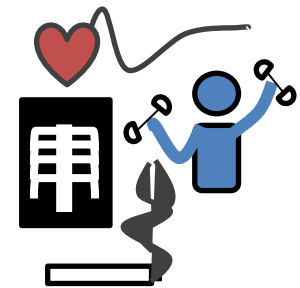# Bias and Confounding

Fpearson@sgul.ac.uk

Objectives

By the end of this session you should:

- Understand the concept of bias
- Be able to give examples of differing types of bias that may be introduced into a study
- Understand the concept of confounding
- Be able to give examples of confounding within studies
- Understand how bias and confounding may affect study results
- Be able to identify ways to control for confounding

Good research

All epidemiological studies should:

- Obtain a sample population that is as representative of the true population of interest as possible
- Aim to acquire the best information possible on parameters of interest amongst their sample population

Causality

- Causality is the central concern of epidemiology

| Obesity | → | CHD |

OR, RR, IRR

Bias

"Systematic, non-random deviation of results and inferences from the truth, or processes leading to such deviation."

(Dictionary of Epidemiology, 3rd ed.)

Bias (including confounding) must be avoided or reduced as much as is possible

– Help to produce accurate, precise, valid and reliable measurements

– Ensure inferences made are true and generalisable

Selection bias

Systematic differences exist between comparison groups in a sample population

Or, when the sample population does not represent the target population
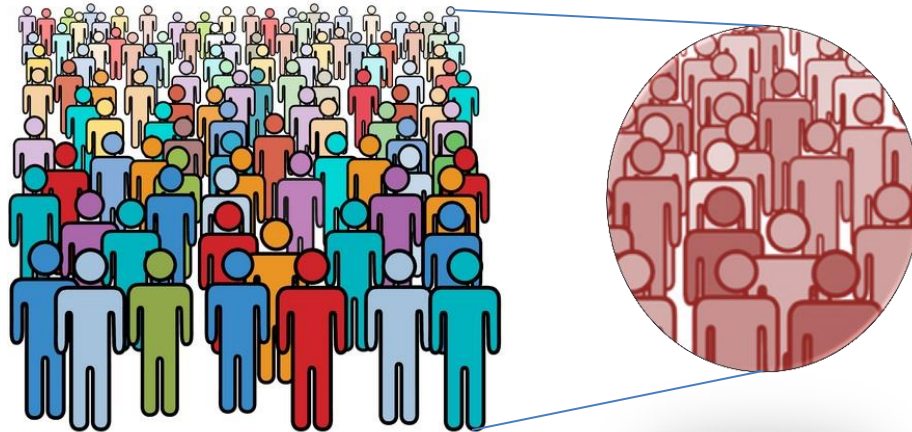
Examples of selection bias

Respondent bias those who agree to be in a study are in some way different from those who refuse so the sample population is not truly representative of the target population

Healthy worker effect employed population have lower mortality and less disease when compared with the general population

Ascertainment bias the cases gathered are not representative of cases in the target population e.g.

- *Healthcare-access bias* patients admitted to an institution are not representative of cases in target population
- *Length-biased* sampling cases with disease of long duration. May not represent the cases originated in the target population, usually have a better prognosis
- *Survivor bias* patients who live longer are more likely to receive treatment. Analysis indicate a false association between that treatment and survival

Examples of selection bias

Selective survival bias *(neyman bias,* incidence-prevalence bias) when 'survivors' are selected if the exposure is related to, or is, a prognostic factors. The sample of cases give a distorted exposure frequency. Bias only occurs if exposure influences mortality due to outcome

- For example in a study of the relationship between smoking and AMI, cases were interviewed 1 week after their AMI. As smokers with AMI die disproportionately, the cases left showed a lower frequency of smoking, and attenuated the association to the null

Spectrum bias when only "definite" cases (not representative of a full "syndrome") are included, and/or only definite healthy controls. Misrepresents conditions in which a differential diagnosis is made increasing diagnostic test sensitivity and specificity

Selection bias – choosing controls

Hospital controls

- ***Berksonian bias*** associations between diseases or between a characteristic and a disease due to differing admission probabilities for those with the disease, without the disease and with the characteristic of interest

- ***Inclusion bias*** when one or more conditions in controls are related with the exposure. Exposure frequency is higher than expected, producing a bias towards the null

Exclusion bias controls with conditions related to an exposure are excluded, whereas cases with the disease as a comorbidity aren't

- In a study of breast cancer controls with CVD (related to use of reserpine) were excluded but cases with CVD were not producing a spurious association between reserpine and breast cancer

Selection bias - controls

Relative/Friend control bias correlation in exposure status between cases and controls leads to biased estimates of an association. However, matching removes bias if the exposure induced risks of disease are constant over time

Matching individual & frequency matching can introduce bias. Can be controlled for in analysis (matched analysis for individual matching and adjustment for variables matched on in frequency matching). However, overmatching is produced when researchers match by a non-confounding variable (associated to the exposure but not to the disease) and can underestimate an association

