Confounding
&
Effect modification

E2040: Introduction to Epidemiology and Environmental Health

2024

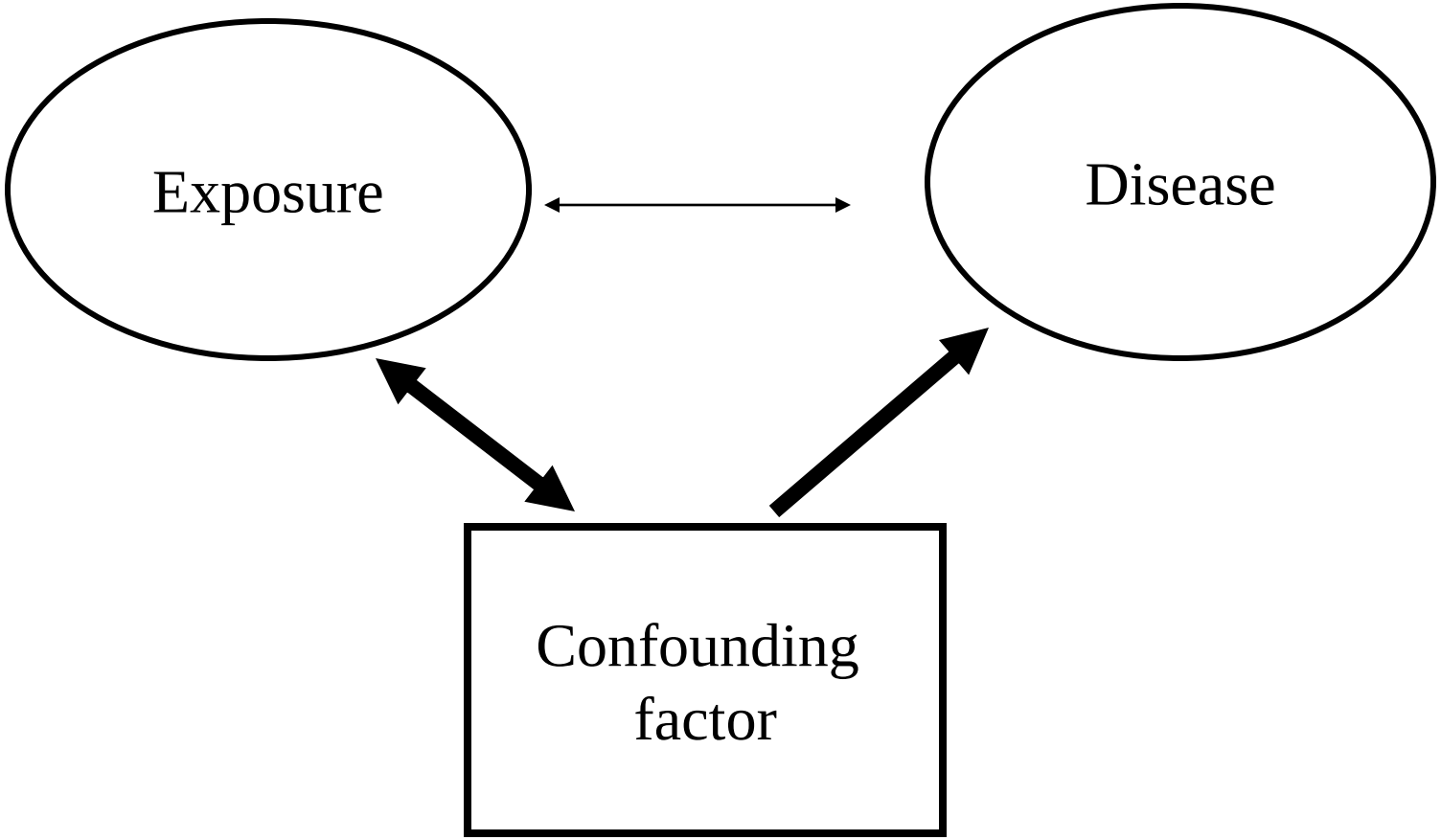
Three major issues in alternative interpretation of observed association:

- Chance (random variation)
- Bias
- Confounding

What is confounding?

- Another factor (alternative explanation) might be causing an observed association: **confounding**

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	
	Alcohol	No alcohol
Cases	50	100
Controls	100	200
Estimated odds ratio	1.0	

The same data stratified by smoking:

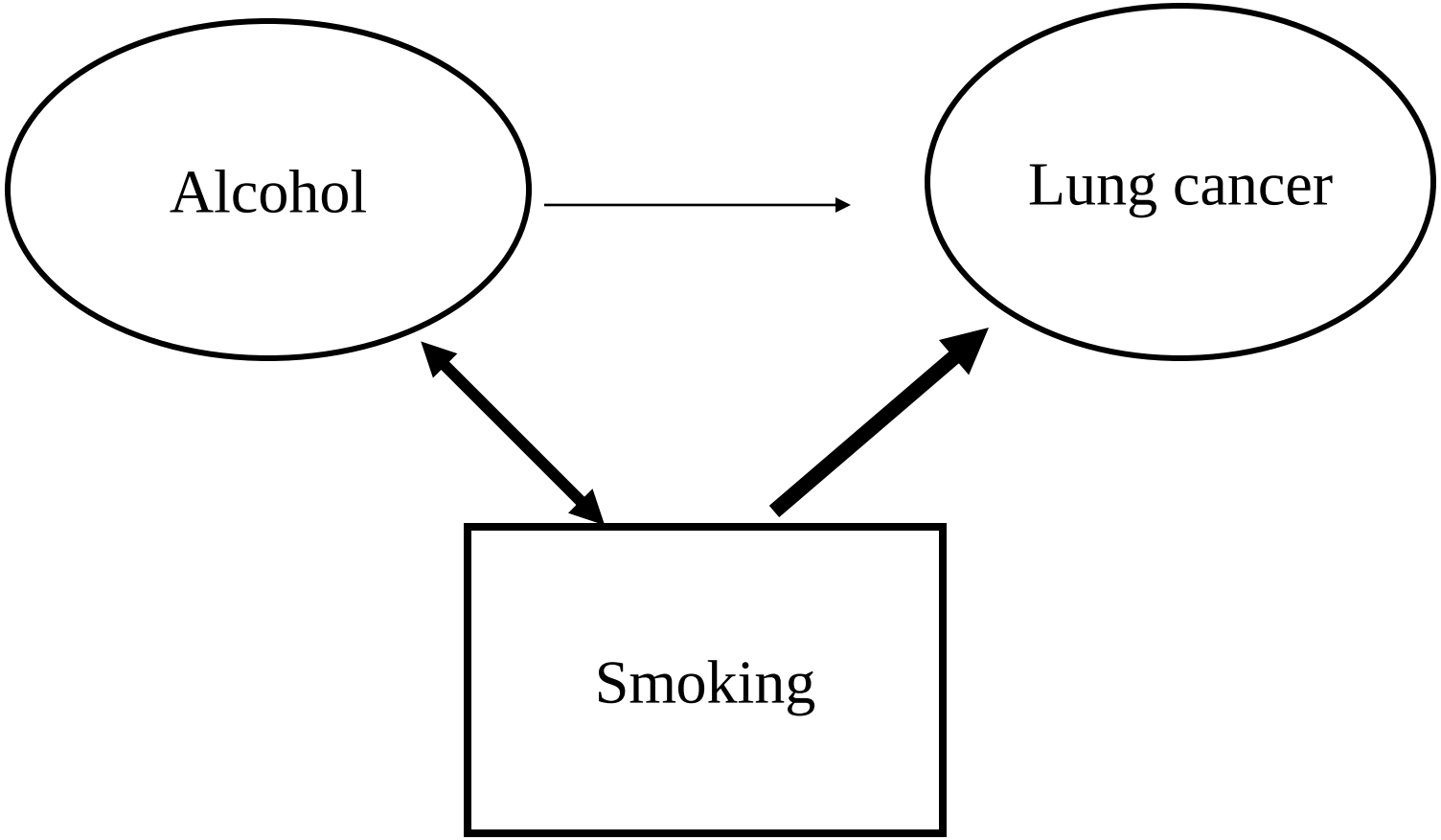
	Non-smokers		Smokers	
	Alcohol	No alcohol	Alcohol	No alcohol
Cases	50	100	400	200
Controls	100	200	100	50
Estimated odds ratio	1.0		1.0	

