Imperial College London

Confounding and effect modification

Marta Blangiardo

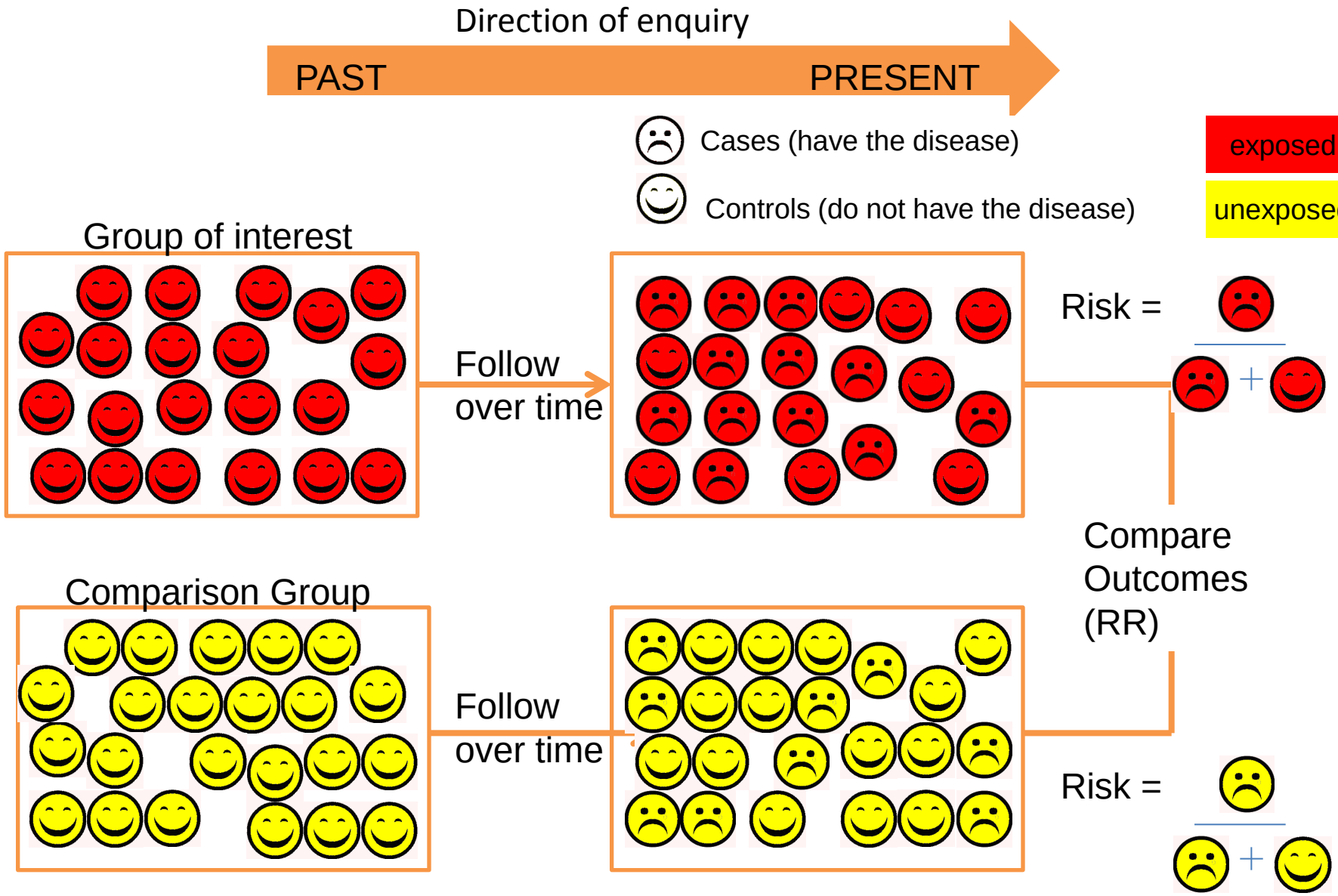
Lecturer in Biostatistics

Imperial College

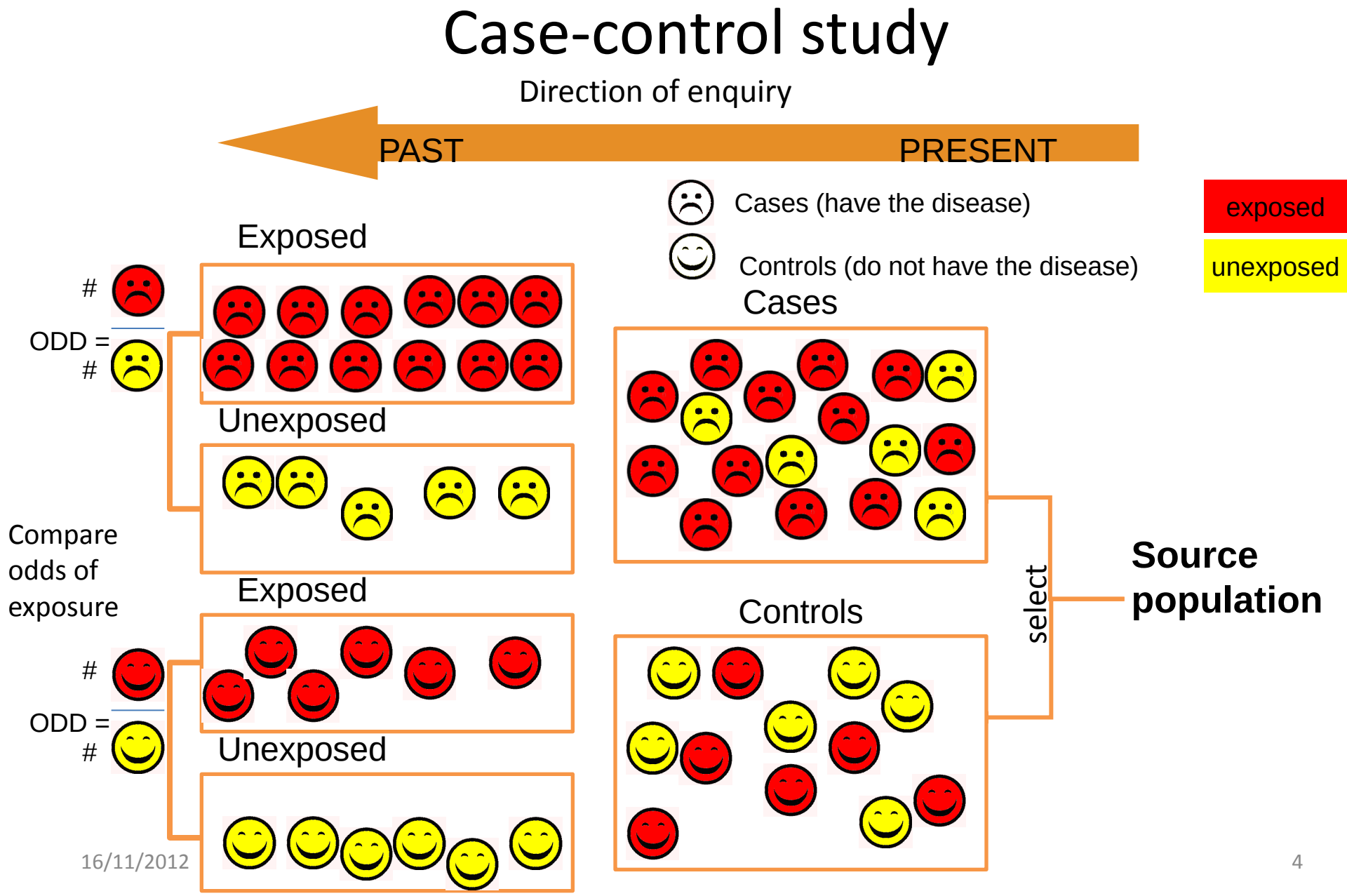
Learning objectives

- To understand the issues of confounding and effect modification in cohort studies and the techniques to deal with these
- To be able to assess the potential confounders in a cohort study and to correct for these in the analysis
- To be able to evaluate the presence of effect modifications and to carry on appropriate analyses to overcome this issue.

