Epidemiologie E350

Confounding a standardizace

Three major issues in interpretation of any epidemiological study

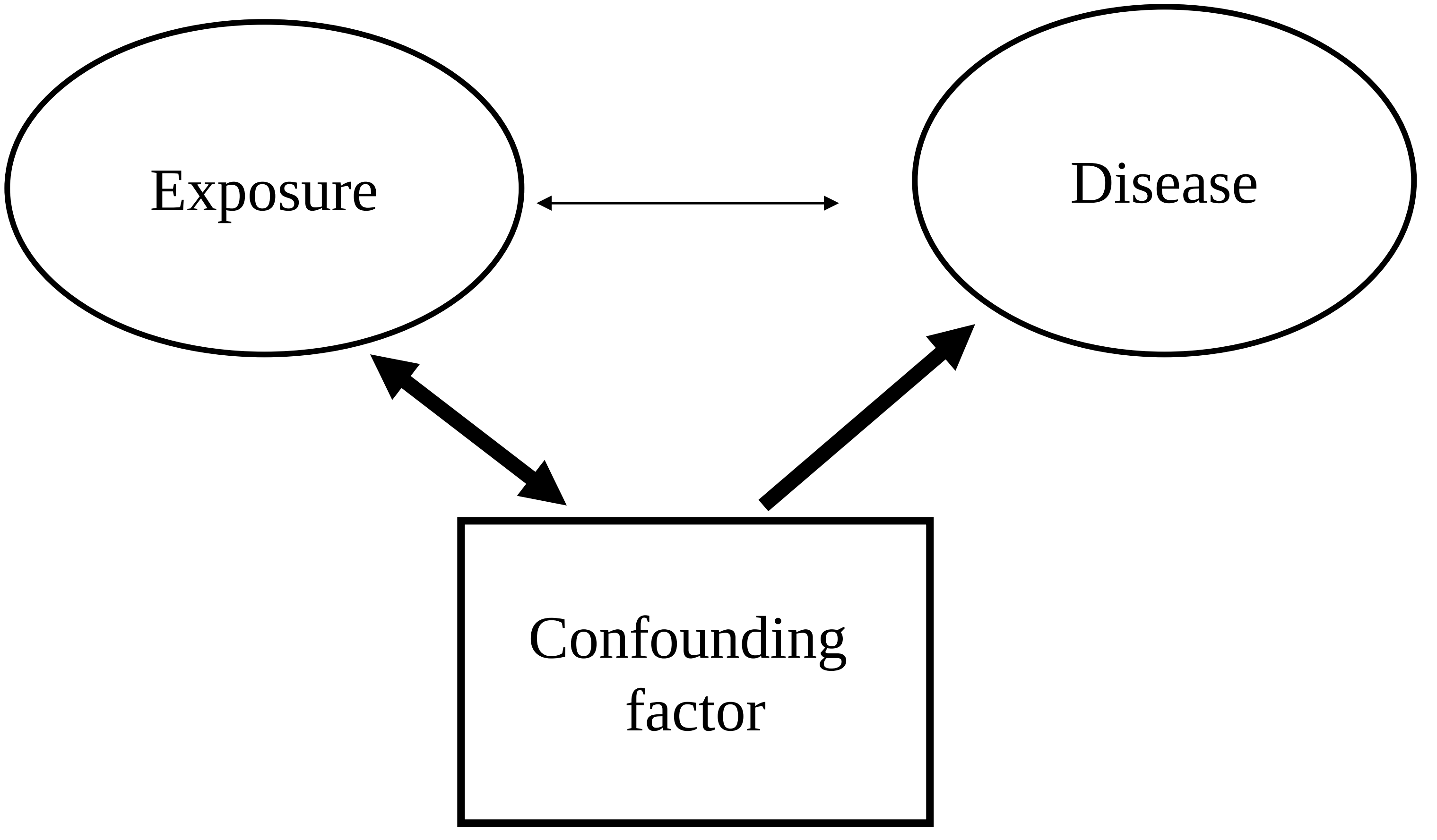
- Chance (random variation) – statistics
- Bias (i.e. systematic error)
- **Confounding**

Confounding

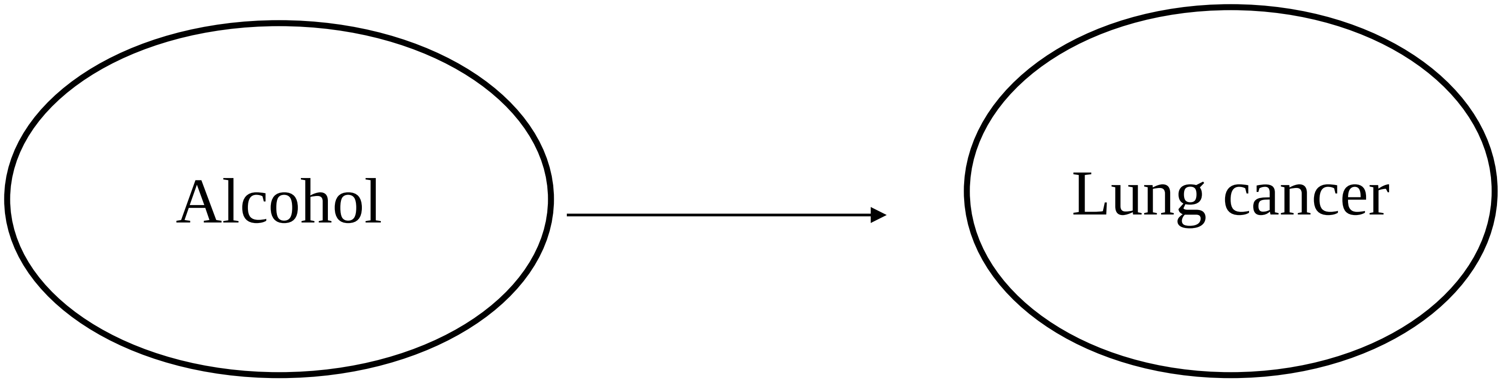
- Situation when a third factor is associated with both exposure and disease

- Association between exposure and disease may not be causal; instead, it is due to a third factor which is associated with both exposure and disease.