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



More explanations of confounding

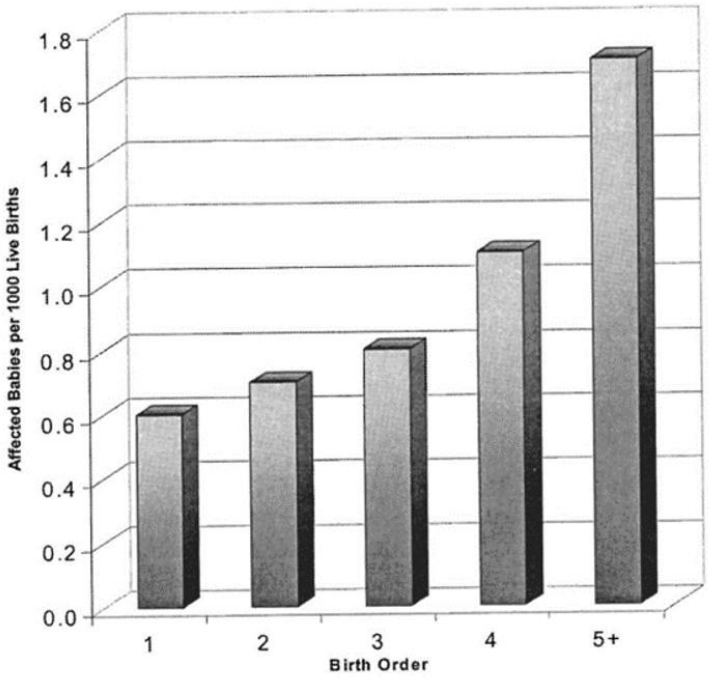
- **Confounding** refers to a situation in which a non-causal association between a given exposure and outcome is observed due to the influence of a third variable, usually referred to as a **confounder**.

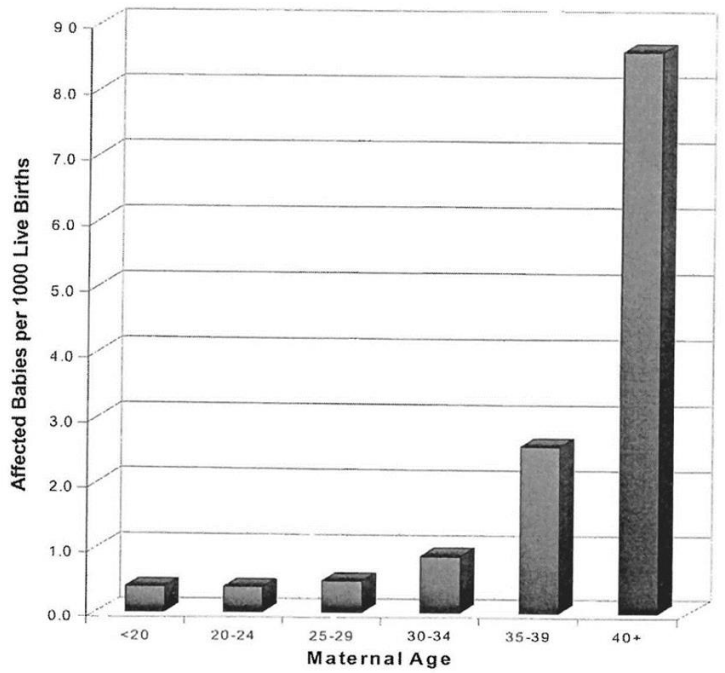
"Confounding is confusion, or mixing, of effects; the effect of the exposure is mixed together with the effect of another variable, leading to bias"