Cohort studies



Case-control study



Can we believe the results?

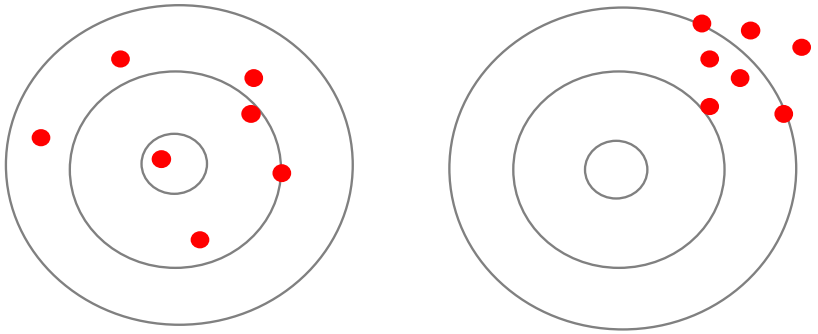
Obesity  Diabetes

RR=1.5

Systematic error

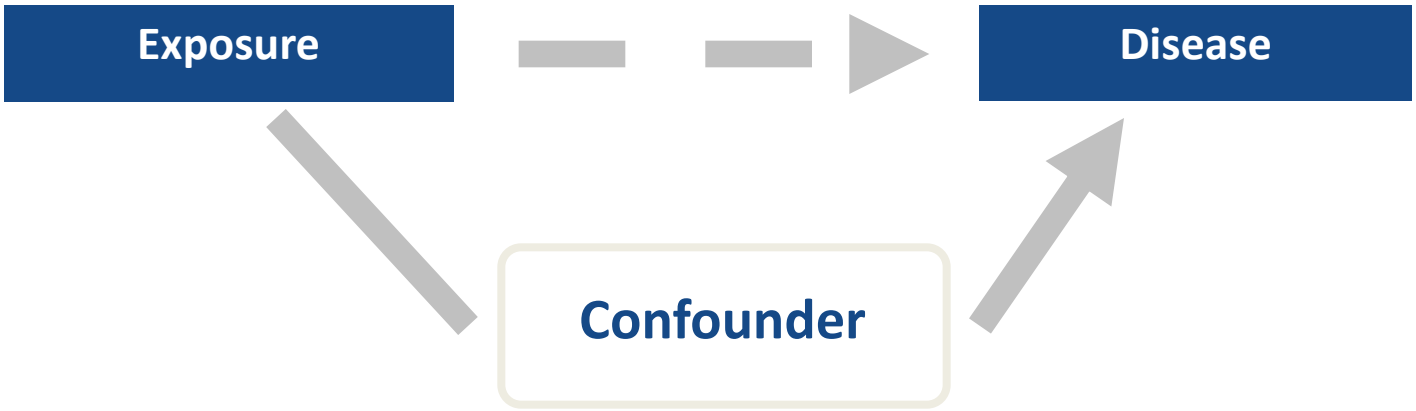
- Does not decrease with increasing sample size