Confounding



Example



Case-control study of alcohol and lung cancer

	Alcohol	No alcohol
Cases	450	300
Controls	200	250

Estimated odds ratio = 1.9

The same data stratified by smoking:

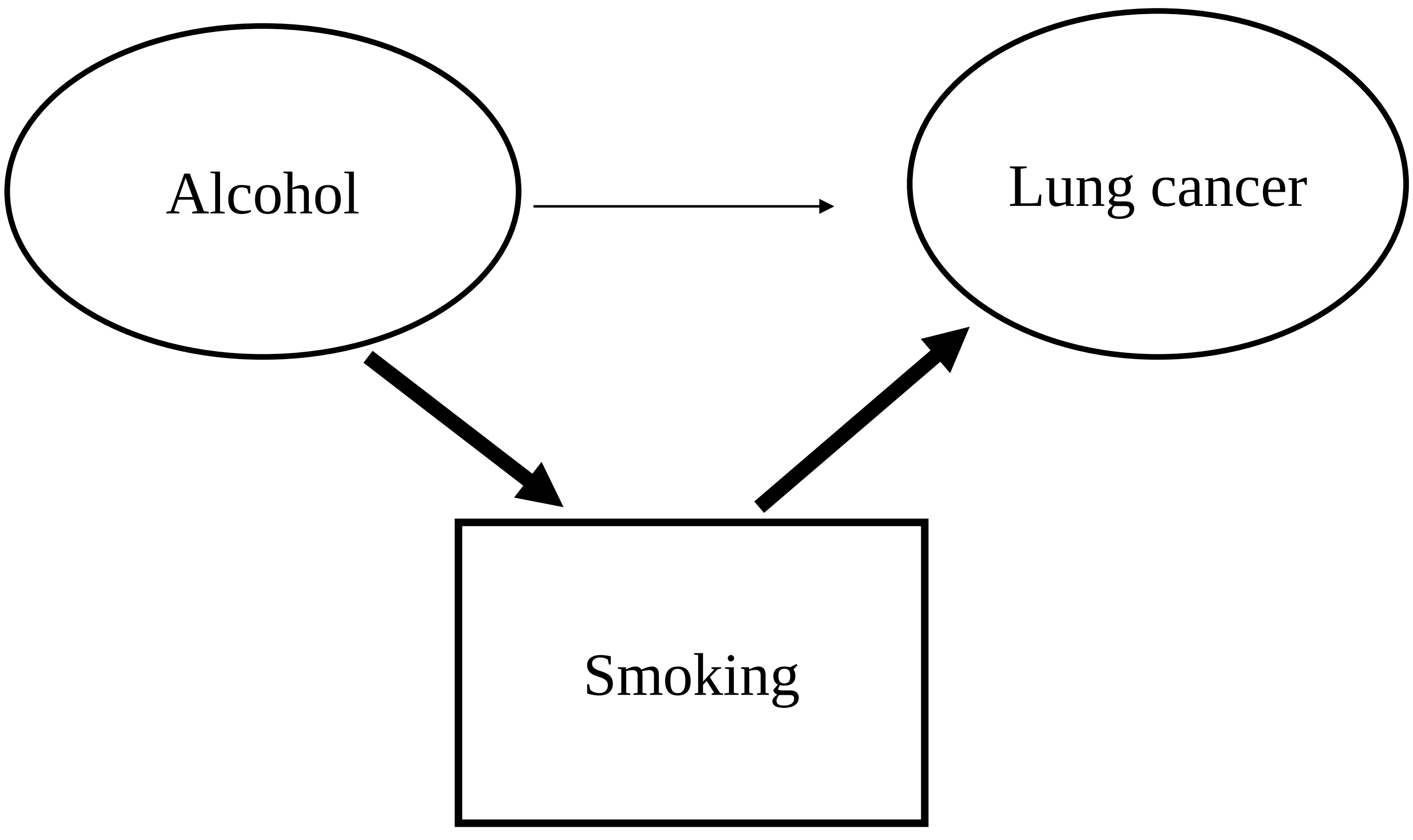
	Non-smokers		Smokers	
	Alcohol	No alcohol	Alcohol	No alcohol
Cases	50	100	400	200
Controls	100	200	100	50
Estimated odds ratio	1.0		1.0	

Alcohol and smoking in controls

	Alcohol	No alcohol
Smokers	100	50
Non-smokers	100	200

Non-drinkers: 1 in 5 were smokers,
Drinkers: 1 in 2 were smokers.

Confounding



Most common confounders:

- Sex (men have higher mortality and more risk factors)
- Age (risk of most diseases increases with age)
- Socioeconomic status (risk of most diseases higher in lower SE groups)
- Ethnic group
- Smoking
- Alcohol
- etc...

Control of confounding

<u>Design</u>

- Randomisation
- Restriction
- Matching

<u>Analysis (if data collected)</u>

- Stratification
- Regression modelling

Step-by-step guide to the stratified analysis

Example

- A study was undertaken to assess whether smokingh increased risk of stomach cancer. Data were collected from 36,000 individuals

	Stomach cancer		
	Yes	No	Total
Smokers	800 (4.0%)	19200	20000
Non-smokers	400 (2.5%)	15600	16000
Total	1200	34800	36000

Example

- $X^2=62.07$ $p<0.001$

$$OR = \frac{\text{Odds(low)}}{\text{Odds(high)}} = \frac{800/19200}{400/15600} = 1.63$$

- 95% CI = 1.44-1.84 (Stata)

- The study found a significantly higher odds of cancer in smokers

But is it real association?