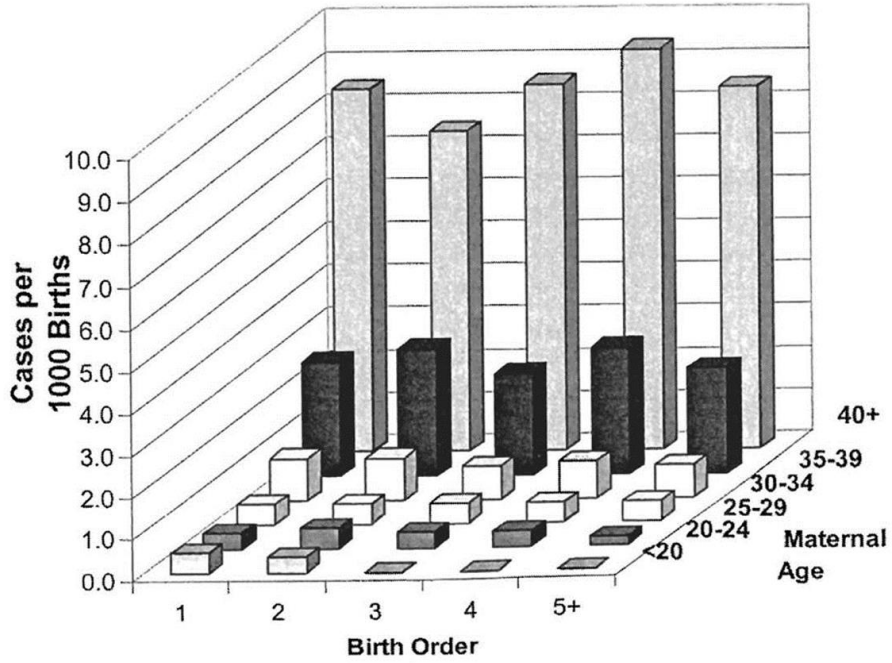
-Rothman, 2002

- Sex (men have higher mortality and more risk factors)
- Age (risk of most chronic diseases increases with age)
- Socioeconomic status (more lifestyle and behavioral risk factors, poorer healthcare access at lower SES)
- Ethnic group (less healthcare access, higher environmental exposures, more discrimination among under-represented groups)
- Smoking
- Alcohol consumption
- Etc.

Another example – birth order and Down Syndrome





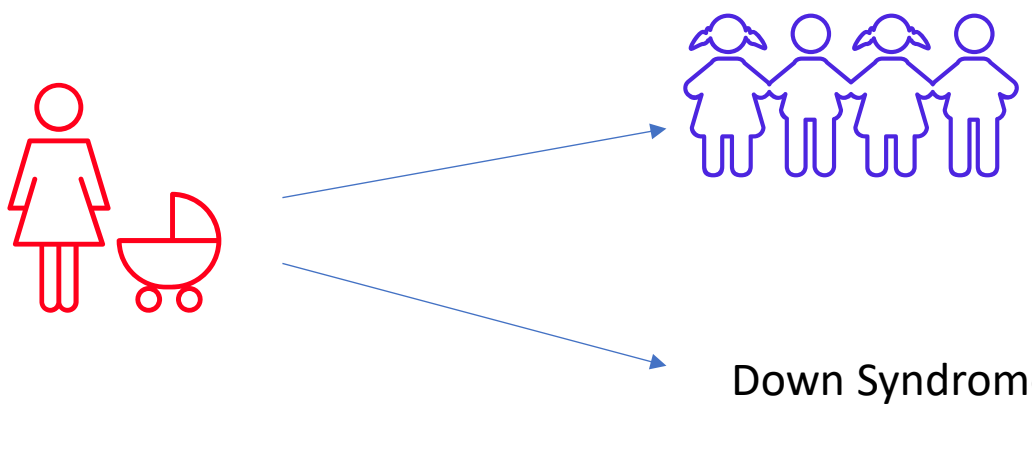


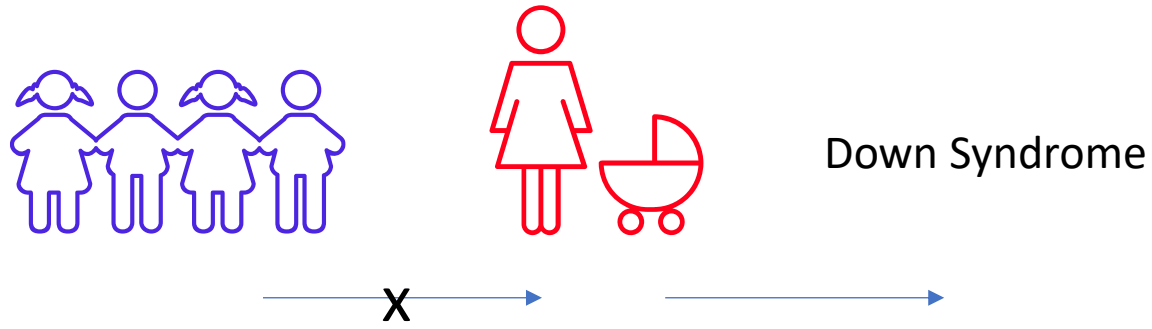
Maternal age confounds the relationship between birth order and Down Syndrome

Stark CR & Mantel N. Effects of maternal age and birth order on the risk of mongolism and leukemia. J Natl Cancer Inst. 1966; 37: 687-98.

A variable must meet three criteria to be a confounder

1. **Must be associated with the exposure**
    - Maternal age is associated with birth order
2. **Must be associated with the outcome**
    - Maternal age is a known risk factor for Down Syndrome
3. **Must not be on the causal pathway between exposure and outcome**
    - Birth order does not cause maternal age





Solving problems at different stages

- At the stage of design
  - Randomization in RCTs helps reduce confounding by observed and unobserved factors
  - Restriction
  - Matching
- At the analysis stage
  - Stratification
  - Adjustment – add potential confounders to statistical models to control their influence on outcome

