Errors in Research

Prof. Dr. Nadhim Ghazal Noaman

Learning Objectives

- To clarify types of errors and its important kinds.
- To demonstrate their consequences.
- To identify how to avoid making them.

Errors in research methods

IN GENERAL :

1. Random errors (chance)

2. Systematic errors (bias)

3. Others ….

Differences between systematic and random errors

- Bias is caused by systematic variation, while chance is caused by random variation.
- Systematic error primarily reduces measurement accuracy, while random error reduces measurement precision.
- It’s possible to avoid systematic error, but random error cannot be predicted although it can be overcome.

CONSEQUENCES of errors in research:

- Can cause distorted results and wrong conclusions.
- Can lead to unnecessary costs, wrong clinical practice and harm ...
- It is therefore the responsibility of all involved stakeholders in the scientific publishing ..

Random errors (chance)

- *Are random in nature …*
- *Not obvious …. Unpredictable …*
- *Beyond the control of the experimenter..*
- *May interfere with the results of the experiment.*
- *Random error is also called as **statistical error. Why?***

AVOIDED BY:

- *A measuring instrument with a higher precision …. Less fluctuations in measurements.*
- *Larger sample size.*
- *Statistical testing.*

Systematic errors (bias)

What is bias?

- Is a **Methodologically demanding** errors, that causing skewed inferences (false conclusions).
- Bias may be introduced at the design or analysis phase of a study.
- Bias … either ***intentionally or unintentionally***.
- Therefore, ***it is immoral and unethical*** to conduct biased research.
- Every scientist should thus be ***aware of all potential sources of bias*** and undertake ***all possible actions*** to reduce or minimize the deviation from the truth.

Major types of systematic error (Bias) include the following

- Selection bias
- Information bias
- Confounding bias

Selection bias

- Is a distorted association due to An error in choosing the individual to take part in study. Selection bias ***addresses internal validity*** of the inferences..

Selection bias

Sampling bias:

- a sample that does not accurately reflect the target population. especially by a nonprobability method.
- The sample content or it's prevalence of exposure are .. Not representative.
- A sample needs to ***be representative*** .. otherwise**..** the conclusions will ***be not generalizable***.
- ***Sampling bias affect external validity …***

Selection bias

- **Sampling bias**
- *Example: Interviewers conducting to select those respondents who are the most accessible and agreeable.*

- *Or a random sample composed of 70% females. This sample would not be representative of the general adult population and would influence the data* ***(need accurate randomization).***

Selection bias

- **Hospital patient bias (Berkson's Bias)**
- The population studied does not reflect the general population.
- May occur in a case-control study, when hospital controls are used. More exposure (like smoking).
- Some patients are less or more likely to enter the study than others.
- In that case, there will be ***under-represented subjects*** and those who are more likely to enter the study will be ***over-represented subjects***… skewed association
- **Is accounted by the:**

✓ **Sampling need to be random.**

✓ **Homogeneity of the population being studied.**

✓ **Size of the sample … Larger.**

Selection bias

Volunteer bias ..

- *E.g. if the aim of the study is to assess the average HsCRP (high sensitive C-reactive protein) concentration in healthy population, SO.. recruits only **healthy volunteer blood donors**, from a general population.*
- ***Volunteers .. are usually individuals who feel healthy. So underestimation of disease.***
- *or a study would include those participants **who might suspect to have** the disease (anemia).. anemic individuals might be over-represented.*

Selection bias

Attrition Bias: losses to follow-up

- i.e. when individuals leave the study before the end of follow-up. This will affect the power of the study according to its level.
- **usually because:**
  - Non eligibility
  - Non response: (from the beginning)
  - Non compliance
  - Drop out (later): die, leave, loss of interest,….

Is accounted by:

- Pilot project
- Education
- Motivation

Selection bias

- **Survivor bias**

subjects who died before the study end point might be missed from the study (cross sectional study).

Selection bias

- ***Non-response bias…***

When the participants can't or won't to answer the survey question.. Skewed results.

