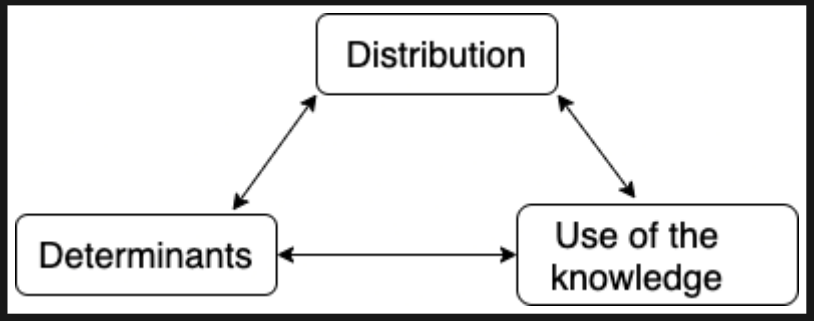
Introduction to Epidemiology: a tour

Epidemiology is the study of distribution and determinants of diseases in populations and use of this knowledge for the improvement in public health

Epidemiology in triads



Epidemiology triad

- Distribution of diseases
- Determinants of diseases
- Use of the information

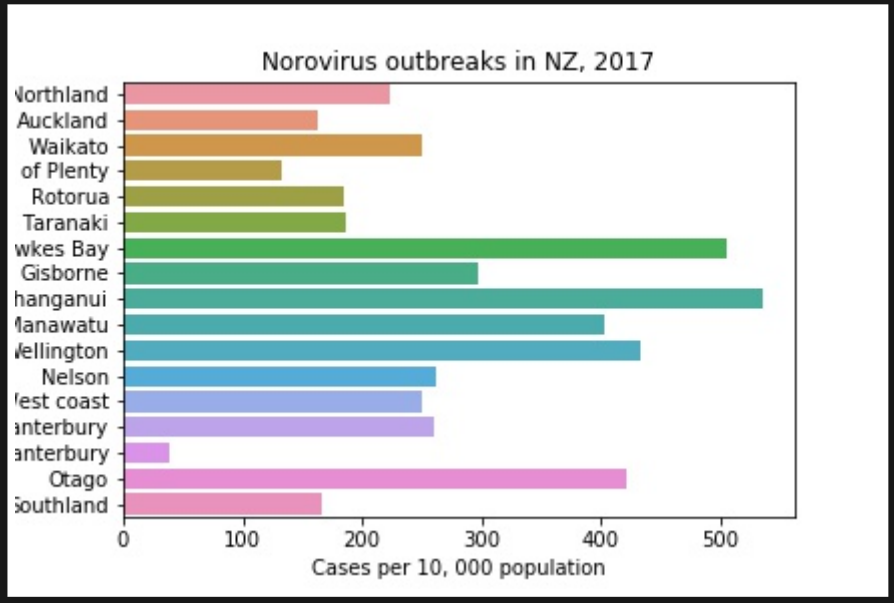
Distribution of diseases

- Person
- Place
- Time

Distribution of diseases person

	Age group	Sex	Case Status	Overseas travel
0	40 to 49	Male	Confirmed	No
1	10 to 19	Female	Confirmed	Yes
2	20 to 29	Male	Confirmed	Yes
3	0 to 9	Male	Confirmed	Yes
4	20 to 29	Female	Confirmed	Yes

Distribution of diseases with place



Distribution of diseases over time

Measures of disease distribution

- Prevalence (proportion)
- Incidence (rate)
- Ratio (standardised mortality ratio)

Concept of prevalence

- total cases of a disease in the population OVER
- total number of people there
- MULTIPLIED by factor of 100 (percent), 1000 or 10, 000
- At a FIXED PERIOD of TIME

Example of prevalence

	Heading	Town A	Town B
0	Total cases	120.0	140.000000
1	Total population	400.0	620.000000
2	Prevalence	30.0	22.580645

Advantage of prevalence

- Simple snapshot of a health condition in the population
- For a single population, at one point in time
- You can use it to compare two or more populations

Limitation of prevalence

- Does not provide any information about how the disease spreads in the community
- Is the disease increasing?
- Is it getting worse?
- Is it getting better?