- Selection bias
- Information bias
- Confounding



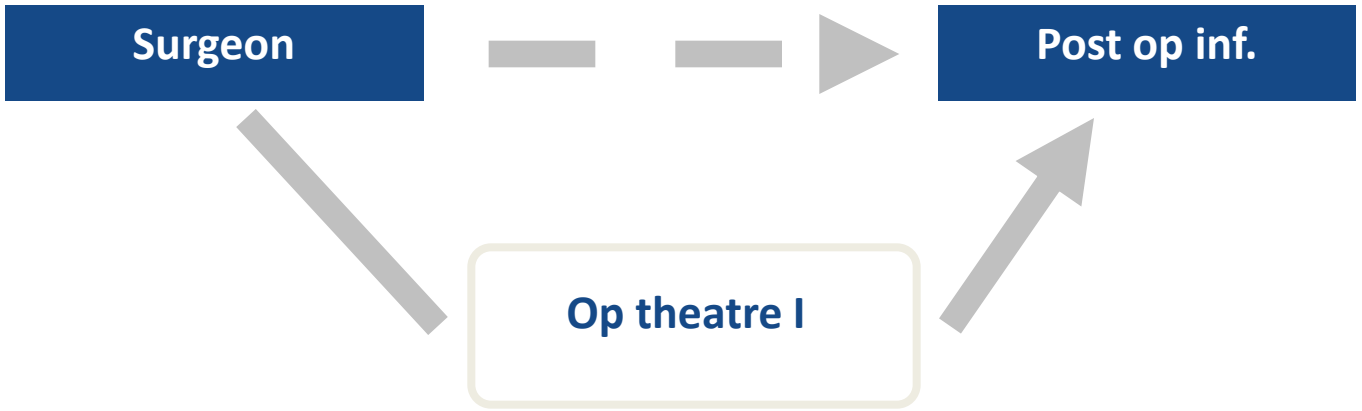
Confunding - 1

"Mixing of the effect of the exposure on disease with the effect of another factor that is associated with the exposure."



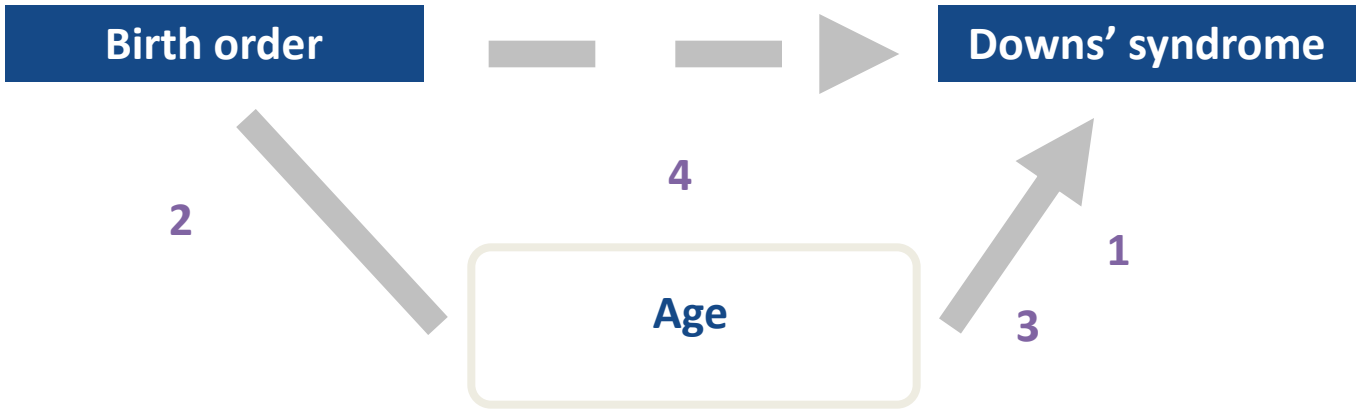
Confounding - 2

- Key term in epidemiology
- Most important explanation for associations
- Always look for confounding factors



