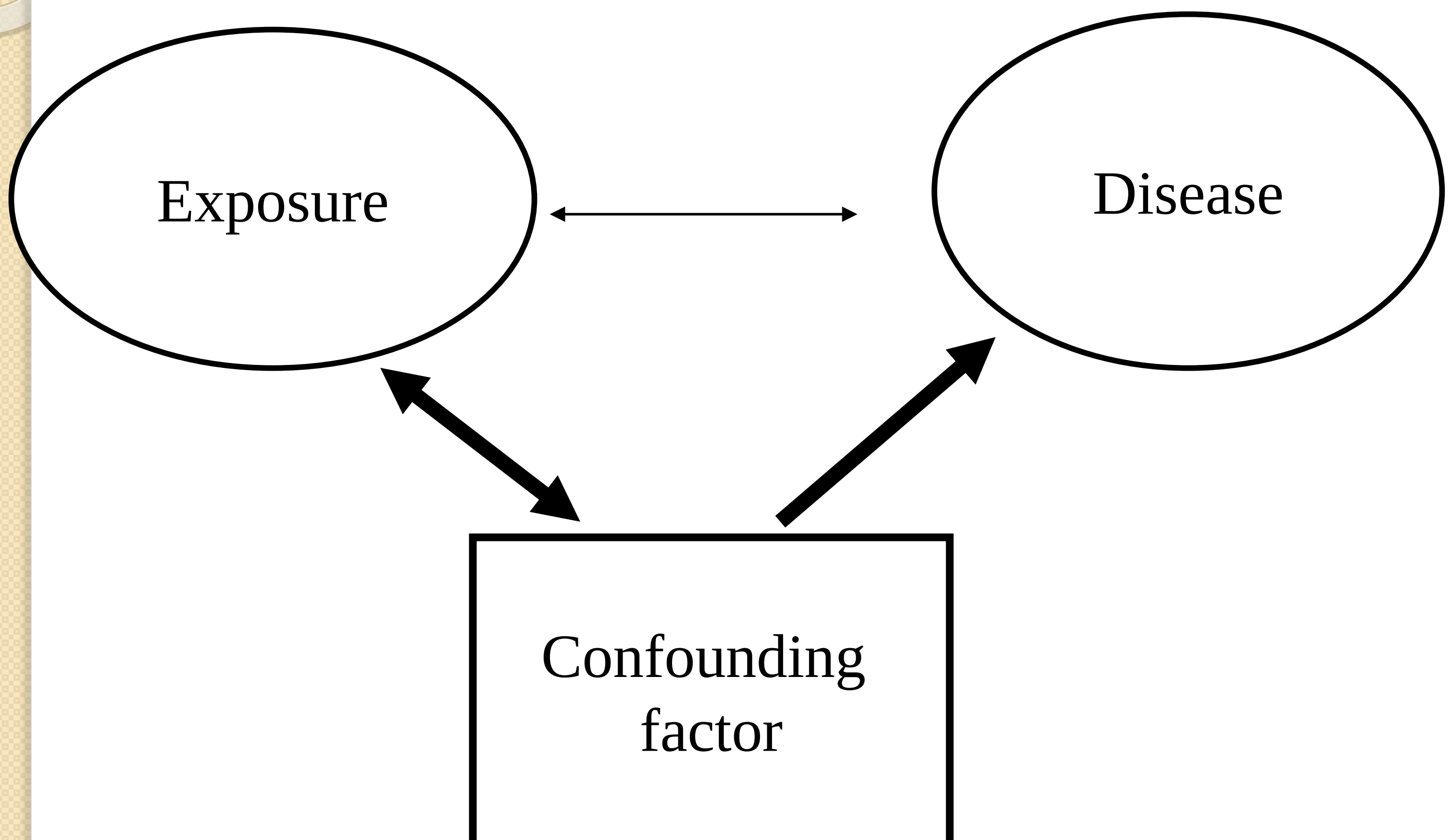
Confounding

- 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



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:

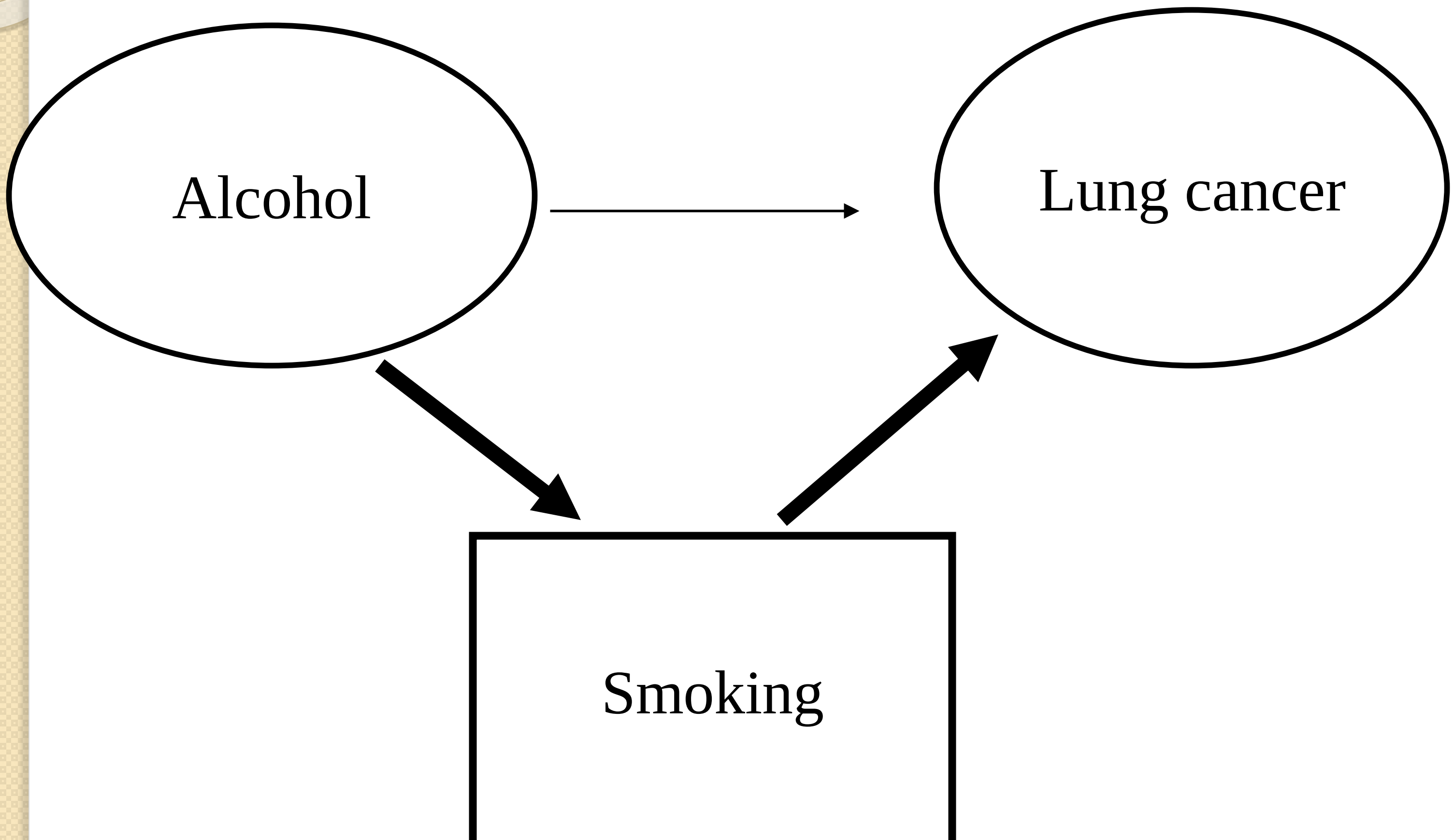
	Non-smokers		Smokers	
	Alcohol	No alcohol	Alcohol	No alcohol
Cases	50	100	400	200
Controls	100	200	100	50
Estimated odds ratio	1.0		1.0	

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:

- Gender (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

Design

- Randomisation
- Restriction
- Matching

Analysis (if data collected)

