Chapter 17
Bias, Confounding, and Effect Modification (Interaction)

Stephen P. Glasser

You're like the Tower of Pisa-always leaning in one direction [1]

Abstract Bias, confounding, and random variation/chance are the reasons for a non-causal association between an exposure and outcome. This chapter will define and discuss these concepts so that they may be appropriately considered whenever one is interpreting the data from a study. Several types of common bias will be discussed (e.g. measurement bias, sampling bias, etc.) and effect modification (interaction) will be explained.

Keywords Bias • Confounding • Effect modification • Interaction

Introduction

Bias, confounding, and random variation/chance are alternate explanations for an observed association between an exposure and outcome. They represent a major threat to the internal validity of a study, and should always be considered when interpreting data. Whereas statistical bias is usually an unintended mistake made by the researcher; confounding is not a mistake; rather, it is an additional variable that can impact the outcome (negatively or positively; all or in part) separately from the exposure. Sometimes, confounding is considered to be a third major class of bias [2] (Table 17.1).

S.P. Glasser, M.D. (✉)
Division of Preventive Medicine, University of Alabama at Birmingham,
1717 11th Ave S MT638, Birmingham, AL 35205, USA
e-mail: sglasser@uabmc.edu

S.P. Glasser (ed.), *Essentials of Clinical Research*, DOI 10.1007/978-3-319-05470-4_17,

Table 17.1 Alternative explanations (other that truth) for observed associations between exposure and outcome

Variable	Description	Correction
Bias	A systematic error in the design, recruitment, data collection or analysis	The most important design techniques for avoiding bias are blinding and randomization
Confounding	A situation in which the effect or association between an exposure and outcome is distorted by the presence of another variable	In the design phase (e.g. during randomization, restriction or matching) or in the analysis phase (by stratification, multivariable analysis and matching)
Effect modification	A variable that differentially (positively and negatively) modifies the observed effect of a risk factor on disease status	Statistical testing for interaction
Random chance	A chance effect (random variation)	Diminishes as sample size gets larger. A small p-value and a narrow CIs are reassuring signs against chance effect-the same cannot be said for bias and confounding

As will be further discussed, when a confounding factor is *known* or suspected, it can be controlled for (albeit never perfectly) in the design phase (e.g. during randomisation, restriction or matching) or in the analysis phase (by stratification, multivariable analysis and matching). The best that can be done about *unknown* confounders is to use a randomised design (see Chap. 3). Bias and confounding are not affected by sample size, but chance effect (random variation) diminishes as the sample size gets larger. A small p-value and a narrow confidence intervals (CIs) are reassuring signs against chance effect but the same cannot be said for bias and confounding [3].

Bias

Bias is a systematic error that results in an incorrect (invalid) estimate of a measure of association. That is, the term bias 'describes the systematic tendency of any factors associated with the design, conduct, analysis, and interpretation of the results of clinical research to make an estimate of a treatment effect deviate from its true value' [3]. Bias can either create or mask an association; that is, bias can give the appearance of an association when there really is none, or can mask an association when there really is one. Bias can occur with all study designs, be it experimental, cohort, or case-control; and, can occur in either the design phase of a study, or during the conduct of a study. For example, bias may occur from an error in the measurement of a variable; confounding involves an incorrect interpretation of an association even when there has been accurate measurement. Also, whereas adjustments can be made in the analysis phase of a study for confounding variables, bias cannot be controlled, at best; one can only suspect that it has occurred. The most important design techniques for avoiding bias are blinding and randomization.