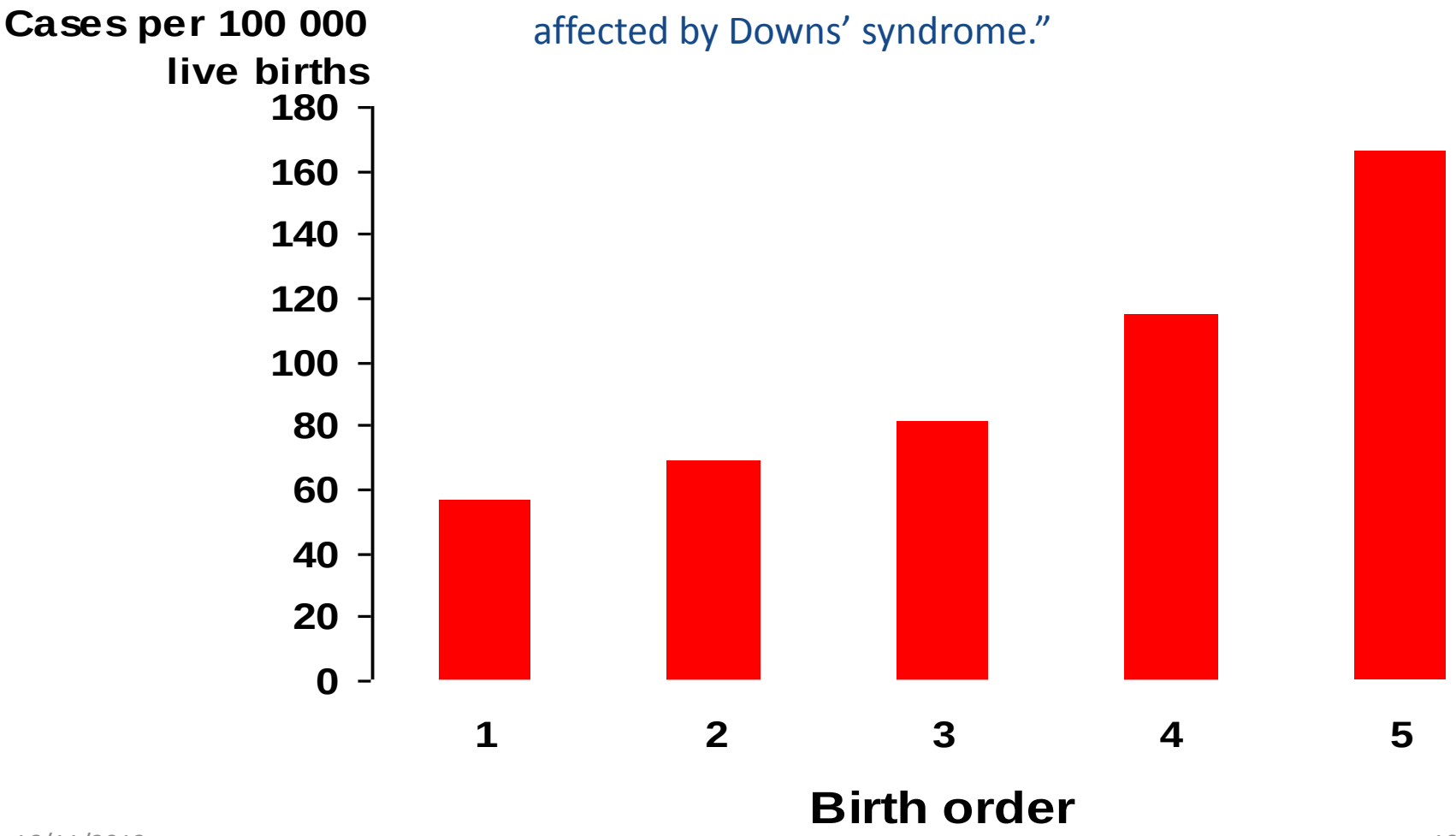
Criteria for a confounder

1 A confounder must be a cause of the disease (or a marker for a cause)
2 A confounder must be associated with the exposure in the source population
3 A confounder must not be affected by the disease
4 A confounder must not be on the causal pathway between exposure and disease

