

Confounding, Effect Modification and Bias

Lesley Rushton

Epidemiology and Public Health

l.rushton@imperial.ac.uk

Objectives

- To explain confounding, the effect it has on study results and how to adjust for it
- To define effect modification and how to estimate it
- To outline the biases that can occur in
 - case-control studies
 - cohort studies

Confounding

- Confounding is the error in the measure of association between a risk factor and disease outcome
- Arises when there are differences in the comparison populations other than the risk factor under study
- Confounding is derived from a Latin word meaning to mix up, a useful idea, for confounding mixes up causal and non-causal relationships
- A confounding factor is
 1. associated with the disease and
 2. differently distributed over various exposure groups

Example

- Imagine that a study follows up people who drink alcohol and observes the occurrence of lung cancer
- A group of people who do not drink and are of the same age and sex provide the comparison group
- The study finds that lung cancer is more common in alcohol drinkers, i.e. there is an association between alcohol consumption and lung cancer.
- Did alcohol causes lung cancer?

Is there likely to be confounding?

- In what other important ways might the study (alcohol drinking) and comparison (no alcohol drinking) populations be different?
- Could the association between alcohol and lung cancer be confounded?
- What might be the confounding variable?
- First key analysis in all epidemiological studies is to compare the characteristics of the populations under study

Examples of confounding

The confounded association	One possible explanation	The confounded factor	The confounding (causal) factor	To check the assumption
People who drink alcohol have a raised risk of lung cancer	Alcohol drinking and smoking are behaviours which go together	Alcohol, which is a marker for, on average, smoking more cigarettes	Tobacco, which is associated with both alcohol and with the disease	See if the alcohol-lung cancer relationship holds in people not exposed to tobacco: if yes, tobacco is not a confounder

