Potential errors in epidemiological studies
L 14

Associate Professor Dr. Eman A. Al-Kamil
Dep. Of Community Medicine
Collage of Medicine
Hashemite University

- Epidemiological investigations aim to provide accurate measures of disease occurrence (or other outcomes).

- However, there are many possibilities for errors in measurement.

- Epidemiologists offer much attention to minimizing errors and assessing the impact of errors that can not be eliminated.

Measurement error, bias and confounding

- What are these ?
- How these impact on observed associations
- Some examples
- Find our own examples
- Sources of error can be:
  - random or
  - systematic.

Random error

- Variability in human population means there is always some error, in gathering information on the exposure, outcome, and any coverable information.
- Greater random error leads to:
  - ➢ lower precision in estimates of association/effect
- Random error can be reduced by large sample size: bigger sample size gives more precise estimates .

Random error

- Random error is when a value of the sample measurement diverges – due to chance alone – from that of the true population value.

- Random error causes inaccurate measures of association.

- Random error can never be completely eliminated since we can study only a sample of the population.

There are three major sources of random error:

1. individual **biological variation**; always occurs and no measurement is perfectly accurate.

2. **sampling error**; is usually caused by the fact that a small sample is not representative of all the population's variables.

- The best way to reduce sampling error is to increase the size of the study group.

3. **Measurement error.**

Measurement error can be reduced by rigid protocols, and by making individual measurements as precise as possible.

- Investigators need to understand the measurement methods being used in the study, and the errors that these methods can cause.
- Ideally, laboratories should be able to document the accuracy and precision of their measurements by systematic quality control procedures.

Systematic error

- Systematic error (or bias) occurs in epidemiology when results differ in a systematic manner from the true values.
- A study with a small systematic error is said to have a high accuracy.
- Accuracy is not affected by sample size.
- The possible sources of systematic error in epidemiology are many and varied.

BIAS

- Bias is a major issue in any epidemiological study design.

- If it is overlooked, it could lead to incorrect conclusions and decision-making.
- Efforts must be taken to eliminate, reduce or at least recognize bias.
- It is not reduced by increasing the sample size

Definition of bias

Any systematic error in the design*, conduct* or analysis* of a study that results in a mistaken estimate of an exposure effect on the risk of disease (distortion of the results),and incorrect conclusions.

- Bias is an error in measuring or collecting information that differs systematically between groups of participants. Bias can result from the design, conduct, or analysis of a study.

- Bias reduces the accuracy (validity) of estimates of effect/association
- Bias can lead to over- or underestimating a true effect/association
- Even when you have reduced random error (increased precision) by having a large study, bias can still affect your results (as well as confounding and other problems).

Sources of bias : Bias may result from the following sources:

1. Faulty study design: selective sample
2. Within subject: position, stress, time of day
3. Within observer (Intra-observer): concentration, fatigue,
4. Between observers (Inter-observer): personality, training, visual acuity, digital preference
5. Methods/techniques: Position, cut off points
6. Apparatus: Type, condition
7. Recording: typing, coding,
8. Interpretation: Cut off point like 140/90 for hypertension
9. Hawthorne bias: Individuals who are aware of being participants of a study behave differently

Types of bias: Selection bias

- Selection bias occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not.
- An obvious source of selection bias occurs when participants select themselves for a study, either because they are unwell or because they are particularly worried about an exposure.

- It is well known, for example, that people who respond to an invitation to participate in a study on the effects of smoking differ in their smoking habits from non-responders; the latter are usually heavier smokers.

- In studies of children's health, where parental cooperation is required, selection bias may also occur.

- If individuals entering or remaining in a study have different characteristics from those who are not selected initially, or who drop out before completion, the result is a biased estimate of the association between exposure and outcome.

- An important selection bias is introduced when the disease or factor under investigation itself makes people unavailable for study.

- For example, in a factory where workers are exposed to formaldehyde, those who suffer most from eye irritation are most likely to leave their jobs.
- The remaining workers are less affected and a prevalence study or the association between formaldehyde exposure and eye irritation that is done only in the workplace may be very misleading.
- In such occupational epidemiology studies this important selection bias is called the healthy worker effect . Workers have to be healthy enough to perform their duties; the severely ill and disabled are usually excluded from employment.
- Similarly, if a study is based on examinations done in a health centre and there is no follow-up of participants who do not turn up, biased results may be produced: unwell patients may be in bed either at home or in hospital. All epidemiological study designs need to account for selection bias.

