International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-4, Issue-9, September 2018
ISSN: 2395-3470
www.ijseas.com

Bias and Confounding

Dr.S.Narayanan[1], Dr.C.Sabarigirinathan[2], Dr.K.Vinayagavel[3], Dr.D.Deepiha[4], Dr.S.BalaSiddharth[5], Dr.M.Saravanapriya[6]

[1](PhD student of Pacific University of Higher Education and Research (PAHER), Udaipur, Rajasthan)

[2](Professor and HOD, Dept of Prosthodontics, Tamilnadu Govt Dental College & Hospital, Chennai, India)

[3](Professor, Dept of Prosthodontics, Tamilnadu Govt Dental College & Hospital, Chennai, India)

[4&6](Post Graduate Student, Dept of Prosthodontics, Tamilnadu Govt Dental College & Hospital, Chennai, India)

[5](MBBS – Private, Chennai, India)

Abstract

Bias is defined as any tendency which prevents unprejudiced consideration of a question. In research, bias occurs when "systematic error is introduced into sampling or testing by selecting or encouraging one outcome or answer over this.
Confounding is a situation in which the effect or association between an exposure and outcome is distorted by the presence of another variable. This article explains in detail about Bias and Confounding.

Introduction

It is important to consider two sources of error when planning research: random error and bias. Bias occurs when the results of a study are systematically different from 'truth'. For example, if the objective of the study is to estimate the risk of disease associated with an exposure, and the result from the study consistently overestimates the risk, the result is said to be biased. Bias should be distinguished from random error, in that random error cannot be associated with a particular cause and tends to 'average out' in repeated sampling. Bias, on the other hand, would repeat the same direction of error in repeated sampling with the same design. Bias results from faulty design. There may be many reasons for bias, and care has to be taken to minimize bias when designing the study, since it is often difficult to separate the true effects from bias. Simply increasing the sample size, on the other hand, can minimize the effect of random error.

Types of bias

Several types of bias exist in research. Sackett et al. have listed 19 types of bias commonly encountered in epidemiological studies. Choi has expanded this list further to 65. Indeed, any type of error introduced into the study, for which a cause can be identified, could potentially be considered a bias by definition (systematic error). Many of these are hard to detect and even harder to avoid. We shall consider three specific types of biases, which are very common in health research.

Selection bias

Selection bias is a distortion of the estimate of effect resulting from the manner in which the study population is selected. This is probably the most common type of bias in health research, and occurs in observational, as well as analytical studies (including experiments).

a. Prevalence-incidence bias

This type of bias can be introduced into a case-control study as a result of selective survival among the prevalent cases. In selecting cases, we are having a late look at the disease; if the exposure occurred years before, mild cases that improved, or severe cases that died would have been missed and not counted among the cases. This bias is not often a problem in cohort studies and experiments, but is quite common in case-control studies.

Example:

The high case-fatality rate in the early stages of clinically manifested coronary artery disease may invalidate the study of possible etiological factors, since the persons available for study as cases are the survivors (severe cases are absent). Likewise, myocardial infarction may be silent. Clinical features may be absent, and the biochemical and electrocardiographic changes in myocardial infarction may return to normal after an infarct (these mild cases will not appear among cases for study). The type of bias introduced into the study may be clear by contrasting a cohort study (where the disease is identified in all its forms) as shown in Table 6.1.

TABLE 6.1 COHORT VERSUS CASE-CONTROL STUDY: ESTIMATES OF THE RELATIVE ODDS OF DEVELOPING CORONARY HEART DISEASE AMONG MEN WITH AND WITHOUT CHOLESTEROLAEMIA

	Cohort Study			Case-control study		
	Developed CHD	Did not develop CHD	Total	CHD present	CHD absent	Total
Highest quartile of serum cholesterol	85	462	547	38	34	72
Lower three quartiles of serum cholesterol	116	1511	1627	113	117	230
Total	201	1973	2174	151	151	302
Odds ratio	2.40			1.16		

b. Admission rate (Berkson's) bias

This type of bias is due to selective factors of admission to hospitals, and occurs in hospital-based studies. Many case-control studies collect cases from hospitals, and identify controls from among patients in the hospital admitted for unrelated events. The diseased individuals with a second disorder, or a complication of the original disease, are more likely to be represented in a hospital-based sample than other members of the general population. The causes of bias include the burden of symptoms, access to care, and popularity of certain institutions (particularly with respect to current practices of admission). Differential rates of admission will be reflected in biased estimates of the relative risks.

This type of bias is more common in observational studies, in particular case-control studies. Since the subjects are randomized after selection, this type of bias is less common among experiments.