Table 17.2 Examples of bias

We all know about the common problems in doing research	
Selecting study participants	Information biases
Selection bias	Recall bias
Non-respondent bias:	Reporting bias
Volunteer or referral bias	Family information bias
External validity	Measurement bias
Sampling bias	Misclassification bias
Ascertainment bias	Reporting bias
Prevalence-incidence bias	End-aversion bias
Berkson bias	Attention bias
Healthy worker effect	
Detection bias: The risk factor investigated itself may lead to increased	
Diagnostic	
Overmatching bias	

Berkson's bias is a type of selection bias which may occur in case-control studies which are based entirely on hospital studies

An example of systematic bias would be a thermometer that always reads three degrees colder than the actual temperature because of an incorrect initial calibration or labeling, whereas one that gave random values within five degrees either side of the actual temperature would be considered a random error [4]. If one discovers that the thermometer always reads three degrees below the correct value one can correct for the bias by simply making a systematic correction by adding three degrees to all readings. In other cases, while a systematic bias is suspected or even detected, no simple correction may be possible because it is impossible to quantify the error. The existence and causes of systematic bias may be difficult to detect without an independent source of information; the phenomenon of scattered readings resulting from random error calls more attention to itself from repeated estimates of the same quantity than the mutually consistent incorrect results of a biased system.

There are many types of bias (Table 17.2), but two common types are; selection and observation bias [5].

Selection Bias

Selection bias is the result of the approach used for subject selection. That is, when the sample in the study ends up being different from the target population, selection bias is a cause. Selection bias is more likely to be present in case-control or retrospective cohort study designs, because the exposure and the outcome have already occurred at time of subject selection. For a case-control study, selection bias occurs when controls or cases are more (or less) likely to be included in study if they have been exposed – that is, inclusion in the study is not independent of the exposure. The result of this is that the relationship between exposure and disease observed among study participants is different from relationship between exposure and

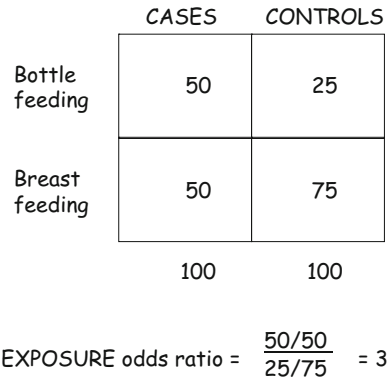

Fig. 17.1 Example of potential selection bias

disease in individuals who would have been eligible but were not included, thus the odds ratio from a study that suffers from selection bias will incorrectly represent the relationship between exposure and disease in the overall study population [2].

A biased sample is a statistical sample of a population in which some members of the population are less likely to be included than others. If the bias makes the estimation of population parameters impossible, the sample is a non-probability sample. An extreme form of biased sampling occurs when certain members of the population are totally excluded from the sample (that is, they have zero probability of being selected). For example, a survey of high school students to measure teenage use of illegal drugs will be a biased sample because it does not include home schooled students or dropouts. A sample is also biased if certain members are underrepresented or overrepresented relative to others in the population. For example, a "man on the street" interview which selects people who walk by a certain location is going to have an over-representation of healthy individuals who are more likely to be out of the home than individuals with a chronic illness. A biased sample causes problems because any statistic computed from that sample has the potential to be consistently erroneous [6]. Bias can lead to an over- or under-representation of the corresponding parameter in the population. Almost every sample in practice is biased because it is practically impossible to ensure a perfectly random sample. If the degree of under-representation is small, the sample can be treated as a reasonable approximation to a random sample. Also, if the group that is underrepresented does not differ markedly from the other groups in the quantity being measured, then a random sample can still be a reasonable approximation.

The word bias in common usage has a strong negative connotation, and implies a deliberate intent to mislead. In statistical usage, bias represents a mathematical property. While some individuals might deliberately use a biased sample to produce misleading results, more often, a biased sample is just a reflection of the difficulty in obtaining a truly representative sample [6].

Let's take as an example the data shown in Fig. 17.1, which addresses the question of whether otitis media differs in bottle-feeding, as opposed to breast feeding. 100 infants with ear infection are identified among members of one HMO, and the

controls are 100 infants in that same HMO without otitis. The potential bias is whether being included in the study as a control is not independent of the exposure, that is, they were not representative of the whole study population that produced the cases. In other words, one could ask the reason(s) that infants were being seen in an HMO in the first place and how many might have had undiagnosed otitis.

So, what are the solutions for selection bias? Little or nothing can be done to fix selection bias once it has occurred. Rather one needs to avoid it during the design and conduct of the study by, for example, using the same criteria for selecting cases and controls, obtaining all relevant subject records, obtaining high participation rates, and taking into account diagnostic and referral patterns of disease. But, almost always (perhaps always) one cannot totally remove selection bias from any study.