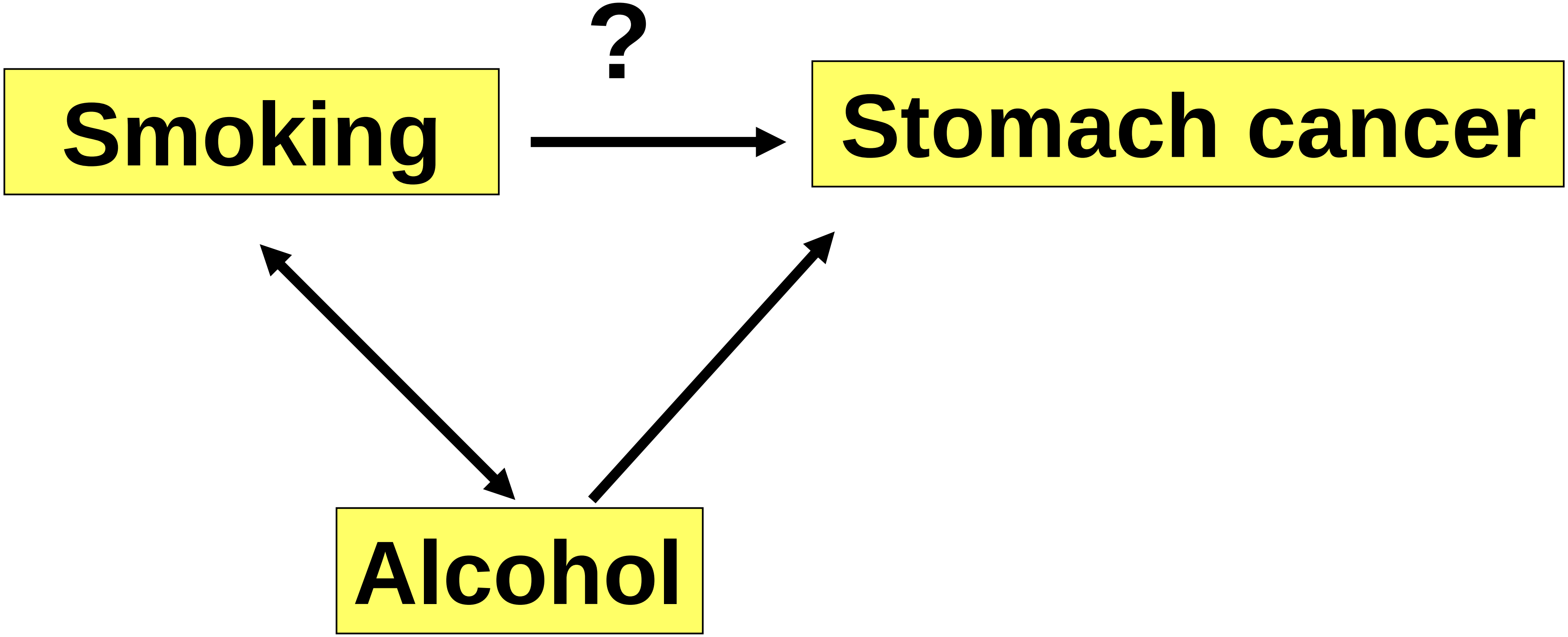
- Smokers are more likely to be drinkers

- Drinking doubles the risk of stomach cancer

?

- THEREFORE

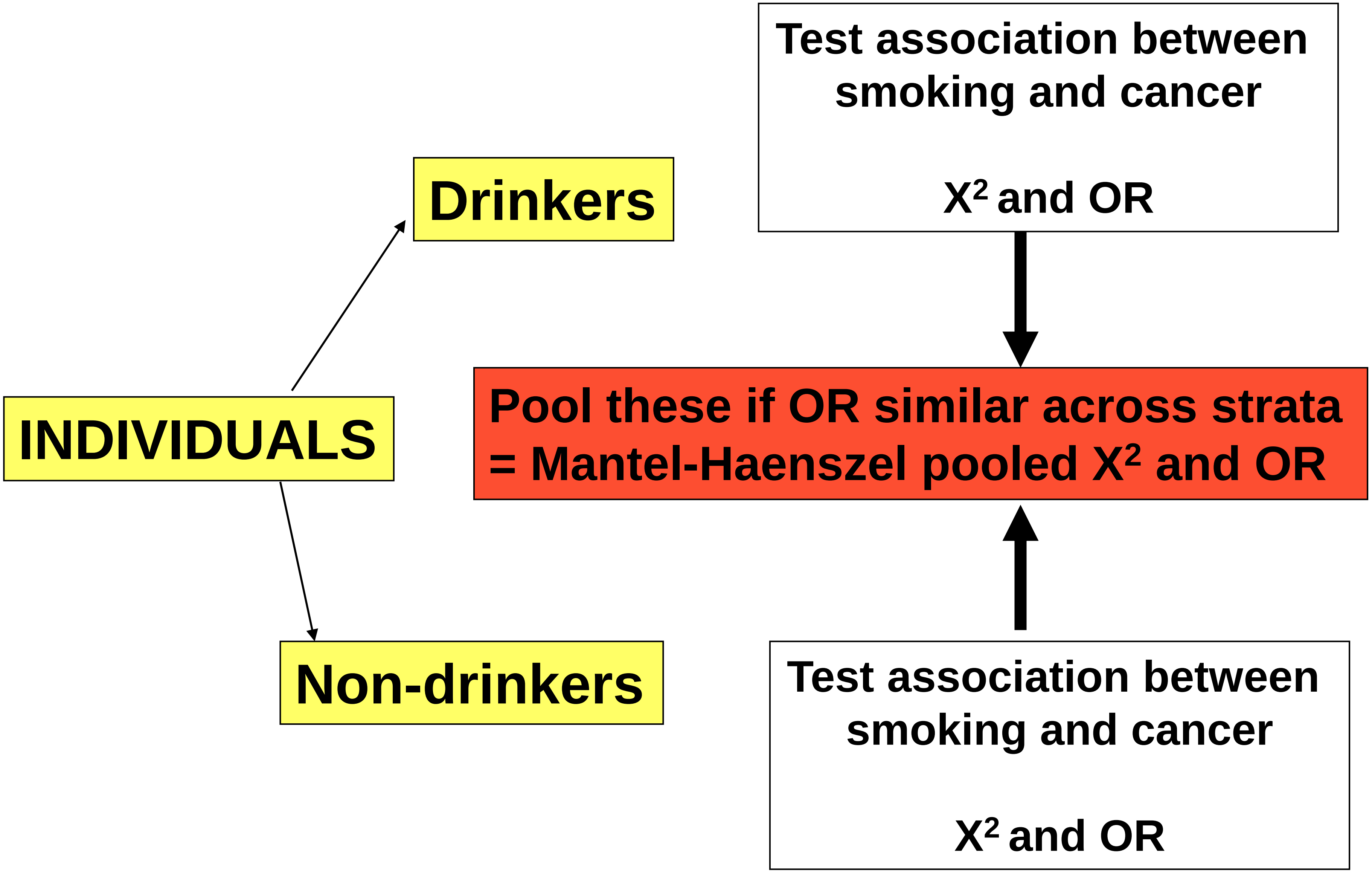
some of the higher risk in smokers could be because they tend to drink more frequently (and have higher risk because of drinking).


