PBIO 504

Case-Control Studies

Types of Epidemiological Studies

- Clinical case series
- Cross-sectional study
- Case-control study
- Cohort study
- ***Randomized clinical trial***

Increasing value of evidence

Case-Control Studies

- Basic strategy: find persons with a specific type of disease (“*cases*”) and compare them to persons without that disease (“*controls*”).
- It is an extension of the clinical case series, with the addition of a comparison group. **Having controls allows for a comparison of the exposure histories of the two groups**.
- This study design is suitable for rare diseases like cancer that have a long period for the disease to develop, as it is relatively quick and lower in costs, compared to a prospective design.

Textbook chapter 11

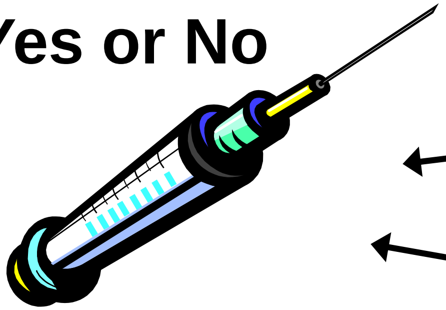
Case-Control Studies

- It follows in logic from effect (cases with disease) to cause (exposure)
- Unlike cohort or cross-sectional studies, subjects are selected for study because they have the disease of interest (case) or they do not have that disease (control)
- The object of study is the difference in exposure between cases and controls

Retrospective Assessment of Risk Factors

Use of injection drugs (increases risk of acquiring HBV or HCV)

Yes or No



1. Start with finding subjects

Liver cancer cases

Controls

2. Look back at risk factors

the past *today* time

Finding the Cases

- This is the easy part of the case-control study, as you simply go to where the cases are found (e.g. a specialty clinic where they get diagnosed and treated) and ask them to participate in the study.
- Use precise criteria for defining the disease
- Decide on including *incident* vs. *prevalent cases*

Defining the Disease

Method 1

- You are interested in studying influenza.
- You ask people if they felt sick last week, and if they say yes, they are the cases for the study.

Method 2

- You are interested in studying influenza.
- You ask people if they had a fever of 101 degrees or higher, with a sore throat and cough last week. If they say yes to all three symptoms, they are the cases for the study.

Choosing the Controls

- General strategy: controls should be from the same source population as the cases *{similar opportunities for exposure to the risk factors}*
- Some common types of controls :

  Random sampling from phone lists, etc…

  Friends or relatives of cases

  Neighborhood controls

  Hospital controls

Examples of Sources of Cases and Controls

Cases	Controls
All cases in the community (population-based study)	Probability sample of the community population
Cases from all area hospitals	Sample of non-cases from those hospitals
Cases from a single hospital	Sample of non-cases from that same hospital
Cases from several scattered hospitals	Sample of people from the same neighborhoods as the cases
Cases selected from any above sources	Spouses, siblings, or friends of cases

Matched and Unmatched Controls

- Controls do not have to be matched to cases on any specific characteristics, although there are often good reasons to do so (see next slide).
- Matched controls can be individually matched (also called pair matching) on one or more characteristics, or they can be frequency matched.

Matching in Case-Control Studies

- Rationale for matching:
  - Efficiency: to insure that the distribution of an external risk factor is similar between cases and controls, e.g. age and/or sex
- An unbiased odds ratio is obtained only if the matching factors are controlled for in the analysis, or if a matched analysis is performed, or if there is no difference between results of matched and unmatched analyses