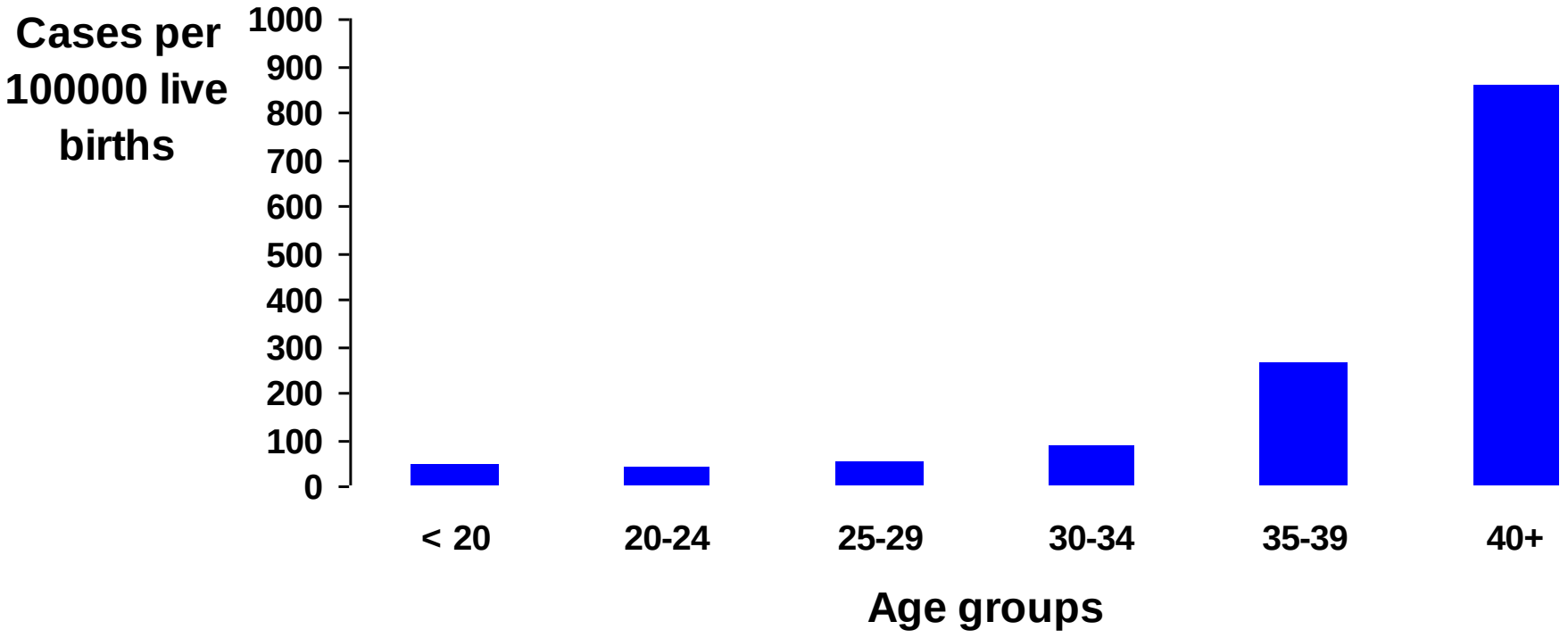


Downs' syndrome by birth order

"Second, third and fourth child are more often affected by Downs' syndrome."



Downs' syndrome by maternal age



What does a confounder do?

Suppose we have the following data:

- Exposure: carry matches
- Disease: Lung Cancer

Risk factor status	Disease status		
	Disease	No disease	***Risk***
Exposed	81	29	= 81 /(81+29) = 0.7364
Not exposed	28	182	= 28 /(28+182) = 0.1333
Relative risk (RR)			= 0.7364 /0.1333 = 5.52

- We might conclude that carrying matches increases highly the chances of getting lung cancer
- In fact, exposed people are 5.5 times more likely to get the disease

What does a confounder do? (2)

- Now imagine that we can observe a third variable (eg smoking status)

Risk factor status	Non Smokers			Smokers		
	Disease	No disease	***Risk***	Disease	No disease	***Risk***
Exposed	1	9	0.10	80	20	0.80
Not exposed	20	180	0.10	8	2	0.80
Relative risk			1.00			1.00

- If we control for smoking status, it appears that there is no effect at all (the risks are the same for exposed and non exposed in both groups)
- How is that paradox possible?
- NB This phenomenon is called "**Simpson's Paradox**"

What does a confounder do? (3)

- The explanation of the paradox is in the fact that smoking status (the confounder) is **associated with the risk factor**

Risk factor status	Non Smokers			Smokers		
	Disease	No disease	***Risk***	Disease	No disease	***Risk***
Exposed	1	9	0.10	80	20	0.80
Not exposed	20	180	0.10	8	2	0.80
Total	21	189	0.10	88	22	0.80

- $RR_{smoke\ /non\ smoke} = \frac{88/(88+22)}{21/(21+189)} = 8$

- The disease is 8 times more likely to be present among smokers than among non smokers!

