

The Online R-FETPV 1st Module : Basic Epidemiology and Surveillance Data Analysis

5 April -28 May 2021



USAID
FROM THE AMERICAN PEOPLE



**Food and Agriculture
Organization of the
United Nations**



Bias, confounding and interaction in Epidemiology

Sangeeta Rao, BVSc, MVSc, PhD
Colorado State University



USAID
FROM THE AMERICAN PEOPLE



**Food and Agriculture
Organization of the
United Nations**



Outline:

- Systematic error
- Random error
- Types of Bias
- Confounder and Interaction
- Examples

Scenario



Depositphotos.com

- You are asked to visit a local dairy herd with 149 dairy cows and collect 30 blood samples as part of a TB surveillance effort to estimate prevalence in the lactating herd.

Select the BEST option

- A. Collect a sample from each of the first 30 cows you see?
- B. Collect a sample from each of the 30 cows the producer has selected from his herd?
- C. Collect a sample from each of the 30 cows you have selected randomly from the herd?

Two Types of Error

Systematic error or Bias

- A process that systematically produces error from the true value (usually in the same direction)

Random error

- Statistical fluctuations (in either direction) in the measured data due to the precision limitations

```
graph TD; A[Systematic error or Bias] --> C((Error)); B[Random error] --> C;
```

Error

Examples

Type of Error	Example	How to minimize it
Systematic error	The weighing scale you use for dogs reads 2 kg less for all measurements.	Systematic errors are difficult to detect and cannot be analyzed statistically, because all the data is off in the same direction (either too high or too low). Spotting and correcting for systematic error takes a lot of care.
Random error	You measure the body temperature of the cow 3 times using the same thermometer and get slightly different values: 38.6, 38.4, 39°C.	Take more data. Random errors can be evaluated through statistical analysis and can be reduced by averaging over large number of observations.

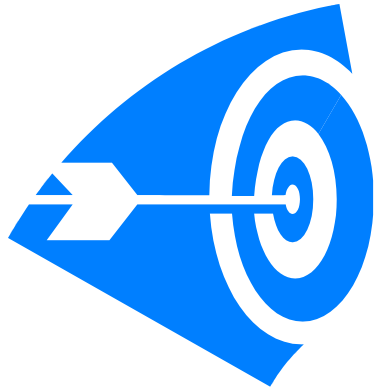
Bias

Systematic (non-random) error in the design and/or analysis affecting data quality and statistical inferences in a study

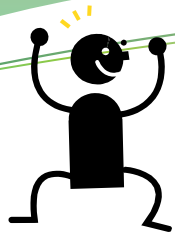
Impact of biases

- Make factor seem important when it is not
- Make factor seem unimportant when it really is
- Under or overestimate the "true risk"

