiple controls per case.
- Maximum efficiency in case-control study:
  - n/(m+n) where m=# cases, n=# controls. (Under the null and when no need to stratify)
  - 1:1 = 50%, 1:2=66%, 1:3=75%, 1:4=80%, 1:5=83%
- Most cost-efficient ratio of controls to cases (under null):

  = sqrt (C_1 / C_0), C_1 = cost of case, C_0 = cost of control

46

UCSF

UCSF

1. Design features of case-control studies

2. Methods used to increase study accuracy

- Matching

- What is the key objective in matching?
- What types of factors should be matched?

47

UCSF

Exercise

- Q: Make a DAG that explicitly accounts for the matching process:
  - E. Exposure
  - D. Disease
  - C. Matching variable. Assume it is a confounder of the E -> D relationship
  - S. Matching (selection) process
- What are the consequences of matching for estimation/ testing of an E -> D effect?

48

UCSF

UCSF

Exercise Answer

- Question: Make a DAG that explicitly accounts for the matching process
- Answer:

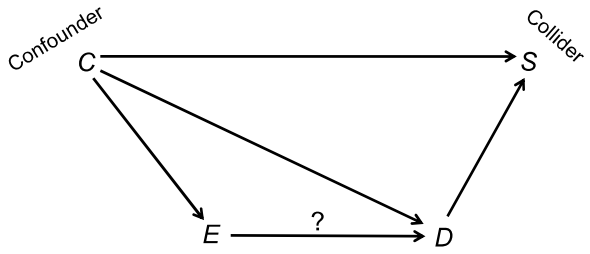


Figure 1 A causal structure.
Note: The question mark denotes the effect of interest.

49

UCSF

1. Design features of case-control studies

2. Methods used to increase study accuracy

- Matching

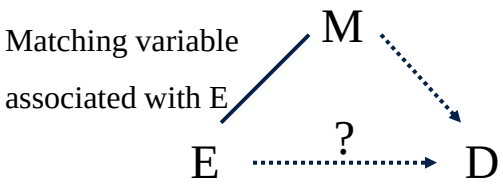
- Selection of reference series (unexposed, or controls) by making them similar to index subjects on distribution of one or more potential confounders.
- This balancing of subjects across matching variables can give more precise estimates of effect with proper analysis.
- Key advantage of matching is not to control for confounding (which is done through analysis), but to control for confounding more efficiently!
- Matching must be accounted for in one's analysis:
  - matching controls to cases on an exposure can introduce selection bias

50

UCSF

UCSF

Case-control matching introduces selection bias



- By matching on M, we have eliminated any association between M and D in the total sample.
- But selection is differential wrt both exposure and disease.
- Exposure distribution (E) of controls is now like the cases'.
- The controls' disease risk falsely elevated by the increased prevalence of another risk factor
- If M-E not associated, then matching will not lead to bias, but may be inefficient.

51

UCSF

Example: Beta-carotene and lung cancer

Base Population

	Smokers		Non-smokers		Total	
Beta-carotene	D	PY	D	PY	D	PY
Low	30	12000	6	20000	36	32000
High	10	12000	6	60000	16	72000
Total	40	24000	12	80000	52	104000
	IDR = 3.00		IDR = 3.00		IDR = 5.06	

Unmatched Case-control Study (All 52 cases and 52 randomly selected controls)

	Smokers		Non-smokers		Total	
Beta-carotene	D	not D	D	not D	D	not D
Low	30	6	6	10	36	16
High	10	6	6	30	16	36
Total	40	12	12	40	52	52
	OR = 3.00		OR = 3.00		OR = 5.06	
			OR_{MH} = 3.00 (1.16 - 7.73)			

Matched Case-control Study (All 52 cases and 52 matched controls)

	Smokers		Non-smokers		Total	
Beta-carotene	D	not D	D	not D	D	not D
Low	30	20	6	3	36	23
High	10	20	6	9	16	29
Total	40	40	12	12	52	52
	OR = 3.00		OR = 3.00		OR = 2.84	
			OR_{MH} = 3.00 (1.31 - 6.88)			

