# DoE 9: Statistics and Epidemiology

## ISES Europe Training Series

Ruairí Weiner, Gerald Bachler

# Legal Notice

**Fair Use Disclaimer**

This presentation may contain copyrighted material used under the fair dealing exceptions in EU copyright law, in accordance with Directive 2001/29/EC and Directive 2019/790. It is intended solely for non-commercial, educational purposes.

**Copyright Notice**

# Overview of All Training Videos

DoE 1) Basic Concepts and Principles in Exposure Science

DoE 2) Fundamentals of Environmental Chemistry and (Eco)Toxicology

DoE 3) Exposure Modelling

DoE 4) Exposure Monitoring

DoE 5) Exposure Assessment and Risk Characterisation

DoE 6) Risk Management and Sustainability Assessment

DoE 7) Relevant Legislative Frameworks

DoE 8) Risk Communication and Stakeholder Engagement

**DoE 9) Statistics and Epidemiology**

Access all videos via: **https://ises-europe.org/**

# Domain of Expertise (DoE) 9:
## Statistics and Epidemiology

**Module 1** — Introduction to Statistics

**Module 2** — Statistics for Exposure Scientists

**Module 3** — Introduction to Epidemiology

# DoE 9: Statistics and Epidemiology

## Module 3

## Introduction to Epidemiology

**Ruairí Weiner**

# Presenters

**Meet Today's Presenter:** **Ruairí Weiner**

- **Current Role:** Teaching Fellow in Research Methods & Statistics, University College Dublin School of Public Health, Physiotherapy and Sports Science
- **Previous Roles:**
  - Teaching data science and statistics, Trinity College Dublin School of Computer Science & Statistics
  - Amyotrophic lateral sclerosis research, Trinity College Dublin School of Medicine
  - Ageing research, The Irish Longitudinal Study on Ageing
  - Programme evaluator for counselling services in third level education
- **Education:**
  - MSc in Applied Social Research

# Context and Disclaimers

## About This Lecture

This lecture provides an **introductory framework**, with some topics simplified for ease of understanding.

## Disclaimer

- The content presented herein does not necessarily reflect the opinions, views, or positions of the presenters' employer or any affiliated organizations.
- References to specific organizations, tools, or entities are for illustrative purposes only and do not imply endorsement or critique.
- While every effort has been made to ensure the accuracy of the information presented, errors or omissions may occur.

# Learning Objectives

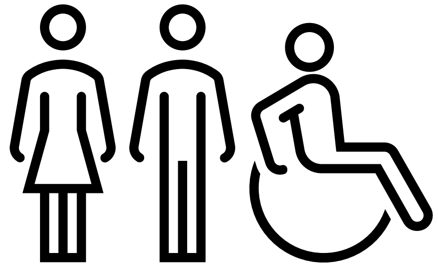
- Understand an overview of the history and aims of epidemiology.
- Describe the basics of measuring exposures and risk.
- Understand the role of epidemiology in linking environmental exposures to negative health outcomes.

