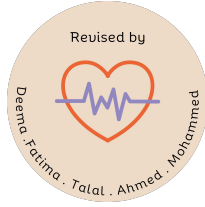


Research
442



Bias in epidemiological studies

Lecture No. 17

Objectives

1. To be able to differentiate between random and systematic error
2. To explain the difference between selection and information bias
3. To give examples for selection and information bias
4. To explain how to reduce bias in the different health research designs

~ This lecture was presented by **Dr. Hani Alghamdi**

~ It is included in the **Midterm Exam**

~ We highly recommended reading the **Ayah** in the first page

Slides

Color code

Original text

Dr. Notes

Important

Golden note



Extra



Editing file

Bias

“deviation of **results** or **inferences** (Information bias) from the truth, or **process** (Selection bias) leading to such deviation” (Grimes and Schulz, 2002)

Probability sampling decreases bias in selection by making the sample results close to population results.

(Total population is taken only in surveillance or if the population is already very small. Those types are usually not research)

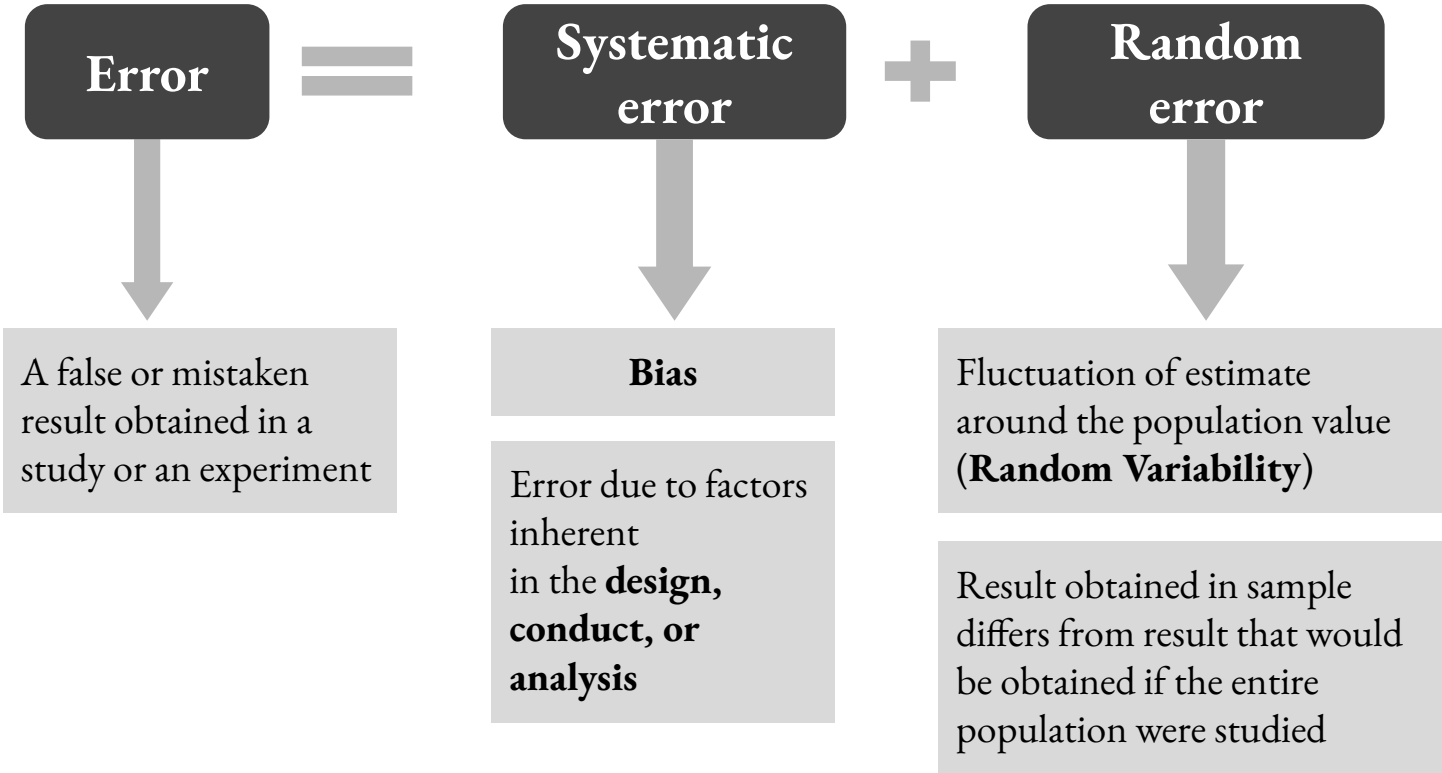
سُورَةُ الرَّاقِعَاتِ

إِنَّهُ لَقُرْآنٌ كَرِيمٌ ﴿٧٧﴾ فِي كِتَابٍ مَّكْنُونٍ ﴿٧٨﴾ لَا يَمَسُّهُ إِلَّا الْمُطَهَّرُونَ ﴿٧٩﴾ تَنْزِيلٌ مِنْ رَبِّ الْعَالَمِينَ ﴿٨٠﴾

المختصر في التفسير

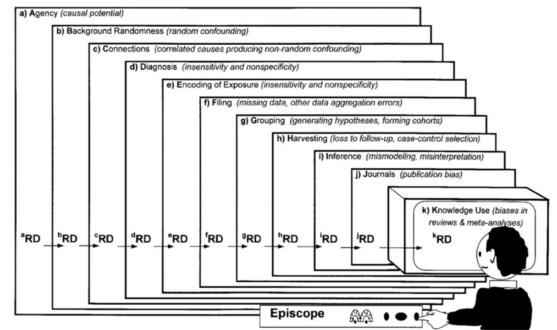
إن القرآن المقروء عليكم - أيها الناس - قرآن كريم؛ لما فيه من المنافع العظيمة.

في كتاب مَصُونٍ عن أعين الناس، وهو اللوح المحفوظ. لا يمسه إلا الملائكة المطهرون من الذنوب والعيوب. مُنْزَلٌ من رب الخلائق على نبيه محمد صلى الله عليه وسلم.



Classification of bias

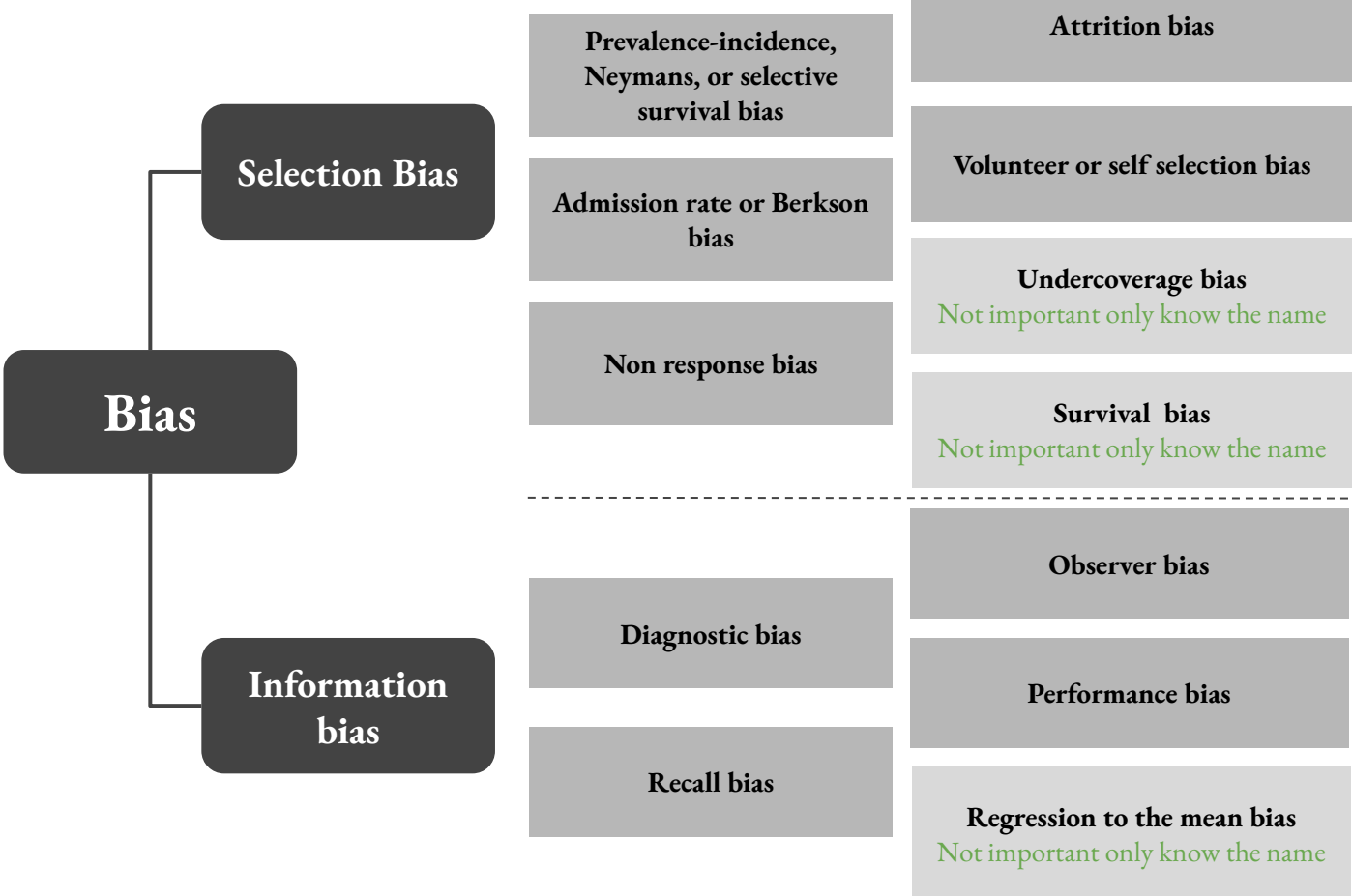
- Based on stages of research
 - Choi (65 types)
 - Sackett (19 types)
- Causal diagram theory (Maclure and Schneeweiss)
- Three main groups (Kleinbaum et. al.)
 - Selection, information, and **confounding**
- Steineck and Ahlbom:
 - Misclassification (similar to information bias), misrepresentation (narrower meaning than selection bias), and analysis deviation



