Harvard Pilgrim Health Care Institute

CLINICAL EPIDEMIOLOGY AND POPULATION HEALTH

Key Points – Bias, Confounding, and Effect Modification

Bias:

Bias is any source of systematic error or misclassification in the determination of the association between the exposure and the outcome of interest. Bias can occur at many different points of a study.

Bias can be categorized according to the phase of study design in which it occurs.

Selection Bias occurs at the time study participants are selected for study entry. The samples being compared are not representative of the same populations at risk. This occurs when criteria for inclusion in one "arm" differ (maybe subtly) from the comparison group. Note: this is not the same as generalizability (external validity). A study may have no selection bias (the two samples differ by exposure but are otherwise representative of the same populations),but may not be generalizable to individuals who do not meet study inclusion criteria. Selection bias is often a concern in case-control studies.

Information, Measurement, or Ascertainment Bias occurs at the time of data collection.
Occurs when there is a lack of comparability in the accuracy or completeness of information between study groups. Examples are many but include:

- recall bias (how subjects provide information).
    - For example, people almost under-report their weight, or smoking status, and over-report their height. (Sometimes this type of bias is specifically labelled "social desirability bias," since people preferentially under-report socially undesirable behaviors or characteristics and over-report desirable ones.
- measurement error (misclassification because of inaccuracy in measurement)

We can further categorize information bias, or misclassification, as differential or non-differential, with respect to how biased measurement is related to other variables under study.

Nondifferential bias means that the frequency of errors is approximately the same in the groups being compared. For the exposure, the bias is nondifferential if it is unrelated to the occurrence or presence of disease. Similarly, misclassification of disease [outcome] is nondifferential if it is unrelated to the exposure. As one example, it is hard to measure long-term diet with complete accuracy; in a longitudinal cohort study of habitual fish consumption and cancer risk, there may be misclassification in exposure. As long as the same method of dietary assessment is used in those who go on to develop cancer as well as those who do not, then that would be an example of nondifferential exposure misclassification. In general, nondifferential misclassification tends to result in estimates of effect that are closer to "null" than the true effect (i.e. a risk difference closer to 0, or a RR/OR closer to 1, than you would expect if there were no nondifferential bias). The measurement error adds noise.

Bias is differential if the error in measurement occurs in only one group under study. For example, imagine that investigators perform a case-control study of whether recent fish consumption is associated with lower risk for sudden death. They ask controls (who are alive) about their recent diet, but for cases (who are dead) they obtain diet recall from a spouse or family member. Diet would be even less well measured in the cases than in the controls, resulting in differential bias in exposure assessment. Differential bias can result in estimates of effect that are smaller, or larger, than the truth, there is no way to predict.

Note: Bias is most often NOT something you can control for or mathematically estimate. You need to assess study designs carefully and think of all the possible biases. Bias is minimized by using carefully designed and deployed studies, but we never rid ourselves of it completely.

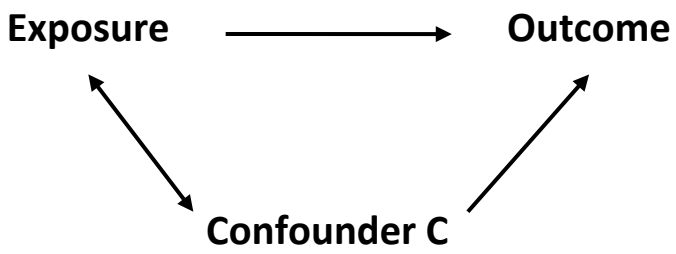
Confounder, Effect Modifier or Intermediate Variable (Mediator)?
Frequently, there are other variables that may actually be causally responsible for some or all of the apparent relationship between an exposure and outcome. We must assess whether each of these additional variables is…

A confounder?
An effect modifier?
An intermediate step in the causal pathway?
None of the above?

