Common biases in collecting data

Cristian Tebé Cordomí, PhD

IGTP - Germans Trias i Pujol Research Institute and Hospital

2023-07-06

Catalogue of Bias

A collaborative project mapping all the biases that affect health evidence.

The Catalogue of Bias Collaboration started in June 2017. The Collaboration meets regularly to develop and refine the contents of the Catalogue and in between work on updating content.

https://catalogofbias.org/

Outline

- Bias, what is it?
- Types of bias (in collecting data)
- What can I do?

Bias, what is it?

Definition

bias *noun*[1]

1. a strong feeling in favour of or against one group of people, or one side in an argument, often not based on fair judgement.
2. an interest in one thing more than others; a special ability.
3. **the fact that the results of research or an experiment are not accurate because a particular factor has not been considered when collecting the information.**

[1] Oxford Dictionaries.

Statistical inference

What we do to estimate the average blood pressure of a population μ?

- we take a random sample of individuals.
- we estimate the average of the blood pressure in our sample.
- and the result should be equal to the true population mean blood pressure.

$$E(X) = \mu$$

Error

$$E(X) \neq \mu$$

Error is a deviation from the true value when measuring or estimating.

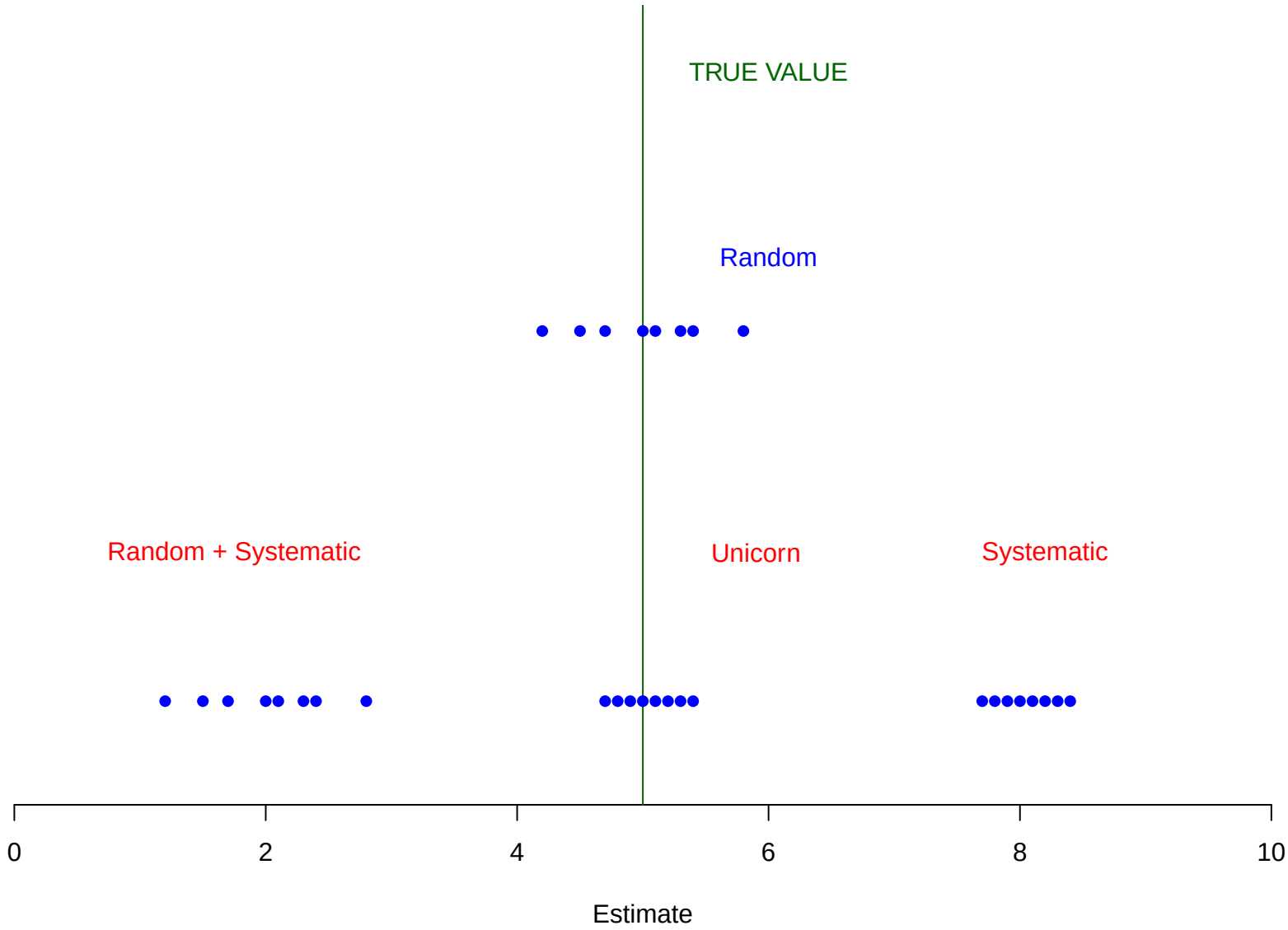
$$\text{error} = E(X) - \mu$$

$$\text{error} = \textit{random} + \textit{systematic}$$

Random component: variability introduced by chance factors in the measurement or estimation process.

Systematic component: consistent deviation from the true value in the same direction across multiple measurements or estimates.

Let's see bias



Let's see bias

What is the percentage of paved roads in Uganda?

What bias is?

A systematic distortion, due to a design problem, an interfering factor, or a judgement, that can affect the conception, design, or conduct of a study, or the collection, analysis, interpretation, presentation, or discussion of outcome data, causing erroneous overestimation or underestimation of the probable size of an effect or association[2].