Types of Matching

Pair Matching

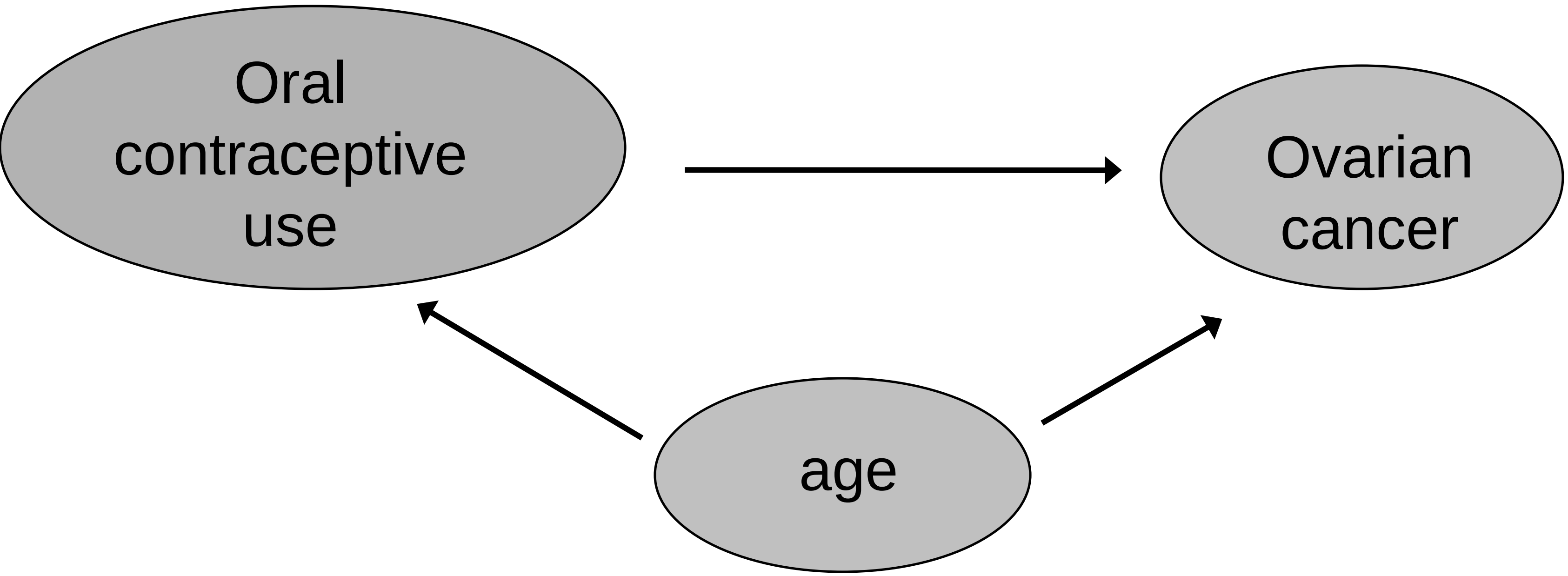
- You must find a control who exactly matches the important features of a specific case.
- The data from the pair must be linked in the statistical analysis.
- <u>Example</u>: a 20-year old male case is matched to a 20-year old male control

Frequency Matching

- You find a group of controls who are similar to the cases on the important features.
- The data are not paired in the statistical analysis.
- <u>Example</u>: 35% of the cases are males of ages 20-25, and we try to get 35% of the controls in that age and sex group. It does not have to be exactly 35%, just close to that figure.

Example of Why Matching the Controls May be Important

Example of matching on age



Age can influence both OC use and cancer, so case and control groups should have a similar age distribution.

So far we have talked about the disease. What about the risk factors in a case-control study?

<u>Examples</u>: chemical exposure, caffeine, smoking

Retrospective Exposure Assessment: how do we find out about past exposures

INDIRECTLY

- Questionnaires
- Medical records
- Factory records
- Census data
- Geographic mapping

DIRECTLY

(BIOMARKERS)

- Stored biological samples from the past
- Fresh tissue (but the usefulness is limited by the half-life of the marker)

Analysis of Case-Control Studies

Note how the table is set up. If the cells are mixed up, your calculation may be wrong

A = number of exposed cases
B = number of unexposed cases
A+B = total number of cases
C = number of exposed controls
D = number of unexposed controls
C+D = total number of controls

		Exposure Yes	No
Disease	Cases	A	B
	Controls	C	D

odds=p/(1-p)

Odds Ratio =

odds of exposure among the cases (A/B) **divided by** odds of exposure among the controls (C/D)

Odds Ratio in Case-Control Studies

Study of oral contraceptive use and ovarian cancer risk

Null hypothesis: OR = 1 (cases and controls have the same odds of having been exposed to the risk factor of interest, OC use)

	Exposed	Not exposed
Cases	119	317
Controls	68	319

Odds Ratio = (A/B) / (C/D) = (119/317) / (68/319) = 1.76

Interpretation: *The odds of having used OC among those with ovarian cancer are* 1.76 *times higher than the same odds among those with no ovarian cancer*