Observation Bias

While selection bias occurs as subjects enter the study, observation bias occurs after the subjects have entered the study. Observation bias is the result of incorrectly classifying the study participant's exposure or outcome status. There are several types of observation bias: recall bias, interviewer bias, loss to follow up, and differential and non-differential misclassification.

Recall bias occurs because participants with and without the outcome of interest do not report their exposure accurately (because they do not remember it accurately) and more importantly report the exposure differently (this can result in an over- or under-estimate of the measure of association). It is not that unlikely that subject's with an outcome might remember the exposure more accurately than subjects without an outcome, particularly if the outcome is a disease. Solutions for recall bias include using controls, who are themselves sick; and/or, using standardized questionnaires that obtain complete information and that mask subjects to the study hypothesis [7].

Whenever exposure information is sought, information is recorded and interpreted. If there is a systematic difference in the way the information is solicited, recorded, or interpreted, interviewer bias can occur. One solution to reduce interviewer bias is to mask interviewers, so that they are unaware of the study hypothesis and disease or exposure status of subjects, and to use standardized questionnaires or standardized methods of outcome (or exposure) ascertainment [8].

Loss to follow up is a concern in cohort and experimental studies if people who are lost to follow up differ from those that remain in the study (which is likely almost always the case). Bias results if subjects lost, differ from those that remain, with respect to both the outcome and exposure. The main solution for lost to follow up is to minimize its occurrence. Excessive numbers of subjects lost to follow up can seriously damage the validity of the study. (See discussion of lost to follow up in Chap. 3).

Misclassification bias occurs when a subject's exposure or disease status is erroneously classified. Two types of misclassification are non-differential (random) and differential (non random). Non-differential misclassification results in inaccuracies

Table 17.3 Questions to consider about bias

Could bias have occurred
Is bias actually present
Is bias large enough to distort the measure of association in an important way
Which direction is the distortion; toward or away from the null

with respect to disease classification that is independent of the exposure; or, with inaccuracies with respect to the exposure that are independent of disease. Non-differential misclassification makes the exposure and non- exposure groups more similar. The probability of misclassification may be the same in all study groups (non-differential misclassification) or may vary between groups (differential misclassification).

Measurement Bias

Let's assume that a true value does in fact exist. Both random and biological variation modifies that true value by the time the measurement is made. Performance of the instrument and observer bias, and recording and computation of the results further modifies the 'true value' and this now becomes the value used in the study. Reliability has to do with the ability of an instrument to measure consistently, repeatedly, and with precision and reproducibility. But, the fact is, that every instrument has some inherent imprecision and/or unreliability. This latter fact negatively impacts one of the main objectives of clinical research, to isolate between-subject variability from measurement variability. Measurement error is intrinsic to research.

In summary, in order to reduce bias, ask yourself these questions:' given the conditions of the study, could bias have occurred? Is bias actually present? Are consequences of the bias large enough to distort the measure of association in an important way? Which direction is the distortion, that is, is it towards the null or away from the null? (Table 17.3) [8].

Confounding

A confounding variable (confounding factor or confounder) is a variable that correlates (positively or negatively) with both the exposure and outcome. One, therefore, needs to control for these factors in order to avoid what is known as a type 1 error, which is a 'false positive' conclusion that the exposure is in a causal relationship with the outcome. Such a false relation between two observed variables is termed a spurious relationship. Thus, confounding is a major threat to the validity of inferences made about cause and effect, i.e. internal validity, as the observed effects should be attributed all or in part to the confounder rather than the outcome.

For example, assume that a child's weight and a country's gross domestic product (GDP) rise with time. A person carrying out an experiment could measure weight and GDP, and conclude that a higher GDP causes children to gain weight. However, the confounding variable, time, was not accounted for, and is the real cause of both rises [9]. By definition, a confounding variable is associated with both the probable cause and the outcome, and the confounder should not lie in the causal pathway between the cause and the outcome. Though criteria for causality in statistical studies have been researched intensely, Pearl has shown that confounding variables cannot be defined in terms of statistical notions alone; some causal assumptions are necessary [10]. In a 1965 paper, Austin Bradford Hill proposed a set of causal criteria [11]. Many working epidemiologists take these as a good place to start when considering confounding and causation.