Hospital admission rate bias: in case control studies based on admitted cases, admitted cases tend to be severe and are likely to have been heavily exposed to risk factors as compared to the general population. This leads to over estimation of association between a risk factor and a disease.

Exclusion bias: If different exclusion or inclusion criteria are used in different study groups.
An example is the exclusion from controls but not from the cases of persons with specific previous exposure or characteristics.
This will lead to under presentation of these exposures among the controls but not among the cases. The same is true if different criteria are used for exposed and non-exposed.

Information Bias

It should be obvious that **if information from a study is incorrectly gathered, the conclusions from the study might be wrong.** There are various kinds of information bias.

Misclassification Bias

- Misclassifying exposure or outcome status is not a good thing.

- Suppose investigators conduct a prospective cohort study with four groups of subjects: those *who exercise vigorously, *moderately, *minimally or *not at all.
- Such groupings might be based on questionnaires completed by subjects and/or interviews with the subjects. **If these methods lead to misclassifications of exposure status, bias may have been introduced into the study.**
- **Outcomes can also be misclassified.** For example, incorrectly concluding that some subjects had an MI when they did not can lead to bias.
- **How much bias depends on the type of misclassification .**

Recall Bias

- This is a bias **unique to case control studies** that rely on **information provided by the subjects.**
- **The notion is that because subjects are aware of their health status as cases or controls, such knowledge might lead to a differential recall of an exposure status.**

- "Thus, a certain piece of information, such as a potentially relevant exposure, may be recalled by a case but forgotten by a control,"
- "Individuals who have experienced a disease or other adverse health outcome tend to think about the possible 'causes' of their illness and thus are likely to remember their exposure histories differently from those who are unaffected by the disease."

Reporting Bias

For a variety of reasons, including issues of social desirability or sensitivity, subjects may not be willing to report an exposure accurately.

When researchers gather baseline characteristic data, subjects may underestimate the amount of alcohol they drink, cigarettes they smoke, illicit drugs they use, etc.

Interviewer Bias

This occurs when data collection methods differ between groups.

In a case control study, for example, an interviewer might ask more inquiring questions of cases than controls, and such inquiring could lead to an overestimate or underestimate of a true exposure status.

Loss-To-Follow-Up Bias

Suppose an RCT has two groups, A and B, followed up to study the outcome being MI.

Suppose 30% of the subjects assigned to group A are lost to follow-up while only 10% of the subjects assigned to group B are lost to follow-up. Subjects assigned to group A might be lost to follow-up because they developed warning symptoms of an MI (unstable angina) and therefore left the study to seek treatment elsewhere.

If this occurs, group A is left with the subjects for analysis, i.e. the subjects who are less likely to develop an MI than those who left the study. The subjects remaining in the group A arm for analysis are therefore not representative of all the subjects originally assigned to the group. This could introduce bias into the study results.

Loss-to-follow-up bias is especially a concern if there is a difference in loss-to-follow-up between the treatment groups. If each group loses the same percentage of subjects, and the subjects leave for the same reasons, the study will not be affected in the same manner .

Measurement bias

- Measurement bias occurs when the individual measurements or classifications of disease or exposure are inaccurate – that is, they do not measure correctly what they are supposed to measure.

- There are many sources of measurement bias, and their effects are of varying importance.
- For instance, biochemical or physiological measurements are never completely accurate and different laboratories often produce different results on the same specimen.

- If specimens from the exposed and control groups are analyzed randomly by different laboratories, there is less chance for systematic measurement bias than in the situation where all specimens from the exposed group are analyzed in one laboratory and all those from the control group are analyzed in another.

- A form of measurement bias of particular importance in retrospective case control studies is known as recall bias.
- This occurs when there is a differential recall of information by cases and controls; for instance, cases may be more likely to recall past exposure, especially if it is widely known to be associated with the disease under study – for example, lack of exercise and heart disease.

- Recall bias can either exaggerate the degree of effect associated with the exposure – as with people affected by heart disease being more likely to admit to a past lack of exercise – or underestimate it – if cases are more likely than controls to deny past exposure.

- If measurement bias occurs equally in the groups being compared, it almost always results in an underestimate of the true strength of the relationship. Such non-differential bias may account for apparent discrepancies in the results of different epidemiological studies.

- If the investigator, laboratory technician or the participant knows the exposure status, this knowledge can influence measurements and cause observer bias .
- To avoid this bias, measurements can be made in a blind or double-blind fashion.
- A blind study means that the investigators do not know how participants are classified.

- A double-blind study means that neither the investigators, nor the participants, know how the latter are classified.

How to control for bias?