Information bias

Also know as observational bias or misclassification. Systematic differences in measurement of exposure, covariate, or outcome variables that results in different quality data between groups
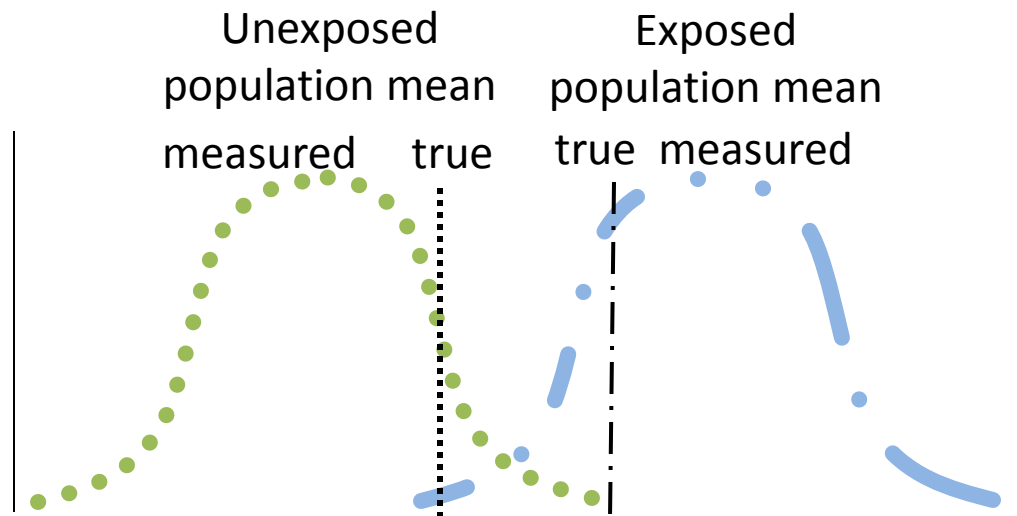
Types of information bias

Interviewer Bias: an interviewer's knowledge may influence the structure of questions and the manner of presentation, which may influence responses

Recall Bias: those with a particular outcome or exposure may remember events more clearly or amplify their recollections

Observer Bias: observers may have preconceived expectations of what they should find in an examination

Loss to follow-up: those that are lost to follow-up or who withdraw from the study may be different from those who are followed for the entire study

Hawthorne effect: an effect first documented at a Hawthorne manufacturing plant; people act differently if they know they are being watched

Types of information bias

Surveillance bias The group with the known exposure or outcome may be followed more closely or for longer than the comparison group
Misclassification bias
Differential
Overestimate exposure for n cases, inflate rates
Underestimate exposure for n cases, deflate rates
Underestimate exposure for n controls, inflate rates
Overestimate exposure for n controls, deflate rates
Non-differential (random) Errors in assignment of group happens in more than one direction. This will bias study findings toward the null.

Bias in RCT's

- *Allocation of intervention*
- *Compliance bias*
- *Contamination bias*
- *Differential maturing*
- *Lack of intention to treat analysis*

Controls for Bias

- Careful study design to minimize chance for bias
- Careful interpretation of results discussing possible bias openly
- Define, a priori, who is a case/control or what constitutes exposure/outcome
- Set up guidelines for data collection using best possible methods
- If possible adjust analyses and use sensitivity analyses
- May be plausible to estimate study findings adjusting for error. Or, to conduct sensitivity analysis to test range of findings consistent with measurement problems
- Don't over interpret results
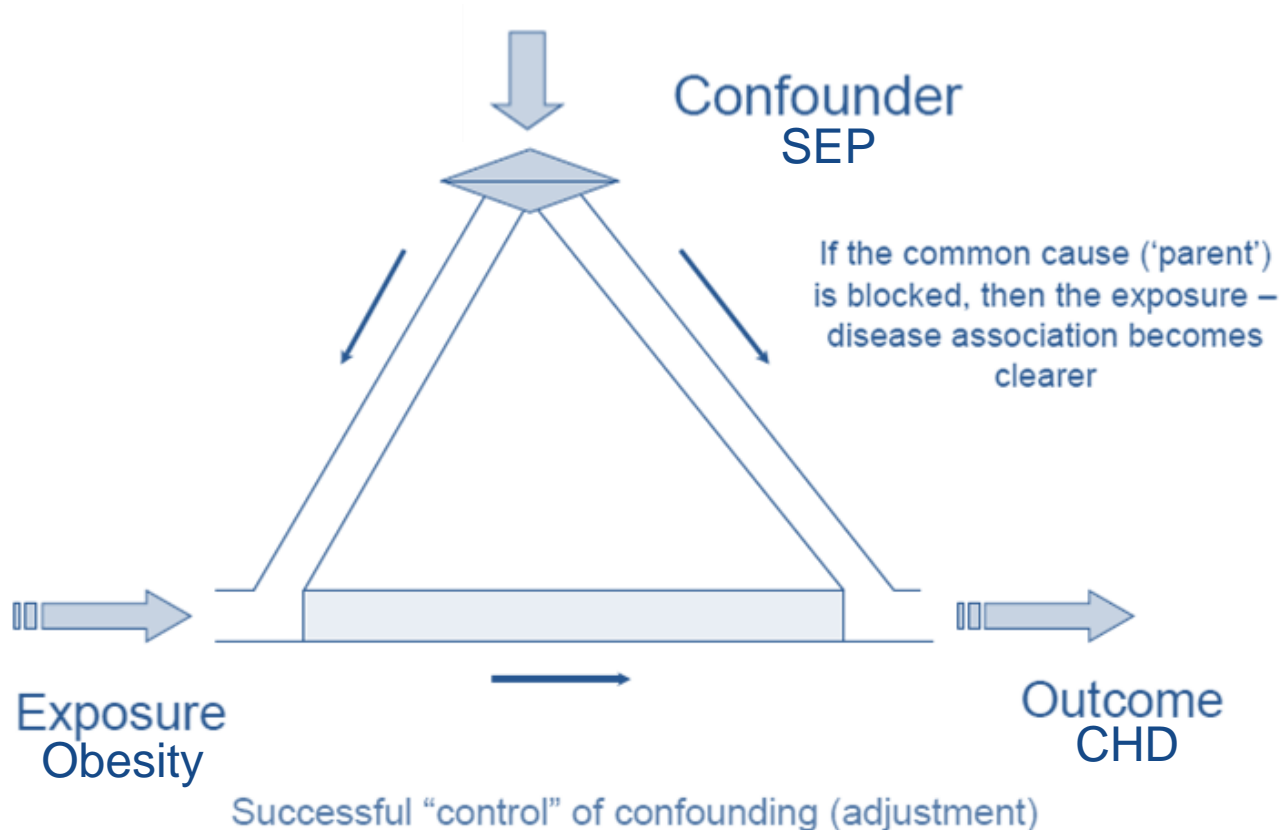
Confounding (a type of bias)

