

Confounding!

Outline of presentation

1. What is confounding

2. Criteria for confounding

3. How to identify confounding

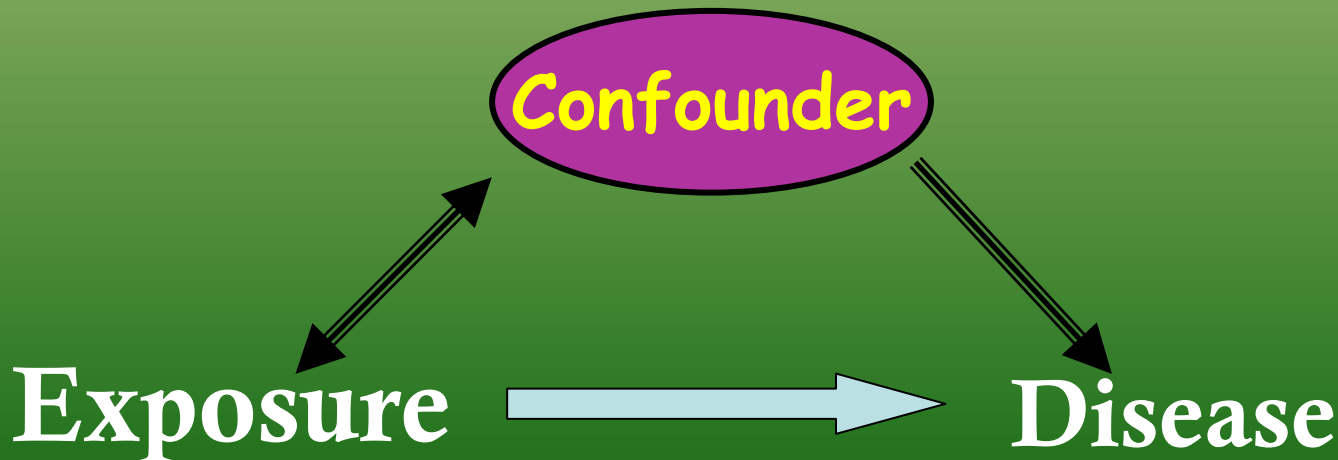
4. How to control confounding

5. Important Tips

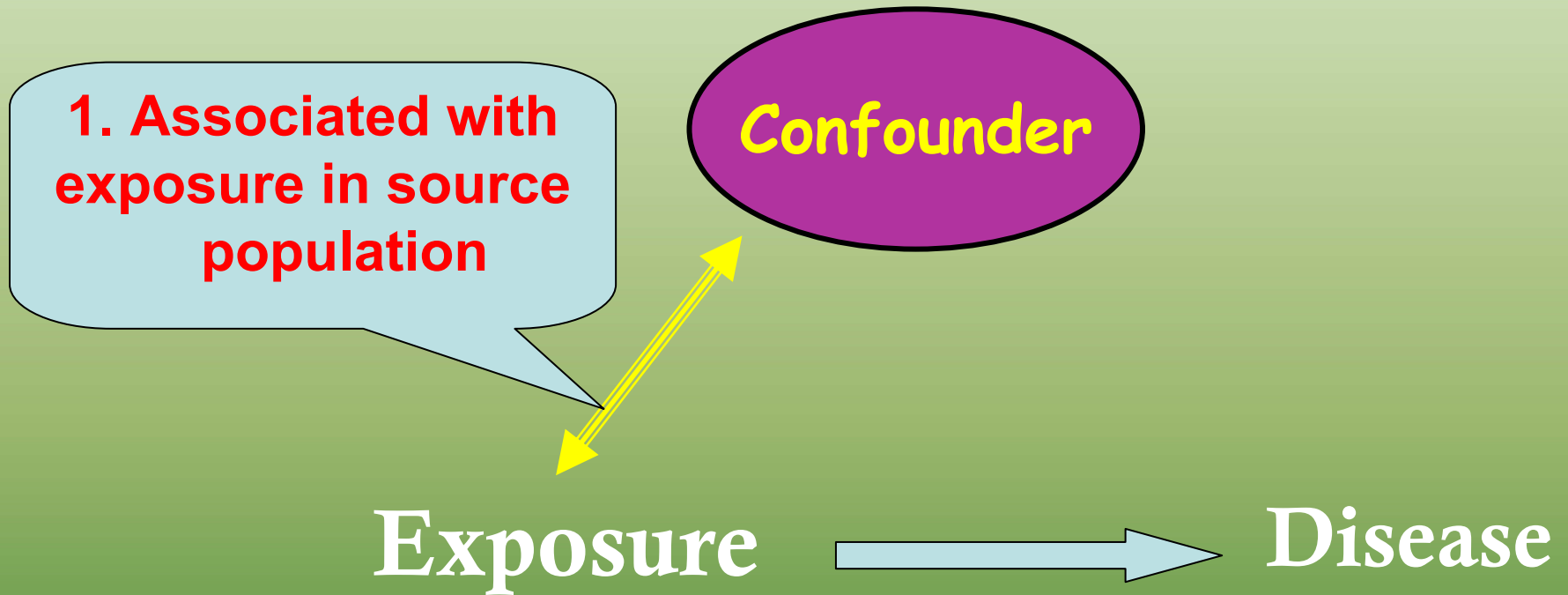
What is confounding?