Example:

Household interviews were performed on random samples of the general population asking about musculoskeletal and respiratory diseases and recent hospitalizations. In the general population, there appeared to be no association between these two disorders (OR = 1.06), but in the subset of the population who had been in hospital during the previous six months, individuals with musculoskeletal disorders were more likely to have respiratory disease than not (OR = 4.06). This occurred because individuals with both disorders were

more likely to be hospitalized than those with only one of the disorders. This finding is illustrated in Table 6.2.

TABLE 6.2 DISEASES OF THE BONE AND ORGANS OF MOVEMENT WITH AND WITHOUT RESPIRATORY DISEASE

		Diseases of bone and organs of movement					
		General population			hospitalized in previous six months		
		Yes	No	Total	Yes	No	Total
Respiratory disease	Yes	17	207	224	5	15	20
	No	184	2376	2560	18	219	237
	Total	201	2583	2784	23	234	257
Odds rat	io	1.06			4.06		

c. Non-response bias

This type of bias is due to refusals to participate in a study. The individuals concerned are likely to be different from individuals who do participate. Non-respondents must be compared with respondents with regard to key exposure and outcome variables in order to ascertain the relative degree of non-response bias.

Non-response bias is common in all types of studies, but is more serious in observational studies. In particular, sample surveys are more prone to this type of bias. If the non-response is similar in the exposure and non-exposure groups (or cases and controls), this may not be a serious problem. Sufficient information about related variables should be included in data collection instruments in order that we can verify the effect of non-response bias on the results. Maximizing the response rate in surveys is one way to minimize this type of bias. In randomized controlled trials, it is possible to collect information on related factors that might shed light on the seriousness of the problem by prospectively collecting information and comparing it.

Example:

In a mailed questionnaire study of the smoking habits of US veterans, it was noted that 85% of non-smokers, but only 67% of cigarette smokers returned the questionnaire within 30 days. Pipe and cigar smokers had an intermediate response rate.

Ascertainment or information bias

Information bias is a distortion in the estimate of effect due to measurement error or misclassification of subjects according to one or more variables. Some specific types of information bias are discussed below.

a. Diagnostic bias

Diagnostic bias may occur due to the performance of a disproportionately high number of diagnostic procedures on cases, as compared with controls. In a cohort study, knowledge of a subject's prior exposure to a possible cause may influence both the intensity and the outcome of the diagnostic process. Knowledge that an individual has worked in the rubber industry, for instance, may lead to a more intensive search for bladder cancer than would occur if the person had worked in another industry.

In a case-control study, if the disease outcome is one with few clinical manifestations, and requires laboratory tests or diagnostic procedures to detect it, the disease may be missed in the control group if they are not adequately examined prior to inclusion in the study. For example, in order to ascertain the presence of endometrial cancer in individuals exposed, or not exposed to estrogen therapy, the same diagnostic procedures must be performed for both groups at the same frequency. This bias can be reduced by having the control group selected from persons who went through the same diagnostic procedures as did the case group, and by using only those with negative results as controls.

Similar bias can also occur in experimental studies, although this is rare, due to the development of, and strict adherence to study protocols that avoid these types of problems. In general, 'blinding' of persons who are reporting tests, by denying them clinical information about which are cases and which are controls (or to which treatment group they have been allocated), and submitting cases and controls to equally rigorous diagnostic preparation, will help reduce this type of bias.

b. Recall bias

An error of categorization may occur if information on the exposure variable is unknown or inaccurate. Ascertainment of exposure to drugs by history alone, recollection by controls of exposure variable, and a more intense search by investigators for exposure variables among cases, may lead to this type of bias. The recall by both cases and controls may differ in both amount and accuracy. Cases are more likely to recall exposures, especially if there has been recent media exposure on the potential causes of the disease.

Example:

In questioning mothers whose recent pregnancies had ended in fetal death or malformation (cases), and a matched group of mothers whose pregnancies had ended normally (controls), it was found that 28% of the former, but only 20% of the latter reported exposure to drugs. This could not be substantiated either in earlier prospective interviews or in other health records.

This type of bias can be avoided by strict adherence to a developed protocol, administered in a standard fashion by 'blinded' investigators, and by using recorded data to supplement information obtained from records and interviews.

Effect of selection and ascertainment bias on odds ratios observed in case-control and cohort studies

The biases mentioned in the previous section can alter the odds ratio, and thus potentially lead to an invalid conclusion. The potential effect is illustrated (in general terms) in Table 6.3.

