Introductory session

Aim

This course will introduce students to more complex epidemiological concepts and advanced methods employed in modern epidemiological research.

Objectives

1. To develop a stronger understanding of causation in epidemiology
2. To recognise important biases and understand how these affect interpretation of findings, as well as how such biases can be dealt with through study design and/or statistical analysis
3. To develop a critical understanding of how quantitative methods can be used to apply effect measures to target populations
4. To develop a critical understanding of the principles underpinning specialist areas of epidemiology, such as life-course epidemiology

Learning outcomes

1. Both critique and design epidemiological research informed by an understanding of counterfactual thinking and causal diagrams, including a critical understanding of the limits of these approaches.

2. Recognise important biases and understand how these affect interpretation of findings, understand how such biases can be dealt with through study design and/or statistical analysis and have a critical understanding of the relative strengths and limitations of different methodological approaches.

3. Critically understand how quantitative methods can be used to apply effect measures to target populations, as well as the assumptions such approaches require.

4. Critically understand the major methodological issues in natural experiment studies, administrative data analyses and life-course epidemiology and relate these to major theories across the wider field (i.e. collider bias, confounding etc).

Emphasis of course

- Understanding assumptions
- Making connections
  - *Different statistical approaches*
  - *Different terminologies within epidemiology*
- Interpretation and partnership
- Get a feel for magnitude and importance of biases

Causal notation

Outline

Causes

Counterfactuals

The notation for counterfactual models

Exchangeability

Randomisation

Conditional exchangeability

Standardisation

Causal notation: what we'd like to know

	A (the ice)	Y|A=0	Y|A=1
Fergus *(f)*	1		1
Graeme *(g)*	1		1

1 = yes - factor present, or outcome happened
0 = no – factor absent, or outcome did not happen

$Y_i|A=1$ means the outcome (Y) given that the exposure status for person i was actually 1 (exposed)

$Y_i^{a=1}$ means the outcome (Y) for person i when their exposure status (perhaps counter to the fact) is set to 1 (ie exposed)

Summing up

Counterfactuals help us to be clear about what we mean by saying that something was causal

It's therefore important to be clear about what counterfactual we are considering

eg "if no-one was obese", or "if everyone had a BMI of 25" are different counterfactuals when we are thinking about the causal effect of abnormally high weight

Summing up

We can’t observe counterfactual outcomes

However if we assume that groups are exchangeable, then we can use the outcome in the other group to estimate the unobserved counterfactual

Randomisation gives us some assurance that the groups are exchangeable

Summing up

In observational research to draw causal conclusions we have to assume that groups are either exchangeable or at least conditionally exchangeable (conditional on some third factor)

We can deal with this by standardising on that third factor or using other similar approaches to adjust for other factors

Summing up

However when randomisation isn't involved, we always have to consider whether there might be other factors that threaten exchangeability

In traditional terms, could our group comparison be confounded?

Conclusion

the counterfactual framework is part of everyday speech and being clear about counterfactuals adds rigour and clarity to discussions about causality

the notation may seem complicated

however it gives us a useful mathematical language to discuss causation

Causal diagrams

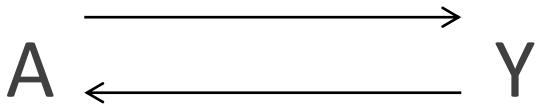
Usefulness of causal diagrams

- Identifies variables relevant to research
- Summarize knowledge
- Visualize assumptions
- Graphic representation of causal network
- Enhance communication among researchers

Directed acyclic graphs (DAGs)

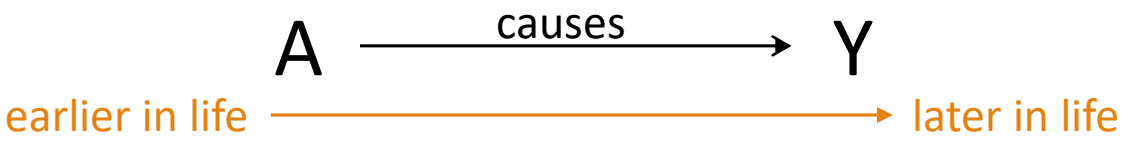
- Directed: Edges (arrowheads) imply a direction