- Stratification
- Regression modelling

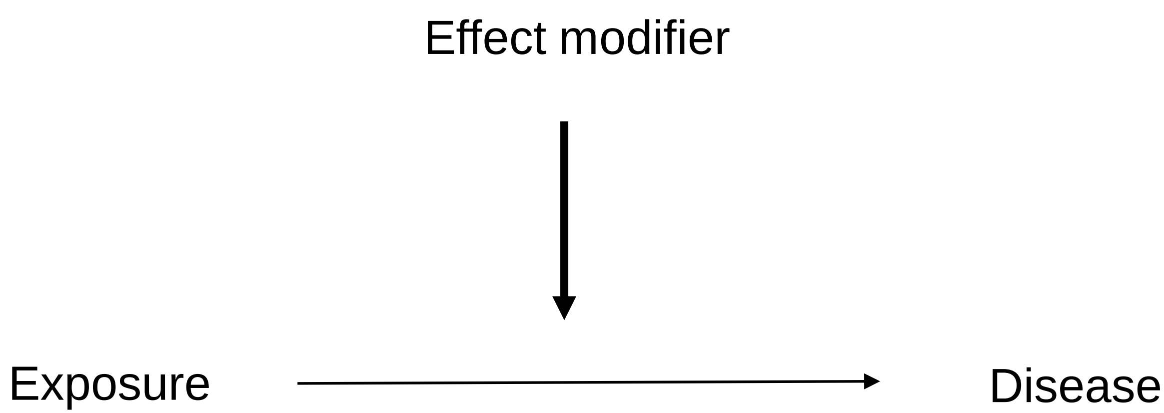
Residual confounding

- 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.

Effect modification (interaction)

- the effect of exposure on disease is dependent on the level of a third factor

Effect modification



Positive and negative effect modification

- Positive:
  - “susceptibility factor” or “vulnerability factor”,
  - its presence (or higher values) strengthens the association between exposure and disease.
- Negative:
  - “resiliency factor” or “buffering factor”
  - its presence (or higher values) weakens the association between exposure and disease

CHD, smoking and age in British doctors study (rates per 100,000)

	Non-smokers	*Heavy smokers*	
	Rate	*Rate*	*RR*
<45	7	104	14.9
45-54	118	393	3.3
55-64	531	1025	1.9

Identification of effect modification

- Stratified analysis
- Compare effect estimates in strata
- Assess differences in effects by significance tests (p-value for heterogeneity)

- Pooled estimates (e.g. standardised) **not appropriate** when there is an interaction

Confounding vs. interaction

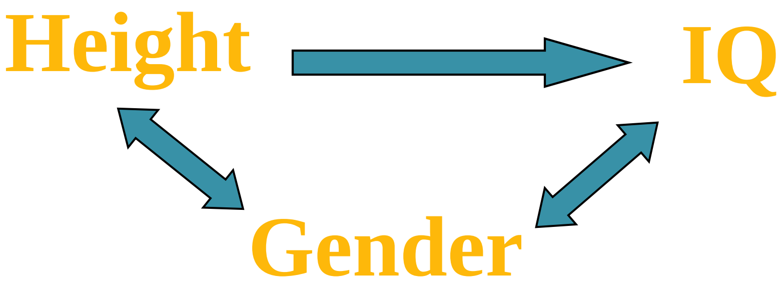
Confounding

- Alternative explanation
- Distorts the "truth"
- Efforts to remove it to get nearer to the "truth"
- When present, stratum specific effects are similar to each other but different from the overall crude effect.

Effect modification

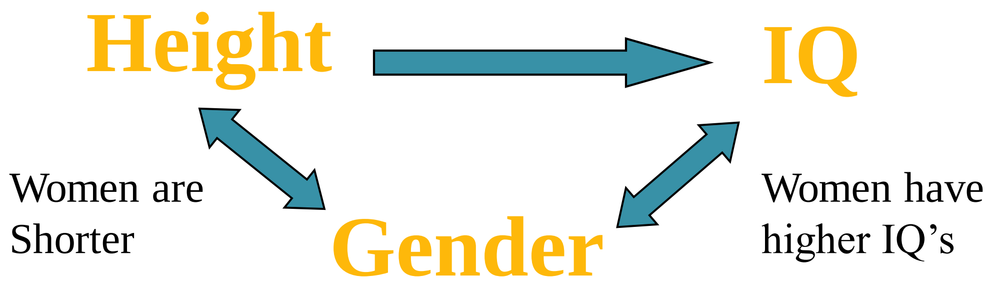
- One factor modifies effect of another factor
- It is genuine, not artefact
- Property of the relationship between factors
- We should detect and describe it but not remove it.

