Errors in research

Arun Sharma[1]

The purpose of medical research is to find the truth, hitherto unknown, but once discovered, will benefit humanity in improving its health. However, there are possibilities of missing out on the truth, if we are not aware of possible mistakes that the researchers can make or errors that may creep in through the practice of research and such errors may lead to a false belief of having discovered the truth where as in reality we might have failed to do so. For example, if at the end of a research work, the researchers conclude that the risk of developing retinopathy increases seven times if the Oral Hygiene Index is poor among diabetic patients. But on critically examining the data and the research methodology, the reviewers of this manuscript discover that there were errors in research methods and the collected data is erroneous, and therefore the 'seven fold increase in risk' is not true. Had the reviewers failed in detecting this error, we would have lived with the belief of this relationship being true. This is the peril of not being correctly able to identify if any error in the research work has crept in knowingly or unknowingly. So, in this paper, I will try to help you detect errors in research and remedies for them, if any.

At the outset, I must mention that the errors that I am talking about, are based on two principles. First, the quantities that we measure in our research subjects vary across and within individuals. Second, most research is carried out on a subset of target population, known as a sample and one subset of sample may have different set of values of the variable then the other subset and so conclusions drawn from two subsets of samples will be different from each other. Since samples are chosen randomly in most of the quantitative research, these are called random errors. Random errors are inevitable and we have to live with it as there is no remedy for this save that the sampling is eliminated and entire target population is included in research, which is impractical in most research scenarios. But it is indeed possible in very rare diseases when data on all patients may be available. The other type of error is known as systematic error. Systematic error is also known as bias. One school of epidemiologists believe that Confounding is also a type of systematic error, but is debated by the other school. It will be discussed later.

To begin with, bias is a systematic error (in contrast to random error), if present, distorts the quantitative result as either an overestimation or an underestimation of the measured parameter.Broadly, bias can be either selection bias or otherwise it can be information bias.

When a bias creeps in due to errors in selection of subjects for the study, it leads to **selection bias**. In sample based research, subjects recruited in the study must be true representative of the target population. But when this tenet is violated during research, it leads to selection bias. It can occur in various scenarios. Suppose you want to recruit cases and controls for a study to determine the association between occurrence of diabetes and poor oral hygiene as its risk factor. The study is being conducted in a government tertiary care hospital in Kolkata. Patients who have diabetes, because of their ill health are unable to maintain good oral hygiene, and are relatively ill educated about the need of maintaining oral hygiene. So these patients will be over representing the cases with poor oral hygiene. On the other hand, controls, who are free of diabetes have better control of their life and therefore may be practicing oral health care not only because they are healthier, but also because in course of treatment they have been instructed to take care of oral hygiene. But as representatives of pool of diabetes patients, the cases may not represent patients who visit private hospitals for treatment. By virtue of having better paying capacity, more educated, it is implied that they are more likely to be better informed about maintaining oral hygiene, and thus oral hygiene status of these diabetic retinopathy cases will be better than that of government hospital attendees. So, both reasons applied together, will result in a bias in selection/ recruitment of cases and controls with respect to their oral hygiene status. In this particular case, since poor oral hygiene will be over represented in retinopathy cases compared to controls, the calculated Odds Ratio will be an overestimation of true Odds Ratio.

The other type of bias is **information bias**. Data collected from subjects recruited in the study pertain to values of different variables in consideration. These values are generated from either answers given by the subjects from their recall of the past events or they could be

[1]Deptt. of Community Medicine, University College of Medical Sciences, New Delhi, India.
Corresponding Author : Prof. Arun Sharma, E-mail: arsharma62@gmail.com
Received on : 03/10/2016, Accepted on : 30/11/2016
Conflict of Interest : None, Financial Disclosure : None

measurements performed on the body or values of estimation of markers obtained from the body. Each such act has a potential threat of bias, and thus results in myriad biases. Some of them are as follows:

When data are collected by asking questions from the subjects abut events occurred in the past, the possibility of getting wrong information can be attributed to several factors, each one of them becoming a possible source of bias. Commonest being, inability of the subjects to recall past events. Such **recall bias** is common when we are interviewing subjects about their habits like drinking, smoking, food habits etc. If the patients of diabetic retinopathy come to know that irregular treatment of coexisting hypertension might increase the risk of retinopathy. These patients will work hard to recall if they missed anti hypertensive drugs and will truthfully report so. But patients who have not yet developed retinopathy, may chose to say that there antihypertensive drug in take was uninterrupted, without giving much thought to it. This differential recall will over assess the irregularity in taking anti hypertensive drugs among patients with retinopathy compared to retinopathy free persons who are controls in this study,thereby resulting in overestimation of Odds Ratio.

Similarly one more information bias is known as **interviewer bias**. This is found more commonly in intervention research, where the interviewer may be interested in establishing a positive effect of a new drug because of business interest. Or, a novice researcher may not put due emphasis on maintaining neutrality while asking questions. For example, while studying effect of occupational exposure to radiation on occurrence of cataract, the interviewer may probe deeper with a cataract patient in assessing radiation exposure and in non cataract controls, he/she may accept the response on face value without cross examining. It has been reported that in intervention research, positive response to the trial drug was ascertained with more emphasis than with that of placebo. Such scenarios are now being avoided with the help of blinding.

Measurement Bias: As the name suggests, this bias occurs due to errors in measurement. For example, while recording blood pressure, if a particular data collector is a little hard of hearing, he may not recognize the Korotkoff sounds as accurately as other data collectors, so his recordings will always be different from that of the others, and he may be consistently recording higher or lower value. This will introduce a systematic error in the recorded blood pressure.

Instrument bias: Similar to measurement bias, instrument bias occurs when the readings taken by a faulty device, which may consistently be under or over estimating. For example, if the spring of a spring balance loses its elasticity partially, it will give underestimate of weight or if the calibration is not standardized, device may give erratic results.

Lead time bias: It is a special type of bias associated with screening and/or diagnostic tests. We emphasize on early diagnosis and treatment. Consider that retinoblastoma can be diagnosed using ophthalmoscope only after 1 year of onset, but it can be diagnosed after 3 months using PETT scan. Now this has increased the length of known disease period by 9 months. If we compare the mortality of the cases in the two scenarios, survival will be considered extended by months for the PETT scan diagnosed cases compared to clinically diagnosed cases. Whereas, there is no true increase in the duration of survival of the patient, it is only that the diagnosis has been 'preponed'. As means of alteration of course of disease may not exist, there may not be true gain in survival time for the patient. Hence this apparently increased survival time is a biased estimate of survival time.

Berkson's bias: Usually case control studies are hospital based. In hospital based studies, cases and controls are selected from the hospital. The controls are expected to represent general population. But, people visiting the hospital and are available for selection as controls may not be true representative of the general population. As a result, a differential selection of controls may occur with respect to the risk factor in question, resulting in an incorrect estimate of the association. To illustrate, suppose a study is being conducted to establish relationship between bad oral hygiene and occurrence of diabetic retinopathy in a certain hospital. The oncology and periodontology departments of this hospital have a reputation in the community, so patients with poor oral hygiene and oral cancers prefer this hospital to others. Hence these patients are more in proportion to other patients who could be randomly representing the general population. As a result, those selected as controls are more likely to have poor oral hygiene (as poor oral hygiene is associated with gum diseases and oral cancer). Thus the putative risk factor in question, poor oral hygiene will be more frequently present among controls, thereby resulting in an underestimation of strength of association (Odds Ratio).

Hawthorne Effect: This is specific to intervention studies. When the experimental subjects are observed for their behavior and they are aware that they are being observed, they tend to behave in a manner which is desirable. Suppose a study is being conducted to see if cognitive functions are improved by giving a medication to alcohol intoxicated persons. If the subjects are aware that they are being observed, they may try extra hard to prove that their cognitive functions are improving. This will result in an overestimate of impact of theintervention.