Question for the class - what data do we need?

- What additional data do we need for that?
- (Question for the class, what do you think?)

Concept of Incidence

- How many NEW cases of the disease in the community
- How many people were AT RISK?
- Over WHAt period of time?

Concept of Person-time

- If you follow 1 person for 1 year,
- You get 1 person-year
- If you follow 100 people for 1 year,
- You get 100 person-year
- If you follow 100 people for 10 years,
- You get 1000 person-years

Question for the class

Let’s say we follow 1000 people for 5 years, which of the following is correct?

- (X) We have 5000 person-years
- (Y) We have 1000 person-years
- (Z) We have 5 person-years

Concept of incidence rate

- Incidence is a rate because it has TIME as DENOMINATOR
- Number of NEW CASES OVER Person-years
- REQUIRES FOLLOW UP of people

Example of incidence rate

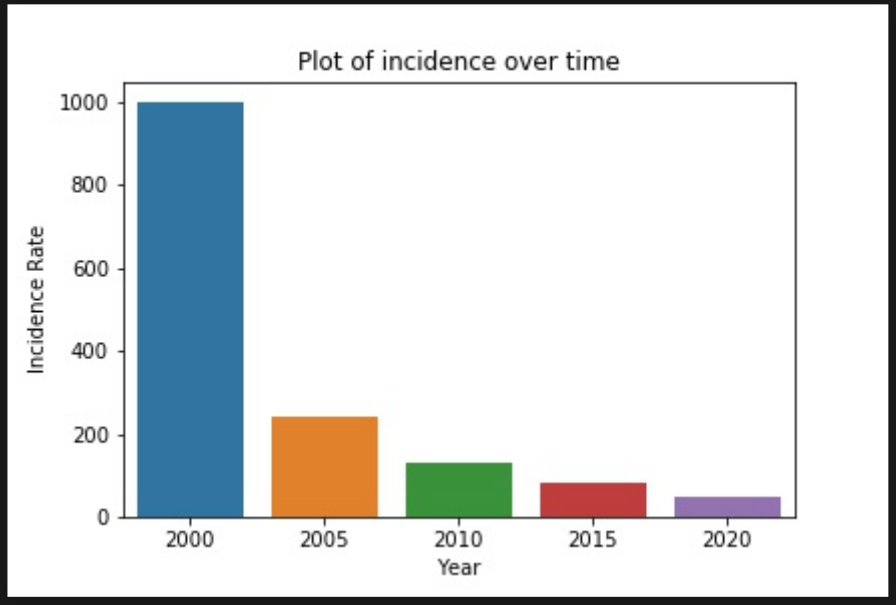
	Year	Person-years	New Cases	Incidence Rate
0	2000	1000	100	1000.0
1	2005	5000	120	240.0
2	2010	10000	130	130.0
3	2015	15000	120	80.0
4	2020	20000	100	50.0