1.Be purposeful in the study design to minimize the chance for bias Example: use more than one control group

2.Clear definition of study population

3.Explicit case, control and exposure definitions.

4.Define, a priorly, who is a case or what constitutes exposure so that there is no overlap CC: Cases and controls from same population , Same possibility of exposure Cohort: selection of exposed and non-exposed without knowing disease status

5. Set up strict guidelines for data collection .Train observers or interviewers to obtain data in the same fashion.

6. Randomly allocate observers/interviewers data collection assignments.

7. Use multiple sources of information.

8.Institute a masking process if appropriate

- Single blind study
- Double blind study
- Triple blind study

9.Build in methods to minimize loss to follow-up.

10. Standardize measurement instruments.

Confounding

- Confounding is another major issue in epidemiological studies.

- In a study of the association between exposure to a cause (or risk factor) and the occurrence of disease, confounding can occur when another exposure exists in the study population and is associated both with the disease and the exposure being studied.

- A problem arises if this extraneous factor – itself a determinant or risk factor for the health outcome – is unequally distributed between the exposure subgroups.

- Confounding occurs when the effects of two exposures (risk factors) have not been separated and the analysis concludes that the effect is due to one variable rather than the other. To be a confounding factor, two conditions must be met,

- Confounding arises because non-random distribution of risk factors in the source population also occurs in the study population thus providing misleading estimates of effect .
- In this sense, it might appear to be a bias, but in fact it does not result from systematic error in research design.
- Age and social class are often confounders in epidemiological studies.

- An association between high blood pressure and coronary heart disease may in truth represent concomitant changes in the two variables that occur with increasing age; the potential confounding effect of age has to be considered, and when this is done it is seen that high blood pressure indeed increases the risk of coronary heart disease.

- confounding may be the explanation for the relationship demonstrated between coffee drinking and the risk of coronary heart disease, since it is known that coffee consumption is associated with tobacco use: people who drink coffee are more likely to smoke than people who do not drink coffee.

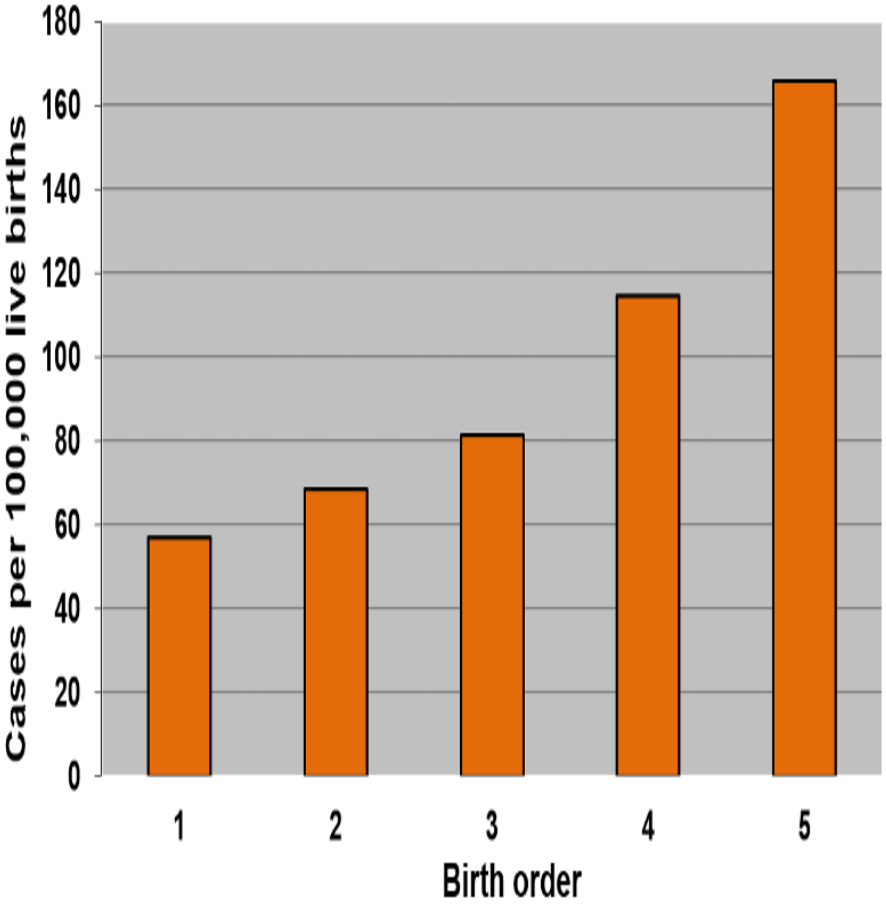
- It is also well known that cigarette smoking is a cause of coronary heart disease.
- It is thus possible that the relationship between coffee drinking and coronary heart disease merely reflects the known causal association of tobacco use and heart disease.

