Case-Control Studies

Zuber D. Mulla

Objectives

1. Describe the essential features of a case-control study.

2. Describe the main advantages and disadvantages of case-control studies.

3. Calculate and interpret the odds ratio.

A major type of observational study

- **Subjects are selected by the presence or absence of the disease or outcome:**
  - **Cases have the disease.**
  - **Controls do not have the disease.**

- **<u>Cases</u> and <u>controls</u> are compared to one another with respect to their exposure(s).**

Hennekens & Buring,
Epidemiology in Medicine

History of Case-Control Studies

- **Were called "retrospective studies"**
- **Sociologists: 1920s and 1930s**
- ***Ex-post facto* effect-to-cause experiments**

Lilienfeld AM and Lilienfeld DE

Criticism

- **Looking backwards from effect to possible cause**

Analyzing Data from a Case-Control Study

	Ill	Well
Exposed	A	B
Not exposed	C	D
	A + C	B + D

Prevalence of exposure in cases= A / (A + C)

Prevalence of exposure in controls= B / (B + D)

Exposure Odds Ratio

- **Odds of exposure in the cases divided by the odds of exposure in the controls**

What is an “odds?”

$$\text{Odds} = \frac{P}{1 - P}$$

- **Ratio of the Probability of an Event (Disease) to the Probability of a Nonevent (No disease)**

$$OR = \frac{\frac{\frac{A}{A+C}}{1-\frac{A}{A+C}}}{\frac{\frac{B}{B+D}}{1-\frac{B}{B+D}}}$$

Probability of exposure among cases

$$\frac{\frac{A}{A+C}}{1-\frac{A}{A+C}}$$

Odds of exposure among cases

$$OR = \frac{\frac{A}{A+C}}{1-\frac{A}{A+C}} \Big/ \frac{\frac{B}{B+D}}{1-\frac{B}{B+D}}$$

$$OR = \frac{\frac{A}{C}}{\frac{B}{D}}$$

Odds Ratio (OR)

	Ill	Well
Exposed	A	B
Not exposed	C	D

OR = AD / BC

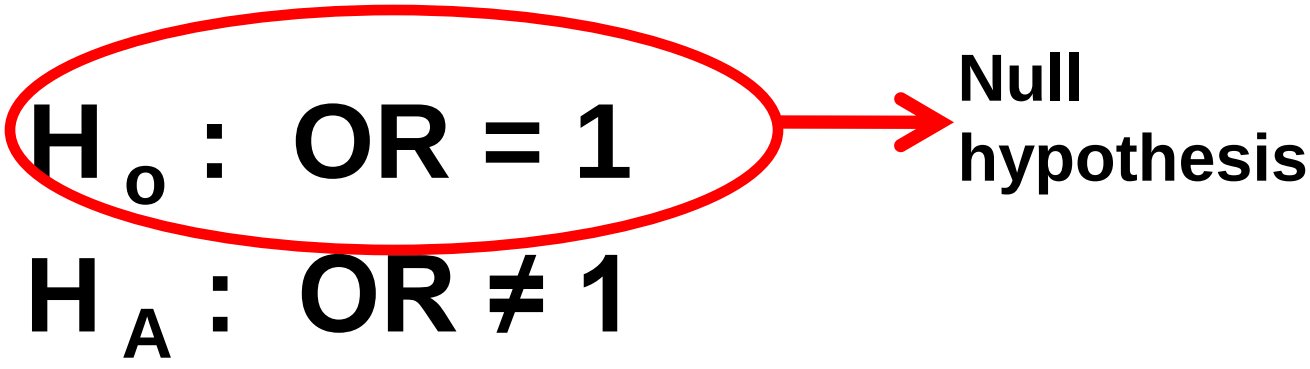
Hypothesis Testing

$$H_o: \ OR = 1$$

$$H_A: \ OR \neq 1$$

$\alpha = 0.05$, Two-tailed test

Hypothesis Testing



Null hypothesis: Usually the hypothesis of "no difference"

Statistical Significance: *P*-Values and Confidence Intervals

For the next few slides we will assume our alpha was set at 0.05. This is done before you conduct your data analysis.

P-Values

If P-value is 0.05 or less (that is, equal to or less than your alpha), then you reject the null hypothesis and conclude the result is statistically significant.

Statistical Significance

If the 95% CI excludes one (e.g., 1.05 – 3.34), then the result is statistically significant and the *P*-value will be 0.05 or less

If the 99% CI excludes one (e.g., 1.05 – 3.34), then the *P*-value will be 0.01 or less

Statistical Significance: *P*-Values and Confidence Intervals

If P-value is greater than 0.05, then you fail to reject the null hypothesis and conclude the result is not statistically significant

For relative risk, prevalence ratio, and odds ratio: If the 95% CI includes one (e.g., 0.80 – 3.34), then the result is not statistically significant.