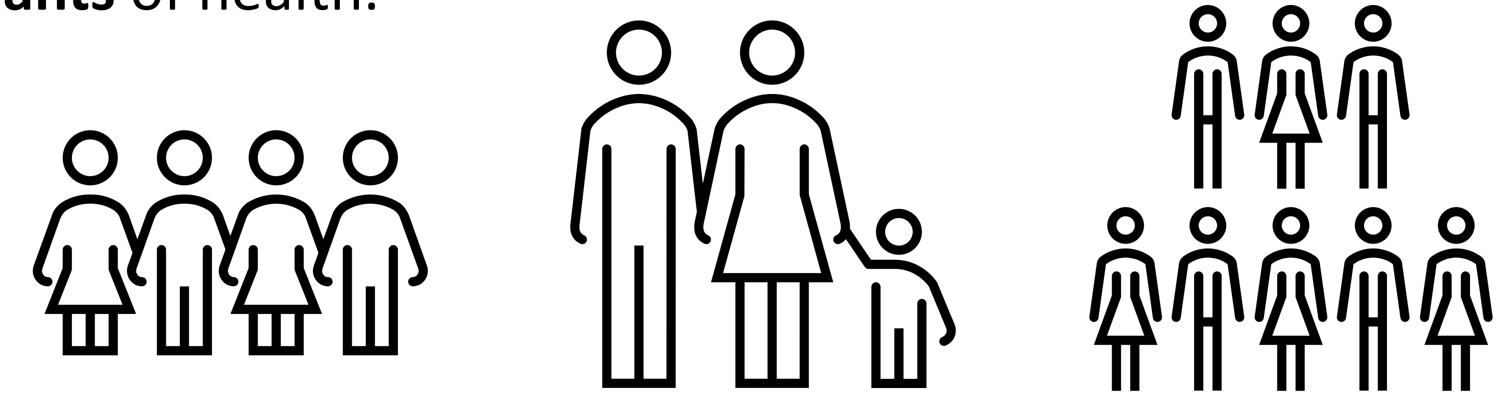
# Content

1. **Definition of Epidemiology**
2. Foundation of Epidemiology
3. Incidence and Prevalence
4. Exposure and Risk
5. Causal Inference
6. Summary

# Defining Epidemiology

- Comes from Greek – ‘*epi*’ means ‘on’ or ‘upon’, ‘*demos*’ means ‘people’ and ‘*logos*’ means ‘study.’
- Essentially: Study of the **population** – its condition.
- According to the US CDC:
  - Study the **distribution** of **health events**.
  - The **determinants** of health.



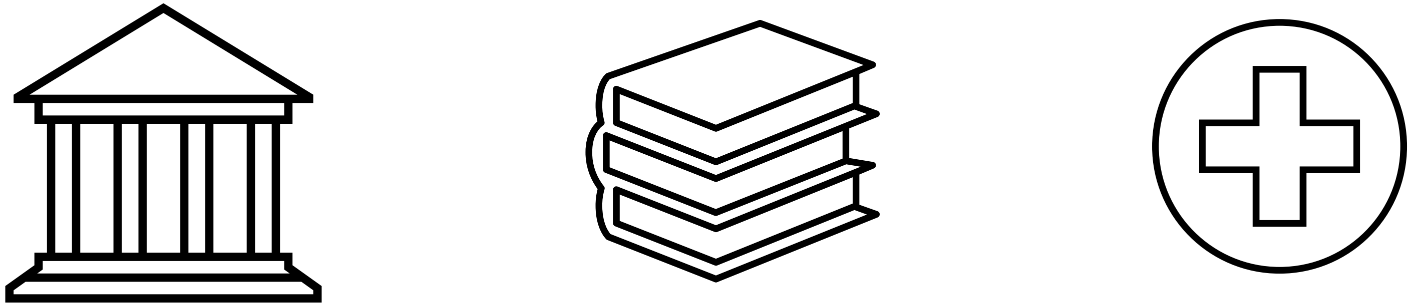


# Defining Epidemiology - Distributions

- **Count health events** – diagnoses, deaths, accidents.
- Map their **distribution** in the **population**
    - Location
    - Age
    - Gender
    - Socio-economic status
- Are **rates** of a health condition/event relatively higher/lower in some **groups**.

# Defining Epidemiology - Determinants

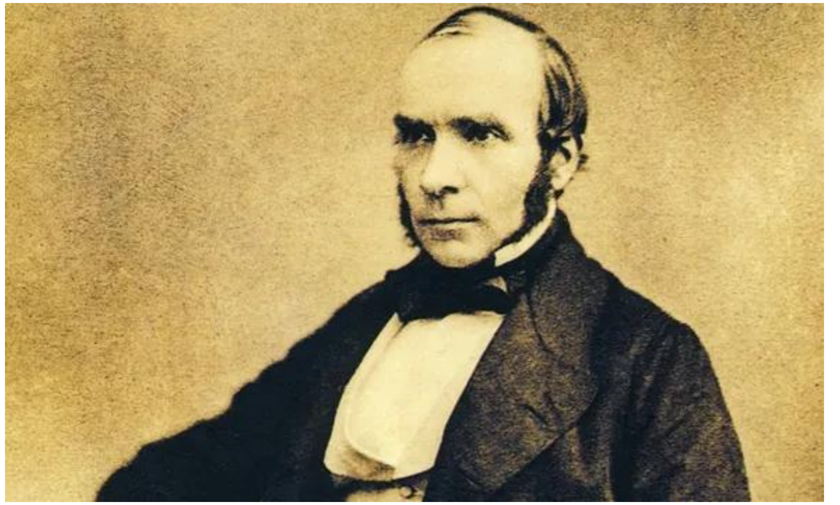
- Mapping health events may reveal **risk factors**.
- Groups who are more at risk or **exposure** that raise risk of a health event.
- Can we **intervene** on any risk factor to reduce rates.
- Need to identify **determinants of health** that can be targeted with **interventions** (policies, treatments, education campaigns).