How to calculate the 95% Confidence Interval for the Odds Ratio

	Exposed	Not exposed
Cases	119	317
Controls	68	319

OR=1.76 and 95% CI: (1.25 ; 2.46)

<u>Lower 95% CI</u> = e { ln (OR) – 1.96*sqrt(1/A + 1/B + 1/C + 1/D) }

= e { ln (1.76) – 1.96*sqrt(1/119 + 1/317 + 1/68 + 1/319) } =

= e { ln (1.76) - 0.336} = e {0.56 – 0.336} = 1.25

<u>Upper 95% CI</u> = e { ln(OR) + 1.96*sqrt(1/A + 1/B + 1/C + 1/D) }

= e { ln (1.76) + 0.336} = e {0.56 + 0.336} = 2.46

Interpreting Confidence Interval

- From the previous example:

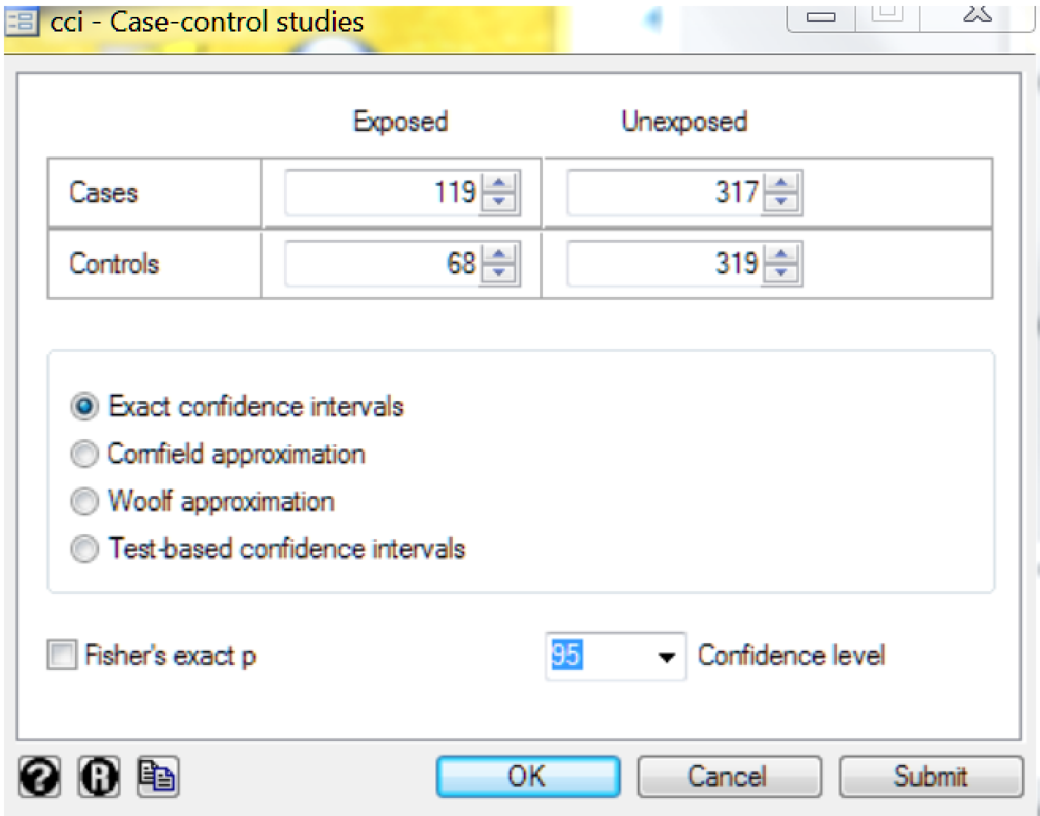
  Odds Ratio = 1.76 for OC use and ovarian cancer