[2] Jeffrey Aronson, Centre for Evidence-Based Medicine in Oxford's Nuffield Department of Primary Care Health Sciences.

Types of bias (in collecting data)

Main types of bias

- Selection bias
- Observer bias
- Misclassification bias
- Hawthorne effect
- Recall bias
- Immortal time bias
- Non-response bias

Selection bias

Definition: selection bias arises when included subjects in a study may differ from the target population.

Impact: selection bias can have an influence on the magnitude and the direction of the measured effect.

Selection bias

Cigarette smoking and dementia:

Some prospective studies that estimated the association between smoking and the incidence of Alzheimer disease and dementia found lower relative rate in smokers.

Selection bias due to censoring by death was one explanation for the lower relative rate of dementia and Alzheimer in smokers with increasing age.[3]

[3] Hernán MA, Alonso A, Logroscino G. Cigarette smoking and dementia: potential selection bias in the elderly. Epidemiology. 2008;19(3):448-450

Selection bias

Hormone replacement therapy (HRT) on coronary heart disease (CHD) in women:

Several studies showed that HRT reduced CHD, but subsequent randomized clinical trial showed that HRT might increase the risk of CHD disease.

The Women in the observational studies on HRT were more health conscious, more physically active, and had higher socioeconomic status than those not on HRT. This self-selection of women led to confounding and a "healthy-user bias".

Selection bias

Prevention:

- Exclusion/ Inclusion criteria
- Baseline comparison
- Intention-to-treat analysis
- Handling missing data
- Target population

Observer bias

Definition: Systematic difference between a true value and the value actually observed due to observer variation.

Impact: observer bias can have an influence on the magnitude of the measured effect.