Find confounders

"A study has found that small hospitals have lower rates of nosocomial infections than the large university hospitals. The local politicians use this as an argument for the higher quality of local hospitals."

WHAT COULD BE THE CONFOUNDER IN THIS SITUATION?
DISCUSS WITH YOUR NEIGHBOURS

Controlling confounding

In the design

- Restriction of the study
- Matching

Before data collection!

In the analysis

- Restriction of the analysis
- Stratification
- Multivariable regression

After data collection!

Restriction

Restriction of the study or the analysis to a subgroup that is homogenous for the possible confounder.

Always possible, but reduces the size of the study.

Stratified analysis

- Calculate crude odds ratio with whole data set
- Divide data set in strata for the potential confounding variable and analyse these separately
- If strata specific RR differs (> 10-20%) from overall RR, then confounding is present and data should be presented separately for the strata

Procedure for analysis

- When two (or more) exposures seem to be associated with disease

1. Choose one exposure which will be of interest
2. Stratify by the other variable
   - Meaning. Making one two by two table for those with and one for those without the other variable (for example, one table for men and one for women)

- Repeat the procedure, but change the variables

Multivariable regression

- Analyse the data in a statistical model that includes both the presumed cause and possible confounders
- Measure the relative risks for each of the exposures, independent from the others
- Logistic regression is the most common model in epidemiology
- But explore the data first with stratification!

Confounder?

Example:

In a study of the association between a fatty diet and CHD is "high cholesterol level" a confounder?

DISCUSS WITH YOUR NEIGHBOURS

Effect modification

- Definition: The association between exposure and disease differ in strata of the population
  - Example: Tetracycline discolours teeth in children, but not in adults
  - Example: Measles vaccine protects in children > 15 months, but not in children < 15 months

Effect modification

Note 1:

Effect modification is not a bias but a description of the effect.

Therefore, if effect modification is present our goal is to detect and to describe it.

Recall:

If confounding is present we try to eliminate it by stratification or multivariate statistical modeling.

Note 2:

Effect modification can be detected through

- Stratification
- Multivariate statistical modeling

Effect modification

What is the difference between confounding and effect modification?

Recall: Smoking is a confounder of the association between alcohol consumption and oral cancer.

Thus, the effect of alcohol consumption on oral cancer is mixed with the effect of smoking on oral cancer. To determine the true effect of alcohol consumption on oral cancer we must separate the two effects through stratification.

Note: The effect of alcohol consumption on oral cancer is the same for smokers and non-smokers, i.e. it does not depend on a person's smoking status.

Effect modification

Smoking is an effect modifier of the association between asbestos exposure and lung cancer.

Thus, the effect of asbestos exposure on lung cancer depends on smoking status.

The crude RR of asbestos exposure shows an average effect for smokers and non-smokers. This information is not very useful. However, if we stratify by smoking status and calculate the stratified RRs of asbestos exposure, we get information about the effect of asbestos exposure on lung cancer separately for smokers and for non-smoker.

The two stratified RRs are different from each other and from the crude RR

Example of detecting effect modification by stratification

Crude 2x2 table

	D	$\overline{D}$	Total
RF	34	48	82
$\overline{RF}$	110	91	201
Total	144	139	283

RR = 0.75

Oral contraceptive use seems to protect against Ovarian cancer.

Risk factor (RF) is oral contraceptive use; Effect Modifier (EM) age; and disease (D) ovarian cancer

Question: Since oral contraceptives have changed over the years, is the protective effect of oral contraceptives true for older and for younger women?

Stratification by age

2x2 table of RF and D for old women

	D	$\overline{D}$	Total
RF	4	23	27
$\overline{RF}$	98	74	172
Total	102	97	199

RR = 0.26 (protective effect)

2x2 table of RF and D for young women

	D	$\overline{D}$	Total
RF	30	25	55
$\overline{RF}$	12	17	29
Total	42	42	84

RR = 1.3 (increase risk)

Effect Modification (Interaction)

Note that the stratified relative risks are different from each other and from the crude RR. This indicates that effect modification is present and that the crude RR is not a complete (or vary useful) description of the effect.