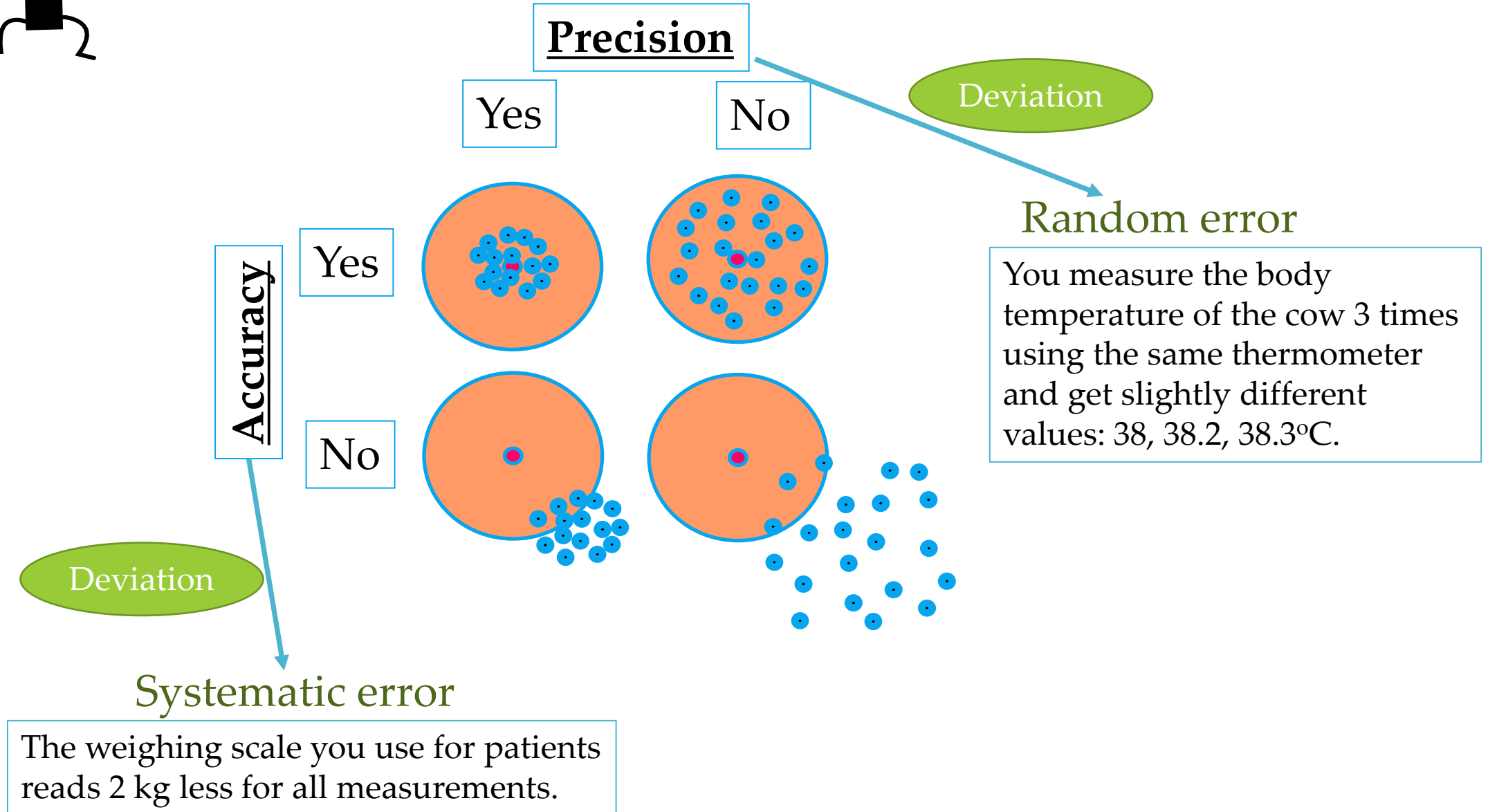
What are *biased data*?