Survival bias/incidence prevalence bias: It is practically difficult to interview patients who are critical. Severely ill patients may die sooner than the less severely ill patients. As a result, such cases will always be missed out in case control studies. In other words, patients who have severe acute episodes of fatal illnesses, will be unlikely to participate in the study, as a result exposure assessment of such patients cannot be done. Hence studies will be biased with respect to exposure among cases.

Attrition bias: In longitudinal studies, where repeated observations are made on patients, patients may be lost to follow up either due to premature death due to other diseases, voluntary withdrawal from the study or migration. When such attrition is disproportionate with respect to cases and controls or intervention and control group, it leads to bias. If withdrawal is proportionate, then there will be no bias.

Publication Bias: There is a common tendency among journal managers to publish positive findings along with reluctance to publish negative findings. As a result, the literature review or meta analysis of published literature becomes biased towards positive association. We are also aware of biases arising out of favoritism in publication. The other type of publication bias emerging recently is from the high cost of publication in some of the journals. Researchers from developing countries do not have adequate funds to pay for publication. Hence good research may remain unpublished.

Social desirability bias: Sometimes answers are given by respondents which are in conformity with the social norms. This is especially true with respect to behavior that finds restricted acceptance in the society. For example, information regarding sex partners, use of banned drugs, missing out on adherence to drug treatment are usually lied about, thus resulting in underestimation of such events.

Confounders: There are certain variables, whose presence may alter the relationship between a set of predictor and outcome variables, resulting in under or over estimation of the association. It is not a bias, because it is a naturally occurring phenomenon, which cannot be removed, we have to live with it. But it is indeed a source of erroneous reporting of the association between a possible risk factor and outcome. Hence it is included in the category of errors. Such variables (also known as confounding variables), are those variables which are associated with the predictor variable as well as the outcome variable. For example, coffee drinking may be a risk factor for peptic ulcer and so is smoking. But if coffee drinkers have more chances of being smokers as compared to non coffee drinkers, then smoking, in this case being associated independently with coffee drinking (predictor variable) as well as peptic ulcer (outcome variable), it will become a confounder.

How to handle bias and confounding

Bias can be prevented. If at the time of conceptualization of the research, we are aware of the possible biases, we may at the time of designing the study ensure that the bias is removed. If we know that there is a risk of selection bias in a given study by virtue of selecting the controls from the same hospital, we may avoid the bias by selecting controls from general population. Regular checking and calibration of measuring instruments will eliminate instrumental bias. Proper training of interviewers and supervision of data collection work will eliminate some of the information bias. Attrition bias is difficult to eliminate, as we cannot force anyone to continue to remain in the study. So the remedy is to over recruit. Suppose the sample size is 100 and there is a possibility of 20% respondents not being able to remain in the study for the entire duration, then we may recruit 125 in this case so that even if 20% are lost to follow up, we are still left with 100. If we apprehend recall bias with respect ot a particular variable, the recall period should be shortened. But these methods of addressing biases are not universal. Each study has to be examined at the stage of developing the methodology to preempt biases and look for methods of its avoidance. If biases are recognized at the analysis stage, it may be corrected. And if it cannot be corrected, it is fair on the part of researchers to apprise the readers of possibility of occurrence of bias in the study so that inferences are drawn keeping the limitations in mind

The confounding effect of a variable can be detected simply by stratifying the data with respect to the confounding variable. Mantel Haenszel Chi square test is a method for adjusting for potential confounders of categorical variety. The data can be stratified for the confounder variable and adjusted measure of association, for example adjusted Odds Ratio for case control studies and adjusted Rate Ratio for cohort studies will help detect the confounding effect. The details of carrying out adjusted analysis are beyond the scope of this write up. The other method of adjustment to address confounding is multivariate regression analysis.

Cite this article as:
Sharma A. Errors in research. Current Indian Eye Research 2016; 3:87-9.