Latin word - *Confundere* –meaning *mixing together*

It means a mixing of effect of an extraneous factor (**confounder**) with effect of interest on outcome –distorting the association between exposure and disease

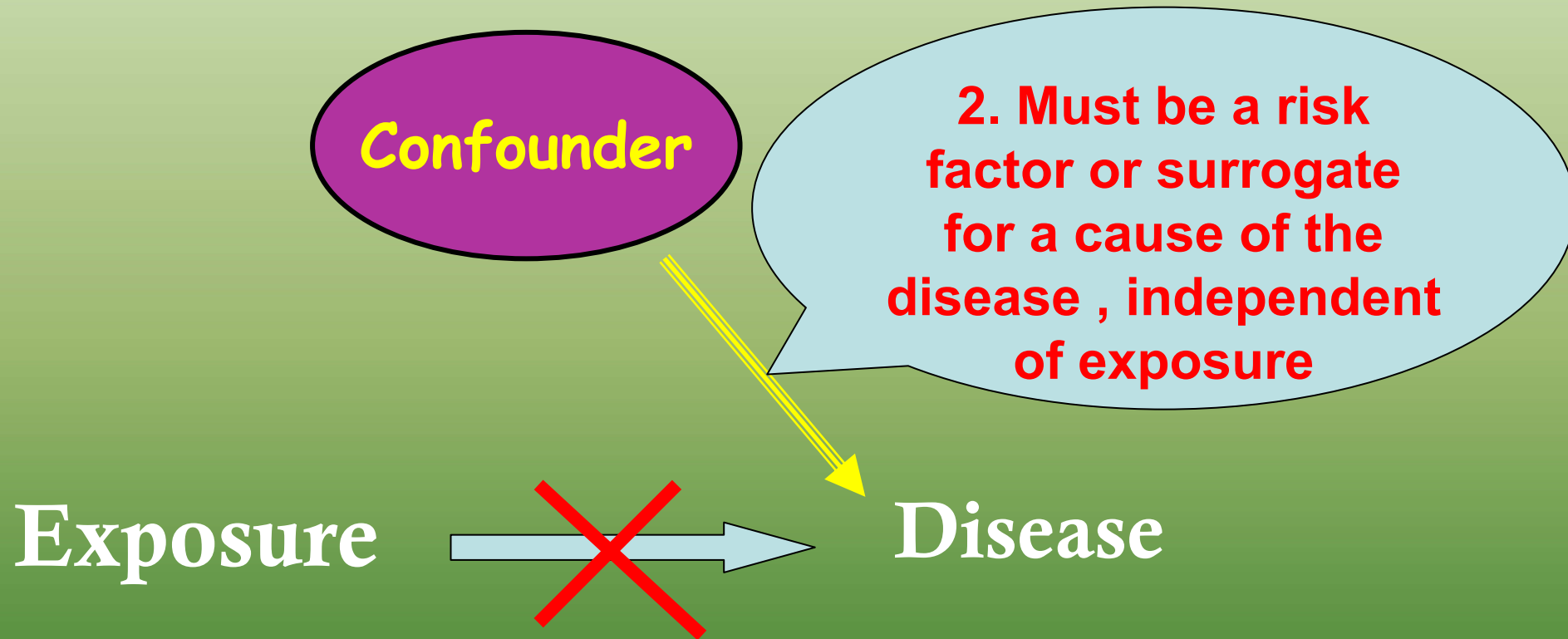


Criteria for confounding

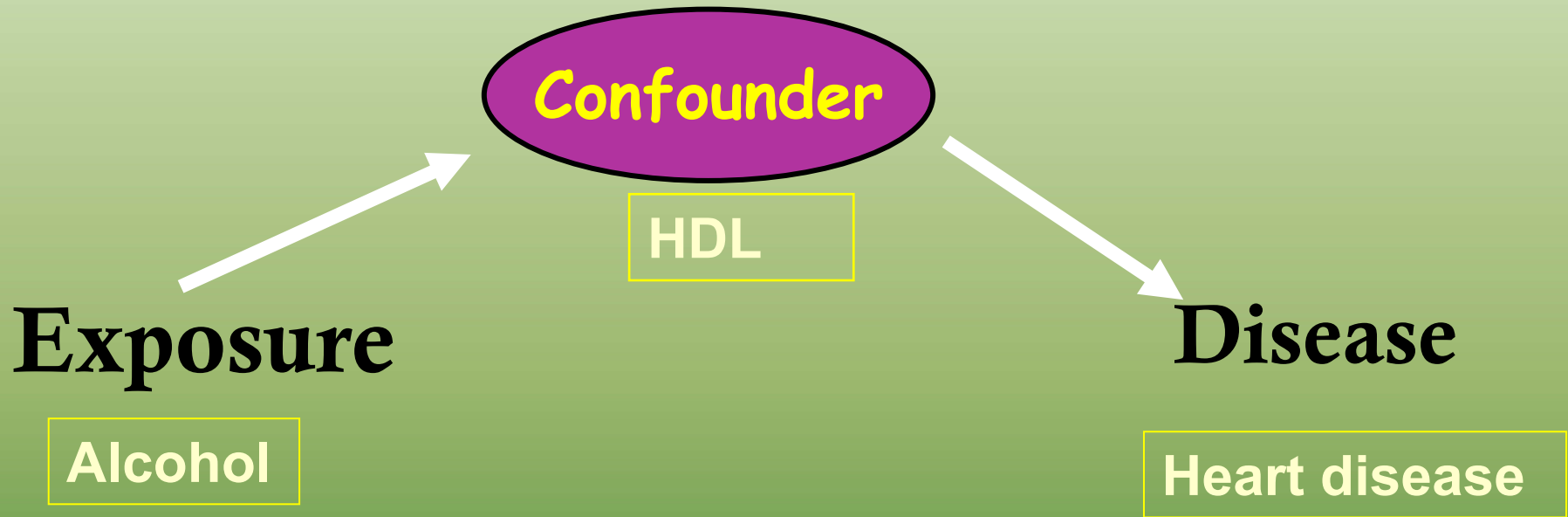


- In cohort –it must be related to exposure at start of study
- In case-control study –association must be present in population which give rise to control

Criteria for confounding



Criteria for confounding



3. A confounder must not be a consequence of exposure – should not be on the causal pathway

4. Must not be the result of disease

A variable can only be confounder if it is distributed differently between compared groups

The observed relationship between the exposure and outcome can be attributed totally or in part to the effect of the confounder

- overestimation or underestimation of the true association between exposure and outcome
- even change the direction of the observed effect.