Confounding
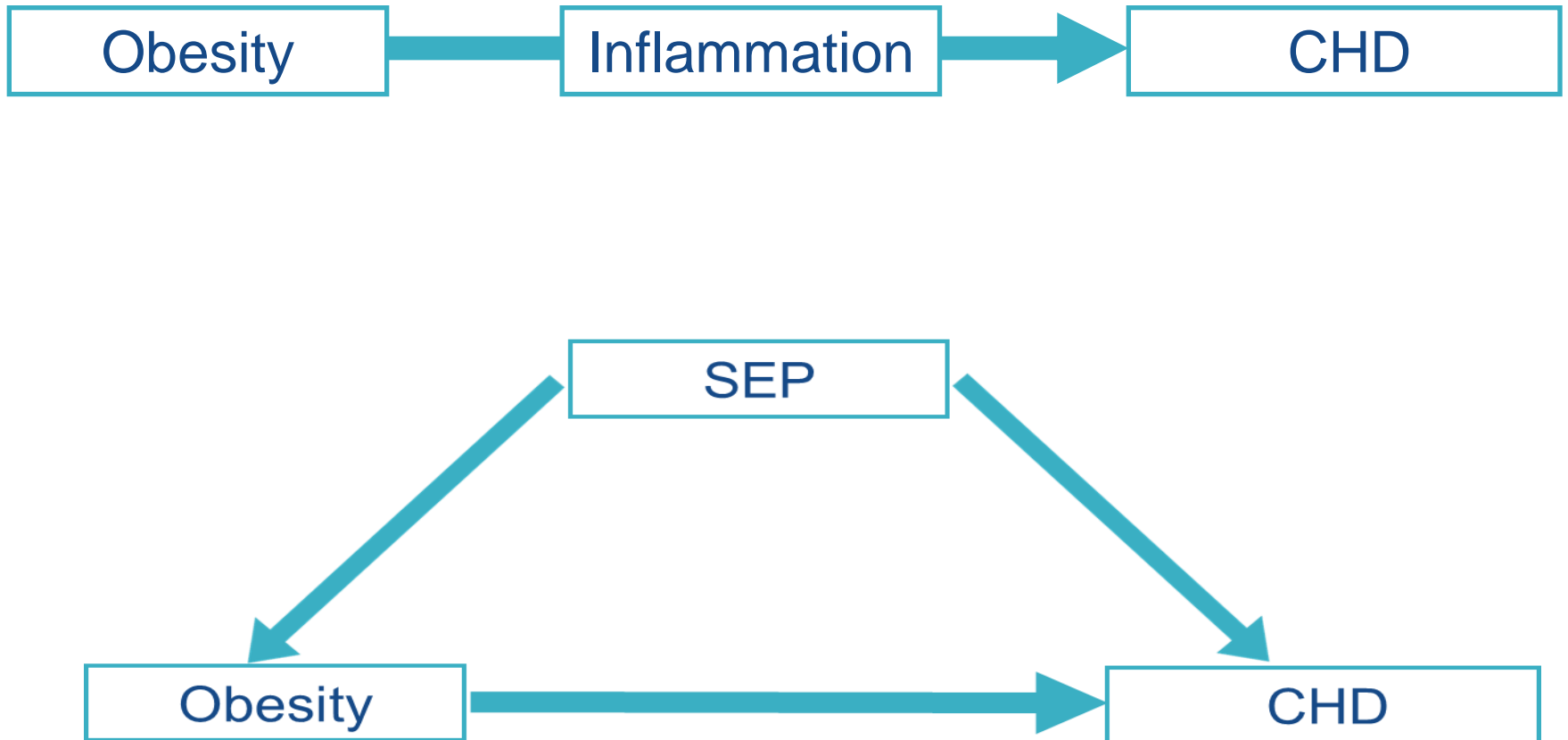
Mixing of effects

"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

Confounder
SEP

If the common cause ('parent') is blocked, then the exposure – disease association becomes clearer