Question for the class 2

Based on the data presented below, is the incidence

- (A) increasing or
- (B) Falling off

Incidence of the disease in the data

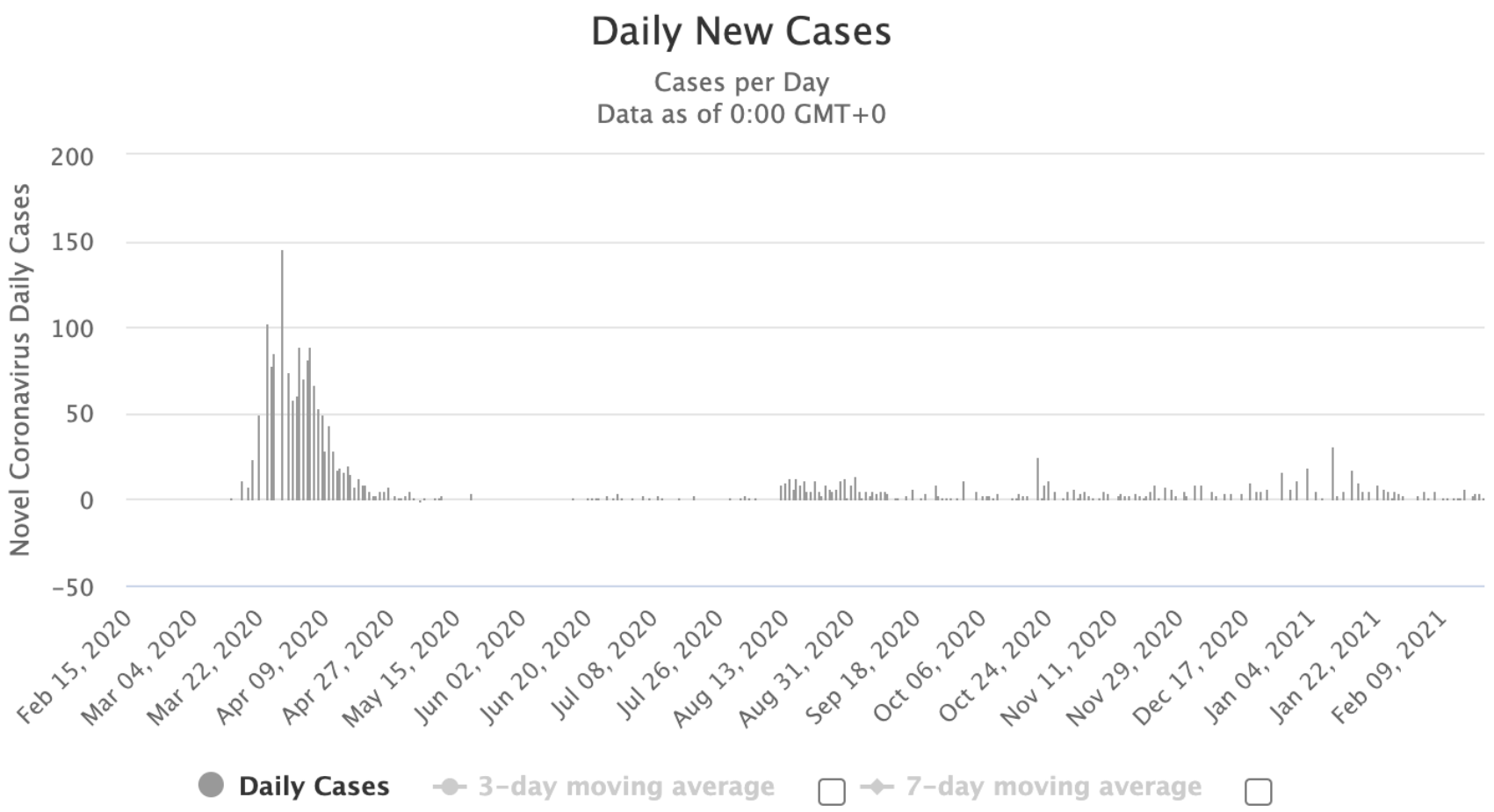


Advantage of Incidence rates

- Used for charting epidemics
- Helps you to understand whether an epidemic is getting better or worse
- Helps to chart data in real time

Example of incidence rate in real life

Daily New Cases in New Zealand



Question for the class - 3

How can we use the information on the COVID Epidemic Curve?

- (1) Test whether the infection is rising or falling
- (2) Get an idea when to introduce lockdown or other containment measures
- (3) Find out whether lockdown or containment measures are working or not
- (4) All of the above

Limitation of Incidence

- We need longitudinal data
- Without follow-up data we cannot estimate incidence
- Incidence is too simple if we want to compare different populations that differ in age groups

Question for class - which population has higher incidence?

Concept of age-standardised rates

	Value
Standardised Count for Population A	6240.0000
Standardised Count for Population B	6630.0000
Standardised Morbidity Ratio (SMR)	1.0625

Question for class - which population has NOW higher incidence?

Summary for distribution of diseases

- Three measures: prevalence, incidence, and standardised ratios
- Prevalence is used for static time
- Incidence is used in the context of person-year as denominator
- Incidnece is used for measuring how disease increases or decreases over time
- Standardised ratios used to compare two different populations

Break for 10 minutes

Determinants of diseases

What does determinants of diseases in populations mean?