There are various ways to modify a study design to actively exclude or control confounding variables [12]:

- Case-control studies assign confounders to both groups, cases and controls, equally. For example if somebody wanted to study the cause of myocardial infarct and thinks that the age is a probable confounding variable, each 67 years old infarct patient will be matched with a healthy 67 year old "control" person. In case-control studies, matched variables most often are the age and sex.
- Cohort studies: A degree of matching is also possible and it is often done by only admitting certain age groups or a certain sex into the study population, and thus all cohorts are comparable in regard to the possible confounding variable. For example, if age and sex are thought to be confounders, only 40–50 years old males would be involved in a cohort study that would assess the myocardial infarct risk in cohorts that either are physically active or inactive.
- Stratification: As in the example above, physical activity is thought to be a behavior that protects from myocardial infarct; and age is assumed to be a possible confounder. The data sampled is then stratified by age group – this means, the association between activity and infarct would be analyzed per each age group. If the different age groups (or age strata) yield much different risk ratios, age must be viewed as a confounding variable. There are statistical tools like Mantel-Haenszel methods that deal with stratified data.

All these methods have their drawbacks. This can be clearly seen in the following example: a 45 year old African-American from Alaska, who is an avid football player and vegetarian, working in education, suffers from a disease and is enrolled into a case-control study. Proper matching would call for a person with the same characteristics, with the sole difference of being healthy – but finding one would be an enormous task. Additionally, there is always the risk of over- and under-matching of the study population. In cohort studies, too many people can be excluded; and in stratification, single strata can get too small and thus contain only a few, non-significant number of samples [4].

An additional major problem is that confounding variables are not always known or measurable. This leads to 'residual confounding' – epidemiological jargon for incompletely controlled confounding. Hence, randomization is often the best solution

since, if performed successfully on sufficiently large numbers, all confounding variables (known and unknown) will be equally distributed across all study groups.

In summary, confounding is an alternative explanation for an observed association between the exposure and outcome. Confounding is basically a mixing of effect such that the association between exposure and outcome is distorted because it is mixed with the effect of another factor that is associated with the disease. The result of confounding is to distort the true association toward the null (negative confounding) or away from the null (positive confounding). It should be re-emphasized, that a variable cannot be a confounder if it is in the causal chain or pathway. For example, moderate alcohol consumption increases serum HDL-C levels that in turn, decreases the risk of heart disease. Thus, HDL-C levels are a step in the causal chain, not a confounder that needs to be controlled [8]. Rather, this latter example is something interesting that helps us understand the disease mechanism. In contrast, because confounding factors are nuisance variables (for example, smoking is a confounder of the effect of occupational exposures (to dyes) on bladder cancer), and therefore does need to be controlled for. That is, when confounders get in the way of the relation you want to study; one wants to remove their effect. Recall that here are three ways of attenuating the effect of a confounder. The first is with the use of a case-control design, in which the confounder is matched between the cases and the controls. The second way of attenuating the effect of a confounder is mathematically, by the use of multivariate analysis. And, the third and best way to attenuate the effect of confounding is to use a randomized design; but, remember "likely to control the effect of a confounder" means just that, it's not a guarantee.

Confounding by indication (treatment selection bias) is a bias frequently encountered in observational epidemiologic studies of drug effects. Because selection of treatments is not random and is determined by patient and physician characteristics, the observed effect is influenced by factors other than the treatment (that is the individuals at most risk are likely to be treated vs those at lesser risk), the resulting imbalance in the underlying risk profile between treated and comparison groups can generate biased results. A simple example is that subjects taking aspirin for primary prophylaxis might actually be found to have a worse outcome than the comparator group not receiving aspirin. But this latter observation might be influenced by the fact that patients taking aspirin might have had a higher disease risk burden. Once we control for disease severity and other confounders that determine who receives aspirin, we have a more accurate assessment of the relative effects of each treatment on outcome.

