Grau d'Estadística

Estadística Mèdica

Francesc Miras

francesc.miras@upc.edu

Universitat de Barcelona – Universitat Politècnica de Catalunya

21st November, 2018

Today's programme

Sources of bias in epidemiological studies

1. Types of bias in epidemiological studies

2. Confounding

3. Selection bias
   - Definition of selection bias
   - Different types of selection bias
   - Final comments

4. Information bias

Types of bias

Definition

Contrary to random errors, which are due to sampling, **bias** is

> *"Any systematic error in design, conduct or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease." (Schlesselman 1982)*

In epidemiology, we can basically distinguish three types of bias:

- **Confounding (bias)** due to non-adjustment of variables associated with both exposure and outcome of interest,
- **Selection bias** due to non-random selection of study participants,
- **Information bias** due to erroneous information.

The problem of confounding

An example for confounding

Consider the following example (from Cobo *et al.* 2003), where the data on the exposure E and disease D come from two different medical centers:

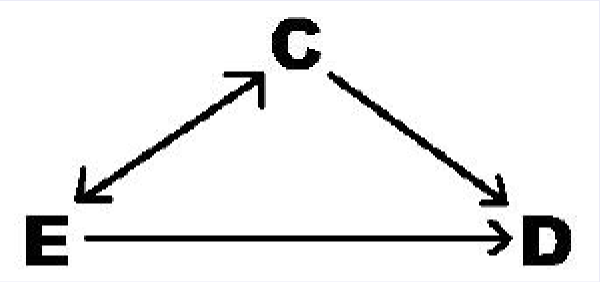
	Center 1			Center 2		
	D	$\bar{D}$	Total	D	$\bar{D}$	Total
E	100	50	150	10	20	30
$\bar{E}$	20	10	30	50	100	150

✍ Which is the odds ratio in each center and which is the overall odds ratio? What do you observe and which are the reasons for that?

Definition of a confounding variable

Definition

C is called a **confounder** (of the association between E and D), if it is both related with E and a prognostic factor of D. As a result, C causes a biased estimation of the association of interest:



Notes:

- A confounder must not be an effect of the exposure (Rothman (2002)).
- A common cause of E and D is a potential confounder.

Direction of the confounding bias

Direction of confounding

In the case of a binary confounder C, the effect measure will be overestimated if either

$$\begin{array}{lccc} & P(D|C=1) > P(D|C=0) & \textbf{AND} & P(E|C=1) > P(E|C=0), \\ \text{or} & P(D|C=1) < P(D|C=0) & \textbf{AND} & P(E|C=1) < P(E|C=0), \end{array}$$

and underestimated otherwise (Hernán & Robins: *Causal Inference.*).

✍ Following, which is the direction of the bias of the non-adjusted OR?

	Center 1			**Center 2**		
	D	$\bar{D}$	**Total**	D	$\bar{D}$	**Total**
E	20	10	30	50	100	150
$\bar{E}$	100	50	150	10	20	30

Strategies to avoid confounding

Can we avoid confounding?

Confounding by a risk factor C can be avoided by

(a) Randomization: exposure is assigned randomly.

(b) Restriction: all study subjects have the same value of C.

(c) A balanced study design, that is, an equal distribution of C among exposed and not exposed subjects.

✍ Using the same probabilities $P(D|C_1)$ and $P(D|C_2)$ as before but with equal distributions of E in each center, which is the non-adjusted OR now?

	Center 1			Center 2		
	D	$\bar{D}$	Total	D	$\bar{D}$	Total
E	80	40	120	26	52	78
$\bar{E}$	40	20	60	13	26	39

Selection bias

Definition

According to Rothman (2002),

> *"Selection bias is a systematic error in a study that stems from the procedures used to select subjects and from factors that influence study participation."*

Its consequence is (Hernán *et al.* 2004)

> *"...that the association between exposure and outcome among those selected for analysis differs from the association among those eligible."*

Selection bias (cont.)

A more formal approach to selection bias

Assume, we are interested in estimating the odds ratio and that the true (unknown) numbers of subjects in the population, according to exposure and disease, are the following:

	Disease	
Exposure	Yes	No
Yes	A	B
No	C	D

Hence, the odds ratio in the population is

$$\mathrm{OR} = \frac{A \cdot D}{B \cdot C}.$$

Selection bias (cont.)

In our study sample, we observe the following numbers of subjects:

	Disease	
Exposure	Yes	No
Yes	$a = w_A \cdot A$	$b = w_B \cdot B$
No	$c = w_C \cdot C$	$d = w_D \cdot D$

Therein, $w_A, \ldots, w_D$ denote the (unknown) selection proportions of subjects in the sample. Hence, we can express the ML estimator of the odds ratio in the following way (Kreienbrock & Schach 1995):

$$\widehat{\mathrm{OR}}_{\mathrm{ML}} = \frac{a \cdot d}{b \cdot c} = \frac{w_A \cdot A \cdot w_D \cdot D}{w_B \cdot B \cdot w_C \cdot C} = \frac{w_A \cdot w_D}{w_B \cdot w_C} \cdot \mathrm{OR}.$$

Selection bias (cont.)

Therefore, the estimation of the odds ratio is free of selection bias if

$$W = \frac{w_A \cdot w_D}{w_B \cdot w_C} = 1.$$

Comments:

- $W = 1$ does not necessarily imply equal selection proportions.
- A cohort study is free of selection bias if

$$w_A = w_B = k \cdot w_C = k \cdot w_D.$$

- A case-control study is free of selection bias if

$$w_A = w_C = k \cdot w_B = k \cdot w_D.$$

Different types of selection bias

Response bias

Response bias is caused by a systematic non-participation of previously selected study participants, that is, when the reason for not participating is associated with either exposure or outcome.

