1.2 Sources of bias

Observational studies have many sources of potential bias

In randomised trials, systematic biases controlled by

randomisation

blinding

For observational studies, potential for bias *at every step* from data collection to reporting

Many of these can be minimized with ’common sense’ and clear thinking, but some are subtle

Here we will discuss biases at the following steps:

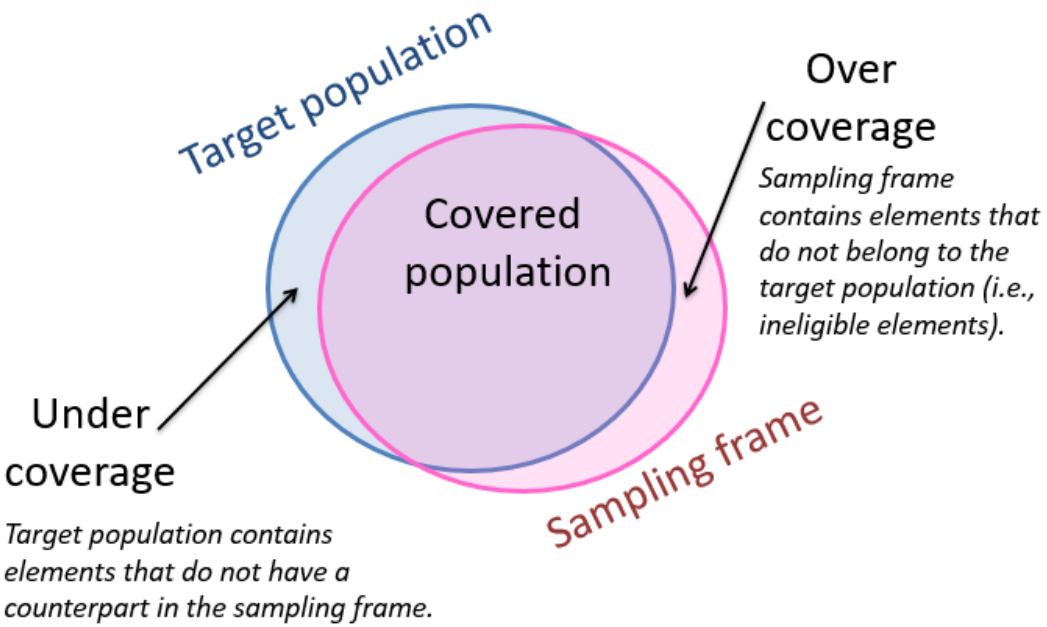
collection, classification/definition, analysis

Bias in data collection

1. observational/sampling bias

The sample not representative of the population, due to imperfections in:

- The population
- The sampling strategy



Bias in data collection

1. observational/sampling bias

Examples of imperfections in the population

Incomplete/missing population information, e.g., cause of death
(may be less problematic where there are electronic records)

“Truncation bias” due to:

- start up of electronic recording
- difficulty in identifying the population of interest

Perinatal epidemiology: miscarriages recorded?
e.g. cut-off for registration of pregnancy week 22 in Sweden

Bias in data collection

1. observational/sampling bias

Examples of imperfections in the sampling

Lack of care in selecting a truly random sample

- Convenience sampling selects easily assessible subjects
- Random sample may not be easy to obtain due to hidden biases: e.g., day of the week, doctor on duty, ….

Berkson's bias (or admission bias)

- Random sample of cases and choosing controls can be challenging (e.g., for hospital-based case)
- Both exposure and outcome are related to the probability of being admitted result in spurious association.

Bias in data collection

1. observational/sampling bias

Sackett (1979)

To assess the association between respiratory and bone diseases, a random sample from general population was collected and analyzed (odds ratio: 4.06; 95%CI: 1.03, 13.48).

A subgroup analysis among those who were hospitalized in the previous six months (odds ratio: 1.06; 95%CI: 0.59, 1.79).

We may have erroneously concluded that there was an association with hospital-based analysis.

Bias in data collection

2. response bias

Although observational/sampling bias could be avoided, we need to be mindful of potential response bias during data collection

This bias is due to a meaningful difference between respondents and non-respondents (e.g., differences in exposure or outcome)

The sample may not be representative of the population of interest due to response bias resulting in potentially erroneous conclusion

A public health survey in Netherlands found respondents tend to have better lifestyle behavior or health status (e.g., smoke less, better mental health); see Cheung et al. (2017).