To avoid :

- e-mail survey,
- double check survey, to eliminate not-at-home respondents.

Information bias

- ***Information bias:*** *from systematic distortions when collecting information about exposures and diseases.*

- ***Observer bias/ Interviewer bias: due to***

1. lack of equal probing for exposure history between cases and controls.
2. lack of equal measurement of health outcome status between exposed and unexposed.

Solutions:

1. Blind data collectors regarding exposure or health outcome status
2. Develop well standardized data collection protocols
3. Train interviewers to obtain data in a standardized manner.
4. Perform pilot studies to identify problems with questionnaires and measuring instruments

Information bias

Memory bias … Recall bias

Recall bias .. either because:

- Unable to remember really
- Don't want to remember (afraid or ashamed)
- Not interested to remember
- Remember but in a defective way (over or under).
- Remember and change the facts (for many reasons)
- Ignorant

Controlled by:

- Use documents or reports
- Link with some events
- Add a case group unlikely to be related to exposure
- Add measures of symptoms or health outcomes unlikely to be related to exposure

Information bias

Measurement Error: is generated by the measurement process itself, and represents the difference between the information generated and the truth. **Its sources:**

- Instruments
- Individual variation (patient)
- Observers

Subjectivity bias: interventional studies

Controlled by:

- Accurate calibration of the instrument
- Masking of the drug
- Blinding: single, double and triple

Information bias

Misclassification bias

- Occurs when a disease of interest **is poorly defined** (not easy detectable)
- **No gold standard** for diagnosis.
- Some subjects are **falsely classified as cases or controls**.
- e.g. Early detection of the prostate cancer in asymptomatic men. Some early prostate cancer cases.. **misclassified as disease-free.**
- It may cause ***under- or over-estimation*** of the accuracy of this new marker.

Confounding

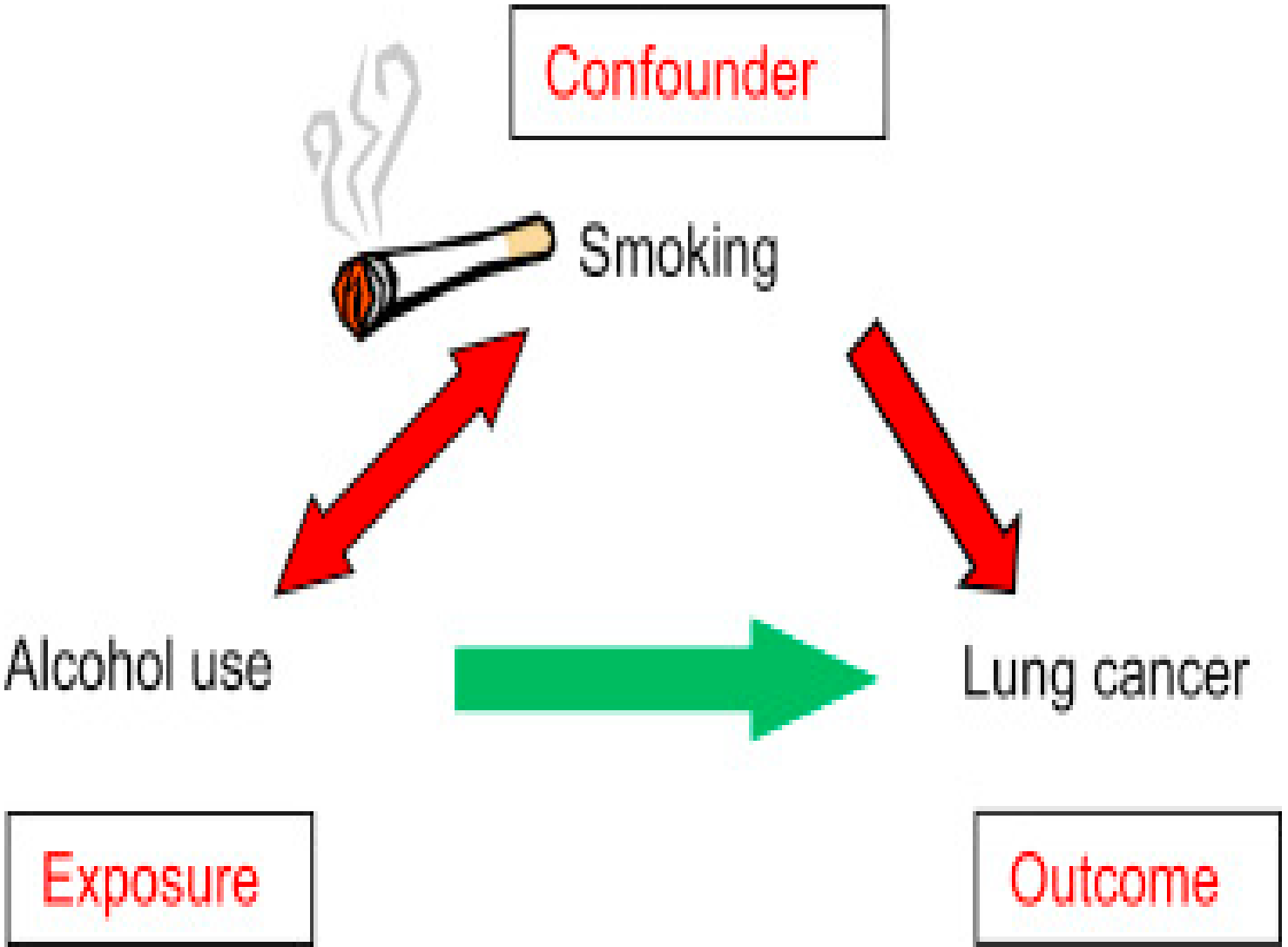
- is the distortion of the association between an exposure and health outcome by an extraneous, third variable called a confounder.
- **Criteria of confounders:**