The stratified analysis indicates that oral contraceptives only protect against ovarian cancer among older women, but slightly increase the risk of developing ovarian cancer among younger women.

Effect Modification (Interaction)

How do we know which variables to consider as potential confounders or effect modifiers?

- Clinical Knowledge
- Biological Knowledge
- Common sense

It is impossible to think of every possible confounder or effect modifier. Therefore, we will practically never be able to determine the "true" RR.

Imperial College London

What is what?

Assume we conducted a cohort study to explore the association between alcohol consumption and liver cancer.

We got the following results.

For the entire study population:

RR alcohol, liver cancer = 3.0

Among males:

RR alcohol, liver cancer = 1.5

Among females:

RR alcohol, liver cancer = 6.0

Assume we conducted a cohort study to explore the association between periodontitis and CVD.

We got the following results.

For the entire study population:

RR periodontitis , CVD = 1.56

Among smokers:

RR periodontitis , CVD = 1.0

Among non-smokers:

RR periodontitis , CVD = 1.0

What is the confounder and what is the effect modifier?
DISCUSS WITH YOUR NEIGHBOURS

Summary RR

We are interested in the association between a risk factor (RF) and a disease (D). We stratify by smoking status to determine whether smoking is a confounder or an effect modifier.

We find that the risk factor-disease RR's in the two strata are almost identical (i.e. 1.5 and 1.4), but are different from the crude RR (i.e. 0.59). We conclude that smoking is a confounder. Which RR should we report as the RR for the risk factor-disease association?

- Reporting the crude RR would mean reporting a biased result.
- Reporting the two stratum specific RR's would unnecessarily complicate the report.

Summary RR

We calculate a summary relative risks, i.e. a weighted average of the stratum specific relative risks.

$$\text{Then } RR_{MH} = \frac{\Sigma (c_k+d_k)/n_k * a_k}{\Sigma (a_k+b_k)/n_k * c_k}$$

Stratum **1**	D	$\overline{D}$
RF	a_1	b_1
$\overline{RF}$	c_1	d_1

n_1

.....

Stratum **k**	D	$\overline{D}$
RF	a_k	b_k
$\overline{RF}$	c_k	d_k

n_k

.....

Stratum **K**	D	$\overline{D}$
RF	a_K	b_K
$\overline{RF}$	c_K	d_K

n_K

Entire Population

RR = 1.17

	D	$\overline{D}$	Total
$\overline{RF}$	500	1000	1500
RF	400	1004	1404
Total	900	2004	2904

Stratum 1

	D	$\overline{D}$	Total
RF	210	120	330
$\overline{RF}$	240	284	524
Total	450	404	854

RR = 1.39

Stratum 2

	D	$\overline{D}$	Total
RF	290	880	1170
$\overline{RF}$	160	720	880
Total	450	1600	2050

RR = 1.36

The two stratum specific odds ratio are similar to each other and different from the crude relative risk. Thus, we can calculate a summary relative risk.

Imperial College London

Summary RR

Then $RR_{MH} = \frac{\Sigma (c_k+d_k)/n_k * a_k}{\Sigma (a_k+b_k)/n_k * c_k} = \frac{(c_1+d_1)/n_1 * a_1 + (c_2+d_2)/n_2 * a_2}{(a_1+b_1)/n_1 * c_1 + (a_2+b_2)/n_2 * c_2}$

$$= \frac{(240+284)/854*210 + (160+720)/2050*290}{(210+120)/854*240 + (290+880)/2050*160} = 1.38$$

	D	$\overline{D}$	Total
RF	210	120	330
$\overline{RF}$	240	284	524
Total	450	404	854

	D	$\overline{D}$	Total
RF	290	880	1170
$\overline{RF}$	160	720	880
Total	450	1600	2050

Note: When the stratum specific RRs are different from each other we conclude that effect modification is present. In this case we Do Not calculate a summary RR, but report the two stratified RRs.

Practical session

- Exercise on confounding

- Exercise on effect modification