A brief overview of bias

Dr James Church
Wellcome Trust Clinical PhD Fellow
Zvitambo Institute for Maternal & Child Health Research
Bias in the everyday
Epidemiological inference

Exposure or Host Characteristic

Is an association observed?

Disease or Other Health Outcome

Is the association valid/true?
Three sources of error

1. Is an association observed?
2. Exposure or Host Characteristic
   - Disease or Other Health Outcome
3. Is the association valid/true?
   - Bias?
4. Random error?
5. Confounding?
What is scientific bias?

• Bias is any trend or deviation from the truth in data collection, data analysis, interpretation and publication which can give rise to false conclusions.

• It does not imply prejudice or deliberate deviation, but the deviation is systematic and non-random.
Bias is bad news!

• Error in the design or conduct of a study
• Not much can be done about it once the study is over!
• Studies have practical and ethical constraints so some bias is almost inevitable.
Bias in three parts
1) Selection bias

**Concerns the people included or compared**

... such that selection of individuals or groups does not achieve randomisation

a. Sampling bias
b. Ascertainment bias
c. Attrition bias (loss to follow-up)

Who is selected and how are they selected?
1) Selection bias

• **Sampling bias**

• When some members of the intended population are less likely to be included than others

• Results in a non-random sample
1) Selection bias

- Sampling bias – pneumonia and alcoholism
- *In the community*

\[
OR = \frac{D_e}{H_e} \div \frac{D_n}{H_n}
\]

\[
OR = \frac{10}{10} = \frac{10 \times 90}{90 \times 10} = 1.0
\]

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Alcoholism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

OR = \frac{D_e}{H_e} / \frac{D_n}{H_n}
1) Selection bias

- Sampling bias – pneumonia and alcoholism
- *In the hospital*

\[
OR = \frac{D_e / H_e}{D_n / H_n}
\]

\[
OR = \frac{20 / 10}{80 / 90} = \frac{20 \times 90}{80 \times 10} = 2.25
\]

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>No</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>Alcoholism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>No</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Sampling bias in case-control studies

Exposed (20)  
Not exposed (10)  
Exposed (15)  
Not exposed (15)

Disease  
(Cases)

No disease  
(Controls)

Population

Time
Sampling bias in case-control studies

Exposures of interest influence the likelihood of an individual becoming a control

Biased assessment of exposure odds in the population from which the cases are drawn
Sampling bias in case-control studies

Examples:

• Case-control study of cancer of the oesophagus and alcohol
  Controls: Men employed in a brewery

• Case-control study of stroke and oral contraceptives
  Controls: Women who attended a family planning clinic

The major problem in case-control studies is the choice of CONTROLS
How to select controls in case-control studies

• Do they reflect all people without the disease?

• Typical sources for control population
  • Hospital based?
  • Population based?
  • Defined subset of population?

• Trade off between convenience and introducing error

• Key to identify potential sources of error
How to select **cases** in case-control studies

- Is the population generalisable to all patients with the disease?
- Is the severity of disease among these patients representative?
- Do cases at different levels of selection have different exposure profiles??

- E.g. epidemiology of hip fracture in Harare
1) Selection bias

- **Ascertainment bias**
- When exposed cases are more (or less) likely to be selected for the study than unexposed cases
- E.g. studies of uterine cancer in the early 1970’s
  - They found a strong association with exogenous oestrogens (HRT)
  - Exogenous oestrogens cause uterine bleeding regardless of whether they cause endometrial cancer
  - Uterine bleeding result in women undergoing gynae investigations and may reveal endometrial cancers that would otherwise have gone undetected
1) Selection bias

- **Attrition bias**
- Systematic difference in withdrawals and exclusions between groups
- Loss to follow up can occur if
  - Treatment has been successful
  - Control group unhappy with lack of progress
2) Information bias

Concerns the measurements made

a. Misclassification
b. Recall bias
c. Observer bias
d. Performance bias
2) Information bias

• **Misclassification**
  • Can occur with anything you measure
  • Applies to exposure and/or disease outcome
    • Know the exposed group so look harder for the disease in this population
    • Know who is a case so probe for more information on exposure

• Non-differential (random)
  • Equal misclassification
  • Bias measure of effect towards null

• Differential (non-random)
  • Non equal misclassification of exposure/outcome
  • Bias measure of effect either way
Information bias

• **Recall bias**
  • When probability of recall is affected by disease status
  • Main form of bias in case-control
  • “Why did it happen to me?”
Information bias

• **Observer bias**
  • Tendency of humans to see what we expect/want to see
  • Can be conscious or unconscious
Information bias

• **Performance bias**
  • Occurs when behaviour change varies depending on group allocation
  • Can apply to participants or caregivers
3) Results bias

**Concerns reporting & dissemination of results**
- a. Outcome reporting bias
- b. Spin or selective focus
- c. Publication bias
- d. Citation bias
3) Results bias

• **Outcome reporting bias**
  • Statistically significant outcomes preferred
  • Subsets of data presented
  • Omission of outcomes
  • Data underreported
3) Results bias

- **Publication bias**
  - Mistaken emphasis on “significant” results i.e. P value < 0.05
  - Leads to overestimation of a treatment effectiveness
  - Small studies may not detect a beneficial effect

“File drawer effect”
Assessing publication bias

• Funnel plots = scatter plot of treatment effect (x-axis) versus standard error of treatment effect (y-axis)
• Funnel asymmetry points to publication bias
• Egger’s test to compute statistically
3) Results bias

• **Spin or selective focus**
  • More commonly associated with public relations & media
  • Can make research seem more convincing than warranted
  • Examples include
    • Detracting from non-significant results
    • Inappropriate use of causal language
    • Abstract article mismatch
3) Results bias

- **Citation bias**

  “The conversion of hypothesis to fact through citation alone”
  - Stephen Greenberg

- Statistically significant results more often cited
- Studies with non-significant results less visible
Cumulative effect of biases at the tail end
How can we mitigate bias?

- Crucially - minimise bias in the design
  - REMEMBER: it can **not** be controlled or adjusted for in the analysis
  - It can be quantified but data rarely available to do this
Examples of mitigating bias in study design

- Blinding of outcome assessors → Observer/detection bias
- Open reporting loss to follow up → Attrition bias
- Careful randomisation → Sampling bias
- Blinding of participants → Performance bias
- Pre-specified trial outcomes → Reporting bias
- Careful choice of control group → Sampling bias
- Intention-to-treat analysis → Attrition bias
Assessing bias of trials in a systematic review

- Tools to summarise risk of bias (RevMan)

![Risk of bias chart]

- Random sequence generation
- Allocation concealment
- Blinding of participants and personnel
- Blinding of outcome assessment
- Incomplete outcome data
- Selective reporting
- Other bias

Legend:
- Low risk of bias
- Unclear risk of bias
- High risk of bias
Assessing bias of trials in a systematic review

- Random sequence generation
- Allocation concealment
- Blinding participants/personnel
- Blinding outcome assessment
- Incomplete outcome data
- Selective reporting

- Selection bias
- Selection bias
- Performance bias
- Observer/detection bias
- Attrition bias
- Reporting bias
Any questions?
Summary

• Research is full of bias
• Bias results in a trend or deviation away from the truth
• Understanding bias and how to detect it allows you to validate and determine quality of scientific research
• Think of bias in three zebra parts
  1) Selection bias
  2) Information bias
  3) Results bias
Thank you for listening!