

Errors in epidemiological studies

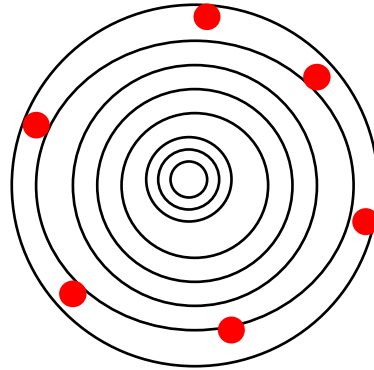
What can go wrong ??

- sampled from the target population ?
- sample large enough ?
- measured precisely ?
- measured accurately ?
- disease correctly diagnosed ?
- absence of disease correctly diagnosed ?
- exposure accurately determined ?
- measurement error same in different groups ?
- groups comparable ?

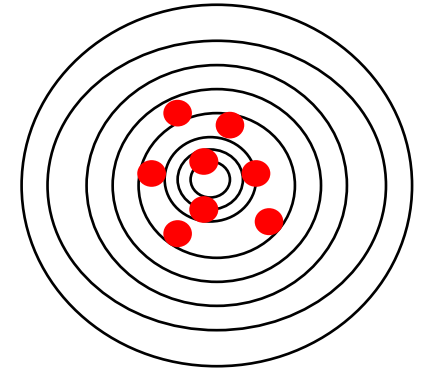
- **Errors can be differential (systematic) or non-differential (random)**
- **Random error: use of invalid outcome measure that equally misclassifies cases and controls**
-
- **Differential error: use of an invalid measures that misclassifies cases in one direction and misclassifies controls in another**
- **Term '*bias*' should be reserved for differential or systematic error**

Types of Error

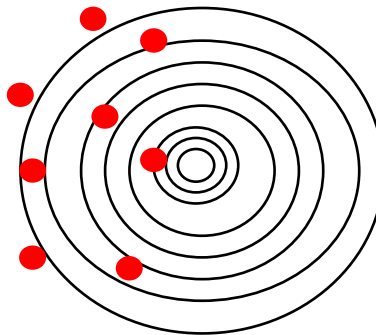
- 'random error'
= imprecision
- 'systematic error'
= bias



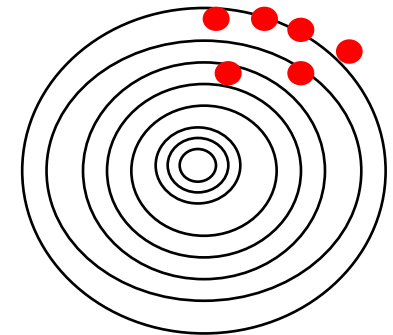
unbiased + unprecise



unbiased + precise



biased + unprecise



biased + precise

Random error is the divergence due to chance alone, of an observation on sample from the true population value, leading to lack of precision in measurement of association

Random error

- **Sources of Random Error**
 - **Sampling error**
 - **Biological variation**
 - **Measurement error**

Sample Size Calculations

Variable to consider

- Required level of statistical significance of the expected result
- Acceptable chance of missing the real effect
- Magnitude of the effect under investigation
- Amount of disease in the population
- Relative sizes of the groups being compared

What is Bias?

- ***Any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth (Last, 2001)**
- ***A process at any state of inference tending to produce results that depart systematically from the true values (Fletcher et al, 1988)**
- ***Systematic error in design or conduct of a study (Szklo et al, 2000)**

What is meant by bias in research?

- **Bias is the term used to describe differences between the study findings and truth**

Bias is a systematic error in an epidemiologic study that results in an incorrect estimation of the association between exposure and health-related event

Systematic Error (Bias)

- **Occurs when there is a tendency to produce results that differ in a systematic manner from the true values**
- **A study with a small systematic error is considered highly accurate**
- **Accuracy is not affected by sample size**
- **Principal biases**
 - **Selection Bias**
 - **Measurement (Classification) Bias**
- **Confounding**

Types of Bias

Selection bias

- Unrepresentative nature of sample

Information (misclassification) bias

- Errors in measurement of exposure of disease

Confounding bias

- Distortion of exposure - disease relation by some other factor
- Types of bias not mutually exclusive
(*effect modification* is not bias)

Selection Bias

- **Occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not**
- **Distortion of effect resulting from the way participants are accepted into studies**
- **Healthy Worker Effect – risk for certain illnesses in industrial working populations is lower than in the general population**

Selection Bias

- Selective differences between comparison groups that impacts on relationship between exposure and outcome
- Usually results from comparative groups not coming from the same study base and not being representative of the populations they come from

Selection Bias Examples

- **Case-control study:**
- **Controls have less potential for exposure than cases**
- **Outcome = brain tumor;**
- **exposure = overhead high voltage power lines**
- **Cases chosen from province wide cancer registry**
- **Controls chosen from rural areas**
- **Systematic differences between cases and controls**

Selection Bias Examples

Self-selection bias:

- You want to determine the prevalence of HIV infection
- You ask for volunteers for testing
- You find no HIV
- Is it correct to conclude that there is no HIV in this location?

