

Lecture 3

Hephaes Chuen Chau

2022-03-03

Discussion on confounder

A confounder C must:

1. Be causally related to the outcome of interest
2. Be statistically associated with the putative risk factor

Matching

- Stratification must be kept at an appropriate level; otherwise too much stratification could cause a loss of variability (in terms of exposure) within stratum
- Regression is often used
- Alternatively (and ideally), increase the sample size and apply regression; matching is not preferred as there could be undesirable effects (overmatching)

Matching in cohort study

- not preferred

Matching in case-control study

- nested case-control study frequently encountered
- circumvent the problem caused by the survival bias involved in the selection of the control
- types of matching:
 - individual
 - frequency (cf <https://www.goldsteinepi.com/blog/matchinginacasecontrolstudyinpractice/index.html>)
 - statistical analysis in individual matching (each case matched with a number of control) :: usually conditional logistic regression is adopted
 - statistical analysis in frequency matching :: usually unconditional logistic regression is adopted
- matching with a distance-ball:
 - for example Euclidean distance-based matching
 - however, the more frequently used type of matching when the number of variables to match is very high is propensity score

Propensity score

- A RCT example with $do(X) \in \{0, 1\}$ representing treatment or not
- too much confounder controlled in logistic regression: power of X , the treatment of interest is reduced
- an alternative approach is adopted such that
 - A logistic regression of X with respect to the confounder is produced
 - The model is used to predict a score of being assigned to a particular X , that is propensity score
 - Subjects with $do(X) = 1$ are matched with subjects in $do(X) = 0$ who have similar propensity score
 - * Downstream analysis is then done with conditional logistic regression

Advantages and disadvantages of matching

- Chapter 2 of Epidemiology, Beyond the Basics
- See p.58 and p.59. of Lecture 1 slides
- residual confounding: insufficient control for the confounders. This problem is especially serious when binning is applied to a confounder (eg, age) for matching.
 - In particular for the variable age, one should be alerted to the possibility of confounding for aging
 - Effect estimates of the variables with which the cases and the control are matched bear no significant meanings; however, the effect estimates of the variables not matched may be important

Measuring the occurrence of a disease

- outcome variables can be categorised into discrete vs continuous variables

Discrete variables

- incidence, prevalence and odds are the three possible types of measurements
- incidence is a more complicated concept than prevalence

Prevalence

- Prevalence measurements can be categorised according to the period of time in which cases are defined
- When time's unit appears in the denominator, the measurement is a rate. Thus, prevalence rate is more properly called prevalence proportions
- We therefore have

$$\text{Prevalence} = \frac{\text{Cases}}{\text{Those who are at risk of being cases} + \text{Cases}}$$

- Note that the time during which the cases and non-cases is observed can be defined to be a point of time or a period of time

Incidence

- the difficulty of arriving at the incidence is the difficulty with which the occurrence of the event is accurately defined
- categorised by the units of the denominator: (a) persons at risk, or (b) person-time units at risk

Hazard rate

-