Cheung et al. (2017) The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health; 17(1):276.

Bias in data collection

2. response bias

Famous mistake of 1936 Literary Digest

Postal survey to predict US presidential election

10 million forms issued and 2 million returned

Predicted Landon would defeat Roosevelt

In fact Roosevelt got 60% of vote!

Only 20% of the participants responded to the postal survey!!

People were sampled from listed phone numbers and owners of registered cars!!

Bias in data collection
2. response bias

Recent examples (2016)

Failure to predict:

Brexit referendum in the UK

Presidency of Donald Trump in the US

Combination of non-response and non-revealing true intentions

Bias in data collection

3. measurement/information bias

Systematic differences in the actual “measurement” of interest between groups (e.g., cases and controls) to be compared

Purpose of randomization in RCT is to eliminate/minimize this, e.g., control bias in respondents with “single blinding”
“Double blinding” also control bias in observer (“observer bias”)

But observational studies are prone to such bias….

Bias in data collection

3. information bias in cohort & case-control studies

In cohort study, participant will know their exposure

In case-control studies know their case and control status

This knowledge may bias the participant's response in a systematic way:

Cohort of siblings of cancer cases and siblings of healthy controls, "exposed" siblings may be more proactive in seeking medical opinion/care

In case-control study of depression, cases may assign higher scores to daily stresses or remember more adverse events in their past **("recall bias")**

Bias in data collection

3. random measurement/classification error

In addition to systematic error, there can also be spurious measurement error, due to random fluctuations in:

biological systems (e.g., BP variation, usual to average 3)

precision of the measuring instrument (called “technical error”)

If measurement error is **non-differential,** will lead to some misclassification, and **dilute** the effect of interest

Awareness of measurement error can lead to better design

Bias in data collection

4. time-related bias

In cohort and case-control studies, as time is also an important part of the design, there are several types of time-dependent bias.

Truncation bias (e.g., register start-up)

Time-window bias (different window when exposure measured for cases and controls)

Length-time bias (e.g., slower growing tumors more likely detected by screening vs. fast-growing may result in patient death before next screening appointment)

Bias in data analysis

4. time-related bias

Lead-time bias (e.g., in cancer screening studies) caused by ‘lead-time’, which corresponds to the time between an early diagnosis (e.g. due to screening) and the time when the disease would have been diagnosed by routine clinical procedures.