Design stage

Randomization helps distribute confounders among experimental groups

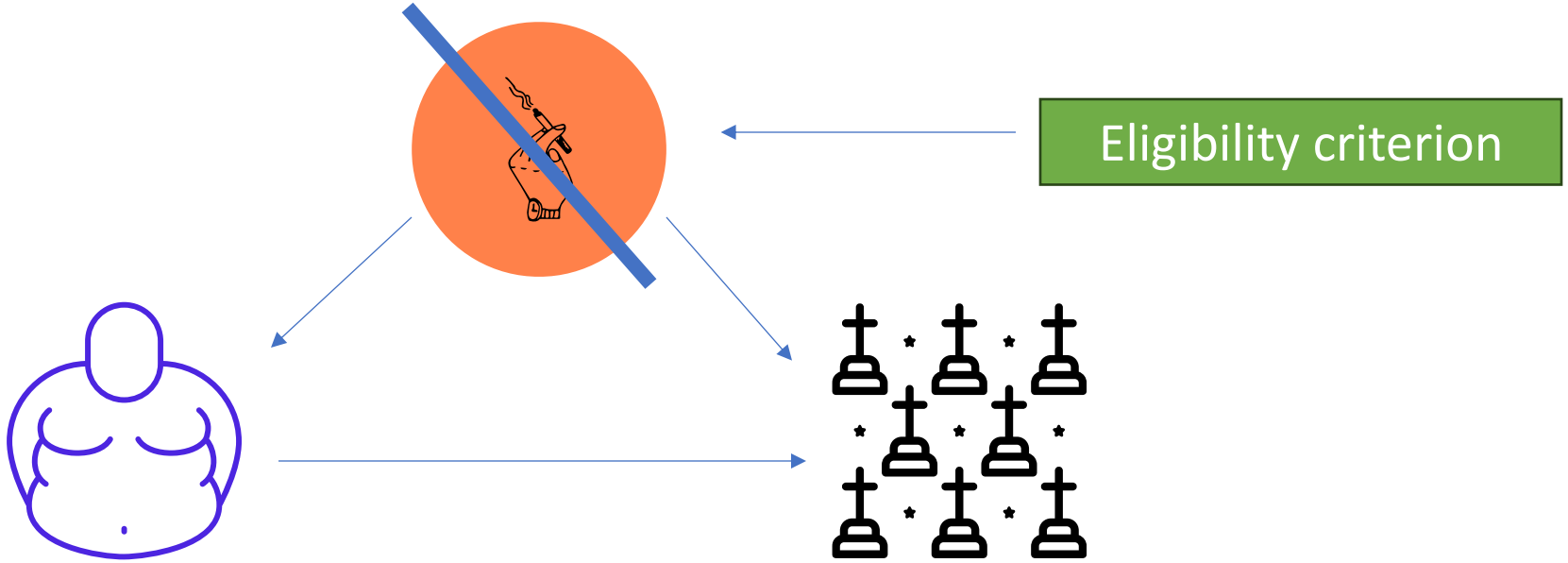
Condition A

Condition B

- If there are enough participants, we hope that randomization will **increase the likelihood that the groups will be comparable** on characteristics about which we may be concerned (such as sex, age, race, and severity of disease).

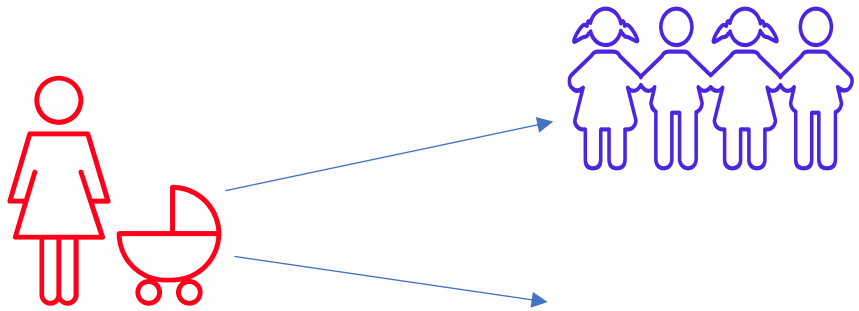
Restriction

- Restricting entry into the study to individuals who have the same value for a particular variable
    - E.g., Restricting study entry to non-smokers
    - E.g., Restricting study entry to women only
- Very effective method for preventing confounding in any type of study design, though has important implications for generalizability of results.



Matching

1. Select variables that could act as confounders
2. Create matched pairs or groups of participants (cases & controls) similar on those variables
3. Conduct study/analysis on pairs/groups



25-30	31-35	36-40	41-45

Analysis stage

First, detect if confounding is present

Crude effect estimate

Does not account for any confounding variable(s)

Adjusted effect estimate

Accounts for confounding variable(s)/potentially confounding variables

Empirical assessment of confounding:

Crude effect estimate ≠ Adjusted effect estimate

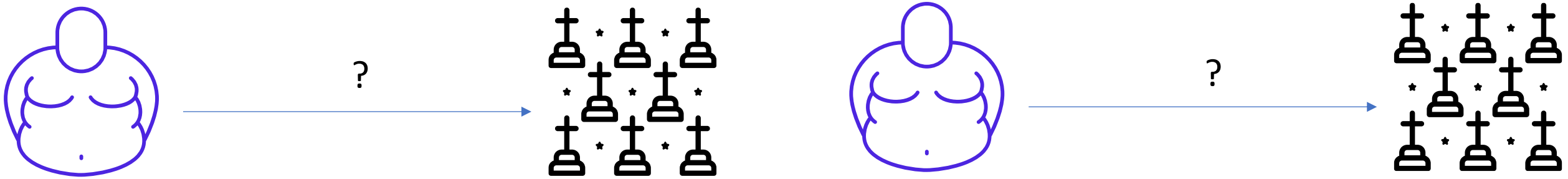
Stratification

The objective of stratified analysis is to set the level of the confounding variable and produce groups within which the confounder does not vary

Then, we evaluate the exposure-disease relationship within each stratum of the confounder

Yes

No



There are limits to stratification

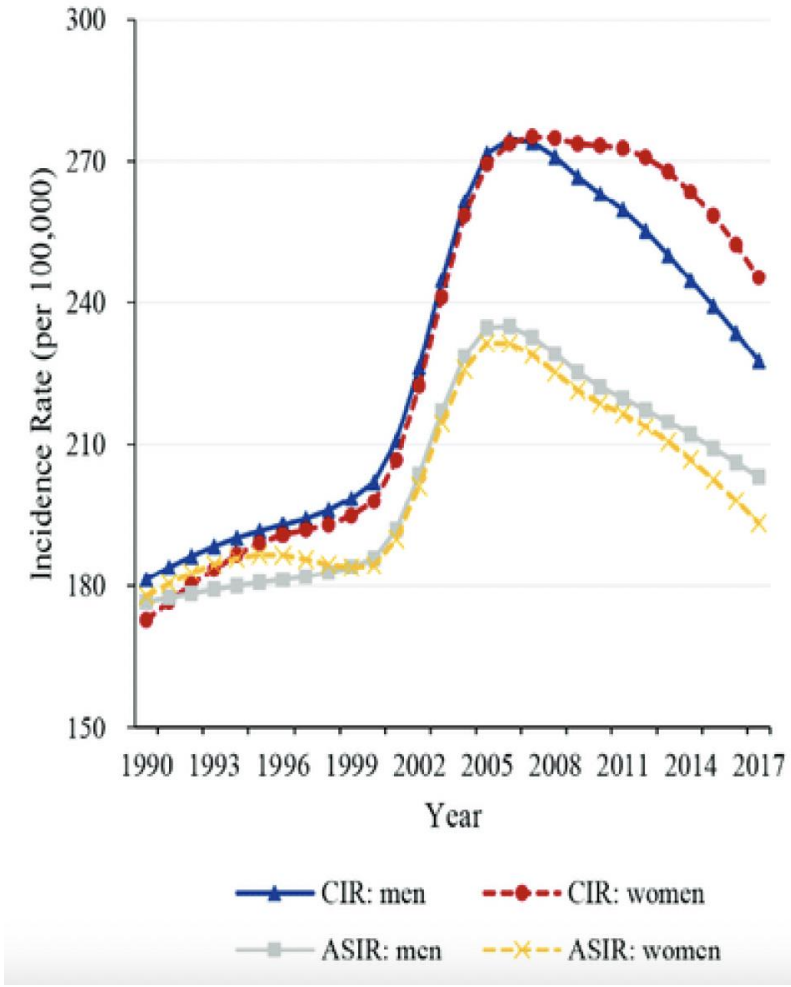
- Can only stratify on categorical variables
- Numerous strata can be problematic
  - Sparse data and imprecise estimates
- Impractical to adjust for multiple confounding variables
  - Controlling for age and gender, if gender is measured with 2 categories and age is measured with 5, end up with 10 strata