- Acyclic: A variable cannot cause itself



Visualization of DAGs

- Presence of arrow:
  - We assume direct causal effect
  - We are not willing to assume that causal effect does not exist
- Absence of arrow = strong assumption:
  - We are willing to assume that causal effect does not exist
- Direction of arrow:
  - Assumed direction of effect
- Time flows from left to right
- We do not distinguish between harmful and protective effects



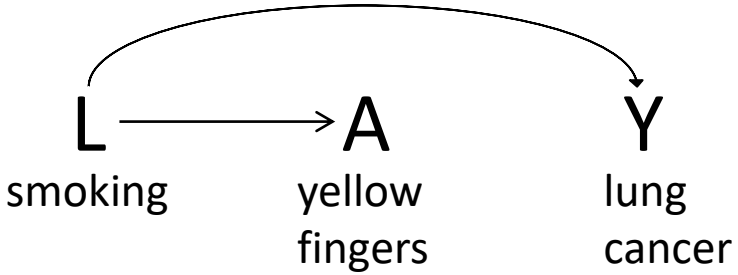
Conditioning

ADJUSTMENT	STRATIFICATION	RESTRICTION
"In our statistical analysis we adjusted our results for smoking" Effect estimates: Unadjusted (crude): ◦ OR 1.22 (1.14-1.56) Adjusted for smoking: ◦ OR 1.15 (1.05-1.25)	"We report our results separately for smoker and non-smoker" Effect estimates: Smoker: ◦ OR: 1.48 (1.36-1.57) Non-smoker: ◦ OR: 1.21 (1.15-1.34)	"We only selected participants that were smokers" Effect estimates: Smoker: ◦ OR: 1.48 (1.36-1.57)

Open and blocked paths – non-collider

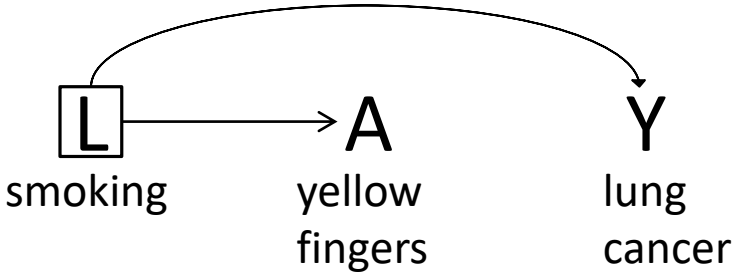
OPEN PATH

- No conditioning on non-collider (common source)



BLOCKED PATH

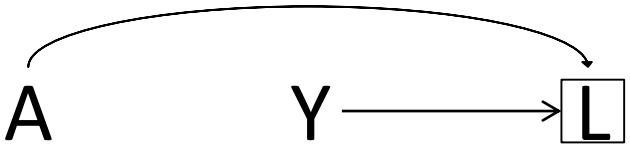
- Conditioning on non-collider (common source)



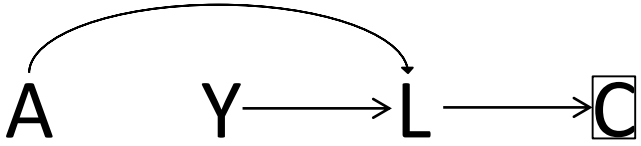
Open and blocked paths – collider

OPEN PATH

- Conditioning on collider (common effect)

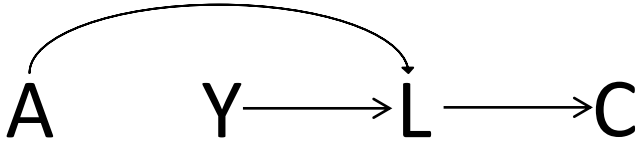


- Conditioning on effect of collider (common effect)

