# Bias in Medical/Clinical Research

### Abstract

Medical research results many times become inconclusive because of some bias. Bias is a process at any stage of inference tending to produce results that depart systematically from the true values. It is any trend in the design, collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth. Therefore, it is important that all sources of bias are considered at the time of planning a study, and all efforts are made to control them. It is well known that 'bias' can produce dramatic change in study results. In clinical research a great deal of effort is aimed at avoiding bias whenever possible and controlling for and estimating its effects when bias is unavoidable.

Various possible sources of bias are known (listed few books). These lists of plausible biases are generally not exhaustive. New ones are identified (or new names for same type of bias are given often). Some steps are suggested to minimize bias are given at the end. If a study is planned, designed, executed, analyzed, interpreted, etc., properly then occurrences of any type of 'bias' are less likely.

Confounding, one important type of bias and other confusing term 'effect modification' is also discussed in more details.

**Keywords:** Bias; Confounding; Effect modification; Steps to minimize bias; List of common 'biases'

**Sanjeev Sarmukaddam***

*Research Professor, India*

***Corresponding author:** Sanjeev Sarmukaddam, Research Professor, Sangeet-Sadhana, 11th Lane, Paramhans Nagar, Off Paud Road, Pune, India, Email: sanjeev.sarmukaddam@gmail.com

## Introduction

Learning from clinical experience (whether during formal research or in the course of patient care) is impeded by mainly two processes: bias and chance. Medical research results many times become inconclusive because some bias is detected after the results are available. A dictionary definition of 'bias' is 'a one-sided inclination of the mind. In statistics, 'bias' is 'systematic error' that can produce results that depart from the true values. That is, bias is a process at any stage of inference tending to produce results that depart systematically from the true values. It is any trend in the design, collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth. Therefore, it is important that all sources of bias are considered at the time of planning a study, and all efforts are made to control them. Other type of error is 'random error'. Random variation can never be eliminated totally; however, one can reduce the role of chance by proper design, adequate sample size, and appropriate analyses. Chance should always be considered when assessing the results of clinical observations. But it is very important to note that these two sources of error–bias and chance–are not mutually exclusive. In clinical research a great deal of effort is aimed at avoiding bias whenever possible and controlling for and estimating its effects when bias is unavoidable. On the other hand, random error resulting from the play of chance is inherent in all observations. Though it can be minimized, it cannot be avoided altogether.

It is well known that 'bias' can produce dramatic change in study results. Few such dramatic effects of bias are shown in an excellent article by Sackett [1] on 'biases in analytic research'.

There is no need of quoting such dramatic effect showing examples as the role of bias is already known to all. In many situations where an investigator is looking for an association between exposure to risk factor and subsequent disease, it is not possible to randomly allocate exposure to subjects (for example, you cannot insist that some people smoke and others do not or randomly expose some industrial workers to radiation). Thus studies relating exposure to outcome are often observational–where the investigator simply observes what happens and does not intervene. Factors (variables) that are related to both the exposure of a risk factor and the outcome are called confounding factors. Other biases (confounding also is one type of bias but it is controversial to keep it separately or treat as one of the biases) also may occur in observational studies. However, since the essential feature of a clinical trial is the random allocation of treatment to subjects may (note that randomization is a sort of insurance and not a guaranty scheme) overrules this possibility. But they should be looked for even in clinical trials especially 'Phase IV' trials as they are more like observational studies (controversy!!). Though in "RTCs" occurrence of bias (including confounding) is less likely, is not 'impossible'. In "observational" (cross-sectional or longitudinal) it is more likely. Therefore, be careful always.

Various possible sources of bias are listed in few (below quoted) references. Those are not mutually exclusive sources generally and the overlap is substantial. Some of the biases in those lists are collection of many biases of similar type and generally the list is not exhaustive. New ones are identified occasionally (or new names for same type of bias are given often). Some steps suggested to minimize bias (in the results) are given at the end. All steps are not applicable in all the situations. Adopt

the ones that are applicable to your setup. Note that if a study is planned, designed, executed, analyzed, interpreted, etc., properly then occurrences of any type of 'biases' are less likely.