Example: Height and IQ – real association or not?



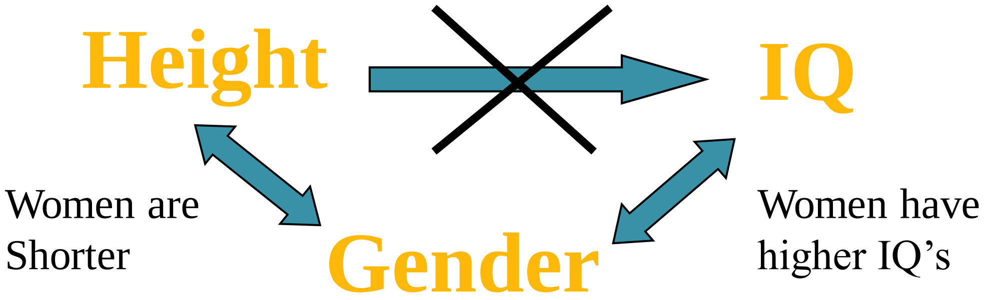
- High negative association between height and IQ

Height and IQ



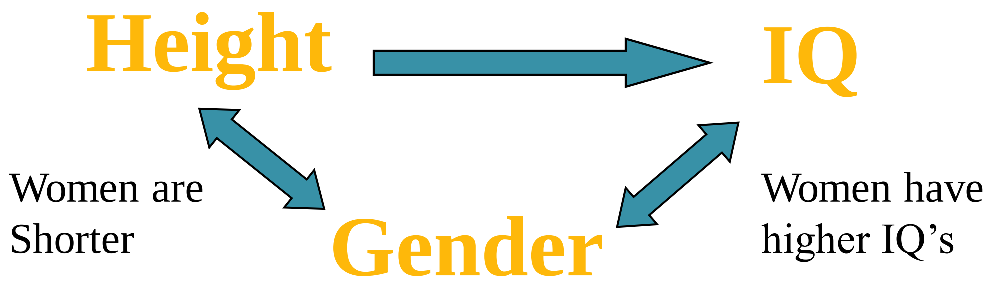
- Find out that Gender is related to Height and that Gender is related to IQ
- Therefore, Gender is a *potential* confounder

Height and IQ



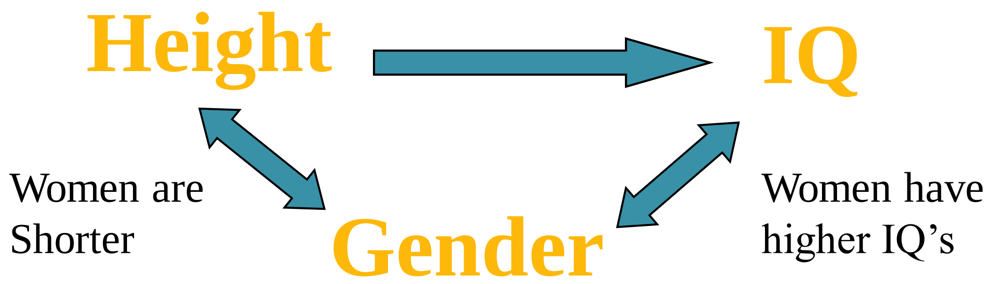
- If after adjustment for Gender there is NO association between height and IQ, then Gender was a confounder