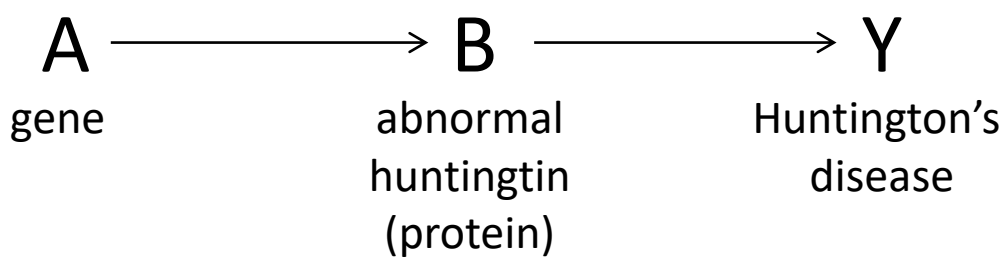


BLOCKED PATH

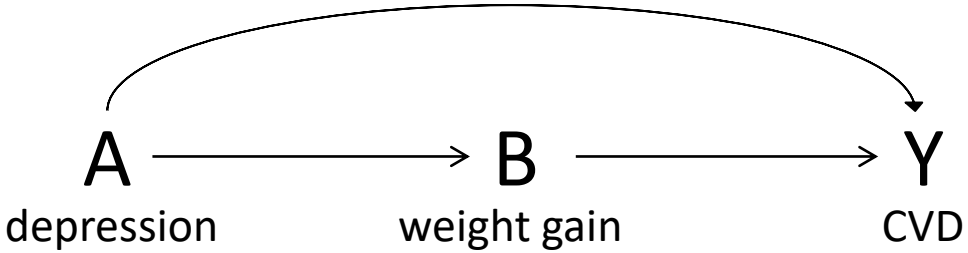
- No conditioning on collider AND no conditioning on effect of collider



Mediation – Estimation of total effect

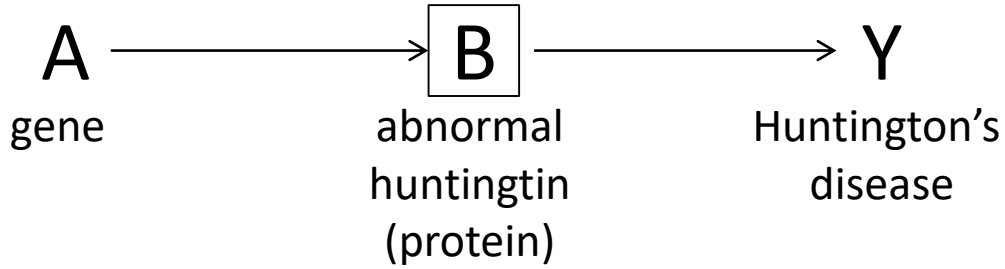


complete mediation

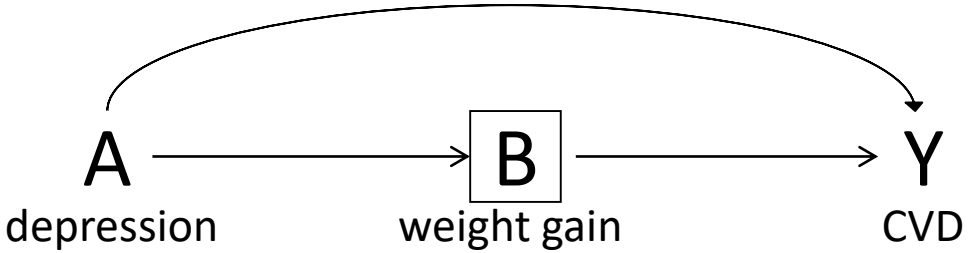


incomplete mediation

Mediation – Estimation of indirect and direct effect



complete mediation



incomplete mediation

Estimation

Estimating causation Lectures 4 and 5

Regression/propensity scores

- Regression models
- Propensity weighting

Assumption
No unmeasured confounding

Non-regression approaches

- **Difference-in-difference**
- **Interrupted time series**
- **Synthetic controls**
- **Regression discontinuity**
- **Instrumental Variables**

Assumptions
~~No unmeasured confounding~~
Various

Routine data

Learning objectives

- Gain an understanding of
  - Sources of routine data that may be useful for public health research
  - Advantages and limitations of routine data
  - Data linkage
  - Ethical considerations when using routine data
  - How to access routine data for research purposes

Scale and interaction

Scale and affect measure modification/interaction

- Scale is everything
- EMM is just about perspective, interaction implies EMM, EMM implies interaction
- No controversy about calculation or statistical definition
- Controversy about interpretation
  - Interaction on additive scale – public health/clinical importance
  - Interaction on relative scale – less common

Competing risks

Learning outcomes

- Understand relationship between rates and risks
- Understand concept of competing risks
- Get a feel for the impact of competing risks on outcomes and treatment effects