1. Must be a known risk factor for the health outcome.
2. Must be associated with the main exposure, but not as a result of the exposure.

All confounders, should be working independently.

Confounding may be seen in which study design?

Confounding



Control of confounding

- **In Analysis phase:**

Stratified analysis (conditional logistic regression) or mathematical modeling.

- **In the Design phase:**

1- Restriction .. to restrict a study population to those without confounder.. (e.g. non-smoker)

2- Matching ..

3- Randomization, can control both known and unknown confounders (correct randomization).

In which design confounding can not be avoided?

Bias in data analysis

Occurs by analyzing data in a way which gives conclusions in favor of research hypothesis. such as by:

- Reporting non-existing data from experiments which were never done (***data fabrication***);
- ***Eliminating data*** which do not support your hypothesis
- Using ***inappropriate statistical tests*** to test your data;
- Performing multiple testing ***"fishing for P"***
- "***Torturing the data***" Use subgroup analyses, until association becomes statistically significant. ***not part of the original research hypothesis***,
- ***Besides being biased, unethical, invalid and illogical, those conclusions are also useless, since they cannot be generalized to the entire population.***

- E.g. . lactate concentration was positively associated with albumin concentration in a subgroup of male patients with a body mass index in the lowest quartile and total leukocyte count below 4.00×10^9/L.

Bias in data interpretation

By interpreting the results, we need:

- *Proper statistical tests were used,*
- *Results were presented correctly and*
- *That data are interpreted only if there was a statistical significance of the observed relationship.*
- *AVOID ..overgeneralization of the study conclusions to the entire general population, even if a study was confined to the population subset;*

Publication bias

- Unfortunately, scientific journals are much more likely ***to accept for publication a study which reports some positive.***
- Ideally, a scientific, well designed study NEED to be published regardless of the nature of its findings.
- What is the benefit of negative findings?
- *However, several journals have already been launched, such as Journal of Pharmaceutical Negative Results, Journal of Negative Results in Biomedicine, Journal of Interesting Negative Results.*