Exposure
Obesity

Outcome
CHD

Successful "control" of confounding (adjustment)

Types of confounding

Positive confounding

When the confounding effect results in an <u>overestimation</u> of the effect (i.e., the crude estimate is further away from 1.0 than it would be if confounding were not present).

Negative confounding

When the confounding effect results in an <u>underestimation</u> of the effect (i.e., the crude estimate is closer to 1.0 than it would be if confounding were not present).

Confounder?
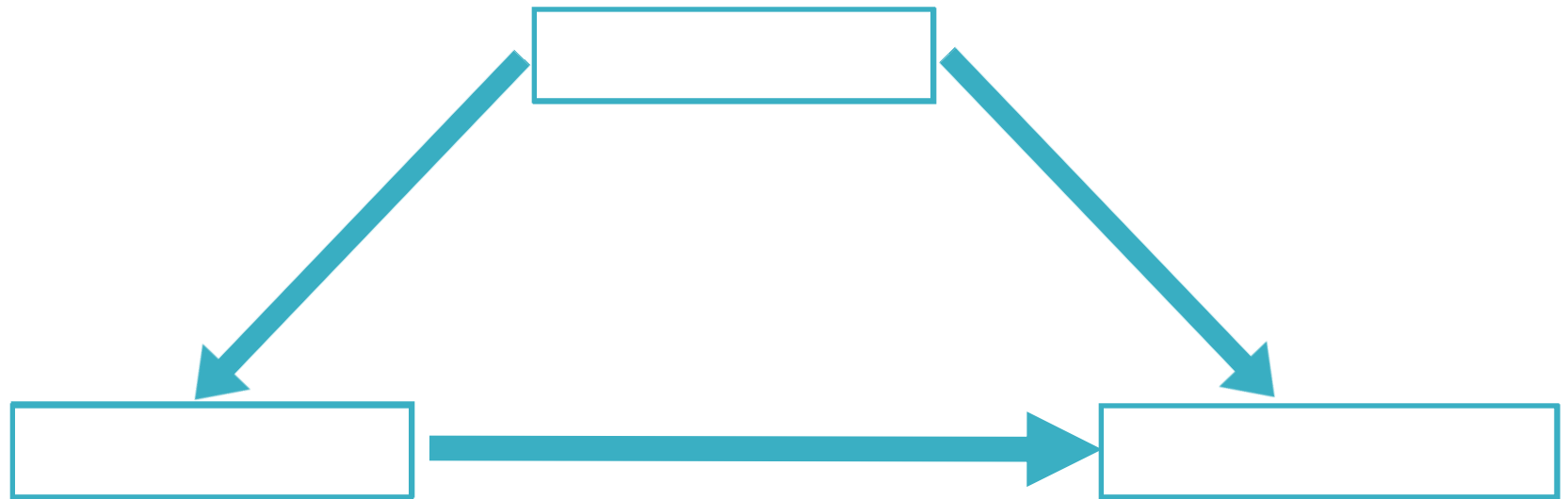
Examples?

Example of negative confounding

An occupational study in which workers exposed to a certain carcinogen are younger than those not exposed.

If the risk of cancer increases with age, the crude association between exposure and cancer will underestimate the unconfounded (adjusted) association.

Age: negative confounder.

Further points on confounding

Confounding is not an all or nothing phenomenon
  A confounding variable may explain the whole or just part of the observed association between a given exposure and a given outcome.
        Crude OR=3.0 … Adjusted OR=1.0
        Crude OR=3.0 … Adjusted OR=2.0
Residual confounding
  Controlling for one of several confounding variables does not guarantee that confounding is completely removed. Residual confounding may be present when:
  - the variable that is controlled for is an <u>imperfect</u> surrogate of the true confounder,
  - other confounders are ignored,
  - the units of the variable used for adjustment/stratification are too broad
The confounding variable may reflect a "constellation" of variables or characteristics
  Occupation (SES, physical activity, exposure to environmental risk factors)
  Healthy life style (diet, physical activity)

Ways to control for confounding

During the design phase of the study:
- Randomized trial
- Cross-over design
- Matching
- Restriction

During the analysis phase of the study:
- Standardization
- Stratification
- Adjustment (direct, indirect, M-H)
- Regression