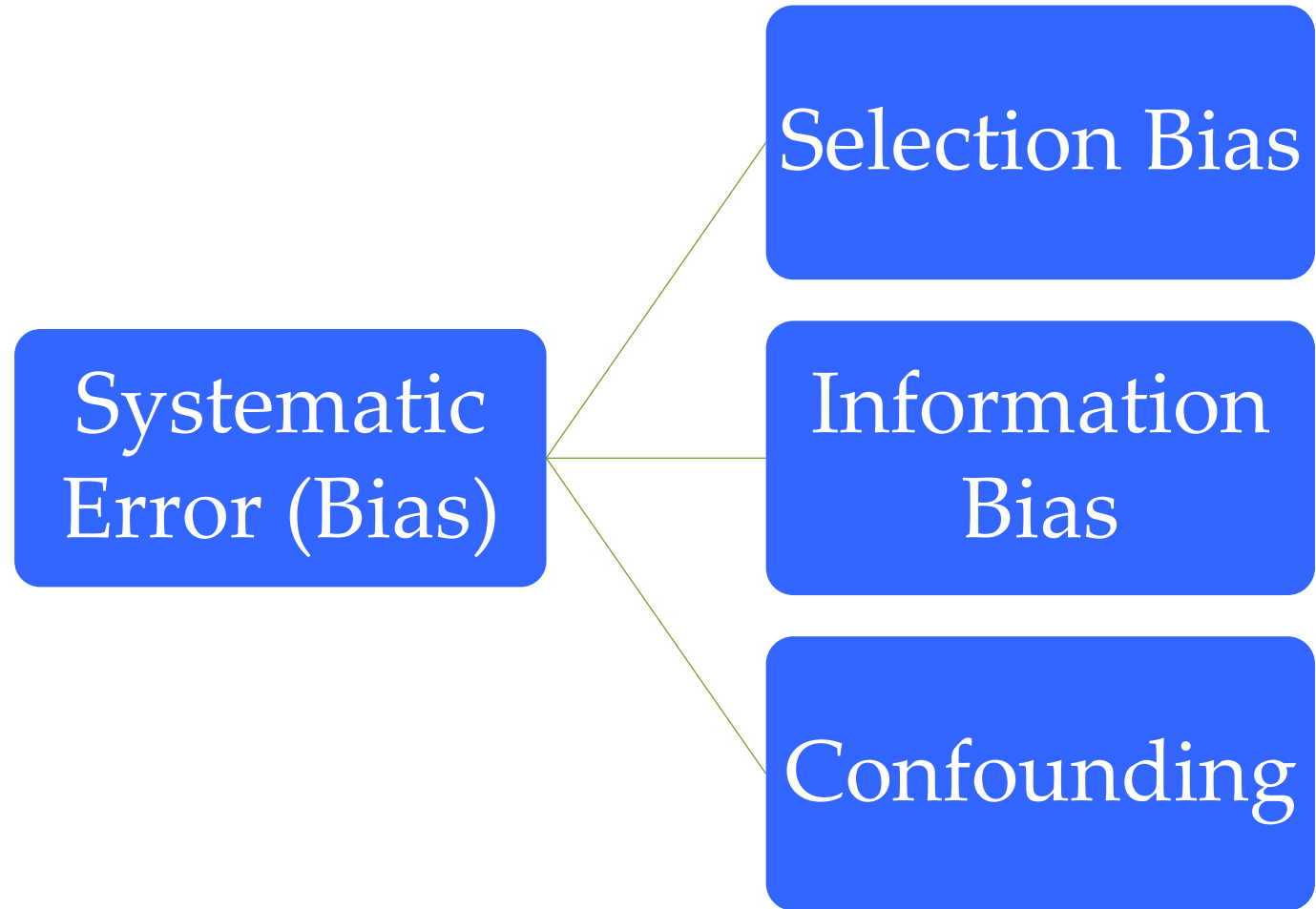
- Using the target analogy:
 - Is bias a lack of...
 - Accuracy ?
(arrows scattered about the bull's-eye)
 - Precision?
(arrows tightly clustered BUT *away* from the bull's-eye)



Quality of Measurements



Types of Systematic Error



Types of Bias

- Selection (Enrollment) Bias
 - The way that study subjects are selected creates bias in outcome or exposure (*internal validity*)
 - Study population is not representative of other populations (*external validity*)
- Information (Measurement) Bias
 - Error in estimation or evaluation
 - Rigor of follow-up, etc.
- Confounding
 - Mixing of the effects of two or more causal factors
 - Distribution is specific to study population
 - The impact is study specific

Where Bias is Introduced



Some examples

- **Bias in Response:** Fancy term for lying or simply forgetting. Questions about the past or sensitive issues often get biased responses (Question to a farmer: “How many times a week do you sterilize the milking utensils?” or “How much money do you earn by selling milk?”).

Information or Measurement Bias

- **Bias due to framing of question:** Questions worded like “Do you agree that no proper biosecurity measure is a risk for the spread of FMD in cattle?” are prompting you to give a particular response. Questions may also be too complicated and confusing.