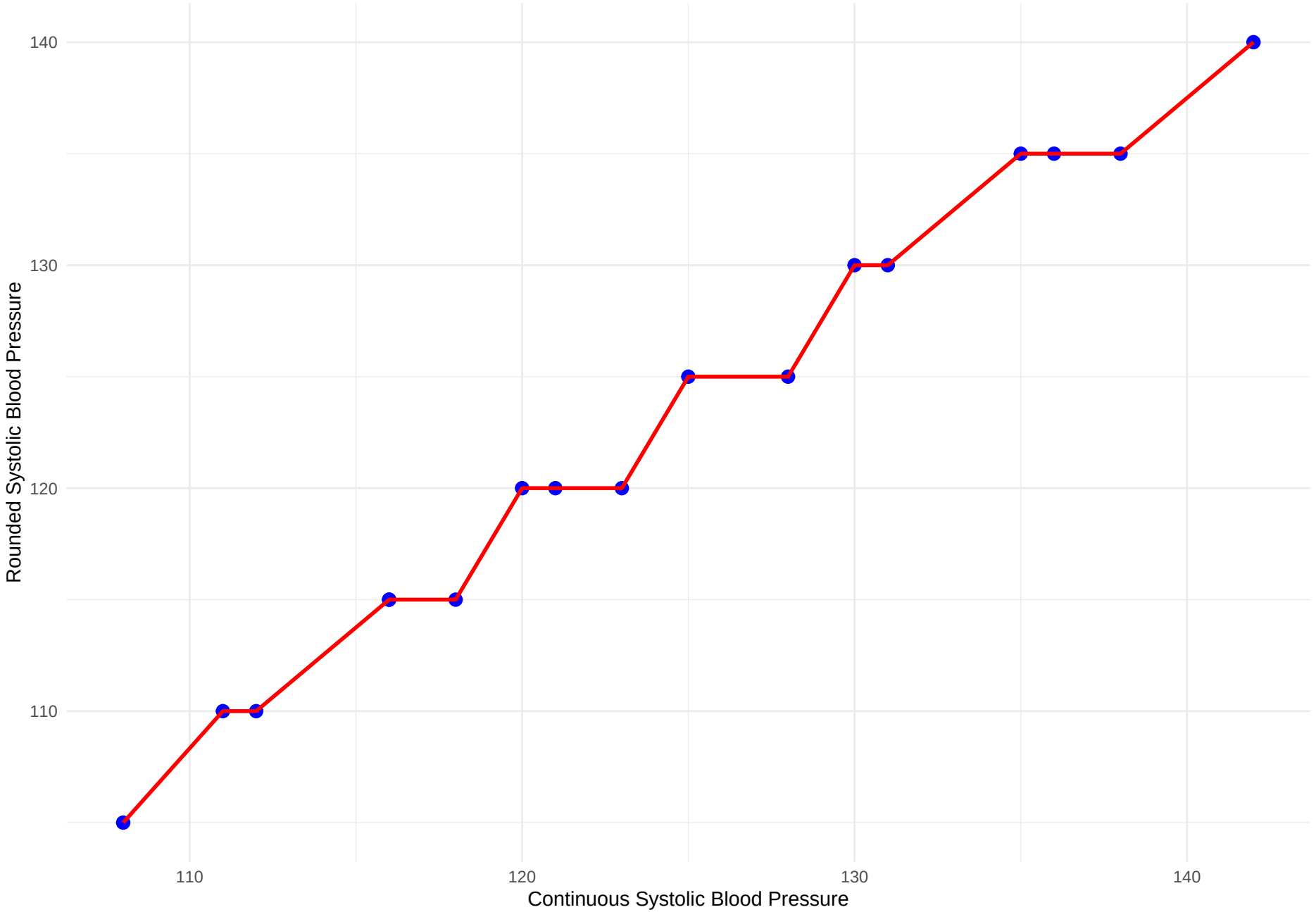
Observer bias

Blood preassure

Clinicians measuring participants blood pressure using mercury sphygmomanometers have been found to round up, or down, readings to the nearest whole number.[4]

[4] Stewart MJ, Padfield PL. Measurement of blood pressure in the technological age. Br Med Bull. 1994;50(2):420-442.

Observer bias



Observer bias

Depression assessment

An epidemiological study is planned to assess the prevalence of depression in a cohort of individuals after hospitalization for COVID-19. A group of psychologists will conduct face-to-face structured interviews and rating scales to collect data on participants' depressive symptoms.

Preconceived notions or expectations about the participants' mental health status may inadvertently influence the way psychologists interpret participants' responses and assess their depressive symptoms.

Observer bias

Prevention:

- Train researchers
- Measures standardization
- Multiple observers
- Blinded observers

Misclassification bias

Definition: Misclassification occurs when individuals are assigned to a different category than the one they should be in.

Impact: This can lead to incorrect associations being observed between the assigned categories and the outcomes of interest.

Misclassification bias

Body mass index (BMI) and mortality

Misclassification of body mass index (BMI) categories arising from self-reported weight and height can bias hazard ratios in studies of BMI and mortality.[5]

[5] Flegal KM, Kit BK, Graubard BI. Bias in Hazard Ratios Arising From Misclassification According to Self-Reported Weight and Height in Observational Studies of Body Mass Index and Mortality. Am J Epidemiol. 2018;187(1):125-134.

Misclassification bias

Prevention:

- Measures standardization
- Caution on categorisation

Hawthorne effect

Definition: Refers to individuals altering their behavior or responses when they know they are being observed, distorting the data collected.

Impact: Hawthorne effect can have an influence on the magnitude of the measured effect.