Assessing measured confounding

Check that the variable meets the criteria of a confounder

If the effect of that variable (on exposure and outcome) is controlled for (e.g., by stratification or adjustment) does the effect measure change?
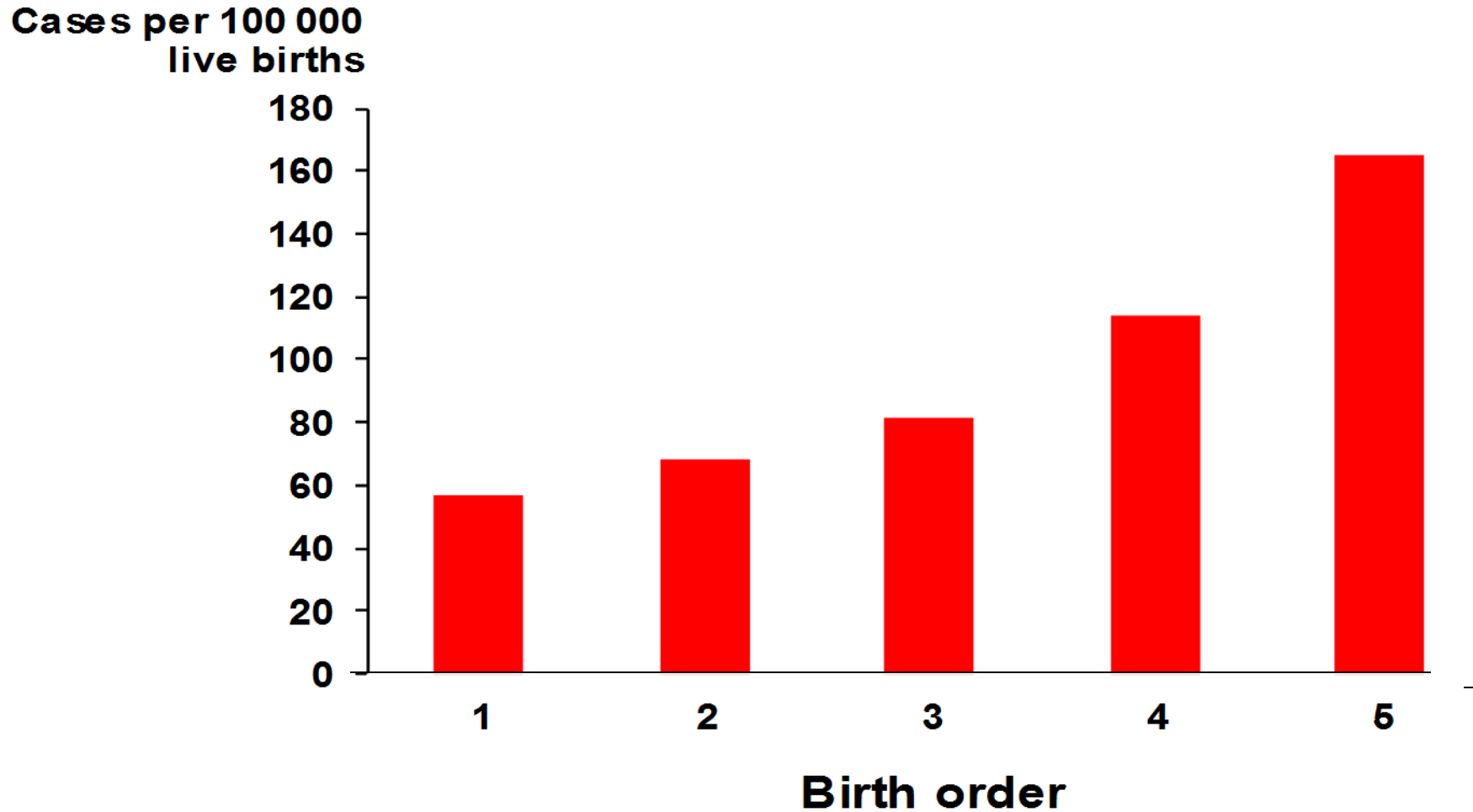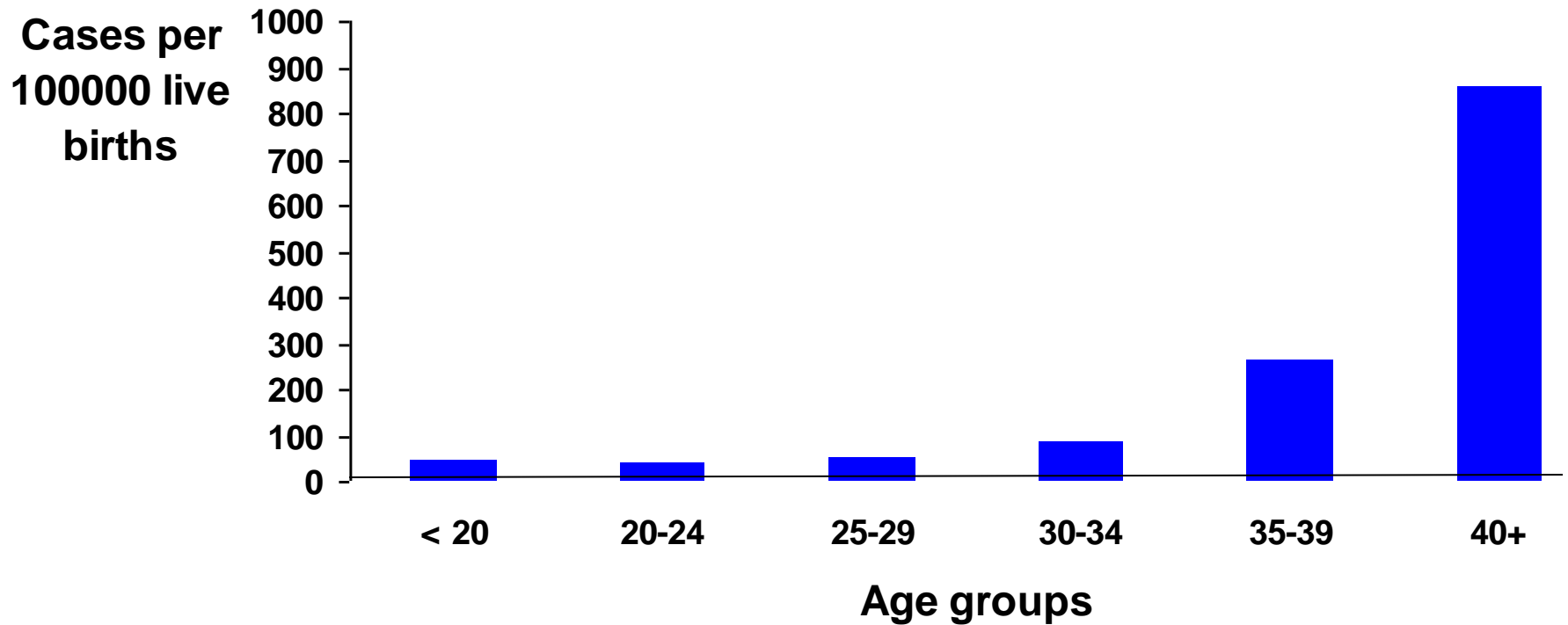
Asessing measured confounding