Sackett [1] identified 56 possible biases that may arise in any analytic research of which over two-thirds are related to aspects of study design and execution. Methodologically inferior trials or studies might produce bias in both directions, thereby causing greater variability in estimates of treatment effects. Empirical evidence [2] shows that estimates of treatment effects would be larger in trial in which

1. Adequate measures had not been taken to conceal treatment allocation;

2. Adequate measures had not been taken to generate the allocation schedule;

3. Some allocated participants had been excluded from the analysis; and

4. Measures had not been taken to implement double-blinding. It is noted in one survey that odds ratios were exaggerated by 41% in trials where treatment allocation was inadequately concealed, and by 30% when the process of allocation concealment was not clearly described. It is well known that the process of 'randomization' is important in eliminating selection bias. The importance of 'blinding' is that it avoids observer bias. Non-blinding studies (according to the same survey) over-estimate treatment effects by about 17%. Trials of poor reporting quality are also known to over estimate the effect of treatment. Bias may also lead to fallacious interpretation of study/trial results. In few books [e.g. 3] a complete chapter devoted to 'statistical fallacies' enumerate indirectly the effect of many such biases.

Some of these biases may also lead to one of the important errors namely of missing data which is frequently encountered in clinical studies. Generally this is neglected which may significantly bias the results of the study, apart from reducing study power. They are a serious problem that undermines the scientific trustworthiness of causal conclusions from clinical trials or observational longitudinal studies. There are imputation methods (imputation is the method in which each missing value in a data set is filled in with a value to yield one complete data set). But choice of the appropriate method is important because there are assumptions involved, example, Last Observation Carried Forward (LOCF) method assumes that the response remains constant at the last observed value. This assumption can be biased if the timing and the rate of withdrawal is related to the treatment (e.g. in the case of degenerative diseases, using the last observed value to impute for missing data at a later point in the study means that a higher observation will be carried forward, resulting in an overestimation of the true end-of-study measurement).

Similarly choice of the appropriate analysis method is also very important. Wrong choice may produce lot of bias [3]. Like the methods of survival analysis are required to analyze duration data (though their use is restricted possibly due to lack of awareness and the intricacies involved). Many instances can be quoted of wrong choice of methodology of data analyses, quoting one common example should suffice to highlight the relevance. In several types of studies we may want to examine the consistency of an observed relation across two or more subgroups of the individuals studied. For example, in a clinical trial we might want to know if the observed treatment difference is the same for young and old patients or for both the genders (males, females) or for different stages of disease at presentation. In such cases we are interested in examining whether one effect is modified by the value of another variable. This may be viewed as the examination of the heterogeneity of an observed effect such as treatment benefit across subsets of individuals. The statistical term for heterogeneity of this type is "interaction" [3]. The medical concept of "synergy" is the same thing (opposition in physiological action is "antagonism"). The statistical term interaction relates to the non-independence of the effects of two variables on the outcome of interest. It is advised very strongly (with reasoning) in the literature that to conclude presence of interaction always "compare effect sizes and not the P values" [4]. Comparing 'P' values alone can be misleading. Comparing confidence intervals is less likely to mislead. However, the best approach is to compare directly the effect sizes using "test of interaction" [5]. Still one can often see the practice of comparing 'P' values alone in such situations.

Remember that unlike chance and confounding, which can be evaluated quantitatively, the effects of bias are far more difficult to evaluate and may even be impossible to take into account in the analysis. For this reason, it is of paramount importance to design and conduct each study in such a way that every possibility for introducing bias has been anticipated and that steps have been taken to minimize its occurrence. It must be clearly kept in mind that tests of statistical significance and confidence intervals evaluate only the role of chance as an alternative explanation of an observed association between an exposure and disease [6]. While an examination of the 'P' value and or confidence interval may lead to the conclusion that chance is an unlikely explanation for the findings, this provides absolutely no information concerning the possibility that the observed association is due to the effects of uncontrolled bias or confounding. All three possible alternative explanations (chance, bias, confounding) must always be considered in the interpretation of the results of every study [7]. One more point of vital importance to be kept in mind is that 'clinical significance is different than statistical significance' [8].

It may be noted that most of the 'biases' fall into one of three broad categories:

a) Selection Bias: (occurs when comparisons are made between groups of patients that differ in determinants of outcome other than the one under study).

b) Measurement Bias: (occurs when the methods of measurements are dissimilar among groups of patients).

c) Confounding Bias: (occurs when two factors are associated i.e. travel together and the effect of one is confused with or distorted by the effect of the other).

**Some steps suggested for minimizing bias:** Develop an unbiased scientific temperament by realizing that you are in the occupation of relentless search for truth. Specify the problem to the minutest detail. Assess the validity of the identified target

population, and the groups to be included in the study in the context of objectives and the methodology. Assess the validity of antecedents and outcomes for providing correct answer to your questions. Beware of epistemic uncertainties arising from limitation of knowledge. Evaluate the reliability and validity of the measurements required to assess the antecedents and outcomes, as also of the other tools you plan to deploy. Carry out a pilot study and pretest the tools. Make changes as needed. Identify all possible confounding factors and other sources of bias, and develop an appropriate design that can take care of most of these biases if not all. Choose a representative sample, preferably by random method. Choose an adequate size of sample in each group. Utilize the knowledge about the population while planning a study. Train yourself and coworkers in making correct assessments. Use matching, blinding, masking, and random allocation as needed. Monitor each stage of research, including periodic check of the data. Minimize non-response and partial response. Double check the data and cleanse it of errors in recording, entries, etc.