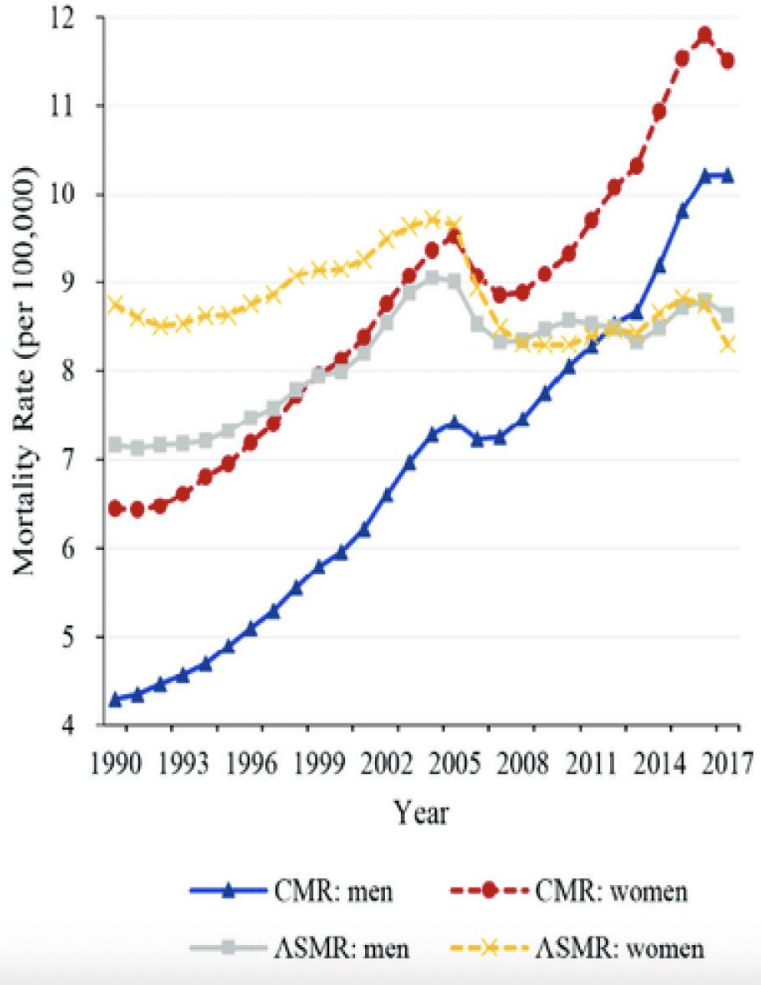
Standardization

- A statistical approach to remove confounding by a common characteristic
  - Age
  - Sex
  - Marital status
  - Education
- The most common standardization is carried out for mortality or disease incidence rates for age & sex
  - Over time
  - Across countries/geographical areas

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.)





What do you observe about crude and age-standardized rates of DM in China?

Another example—compare all-cause mortality between Sweden & Panama

	Sweden			Panama		
Age group	Number deaths	Population	Mortality rate / 1000pyrs	Number deaths	Population	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)

Another example—compare all-cause mortality between Sweden & Panama

	Sweden			Panama		
Age group	Number deaths	Population	Mortality rate / 1000pyrs	Number deaths	Population	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)

	Sweden			Panama		
Age group	Number deaths	Population	Mortality rate / 1000pyrs	Number deaths	Population	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

How can this be?

Another example—compare all-cause mortality between Sweden & Panama

	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)

	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

How can this be?

Sweden has an older population (17% vs. 5% of people older than 60 years) and mortality increases with age.

- If the number of potential confounders is large, multivariate analyses (regression analysis) offer the only real solution
- Can handle several confounders simultaneously
- Uses statistical regression models
- Always done with statistical software (SAS, Stata, R)

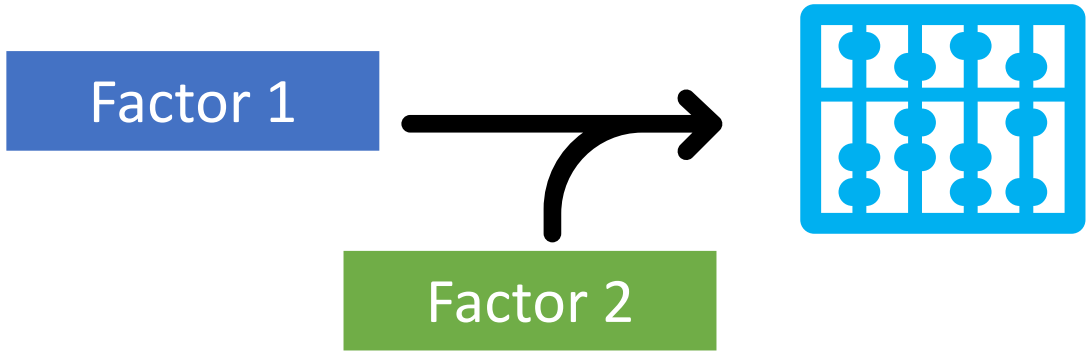
Residual confounding

- Unmeasured confounders or error in the measurement of observed confounders may lead to “residual” confounding
  - Confounding remains or is imperfectly accounted for
- Possibility of residual confounding cannot be completely eliminated in observational studies

Effect modification

When are we concerned with an interaction?

- When we have TWO exposures we are interested in and want to see if the joint effect of these two exposures on the outcome differs from the effect of either exposure independently
- E.g., Drinking and driving are independent causes for injury, but together they increase the risk more than either exposure independently



- Synergistic – effect of the two is more potent than either one alone
- Antagonistic – effect of one is diminished by the other

Identifying interactions – what is the combined effect of factors A & Z on outcome Y?