52

UCSF

UCSF

1. Design features of case-control studies

2. Methods used to increase study accuracy

- Matching

Types of matching

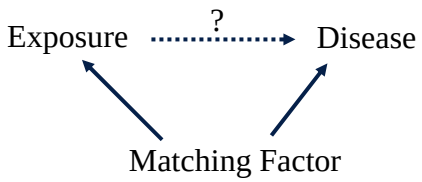
- Individual – one or more comparison subjects is selected for each index subject (fixed or variable ratio)
- Category – select comparison subjects from the same category the index subject belongs to (male, age 35-40)
- Frequency – Total comparison group selected to match the joint distribution of one or more matching variables in the comparison group with that of the index group (~category)
- Caliper – select comparison subjects to have the same values as that of the index
  - Fixed caliper – criteria for eligibility is the same for all matched sets (age ± 2 years)
  - Variable caliper – criteria for eligibility varies among the matched sets (select on value closest to index subject, i.e. nearest neighbor)

53

UCSF

Appropriate matching

(Matching factor is a confounder)



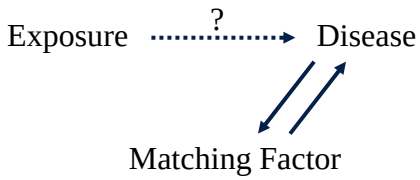
	Analysis	
Study Design	Stratified	Crude
Match	Valid/Max Precision	BIAS
No matching	Valid/Reduced Precision	BIAS

54

UCSF

UCSF

Unnecessary matching:
(Matching factor is unrelated to exposure)



	Analysis	
Study Design	Stratified	Crude
Match	Valid/Reduced Precision	Valid/Max Precision
No matching	Valid/Reduced Precision	Valid/Max Precision

55

UCSF

Overmatching

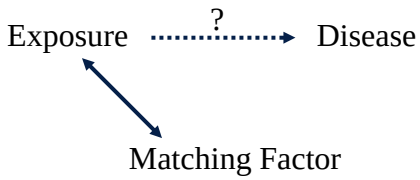
1. Loss of information due to matching on a factor that's only associated with exposure (non-confounder). Still need to undertake stratified analysis to address selection bias, but this was unnecessary.
2. Irreparable selection bias due to matching on factor affected by exposure or disease.

56

UCSF

UCSF

Overmatching:
(Matching factor is associated with exposure)

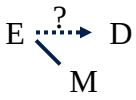


Study Design	Analysis: Stratified	Analysis: Crude
Match	Valid/Reduced Precision	BIAS
No matching	Valid/Reduced Precision	Valid/Max Precision

57

UCSF

Overmatching



Case-Control (unmatched)

	E	not E	
Cases	44	110	154
Controls	26	130	156

OR_{crude} = 2.00 (1.16 - 3.46)

OR_{MH} = 2.00 (1.09 - 3.66)

Case-Control (matched)

	E	not E	
Cases	44	110	154
Controls	29	125	154

OR_{crude} = 1.72 (1.01 - 2.94)

OR_{MH} = 2.00 (1.09 - 3.64)

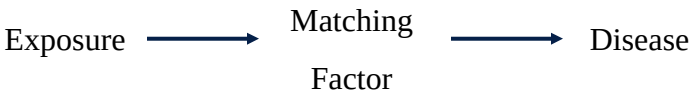
Study Design	Analysis: Stratified	Analysis: Crude
Match	Valid/Reduced Precision	BIAS
No matching	Valid/Reduced Precision	Valid/Max Precision

58

UCSF

UCSF

Matching on a intermediate variable



	Analysis	
Study Design	Stratified	Crude
Match	BIAS	BIAS
No matching	BIAS	Valid/Max Precision

59

UCSF

Summary

- When might matching be a good strategy?
- Decision should reflect cost / benefit tradeoff
- Costs:
  - Cannot estimate effect of matching variable on disease.
  - May be not cost effective if limits potential study subjects.
  - Might overmatch.
- Benefits:
  - May provide more efficient study and manner to control for potential confounding.
- Compare sample sizes needed to obtain a certain level of precision with matching versus no matching (assuming correct analysis)
- *One should not automatically match!*

60

UCSF

UCSF

1. Design features of case-control studies

3. Exposure Classification

- Same principles as discussed for cohort studies
- Cases' exposure should be classified as of the time of diagnosis or disease onset, accounting for induction time hypotheses
- Controls should be classified according to their exposure status at the time of selection, accounting for induction time hypotheses

61

UCSF

1. Design features of case-control studies

4. Timing of Exposure Classification

- Selection time does not necessarily refer to the time at which a control is first identified
  - For hospital-based controls, selection time may be date of diagnosis for the disease that resulted in the current hospitalization
  - Date of interview is often used if there is not an event analogous to the cases' date of diagnosis
- Interviewers should be blinded to case-control status whenever possible

62

UCSF

2. Special types of case-control studies

Variations in case-control study designs

- Nested case-control
- Case-cohort
- Two-stage case-control study
- Counter-matched studies

63

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Nested case-control study

Table 1. Descriptive characteristics [*n* (%)] of cases and controls identified during follow-up (1986–2006).

