Bias

Bias refers to errors in the design or conduct of a study that makes the study results different from the truth. To put it simply, bias occurs when the wrong subjects are studied or the wrong information is obtained form some study subjects. Therefore, a biased study does not give a true conclusion about the situation and associations between exposures and outcomes.

Although some authors tend to consider confounding as a type of bias, but it is better to discuss them separately. Confounding is not related to errors in the study leading to including wrong subjects or recording wrong information; it is a legitimate condition that could happen in any study when there are associations between some exposures and outcomes.

There is no cure for bias at the analysis stage; the only solution is prevention during design. If results of a study are grossly biased, they will be useless; no matter how much analysis is done on them. To avoid such a situation, potential sources of bias should be identified early in the study designs stage and appropriate precautions should be taken to minimize bias.

Types of Bias

Bias can broadly be classified in two types: selection bias and information bias. Information bias is further classified to reporting bias and observer bias. Each of these two types is related to a different stage of the study design but both lead to the same result i.e. error in the truthfulness of the results.

Selection bias

As its name implies, selection bias is related to the way participants are selected to the study and this bias can occur in either of two situations:

1. People who are selected to participate in the study are not a representative sample of the reference population. This could happen either because of using non-random methods for sample selection or when response rate is low. In descriptive studies, if the studied population does not represent the reference population properly, the study estimate will not be a true reflection of the reference population and this error in the sample estimate occurs due to selection bias. For example if wanted to study prevalence of malnutrition amongst children under 5 in Kurdistan and we decided to do this in the kindergartens and nurseries. Clearly, these children will not be a true representation of all Kurdish children since poorer families, where malnutrition is likely to be higher, are less likely to send their children to pre-school education. In such a case the sample estimate will be an under-representation of reality due of selection bias. Even if the original sample was representative of the target population but a large proportion of the sample refused to participate, the sample may no more be representative.

2. When the study compares two or more groups, these comparison groups are not really comparable. In analytical studies the comparison groups should be comparable in terms of exposures and outcomes apart from those under study. If these groups are

grossly different, the comparison will not be fair and the resulting effect estimation will be wrong due to selection bias. For example if we wanted to study the risk factors of HIV by open testing. It is possible that people who are at a higher risk of HIV will be less likely to participate in this study than people are at lower risk leading to bias in the estimated effect of risk potential factors.

Selection bias in cross-sectional surveys

In Surveys selection bias occurs when the sample is not representative of the target population. Selection bias in this case could lead to either over-estimation or under-estimation. If we wanted to estimate prevalence of pneumonia through patients admitted to hospital, selection bias is likely because the hospital admissions do not fairly represent all population and our finding will be an under-estimation. Our sample may not include mild cases of pneumonia or sever cases who die before hospitalization.

If we wanted to do a study on prevalence of adult obesity in Sulaymaniyah through primary health care centers selection bias is likely to be a problem. People attending health centers are more likely to be obese because obesity is associated with certain illnesses who may lead to PHC attendance. If such a situation is true, the estimated prevalence will be an over-estimation of reality. However, if obesity is associated with higher socioeconomic status, then it is likely that obese people will attend private clinics and less of them attend PHCS. If such a situation is true, then the estimated prevalence will be an under-estimation of reality. Making sure that the study population is a good representation of the target population and then following random selection methods will help eliminate selection bias.

Selection bias in case-control studies

In case-control studies, selection bias could happen in selection of cases or selection of controls. In a case-control study comparison is made between the exposure status of cases and controls, therefore these two groups have to be similar in respect to their exposures. Any situation that leads to inclusion/ exclusion of cases or controls with more or less likely chances of exposures could lead to selection bias.

1. **Bias in selection of cases** happens when cases are not representative of all cases within a defined population. if we include cases who have had more chances of being exposed to the exposure of interest than other cases in the population, this leads to over-estimation of the effect of the exposure. On the other hand, if we exclude such cases, the strength of the association will be weaker. For example if we undertake a case-control study to investigate risk factors of diarrhea in children and we select our cases at a PHC center, selection bias may be problem. Since socioeconomic status is a potential risk factor, cases from a PHC may not represent all cases in the population because people with diarrhea with lower socioeconomic status are more likely to present to a PHC than those from higher classes. This exclusion of cases with higher socioeconomic classes may lead to over-estimation of the association. To avoid this type of bias we have to make sure that the cases we select are not more or less likely to be exposed to the exposure of interest than all cases in the population.

