MA50259: Statistical Design of Investigations

Dr. Sandipan Roy

Lecture 8: Case-Control Studies

Controlled Cohort studies

- **Controlled Cohort study:** Two (or more) groups of individuals are followed over time and occurrences of the disease are counted within each group.

Exposure	Disease	Not Disease	Total
Exposed	a	b	$n_1 = a + b$
Non-Exposed	c	d	$n_0 = c + d$

- For each individual, the **exposure category is fixed** and the **disease outcome is as random**

Case-control studies

- A sample of **cases (individuals who have the disease)** is selected. Then a second group of individuals who are not cases (individuals who do not have the disease) is selected. These are the **controls**
- In both groups, the disease **outcome is known** but the **exposure category is treated as random.**
- After cases and the controls have been selected, their previous exposures are ascertained

Exposure	D (cases)	Not D (controls)
Exposed	a	b
Non-Exposed	c	d
Total	$m_1 = a + c$	$m_0 = b + d$

Case-control studies

- Controlled cohort study: n_0 and n_1 fixed in advance
- Case-control study: m_0 and m_1 fixed in advance
- In a cohort study, the disease outcome is determined by following the exposed and not-exposed groups forward in time
- In a Case-control study the exposure categories are determined by looking back in time at the histories of cases and controls
- Proportions of cases and controls with the exposure E are compared
- If there is no association between E and D, the the proportions exposed in cases and controls should be similar.

Ex 1. Doll & Hill Case-control study

	Lung cancer	Control	Total
Smoked	647	622	1269
Never smoked	2	27	29
Total	649	649	1298

$$\widehat{RR} = \frac{647/1269}{2/29} = 7.39\,, \qquad \widehat{OR} = \frac{647(27)}{622(2)} = 14.04$$

Odds ratio for smoking is 14.04 with 95% CI $(3.33, 59.31)$

Ex 1. Doll & Hill Case-control study

- The 649 cases were male patients admitted to hospital with lung cancer
- For each case a control was selected of same sex and age admitted to hospital for other disease
- If more than one eligible control, the first on the list was chosen

Ex 1. Doll & Hill study (Hypothetical 1)

	Lung cancer	Controls×10	Total
Smoked	647	6220	6867
Never smoked	2	270	272
Total	649	6490	7139

$$\widehat{RR} = \frac{647/6867}{2/272} = 12.81\,, \qquad \widehat{OR} = \frac{647(270)}{6220(2)} = 14.04$$

Ex 1. Doll & Hill study (Hypothetical 2)

	Lung cancer×10	Control	Total
Smoked	6470	622	7092
Never smoked	20	27	47
Total	6490	649	7139

$$\widehat{RR} = \frac{6470/7092}{20/47} = 2.14\,, \qquad \widehat{OR} = \frac{6470(27)}{622(20)} = 14.04$$

Ex 3. Doll & Hill study (extras)

- The 649 cases were male patients addmitted to hospital with llung cancer
- For each case a control ws selected of same sex and age admitted to hospital for other disease
- If more than one eligible control, the first on the list was chosen
- Odds ratio for smoking is 14.04 with 95% CI $(3.33, 59.31)$
- How different are the cases and controls?
- Does any difference matter for the association?
- Which alternative explanations are for the association

Case-control studies: estimation

- The relative risk (risk ratio) cannot be estimated in a case-control study as it depends on the number of cases and controls selected
- The odds ratio does not depend on the number of cases and controls selected
- The odds ratio only depends on the proportion of cases and controls with the exposure

Case-control studies: estimation

Exposure	D (cases)	Not D (controls)
Exposed	a	b
Non-Exposed	c	d
Total	$m_1 = a + c$	$m_0 = b + d$

- $$\widehat{OR} = \frac{\widehat{odds}\,(E|D)}{\widehat{odds}\,(E|D^c)} = \frac{a/c}{b/d} = \frac{a/b}{c/d} = \frac{\widehat{odds}\,(D|E)}{\widehat{odds}\,(D|E^c)} = \frac{ad}{cb}$$

- Confidence intervals for the odds ratio are calculated in the same way as in a cohort study (see Lecture 3)

Case-control studies

- Main issue with case-control studies: **How to select the controls?**

  + Selected from the population that gave rise to the cases

  + Controls should have the same opportunity as the cases to become exposed

Example 2: Physical exertion and heart attacks

Planning of a case-control study of the impact of physical exertion on triggering heart attacks

- How to select the cases and the controls?
- Need to take into account potential confounder variables? If so, which ones?
- Is stratification of confounders a problem?
- How to account for confounding from the outset?

Example 3: Autism and MMR

- In 1998 a hypothesis was formulated that measles, mumps and rubella (MMR) vaccine may be associated with the development of autism
- Controversy led to a drop of MMR vaccination rates in some parts of the UK
- Many studies, none of which could confirm the original hypothesis
- One study: 1294 cases of pervasive developmental disorder
- 4469 controls selected from GP records
- Up to five controls were matched to each case (same sex, birth date within one year)
- Could this procedure be considered as stratification? If so how many strata?
- Can we use Mantel-Haenszel to obtain a summary effect?

Example: Cirrhosis and heavy drinking

- Case-control study to explore the association between Cirrhosis and heavy drinking
- What is the time sequence in case-control study?
- How good a case-control study would be to establish causality here?
- Other problems?

Example: Sudden infact death syndrome

- Cases are deaths from SIDS
- Mothers are asked to report medical events
- Any problems here?

Case-control studies

- Advantages

```
+ Quicker and cheaper  than cohort studies
+ Typically require much smaller sample sizes
+ Particularly suited to study rare diseases
+ Many risk factors can be studies simultaneously
```

- Disadvantages

```
+ No time sequence: forbids causality
+ Only one disease outcome
+ No estimates of the relative risk
+ Likely to suffer from bias error from the way
the controls are sampled
```