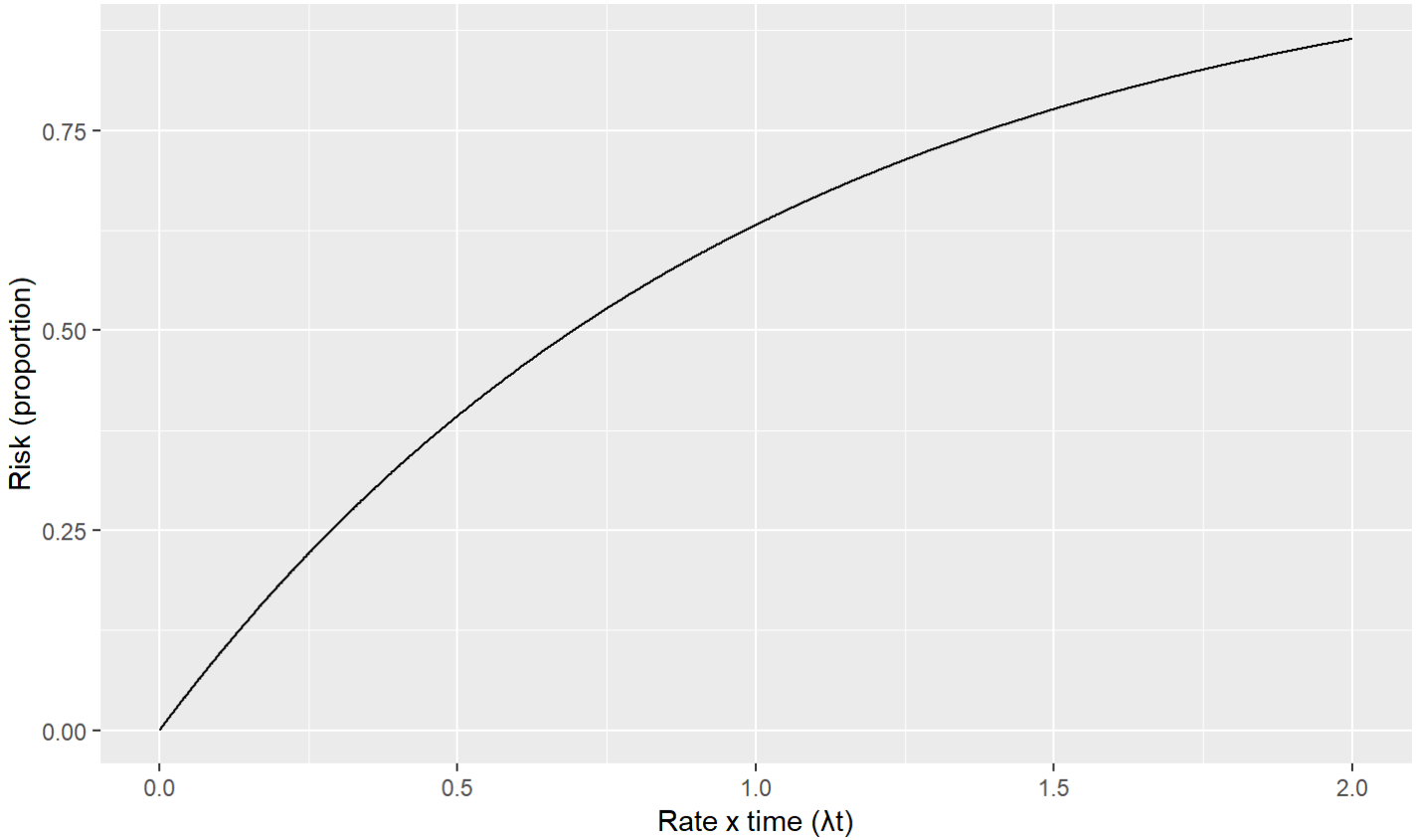
Overview

- Spend some time on rates and risks
- Proceed to competing risks
- Discussion, equations and interactive plots (via a Shiny app)

Cancer

- Trial of a cancer treatment
  - Death from relapse
  - Death from non-relapse related causes