Characteristic	Cases (*n* = 137)	Controls (*n* = 863)	*p*-Value[a]
Radiation dose, mGy [mean ± SD (range)][b]	132.3 ± 342.6 (0–3220.0)	81.8 ± 193.7 (0–2600.0)	0.119[c]
Year of birth			0.988
1923–1929	10 (7)	67 (8)	
1930–1939	38 (28)	222 (26)	
1940–1949	43 (31)	285 (33)	
1950–1959	37 (27)	234 (27)	
1960–1965	9 (7)	55 (6)	
Areas of study			0.938
Cherkasy Oblast	7 (5)	60 (7)	
Chernihiv Oblast	11 (8)	77 (9)	
Dnipropetrovsk Oblast	26 (19)	155 (18)	
Kharkiv Oblast	17 (12)	107 (12)	
Kyiv Oblast	27 (20)	183 (21)	
Kyiv City	49 (36)	281 (33)	
Type of residence at time of interview			0.090
Urban	101 (74)	680 (79)	
Rural	19 (14)	151 (18)	
Other	10 (7)	32 (4)	
Unknown	7 (5)	0 (0)	
Age at first exposure (years)			0.970
20–34	31 (23)	207 (24)	
35–41	36 (26)	221 (26)	
42–49	40 (29)	239 (28)	
50–63	30 (22)	196 (23)	
Education			0.474
≤ 8 years	16 (12)	131 (15)	
High school	46 (34)	341 (40)	
Trade school	34 (25)	200 (23)	
College	34 (25)	188 (22)	
Unknown	7 (5)	3 (0)	

64 Zablotska et al. 2013.

UCSF

UCSF

Model description	Cases (*n*)	ERR/Gy (95% CI)	*p*-Value[a]	*p* Interaction[b]
All cases	137	1.26 (0.03, 3.58)	0.041	
Excluding cases with direct interviews	117	2.38 (0.49, 5.87)	0.004	
< 2 years from start of chemotherapy				
Leukemia subtype				
Non-CLL	52	2.21 (0.05, 7.61)	0.039	0.888
CLL	65	2.58 (0.02, 8.43)	0.047	

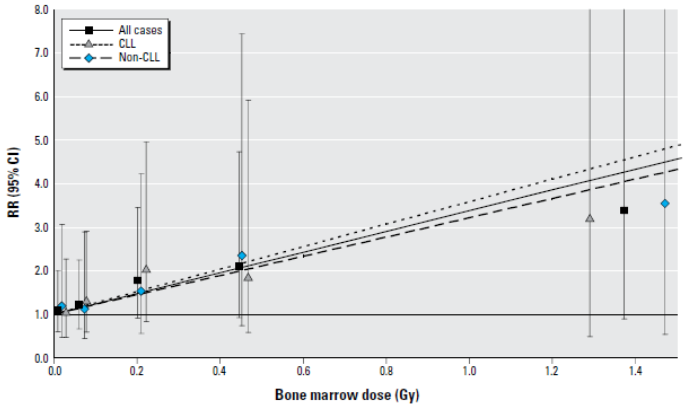


Figure 1. RRs (95% CIs) of leukemia by categories of radiation dose and fitted linear dose–response models. For display purposes, we added offsets to category mean doses on the abscissa coordinate to separate the overlapping estimates (10 mGy for non-CLL and 20 mGy for CLL analyses, respectively).

65 Zablotska et al. 2013.

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Case-cohort studies

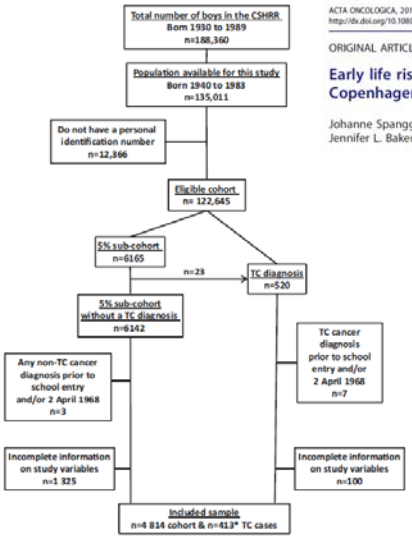
- Advantages:
  - Could conduct several analyses from a single cohort using the same control group (since we are taking a sample from the cohort at enrollment, the control group could be compared with any number of case groups)
- Disadvantages:
  - Need a larger sample of controls since some of them could be later sampled as cases, decreasing precision compared with a typical case-control study
    - If incidence rate is 20%, then need to select 25% more controls compared to the typical study to ensure that there will be as many controls who never become cases in the study
    - Not very important if the disease is uncommon, i.e. incidence rate<5%

