



Perspective

Statistical fallacies & errors can also jeopardize life & health of many

Dealing with vitalities such as disease and death, medical research is a delicate endeavour, and required to be carried out with utmost responsibility. Practitioners believe the research results, particularly when published in a reputed journal, and use them in future cases. While much of this research has helped in improving the health across the world, a few researchers realize that imprecise research can also imperil life and health of a large number of people. Amidst several factors, this can happen when the results are based on misuse and abuse of statistical methods to arrive at conclusions.

Research results, when wrong by just one per cent and adopted for practice on millions of patients, can threaten health and life of thousands of people. Substandard research can jeopardize many lives. If a trial finds higher efficacy of a new regimen, practitioners will obviously adopt the new and discard the old. However, if the trial is later found to have faulty design, faulty data or faulty analysis, life and health of many may have already been compromised. Type I and Type II errors are considered genuine statistical errors, but these also can have far-reaching implications on health of the people.

Statistical methods provide an appropriate tool for measuring uncertainties, and in some cases, to control them¹. However, that does not help make it an exact science. The 95 per cent statistical confidence intervals that exclude five per cent unlikely values epitomizes the vagaries of statistical science. It may surprise some of us that statistical methods are able to find a Gaussian pattern in random variations too, and it is routinely exploited to draw conclusions. However, the probabilities remain a sheet anchor of these methods, and the conclusions remain inexact. A few realize that probabilities work in the long run, just as insurance do, but can miserably fail in individual cases. That is where the vagaries lie. The problem is compounded by

intentional and unintentional fallacies that creep into medical data and their interpretation.

Statistical fallacies

A common misinterpretation is considering mere association or correlation as evidence of cause-effect². The incidence of cardiovascular diseases in India is negatively correlated with birth rate, but it has no causal implication. Counterfactuals are important ingredient in empirical reasoning but many times ignored. Sometimes, the distinction between necessary and sufficient is lost in arguments and that can result in inappropriate conclusions.

An interesting quote is 'Head in an oven and feet in a freezer, and the person is comfortable, on average'. Nonsense of such assertions is apparent, but that is what seems to be passed on to the reader in some medical researches³. Summary measures such as mean and proportion, when based on the aggregated data, can be deceptive. Consider case fatality in cancer patients in a general hospital and a specialized cancer hospital, both with the same case fatality in aggregate⁴. Cancer hospital would receive patients predominantly in advanced stages, and in them, its performance may be markedly better than in a general hospital. If only the aggregate percentage is reported, this distinction is lost. Standardization is advocated to avoid such discrepancies but not done in many reports.

Many medical researchers look at the difference or gain in medical parameters after the treatment compared with values before the treatment. In their keenness, they sometimes forget that a gain of 3 g/dl in Hb level over pre-treatment value 8 g/dl has a different meaning than the same gain over the pre-treatment value of 11 g/dl⁵. It is relatively easy to affect a rise over lower Hb values than over higher values - thus, stratification or covariance analysis is required.

It is generally believed that statistics is the science that crunches numbers. As Gregg Easterbrook said,

torture the data and they will confess to anything⁶. With online access, data availability has multiplied manifold. Google chief economist rightly predicts that data scientist's job is becoming sexiest of all⁷ as availability of enormous data requires skill to extract relevant messages. Expansion in this skill has not kept pace with the rapid rise in the availability of data. Semi-skilled professionals draw conclusions, some of which are of dubious quality. While calculations can be done by a computer, interpretation of statistical results requires skill.

Among other fallacies is cherry-picking the statistical indices for communicating a result. Research results can be provided in terms of actual blood sugar levels or as prevalence of diabetes, in terms of Hb level or anaemia. All quantitative measurements can be converted to qualities. The summary measure for quantities generally is mean and for qualities is proportion. A researcher can try out both and report the one that suits a particular hypothesis. The protocol should specify the indicators with justification and should be adhered to in the analysis and communication.

Statistical errors

We know that medical errors in terms of misdiagnosis, missed diagnosis, negligent care, treatment errors, prognostic misjudgement, *etc.* can cause misery and death. Statistical errors of Type I and Type II are not just known and acknowledged but also accepted. Not many realize that genuine Type I error means an ineffective regimen is proclaimed effective and many deaths can occur due to this error. Similarly, Type II error means an effective treatment is denied to the patient, and this also can cause deaths. Misdiagnosis and missed diagnosis can legitimately occur due to statistical errors, but these can be minimized by choosing a right design and an adequate sample size.

Scientists debate about the validity of the conventional cut-off 0.05 for P values⁸, which is so commonly used by statisticians and medical researchers alike. This is the chance that a random sample of subjects happens to provide wrong evidence of efficacy of a regimen when none is actually present. Use of confidence intervals is advocated instead of P values, but there is no escape in some situations. If the objective is to establish that one treatment is better than the other by at least three per cent, statistical significance is the only way to rule out sampling fluctuations. The values between 0.04 and 0.06 should be interpreted with caution. The same

kind of precaution that is always taken for patients with borderline values. In any case, P values should not be taken too seriously. In most practical research setup, these arise not just from random fluctuations but also incorporate chance of errors due to faulty design and faulty data. In addition, the P values must be complemented by biological plausibility and, of course, common sense.