Information or Measurement Bias

- **Bias due to under-coverage:** Sometimes parts of the population are left out in the process of choosing the sample. For example, “sampling only the dogs visiting hospital for mange, the stray dogs are not included in the population”.

Selection or Enrollment Bias

Dealing with Bias

- Bias may not be immediately apparent so plan carefully
- Design stage
 - Appropriate design and selection of subjects and tests
 - Matching groups
- Analysis stage
 - Stratification
 - Statistical Analysis
 - Forcing confounders into multivariate models e.g. logistic regression
- Caution must be used when adjusting for bias and an and/or statistician should be consulted

Basic Approaches for the Control of Bias

- Randomization
 - Selection
 - Treatment allocation
- Blinding
 - Single vs. Double vs. Triple
 - Participant vs. owner / manager
- Removing/Controlling Confounders
 - Restriction
 - Matching
 - Analysis

Blinding or Masking

- Bias is a particularly challenging problem when dealing with human subjects because of the *placebo effect*.
- **Single Blinding** → conceal treatment allocation from subject (owner)
- **Double Blinding** → also conceal allocation from investigators
 - Administering treatments
 - Making measurements (pre/post) or assessing outcome
- **Triple Blinding** → data analysis is conducted blinded
- However, subjects must be informed that they will get one of several treatments and must consent to that condition (it would be unethical otherwise).

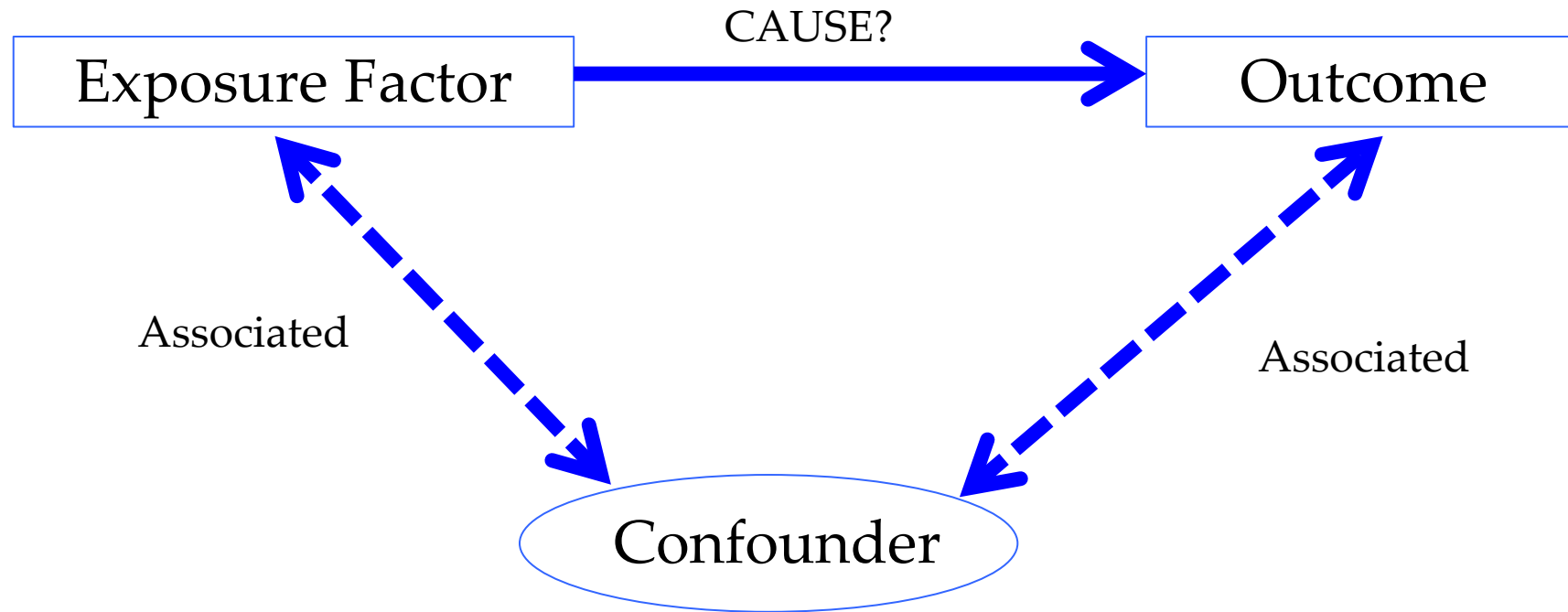
Confounding and Interaction

```
graph TD; A[Confounding and Interaction] --> B[Situation in an analysis where the effect or association between a factor and outcome is "distorted" by the presence of another factor. It is a type of bias.]; A --> C[Situation in an analysis where a factor positively or negatively "modifies" the observed effect of a risk factor on disease status.];
```