Every research must have some degree of error

We will study WHO classification of bias (next slide) which has 2 main types: Information and selection bias

WHO Classification of Bias



Selection Bias

Bias due to factors related to the study's participants, such as the **choice of the population** or **sampling and requirements methods**.

Sampling bias

- Occurs when some members of the intended population are **less likely to be included** than others.

Attrition bias

- Occurs when participants who **drop out** of a study differ from the ones who remain.

Volunteer (self selection) bias

- People with specific characteristics are **more likely to participate** than others.

Non-response bias

- People who **refuses to participate** or drop out systematically differ from those who take part.

Survivorship bias (Not important only know the name)

- Successful observations or people are more likely to be represented in the sample than unsuccessful ones.

Undercoverage bias (Not important only know the name)

- Some members of a population are inadequately represented in the sample.

Sampling bias is under selection bias.

Volunteer/self selection, non response, survivorship, and undercoverage bias falls in both selection and sampling bias.

Prescreening/advertising and healthy user bias only falls under sampling bias.

Attrition bias falls under selection bias only.

1. Know all names of **Neyman** and volunteer and **Berkson** bias!!

In volunteer bias: When the participants are **free** to choose to participate or not.

First you choose randomly, select the sample then allow for freedom.

How to minimise selection bias

1. Observational studies:
 - a. Non-probability sampling → Matching (unexposed group comparable to exposed as possible as you can) **Minimise volunteer bias**
2. Experimental studies:
 - a. Randomisation (random allocation)
 - b. Blinding (masking) **(If not randomised, it must be blinded)**
3. Sampling bias:
 - a. Probability sampling

Sampling Bias



This type of bias occurs when **some members of a population are systematically more likely to be selected in a sample than others**

Volunteer (self selection) bias

- People with specific **characteristics** are **more likely to participate** than others.

Non response bias

- People who **refuses to participate** or drop out **systematically differ** from those who take part.

Prescreening or advertising bias (Only under sampling bias)

- Bias due to the way participants are pre-screened or where a study is advertised.