Analyze the data with proper statistical methods. Use standardized or adjusted rates where needed, do the stratified analysis, or use mathematical models such as regression to take care of biases that could not be ruled out by design. Interpret the results in an objective manner based on evidence. Report only the evidence based results – enthusiastically but dispassionately. Exercise extreme care in drafting the report and keep comments or opinions separate from the results. Bias and other aspects of design can be very adequately taken care of if you could imagine yourself presenting the results a couple of years hence to a critical but friendly audience. Consider what your colleagues could question or advise at that time, consider their reaction when you

conclude that the results are significant and also if you conclude that the results are not significant. Remember that statistical significance and non-significance are equally important.

In short, always remember (as said earlier) that if a study is planned, designed, executed, analyzed, interpreted, etc., properly then occurrences of any type of 'biases' are less likely. Further ask 'can there be non-causal explanations of the results? Are there any confounding factors that have been missed? Whether chance or sampling error could be an explanation?'. Such consideration will help you to develop proper design, and to conduct the study in an upright manner.

Measures of association must be interpreted in terms of the potential for confounding effects of extraneous variables in the design. Confounding is introduced when extraneous variable(s) interfere with the observed association between the exposure (i.e. risk factor) and outcome. A confounder is a variable that is

I. Independently of the exposure, is a risk factor for the disease, and

II. Associated with the exposure, and

III. Is not part of the causal link between the exposure and the disease.

To illustrate this concept, consider the hypothetical data (showing the association between the use of oral contraceptive and myocardial infarction with confounding by age) displayed in following table (similar to table 15.3 in "Foundation of Clinical Research : Applications to practice" 2nd edition by L.G. Portney & M.P.Watkins, 2000, Prentice Hall, New Jersey) Table 1:

**Table 1:** Odds ratios-association between 'passive smoking' and 'cancer' with potential confounding variable 'personal smoking'.

| About Nature of Sample | Exposure Status | Cancer | | | Odds Ratio (Crude/Raw) | Odds Ratio (Mantel-Haenszel Pooled-Adjusted for Self-Smoking) |
| --- | --- | --- | --- | --- | --- | --- |
| | | Cases | Controls | Total | | |
| Total Sample | Passive Smoking Present | 281 | 210 | 491 | 1.63 | 1.63 |
| | Passive Smoking Absent | 228 | 279 | 507 | | |
| | Total | 509 | 489 | 998 | | |
| Smokers | Passive Smoking Present | $120(=a_1)$ | $80(=b_1)$ | 200 | 2.09 | |
| | Passive Smoking Absent | $111(=c_1)$ | $155(=d_1)$ | 266 | | |
| | Total | 231 | 235 | $466(=n_1)$ | | |
| Non-Smokers | Passive Smoking Present | $161(=a_2)$ | $130(=b_2)$ | 291 | 1.31 | |
| | Passive Smoking Absent | $117(=c_2)$ | $124(=d_2)$ | 241 | | |
| | Total | 278 | 254 | $532(=n_2)$ | | |

The odds ratio [between oral contraceptive and myocardial infarction] for the total sample is 2.2. Therefore, women who use oral contraceptive (OC) have greater than twice the risk of myocardial infarction (MI) over those who do not use OC; however the question [often raised] is about the potential confounding effect of age in the analysis. Let us now examine the role of 'age'.

I.   First, we know that 'age' is generally a risk factor associated with MI. In fact, in the above data, among the nonusers of OC, the proportion of cases is greater for older subjects [88 / 183 = 0.48] than for younger subjects [26 / 85 = 0.31]. This suggests that 'age' is a risk factor for MI independent of OC use.

II.  Second, the data also show that some relationship exists between 'age' and use of OC; among the controls, there is a higher proportion of OC users in the young 'age' category [17 / 76 = 0.22] as compared with old 'age' category (7 / 102 = 0.07].