*Situation in an analysis where the effect or association between a factor and outcome is "distorted" by the presence of another factor.
It is a type of bias.*

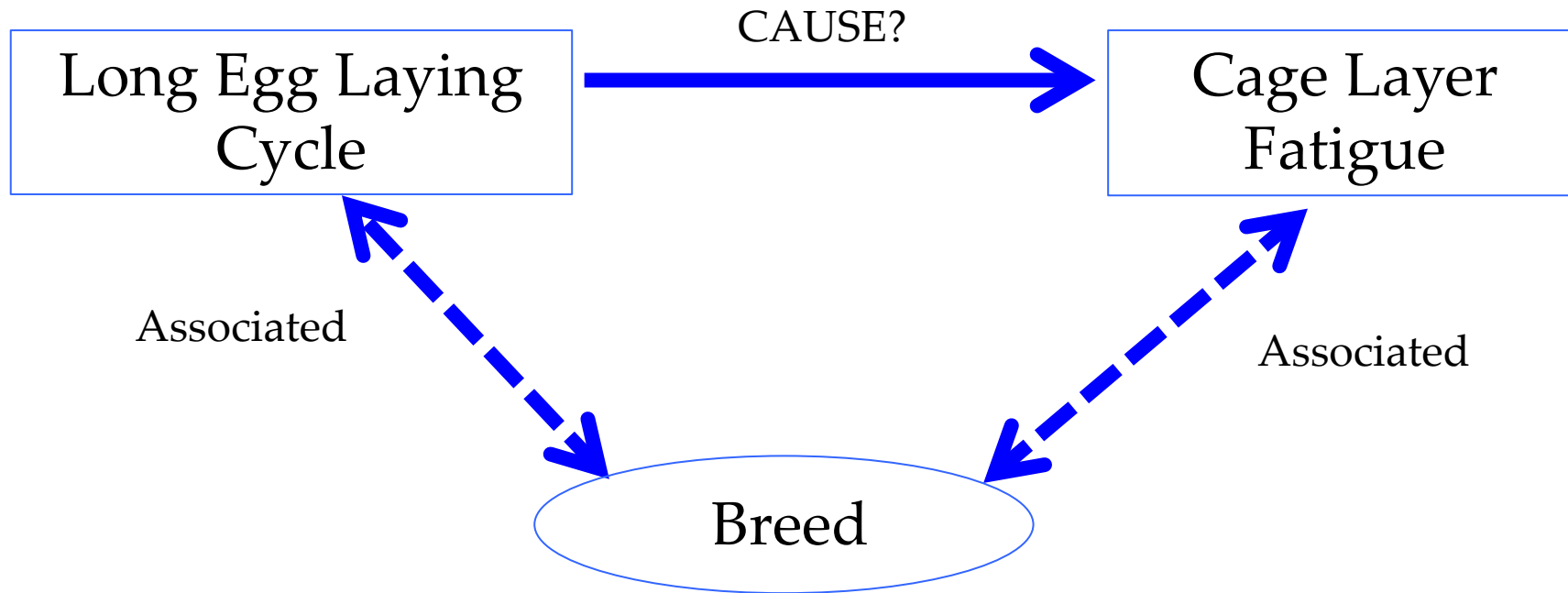
Situation in an analysis where a factor positively or negatively "modifies" the observed effect of a risk factor on disease status.