Healthy user bias (Only under sampling bias)

- Volunteers for preventative interventions are more likely to pursue health-boosting behaviors than others

Survivorship bias (Not important only know the name)

Undercoverage bias (Not important only know the name)

Prevalence-incidence bias (Neyman or selective survival bias)

- Exclusion of individuals with severe or mild disease resulting in a systematic error in the estimated association or effect of an exposure on an outcome.
- Occurs due to the timing of when cases are included in a research study.
- This bias can occur in **both cross sectional and (prevalent) case-control studies**.
- Example: A case-control study investigating pneumonia that only enrolls cases and **controls admitted to a hospital**. Those with pneumonia **who died** prior to admission will not be included the sample. The selected sample will, therefore, include moderately severe cases, but not fatal cases.

Example: measuring knowledge about DM in a patient who was recently diagnosed (not a lot of knowledge) and another patient who had DM for more than 10 years (very knowledgeable)

Controls should be exposed and in the SAME environment, but do not have the outcome.

Admission rate or Berkson bias

- Arises when the variables under study are affected by the **selection of hospitalized** subjects leading to a bias between the exposure and the disease under study.
- First described for **case-control studies**.
- Example (**Doctor haven't read the example**): Berkson described assessment of the relationship between gallbladder disease (as a possible cause) and diabetes. Because the study involved participants attending a clinic, whose attendance (overall) was affected both by gallbladder disease and by diabetes, this biased the association between gallbladder disease and diabetes. (Berkson, 1946).

Information bias

Bias as a result of **incorrect measurement** or **classification of key variables**

When it occurs?

- When information used in a study is either measured or recorded inaccurately.

How?

- The researchers use **different methods** to assess outcomes in each group (e.g., 2 different questionnaire)
- The independent variable and/or the dependent variable are **recorded inaccurately**
- **Instruments** for objective measurements (e.g., weight) is **not correctly calibrated**, results are registered incorrectly, or data are switched during the **data entry or data cleaning phase**
- In clinical trials, the study does not have double blind design

Causes?

1. **Non-differential misclassification**
2. **Differential classification**

1. **Non-differential misclassification:**

- **Equally inaccurate** measurements in all study groups.
- Study participants in **both** comparison groups have difficulty accurately **remembering something** that is not objectively verifiable

Why it is problematic?

- It tends to make the groups **appear more similar** than they really are
- It causes researchers to **underestimate the association** between variables

2. **Differential classification:**

- Measurement difference that exists between study groups

Why?

- Participants in the case group may be able to **recall** past exposure to risk factors more accurately than the healthy control group

So?

- It can cause either an **underestimate or an overestimate** of the association between variables

	Shows symptom Y	Doesn't show symptom Y
Has new disease X		
Doesn't have new disease X		

Scribbr ★ has the disease + doesn't have the disease

Studies of **rare or newly discovered** diseases that do not have **uniform diagnostic criteria** are at risk for information bias.

In the **absence of a common standard**, people who do not have a disease may be classified as having it, and vice versa.

Information bias is also known as **measurement bias or misclassification.**

To avoid information bias: Questionnaire should be valid and reliable.

Both groups can not recall in non-diff bias.

Types of Information Bias

Recall bias

participants in one of the study groups can recall past events or behaviours better than those in the other.

Observer bias

researchers are aware of the **hypothesis under investigation** or know which group each participant is assigned to.

(RCT, focus on the intervention group; In qualitative, focus on certain sample)

Performance bias

researchers or participants in a study modify their behaviour or responses because they are aware of group allocation.

Patient knows if on medication or placebo.

Regression to the mean bias

Doctor skipped

a variable that shows an extreme value (outlier) on its first measurement will tend to be closer to the mean on a second measurement.

Diagnosis bias

(Diagnostic suspicion bias; Not important)

When biased considerations along with inadequate individual thought patterns result in misdiagnosis.

How to minimise information bias?