For example,

- In a cross-sectional study on the association between pollution and coronary heart diseases, non-exposed healthy people might be less interested in participating. As a consequence, w_D would be systematically too low causing an underestimation of the odds ratio.
- During a cohort study on a highly toxic substance, some exposed participants might die because of this exposure but without developing the disease of interest. As a consequence, w_A would be systematically too low leading to an underestimation of the odds ratio.

Different types of selection bias (cont.)

Migration bias

Migration bias can occur in a cohort study, if previously exposed subjects change exposure status because of illness (or the other way round) and data are analyzed using the present classification.

Remember the example from the Framingham Heart Study:

Prevalence and incidence data on CHD in males

	Incidence (10 years)			**Prevalence (after 10 y.)**		
Cholesterol	CHD	No CHD	Total	CHD	No CHD	Total
High	85	462	547	38	371	409
Low	28	516	544	33	347	380

From: Jewell (2004); Source: Friedman et al. (1966).

✍ Which selection proportions might be affected here and which is the consequence?

Different types of selection bias (cont.)

Membership bias: the Healthy Worker Effect

The **Healthy Worker Effect** can occur in studies of occupational diseases since workers usually exhibit lower overall disease rates than the general population because severely ill and disabled people are excluded from employment.

A hypothetical example from Rothman (2002):

		General population		
	Exposed Workers	**Workers**	**Non-workers**	**Total**
Deaths	50	4500	2500	7000
Person-time	1000	90000	10000	100000
Mortality rate	0.05	0.05	0.25	0.07

Different types of selection bias (cont.)

Prevalence-Incidence Bias

The **prevalence-incidence bias** can be present, if the exposure of interest is related to disease duration.

For example, exposure E may not cause the disease D but prolong disease duration. As a consequence, in a cross-sectional study on the effects of E on D, nonexposed cases have lower probability to be observed than exposed cases. That is, w_C would be systematically too low causing an overestimation of the odds ratio.

To prevent this kind of bias, cases with disease onset long ago might not be included in this study.

Different types of selection bias (cont.)

Further types of selection bias

- **Detection bias**

  Can occur in case-control studies (Vineis 2002),

  > *". . . in which the risk factor investigated leads to special diagnostic procedures and thus increases the probability that the disease is identified (in contrast to unexposed subjects)."*

- **Admission-rate bias (Berkson's bias)**

  In case-control studies carried out in hospitals, controls might be not representative with respect to the exposure of interest.

- **Self-selection bias**

  If participants are chosen on a volunteer basis for a study, individuals worried about the exposure or the disease of interest might be more prone to take part in the study.

Selection bias (cont.)

Comments

- Selection bias could be prevented by random selection. However, in epidemiologic studies, random selection may not be feasible.
- Even an appropriate design and confounding adjustment cannot exclude the possibility of selection bias.
- Selection bias is mainly a problem of case-control studies, since it may be difficult to obtain a random sample of controls.
- According to Hernán *et al.* (2004), all types of selection bias have in common that exposure and disease have a common effect.
  They propose the use of inverse probability weighting to overcome the problems of selection bias.
- As statisticians, if we cannot influence the study design, we should be aware of a possible selection bias when analyzing the data or presenting the results.

Information bias

Definition

Information bias or misclassification bias is defined (Rivas-Ruiz *et al.* 2012)

> *"... as bias introduced by the method used to collect information referred to exposure, the results or other confounding or effect-modifying variables."*

It does not depend on the study type, but on the way, information and data are collected and may arise, among others, from

- reversing the order of exposure and disease in case-control studies,
- misclassification of disease status or wrong quantification of exposure status,
- measurement error, for example, when people get tired of answering during a very long questionnaire.

Information bias (cont.)

Two types of information bias can be distinguished:

1. **Non-differential** information bias
   - Misclassification of disease or exposure affects all study groups equally.
   - Leads to an underestimation of the effect measure of interest.
2. **Differential** information bias
   - Misclassification of disease is not independent of exposure and viceversa.
   - **Recall bias:** Can occur in case-control studies if cases and controls misclassify exposure in the past differently.
   - **Interviewer bias:** Can occur in case-control studies if the interviewer treats cases and controls differently.
   - **Diagnostic suspicion bias:** Can occur in cohort studies if the procedure to diagnose the outcome depends on exposure status.

References

- Cobo E., S. Videla, and B. Clotet (Editores) (2003). *Investigación Clínica y Estadística. Una visión multidisciplinar con aplicaciones en estudios de VIH/Sida*. Fundació Lluita contra la SIDA.
- Friedman, G., W. Kannel, T. Dawber, and P. McNamara (1966). Comparison of prevalence, case history and incidence data in assessing the potency of risk factors in coronary heart disease. *American Journal of Epidemiology 83*, 366–378.
- Hernán, M., S. Hernández-Díaz, and J. Robins (2004). A Structural Approach to Selection Bias. *Epidemiology 15(5)*, 615–625.
- Jewell, N. (2004). *Statistics for Epidemiology*. Chapman & Hall/ CRC.
- Kreienbrock, L. and S. Schach (1995). *Epidemiologische Methoden*. Gustav Fischer Verlag Stuttgart·Jena.

References

- Rivas-Ruiz, F., Pérez-Vicente, S., González-Ramírez, A. (2013). Bias in clinical epidemiological study designs. *Allergologia et immunopathologia 41(1)*, 54–59.
- Rothman, K. (2002). *Epidemiology: an introduction.* Oxford University Press.
- Schlesselman, J. (1982). *Case control studies: Design, conduct, and analysis.* Oxford University Press, New York.
- Vineis, P. (2002). History of bias. *Sozial-und Praventivmedizin 47*, 156–161.