Controls

- **Choice of the most appropriate control group is one of the most difficult aspects of study design**

Control Selection

- **Controls should be selected from the same population, the source population or study base, that gave rise to the cases.**

  – **Rothman and Greenland, *Modern Epidemiology* text (p. 97, 2nd edition)**

Control Selection

- **"…the controls must be selected to represent not the entire nondiseased population but the population of individuals who would have been identified and included as cases had they also developed the disease."**
  - **Hennekens & Buring, *Epidemiology* in *Medicine* text**

Example

- **Cases of acute lymphocytic leukemia from El Paso County**

- **Controls: _____**

Example

- **Cases of acute lymphocytic leukemia from El Paso County**

- **Controls: <u>El Paso County</u>**

Types of Controls

- **Hospital**
- **Probability sample (RDD)**
- **Friend**
- **Neighborhood**

Explanations when you Observe or Don't Observe an Association

- ***Truth***

- **Chance**
- **Bias**
- **Confounding**

Hennekens & Buring,
Epidemiology in Medicine

Bias

- **An alternative explanation for an observed association is the possibility that some aspect of the design or conduct of a study has introduced a bias into the results**

- ***Systematic error***

Bias

- **Recall bias**
- **Selection bias**

Explanations when you Observe or Don't Observe an Association

- *Truth*

- **Chance**
- **Bias**
- **Confounding**

Methods to Control for Confounding

- **Randomization**
- **Matching**
- **Restriction**
- **Multivariate modeling**
- **Stratified analysis**

Matching

- **Unlike randomization and restriction, which are used to control confounding in the design stage of a study, matching is a strategy that must include elements of both design and analysis.**

Definition of Matching in a Case-Control Study

- **The selection of a series of controls that is identical, or nearly so, to the cases with respect to the distribution of one or more potential confounders (adapted from *Modern Epi.*, 2nd edition by Rothman and Greenland)**

Matching

- **When matching in the design is combined with the appropriate analysis, then control of confounding by the matching factors is achieved.**

To summarize…

- ***A matched design requires a matched analysis.***

Warning

- ***Failure to control a matching factor can lead to biased measures of association.***

Types of Matching

- **Frequency**
- **Individual**

Frequency Matching

- **The control group is selected so that it will have the same frequency distribution of the matched variables as the cases.**

Frequency Matching

- **If 60% of the cases are white and 40% are black, then controls would be selected so that 60% are white and 40% are black.**

Individual Matching

- **Each case is matched with one or more controls (matched on one or more potential confounders).**

- **When 1 control is matched to 1 case, the technique is called “pair matching.”**

Individual Matching

- **An 8-year-old white female case would be matched with an 8-year-old white female control**

- **How many factors were matched on?**

Multiple Controls per Case

- **When the number of cases is limited or fixed, you can increase the ability of the study to detect an association if one truly exists by increasing the number of controls per case (R:1 matching)**
- **4:1**

Multiple Controls per Case

- **Let's say you wanted a 2:1 control: case ratio…**

- **Then, an 8-year-old white female case would be matched with two 8-year-old white female controls**

Analysis of Matched Data

- **Frequency**

- **Individual matched data**

Analysis of Individually Matched Data

- **If you have 1 control per case (that is, pair matched data), then organize your data in a special 2 x 2 table…**

Analysis of Pair Matched Data

Exposure of Controls

Exposure of Cases

	Exposed	Not exposed
Exposed	**A**	**B**
Not exposed	**C**	**D**

Matched odds ratio = B / C

Pair Matched Data

- **Note that the cells in which both the case and control had the same exposure status (both exposed, cell A; and both unexposed, cell D) don’t contribute to the matched odds ratio**

Pair Matched Data

- **We only use the discordant pairs**

- **Only the cells in which the exposure status was different for the case and control contribute information**

Pair Matched Data

- **What about a *P*-value or CI?**
- **Use McNemar's test (chi-square test for matched-pair data)**

More on Pair Matched Data

- **The possible outcomes can be represented by four 2 x 2 tables**

- **Each 2 x 2 table represents a pair (just two subjects)**

The 4 Possible Outcomes with Pair Matched Data

	Case	Control
Exposed	1	1
Not exposed	0	0
	1	*1*

	Case	Control
Exposed	1	0
Not exposed	0	1
	1	*1*

	Case	Control
Exposed	0	1
Not exposed	1	0
	1	*1*

	Case	Control
Exposed	0	0
Not exposed	1	1
	1	*1*

Interesting Fact

- **The OR for matched pair data (B / C) is also the Mantel-Haenszel OR applied to matched pair data**