6.3 EFFECT OF BIAS ON ODDS RATIOS OBSERVED IN CASE-CONTROL AND COHORT STUDIES

Type of bias	Effect on odds ratio	
	Case-control	Cohort
Selection bias		
Prevalence-incidence	⇑ or ⇓	unlikely
Berkson's bias	⇑ or ⇓	not applicable
Non-response	⇑ or ⇓	⇑ or ⇓
Measurement bias		
Diagnostic	⇑	⇑
Recall	⇑	not applicable

The prevalence-incidence bias can either increase or decrease the odds ratio in a case-control study, but this is unlikely to occur in a cohort study or experiment. The non-response bias can influence both case-control and cohort studies, as well as experiments, and can occur in either direction. Selection biases are the most difficult to avoid. The prevalence-incidence bias cannot be prevented in a casecontrol study, but is at least partially measurable. The admissionrate bias is neither preventable nor measurable. Non-response bias can be both prevented and measured.

Of the ascertainment biases, the diagnostic bias will inflate the odds ratio in both case-control and cohort studies. Recall bias will also inflate the odds ratio in a case-control study, but is not applicable to a cohort study. Both of these biases are preventable.

Selection biases make it impossible to generalize the results to all patients with the disorder of interest, while the measurement biases influence the validity of the study conclusions.

Since biases are difficult to control in most cases, care should be taken to prevent their occurrence by the choice of appropriate design, development of strict protocols and adherence to these protocols. In the worst case, when these biases cannot be prevented, the potential biases should at least be measured, and possible statistical adjustments of results considered.

Confounding

Confounding is a special type of bias. The effect of the factor under consideration is mixed up with effects of other factors not directly relevant to the study question. An exposure, E is said to be confounded with another factor, C with respect to its effect on a disease, X, if both C and E are associated with the disease, and C and E are associated with each other. The confounding is manifested in the study results when the factor, C appears unequally among the exposed and unexposed groups; the comparison of disease incidence or prevalence in the two groups is mixed with the different presence of the factor, C. This is the only type of bias that can often be corrected (if appropriate measures have been taken during the study) by statistical adjustments.

An important consideration when dealing with confounding is that both factors are potential risk factors for the disease; which one is the cause and which is confounding depends on the study objective. For example, when studying the effect of exposure to asbestos dust (working in asbestos mines) on lung cancer, cigarette smoking is a confounder. We know that cigarette smoking is closely associated with lung cancer, and that miners tend to smoke more often than nonminers. On the other hand, if the question of interest was the association of smoking and lung cancer, exposure to asbestos dust could be a confounder.

Confounding is a form of bias, and therefore affects the validity of the study; estimates of the risk coefficients may be systematically higher (or lower) than the true risk. Adjusting for confounding will improve the validity but reduce the precision of the estimates. Since it is possible to adjust statistically for confounding, if information on the potentially confounding variables has been collected, there is a tendency to adjust for all potential confounders. This is counterproductive: one would lose statistical power (precision) and might not gain much in terms of validity if the factors considered were not confounders. Before adjusting for confounders therefore, both conditions for confounding should be verified. For a detailed discussion on confounding, see Kleinbaum, Kupper and Morganstern.

When designing a research project, therefore, careful consideration should be given to what are the risk factors of interest, and what could be potential confounders (known risk factors that are of no particular interest in the present study, and that might have an association with the hypothesized risk factors). Being a type of bias, it is best to avoid the problem if we can, and to collect relevant information if we cannot avoid the problem.

Example:

Suppose that one wants to investigate a postulated causal connection between alcohol consumption and myocardial infarction. Smoking is known to be a cause of this disease; alcohol intake and smoking are known to be correlated. Suppose that alcohol consumption is in fact not a cause of myocardial infarction. By virtue of its association with smoking, however, alcohol intake would be found to be associated with, and apparently to increase

the risk of the disease. One might even find an apparent dose-response relationship between alcohol intake and myocardial infarction, since heavy drinkers are often heavy smokers as well. In order to disentangle the effects of smoking and alcohol intake, one may stratify the subjects (both cases and controls) into smoking and non-smoking groups, and within each subgroup, look for an association between alcohol intake and myocardial infarction. Table 6.4 illustrates the effect of confounding in this situation.

TABLE 6.4 RELATIONSHIP OF ALCOHOL CONSUMPTION TO MYOCARDIAL INFARCTION (MI)

A. Ignoring smoking

Alcohol intake	MI	Control
Yes	71	52
No	29	48
Total	100	100

Odds ratio = 2.26, $\chi^2 = 7.62$, P = 0.006 (2-sided).