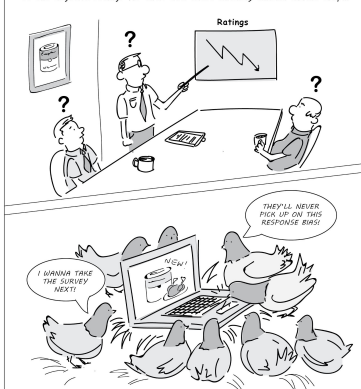
Raised from data collection and measurement:

1. Verify collected information (Recheck a random small amount of data manually)
2. Double blinding (masking) → if not possible?
 - a. Develop a **protocol** for data collection, measurement, and interpretation of information
3. Standardised questionnaire
4. Calibrated instrument

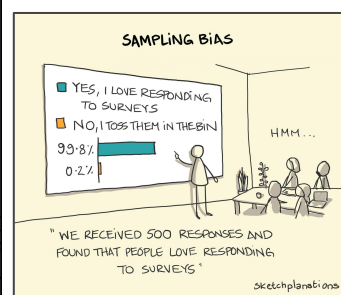
Recall bias can occur in case control and cross-sectional bias

Observer bias: e.g., In the medication group → more questions and attention, in placebo group → less attention.

After conducting an online survey, the Campbell's Soup Company is baffled at the negative ratings for their new extra-chickeny chicken noodle soup...



Little did they know, voluntary surveying had made their sample anything but representative of the population...



Know how to minimise information bias!
Interim analysis: used in clinical trials, conducted before data collection has been collected, in which the primary research question is addressed. It is not used to minimise information bias

Q1: A researcher wants to study the effects of a new educational program on student achievement. The researcher recruits participants by posting a flyer in the school library.

- A. Recall bias
- B. Sampling bias
- C. Non response bias
- D. Observer bias

Q2: A researcher wants to study the effectiveness of a new tutoring program on student achievement. The researcher recruits participants from a group of students who are struggling in math. The researcher follows the participants for 6 months, but many of the participants drop out of the program before the end of the study.

- A. Sampling bias
- B. Non-response bias
- C. Attrition bias
- D. Neyman's bias

Q3: A researcher wants to study the student satisfaction with the school cafeteria. The researcher sends out a survey to all students in the school. 15% of students respond to the survey.

- A. Non-response bias
- B. Attrition bias
- C. Sampling bias
- D. Volunteer bias

Q4: Sackett (1979) interviewed a random samples of the general population to determine the presence or absence of respiratory disease and locomotor disease. He then looked at the same thing for those within the sample who had been hospitalised in the previous six months. The results are in the table:

- A. Neyman's bias
- B. Berkson's bias
- C. Non response bias
- D. Volunteer bias

Bias in Analytic Research 53

TABLE 2. THE RELATIVE ODDS OF DISEASE OF THE BONES AND ORGANS OF MOVEMENT WITH AND WITHOUT RESPIRATORY DISEASE*

		In the general population			In the subset who were in the hospital in the prior 6 months		
		Disease of bones and organs of movement		Totals	Disease of bones and organs of movement		Totals
		Yes	No		Yes	No	
Respiratory disease	Yes	17	207	224	5	15	20
	No	184	2376	2560	18	219	237
	Totals	201	2583	2784	23	234	257
		Relative odds (cross-products): 1.06			Relative odds (cross-products): 4.06		

*Adapted from Roberts *et al.* [3].

In the hospital sample, people with respiratory disease are much more likely to suffer from locomotor disease (Relative odds 4.06). But, if we looked at the general population, we would conclude there is no association between the two diseases (Relative odds 1.06)



Q5: In a study examining the relationship between mobile phone usage and the risk of brain tumours, participants are interviewed about their past mobile phone usage habits, including the number of hours they spent on their phones each day over the past 10 years. The study finds a statistically significant increase in the risk of brain tumours among participants who report higher mobile phone usage.

- A. Recall bias
- B. Sampling bias
- C. Admission rate bias
- D. Diagnostic bias

Q6: A researcher is conducting a study to examine the impact of a new stress reduction program on overall well-being. The researcher recruits participants by sending out an email to all employees in a large corporation, inviting them to participate in the program. The study finds that those who choose to participate are generally more health-conscious and proactive in seeking stress reduction solutions.

- A. Healthy worker bias
- B. Volunteer bias
- C. Attrition bias
- D. Neyman's bias

Q7: Scenario 7

A company is conducting a long-term study to assess the health effects of exposure to certain chemicals used in its manufacturing processes. Over several years, the study consistently shows that the employees who remain in their jobs have lower rates of respiratory issues and other health problems compared to the general population.