Hawthorne effect

Use of antiseptic handrub (AHR):

Medical personnel were monitored in 2 periods regarding compliance with AHR use when there were indications for AHR use. In the first period, the personnel had no knowledge of being observed. The second observation period was announced to the staff of the intensive care units in advance and information about what the observer would be monitoring was provided.[6]

[6] Eckmanns T, Bessert J, Behnke M, Gastmeier P, Ruden H. Compliance with antiseptic hand rub use in intensive care units: the Hawthorne effect. Infect Control Hosp Epidemiol. 2006;27(9):931-934.

Hawthorne effect

Use of antiseptic handrub (AHR):

Data were collected from 2,808 indications for AHR use. The overall rate of compliance was 29% (95% confidence interval, 26%-32%) in the first period and 45% (95% confidence interval, 43%-47%) in the second period.

The Hawthorne effect has a marked influence on compliance with AHR use, with a 55% increase of compliance with overt observation.

Hawthorne effect

Drug adherence:

Let's consider a study to measure adherence to a new drug in subjects with type 2 diabetes mellitus. Patients' medication compliance is measured for 3 months using electronic pill bottles.

The Hawthorne effect may occur when patients are aware that their medication adherence is being monitored and modify their behavior improving their adherence during the study period.

Hawthorne effect

Prevention:

- Unawareness/ Blinding
- Control group

Recall bias

Definition: Recall bias arises when subjects do not remember previous events or experiences accurately or omit details: the accuracy and volume of memories may be influenced by subsequent events and experiences.

Impact: Recall bias may increase or decrease the strength of observed associations. Mainly on the role of specific risk factors.

Recall bias

Example:

Let's consider a case-control study to examine the relationship between mobile use and the occurrence of a glioblastoma, a type of brain tumor. Researchers conduct face-to-face interviews with individuals and ask them to recall their past exposure to mobile.

Recall bias may occur when cases (subjects with glioblastoma) accurately remember and report their past mobile phone use. Whereas controls (subjects without glioblastoma) are less accurate or under report their mobile phone exposure.

Recall bias

Prevention:

- Train researchers
- Measures standardization
- Objective measures (non self-reported)
- Prospective studies

Immortal time bias

Definition: Refers to individuals of a cohort study that cannot experience the outcome during some period of follow-up time.

Impact: Immortal time bias can have an influence on the magnitude and direction of the measured effect.