Height and IQ



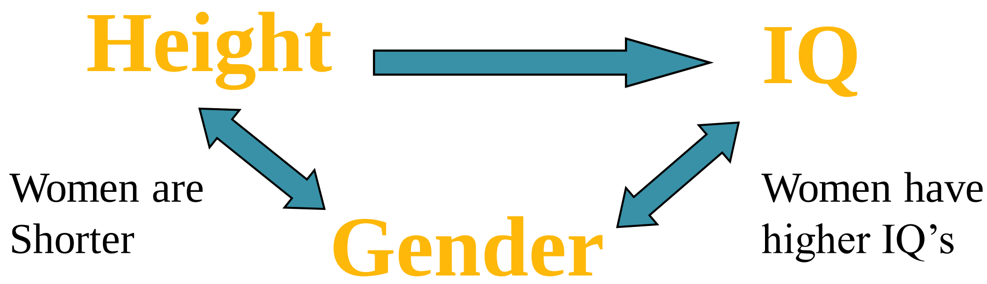
- If after adjustment for Gender there is still a strong negative association between Height and IQ, then Gender is not a confounder

Height and IQ



- If after adjustment for Gender there is still an association between Height and IQ, but the nature and/or strength of the association changes with Gender, then Gender is an **Effect Modifier**.

Height and IQ



- If there is no association between Gender and IQ, then Gender cannot be a confounder
- Likewise, if gender is not associated with height, then Gender cannot be a confounder
- The confounder must be related to both the cause and the effect

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

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

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

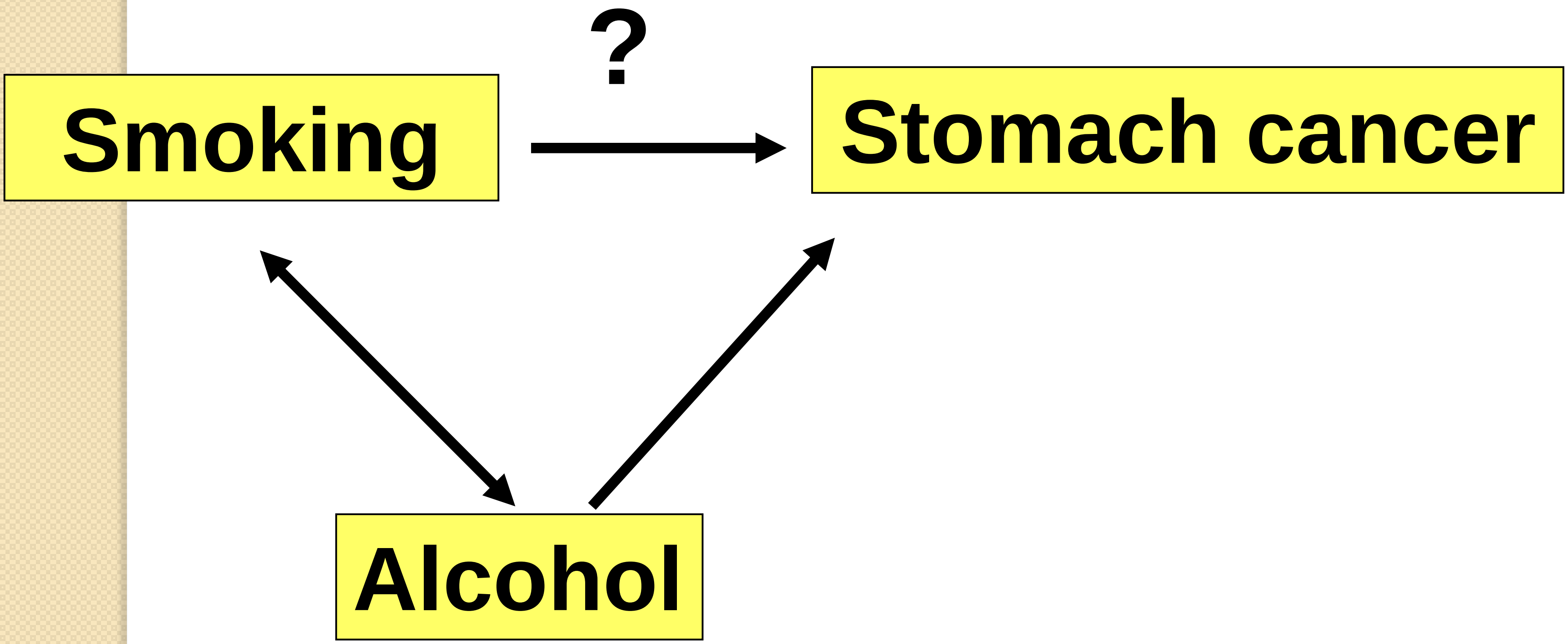
- 95% CI = 1.44-1.84 (Stata)
- The study found a significantly higher odds of cancer in smokers

But is it real association?