How do we know that X causes Y?

Observe facts --> Frame theories --> Test with new facts

What is meant by exposure and outcome

What is association?

- If high levels of E leads to high levels of O, then
- We call that a positive association, or RISK
- That is, those with high levels of E will have high incidence of O
- Example: people who smoke lots of cigarettes end up with lung cancer
- We say cigarette smoking is associated with lung cancer
- Or, we say Smoking is a RISK for Lung Cancer

How do we measure associations?

- Risk Ratio (RR) Or
- Odds Ratio (OR)
- Risk Ratio = Risk of Disease among Exposed OVER Risk of Disease among non-Exposed
- Odds Ratio = Odds of Exposure among Diseased OVER Odds of Exposure among those without the disease

Decisions about RRs and ORs

- If RR > 1, the risk is high, OR, association is positive
- Otherwise, if RR = 1, then we cannot say anything
- Else, if RR < 1, there is BENEFICIAL effect!

Question for the Class - Is this high risk?

You conducted a study on cigarette smoking and risk of lung cancer, and found RR = 2.50; what would you say?

- (1) Cigarette smoking increases the risk of lung cancer
- (2) Cigarette smoking has no effect on lung cancer

Establishment of cause and effect

- If we have to show exposure E is a cause of disease O, then
- Show that E has TRUE or REAL association with O
- And show that,
- that Association is one of CAUSE and EFFECT

Four things to consider

- Valid Association: did not occur due to chance (rule out chance)
- Observed association could not be due to biases (eliminate biases)
- Observed association cannot be due to a third factor (confounding)
- Examine causal factors using Hill’s criteria
- Examine counterfactual theories of causation

Rule out the play of chance for exposure disease relationship

- Test with hypothesis testing
- Null Hypothesis: that there is equal chances of disease with and without exposure
- Alternative hypothesis: risk of disease is higher with exposure
- Always test with null hypothesis
- Test p-values and 95% confidence interval

What is meant by null hypothesis? (Example: smoking and lung cancer)

- Exposure and disease outcomes are unrelated
- People who smoke and who are non-smokers get lung cancer at the same rate
- Rate of lung cancer among smokers = Rate of smoking among non-smokers
- All epidemiological research is about disproving the null hypothesis

Question for the class - what is the correct null hypothesis

Imagine you are investigating whether smoking cause lung cancer. What is the correct null hypothesis?

p-value and 95%Confidence Interval

- You completed a study of an Exposure E and outcome O
- You were to repeat the study a 100 times!
- You found a measurement of RISK (RR) of 2.5
- You found a p-value of 0.02
- You found 95% Confidence Interval of 2.1 - 4.2
- What do these things mean?

Question for the class - what does RR of 2.5 mean?

- (A) E is high risk for O
- (B) E has no association with O

Concept of p-value

- Assuming that the null hypothesis is TRUE,
- and suppose you ran this study 100 times,
- Then only in 2 out of those 100 studies,
- you might get the kind of high risk RR you got
- You can reject the null hypothesis

Concept of 95% Confidence Interval

- If you repeated this study 100 times,
- In 95 out of 100 times,
- You might get an RR value between 2.1 and 4.2
- Most likely 2.5

Question for the class - What does RR of 2.1 mean?

- (A) E is high risk for O
- (B) E has no association with O

How do we rule out the play of chance?

- Select a large enough sample suitable for the effect you want to study
- Perform sample size estimation and power calculation ahead of the study
- Deal with it during the planning of the study

What is bias?

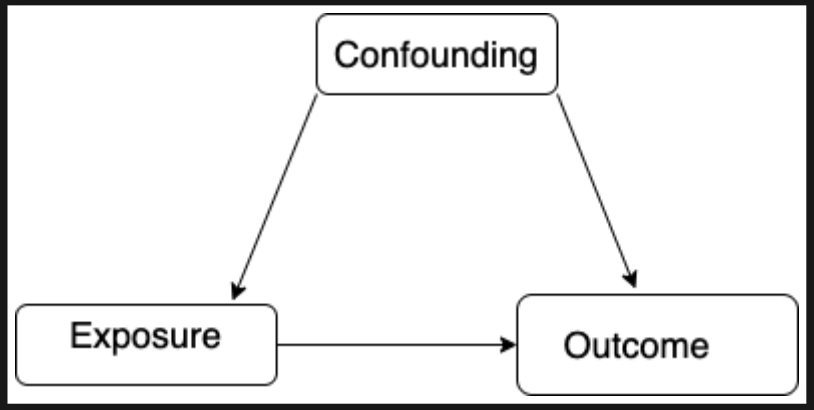
- Bias = Systematic errors in observation or conduct of the study
- Selection Bias: Where the groups are not comparable the way they were identified
- Response Bias: When the participants of the study provide erroneous information