66

UCSF

UCSF

Example, case-cohort design



ACTA ONCOLOGICA, 2017
http://dx.doi.org/10.1080/0284186X.2016.1266085

Taylor & Francis

ORIGINAL ARTICLE

Early life risk factors for testicular cancer: a case-cohort study based on the Copenhagen School Health Records Register

Johanne Spanggaard Piltoft[a], Signe Benzon Larsen[a], Susanne Oksbjerg Dalton[a], Christoffer Johansen[a,b], Jennifer L. Baker[c], Luise Cederkvist[a] and Ingelise Andersen[d]

67 Piltoft et al. 2017

UCSF

How to analyze case-cohort data?

- Cox proportional hazards model
- Need to adjust variance
  - Several methods exist

TABLE 3. Parameter estimates and standard errors for log (exposure + 1) using the full-cohort, case-cohort, and nested case-control designs

Method	Total *n*	$\hat{\beta}$	Naive SE	Robust SE
Full Cohort	679	0.766	0.174	0.170
Case-cohort Prentice [2] method	182	0.712	0.175	0.241
Case-cohort Barlow [8] method	182	0.740	0.176	0.259
Case-cohort Self and Prentice [7] method	182	0.749	0.176	0.265
Nested case-control 1 case : 3 controls	190	0.913	0.274	

Straightforward with no adjustment of the variance

Controls are weighted in proportion to the sampling scheme from the underlying cohort

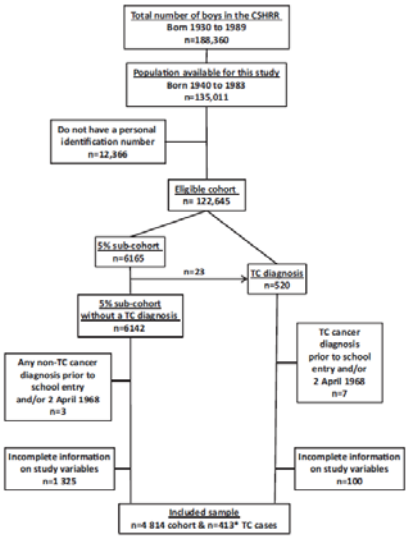
Poor performance!

68 Barlow et al. 1999

UCSF

UCSF

Example, case-cohort design



ACTA ONCOLOGICA, 2017

Taylor & Francis

ORIGINAL ARTICLE

Early life risk factors for testicular cancer: a case-cohort study based on the Copenhagen School Health Records Register

Johanne Spanggaard Piltoft[a], Signe Benzon Larsen[a], Susanne Oksbjerg Dalton[a], Christoffer Johansen[a,b], Jennifer L. Baker[c], Luise Cederkvist[a] and Ingelise Andersen[d]

Table 2. Hazard ratio for the age- and multiple-adjusted associations between early life risk factors and testicular cancer among 5227 men from the Copenhagen School Health Records Registry.

Risk factor	Age adjusted HR (95% CI)	Multiple adjusted[b] HR (95% CI)
Cryptorchidism (any kind)	3.60 (2.79–4.65)	3.46 (2.67–4.48)
Birth weight (g)		
<2500	1.71 (1.16–2.50)	1.48 (1.00–2.17)
2500–2999	1.28 (0.96–1.71)	1.22 (0.91–1.63)
3000–3999	Reference	Reference
4000–4499	0.93 (0.69–1.25)	0.98 (0.73–1.33)
4500+	0.82 (0.49–1.36)	0.90 (0.54–1.49)
Birth order		
1	Reference	Reference
2	1.16 (0.94–1.44)	1.16 (0.94–1.44)
3	0.98 (0.72–1.34)	0.96 (0.71–1.31)
4+	0.68 (0.44–1.06)	0.70 (0.45–1.09)

CI: confidence interval; HR: hazard ratio. [a]p-Value for likelihood ratio test; [b]All models adjusted for the other covariates and stratified by birth cohort.