III. Third, 'age' cannot be considered a causal link between OC use and MI.

Therefore 'age' meets all the three [above mentioned] criteria for a confounding factor. We can evaluate the possible confounding by 'age' in the analysis of the data by stratifying the sample into younger and older 'age' groups: 'under 40' and '40 or above'. Odds ratio for these 'age' strata each is 2.8 i.e. each stratum shows the same risk associated with OC use regardless of age. The individual estimates for each group (i.e. strata specific) are considered "unconfounded" for age; however, these unconfounded estimates are different from the overall or crude odds ratio of 2.2. This difference tells us that age does affect the risk estimate; that is, age is a confounding variable. Because OC users tend to be younger, and younger women tend to have fewer MIs, the crude odds ratio was an underestimate of the risk of MI associated with OC use. When 'age' is taken into account (pooled estimate by M-H is described below), the actual risk is higher. If there was no discrepancy between the crude and unconfounded estimates, there would be no confounding. The degree of discrepancy is indicative of the extent to which 'age' confounded the original data [one possible measure could be 'percentage underestimation' which for these data is {{(2.8-2.2)/2.2}×100} = 27.27%].

To evaluate the effect of confounding in an analysis, the researcher must collect information on the potentially confounding variable(s). If the investigator in this example had not collected data on subjects' ages, the preceding analysis would not have been possible. Although 'age' is continuous variable, for such analysis we have to make few strata (here only two are made but one can make more 'clinical meaningful' strata, remember that an estimate will change for different cut-off). The researcher must be able to predict what variables are possible confounders. It is possible that several confounding factors will be operating in one study. In addition to controlling for confounding in the analysis, researcher can use design strategies, such as matching or homogeneous subjects, to control for these effects. For instance, if we were to restrict the subjects to women under 40, age could not be a confounding factor.

When data are stratified, and separate risk estimates (here odds ratio, but in prospective studies-relative risks) are calculated for each stratum, it is possible to report each estimate; however, it is usually more useful to calculate a single overall estimate that reflects the association between risk factor (exposure) and disease with the confounding factor taken into account. Most commonly used procedure to accomplish this is by Mantel-Haenszel [set of Mantel-Haenszel pooled risk estimates provide a weighed summary value that can be used to report risk associated with a specific exposure (risk factor) adjusted for the confounding variable].

Pooled Odds Ratio by Mantel-Haenszel=$OR_{M-H}$=[{Sum{($a_i d_i$)/N} / Sum{($b_i c_i$)/N}]

($OR_{M-H}$={[($a_1 d_1$) / ($n_1$)] + [($a_2 d_2$) / ($n_2$)]} / {[($b_1 c_1$) / ($n_1$)] + [($b_2 c_2$) / ($n_2$)]}]=2.8 for above data)

When the Mantel-Haenszel estimate differs from the crude estimate, it is the Mantel-Haenszel estimate that should be reported. It is most appropriately used when the stratum specific estimates are uniform that is when there is no 'effect modification'.

Confounding variables can be thought of as 'nuisance' variables, they may or may not be present depending on the source population and how subjects are chosen. However, sometimes a third factor will interact with the exposure (risk factor) and disease variables in such a way as to present a constant effect and such a variable is called an 'effect modifier' which is generally a natural phenomenon that exists independent of the study design & will always be a factor in interpretation of risk. Effect modifiers tend to be biologically related to the variables being studied. For example, suppose we wanted to look at the association between exposure to asbestos and development of lung cancer. We prospectively follow a group of asbestos workers and a group of workers in a different industry for, say, 15 years. Assume we found that asbestos is a risk factor for lung cancer, with a relative risk of 4.5. We also have collected data on the subjects' smoking habits as well because we know that smoking is also a risk for lung cancer. We can stratify our subjects according to smokers & non-smokers and look at the relative risk associated with asbestos exposure for each group. Suppose we find that the risk associated with asbestos for smokers is 5.0, whereas the risk for non-smokers is 1.3. This would tell us that smoking is an effect modifier-that the effect of asbestos is exacerbated for a smoker in terms of risk for lung cancer. The fact that the stratum specific risk estimates are different for smokers and non-smokers indicates that smoking interacts with asbestos as an effect modifier. Note that this is not the case with confounding variable, as illustrated in the above example of MI and OC use.

Assumption made in the estimation of a common odds ratio (by above M-H method) is that the strength of association is the same in each stratum. If the underlying odds ratio is different in the various strata, then it makes little sense to estimate a common odds ratio. Suppose we are interested in studying the association between a disease variable 'D' and an exposure variable (risk factor) 'E', but are concerned about the possible confounding effect of another variable 'C'. Then we stratify the study population into 'k' strata according to the variable 'C' and compute the odds ratio

relating disease to exposure in each stratum. If the underlying (true) odds ratio is different across the 'k' strata, then there is said to be 'interaction' or 'effect modification' between 'E' and 'C' and the variable 'C' is referred to as an "effect modifier". In other words, if 'C' is an effect modifier, the relationship between disease and exposure (risk factor) differs for different levels of 'C'.