Figure 4.3

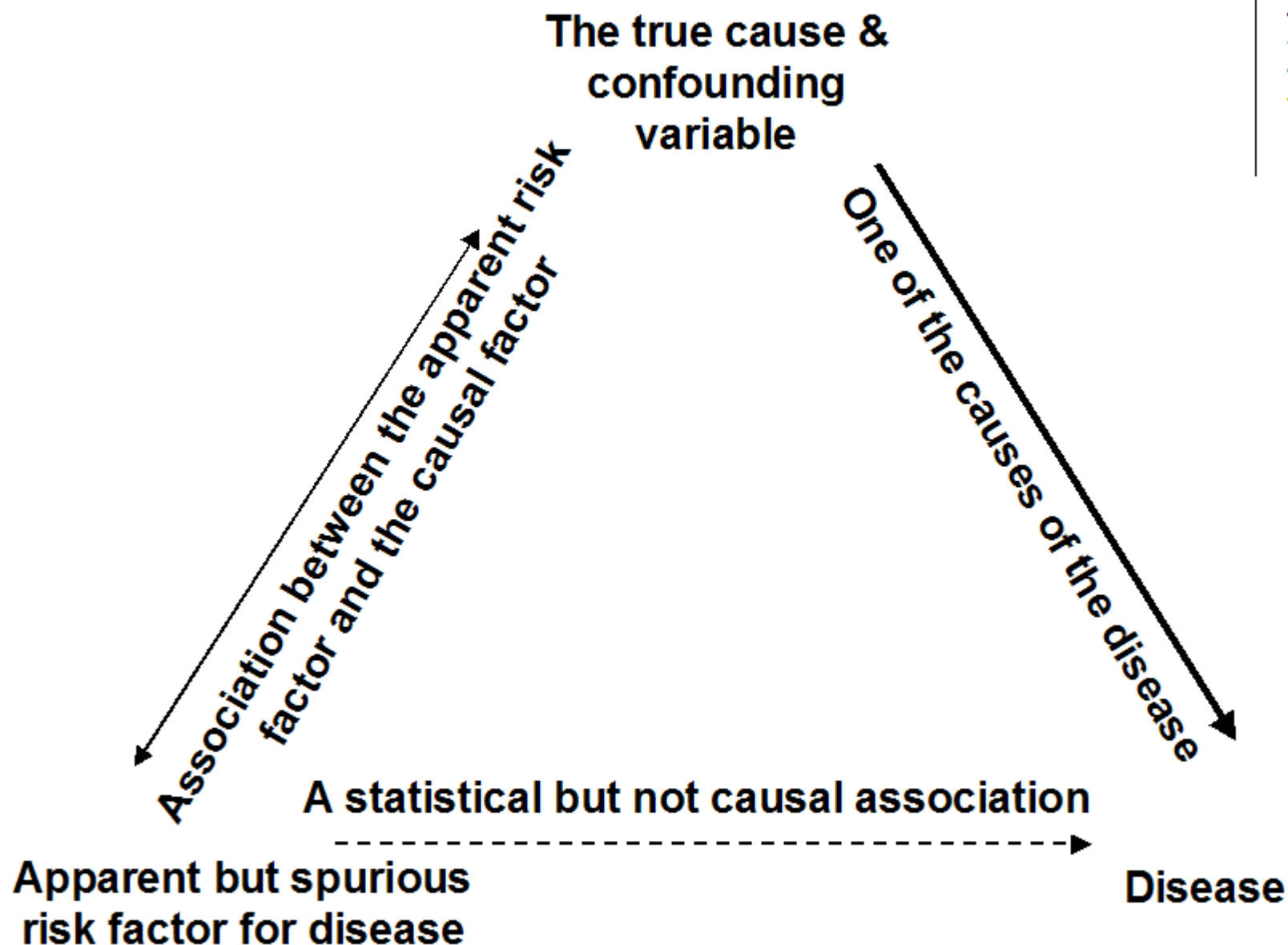
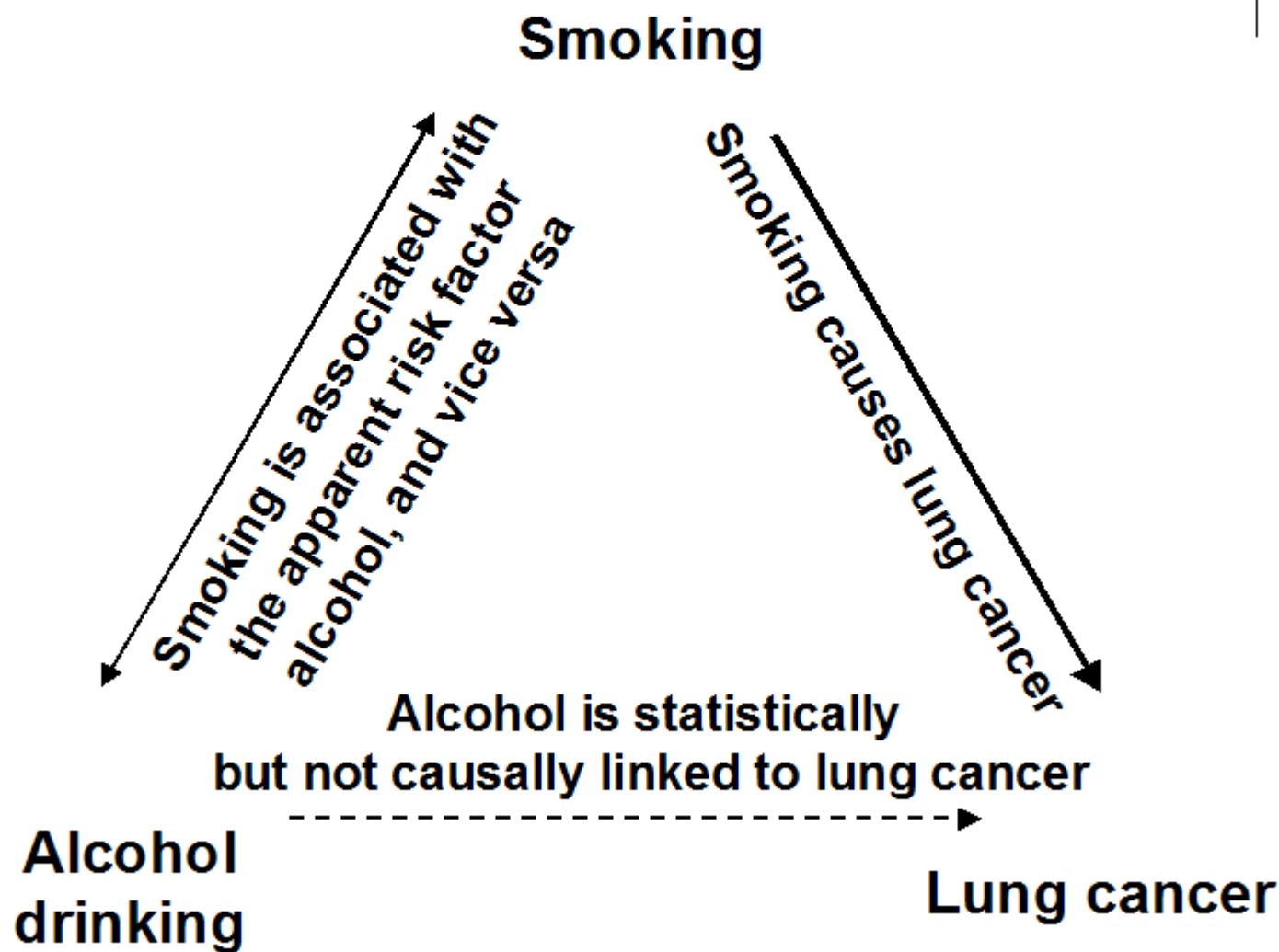


Figure 4.4



Possible actions to control confounding

- Study Design : Randomise individual subjects or units of populations to avoid selection bias
- Study Design : Select comparable groups/ restrict entry into study
- Study Design : Match individuals or whole populations
- Analysis : Analyse subgroups separately
Analysis : Adjust data statistically

Problems with matching

- Need to know about confounders before the study
- Bias arises if matching factor is correlated with exposure of interest, forces controls to be more like the cases than they otherwise would be
- Matching on many variables => virtually identical cases and controls
- May make study more costly and time consuming
- In a cohort study may not be able to identify a closely matched comparison population

Stratified analyses

- Odds ratios are obtained for different values or **strata** of the factor of interest.
- e.g. obtain an estimate of the effect of exposure on a disease separately for males and females.
- Can also obtain a pooled result across the strata, e.g. for the two sexes, where the confounding effect of the factor, gender, has been adjusted for.