# Content

1. Definition of Epidemiology
2. **Foundation of Epidemiology**
3. Incidence and Prevalence
4. Exposure and Risk
5. Causal Inference
6. Summary

# Foundation of Epidemiology



- Founding figure: John Snow
  - Victorian obstetrician.
  - Noticed things that appeared to be making people sick.
  - Devised ways to prove his suspicions.
- Tracking and mapping disease outbreaks.
- Identifying explanations for outbreaks.

# John Snow – Arsenic Gas in Autopsy

- While studying, Snow noticed **gastrointestinal** complaints in medical students conducting **autopsy**.
- Recent switch to arsenic embalming, though arsenic known **toxin**.
- Conducted **experiments** which showed **arsenic gas** is released from embalmed bodies.
- **Policy change:** Medical school ceases arsenic embalming and cessation of arsenic candle sales.
- Early example of observation & experimentation to demonstrate danger of **environmental exposure**, leading to **effective intervention**.

# John Snow – Cholera in London I

- Mid-19th century London faced repeated, deadly cholera **outbreaks**.
- **Germ theory** was not yet accepted; **miasma theory** blamed “foul air,” noting fewer cases at higher altitudes.
- **Drinking water** came from the Thames, heavily contaminated with sewage – Snow suspected this as the true source.
- Snow **mapped** cholera **cases** and **water pump locations**.

# John Snow – Cholera in London II

- Living near pumps drawing from the lower, more polluted **lower Thames** greatly increased cholera **risk**.
- This water served both poor and rich across London, **explaining seemingly random outbreaks**.
- During a severe outbreak, Snow traced nearly all cases to the **Broad Street pump** supplying contaminated water.
- A nearby workhouse and brewery were **spared** because they used their own well or beer.
- Some cases in surrounding areas were from people who travelled to the Broad Street pump for its preferred taste.

# Content

1. Definition of Epidemiology
2. Foundation of Epidemiology
3. **Incidence and Prevalence**
4. Exposure and Risk
5. Causal Inference
6. Summary

# Incidence and Prevalence

- **Counting** health events is the most fundamental exercise in epidemiology.
- **Comparing** change over time and numbers between groups is essential but a raw count doesn't account for population differences.
- Event numbers must always be reported relative to **the population at risk** to make fair comparisons.

$$\frac{number\ of\ cases}{population\ at\ risk}$$

# The Bathtub of Epidemiology

- **Prevalence** refers to the current active cases (or during a set period).
- **Incidence** refers to new onset cases in a given period.
- **Prevalence** increases when **incidence** is higher than the combined rate of **recovery** and **mortality**.
- There is a **time delay** between a decrease in incidence and a decrease in prevalence as today's cases will be tomorrow's recoveries.


Source : HSC Public Health Agency (2020)
https://www.publichealth.hscni.net/node/5277

# Content

1. Definition of Epidemiology
2. Foundation of Epidemiology
3. Incidence and Prevalence
4. **Exposure and Risk**
5. Causal Inference
6. Summary

# Exposure and Risk

- In epidemiology we try to track if an **exposure** is associated with an **increased risk** of a negative health outcome.
- This allows planning for future **health care provision** and targeting of **control measures** to mitigate harm.
- If we find exposure to **poor air quality** is associated with reduced **cognitive performance,** we may examine causes and attempt interventions to improve this.
- We may also examine if some **groups** are **more at risk** than others – urban/rural, white collar/blue collar, etc.

# Relative Risk

- We typically **compare the proportion** of people **exposed** to a hazard who have the negative health **outcome** with those who were **not exposed**.
- We call this **relative risk (RR)** where a RR = 1 means no increased risk, an RR > 1 means the exposure comes with increased risk of the outcome, and an RR < 1 indicates reduced risk.

| | Disease | No disease |
|---|---|---|
| Exposed | A | B |
| Not exposed | C | D |

$$relative\ risk\ (RR) = \frac{A/(A+B)}{C/(C+D)}$$

# Content

1. Definition of Epidemiology
2. Foundation of Epidemiology
3. Incidence and Prevalence
4. Exposure and Risk
5. **Causal Inference**
6. Summary