How to identify confounding

**Compare the crude measure of effect
(RR or OR)**

To

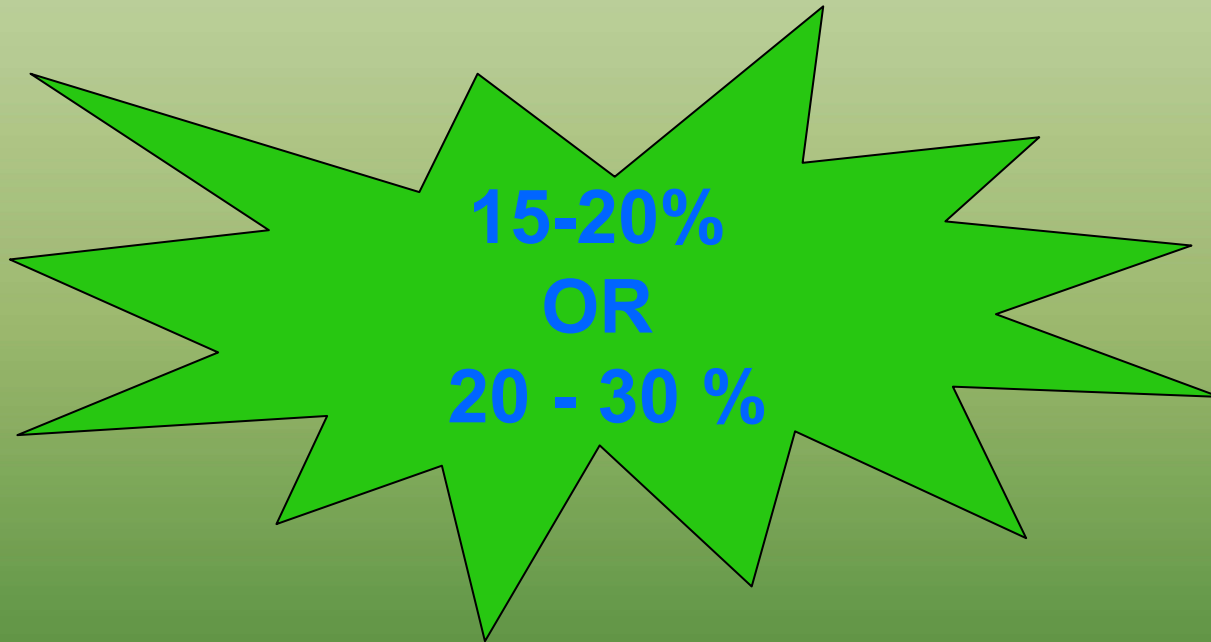
**Adjusted (weighted) measure of effect
(Mantel Haenszel RR or OR)**

Any statistical test to help us?

**Hypothesis testing for
confounding?**

Answer – Definite NO!

Confounding can be judged present when adjusted RR/OR-MH is different from crude RR/OR by about:



****Provided stratum specific RR/OR are homogenous (similar)**

If not –

INTERACTION!

Example of Confounding

Age on obesity and risk of mastitis

	Mas +ve	Mas -ve	Total	Risk
Obese	50	150	200	0.25
Normal	30	170	200	0.15

C-RR = 1.7

Adj. RR = 1

Crude vs Adj. RR

= 41%

Young cows **RR = 1**

	MF+	MF -	Total	Risk
Obese	5	45	50	0.10
Normal	15	135	150	0.10

Old cows **RR = 1**

	MF+	MF -	Total	Risk
Obese	45	105	150	0.30
Normal	15	35	50	0.30

Age on oral contraceptive (OC) and myocardial infarction (MI)

OC-use	MI - cases	Control
Yes	45	24
No	114	154

Crude OR = 2.5

Adj. OR = 3.2

Crude vs Adj. OR
= 28%

AGE <40

OR = 3.2

OC-use	MI - cases	Control
Yes	24	17
No	26	59

AGE >=40

OR = 3.2

OC-use	MI - cases	Control
Yes	21	7
No	88	95

Example of non-confounder

Breed on obesity and risk of mastitis

	MF+	MF -	Total	Risk
Obese	240	1760	2000	0.12
Normal	92	1908	2000	0.046

Crude RR = 2.6

Adj. RR = 2.8

Breed 1

RR = 2.6

	MF+	MF -	Total	Risk
Obese	211	789	1000	0.21
Normal	82	918	1000	0.08

Breed 2

RR = 2.9

	MF+	MF -	Total	Risk
Obese	29	971	1000	0.029
Normal	10	990	1000	0.01

Confounding Vs. Interaction

CONFOUNDING

Belongs to study

- Weighted RR different from crude RR-but strata specific RR are homogenous
- Distortion of effect & creates confusion in data

INTERACTION

Belongs to nature

- Different effects in different strata
- Useful
- Increases knowledge of biological mechanism

Control for Confounding

Study Design Phase

1. Randomisation
2. Matching
3. Restriction (exclusion)

At Analysis

1. Stratified analysis
2. Multivariable analysis

Study Design Phase

1. Randomization

Every individual has the same chance of entry into a group –only way to equalize all factors

Sufficient sample size, randomization is likely to control both known and unknown confounders (no guarantee!)