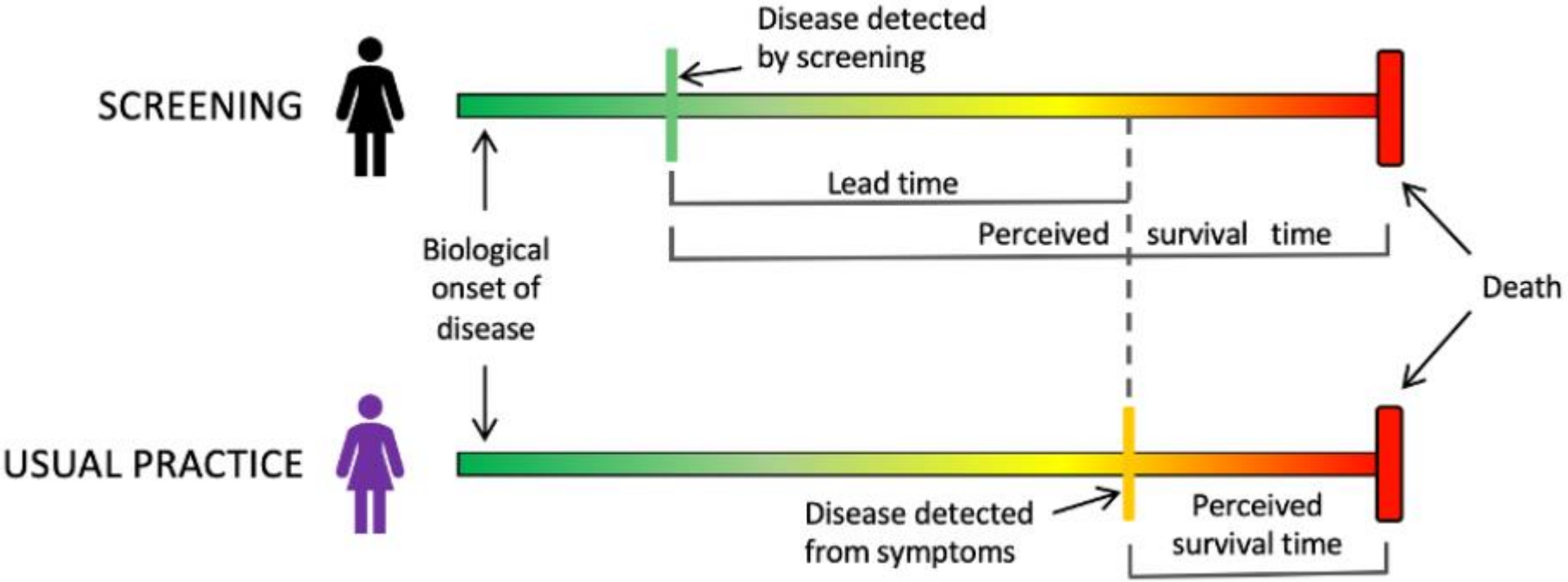


Figure: https://catalogofbias.org/biases/lead-time-bias/

Bias in data analysis

4. time-related bias

Immortal time bias occurs when participants cannot experience the outcome during a specific period of follow-up (i.e., 'immortal time')

When 'immortal time' is misclassified or excluded during analysis, it results in a biased association