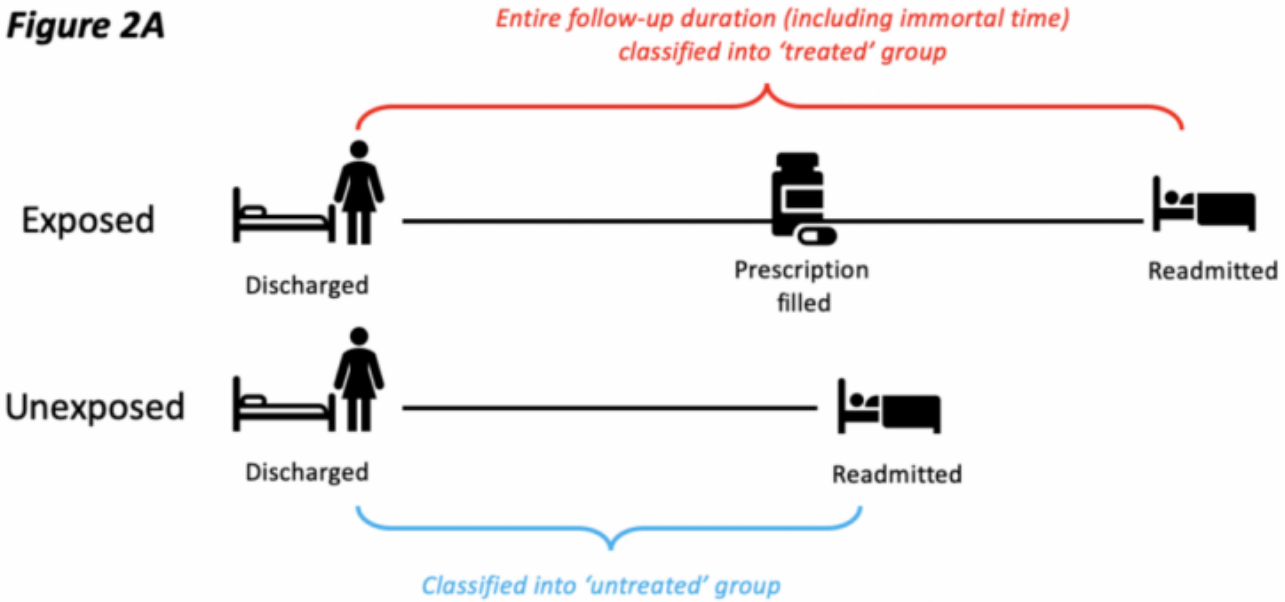
Immortal time bias

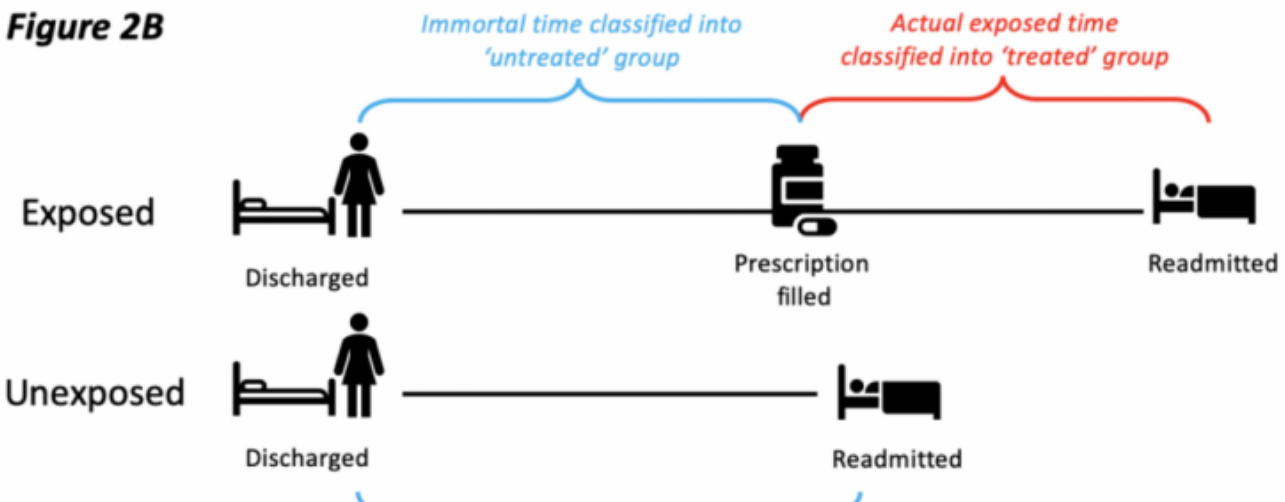
Example:

Inhaled corticosteroids could effectively prevent readmission and mortality in patients previously hospitalised with COPD.