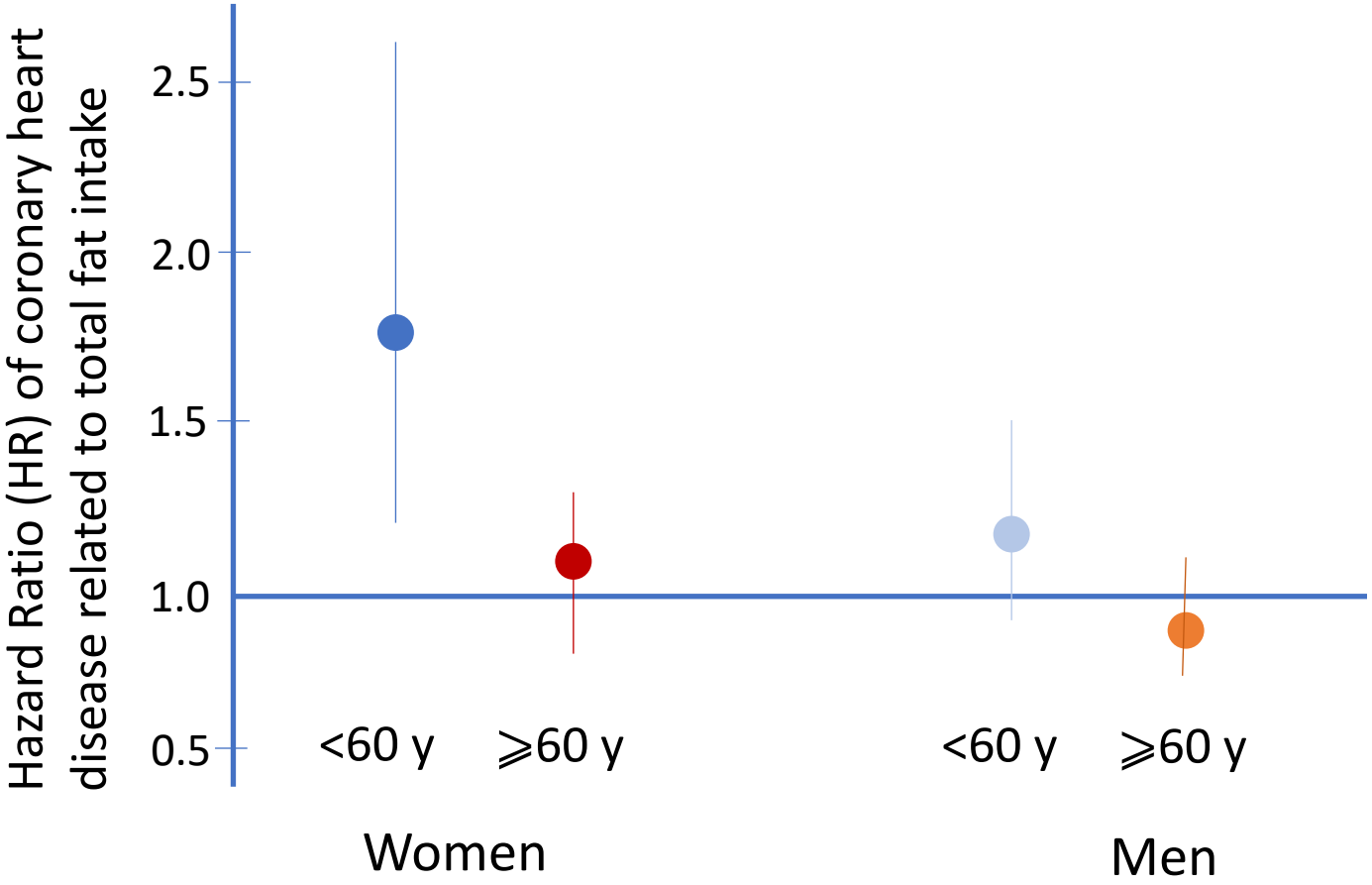
- Interaction = joint effect of two exposures
- When the rate of disease in the presence of two or more risk factors differs from the rate expected to result from their individual effects.
- **Positive interaction:** The effect of two risk factors combined is greater than what we would expect (also called synergism)
- **Negative interaction:** The effect of two risk factors combined is less than what we would expect from either risk factor independently (also called antagonism).

Effect measure modification (EMM) is a similar concept

- We are concerned with EMM when we have an exposure and outcome and wish to examine whether the relationship between the two differs by levels (strata) of a third variable
- Effect modification occurs when the effect of a risk factor (X) on an outcome (Y) differs in strata formed by a third variable (Z)
- Effect of exposure on disease is modified depending on the value of a third variable called an “effect modifier”
- Effect measure (e.g., risk difference, risk ratio) differs across different levels/strata of the third variable

Effect measure modification (EMM)