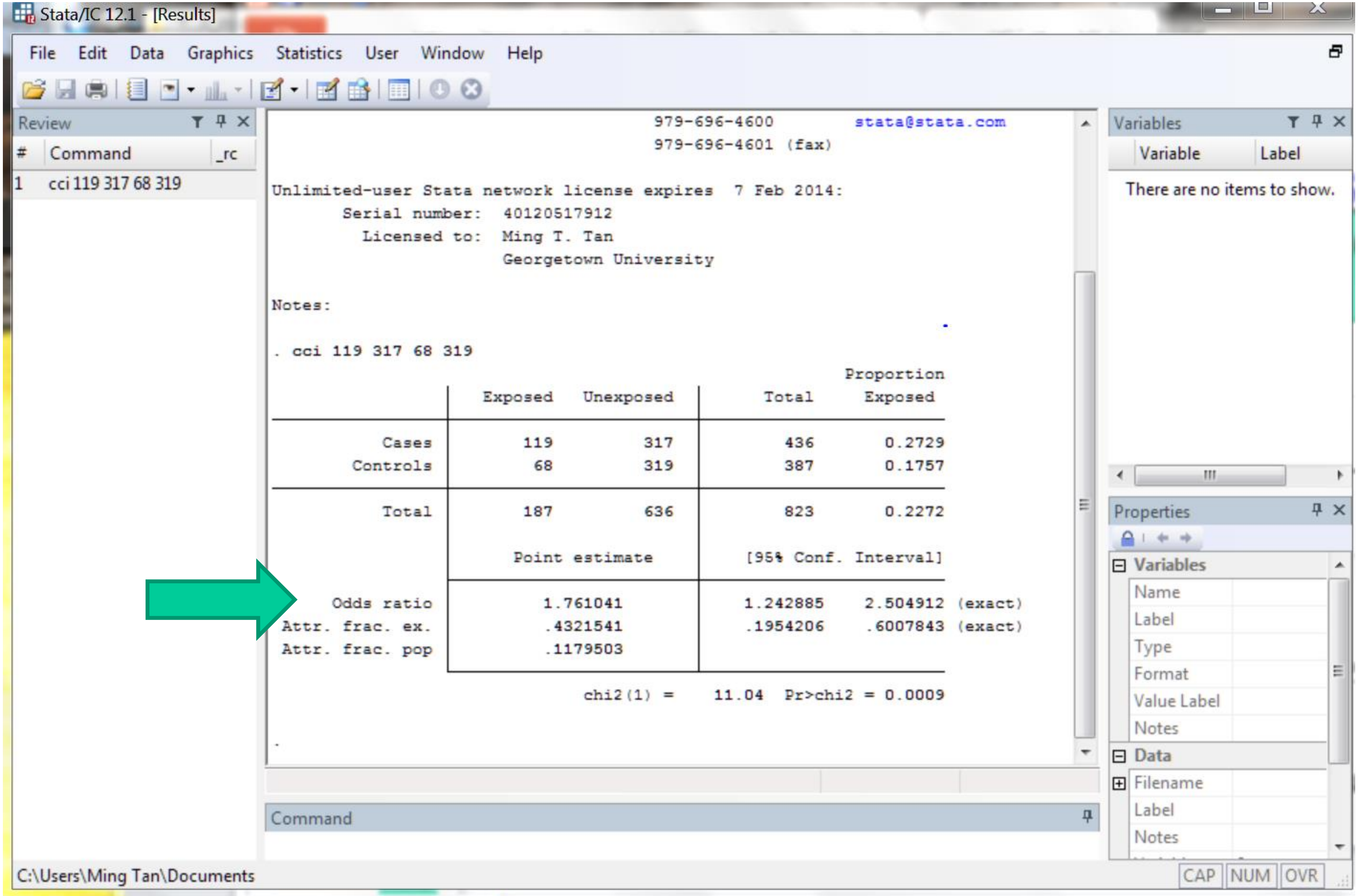
  95% CI : (1.24 ; 2.46) statistically significant association

  The null value is not in this interval and we can reject the null hypothesis OR=1

- We are 95% confident that the true OR is between 1.24 and 2.46, or that we would observe an OR as low as 1.24 or as high as 2.46 for this association (note that means no association).