?
Smoking
Stomach cancer
Alcohol

Confounding

- We say that alcohol is a **confounding** variable because it is related both to the outcome variable and to exposure (smoking)

- Ignoring alcohol in the analysis leads to **misleading** results


Test association between smoking and cancer
X^2 and OR
Drinkers
INDIVIDUALS
Pool these if OR similar across strata
= Mantel-Haenszel pooled X^2 and OR
Non-drinkers
Test association between smoking and cancer
X^2 and OR

Example

DRINKERS	Stomach cancer		
	Yes	No	Total
Smokers	660	13200	13860
Non-smokers	270	7800	8070
Total	930	21000	21930

DRINKERS	Stomach cancer		
	Yes	No	Total
Smokers	140	6000	6140
Non-smokers	130	7800	7930
Total	270	13800	14070

Example

DRINKERS	Stomach cancer		
	Yes	No	Total
Smokers	660 (4.76%)	13200	13860
Non-smokers	270 (3.35%)	7800	8070
Total	930	21000	21930

NON-DRINKERS	Stomach cancer		
	Yes	No	Total
Smokers	140 (2.28%)	6000	6140
Non-smokers	130 (1.64%)	7800	7930
Total	270	13800	14070

Stratum specific calculations

DRINKERS:

X^2=25.19 p<0.001

OR (95% CI) = 1.44 (1.25-1.67)

NON-DRINKERS

X^2=7.55 p=0.006

OR (95% CI) = 1.40 (1.09-1.79)

Interpretation

- Stratum specific OR are lower than the crude OR (1.44 and 1.40 vs 1.63)
- Stratum specif OR are similar to each other
- This means that it is logical and sensible to pool them
- If they are different (very different) – we should consider drinking to be an EFFECT MODIFIER (the effect of smoking on cancer is modified by drinking status)

Steps for dealing with possible confounders

1. Calculate crude X^2 and OR – DONE (X^2 signif. and OR calculated)

2. List possible confounders – we have chosen alcohol in our example

3. Determine whether they are possible confounders
   a. Association with exposure
   b. Association with outcome
   c. Not on causal pathway

Steps for dealing with possible confounders

4. Do stratified analysis by possible confounder

5. Calculate pooled X^2 and OR (= look at the association that is adjusted for confounder)

6. If crude OR and pooled OR different – conclude that variable is a confounder

Summary of results

- Results are best summarized in the table

Association between smoking and cancer	OR	P-value	Conclusion
Crude assoc.	1.63	<0.001	Odds of cancer 1.63 times higher if smoker
Stratified anal.			
Drinkers	1.44	<0.001	Odds of cancer 1.44 times higher if smoker
Non-drinkers	1.40	0.006	Odds of cancer 1.40 times higher if smoker
Adjusted for drinking	1.43	<0.001	Confounded. Odds of cancer 1.43 times higher rather than 1.63 times higher if smoker

Interpretation of results

- There is still an association between smoking and cancer but less strong than originally showed (in crude analysis)
- The confounding variable (drinking) made the association between smoking and cancer look stronger that it is.
- There is NO STATISTICAL TEST to help you decide whether change in odds ratios (1.63 to 1.43 in our example) is large enough to say that variable is confounder.

Residual confounding

- Unmeasured confounding factors or measurement error in confounding factors may lead to residual confounding.
- The possibility of residual confounding cannot be completely eliminated in observational studies.