Many examples can be cited from medical literature when more than one statistical test is done on the same data, each at level 0.05⁹, without realizing that this inflates the error rate. Thus, false significance is achieved that does not replicate in actual practice. This can affect health of many when the results are unsuspectingly used on a large number of patients. Statistical procedures such as Bonferroni and Tukey are used to control the probability of Type I error to the specified level but sometimes ignored.

Many statisticians and medical professionals alike fail to make a distinction between statistical significance and medical significance. Statistical inference heavily depends on size of the sample. Surprise for many is that a difference of one per cent in efficacy of two regimens can be statistically significant in large-scale studies. This may not have any medical significance. Reverse can also happen. Freiman *et al*¹⁰ studied 71 negative trials and observed that the sample size was too small to detect 25 per cent improvement in the outcome. Had the sample size been adequate, adverse outcome for many could have been avoided.

Conclusion

Statistical fallacies and errors are summarized in the Table. More serious are errors due to negligence. Uncounted deaths occur due to wrong conclusion arrived by inappropriate analysis and inaccurate data. Many times, these go unnoticed. Although this can happen with fully trained statistical professionals, there are a large number of researchers who have little or no expertise and training in statistical analysis. With the ready availability of statistical software, anybody can think of himself/ herself as statistical expert and do the analysis. Whereas medical professionals are trained for endless years in the business of saving lives and reducing suffering, neither statisticians nor medical researchers are trained for data analysis with the same rigorousness. Medical biostatistics too is in the business of saving lives and reduces suffering, but only a few realize this to be so.

Table. Statistical fallacies and errors and their solution

Statistical fallacy/error	Solution
Using probability for a single case without sufficient caution	Use probability for (large) group and use sufficient precaution while using it for a single case
Using correlation as an indication of cause-effect	Use counter-factual and distinguish between necessary and sufficient factors
Using aggregate rates for comparison of groups	If the composition of groups varies, use standardized rate
Using gain by an intervention for inference disregarding baseline values	Use stratification or covariance analysis
Data analysis and interpretation by inadequately skilled professionals	Statistical analysis and interpretation requires rigorous training and experience
Cherry-picking the statistical indices for communicating a result	The protocol should specify the indicators with justification and should be adhered to in the analysis and communication
Type I and Type II errors	Choose a right design and an adequate sample size
Strictly adhering to the conventional cut-off of $P=0.05$ for statistical inference	Interpret P between 0.04 and 0.06 with caution and conclude that further work is needed
Forgetting that the P values incorporate chance of errors due to faulty design and faulty data	Ensure that the design is adequate, and the data are correct
Using many statistical tests, each at level 0.05	Use Bonferroni correction or Tukey method to control the overall chance of error
Failure to make a distinction between statistical significance and medical significance	Assess medical significance separately from statistical significance

Financial support & sponsorship: None.

Conflicts of Interest: None.

Abhaya Indrayan

Biostatistics Consultant, Max Healthcare, Saket, New Delhi 110 017 & Former Professor, Department of Biostatistics & Medical Informatics, University College of Medical Sciences, Delhi 110 095, India
a.indrayan@gmail.com

Received May 5, 2018

References

- Indrayan A, Malhotra RK. *Medical biostatistics*, 4th ed. Chapman and Hall: CRC Press; 2018.
- Czarnecka A, Milewski K, Zielińska M. Asymmetric dimethylarginine and hepatic encephalopathy: Cause, effect or association? *Neurochem Res* 2017; 42 : 750-61.
- Ludwig EG, Collette JC. Some misuses of health statistics. *JAMA* 1971; 216 : 493-9.
- Silber JH, Rosenbaum PR. A spurious correlation between hospital mortality and complication rates: The importance of severity adjustment. *Med Care* 1997; 35 : OS77-92.
- Hashimoto M, Fujii T, Hamaguchi M, Furu M, Ito H, Terao C, et al. Increase of hemoglobin levels by anti-IL-6 receptor antibody (tocilizumab) in rheumatoid arthritis. *PLoS One* 2014; 9 : e98202.
- Easterbrook G. Science. Available from: https://www.todayinsci.com/E/Easterbrook_Gregg/EasterbrookGregg-Quotations.htm, accessed on May 1, 2018.
- Why “Data Scientist” is being called the Sexiest Job of the 21st Century; 7 May, 2015. Available from: <https://www.import.io/post/why-data-scientist-is-being-called-the-sexiest-job-of-the-21st-century/>, accessed on May 1, 2018.
- Wasserstein RL, Lazar NA. The ASA’s statement on P-values: Context, process, and purpose. *Am Stat* 2016; 70 : 129-33.
- Naqash A, Ara R, Bader GN. Effectiveness and safety of ferric carboxymaltose compared to iron sucrose in women with iron deficiency anemia: Phase IV clinical trials. *BMC Womens Health* 2018; 18 : 6.
- Freiman JA, Chalmers TC, Smith H Jr., Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *N Engl J Med* 1978; 299 : 690-4.