"Second, third and fourth child are more often affected by Downs' syndrome."
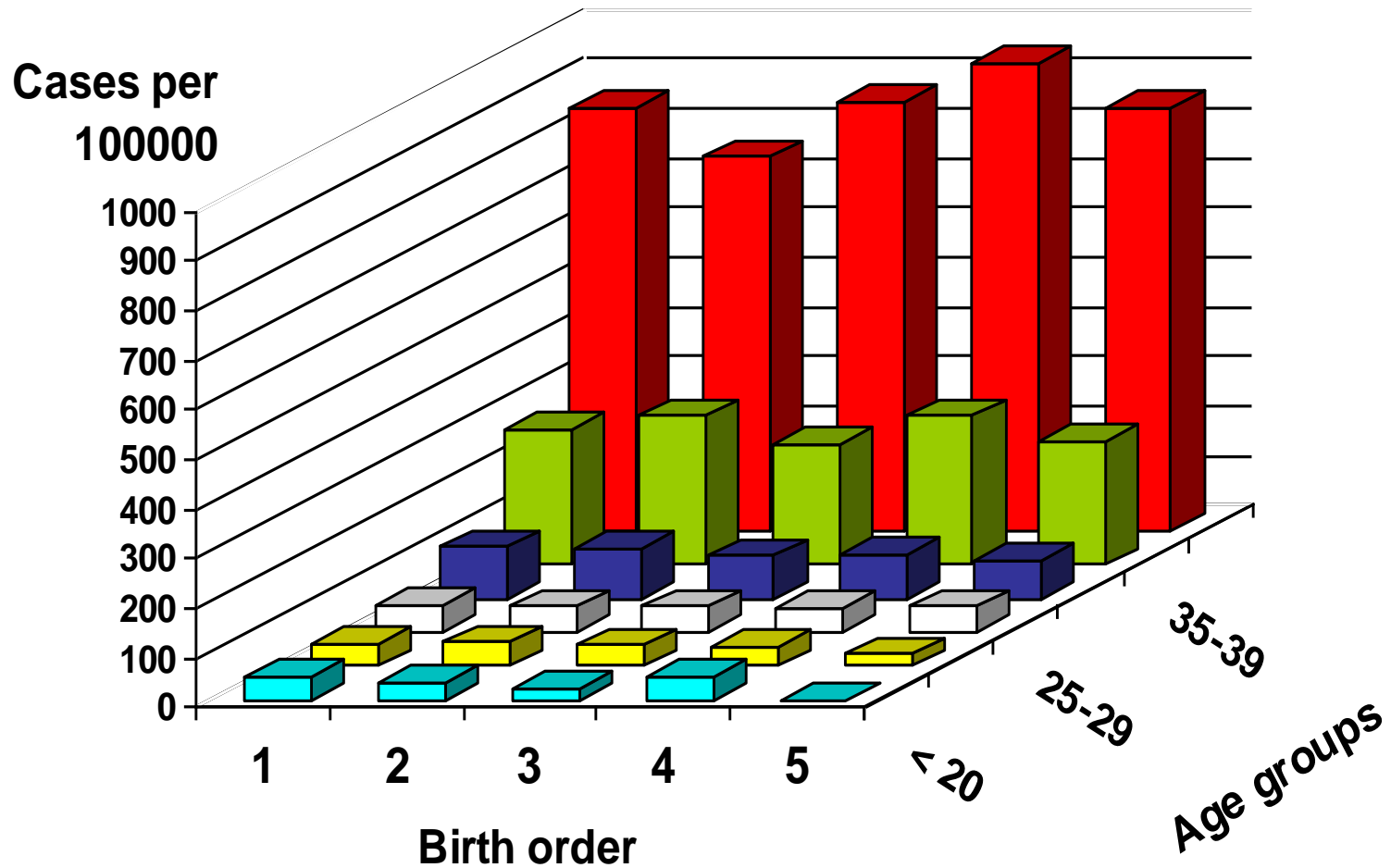
| Many Children | ➡ | Down Syndrome |