Standardisation

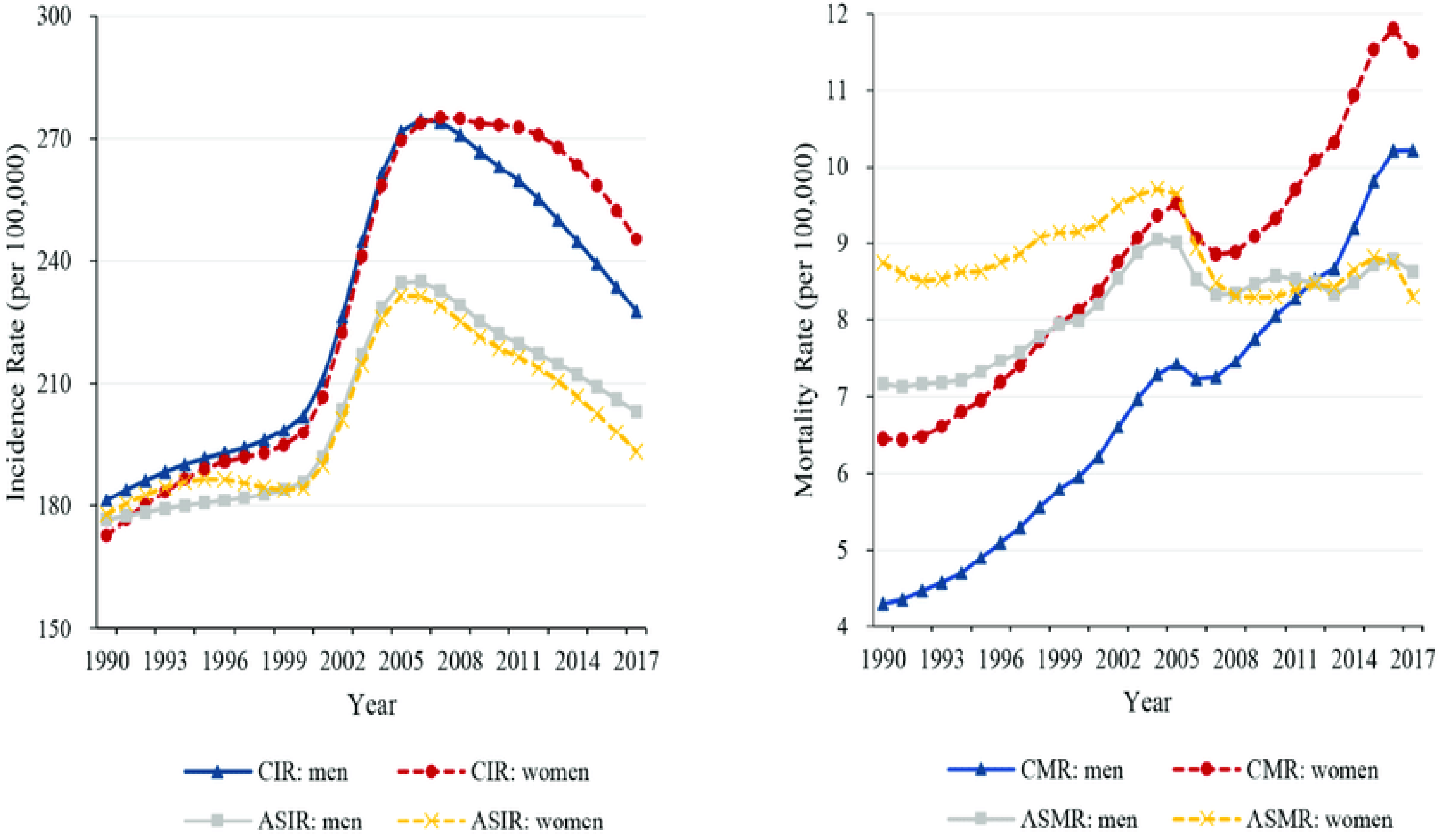
Standardisation in epidemiology

- A numerical (quantitative / statistical) approach to remove confounding by a common characteristic
  - Age
  - Sex
  - Marital status
  - Education

- The most common is standardisation of mortality or incidence rates for age and sex

Trends in crude and age-standardized rates for diabetes mellitus in men and women, China, 1990-2017

(Int J Env Res Public Health. 16. 158. 10.3390/ijerph16010158.)



Crude vs. standardised trends:

- Trends more dramatic for crude rates
- Diabetes strongly associated with older age
- Chinese population is ageing very fast
- Many more “old” people (e.g. 65+) in 2017 than in 1990
- Population ageing distorts the comparisons over time
- Age acts as confounding

Example

Comparison of all-cause mortality rates between Sweden and Panama, 1962

	Sweden			Panama		
Age group	Number deaths	Populati on	Mortality rate / 1000pyrs	Number deaths	Populati on	Mortality rate / 1000pyrs
All ages	73555	7496000	9.8	8281	1075000	7.7

Sweden has mortality rate higher than Panama (9.8 vs 7.7)

Example

	Sweden			Panama		
Age group	Number deaths	Populati on	Mortality rate / 1000pyrs	Number deaths	Populati on	Mortality rate / 1000pyrs
All ages	**73555**	**7496000**	**9.8**	**8281**	**1075000**	**7.7**
0-29	3523	3145000	1.1	3904	741000	5.3
30-59	10928	3057000	3.6	1421	275000	5.2
60+	59104	1294000	45.7	2956	59000	50.1

All age-specific mortality rates are lower in Sweden than in Panama

WHY?

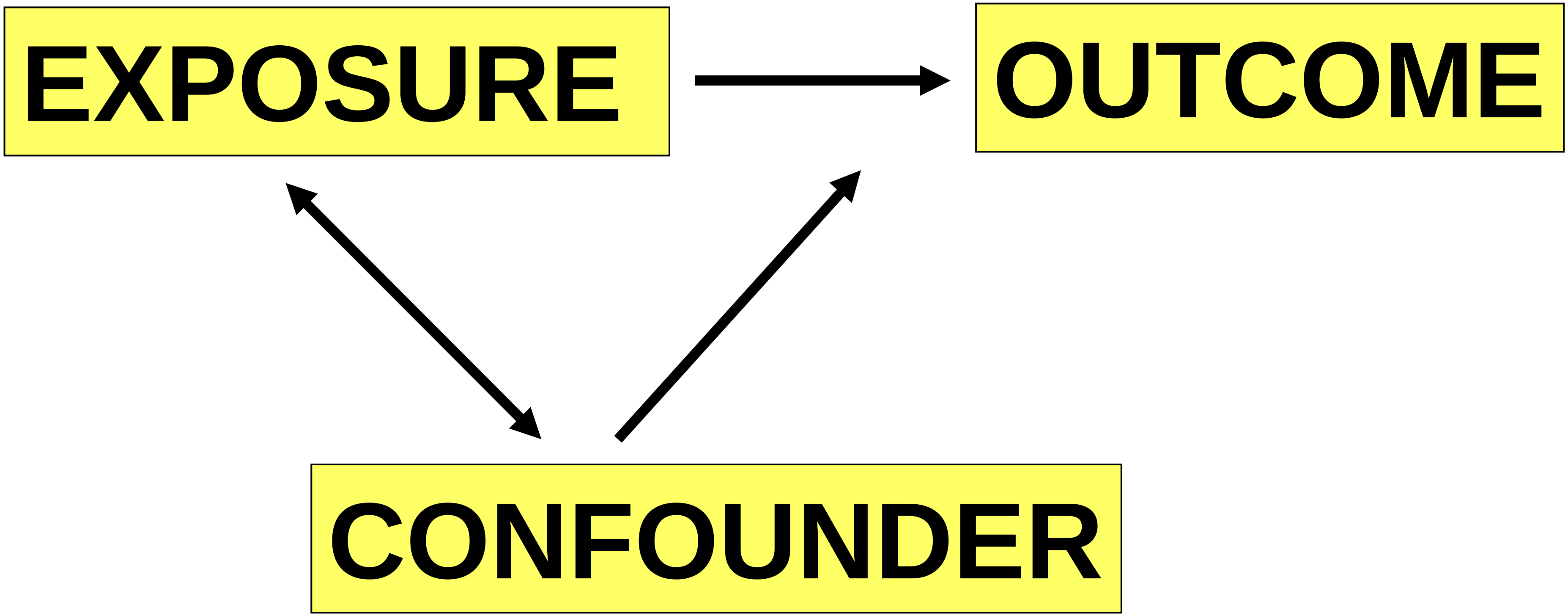
WHY?