Study design phase

2. Matching

Select study subjects so that the potential confounder/s are distributed in an identical manner among the exposed and non-exposed groups (cohort) and cases and controls (case-control study)

***In case-control –matching does not in itself control confounding, matched analysis must also be performed**

– crude OR from matched case-control study will be biased – should use an adjusted estimate using matched analysis.

Study Design Phase

3. Restriction (exclusion)

Restrict admissibility criteria for study subjects and limit entrance to individuals who fall within a specified category of confounder

-But it limits generalisability

Analytical control

1. Stratified analysis

Stratification is a technique in which data are stratified by the levels of confounding factor and the relative risk estimates are compared between the different strata and come up with summary effect estimate.

a) Pooling – Pooling of stratum specific estimate into one estimate -Mantel-Haenszel technique

1. Stratified analysis

b) Standardization

Involves taking a weighted average of disease occurrence across strata and then comparing the standardized occurrence measure between exposed and non-exposed with no assumptions of uniformity of effect

Standardization

1. Direct method

Stratum-specific rates of study population are applied to the confounder (age) distribution of standard population

Direct Standardization

Std Ref. Pop. TAR dist. of:

Beef = 40%

Dairy = 60%

$$\text{Direct Std. rate} = \sum T_{sj} * I_{j_j}$$

TB-Herd Rate

Type	No. of cases	No. of Herd-yr	Obs. Rate (T _{sj})	Ref. Pop Distrib. (I _j)	Product (T _s * I _j)
Region A					
Beef	17	550	0.031	0.4	0.012
Dairy	41	450	0.091	0.6	0.055
	58	1000	0.058		0.067
Region B					
Beef	10	500	0.020	0.4	0.008
Dairy	120	1500	0.080	0.6	0.048
	130	2000	0.065		0.056

Standardization

2. Indirect method

Stratum-specific rates from standard population are applied to confounder (Age) distribution of study population

Indirect Standardization

Std Incidence rate of Pop

Beef herd = 0.025 cases/herd-yr

Dairy herd = 0.085 cases/herd-yr

Overall = 0.06 cases/herd-yr

SMR = Obs / Expected

**Indirect Std. rate
= Overall rate * SMR**

Type	No. of cases	No. of Herd-yr	Prop. in each Stratum (Hj)	Ref. Pop Std. rate (Is)	Product (Hj * Is)
Region A					
Beef	17	550	0.55	0.025	0.014
Dairy	41	450	0.45	0.085	0.038
	0.058	58	1000	SMR = 1.12	0.052
Region B					
Beef	10	500	0.25	0.025	0.006
Dairy	120	1500	0.75	0.085	0.064
	0.065	130	2000	SMR = 0.93	0.070
Indirect Standardized Rate = 0.067					
Indirect Standardized Rate = 0.056					

Analytical control

2. Multivariate analysis

- Any analysis technique that simultaneously adjusts for several variables
- simultaneously control potential confounders.

- Once included in model – effect of other variables would be minus the effect of confounder

- Multiple linear regression – continuous outcomes
- Logistic regression – categorical outcomes

The adjusted measures of association we obtain from multivariable models (multiple factors included) are “**direct effects**” and not “**total effects**”.

Thus for the causal association, we will have “**over-controlled**” for intervening variables (and perhaps effects of cause)

Therefore the need of *a priori* knowledge of potential confounders and

the “**CAUSAL DIAGRAM**”

Important Tips

1. List potential confounders and effect modifiers right at the start
- “No control without information - WITH ONE EXCEPTION – randomization in experimental studies”
2. If possible control confounder/s at design phase
3. Stratify data according to potential confounders or effect modifiers and estimate measures of association
5. Check for interaction (effect modification)
6. If interaction is present, show the data by stratum
7. If not check for confounding
8. If confounding is present show adjusted measure of effect
9. If no confounding, show crude data

Important Tips

10. In multivariable analysis – draw a “Causal Diagram”- to identify variables to be controlled or not in the model
11. Known potential confounders for primary predictors of interest should be included in the model, regardless of significance and do not include intervening variable.

Important Tips

- 12. Determination of presence of confounding demands an understanding how things work**
- mechanisms, trade-offs, processes and dynamics, cause and effect, etc.**

“MATTER OF SCIENTIFIC JUDGEMENT!”

I hope my talk on
confounding was not
confounded!!