Confounding vs. Effect Modification

As discussed above, confounding is another explanation for apparent associations that are not due to the exposure. Also recall, that confounding is defined as an extraneous variable in a statistical or research model that affects the outcome measure, but has either not been considered or has not been controlled for during the study.

The confounding variable can then lead to a false conclusion that the outcome has a causal relationship with the exposure. Consider the example where coffee drinking is found to be associated with myocardial infarction (MI). If there is really no effect of coffee intake on MI but more coffee drinkers smoke cigarettes than non coffee drinkers, then cigarette smoking is a confounder in the apparent association of coffee drinking and MI. If one corrects for smoking, the true absence of the association of coffee drinking and MI will become apparent.

Effect modification (also referred to as interaction) is sometimes confused with confounding but with effect modification an apparent association between an exposure and outcome is "shared" with the confounder. Clinically, this can be expressed by understanding that the relationship between the exposure and outcome is different among different subgroups, or that there is a change in the magnitude of an effect according to some third variable. Referring back to the example above, let us say that coffee drinking and smoking impact on the outcome (MI). If one corrects for smoking, and there is still some impact of coffee drinking on MI, some association is imparted by cigarette smoking. In the hypothetical example above, let's say we find a RR of 5 for the association of coffee drinking and MI. When cigarette smokers are eliminated from the analysis and smoking is a confounder, the RR will be 1. In the case of effect modification where both coffee drinking and smoking equally contribute to the outcome (i.e. both smoking and coffee drinking have an equal impact on the association) the RR for each will be 2.5.

Summary

When examining the relationship between an explanatory factor and an outcome one is interested in identifying factors that may modify the factor's effect on the outcome (effect modifiers). We must also be aware of potential bias or confounding in a study because these can cause a reported association (or lack thereof) to be misleading. Bias and confounding are related to the measurement and study design. To review:

- **Bias**: A systematic error in the design, recruitment, data collection or analysis that results in a mistaken estimation of the true effect of the exposure and the outcome.
- **Confounding**: A situation in which the effect or association between an exposure and outcome is distorted by the presence of another variable. *Positive* confounding (when the observed association is biased away from the null) and *negative* confounding (when the observed association is biased toward the null) both occur.
- **Effect modification**: a variable that differentially (positively and negatively) modifies the observed effect of a risk factor on disease status. Different groups have different risk estimates when effect modification is present.

If the method used to select subjects or collect data results in an incorrect association, think bias, If an observed association is not correct because a different

(lurking) variable is associated with both the potential risk factor and the outcome, but it is not a causal factor itself, think confounding; and, if an effect is real, but the magnitude of the effect is different for different groups of individuals (e.g., males vs females or blacks vs whites), think effect modification.

References

1. Cited in "Quote Me" How to add wit and wisdom to your conversation. Compiled by Edward Breslin J. Ontario: Hounslow Press; 1990. p. 44.
2. Tripod.com. Available at: http://dorakmt.tripod.com/epi/bc/html
3. Dorak MT. Bias and confounding. Available at: http://www.dorak.info/epi/bc.html
4. Systematic error. In: Wikipedia. Available at: http://en.wikipedia.org/wiki/Systematic_bias
5. Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. Br Med J. 2002;325:1437–8.
6. Sampling bias. In: Wikipedia. Available at: http://en.wikipedia.org/wiki/Biased_sample
7. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32:51–63.
8. Aschengrau A, Seage GR III. Essentials of epidemiology in public health. 3rd ed. Available at: http://publichealth.jbpub.com/aschengrau/ppts/confounding.ppt10
9. Confounding. In: Wikipedia. Available at: http://en.wikipedia.org/wiki/Confounding_variable
10. Pearl J. Causality: models, reasoning, and inference. New York: Cambridge University Press; 2002.
11. Bradford Hill A. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295–300. PMC1898525.
12. Hennekens C, Buring J, Mayrent SL. Epidemiology in medicine. Philadelphia: Lippincott Williams & Wilkins; 1987.