Important question that "how can we detect if any variable (say 'C') is an effect modifier?" is addressed by 'Woolf' test. Suppose we have a dichotomous disease variable 'D' and exposure 'E'. We stratify our study population into 'k' strata according to a confounding variable 'C'. Let $OR_i$ = underlying odds ratio in the $i^{th}$ stratum. To test the hypothesis

$H_0$: $OR_1$ =......= $OR_i$ =......= $OR_k$ versus $H_1$: at least two of the $OR_i$ are different with a significance level α, use the following procedure:

Test statistic $X^2hom = Sum\{w_i(ln\ OR_i - ln\ OR)^2\}$ which follows Chi-square distribution under $H_0$, $ln\ OR_i$ is the log odds ratio;

$w_i = \{(1/a_i) + (1/b_i) + (1/c_i) + (1/d_i)\}^{-1}$; i goes from 1 to k

$ln\ OR = \{Sum\ wi\ (ln\ OR_i) / Sum\ w_i\}$; i goes from 1 to k (i.e. Sum over all the categories).

Example: In one study on effect of passive smoking on cancer risk, potential confounding variable being 'smoking' by the test subjects themselves (i.e. personal smoking), because personal smoking is related to both cancer risk and spouse smoking (=passive smoking) Table 2.

For this data set $ln\ OR_1 = 0.739$, $w_1 = 27.55$, $ln\ OR_2 = 0.272$, $w_2 = 32.77$ and thus $\chi^2hom = 3.27$; df=1; P >0.05. Thus there is no significant effect modification.

**Table 2:** Odds ratios-association between 'use of oral contraceptive' and 'myocardial infarction' with confounding by age.

| About Nature of Sample | Exposure Status | Myocardial Infarction | | | Odds Ratio (Crude/Raw) | Odds Ratio (Mantel-Haenszel Pooled Estimate-Adjusted for Age Categories) |
|---|---|---|---|---|---|---|
| | | Cases | Controls | Total | | |
| Total Sample | OC User | 39 | 24 | 63 | 2.2 | 2.8 |
| | Non User | 114 | 154 | 268 | | |
| | Total | 153 | 178 | 331 | | |
| < 40 Years Old | OC User | $21(=a_1)$ | $17(=b_1)$ | 38 | 2.8 | |
| | Non User | $26(=c_1)$ | $59(=d_1)$ | 85 | | |
| | Total | 47 | 76 | $123(=n_1)$ | | |
| ≥ 40 Years Old | OC User | $18(=a_2)$ | $7(=b_2)$ | 25 | 2.8 | |
| | Non User | $88(=c_2)$ | $95(=d_2)$ | 183 | | |
| | Total | 106 | 102 | $208(=n_2)$ | | |

There is a test (called 'Chi-square M-H' as it is again given by Mantel & Haenszel) to assess the significance of pooled (i. e. adjusted for confounder) odds ratio and a method to estimate its [pooled (i.e. adjusted for confounder) odds ratio] but are not discussed (or given) here to avoid lot of mathematics. However, interested readers can refer to excellent book by Rosner ("Fundamentals of Biostatistics" 5th edition, Bernard Rosner, Duxbury Thomson Learning, CA, 2000) and note its availability in software 'CIA' ["Confidence Interval Analysis" software accompany excellent book: Altman DL, Machin D, Bryant TN, and Gardner MJ. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines' 2nd* edition, BMJ Books, London, 2003].

## Acknowledgement

None.

## Conflict of Interest

None.

## References

1. Sackett DL (1979) Bias in analytical research. J Chronic Dis 32(1-2): 51-63.

2. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. JAMA 273(5): 408-412.

3. Sarmukaddam SB (2006) Fundamentals of Biostatistics. Jaypee Brothers Medical Publishers Ltd. New Delhi, India.

4. Matthews JN, Altman DG (1996) 'Interaction 2 : compare effect sizes not P values'. BMJ 313(7060): 808.

5. Altman DG, Bland JM (2003) 'Interaction revisited: the difference between two estimates'. BMJ 326(7382): 219.

6. Indrayan A, Sarmukaddam SB (2001) Medical Biostatistics. Marcel Dekker, Inc., New York, USA.

7. Sarmukaddam SB (2010) Biostatistics Simplified. Jaypee Brothers Medical Publishers Ltd. New Delhi, India.

8. Sarmukaddam SB (2012) Interpreting "statistical hypothesis testing" results in clinical research. J Ayurveda Integr Med 3(2): 65-69.