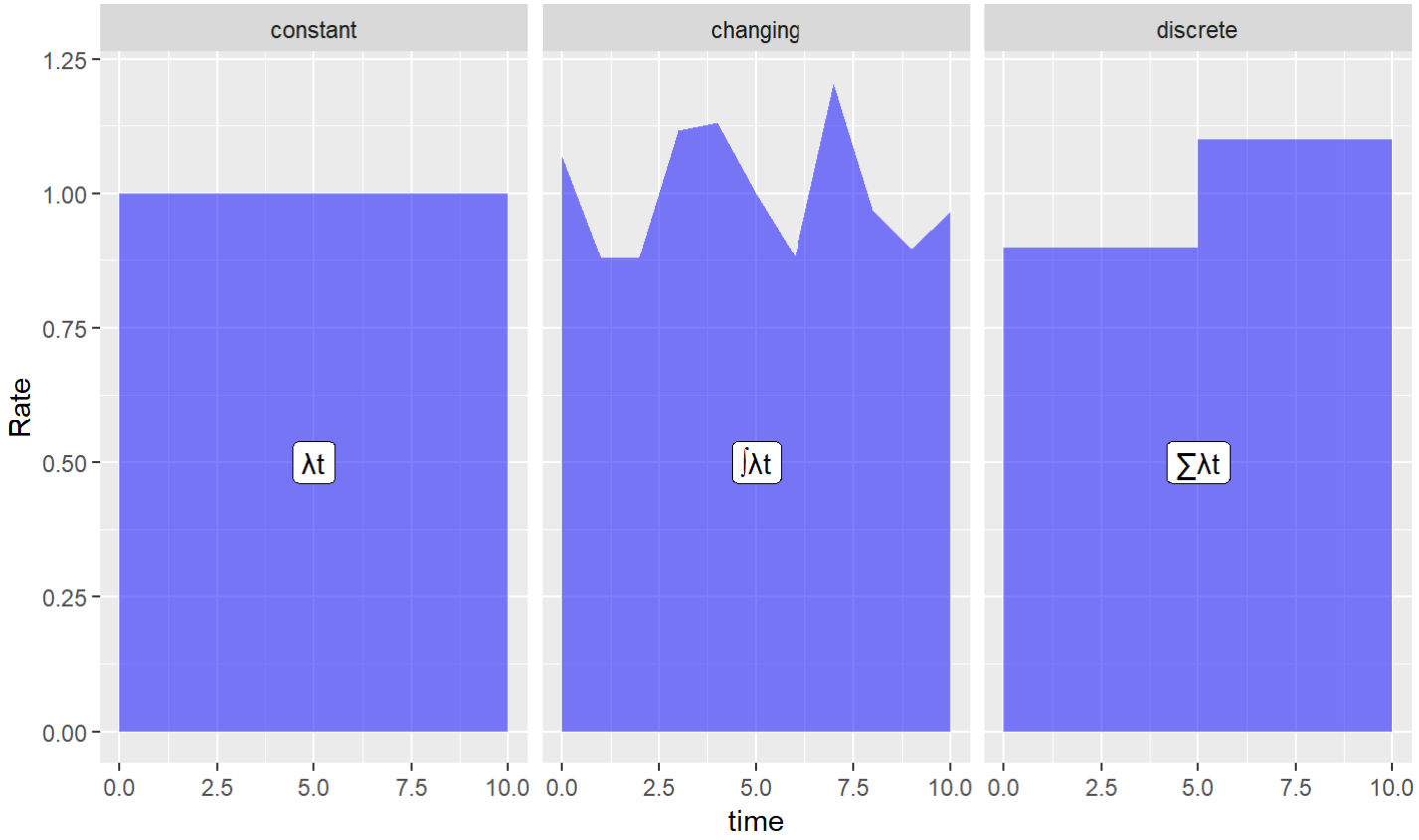
Risk and rates



Summary: without competing risks

- Rates lie between 0 and infinity
- Risks lie between zero and 1
- Risks can be estimated from rates (and vice versa)
- doubling rate and doubling time have the same effect on the risk
- The relationship between rates and risk is non-linear
- The relationship between time and risk is non-linear
- If the rate ratio is constant over time, the risk ratio will attenuate over time
- Rate ratios are **NOT** risk ratios

Area under the curve



Area under the curve

Cause-specific hazard ratios

- Can also use Cox regression to estimate the cause-specific hazard
- Same model, different interpretation
- Cannot directly translate to risk
- Instead combine the cause-specific hazards using the equation in previous slide to estimate the risk of each outcome
    - R packages such as mstate and msm allow combination of different models
    - Rely on simulation or bootstrapping to get 95 % confidence intervals