- Smokers are more likely to be drinkers
- Drinking doubles the risk of stomach cancer

?

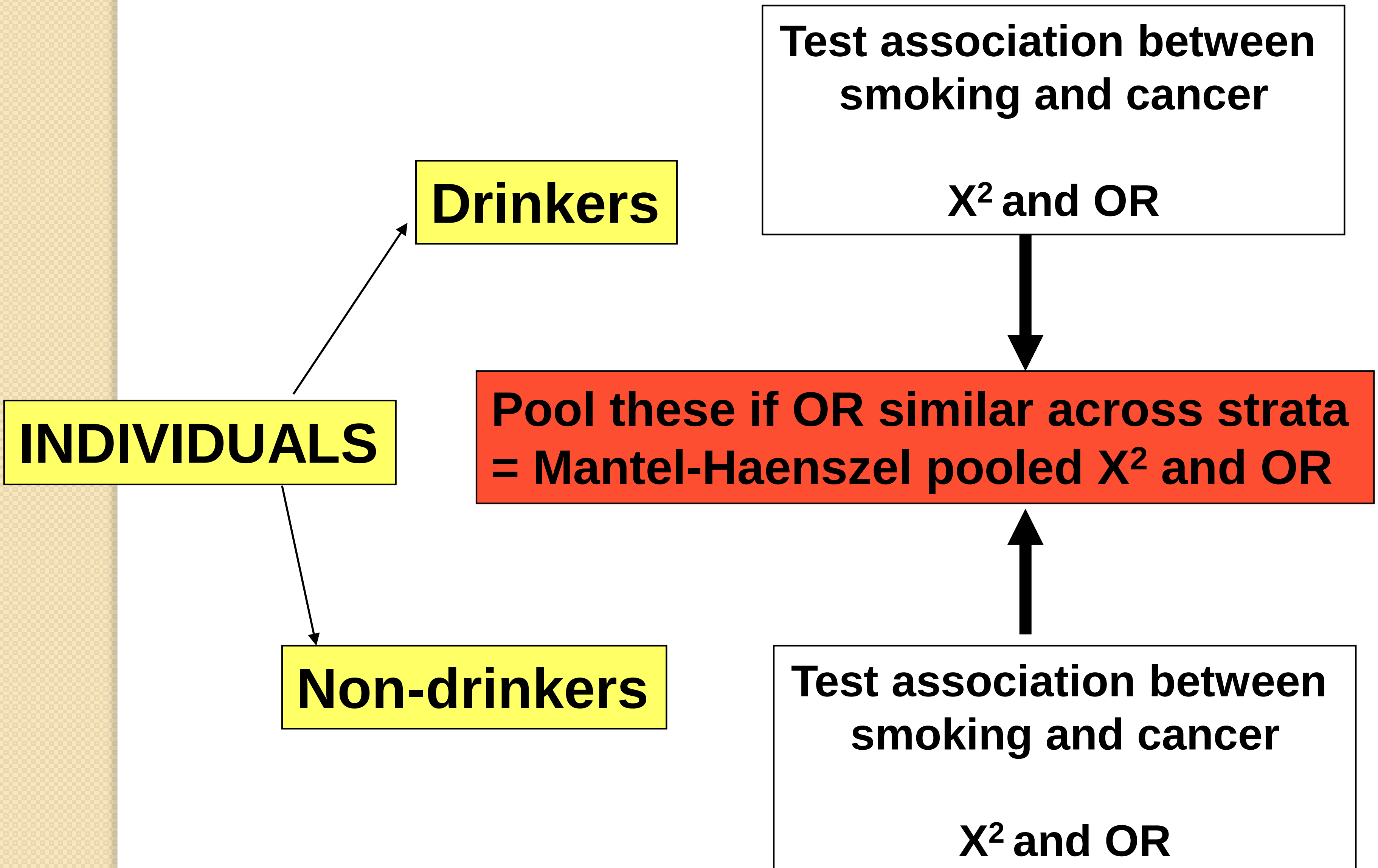
- THEREFORE some of the higher risk in smokers could be because they tend to drink more frequently (and have higher risk because of drinking).