2. **Bias in selection of controls** happens if controls are not representative of the **population** which produced the cases. The control group should be a fair representation of the reference population in such a way that if a control had had the outcome of

interest, they would have been included as a case. The best strategy to avoid this type of bias is to select the controls from the reference population i.e. community-based controls. However, frequently and for practical reasons, we select controls from health settings e.g. people attending a hospital fro other health problems (hospital-based controls). This is not a good strategy because people attending a hospital are not a good representation of the population and therefore selection bias is likely. If we decided to do this, we have to keep in mind that certain exposures of interest may be associated with been in hospital which could lead to over- or under-estimation of the association. For example if we do a case-control study on lung cancer by selecting hospital-based controls, we may find weak association with smoking because hospital controls may be more likely to smoke than the reference population as some hospital admissions may be related to smoking.

Selection bias in cohort studies

A cohort study involves comparison of exposed and unexposed groups followed up for a certain period of time. Selection bias could happen if either the two groups are not truly comparable due to poor choice of the unexposed group; or if there are differences of the follow-up process of the two groups.

1. **Bias due to poor choice of the unexposed group** happens when the unexposed groups differs from the exposed group in relation to exposures other than the exposures of interest. Ideally the unexposed group should be as much similar to the exposed group as possible in relation to factors not under study because if they differ in other factors which could be related to the outcome, the results will be biased. For example if we wanted to compare morbidity and mortality rates of workers of a cement factory with those of the unexposed general population. The general population may not be a good comparison group for cement workers because of workers generally may be healthier than the general population (the healthy worker effect). It would be better to compare the cement workers with workers from another factory. This bias could be avoid by carefully choosing the unexposed group.

2. **Bias due to differences in follow-up** happens when the exposed and unexposed groups are not followed in the same way and outcomes are not ascertained similarly for both groups. Both groups must be followed up for similar periods of time and the same effort must be done to ascertain outcomes in both groups. Incomplete follow-up for more people from either group than the other may cause some outcome events to be missed leading to selection bias. For example, in a study comparing incidence of suicide attempts in the youth, more males than females might migrate out the county and hence be lost to follow-up. Some of these migrants might have developed suicide attempts yet missed from the study. This situation leads to under-estimation of incidence rate in males. To avoid this type of bias, all efforts must be done to maximize follow-up and ascertainment of cases in a high proportion of both groups such as more than 80%.

Information Bias

Information bias is any error in the measurement of exposures or outcomes that results in systematic differences in the accuracy of information collected from the study participants. Errors in measurement are always a possibility, but when such errors are not distributed

randomly between participants or they are too many, they result in systematic differences in the accuracy of exposure and/or outcome data between the comparison groups and therefore the study findings will be biased and may even be useless. For example in a case-control study, information bias could lead to systematic differences in accuracy of exposure data between cases and controls. In a cohort study, information bias could lead to systematic differences in accuracy of outcome data between exposed and unexposed groups.

When the exposures and outcomes are assigned to the wrong category, misclassification occurs. For example when a disease even is labeled as non-event or when an exposed subject is labeled as unexposed, this is misclassification i.e. the subject has been placed in the wrong category. When misclassification is not random, i.e. when it affects the two comparison groups in a grossly different way, it is called differential misclassification which biases study findings. For example if in a case-control study on maternal education and child injury, 30 mothers of cases vs. 5 of controls were assigned to the wrong category of education, this is differential misclassification.

Information bias can be divided into two main types of reporting bias and observer bias

1. **Reporting bias** is bias due to errors of people (subjects) providing the information i.e. the study participants. Reporting bias can affect either the exposure or the outcome or both. Reporting bias affecting the exposure happens when subjects with a specific health outcome report previous exposures with a different degree of accuracy to those without the outcome. Reporting bias affecting the outcome happens when subjects who have experienced a specific exposure report the outcome of interest with a different degree of accuracy to those who have not experienced the exposure.

   In a cross-sectional survey, mothers of vaccinated children may not remember the vaccination status of their children and report it wrongly. In a case-control study investigating risk factors of burns in pre-school children, mothers of cases may report their children as hyperactive more frequently than mothers of controls subconsciously to defend themselves that they are not to blame for what happened to their children. In a case-control study on leukemia, cases could be more likely to report exposure to chemical weapons than controls either because they remember better than controls or they are affected by their awareness of the potential link. In general, people with a disease may remember exposures better because they might have been asked previously about possible exposures, and they might have spent some time thinking about the cause of their illness. This type of reporting bias is called **recall bias.**

   In a cohort study about a new promising medication, people receiving the new drug may be more likely to interpret any improvement in their health as the effect of the drug, compared with people receiving a standard drug.