Under Statistics, choose Epidemiological and related → Tables for epidemiologists → Odds ratio calculator




Stata/IC 12.1 - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command _rc
1 cci 119 317 68 319
979-696-4600 stata@stata.com
979-696-4601 (fax)
Unlimited-user Stata network license expires 7 Feb 2014:
Serial number: 40120517912
Licensed to: Ming T. Tan
Georgetown University
Notes:
. cci 119 317 68 319
Proportion
Exposed Unexposed Total Exposed
Cases 119 317 436 0.2729
Controls 68 319 387 0.1757
Total 187 636 823 0.2272
Point estimate [95% Conf. Interval]
Odds ratio 1.761041 1.242885 2.504912 (exact)
Attr. frac. ex. .4321541 .1954206 .6007843 (exact)
Attr. frac. pop .1179503
chi2(1) = 11.04 Pr>chi2 = 0.0009
Command
Variables
Variable Label
There are no items to show.
Properties
Variables
Name
Label
Type
Format
Value Label
Notes
Data
Filename
Label
Notes
C:\Users\Ming Tan\Documents
CAP NUM OVR

OR of Dz Exercise

Study of oral contraceptive use and ovarian cancer risk

Null hypothesis: OR = 1 (same odds of Dz among exposed and unexposed)

	Cases	Controls
Exposed	119	68
Not Exposed	317	319

Exercise: calculate and interpret the OR of Dz in this example

Be careful when setting up the table!

- The 2 x 2 contingency table should be set up so that cell A (*the top, left hand cell*) represents those who have the disease and were exposed.

- This stands for any arrangement of the table with either Dz or Exp on rows or columns.

- Any other arrangements of the 2 x 2 table will give incorrect results.

Odds Ratio of Disease

		Disease status	
		Cases	Controls
Exposure status	Exposed	A	B
	Unexposed	C	D

Odds ratio = AD/BC

odds of disease among the exposed (A/B) **divided by** odds of disease among the unexposed (C/D)

Interpreting the Odds Ratio for Dz

OR = 1 The estimated disease (dz) risk is the same for exposed and nonexposed

OR > 1 The estimated disease risk is greater for exposed than nonexposed

OR < 1 Exposure protects against disease occurrence (or exp. reduces dz risk)

Odds Ratio: risk in cases relative to controls

O.5	1.0	2.0
Exposure protects by 50%	No association	Exposure increases dz risk by two fold

Hypothetical OR and 95% CI examples: how to interpret

Odds ratio	***Lower CI***	***Upper CI***
6.6	3.1	14.7
1.9	0.8	2.1
1.6	1.1	3.3
0.5	0.2	0.8

For each row, discuss how to interpret the size of the odds ratio, and whether we should accept or reject the null hypothesis.

What factors influence the size of the odds ratio and confidence interval?

- Sample size: smaller studies have wider CI. Why is this true?
- Measurement error: if the exposure is measured imprecisely, the CI will be wide, and the OR will be closer to 1.0. Why?
- Case misclassification: if some cases are not truly cases, the results will be biased. Why is this so? Is this also true if some controls are incorrectly assigned as controls?

Advantages and Disadvantages of the Case-Control Method

Advantages

- Fast (1-5 years)
- Easy to find cases
- Can generate hypotheses
- Often the only practical design given the rare disease issue

Disadvantages

- Infering exposures
- Hard to choose controls
- Temporality problem*
- Interviewer bias*
- Recall bias*
- Selection bias*

* covered later in the course

Advantages of Case-Control Studies

- Most efficient design for rare diseases
  - Cohort design requires very large populations at risk to obtain sufficient numbers of new cases of a rare outcome
- Considerably smaller study population than cohort study of same outcome
  - Allows more intensive evaluation of exposures of cases and controls

Disadvantages of Case-Control Studies

- Does not yield an estimate of incidence rate or cumulative incidence, since size of the at-risk population not known
- Subject to recall bias if exposure is measured by interviews and if recall of exposure differs between cases and controls; avoided if historical records of exposure are available

Disadvantages of Case-Control Studies

- Selection of population controls from the appropriate source population can be difficult to achieve; thus, higher probability of <u>selection bias</u>
- Not efficient for studying rare exposures (less than 10% of controls are exposed) because very large numbers of cases and controls are needed to detect effects of rare exposures

Assignment

-Study Chapter 7

-Answer the Review Questions for Chapter 7 (do not submit)

-Read the Article Posted

Submit Homework 7 by 10/21/19