Selection Bias Examples

- Healthy worker effect:
- Another form of self-selection bias
- “self-screening” process – people who are unhealthy “screen” themselves out of active worker population
- Example:
 - - Course of recovery from low back injuries in 25-45 year olds
 - - Data captured on worker’s compensation records
 - - But prior to identifying subjects for study, self-selection has already taken place

Selection Bias Examples

- Also occurs before subjects are identified for study
- Diagnostic or workup bias:
- Diagnoses (case selection) may be influenced by physician's knowledge of exposure
- Example:
 - - Case control study – outcome is pulmonary disease, exposure is smoking
 - - Radiologist aware of patient's smoking status when reading x-ray – may look more carefully for abnormalities on x-ray and differentially select cases
- Legitimate for clinical decisions, inconvenient for research

Sources of Selection Bias

- **Volunteers for studies are almost always selective**
- **Paid participants may be selectively different from the general population**
- **Hospital and clinical data are based on a selective population**
- **Disease or factor under investigation makes people unavailable for study**

Measurement Bias

- **Occurs when individual measurements or classifications of disease or exposure are inaccurate**
- **Sources:**
 - **Quality of laboratory analysis**
 - **Recall bias**

- **Information / Measurement / Misclassification Bias**
- Method of gathering information is inappropriate and yields systematic errors in measurement of exposures or outcomes
- If misclassification of exposure (or disease) is unrelated to disease (or exposure) then the misclassification is *non-differential*
- If misclassification of exposure (or disease) is related to disease (or exposure) then the misclassification is *differential*
- Distorts the true strength of association

Sources of information bias

- **Subject variation**
- **Observer variation**
- **Deficiency of tools**
- **Technical errors in measurement**
-

Information / Measurement / Misclassification Bias

- **Recall bias:**
- **Those exposed have a greater sensitivity for recalling exposure (reduced specificity)**
- **- specifically important in case-control studies when exposure history is obtained retrospectively**
- **cases may more closely scrutinize their past history looking for ways to explain their illness**
- **- controls, not feeling a burden of disease, may less closely examine their past history**

Information / Measurement / Misclassification Bias

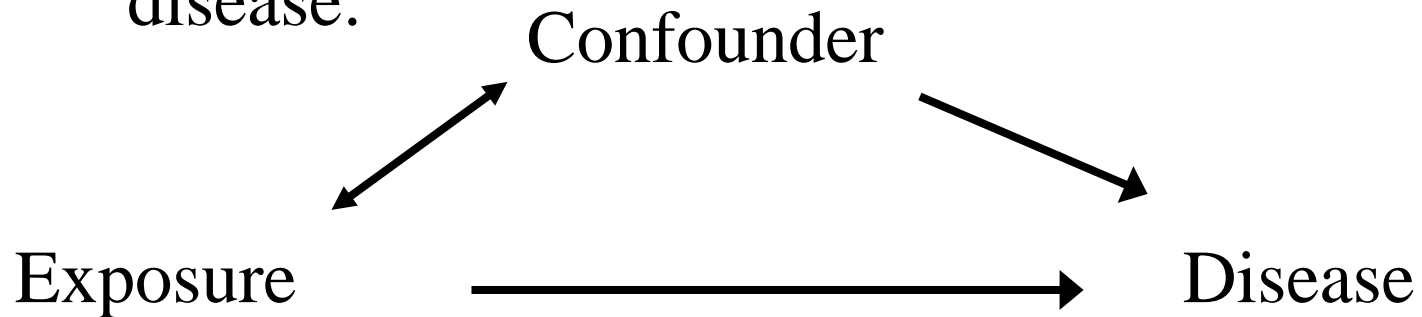
- Reporting bias:
- Individuals with severe disease tends to have complete records therefore more complete information about exposures and greater association found
- Individuals who are aware of being participants of a study behave differently (Hawthorne effect)

Confounding

- **Occurs when another exposure exists in the study population and is associated with both the disease and the exposure being studied**
- **When the effects of two exposures (risk factors) have not been separated, and incorrect conclusions are drawn that the effect is due to one rather than the other variable**

Confounding

- Effect of a third variable that accounts for all or some of the association between exposure and disease:



Definition of Confounding

A confounder:

- 1) Is associated with exposure**
- 2) Is associated with disease**
- 3) Is NOT a consequence of exposure (i.e. not occurring between exposure and disease)**



Confounding

- **Confounder is not a result of the exposure**
 - e.g., association between child's birth rank (exposure) and Down syndrome (outcome); mother's age a confounder?
 - e.g., association between mother's age (exposure) and Down syndrome (outcome); birth rank a confounder?

Confounding

- **To be a confounding factor, two conditions must be met**
- **Be associated with exposure**
 - **without being the consequence of exposure**
- **Be associated with outcome**
 - **independently of exposure (not an intermediary)**

Controlling for confounders

- **RESTRICTION** of subjects according to potential confounders (i.e. simply don't include confounder in study)
- **RANDOM ALLOCATION** of subjects to study groups to attempt to even out unknown confounders
- **MATCHING** subjects on potential confounder thus assuring even distribution among study groups