69 Piltoft et al. 2017

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Two-stage case-control studies

Table 2. Data from hypothetical two-stage case–control study.[a]

Exposure	Stage I Cases	Stage I Controls
Factory workers	40	20
Employed elsewhere	460	480
Total	500	500

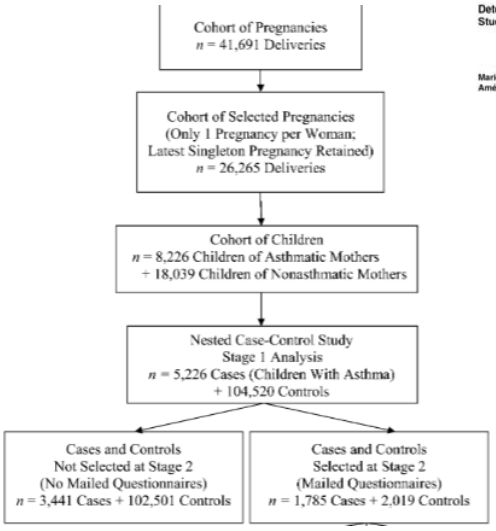
	Stage II Cases		Stage II Controls	
Exposure	Factory	Elsewhere	Factory	Elsewhere
Smokers	30	37	10	42
Nonsmokers	10	133	10	128
Total	40	170	20	170

- Cases and controls randomly sampled from the community
- Exposure tied to employment status → use existing data
- Stage II: interview all exposed (all workers) and a sample of unexposed
- Use the information on the disease-exposure relationship from Stage I (1,000 subjects) and the confounder information from Stage II (340 subjects)
- Adjust exposure-outcome estimate and its variance for the method of sampling in Stage II

70 Tager 2000

UCSF

UCSF



American Journal of Epidemiology

DOI: 10.1093/aje/kwn260
Advance Access publication November 24, 2008

Original Contribution

Determinants of the Incidence of Childhood Asthma: A Two-Stage Case-Control Study

Marie-Josée Martel, Évelyne Rey, Jean-Luc Malo, Sylvie Perreault, Marie-France Beauchesne, Amélie Forget, and Lucie Blais

71 Martel et al. 2008

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Counter-matched studies

- Nested case-control study uses only information for the cases and controls actually used in the case-control study → lower precision of the exposure effect estimates
- Additional loss in efficiency because of matching on potential confounders (exposure-concordant pairs (risk sets) do not contribute to the estimation of the exposure effect)
- The counter-matching design uses stratified sampling of exposures (not confounders) and adjusts for the sampling strategy by weighting the probability of being sampled from a given exposure stratum in the risk set at the time the case appears.

Table 3. Situations in which countermatching designs may be helpful.[a]

Situation	Goal	Comments
Crude exposure surrogate available on all subjects; more detailed exposure available on limited sample	Assess the effect of the detailed exposure variable	More efficient (↓ variance) than nested case–control The more accurate the surrogate the closer the efficiency to the full cohort relative to a simple nested design Exposure strata may be optimal when based on case distribution of true exposure and equal-size strata
Exposure data available on all subjects and confounders to be collected on a subset	Assess confounder-adjusted exposure effect	Confounder effect and variance seemed to be captured better with 1:3 countermatched design than with 1:1 Confounder effect may not be estimated any more accurately or precisely than simple nested design, but the main exposure effect more is accurate and the variance smaller
Exposure data available on everyone, and another covariate is collected on the sample	Investigate interaction between exposure and covariate	For most situations in which exposure is associated with disease, countermatching may be more efficient for the estimation of the interaction parameter (↓ variance)
Exposure 1 data available on all subjects and exposure 2 data available only on a sample of subjects	Investigate exposure 1 and main effect of exposure 2	Low efficiency Proposes strategy for use of additional randomly sampled controls, which improves the efficiency of estimation of the exposure 2 while retaining the countermatching efficiency for the primary exposure (exposure 1)

72 Tager 2000

UCSF

UCSF

Available online http://breast-cancer-research.com/content/6/3/R199

Research article

Open Access

Study design: Evaluating gene–environment interactions in the etiology of breast cancer – the WECARE study

Jonine L Bernstein[1], Bryan Langholz[2], Robert W Haile[2], Leslie Bernstein[2], Duncan C Thomas[2],

WECARE - Women's Environment, Cancer, and Radiation Epidemiology.

