Week 3 – Formulating hypotheses, error, and bias

Definitions

- Determinants of health = exposure = treatment = X (eg. Smoking, exposure to UV light)
- Health related states and events = outcome = disease = Y (eg. Cancer, mental illness)
- Descriptive epidemiology: hypothesis generation (should be quantitative, specific and feasible)
- Analytical epidemiology: hypothesis testing
- Null hypothesis: no association between exposure and outcome (if true, retain the null)
- Alternative hypothesis: there is an association between exposure and outcome (if true, reject the null)

What factors affect an association between exposure and outcome?

- Accuracy, bias, confounding, chance.
- Bias: the difference between results and population value due to incorrect measurements being taken or measurements being taken on a non-representative sample, repeating the study would reach the wrong answer on average, used to judge the quality of evidence
    - Selection bias: systematic difference between the baseline characteristic of the groups compared (eg. Measure for morbidity at an office is flawed because people need to be relatively healthy to be working in the first place, therefore the sample will indicate that morbidity is lower than it actually is)
    - Measurement bias: a systematic error in the measure of information on the exposure or outcome (eg. Using an instrument that is incorrectly calibrated)
    - Recall bias: systematic error caused by the differences in the accuracy or completeness of the recollections retrieved by study participants regarding events or experiences from the past (eg. People who have the disease may search their memories more thoroughly than unaffected controls, people are also more likely about risk factors that are embarrassing)
- Confounding: situation in which a non-causal association between a given association is observed due to the influence of a third variable (eg. Coffee drinkers are more likely to develop lung cancer – this is not because coffee is inherently a risk factor for lung cancer, but due to coffee drinkers being more likely to smoke, which is a risk factor for lung cancer), how to control for confounding?
    - Design stage
        - Randomisation: equal distribution of groups
        - Restriction: only non-smokers, males etc
        - Matching: match for age, sex, social class (effective but time consuming and expensive)
    - Analysis stage
        - Stratification: two tables of exposure vs outcome – one for each level of confounder
        - Statistical modelling: can adjust for multiple things
- Bias creates an association that is not true, whereas confounding describes an association that is true, but potentially misleading
- Bias cannot be quantitatively evaluated, but confounding and chance can

Exposure

- Can be measured quantitatively – level and duration
- Types of exposure

    - Cause acute effects soon after exposure starts (London smog)
    - Effects only after long period of exposure (cadmium, noise)
    - Can be exposure level (number of cigs/day) or combined duration and exposure (pack-years)

Dose

- Individual dose: one-person dose, can vary over time, exposure and distribution
- Population dose: sum of individual doses, dose commitment/population dose
- Dose-effect: relationship between dose and severity of symptoms, effects in one individual
- Dose-response: as dose increases, % of cases with condition increases, looks at population

Error

- Random error – due to chance alone
    - Sampling error: reduce by increasing sample size
    - Measurement error: reduce by strict protocol, precise measurements
- Systematic error – results differ from true values
    - Selection bias: characteristics of participants different to population
    - Measurements bias: labs produce difference results
    - Recall bias: cases more likely to recall past exposure than controls

Reliability and validity

- Validity: test capable of measure what it is intended to measure (questions not open to interpretation)
    - No systematic error, random error small
    - Internal validity: are results correct for the group being studied?
    - External validity: how do results generalise/apply to those not in it?
- Reliability: stability and reproducibility over time
    - Stability: consistent answers to questions over time, if you ask someone a question will they provide the same answer three weeks later?
    - Reproducibility: will same answers results from the same questions if the interviewer is changed?