Case-control study

Example: Blood Pressure (BP) and Myocardial Infarction (MI)

Age is a confounding factor because BP is different at different ages (higher the older one becomes) and the risk of a MI also increases as one ages.

Prevalence of Myocardial Infarction by Systolic Blood Pressure and Age.

	MI cases	MI negative	Total	
Age < 60				
SBP ≥ 140	9	115	124	
SBP < 140	6	73	79	
Total	15	188	203	
				OR = 0.95
Age ≥ 60				
SBP ≥ 140	20	596	616	
SBP < 140	21	1171	1192	
Total	41	1767	1808	
				OR = 1.87

Source: unpublished data from the Israeli Heart Disease Study

The table above shows a stratified analysis by age. The risk of MI associated with raised BP increases with increasing age.

When the OR is estimated for separate strata the Mantel-Haenszel procedure can be used to combine these to give an overall estimate (see Appendix for calculations)

$$\text{OR (overall)} = 1.57$$

i.e. OR relating higher SBP to MI is 1.57 after removing part of the confounding effect of age. (NB. effect is not eliminated entirely because within age classes ≥ 60 and < 60 the people with higher BP are likely to be older than others)

Cohort Study Example

When calculating the relative risk we make the assumption that the ratio of disease rates between the cohort and the comparison populations is constant across different categories of potential confounding variables.

British Doctors Study

Deaths from coronary heart disease (CHD) among British male doctors*

Age group (yrs)	No.person years (1000's)		No. of CHD deaths		CHD rates†		Rate ratio
	Non-smoker	Smoker	Non-smoker	Smoker	Non-smoker	Smoker	
35-44	18.790	52.407	2	32	0.11	0.61	5.73
45-54	10.673	43.248	12	104	1.12	2.40	2.14
55-64	5.710	28.612	28	206	4.90	7.20	1.47
65-74	2.585	12.663	28	186	10.83	14.69	1.36
75-84	1.462	5.317	31	102	21.20	19.18	0.90
Total	39.220	142.247	101	630	2.58‡	4.43‡	1.72

* From Doll and Hill (1996) as quoted by Breslow and Day (1987)

† Per 1000 person-years

‡ Average rates over entire age range

Effect modification

- An effect modifier is a factor in which the effect of exposure is stronger in some strata than others.
- if association of blood pressure and myocardial infarction is different at different ages then age is an effect modifier
- Say there is an **interaction** between exposure and factor which is an effect modifier, e.g. interaction between age and BP in relation to the risk of myocardial infarction
- In the table the OR is higher at age >60 than age <60
⇒ high BP is a greater risk for MI at age over 60 i.e. age is an **effect modifier**

Note The interaction between age and BP is separate and distinct from the confounding between age and BP. If we do not properly control for confounding we overstate the effect of high BP on mortality risk (because those with high BP are older than those without it).

Error and bias in epidemiology

- Error and bias in epidemiology focus on
 - (a) selection (of population)
 - (b) information (collection, analysis and interpretation of data)
 - (c) confounding
- Different epidemiological study designs share most of the problems of error and bias
- Need to distinguish between random error and systematically occurring error
- Bias due to confounding may be important

Information bias: Measurement errors

- Information bias is caused by inaccurate measurement of variables. Causes include:
- Recall bias and the need to estimate exposure retrospectively
- Inaccurate observation by the investigator
- Imprecise measurement tools
- Biological variation that results in the use of a summary measure
- Some variables have natural variation so great that making estimates is extremely difficult, for example, in diet, alcohol consumption, and the level of stress

Misclassification bias

- Measurement errors which occur unequally in the comparison populations are called **differential misclassification errors or bias**. They can cause the risk estimate to be biased up or down
- **Non-differential errors or biases**, occurring in both comparison populations, may be more likely to occur. These tend to reduce the risk estimate

Selection Bias

- Study group does not reflect general population with regard to age, smoking etc.
- Can be because of:
 - refusal to participate
 - records incomplete
 - effects of volunteering
 - inclusion of those with incipient disease
 - distribution of confounders

Follow-up Bias

- Internal validity not affected if loss to follow-up equal in exposed and non-exposed.
- Loss to follow-up can lead to misclassification if exposure data is being collected concurrently with disease occurrence

Post hoc Bias

- Use of study data to make observations which were not part of original study intent, i.e. interesting relationships not originally anticipated.
- Treat as hypothesis generating, to be studied with **new** data.

Reducing bias

- Selection bias
 - Careful selection
 - Characterize differences between respondents and non-respondents
- Information bias
 - Use well defined, precise measurements with known sensitivity and specificity
- Follow-up bias
 - Intensive follow-up
 - Compare baseline characteristics of those followed up and lost
- Confounding bias
 - Stratified analysis
 - Adjustment in the analyses