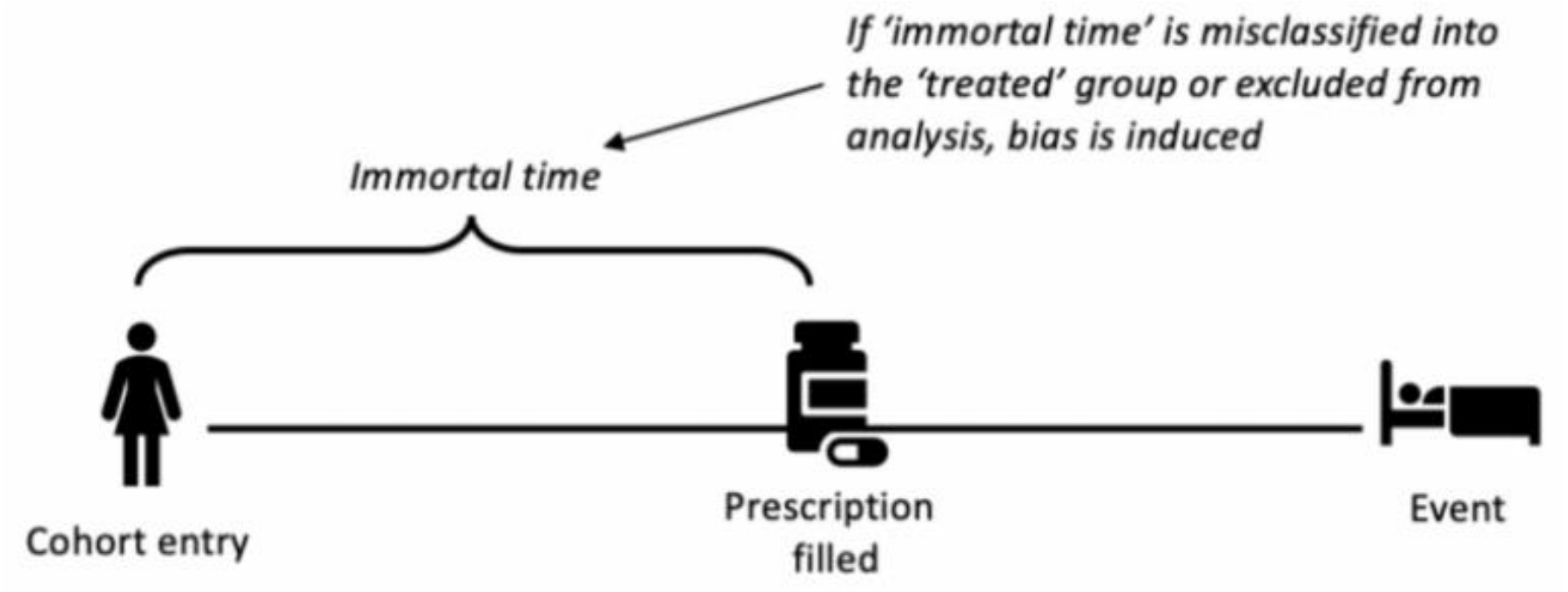


Figure: https://catalogofbias.org/biases/immortal-time-bias/

Bias in data analysis

5. Confounding bias

Although an observational study may have avoided the biases discussed previously, we need to ensure there is no bias due to the presence of a confounding factor

A confounder is a variable that influences both the exposure and the outcome, resulting in an incorrect conclusion on the relationship between the exposure and the outcome

The effect of the exposure on the outcome could be exaggerated or diluted

Bias in data analysis

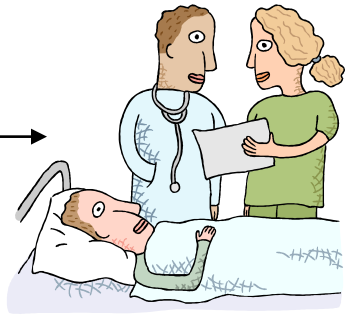
5. Confounding bias

EXPOSURE

(drinking)

OUTCOME

(lung cancer)



CONFOUNDNG VARIABLE

(cigarette - smoking)

List of biases (partial list)

Table 1 Alphabetical list of biases, indicating their type and the design where they can occur

Specific name of bias	Group of bias	Subgroup of bias (next level to specific name)	Type of design affected
Allocation of intervention bias	Execution of an intervention		Trial
Apprehension bias	Information bias	Observer bias	All studies
Ascertainment bias	Selection bias	Inappropriate definition of the eligible population	Observational study
Berkson's bias	Selection bias	Inappropriate definition of the eligible population	Hospital based case-control study
Centripetal bias	Selection bias	Healthcare access bias	Observational study
Citation bias	Selection bias	Lack of accuracy of sampling frame	Systematic review/meta-analysis
Competing risks	Selection bias	Ascertainment bias	All studies
Compliance bias	Execution of an intervention		Trial
Confounding by group	Confounding		Ecological study
Confounding by indication	Confounding		Case-control study, cohort study
Contamination bias	Execution of an intervention		Trial, mainly community trials
Detection bias	Selection bias	Uneven diagnostic procedures in the target population	Case-control study
Detection bias	Information bias	Misclassification bias	Cohort study
Diagnostic/treatment access bias	Selection bias	Healthcare access bias	Observational study
Diagnostic suspicion bias	Selection bias	Detection bias	Case-control study
Diagnostic suspicion bias	Information bias	Detection bias	Cohort study
Differential maturing			Trial
Differential misclassification bias	Information bias	Misclassification bias	All studies
Dissemination bias	Selection bias	Lack of accuracy of sampling frame	Systematic review/meta-analysis
Ecological fallacy	Information bias		Ecological study
Exclusion bias	Selection bias	Inappropriate definition of the eligible population	Case-control study
Exposure suspicion bias	Information bias	Recall bias	Case-control study
Family aggregation bias	Information bias	Reporting bias	Observational study
Friend control bias	Selection bias	Inappropriate definition of the eligible population	Case-control study
Hawthorne effect	Information bias		Trial
Healthcare access bias	Selection bias	Ascertainment bias	Observational study
Healthy volunteer bias	Selection bias	Non-response bias	Observational study
Healthy worker effect	Selection bias	Inappropriate definition of the eligible population	Cohort study (mainly retrospective)

Delgado-Rodríguez et al. (2004). Bias. J Epidemiol Community Health; 58(8):635-41.

Useful links/resources

Catalogue of Bias:

https://catalogofbias.org/biases

European Network of Centres for Pharmacoepidemiology & Pharmacovigilance (Chapter 6: Methods to address bias and confounding):

https://www.encepp.eu/standards_and_guidances/methodologicalGuide6.shtml