Modelling cumulative incidence directly

- Can also estimate the cumulative incidence directly using Fine and Gray model
- Produces regression coefficients for effect of a predictor on the sub-distributional hazard rate
- Unlike the hazard rate from a Cox model this has no natural interpretation

Direction of effect on cause-specific hazard rates and risk

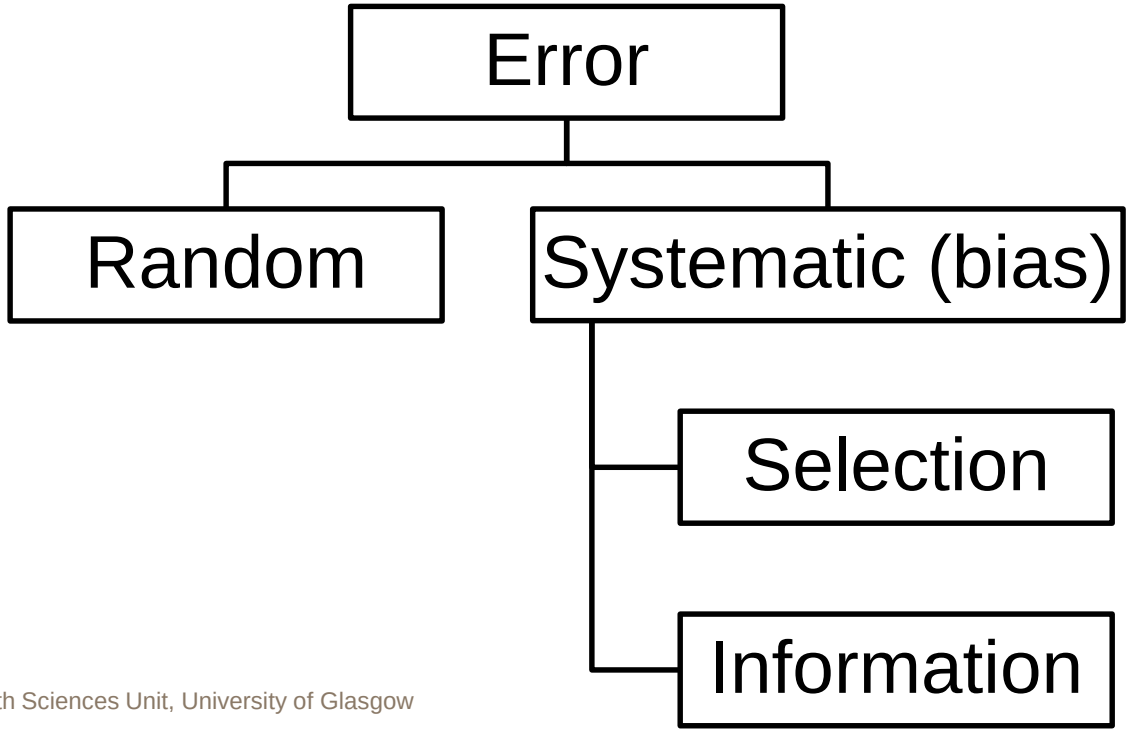
Prognostic score	Relapse Cox	Relapse Fine and Gray	Death Cox	Death Fine and Gray
Very low	1	1	1	1
Low	1.01 (0.81-1.27)	0.93 (0.75-1.16)	1.57 (1.25-1.97)	1.56 (1.24-1.96)
Medium	1.28 (1.03-1.59)	1.07 (0.87-1.33)	2.01 (1.61-2.52)	1.94 (1.55-2.42)
High	1.57 (1.25-1.99)	1.17 (0.93-1.48)	2.68 (2.12-3.37)	2.48 (1.96-3.12)
Very high	2.67 (2.06-3.47)	1.55 (1.19-2.02)	3.98 (3.09-5.13)	3.27 (2.5; .1.22)

Measurement bias

Aims/Learning Outcomes

Recognise important biases and understand:

- how these affect interpretation of findings
- how such biases can be dealt with through study design and/or statistical analyses.