Downs' syndrome by birth order
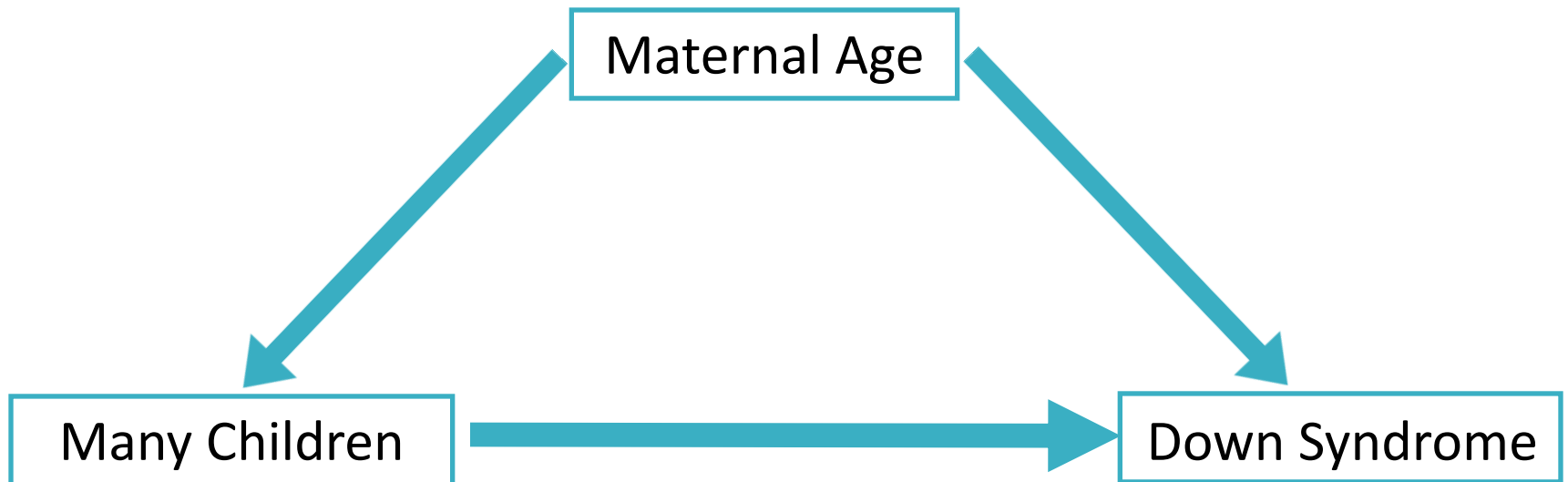
Downs' syndrome by maternal age

Downs' syndrome by birth order and maternal age groups

Plausible confounder

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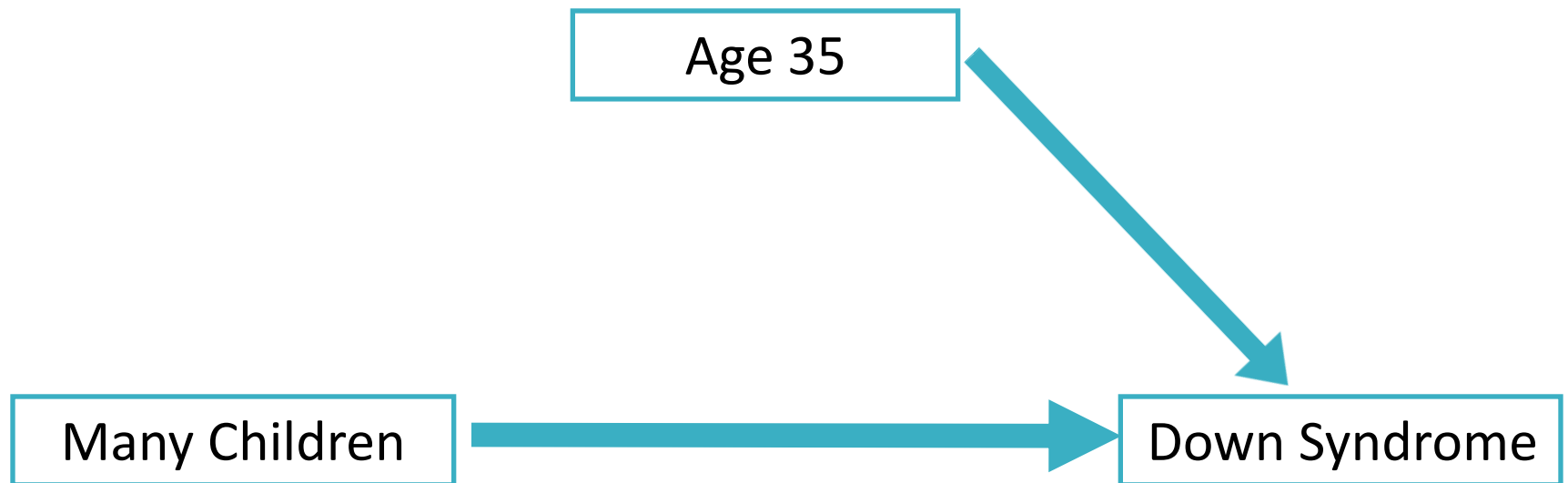
Restriction