Confounding



A confounding factor is associated with exposure of interest and the outcome.

Example



- Confounding occurs when the effects of two exposures have not been separated

Breed is confounding variable: True risk for Cage Layer fatigue is associated with long egg laying cycle; much of that risk likely disappears after accounting for breed

Example

A study hypothesizes that stray dogs that are in close vicinity to street dumpsters (garbage) are more likely to get rabies.

2x2 table:

		Rabies	No Rabies
Vicinity to dumpster	Ye s		
	No		

OR = ???

Case-control study:

100 cases of Rabies, 78% are close to dumpster

100 controls (no rabies), 38% are close to dumpster

Case-control study:


100 cases of Rabies, 82% are close to dumpster

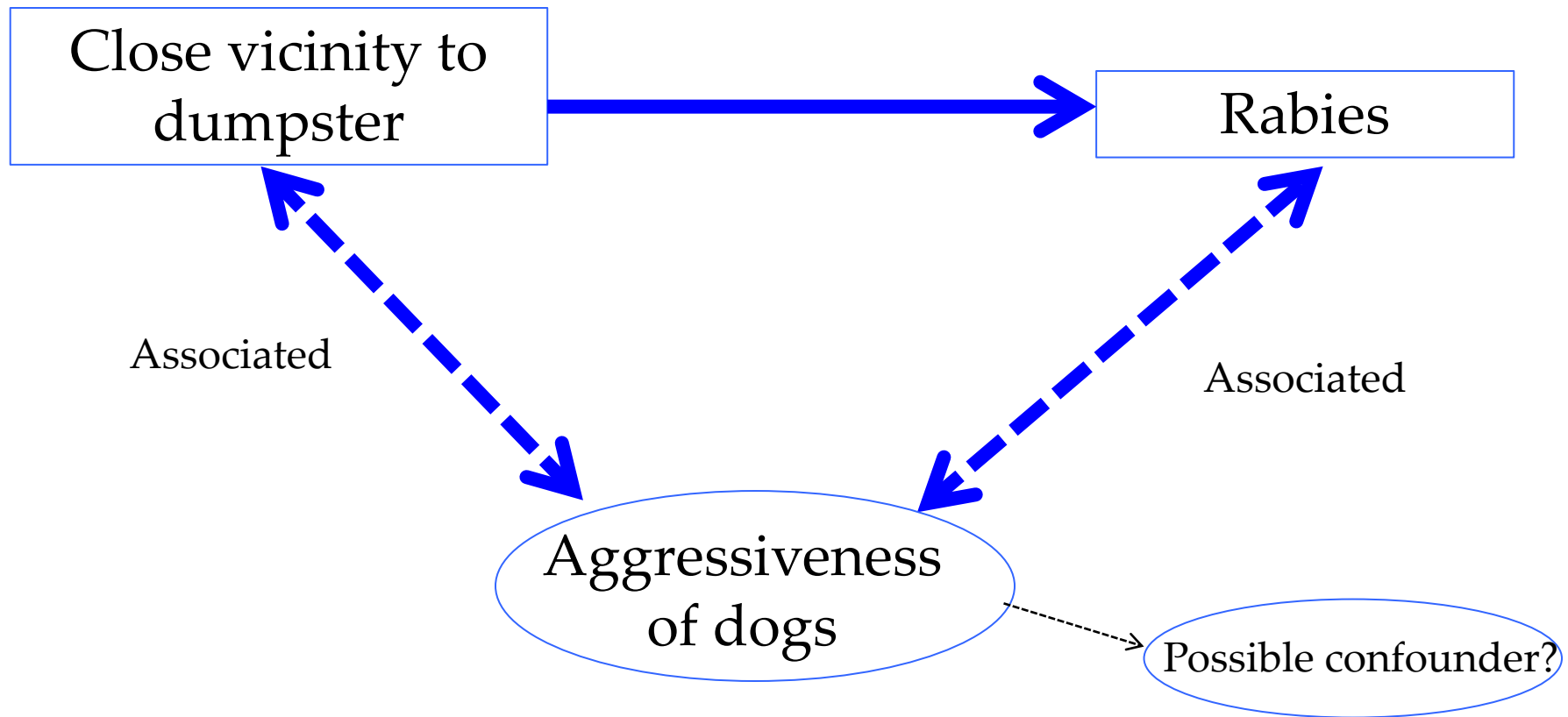
100 controls (no rabies), 34% are close to dumpster