Types of Misclassification

Non-differential misclassification

- Misclassification is not related to exposure or disease status
- Probability of incorrect information is equal in the two group

Differential (systematic) misclassification

- Misclassification results in incorrect exposure being recorded in more cases than controls, or vice versa
- One group has more incorrect information than the other

Applying epidemiological estimates

Population Attributable risk

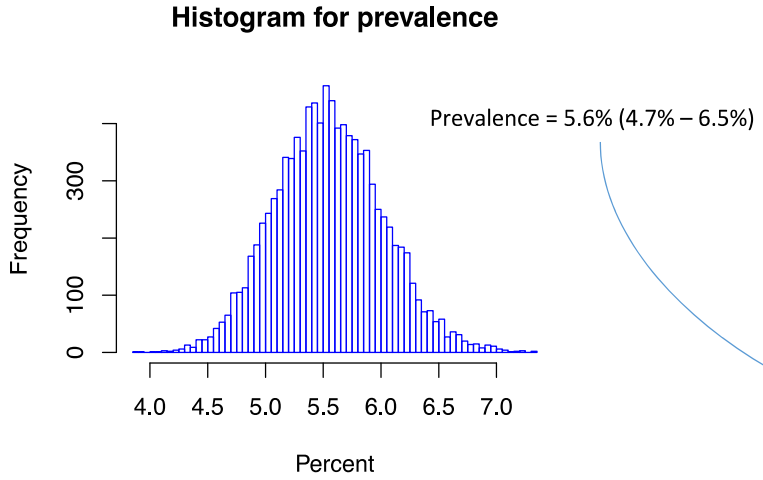
Population Attributable Risk (or Population Attributable Fraction) indicates the number (or proportion) of cases that would not occur in a population if the factor were eliminated

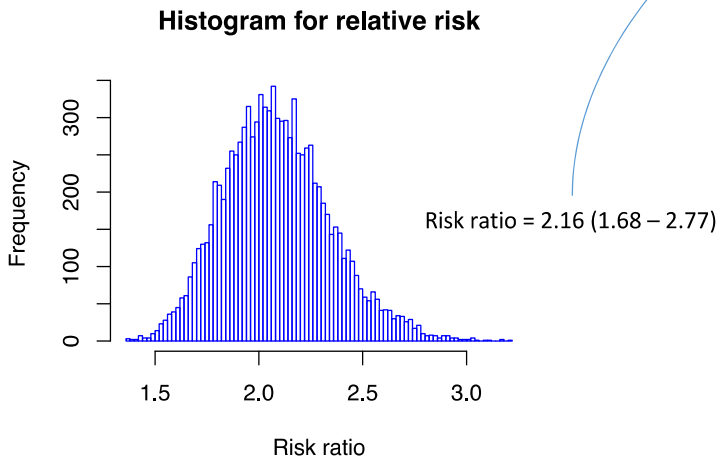
So for this metric we need one additional parameter?

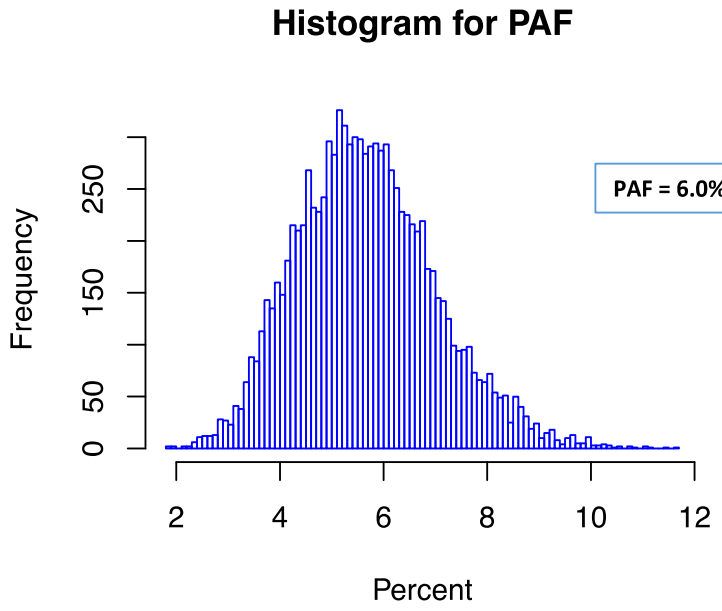
Prevalence

$$\mathbf{PAR = P_e (RR_e\text{-}1) \ / \ [1 + P_e (RR_e\text{-}1)]}$$

Kenya
Illustration for deriving population attributable fraction (PAF)







$$Population\ Attributable\ Fraction = \frac{Prevalence * (RR - 1)}{1 + Prevalence * (RR - 1)}$$