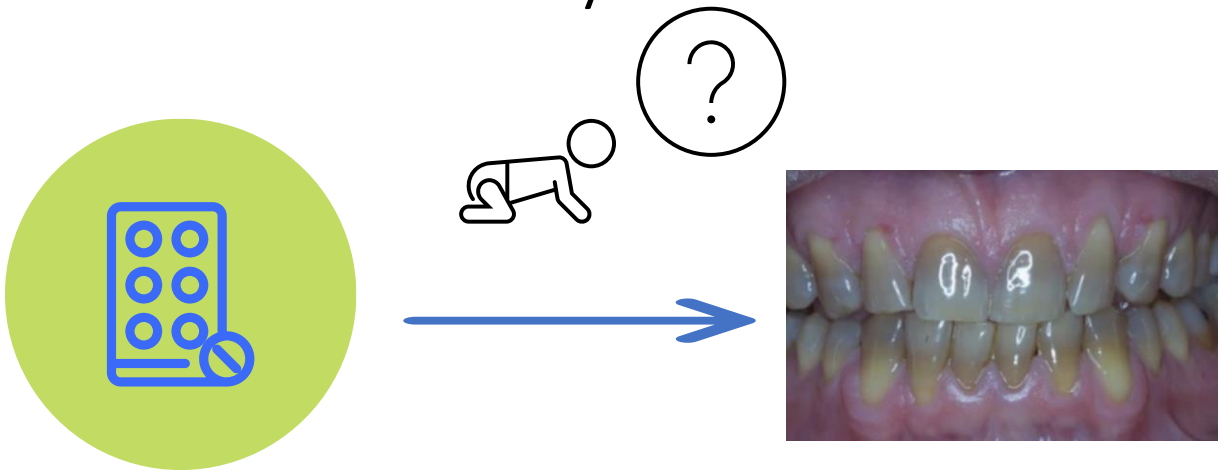
Example of EMM



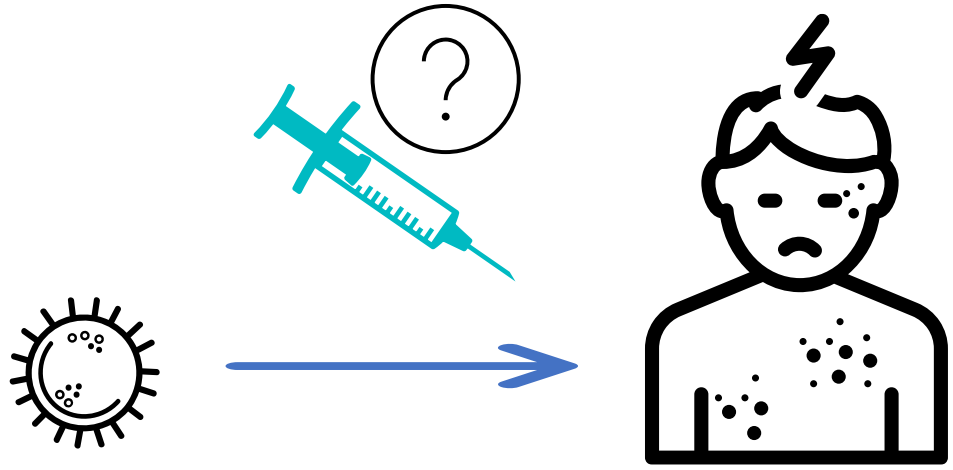
Jakobsen MU et al. Dietary fat and risk of coronary heart disease: possible effect modification by gender and age. *American Journal of Epidemiology* 2004; 160: 141-9.

Other examples of effect modification

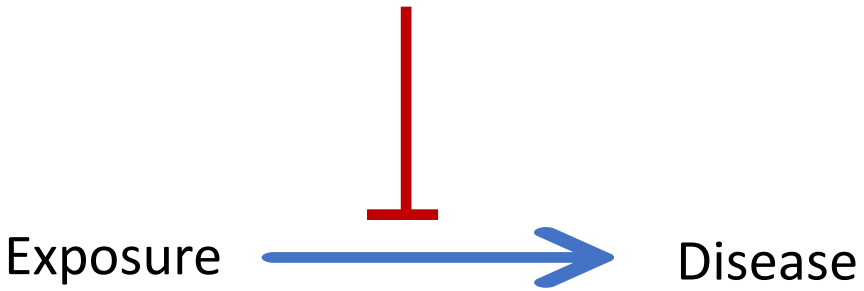
- Example 2
  - Exposure to the antibiotic tetracycline is related to the discoloration of teeth
  - This discoloration occurs when children up to 8 years of age take tetracycline
  - Discoloration is not observed when adults take tetracycline

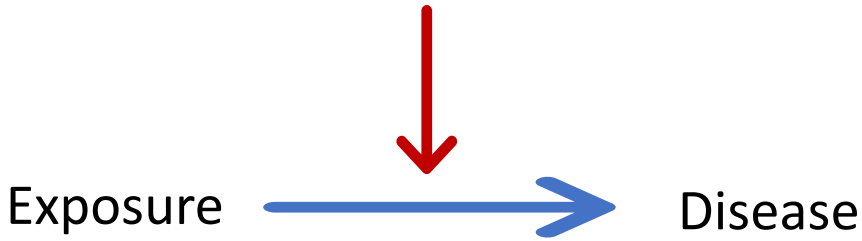


- Example 3
  - Individuals exposed to the measles virus will develop measles infection
    - Unless they have prior history of measles
    - Unless they have been vaccinated



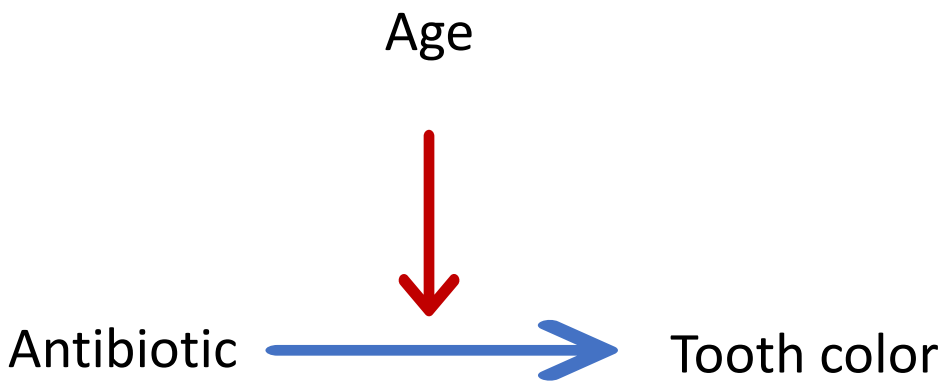
Notation in epidemiology to represent EMM



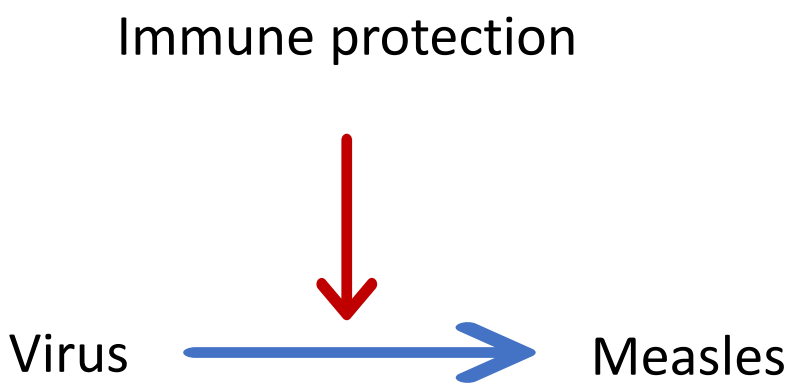


Going back to our examples

- Example 1
  - Age is the effect-modifier



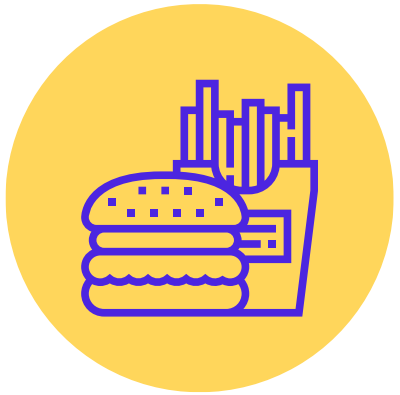
- Example 2
  - Immune status is the effect-modifier



Stratification aids in understanding interaction/ EMM

- Stratification is essential to understanding interaction and EMM
- Creating 2x2 tables (“cross-tabulating”) for the exposure-disease relationship by categories of another variable
- E.g., young/old, smokers/non-smokers

Yes	No



Yes	No

Yes	No

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

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

Reciprocal nature of effect modification

- For any given outcome and two predictor variables, it is a purely arbitrary decision which predictor variable will be the exposure, and which the potential effect modifier.

- Effect modification is reciprocal. In any of examples, the exposure and other factor (or variable) could have be labelled the other way round, and the same effect would still have been seen.