How can you eliminate biases in the study design?

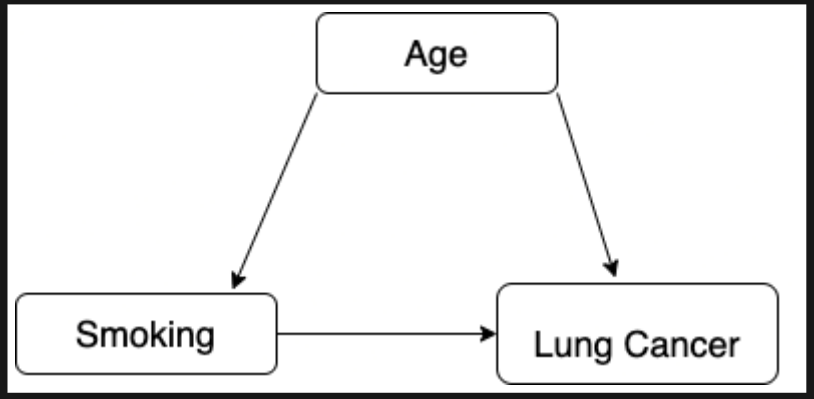
- In the study design phase,
- The investigator should be careful about selection of the sample
- In experimental studies, use randomisation and blinding
- Train the data collectors in the study

Tests of association: control for confounding variable

What is a confounding variable?



Example of a confounding variable



Explanation why Age is a confounding variable

- In the study on the association between smoking and lung cancer,
- Age is a confounding variable
- Old people tend to smoke more than younger people, AND
- Old age is also a risk factor for ANY cancer!
- Smoking CANNOT CAUSE Aging!

How do we control for confounding variable?

- Matching
- Multivariable analysis
- Allocation to groups being compared using Random Numbers Table

How do we find causal linkage using Bradford-Hill Criteria

Hill's Criteria of Causation Applied to Subluxation

	Criteria	Result
1	Strength	There were no studies that found a relative risk or odds ratio linking subluxation
2	Consistency	Subluxation has not been noted to be consistently found across any studies in different people, places, circumstances or time.
3	Specificity	There were no studies that linked disease with subluxation of any specificity. Other exposures (variables) or explanations can be given to the disease complex.
4	Temporal sequence	There were no studies suggestive of a temporal sequence linking subluxation with disease
5	Dose response	There were no studies found linking incidence of disease with magnitude of the subluxation
6	Experimental evidence	There were no consistent studies demonstrating subluxation in the animal model
7	Biological plausibility	No studies were found that offered reproducible evidence to suggest a biological plausibility of the subluxation construct.
8	Coherence	There were no studies that indicated a credible level of coherence
9	Analogy	There were no studies suggestive of a casual association via a similar agent.