		Rabies	No Rabies
Vicinity to dumpster	Yes	78	38
	No	22	62

$$OR = \frac{78 \times 62}{22 \times 38} = 5.78$$

Close vicinity to dumpster increases the likelihood (odds) of rabies almost 6-fold!

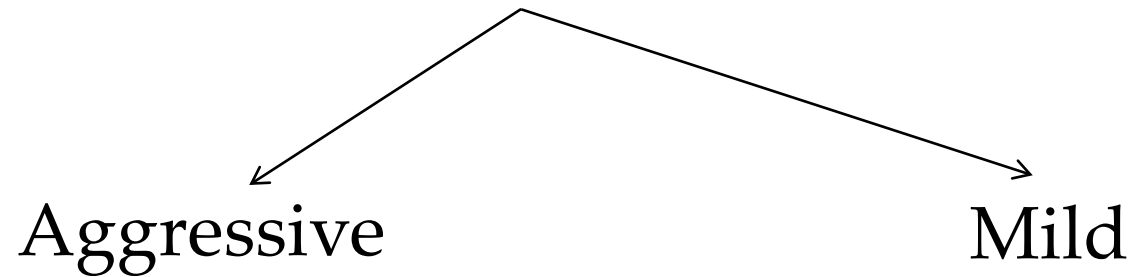
- 
- It is true that dogs close to dumpster are at increased risk of rabies
 - But it is *not* true that close vicinity to dumpster *causes* rabies.
 - Differentiate between 'association' and 'causation'!



A confounding factor is associated with exposure of interest and the outcome.

Now...

- Stratify by Confounder
- In this case, 'aggressiveness' is the *confounder*
- The analyses should be stratified by 'aggressiveness'.



Stratify by Confounder:

Aggressive dogs

		Rabies	No Rabies
Vicinity to dumpster	Yes		
	No		

n= 120 aggressive

Mild dogs

		Rabies	No Rabies
Vicinity to dumpster	Yes		
	No		

n= 80 mild

Stratify by Confounder:

		<i>Aggressive dogs</i>		<i>Mild dogs</i>	
		Rabies	No Rabies	Rabies	No Rabies
Vicinity to dumpster	Yes	75	23	3	15
	No	15	7	7	55

$$OR_{aggressive} = \frac{75 \times 7}{15 \times 23} = 1.52$$

$$OR_{mild} = \frac{3 \times 55}{7 \times 15} = 1.57$$

If the stratified odds ratios are similar to each other and are >10% different than the crude OR (=5.78), we can suspect confounding.

If the stratified odds ratios are different than each other, we can suspect interaction.

After Stratifying by aggressiveness

- The association between vicinity to dumpster and rabies in dogs is completely confounded by their aggressiveness.
- The association between vicinity to dumpster and rabies was evident due to a third variable (aggressiveness in dogs).

Specification and Matching

- *Specify/ Restrict*: only study those with (or without) the confounder
 - Look at vicinity to dumpster and rabies only in aggressive dogs.
- *Match*: make sure that a case of rabies is matched to a control on a confounder (aggressiveness).

Scenario

- You are interested in assessing whether 'off-site' heifer rearing reduced the incidence of ringworm compared to 'on-site' rearing. What confounding factors (or other biases) might need to be considered in the design and analysis of such a study?