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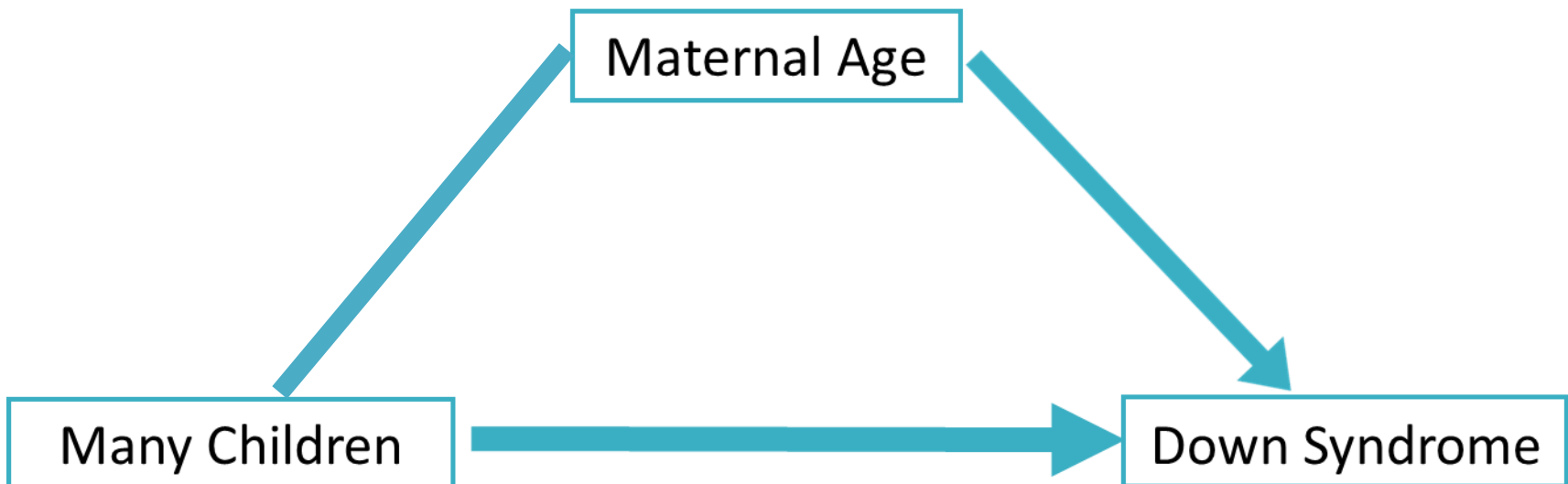
Will dramatically reduce the size of the study

In this example assess only mothers above a certain age

Age 35

Many Children → Down Syndrome

Matching

Select controls identical to the cases with respect to distribution of one or more potential confounders.

Disadvantages of matching

Control group are no longer representative of the target population

- More complex "matched" analysis needed

Cannot study whether any matched factors have a causal effect

Pragmatically becomes more difficult to find controls

Adavntages of matching

Taking a random sample from the target population is not always possible

Pragmatically can be a quick, easy and cost-effective way to select controls

- Matched on "social factors": Friend controls, family controls, neighbourhood controls
- Matched on time: Density case-control studies

Can improve statistical efficiency of study

Can control for confounding due to factors that are difficult to measure

Stratified analysis

Calculate crude odds ratio with whole data set

Divide data set in strata for the potential confounding variable and analyse these separately

Calculate adjusted (OR_{mh}) odds ratio

If adjusted OR differs *(> 10-20%)* from crude OR, then confounding is present and adjusted OR should be reported

Stratified analysis

Crude

	Lung Ca	No Lung Ca
Matches	820	340
No Matches	180	660

$OR_{crude} = 8.8$

Stratified   Smokers   Non-Smokers

	Lung Ca	No Lung CA
Matches	810	270
No Matches	90	30

	Lung Ca	No Lung CA
Matches	10	70
No Matches	90	630

$OR_{CF+} = OR_{smokers} = 1.0$

$OR_{CF-} = OR_{non-smokers} = 1.0$

Multivariable regression

Analyse the data in a statistical model including

Both the presumed cause and possible confounders

Just the presumed cause

This will measure the effect measure for each of the exposures, independent from the others

Use Wald test and LRT test to assess model fit

But explore the data first with stratification!