B. By smoking status

Alcohol intake	Non-smokers		Smokers	
	MI	Control	MI	Control
Yes	8	16	63	36
No	22	44	7	4
Total	30	60	70	40
Odds ratio	1.0		1.0	

The statistically significant elevation in risk (OR = 2.26, P<0.01) in the analysis that ignores smoking is spurious. Among non-smokers, the estimated OR for myocardial infarction being associated with alcohol intake is 1.0, with an identical estimate among smokers. The effect in Table A is therefore said to be due to confounding with smoking. One may regard the subgroup specific ORs in Table 6.4 as representing the effect of alcohol, 'adjusted for smoking' on the risk of myocardial infarction. Conceptually, the effect of smoking has been held constant, although not in an experimental sense. If the two ORs (with and without smoking) were not the same, a pooled estimate of the effect of alcohol intake on MI would not be easy. Statistical methods of adjustment

incorporate the use of standardization (using some hypothetical population as standard, so that both exposure and non-exposure groups have a similar distribution of the confounding factor). A common method of such standardization is the MantelHaenzel adjusted odds ratio (see Kleinbaum, Kupper and Moganstern for details).

Options for control of confounding in observational studies

Several methods are available for the control of confounding, either by preventing confounding or by adjusting for it in the analysis.

1. Restriction by study design

This approach to control simply involves specifying narrow ranges of values for one or more extraneous variables in determining admissibility into the study (e.g. restriction to white males only, or to ages between 40 and 50 years). The restriction applies to both index and comparison groups (cases and controls, or exposed and unexposed). This has the effect of removing the confounding variables and retaining a relatively homogeneous group for comparison. The disadvantage of this approach is that the generalizability of the study is limited to the narrow group included in the study. While the study would have external validity to the narrowly defined population, it would not be very useful for the general population of interest.

2. Matching

Matching involves the use of constraints in the selection of the comparison groups, so that the index and comparison groups have similar distribution with respect to the potentially confounding variable.

A common example is when the controls are selected to match the cases for age and gender. By making such a choice, age and gender will no longer be confounding variables (even though they may be associated with the disease, the association of the exposure to the disease is not confounded by these). While this is less restricted than selecting a narrow population of interest, it imposes the restriction that the population of interest is limited to what has been observed in the index group.

Analysis of results from matched studies will need to incorporate the matching design (the two groups are not statistically independent) and often precision is reduced. For example, if 100 cases and 100 controls are used in a matched study, this is like having only 100 observations (100 matched pairs) and the statistical power is approximately 60% compared with the unmatched study with 100 cases and 100 controls. Therefore, matching has to be done judiciously. In addition, matching for several variables simultaneously can lead to serious ‘overmatching’ in that any potential association gets washed out, and results are never statistically significant.

3. Stratification in the analysis without matching

This option essentially involves restriction of the analysis (rather than the sampling scheme) to narrow ranges (strata) of the extraneous variable. Pooling of the results from the various strata may be possible, if there is no interaction between the two factors. An example was presented in Table 6.4.

4. Mathematical modelling in the analysis

This approach involves the use of advanced statistical methods of analysis, such as multiple linear regression, logistic regression, etc. This is a form of stratification in the analysis and pooling of the information, except that the stratification and pooling is done under the assumption of some mathematical form of relationship. Specific types of relationships may be explored by these methods, and these can be statistically more powerful than the individual stratified analysis. For more details, see Kleinbaum, Kupper and Morganstern, or Hosmer and Lemeshow.

Recommendations for minimizing bias in analytical studies

1. Cases should be limited to incident cases, and should be chosen as homogeneous entities or as random samples of all cases.
2. Definitions, ascertainments and exclusions must always be made explicit, and this should be done in advance.
3. At least two control groups should be chosen:

a. a hospital-based group, preferably from among patients who have undergone the same diagnostic procedures as the cases; controls may either be matched to the cases, preferably on a stratified basis, or chosen as a random sample of potential controls;

b. a community-based control group.

4. Analysis should be complete. All known potential confounders, if not already considered in the matching process, should be the subject of analysis by stratification or multivariate techniques.

References and further reading

Fletcher H.R., Fletcher W.S., Wagner H.E. Clinical epidemiology: the essentials. London, Williams and Wilkins, 1982.

Hosmer D.W., Lemeshow S. Applied logistic regression. New York, John Wiley and Sons, 1989.

Kleinbaum G.D., Kupper L.L., Morganstern H. Epidemiologic research: principles and quantitative methods. New York, Van Nostrand Reinhold, 1982.

Sackett D.L., Haynes R.B., Guyatt G.H., Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston, Little, Brown and Company, 1991.