   Reporting bias may be exaggerated when people are aware of the exposure-disease association under study. For example, in the late 1980s there was public concern in the UK about the risk of *Salmonella enteritidis* gastroenteritis from eating eggs. In a case-control study of causes of *S. enteritidis* gastroenteritis, cases would be likely to exaggerate their egg consumption, or to remember it better and report it more completely than controls.

Reporting bias may also happen when participants feel they have to give the answer that they think will please the investigator or they think is the right answer. For example in a study about hand washing and diarrhea, subjects may over-report their hand washing habits. Reporting bias may also happen when participants try to conceal embarrassing information such as in a study with number of sexual partners in a study on STD and sexual activity.

Reporting bias can be minimized by using, if possible, data from objective sources; keeping subjects unaware of the association under study, if possible and; and by using standard and uniform methods of data collection from all groups and encouraging theor participation and remembering.

2. **Observer bias** is bias due to errors of the person recording information from participants i.e. the interviewer or the investigator. Observer bias can affect either the exposure or the outcome. For example in a case-control study observer bias happens when the accuracy of exposure data recorded by the investigator differs systematically between cases and controls. In a survey observer bias happens when the accuracy of exposure or outcome data recorded by the investigator differs systematically between subjects in different groups. In a cohort or intervention study, observer bias happens when the accuracy of outcome data recorded by the investigator differs systematically between exposed and unexposed groups.

   Observer bias can be a deliberate or careless action by un-committed data collectors, but it is more likely to occur if the assignment of exposure or outcome depends on a subjective judgment. For example *a*sbestos is recognized cause of mesothelioma. In a cohort study investigating risk factors for mesothelioma, the histologist might be more likely to classify a doubtful histological specimen as mesothelioma if he knew that the patient had been exposed to asbestos. In a case-control study on BCG effectiveness, interviewer might be less likely to observe presence of a BCG scar if they knew that the subject had TB.

   An Interviewer, who is aware of the hypothesis being tested, may interpret the participant's responses differently depending on whether the participant is a case or a control. This type of observer bias is also known as **interviewer bias**. The best way to avoid this bias is by making the interviewer unaware of the subject's disease status (in a case-control study) or subjects exposure status (in a cohort or intervention study). This strategy is called **blinding** which means blinding or masking the observer to the outcome or exposure status of the participant. Blinding of observers is simple when exposure or disease category is not overtly obvious but determined by another method such as laboratory tests. The research participants can also be blinded to the exposure (such the type of drug they are given) and this protects against reporting bias (such as mis-reporting side effects of the drug). **Double-blinding** is the situation when both observers and participants are blinded to the exposure under study (such as the drug) which in addition to reporting bias, protects against observer bias since it prevents the observers who classify outcome from being influenced by knowledge of the exposure status of the participant.

Minimizing Bias

Bias is not completely avoidable but we can take certain precautions to minimize it by careful design of the study especially in relation to selection of the study population and the methods of data collection.

Minimizing selection bias

The following precautions should help minimize selection bias:

1. Making sure that study participants are representative of the target population.
2. Making sure that response rates are as high as possible. If response rate is low, non-participants may be systematically different from participants which may mean that participants are not representative of the target population even if the original study population was. In such a situation, non-participants should be examined against participants for any systematic differences.
3. In case-control studies, making sure that the controls represent the population which produced the cases. We can check this by asking, "if a control had the outcome of interest, would they be a case?" The answer should be yes.
4. In cohort studies, maximizing complete follow-up and ascertainment of outcome. If a large proportion of subjects were lost to follow-up selection bias may occur. If loss to follow-up is high we must try to ascertain the outcome for a random number of those lost and check it against the outcome of those who remained. Systematic differences between these two groups indicates selection bias.

Minimizing information bias

A. Study design
   1. Binding of study participants if possible to the exposure and hypothesis (recall bias).
   2. Blinding of observers if possible to exposure or outcome (interviewer bias).

B. Data collection
   1. Using objective records to measure exposure if possible rather than relying on recall which may be less accurate.
   2. Collecting data on exposure and outcome as soon as possible after they have occurred as recall will be more accurate.
   3. Using standard questionnaires and trained interviewers to ensure uniformity in obtaining data.
   4. Proper questionnaire design to help more accurate data collection such as using closed questions.
   5. Using automated measuring devices to reduce observer bias such as automated blood pressure devices

*****