Confounding:
Confounding occurs when an apparent association between an **Exposure** and an **Outcome** is actually the result of a third factor—the **Confounder.**
A confounder should meet these 3 criteria:

- It is associated with the exposure under study
- It is a cause or correlate of the outcome under study, independent of the exposure
- It is not a natural intermediate step between an exposure and outcome, nor is it naturally upstream of the exposure or downstream of the outcome.

Visually, a confounder has this relationship to the other variables:



Confounding can be thought of as a distortion of the effect of one risk factor (exposure or predictor) by the presence of another. A confounding factor may mask an actual association, or falsely demonstrate one when it does not exist. Confounding is a nuisance – it hampers our ability to identify the true relationship between an exposure and an outcome. We want to minimize confounding, which can be done either at the time the study is designed (*e.g.,* restriction, matching, randomization) or when it is analyzed. Either way, we need to measure the potential confounder accurately in order to account for it. Two methods for dealing with confounding in the analysis are stratification and multivariable analysis (see below).

Effect Modification:
Occurs when the exposure-outcome relationship between two variables is different depending on the level (value) of a third variable.

For example, the relationship between smoking and lung cancer is stronger for men than for women. In this case, sex is a variable that we want to highlight and display (not minimize, as we would with confounding). We would want the researchers to present sex-specific results.

Detecting and Addressing Confounding and Effect Modification: Stratification
Stratify your sample by the variable that you suspect is a confounder or effect modifier, and assess the association between the predictor and outcome (for example, by calculating a Relative Risk).

1. If stratum-specific RRs/ORs are equal to each other AND equal to the crude RR/OR, the suspect variable is neither a confounder nor an effect modifier.
    - You can rely on the crude RR/OR.
2. If stratum-specific RRs/ORs are equal to each other but are <u>different</u> from crude RR/OR then the third variable is a confounder.
    - You cannot rely on the crude RR/OR.
    - You would report an "adjusted RR" or "adjusted OR" (which is the association of the predictor and outcome accounting (or controlling) for the confounder.
3. If stratum-specific RRs/ORs are <u>different from each other,</u> then effect modification is present.
    - The crude RR/OR is not telling the whole story.
    - You need the stratum-specific RRs/ORs to understand the relationship of the exposure and outcome at each level of the effect modifier.
    - You wouldn't highlight the result as an adjusted RR or OR because effect modification is not something you want to explain away, you want to highlight and understand the relationship in each group.

Note: if the "third variable" in question is a biologic intermediate or downstream consequence of a predictor, then it is not appropriate to stratify on that variable. For example, if we examine the association between number of alcohol drinks and motor vehicle crashes, we wouldn't try to control for blood alcohol level - it's on the causal pathway (drinking causes increased blood alcohol which in turn causes crashes.) It is neither a confounder nor an effect modifier – it is a <u>mediator</u>.

A variable can be a confounder in one context/research question or an effect modifier in another. It all depends on what you are asking and in whom. Sometimes it is acting as both confounder and effect modifier at the same time, but this is less common. Furthermore, this shouldn't matter because if it is an effect modifier, you'd want to stratify on it, and in that case you are also accounting for confounding.

Confounding can be addressed in several ways:

1. <u>In the study design stage</u>
    - Restriction:
        - Restrict the sample to those without the potential confounder or restrict to ensure the groups are similar with respect to the potential confounder – e.g. if sex and BMI are potential confounders, restricting to men with normal BMI (20-24.9 kg/m^2).
        - Downside: can lose generalizability, can't evaluate impact of variable you have restricted, may limit your sample.
    - Matching
        - Matching the two groups by potential confounders – e.g. for <u>every</u> case you have a control (or>1 control) matched by those factors.
        - Downsides: Can be time consuming and expensive, you can't evaluate impact of variable you are matching on, can limit sample size.
    - Randomization
        - Randomize patients to exposure
        - Downside: May not be practical. May not be ethical. Can be time consuming and expensive

2. <u>In the analysis stage</u>
    - Stratification:
        - Analyze the association of your exposure with your outcome within strata of the potential confounder
        - Downside – can only stratify on a small handful of variables, and stratification doesn't work for continuous variables
    - Statistical adjustment – more on this in the session on multivariable modeling