# Causal Inference

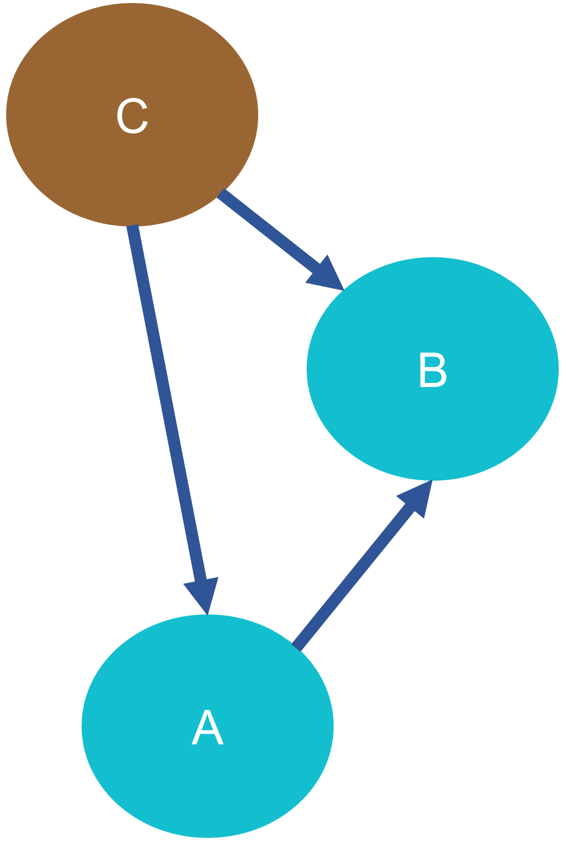
- It is one thing to say an **exposure** raised the **risk** of an **outcome**, but it is another thing to say an exposure **caused** an outcome.
- If the exposed group and non-exposed group are **identical** in all but the exposure, then we can conclude the exposure caused any differences in the outcome.
- In reality we can never have two truly identical groups so we can either:
  a) randomly assign individuals to an exposure so any group differences are **random**.
  b) **control** (statistically) for any differences.

# Statistical Controls

- My exposed group are **lower income** than my nonexposed group.
- I worry this would cause my exposed group to have poorer health and **bias my estimate** of the exposure effect.
- I could **break down** individuals by income band and see if richer exposed individuals have poorer health than richer nonexposed individuals.
- This would be an attempt to **control statistically** for income.

# Confounds and Directed Acyclic Graphs

- In epidemiology we use **directed acyclic graphs (DAGs)** to describe causal relations and to decide what to **control** for.
- Each variable gets a node and if variable A **causes** variable B, we draw an **arrow** originating at node A and terminating at node B.
- Here C **causes both** A and B.
- If we don't control for C, it will bias the estimate of A's effect on B.
- We call C a **confound** and we always control for variables like C.



# Randomised Controlled Trials (RCTs)

- To get the right answer through **statistical controls**, we have to correctly identify **all possible confounds**.
- In RCTs individuals are randomly assigned to the treatment/exposure meaning any differences in outcome are either **random chance** or **caused by the treatment**.
- The rate of random positive findings is known so we can assign a **probability** that a **positive finding occurred through chance**.
- This ability to **control for unknown biases**, makes RCTs the **gold standard** in causal inference.

# Content

1. Definition of Epidemiology
2. Foundation of Epidemiology
3. Incidence and Prevalence
4. Exposure and Risk
5. Causal Inference
6. **Summary**

# Summary: Key Takeaways

- **Epidemiology:**
  The study of patterns of exposure and health events in populations.
- **Incidence and prevalence:**
  Metrics that track rates of health outcomes and exposures, helping identify trends and plan interventions.
- **Exposure and risk:**
  We identify factors that increase the chance of negative health outcomes to understand causes and guide prevention.
- **Causation and controls:**
  Elevated risk alone doesn't prove causation; we control for biases to determine true causal links.
- **Randomised controlled trials:**
  Well-designed trials minimise bias and are the gold standard for testing intervention effects.

# Consequent Modules

**Future training videos**:

- Additional legislations may be covered by ISES Europe in specialised training videos.

# Thank You!

We appreciate your participation and attention

Congratulations on completing the ISES Europe training videos.

Access all videos via: https://ises-europe.org/

# Further Reading

**Introduction to Epidemiology**

*Centres for Disease Control and Prevention (CDC)*

https://www.cdc.gov/training-publichealth101/php/training/introduction-to-epidemiology.html

**Exposure Assessment in Environmental Epidemiology**

*National Research Council (US – NRC)*

https://www.ncbi.nlm.nih.gov/books/NBK233635/