Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/injms

Review Article

Biases in epidemiological studies: How far are we from the truth?

*Gunjan Kumar, Anita Shankar Acharya**

Department of Community Medicine, Lady Hardinge Medical College, New Delhi, India

ARTICLE INFO

Article history:
Received 22 July 2013
Accepted 8 August 2013

Keywords:
Bias
Epidemiology
Confounding factors
Selection bias
Information bias

ABSTRACT

In any research occurrence of 'Bias' is inevitable. Bias is an error that occurs in a systematic way during the design, implementation or interpretation of the study that deviates the results away from the actual facts. Biases are broadly of two types, i.e. selection bias and information bias. Selection bias occurs when the population selected for the study is not representative of the target population. Information bias occurs when there is a systematic difference in the way data is gathered from the subjects. Confounding is another factor which though not a bias may also cause deviation from truth. There are various methods of minimising bias like adhering to a strict protocol, good sampling methods, proper training of the researchers, double entry of the data and minimising follow up, whereas confounding can be minimised by randomisation, restriction, matching and statistical modelling. This article describes the various types of biases, their occurrence and how they can be minimised so that the research output is robust.

1. Introduction

Bias literally means "Distortion of statistical result".[1] For the purpose of epidemiological studies bias has been defined as "Deviation of results, or inferences from the truth, or processes leading to such deviation"[2] or "Any systematic error in design, conduct or analysis of study that results in mistaken estimate of an exposure's effect on risk of disease".[3] Bias is different from random error that occurs only due to chance. Random errors are more or less nullified in the long run, but systematic errors tend to produce results different from truth. With increasing interest in the area of research, bias becomes an important problem that may not let us reach the truth. This article has been written to provide an overview of "BIAS", how it is introduced in studies, and what we can do to minimise bias.

2. Methods

A thorough literature search using key words "bias", "confounding", "epidemiology" was conducted on various web based platforms and libraries. The full texts of articles and

* *Corresponding author*. 38, Engineer's Enclave, Pitampura, New Delhi -110034, India.
E-mail address: anitaacharya29@gmail.com (A.S. Acharya).
http://dx.doi.org/10.7713/j.ijms.2013.0039

books were looked into and reviewed. Cross-references were looked into and sought. The information has been summarised for the benefit of the researchers.

2.1. *Principles and type of biases*

As we know, bias deviates the results from the truth, which brings in lack of internal validity in the study. Validity is an expression of the degree to which a test is capable of measuring what it intends to measure.[4] A study is valid if its results correspond to the truth. Internal validity is to apply your results to the particular group of population being studied. Presence of internal validity does not always promise external validity which means that the results are applicable to the universe. Biases can be broadly classified into two types[2,5]:

1. Selection Bias
2. Information Bias/Measurement Bias

3. Selection bias

Selection bias can occur due to (a) Choosing an ill-defined study population, (b) During sampling, (c) Due to diagnostic procedures, and (d) During the conduct of the study. Selection Bias is a distortion in the estimate of association between risk factor and disease that results from how the subjects are selected for the study. Selection bias could occur because the sampling frame is sufficiently different from the target population or because the sampling procedure cannot be expected to deliver a sample that is a mirror image of the sampling frame. Selection bias can occur due to various reasons that are enlisted below[3]:

3.1. *Ill defined study population*

In order to focus the research, the researcher needs to define the population, whether the whole population is being studied or a sample is being studied. Undefined population leads to selection of a non-representative sample leading to a bias. This bias is mainly important in the hospital based/institution based studies when the subjects involve only the people reporting to a certain facility.[3] Few examples are enlisted below:

a. **Health care access Bias** which occurs when cases admitted to a certain health facility do not represent the cases that originate in a community. May occur due to web of causes when people due to lack of awareness, cultural and geographical factors are not able to access a particular facility (Access Bias), when admissions are based on interest of a clinician (Popularity Bias), when people are attracted to success of certain clinician (Centripetal Bias) or if health care is organisation based on various levels in which "easy/manageable" cases may be managed at the primary level and "difficult/unmanageable" cases may be sent to the secondary or tertiary level. Length bias sampling means cases with diseases of long duration are more easily included in surveys.[3] For example, in a study including all types of thyroid cancers, papillary cancer having the maximum survival will be enrolled the most and anaplastic cancer, having the least survival time will be enrolled the least.
b. **Neyman Bias (Late Look Bias)** more commonly known as "selective survivor bias" occurs when the exposure of interest is a prognostic determinant. For example if we are studying an association between diabetes mellitus and renal failure, in cases being interviewed one month after occurrence of renal failure and if renal failure patients with diabetes die more frequently, the remaining cases of renal failure would show lower frequency of diabetes mellitus, thus under-evaluating the association between renal failure and diabetes.
c. **Spectrum Bias** occurs when only "clear" or "definite" cases are included not representing the whole spectrum of disease.[5,6] For example, a patient with Systemic Lupus Erythematosus (SLE) might have only facial ushing and may be diagnosed as heat dermatitis and not enrolled in the patients of SLE as it was not considered as a differential.
d. **Survivor-treatment selection Bias** occurs when patients who live longer have more probability to get a particular treatment.[3,6] For example, a retrospective analysis of breast cancer would yield a positive association between treatment and survival.
e. **Berksonian Bias** occurs mainly in hospital based case-control studies when hospitalisation of cases and controls differ.[7,8] For example, in a case control study of pancreatic cancer and coffee drinking, controls were selected from gastroenterology patients in same hospital. However, gastroenterology patients are less likely to drink coffee than the rest of the population because of their disease. Hence the Odds Ratio for coffee drinking was artificially increased due to the under-representation of coffee drinkers among controls.
f. **Image based selection Bias** occurs when inclusion of subjects depends upon their having undergone a certain imaging study. This is a common bias in radiology studies where inclusion of the subject depends upon the availability of imaging studies.[9]
g. **Language Bias** occurs while conducting systematic reviews and meta-analysis where it is common to exclude reports in other languages than English.[6]
h. **Publication Bias** occurs most commonly in systematic reviews and meta-analysis. It occurs due to the in uence of the study results on chances of publication. Studies with positive results are more likely to be published than studies with negative results leading to a preponderance of false positive results in the literature.[3,6,10]
i. **Exclusion Bias** occurs when controls with conditions related to the exposure are excluded where as cases with diseases as co-morbidities are kept in the study. For example, a case-control study was conducted in the surgical patients in a hospital at Helsinki for the association between reserpine and breast cancer. Controls with certain diseases like cholecystectomy, thyroidectomy, cardiovascular surgeries (a common co-morbidity and related to the use of reserpine) were excluded because at that time it was thought that if they are included then use of reserpine would be artificially higher in the controls so even if the use of reserpine is high in breast cancer cases,