?
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


INDIVIDUALS
Drinkers
Non-drinkers
Test association between smoking and cancer
X2 and OR
Pool these if OR similar across strata = Mantel-Haenszel pooled X2 and OR
Test association between smoking and cancer
X2 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)

- 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

Mantel-Haenszel pooled X^2 and OR

$$OR = \frac{\sum \left(a_i d_i / n_i \right)}{\sum \left(b_i c_i / n_i \right)}$$

```
. mhodds cancer smok, by(drink)

Maximum likelihood estimate of the odds ratio
Comparing smok==2 vs. smok==1
by drink

-------------------------------------------------------------------------
    drink | Odds Ratio   chi2(1)         P>chi2       [95% Conf. Interval]
----------+--------------------------------------------------------------
        1 |   1.444444    25.19          0.0000        1.25020    1.66886
        2 |   1.400000     7.55          0.0060        1.10001    1.78181
-------------------------------------------------------------------------
   Mantel-Haenszel estimate controlling for drink
    ----------------------------------------------------------------
     Odds Ratio     chi2(1)         P>chi2        [95% Conf. Interval]
    ----------------------------------------------------------------
       1.433140       32.73          0.0000        1.266074   1.622251
    ----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1)    =    0.05
                                     Pr>chi2    =  0.8274
```

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.

Effect modification

- We still need to check one important aspect of M-H analysis – we make the assumption that the association between exposure and the outcome is the same in each level of confounding factor
- If this is NOT true, then you cannot combine stratum specific ORs into one pooled estimate
- If the exposure-outcome association varies in different levels of third variable we say that such third variable modifies the effect of exp on outcome

Effect modification

- Third variable can be called EFFECT MODIFIER

- Effect modification = interaction = heterogeneity between strata

- Testing for effect modification – Kirkwood and Sterne, 186-187

- We will look back to STATA output

```
. mhodds cancer smok, by(drink)

Maximum likelihood estimate of the odds ratio
Comparing smok==2 vs. smok==1
by drink

-------------------------------------------------------------------------
    drink | Odds Ratio   chi2(1)         P>chi2       [95% Conf. Interval]
----------+--------------------------------------------------------------
        1 |   1.444444    25.19          0.0000        1.25020    1.66886
        2 |   1.400000     7.55          0.0060        1.10001    1.78181
-------------------------------------------------------------------------
   Mantel-Haenszel estimate controlling for drink
    ----------------------------------------------------------------
     Odds Ratio     chi2(1)         P>chi2        [95% Conf. Interval]
    ----------------------------------------------------------------
       1.433140      32.73          0.0000        1.266074   1.622251
    ----------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1)   =    0.05
                                     Pr>chi2   =  0.8274
```

Example

- STATA = test of homogeneity (NULL hypothesis is that stratum specific ORs are homogenous)

- Our example – test of homogeneity: $p=0.83$

- We can assume that stratum specific estimates are same or similar and we can use pooled estimate

When is effect modification important?

- If we find that stratum specific odds ratios are not homogenous (p-value for test of homogeneity <0.05) we cannot report pooled estiamte
- We need to report stratum specific results!
- Test for homogeneity has low power; → a large p-value does not establish the absence of effect modification. Small p-value however suggest that effect modification is substantial

How to examine effect modification

- Always examine stratum specific odds ratios – how different do they look?
- If there is clear evidence of effect modification, report the exp-outcome association separately for each stratum
- If there is moderate evidence of effect modification, report both M-H OR and stratum specific OR
- If no evidence of effect modification, use M-H OR