- 700 women with asynchronous (2nd primary) bilateral breast cancer were individually matched to 1400 controls with unilateral breast cancer on date and age at diagnosis of the 1st primary breast cancer, race, and registry region, and counter-matched on radiation therapy (RT)
- Each triplet (1case:2controls) comprised two women RT+ and one woman RT-
- Optimized the informativeness of the collected dosimetry data by increasing the variability of radiation dose within the case–control sets and enhanced ability to detect radiation–genotype interactions
- Counter-matching on exposure (or a correlate of exposure) is more efficient than simple random sampling for assessing exposure dose–response and gene–exposure interactions

73

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Counter-matched studies: effects on statistical power

Statistical power (percentages) to detect a log-linear trend from various designs for 700 risk sets

		Random sampling		Counter-matching		
				CM 1:2		CM 2:1
2 Gy rate ratio	All[a]	NCC 3	NCC 4	PN = 100%, PP = 100%	PN = 90%, PP = 90%	PN = 100%, PP = 100%
1.4	79	61	63	70	67	63
1.5	87	69	78	84	80	78
1.6	97	86	93	95	92	92
1.7	100	95	97	99	96	97

CM 1:2, counter-matched design in which one member of the triplet is registry radiation treatment negative (RRT–) and two are registry radiation treatment positive (RRT+); CM 2:1, counter-matched design in which two members of the triplet are RRT– and one is RRT+; NCC 3, nested case-control study in which each case is randomly matched to two controls; NCC 4, nested case-control study in which each case is randomly matched to three controls; PN, predictive negative value; PP, predictive positive value. [a]Full cohort power where radiation dose is available for all controls in the risk set and used in the analysis; this is the upper limit on any case-control analysis.

74 Bernstein et al. 2004

UCSF

Contralateral breast cancer after radiotherapy among *BRCA1* and *BRCA2* mutation carriers: A WECARE Study Report

Jonine L. Bernstein[a,*], Duncan C. Thomas[b], Roy E. Shore[c,d], Mark Robson[e],

Effect of radiation exposure on risk of developing contralateral breast cancer by *BRCA1 or BRCA2* carrier status.

Radiation exposure	No *BRCA1/BRCA2* mutation				*BRCA1* or *BRCA2*			
	Case *n* (%)	Control *n* (%)	Rate ratio (RR)[a]	95% confidence interval (CI)	Case *n* (%)	Control *n* (%)	RR[a]	95% CI
No	261 (51)	229 (50)	1.0		40 (42)	9 (41)	1.0	
Yes	246 (49)	908 (50)	1.1	0.9–1.4	56 (58)	53 (59)	1.4	0.6–3.3
Estimated dose to contralateral breast								
No	261 (52)	229 (50)	1.0		40 (42)	9 (41.0)	1.0	
0 < 1.0 Gy	128 (25)	502 (27)	1.0	0.8–1.3	35 (36)	27 (30.0)	1.9	0.7–4.8
⩾1.0 Gy	118 (23)	406 (23)	1.2	1.0–1.6	21 (22)	26 (29.0)	1.0	0.4–2.8

The effect of carrying *BRCA1* or *BRCA2* mutations on risk of developing contralateral breast cancer.

	Mutation carrier status	Cases (%)	Controls (%)	Rate ratio[a]	95% confidence interval (CI)
BRCA1 or *BRCA2*	No	507 (84)	1137 (95)	1.0	
	Yes	96 (16)	62 (5)	4.5	3.0–6.8

Conclusion: No clear indication that carriers of deleterious BRCA1/ BRCA2 mutations were more susceptible to the carcinogenic effects of radiation than non-carriers. **No G x R interaction**.

75 Bernstein et al. 2013

UCSF

2. Special types of case-control studies

Variations in case-control study designs

Counter-matched studies

- Advantages:

– substantial gains in statistical efficiency compared to simple random sampling at a first stage of analyses focused by design on variables correlated with the counter-matching variable

- Disadvantages:

– will lose efficiency at a second stage of analyses aimed at variables independent of the counter-matching variable and not conceived as a part of the initial study

  – E.g., counter-matching is more efficient than random sampling for assessing the main effect of R and R x G interaction effects, but less efficient for estimating the main effect of G

76

UCSF

UCSF

Summary

- Case-control studies should be conducted in such a way that they approximate a cohort study design
- Subject selection and exposure and outcome assessment play crucial role in getting valid results from case-control studies which would approximate the results from a cohort study design
- Multiple opportunities for selection and measurement/information bias, which could undermine or even overwhelm study findings

77

UCSF