Sweden has an older population structure than Panama

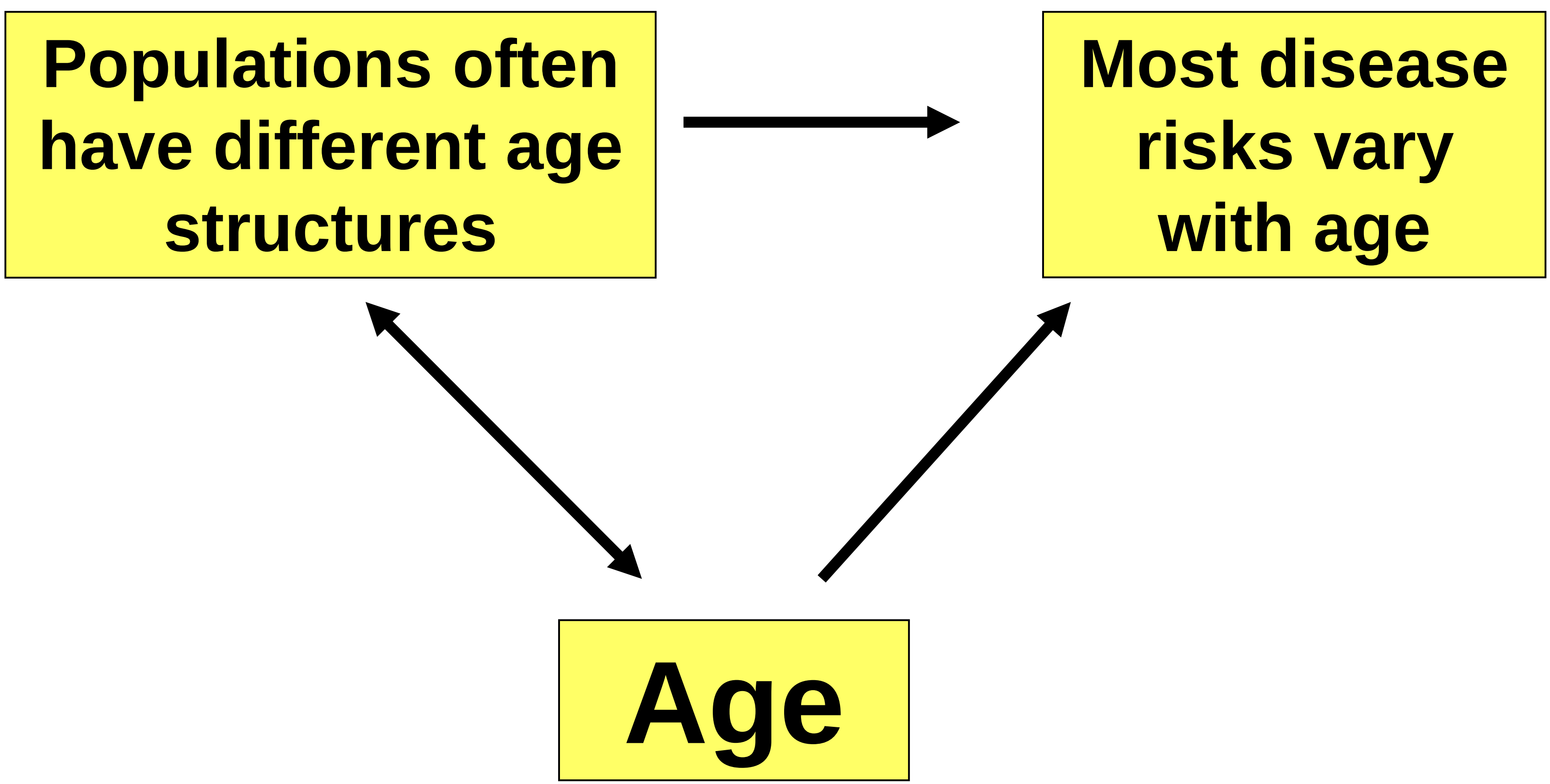
Age group	Sweden	Panama
0-29	42%	69%
30-59	41%	26%
60+	17%	5%

… and mortality increases with age

Age is a confounding factor



Age is a confounding factor



Confounding in epidemiological studies

- At the design stage
  - Randomisation
  - Restriction
  - Matching
- At the analysis stage
  - Statistical modelling
  - Stratification

Summarising stratum specific measures of effect

- We want to **summarise** the effect of E on the risk of D, allowing for the confounding effect of C.
- In order to get this adjusted rate ratio (or odds ratio, or risk ratio) we pool the stratum-specific rate ratios (or odds ratios, or risk ratios).
- A common method of doing this = the Mantel-Haenszel method (known to us from session 6)
- Another major method which uses the principle of stratification is **standardisation**. This method is commonly used when comparing rates.

Example – cont.

- Ideally, we want to have summary measure for each population which has been controlled for different age structure

- Two possibilities:

  - DIRECT standardisation
  - INDIRECT standardisation

Direct vs. Indirect Standardisation

DIRECT

↓

Uses

STANDARD POPULATION STRUCTURE

INDIRECT

↓

Uses

STANDARD SET OF AGE-SPECIFIC RATES

Direct standardisation

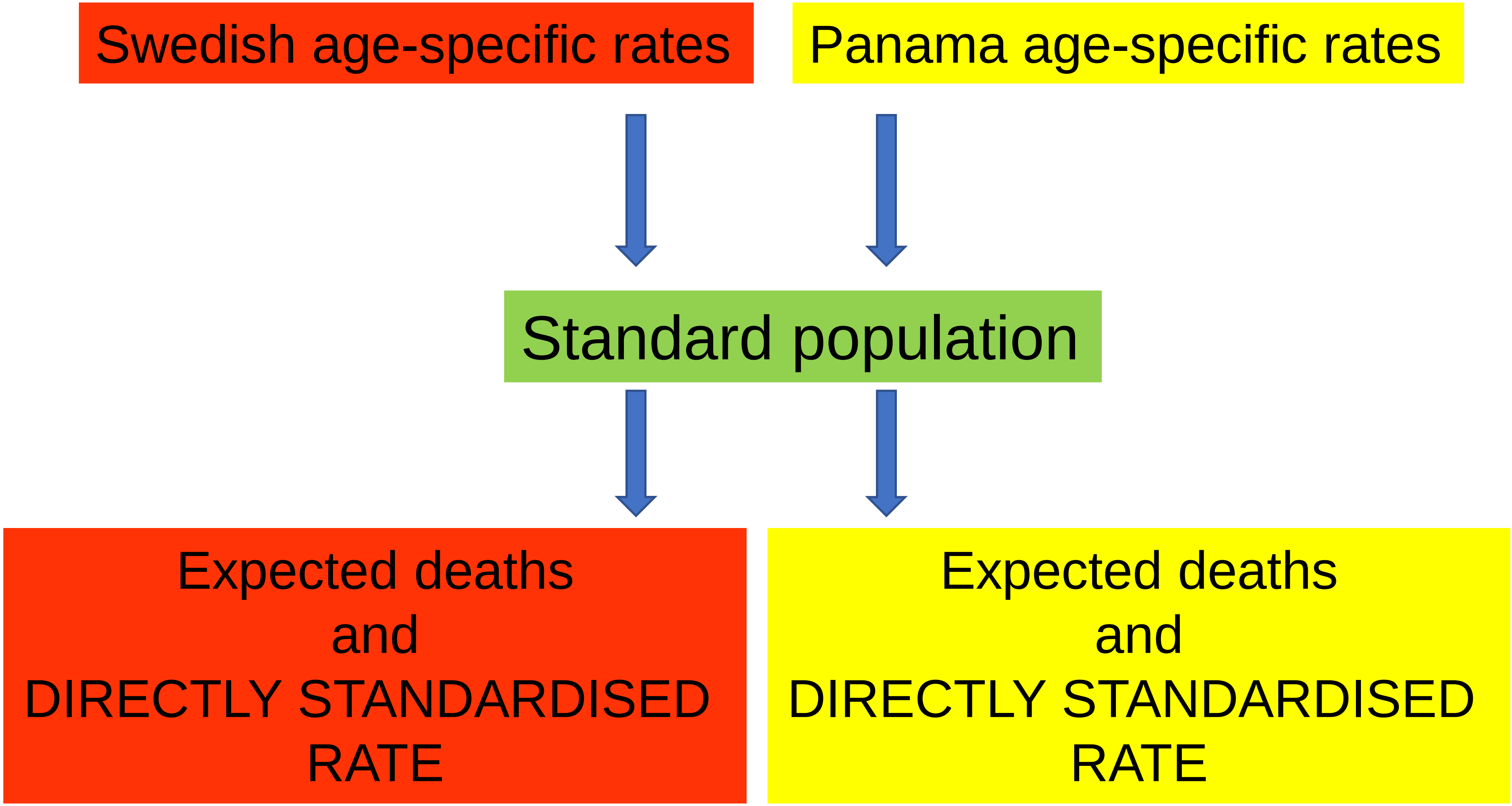
We have “standard population” = hypothetical population with known age structure

Age (years)	Population
0-29	56,000
30-59	33,000
60+	11,000
All ages	100,000

Q1: how many deaths would be expected in Sweden if it had the same age distribution as this standard population

Q2: how many deaths would be expected in Panama if it had the same age distribution as this standard population

Direct standardisation



Example

	Sweden			Panama		
Age group	Number deaths	Populati on	Mortality rate / 1000pyrs	Number deaths	Populati on	Mortality rate / 1000pyrs
All ages	**73555**	**7496000**	**9.8**	**8281**	**1075000**	**7.7**
0-29	3523	3145000	1.1	3904	741000	5.3
30-59	10928	3057000	3.6	1421	275000	5.2
60+	59104	1294000	45.7	2956	59000	50.1

Age specific rates in Sweden (per 1000 pyrs)

0-29	1.1
30-59	3.6
60+	45.7

Age specific rates in Panama (per 1000 pyrs)

0-29	5.3
30-59	5.2
60+	50.1

Standard population	
0-29	56,000
30-59	33,000
60+	11,000

Age	Expected deaths
0-29	0.0011 x 56,000=61.6
30-59	0.0036 x 33,000=118.8
60+	0.0457 x 11,000=502.7
TOTAL	683.1

Age	Expected deaths
0-29	0.0053 x 56,000=296.8
30-59	0.0052 x 33,000=171.6
60+	0.0501 x 11,000=551.1
TOTAL	1019.5

Age-adjusted rates

- Sweden:
  - 683.1/100,000=6.8 per 1,000 person years
- Panama:
  - 1019.5/100,000=10.2 per 1,000 person years

These rates can be interpreted as the mortality rates that these two countries would have if their age distributions were changed from what they actually were to the age distribution of the standard.

Direct standardisation

- A weighted average of the age-specific rates
- Weights = population in strata of standard population
- Weights are the same = Age-standardised rates can be directly compared

- We can calculate age-standardised rate ratio:

10.2/6.8=1.5

What standard population?

Age distributions of the standard populations used for age standardization

Age group	World ASR (W)	European ASR (E)
0-4	12 000	8 000
5-9	10 000	7 000
10-14	9 000	7 000
15-19	9 000	7 000
20-24	8 000	7 000
25-29	8 000	7 000
30-34	6 000	7 000
35-39	6 000	7 000
40-44	6 000	7 000
45-49	6 000	7 000
50-54	5 000	7 000
55-59	4 000	6 000
60-64	4 000	5 000
65-69	3 000	4 000
70-74	2 000	3 000
75-79	1 000	2 000
80-84	500	1 000
85+	500	1 000
Total	100 000	100 000