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

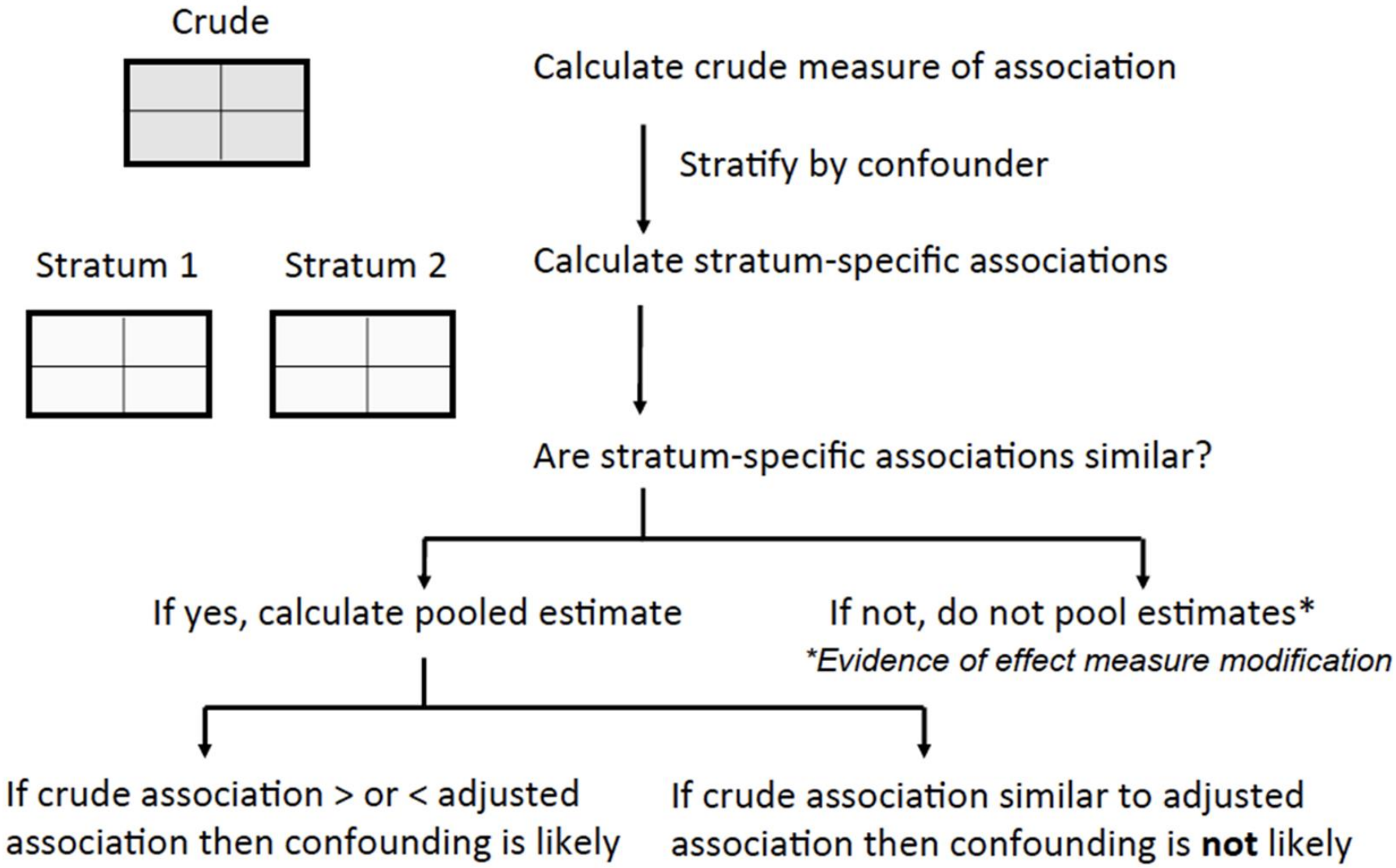
CHD, smoking and age in British doctors study (rates per 100,000)

	Non-smokers	*Heavy smokers*
	Rate	*Rate*
<45	7	104
45-54	118	393
55-64	531	1025
RR	75.9	9.9

How does interaction/EMM differ from confounding?

- Confounding
  - An alternative explanation for observed relationship
  - Distorts the “truth”
  - Epi attempts to remove it to get nearer to the truth
  - When it is present, stratum-specific effects are similar to each other but different from overall crude effect
- Interaction / EMM
  - One effect modifies the effect of another factor
  - It is genuine, not an artefact
  - Property of the relationship between factors
  - We should detect and describe it but not remove it.

Interaction vs. confounding



Let’s work through an example

Question: Does fat consumption modify the association between smoking and the risk of myocardial infarction (heart attack)?

Step 1: calculate crude measure of association

Smoking status	Heart Attack	No Heart Attack	Total
Smokers	42	158	200
Non-Smokers	21	175	196
Total	63	333	396

Calculate the Odds Ratio (OR), what is it?

Let's work through an example

Smoking status	Heart Attack	No Heart Attack	Total
Smokers	a=42	b=158	200
Non-Smokers	c=21	d=175	196
Total	63	333	396

$$OR = \frac{ad}{bc}$$

$$OR = 2.22\ [95\%\ CI: 1.26, 3.91]$$

Let's work through an example

Question: Does fat consumption modify the association between smoking and the risk of myocardial infarction (heart attack)?

Step 2: calculate associations within strata

1: Dietary fat intake <30% of calories

Smoking	+ Heart Attack	- Heart Attack	Total
Smokers	12	133	145
Non-Smokers	11	123	134
Total	23	256	279

$1.01\ [95\%\ CI: 0.43, 2.37]$

2: Dietary fat intake >30% of calories

Smoking	+ Heart Attack	- Heart Attack	Total
Smokers	30	25	55
Non-Smokers	10	52	62
Total	40	77	117

$6.29\ [95\%\ CI: 2.64, 14.75]$

What does this mean?

Crude OR = 2.22

Stratum specific ORs

- Dietary fat <30% = 1.01 (0.43, 2.37)
- Dietary fat >30% = 6.29 (2.64, 14.75)

Is there effect measure modification? Is there confounding?

https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC12.pdf

Heterogeneity vs. homogeneity of effects

When effect estimates are different in strata of the potential effect modifier → heterogeneity is present

Strata 1: OR=1.8

Strata 2: OR=5.7

When effect estimates are similar in strata of the potential effect modifier → homogenous effect estimates

Strata 1: OR=2.3

Strata 2: OR=2.5

What have we learnt today?

- Principle of confounding
- Principles of effect modification
- Step-by-step method of tidentifying confounding and effect modification by stratification
- Interpretation of results involving confounding and effect modification