- A. Non-response bias
- B. Attrition bias
- C. Pre-screening bias
- D. Healthy worker bias

Q8: Suppose that a case-control study is carried out to study the relation between tobacco smoking and acute myocardial infarction (AMI), being cases interviewed one week after the coronary attack. (smoker patients with AMI die more frequently)

- A. Berkson bias
- B. Neyman bias
- C. Healthy worker bias
- D. Volunteer bias



Q9: A psychology researcher is conducting a study on the effects of a new stress-relief intervention on participants' anxiety levels. The researcher knows that the study hypothesis predicts that the intervention will significantly reduce anxiety. Participants are divided into two groups: one receiving the new intervention, and the other receiving a placebo. During the study, the researcher inadvertently observes the intervention group more frequently and spends more time with them, believing that the intervention is effective. The placebo group, on the other hand, receives less attention because the researcher assumes the treatment is less likely to be effective. This difference in attention and interaction with the two groups leads to unequal data collection and, in turn, may bias the results in favor of the intervention group, making it appear more effective than it actually is.

- A. Recall bias
- B. Sampling bias
- C. Observer bias
- D. Neyman bias

Q10: A public health department is conducting a survey to understand the eating habits and nutritional choices of residents in a specific city. They distribute flyers and post online announcements inviting residents to participate voluntarily in the survey. The survey is entirely self-selecting, meaning residents decide for themselves whether they want to participate or not.

- A. Non response bias
- B. Volunteer bias
- C. Attrition bias
- D. Berkson bias

Extra

Q11: Research looking for construction workers who have disabilities due to their work, but chooses the sample from a construction site. (Knowing that disabled construction workers won't be able to continue in this type of job)

- A. Non-response bias
- B. Neyman bias
- C. Volunteer bias
- D. Healthy worker bias

Q12: Researcher asks "Which one serves the best coffee?" with options of: Tim hortons, starbucks, dunkin, and half million. The researcher did not provide a fair comparison because of difference of prices, locations, number of branches, and specification of which type of coffee.

Is this selection or information bias?

- A. Selection bias
- B. Information bias

Q13: Researcher asks "Which one serves the best coffee?" with options of: Tim hortons, starbucks, dunkin, and sultan coffee. The researcher did not provide a fair comparison because the researcher did not specify which type of coffee since sultan coffee would only have arabic coffee. Is this selection or information bias?

- A. Selection bias
- B. Information bias

Q14: Researcher asks "Which one serves the best coffee?" with options of: Tea colors, tea world, and dunkin. The researcher did not provide a fair comparison because from the shops names, two of them serve tea and dunkin serves coffee, but the researcher is asking about coffee. Is this selection or information bias?

- A. Selection bias
- B. Information bias

Q15: When the doctor presented the QR code to participate in the questions, this sample does not include the students who are not in class.

- A. Sampling bias
- B. Recall bias

القارة:
عبدالله الشهري
وهي المتحفي

نواف التركي

ريان الفنامي

الأعضاء:

رغد النظيف
ريما الجريبة
شهد البخاري
نوف الضلعان
أمير الاحمري
وعد ابونخاع
نراء الهويش
في الدوسري
منار الزهراني

عبدالله التركي
محمد الزير
عثمان الدريهم
عبدالعزیز القحطاني
ناصر الفيت
سعد السهلي
رائد الماضي
سعود الشعلان
عبدالله المياح
عبدالله النجريس
تركي العتيبي
عبدالله القرني
عامر الفامري
سعد الاحمري
معاذ آل سلام
محمد الحصيني

MCQ:

Q1: In a control study, patients with neonatal congenital deformity were asked about a medication they took during pregnancy. Most of them reported severe side effects but this medication was known to be safe during pregnancy. Which type of bias?

- A. Selection bias
- B. Survival bias
- C. Recall bias
- D. Interviewer bias

Q2: How to minimise selection bias?

- A. Calibrate instruments
- B. Standardised questionnaire
- C. Verify information
- D. Matching

MCQ:

Q3: A study wanted to measure how many residents suffer from rabies. The researchers used telephone survey and saw that there is 0% of residents with rabies. What type of bias might have affected the study?

- A. Volunteer
- B. Neyman bias
- C. Non response bias
- D. Recall bias

Q4: How to minimise information bias?

- A. Matching
- B. Masking
- C. Randomisation
- D. Probability sampling