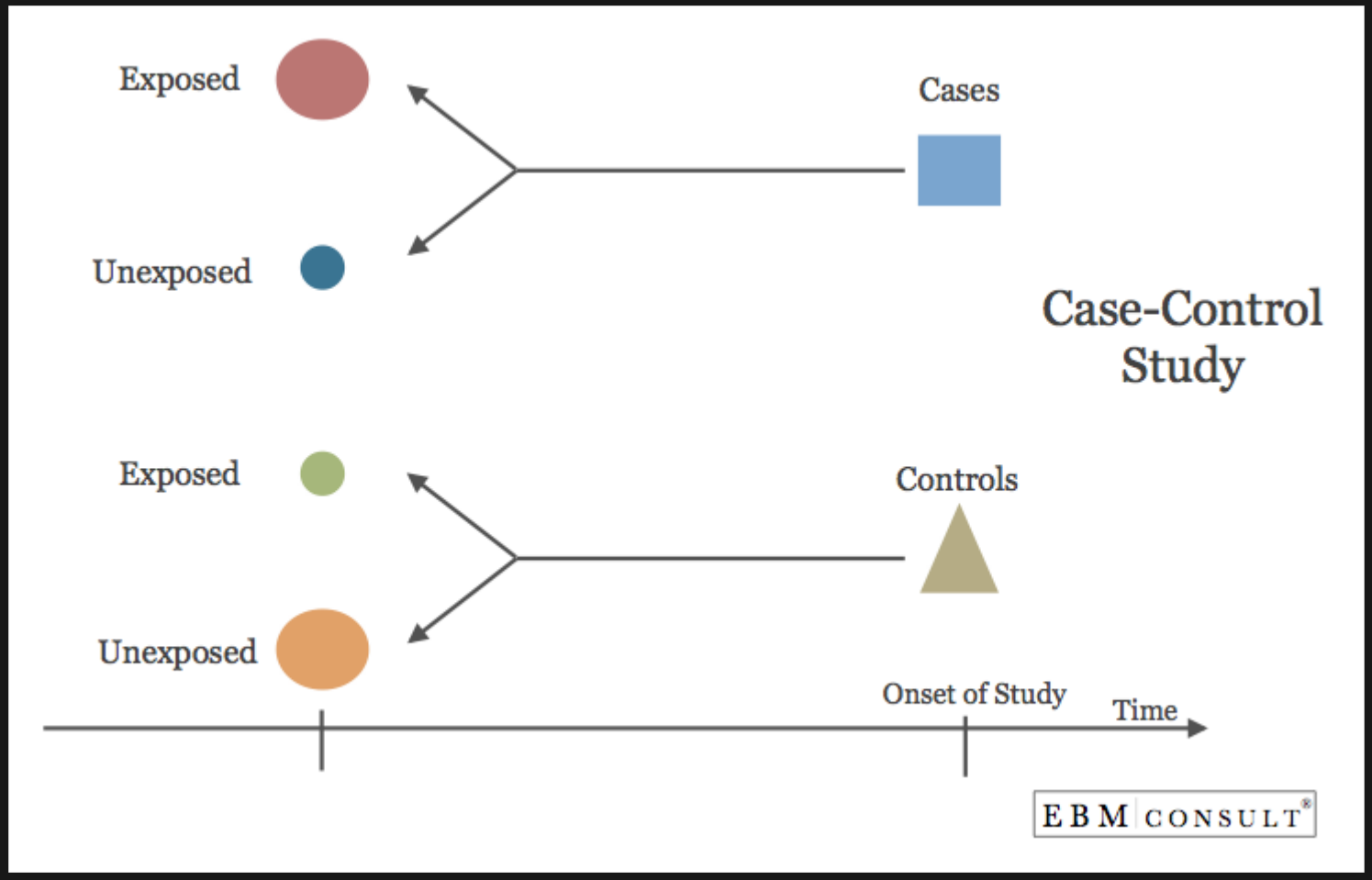
Break for a couple of minutes

Putting these ideas together: study designs

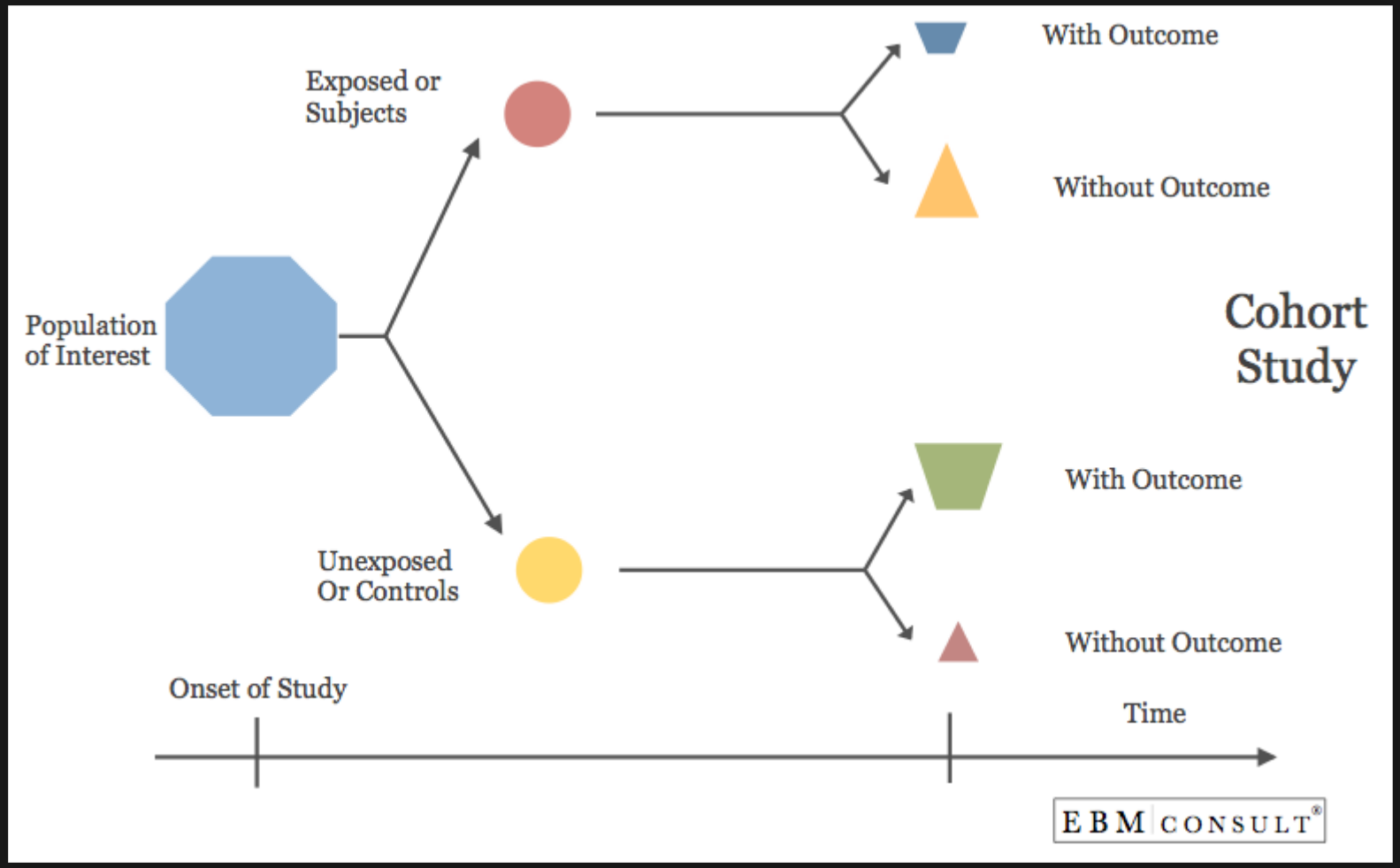
Epidemiological study designs

- Single Case studies
- Case series (used in Epidemic Surveillance)
- Cross-sectional surveys
- Case control studies (Most widely used study designs)
- Cohort studies

Case control studies



Cohort studies



Rounding up everything we learned with Snow's Cholera investigation

Cholera Epidemic

https://www.youtube.com/watch?v=KvHL0dHj3RM

The ghost map


The Broad
Street Well
GREAT MARLBOROUGH STREET
ARGYLL PLACE
ARGYLL STREET
REGENT STREET
CIRCUS
MADDOX STREET
NEW BURLINGTON STREET
GOLDEN SQUARE
WARWICK STREET
BROAD STREET
PORTLAND STREET
WORK HOUSE
SOHO SQUARE
OLD COMPTON
RICHMOND BUILDINGS

Snow’s table

Water Company	# of Houses Served	# of Cholera Deaths	Death Rate per 10,000 Houses
Southwark & Vauxhal	40,046	1,263	315
Lambeth	28,107	98	37
The rest of London	256,423	1,422	59

Summary