- In this situation, smoking confounds the apparent relationship between coffee consumption and coronary heart disease because smoking is correlated with coffee drinking and is a risk factor even for those who do not drink coffee.

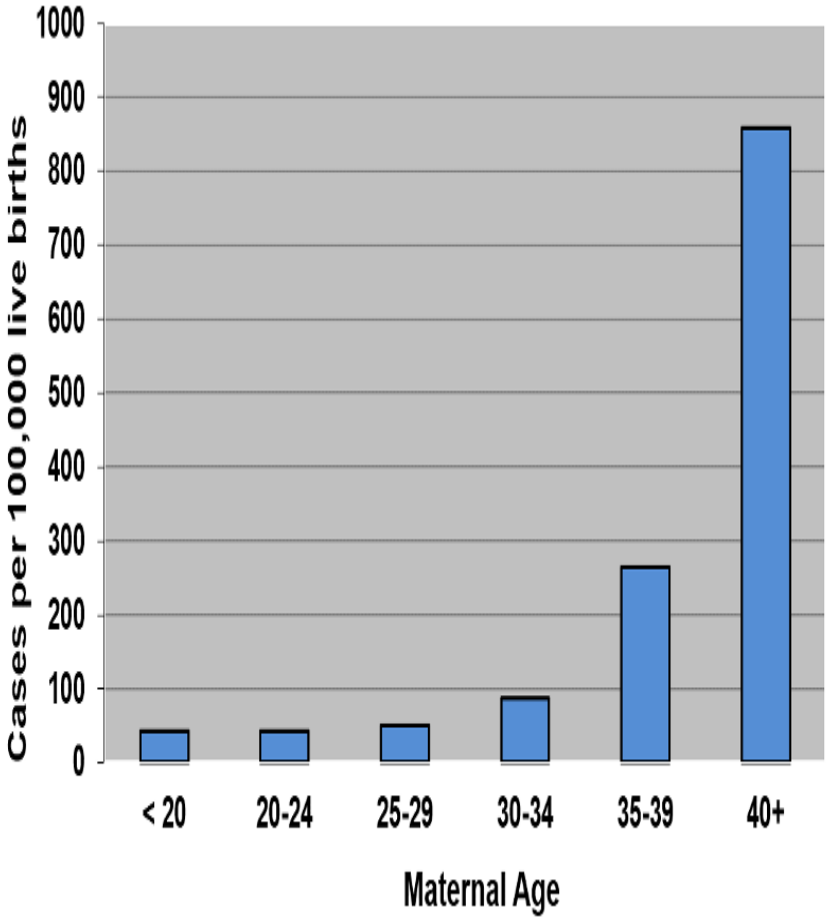
- Confounding can elevate, reduce or reverse an observed association.
- The difference between bias and confounding :

Bias creates an association that is not true, but confounding describes an association that is true, but potentially misleading.

The incidence of Down's syndrome by birth order



The incidence of Down's syndrome by maternal age



The control of confounding

Several methods are available to control confounding, either through study design or during the analysis of results.

The methods commonly used to control confounding in the design of an epidemiological study are:

1. Randomization
2. Restriction
3. matching.

At the analysis stage, confounding can be controlled by:

1. Stratification
2. statistical modeling.

Randomization

In experimental studies, randomization is the ideal method for ensuring that potential confounding variables are equally distributed among the groups being compared.

The sample sizes have to be sufficiently large to avoid random misdistribution of such variables.

Randomization avoids the association between potentially confounding variables and the exposure that is being considered.

Restriction

One way to control confounding is to limit the study to people who have particular characteristics. For example, in a study on the effects of coffee on coronary heart disease, participation in the study could be restricted to nonsmokers, thus removing any potential effect of confounding by cigarette smoking.

Matching

Matching is used to control confounding by selecting study participants so as to ensure that potential confounding variables are evenly distributed in the two groups being compared. For example, in a case-control study of exercise and coronary heart disease, each patient with heart disease can be matched with a control of the same age group and sex to ensure that confounding by age and sex does not occur.

Matching has been used extensively in case-control studies, but it can lead to problems in the selection of controls if the matching criteria are too strict or too numerous; this is called overmatching.

Matching can be expensive and time-consuming but is particularly useful if the danger exists of there being no overlap between cases and controls, such as in a situation where the cases are likely to be older than the controls.


Thank You!