Analysis of Individually Matched Data

- **If you have 2 or more controls per case, refer to the text by Schlesselman for the formulas for matched odds ratios and CIs**

- **You can also use conditional logistic regression (but not unconditional)**

Advantages of Matching

- **In some situations matching is a useful method for improving study efficiency in terms of the amount of information per subject studied**

- **Sparse-data problem**

Disadvantages of Matching

- **"Although usually intended to control confounding, matching does not attain that objective in case-controls studies. Instead, it superimposes over the confounding a selection bias that behaves like confounding…the bias can be removed by treating the matching variable as a confounder."**

See *Modern Epidemiology, 2nd edition*, Rothman & Greenland

Disadvantages of Matching

- **Time consuming**

- **Inability to evaluate the effect of the matching variable on the outcome since the distribution of that variable has been forced to be identical**

True or False?

- **The phrase “retrospective study” is synonymous with “case-control study?”**

Use of the Terms “Retrospective” and “Prospective”

- Retrospective case-control study
- Prospective case-control study

- Retrospective cohort study
- Prospective cohort study

- **Most case-control studies are retrospective. That is, cases have already occurred.**

- **Whether or not retrospective or prospective, you start with the outcome/disease and look back in time**

Key to the next 3 slides (adapted from Hennekens & Buring)

= Present

= Absent

} ***Basis on which groups are selected at start of study***

? = To be determined

= The investigator at the beginning of the study

Case-Control Study

Exposure Disease

?

?

Prospective Cohort Study

Exposure

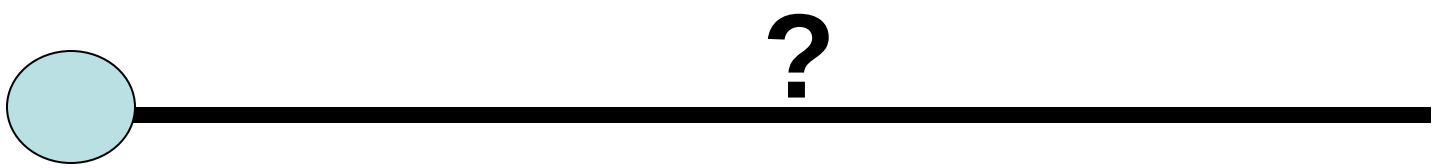
Disease

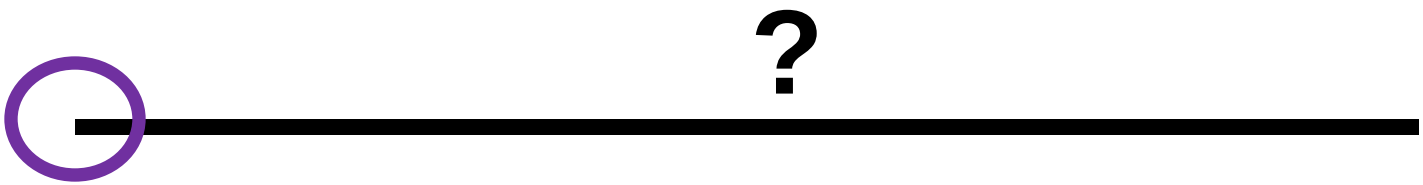
?

?

Retrospective Cohort Study

Exposure Disease





Retrospective Case-Control Study

- **In September of 2014 you identify cases of epithelial ovarian cancer diagnosed in 2013**

- **Look back in time: Up to 1 y before ca dx, did you _____ ?**

Prospective Case-Control Study

- **In September of 2014, you decide to conduct a case-control study**
- **No cases identified yet**
- **From October 2014 - June 2015**
- **Look back in time: Up to 1 y before ca dx, did you _____ ?**

Excellent references

- ***Case-Control Studies: Design, Conduct, Analysis***

By James J. Schlesselman

Oxford Univ. Press, 1982

Excellent references

- ***Statistical Methods in Cancer Research, Volume I—The Analysis of Case-Control Studies***

By N.E. Breslow & N.E. Day

IARC, 1980

Advantages of the Case-Control Study Design

- **Can study rare diseases**
- **Multiple exposures**
- **Usually cheaper than prospective cohort studies**

Disadvantages of the Case-Control Study Design

- No incidence (usually)
- More prone to selection and recall bias than other designs
- Temporality may not be intact

Cited Reference

Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.

Lilienfeld AM, Lilienfeld DE. A century of case-control studies: progress? J Chron Dis 1979; 32: 5-13.

Rothman KJ, Greenland S. Modern epidemiology second edition. Philadelphia: Lippincott Williams & Wilkins, 1998.