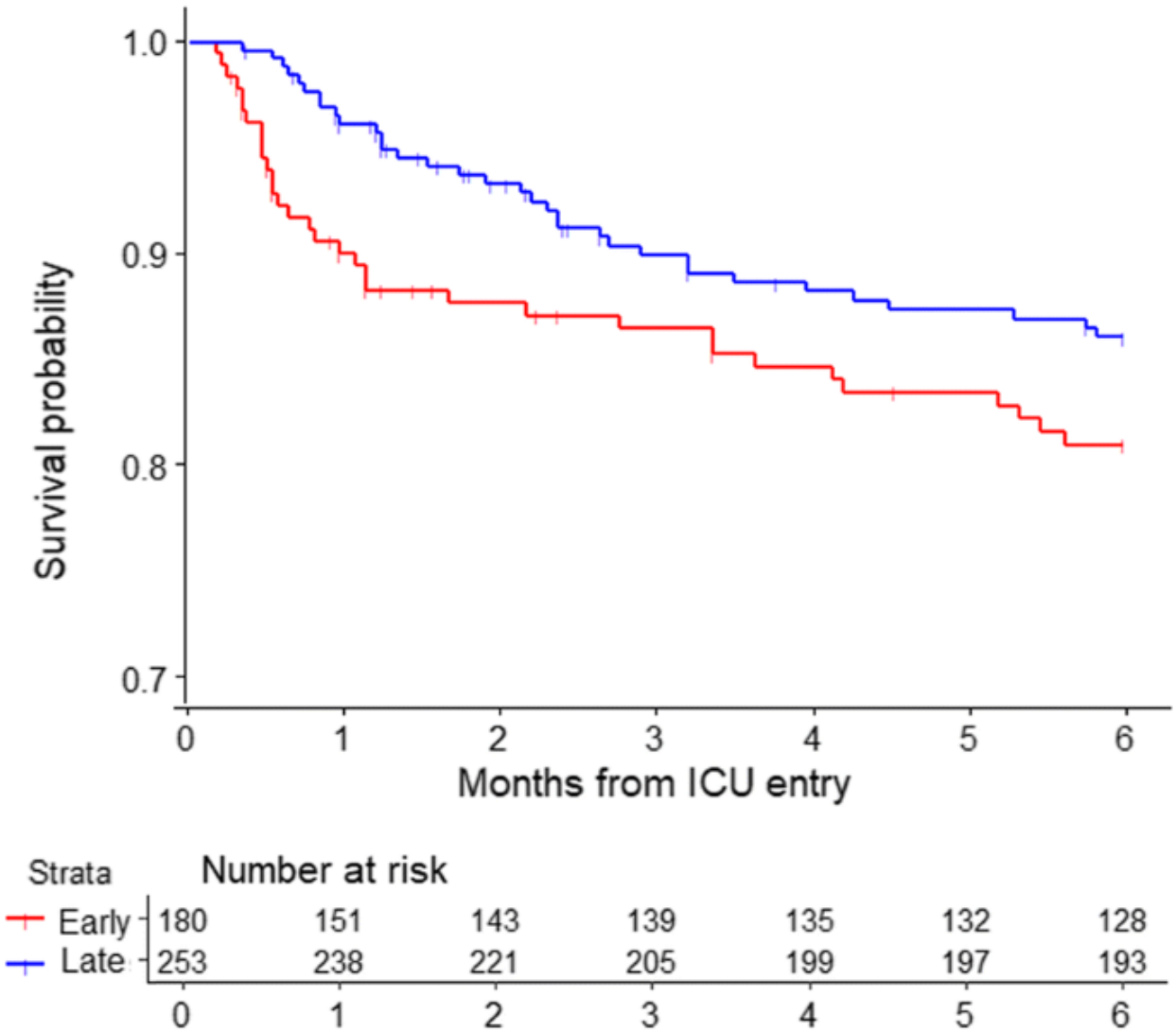
Participants entered the cohort on the day they were discharged and were then assigned to the treated group if they filled a prescription for a corticosteroid within the first 90 days from discharge. By design, participants allocated to the treated group could not have died or been readmitted between the time of entering the cohort and the time of filling their first prescription. In effect, they contributed 'immortal time' to the treated group.

Immortal time bias





Immortal time bias



Immortal time bias

Prevention:

- Exposition have to start baseline in both groups (time 0)

Non-response bias

Definition: occurs when non-responders from a sample differ in a meaningful way to responders (or early responders).

Impact: Errors in the estimation of population characteristics due to under-representation.

Non-response bias

Example:

Women, older people and people with higher levels of education were more likely to participate in a community cardiovascular follow-up study[7].

Non-response was determined to contribute to underestimated health risks.

[7] Jooste PL, Yach D, Steenkamp HJ, Botha JL, Rossouw JE. Drop-out and newcomer bias in a community cardiovascular follow-up study. Int J Epidemiol. 1990;19(2):284-289

Non-response bias

Prevention:

- Survey design
- Non-volunteer participation
- Incentives to participate
- Characterising non-respondents
- Weighting analysis

What to do?

What to do?

What to do?

Study design

- **Protocol, protocol, protocol!!!**
- Randomization
- Blinding
- Use always a control group when a comparison is required
- Prospective data is preferred to retrospective (when ever is possible)
- Standardize how data will be collected, with special attention to outcomes and interventions
- Conflict of interest disclosure

What to do?

Analysis

- **Statistical analysis plan... SAP, SAP, SAP!!!**
- Data quality control
- Blinding **!?**
- Stratification and adjustment
- DAG's
- Handle missing data and try to avoid complete case analysis
- Sensitivity analysis
- Follow reporting guidelines: EQUATOR NETWORK

What's next?

Bias more related to the analysis

- Wrong sample size bias
- Sampling bias
- Reporting bias
- Confounding bias
- Analytical bias

Message for take-home

Bias can never be completely eliminated, but you can significantly reduce its impact on your results.

Thank you very much!!!