WHO standard populations

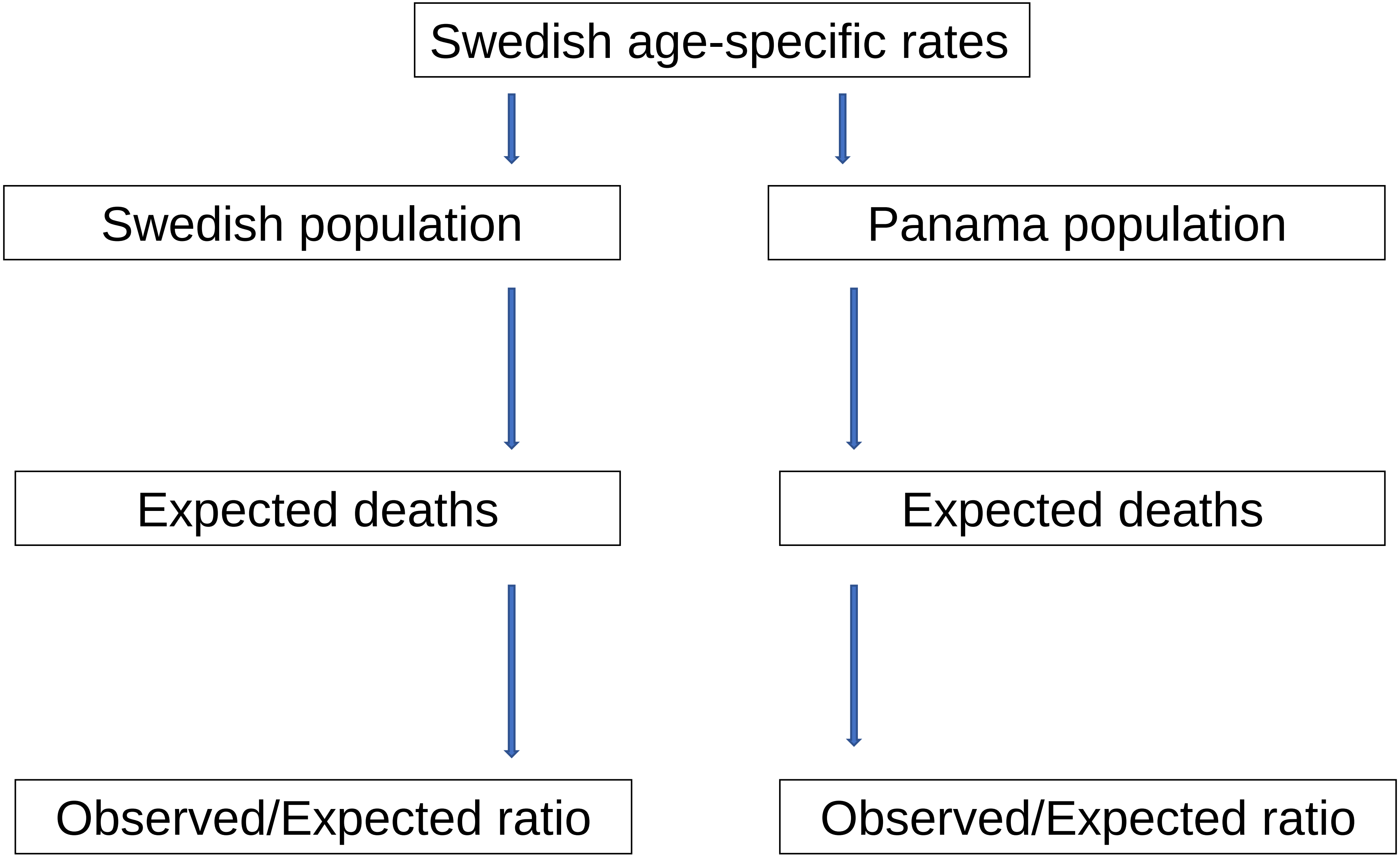
Indirect standardisation

- Let’s assume that the total number of deaths for Panama is known but their distribution by age is not available
- It is not possible to use the direct method of standardisation.

	Sweden			Panama		
Age group	Number deaths	Populati on	Mortality rate / 1000pyrs	Number deaths	Populati on	Mortality rate / 1000pyrs
All ages	**73555**	**7496000**	**9.8**	**8281**	**1075000**	**7.7**
0-29	3523	3145000	1.1	NA	741000	-
30-59	10928	3057000	3.6	NA	275000	-
60+	59104	1294000	45.7	NA	59000	-

Indirect standardisation

- It is possible to calculate how many deaths would be expected in Panama and in Sweden if both these countries had the same age-specific mortality rates as Sweden
- Swedish age-specific rates will be taken as a set of standard rates



Swedish age-specific rates
Swedish population
Panama population
Expected deaths
Expected deaths
Observed/Expected ratio
Observed/Expected ratio

Age-spec rates in Sweden	
0-29	1.1
30-59	3.6
60+	45.7

Swedish population	
0-29	3,145,000
30-59	3,057,000
60+	1,294,000

Panama population	
0-29	741,000
30-59	275,000
60+	59,000

Age	Expected deaths
0-29	0.0011 x 3,145,000=3,523
30-59	0.0036 x 3,057,000=10,928
60+	0.0457 x 1,294,000=59,104
TOTAL	73,555

Age	Expected deaths
0-29	0.0011 x 741,000=815.5
30-59	0.0036 x 275,000=990.0
60+	0.0457 x 59,000=2,696.3
TOTAL	4,501.4

SWEDEN	PANAMA
Total expected deaths (E) = 73,555 Total observed deaths (O) = 73,555	Total expected deaths (E) = 4,501 Total observed deaths (O) = 8,281
O/E (%) = 100	O/E (%) = 184

STANDARDISED MORTALITY RATIO
SMR
(rate ratio)

The SMR for Panama is equal to 184 = the number of observed deaths was 84% higher than the number we would expect if the Panama had the same mortality experience as Sweden.

Comparison of the methods

- Direct method uses STANDARD population structure
- Indirect method uses STANDARD set of age-specific rates

Data needed for each study population

- Direct method: number of cases by age group, population numbers by age group (to be able to calculate age specific rates)
- Indirect method: total number of cases only, population number by age group

Method

- Direct method: select standard population, apply age specific rates to standard population
- Indirect method: choose standard age-specific rates and apply them to each study population

Which method preferable?

- Decision depends on what data are available
- The direct method requires stratum-specific rates (e.g. age-specific rates) in all the populations under study whereas the indirect method only requires the total number of cases
- If stratum-specific rates are not available for the study population, the indirect method may provide the only feasible approach

Which method preferable?

- Indirect method preferred when there are small numbers in age-specific groups. Rates in direct adjustment would be based on these small numbers and would be subjected to substantial sampling variation.
- With indirect adjustments the summary rates are more stable because we can choose the most stable rates as the standard rates

STANDARDISED MEANS

- Same principle as with proportions/rates
- If continuous variable is related for example to age and age structure differs in 2 populations the comparisons of means of continuous variable might be misleading

SUMMARY

- Confounding is hugely important issue in epidemiology
- Common alternative explanation for observed association
- Can be controlled by design or analysis

Adjustment = analytical approach to control for confounding

Standardisation - uses stratification method

- Two types of standardisation
- Direct x Indirect standardisation
- Standardized rates
- Standardized means