

Review Article

Suggested Statistical Reporting Guidelines for Clinical Trials Data

Jaykaran Charan, Deepak Saxena¹

ABSTRACT

Statistical methods are widely accepted tools for analyzing data and drawing appropriate conclusions. However, inappropriate statistical methods may lead to unsuitable or false conclusions. Statistical quality of articles published in Indian medical journals presents with statistical fallacies and very few attempts have been done to document it till date. Issues related to inappropriate use of statistic in interpreting clinical trial raises serious doubt about the validity of the results published and there generalization to patient population. Clinical trials are considered to have one of the highest evidences in terms of efficacy of any intervention and hence results reported in these clinical trials help in appropriate decision making in clinical practices. Present manuscript is an attempt to have an insight into issues related to inappropriate use of statistic during clinical trials and suggests possible methods by which such fallacies can be averted.

Key words: *Statistics, clinical trials, multiple end points, type 1 error*

INTRODUCTION

Statistics are widely accepted tools for analyzing data and drawing appropriate conclusions but use of inappropriate statistical methods leads to an inappropriate or false conclusion. It has been observed that qualities of statistics used in articles published in various Indian medical journals are inappropriate.^[1,2] A serious doubt is expressed on the validity of the results published and there generalization to patient population because in absence of use of appropriate statistical methods. Clinical trials are considered to have one of the highest evidences in terms of efficacy of any intervention and hence results reported in these clinical trials helps for

decision making in clinical practice. However, Clinical trials published in various medical journals in India also presents with inappropriate use of statistical tests and drawing false conclusions. There is a definite lack of such studies narrating inappropriate use of statistic in clinical trials published in Indian medical journals.^[3,4] The present review is an attempt to highlight issues related to inappropriate use of statistics in clinical trials published in various Indian medical journals. We describe various types of inappropriate statistical methods used by authors under the following subheadings:

SUGGESTED STATISTICAL REPORTING GUIDELINES FOR CLINICAL STUDY RESULTS

Study design in clinical trial is an important phase as type of study not only affects the smooth conduction of trial but also has an impact on validity and reliability of clinical trial conducted. A qualified statistician or an appropriate knowledge and orientation about basic statistics is required to avoid few statistical flaws.

Access this article online	
Website: www.ijpm.info	Quick Response Code 
DOI: 10.4103/0253-7176.96152	

Department of Pharmacology, Government Medical College, Surat, 'Department of Epidemiology, Indian Institute of Public Health, Ahmedabad, Gujarat, India

Address for correspondence: Dr. Jaykaran Charan
Department of Pharmacology, Government Medical College, Surat, Gujarat, India. E-mail: drjaykaran@yahoo.co.in

During this phase of clinical trial, aim of the study, primary and secondary endpoints should be formulated clearly. However, it is observed that division between primary and secondary endpoints are not reported in many of the clinical trials published in Indian medical journals.^[4] Division between primary and secondary endpoint is very important to prevent type I error (false positivity) which arises because of multiple endpoints.^[5,6] Equally important is that aims and objectives of the trial should be clearly framed in an unambiguous language.

Reporting of methods for calculation of sample size and non reporting of sample size estimation is an important issues.^[4] Many of the clinical trials published in Indian and western medical journals are deficient in reporting of sample size calculation.^[4,7,8,1,2] Help of statistician or readymade computer software can be taken for the calculation of appropriate sample size. Power calculation should also be done during designing phase. Power of study should be adjusted so that any worthwhile effect if present can be detected. Small sample sizes are associated with Type-II errors and hence raise serious questions on validity of study. It is observed that many negative clinical trials published in Indian medical journals on analysis found to be underpowered because of less sample size.^[8] During calculation of the desired sample size, appropriate wattage for components like Type-I error, Type-II error, effect size. Such parameters should also be considered and mentioned in detail so that it can be a guiding tool for readers to emulate and reader can calculate the sample size on the basis of information provided in the article. Sample size calculation should also include probable loss to follow up rate. It is equally important to understand and perceive that formula of sample size calculation is different for different type of study designs.

During designing the trial, methods of analysis of data should be selected, narrated and adhered. It is always mandatory to fulfill certain assumptions to apply statistical tests. All care should be taken to fulfill these assumptions during collection of data itself. In clinical trials one such important assumptions is randomization. Proper method of randomization should be selected based on study objectives. Statistical tests lose its validity if applied in a non randomized sample. Adequate information related to randomization procedure and sampling method should be included in the manuscript.^[9]

Baseline statistical comparison between groups (treatment and control) is commonly used in randomized controlled trial published in different medical journals.^[4] Inappropriate comparison of baseline parameters including demographical factors just to avoid partially or inherently heterogeneous group is

very frequently done by researchers which in truth can not be compared.^[9] If the group is heterogeneous, then difference in the treatment outcome can not be considered because of intervention unless confounding factors are adjusted using sophisticated multivariate analysis.^[10] So knowledge about baseline parameters should be clear so as to see the imbalance which may affect the result but there is no need to compare the baseline parameters by statistical tests and there is no need to report *P* values.^[10] In a true randomized trial each patient has equal probability of being assigned to either the treatment group or control group therefore any difference obtained at the baseline can be because of chance variation.^[10] It should also be perceived that unlike other research methods in randomized trials difference in the baseline parameters does not indicate bias.^[10] Equally true is that just by showing that there are no statistically significant differences between groups at baseline, it can not be concluded that groups are equivalent, particularly in studies with small sample size which lack statistical power.

USE OF INAPPROPRIATE STATISTIC DURING DATA ANALYSIS

One of the fallacies during clinical trials is use of inappropriate or wrong statistical tests.^[1-5] Number of reasons cited for considering a test to be inappropriate. As mentioned earlier before using the particular statistical tests, all the assumptions for that statistical test should be fulfilled. In many of the clinical trials published in Indian medical journals it is observed that distribution of data was not checked which is an important prerequisite before applying parametric statistical tests.^[1,2] Biological variables are usually more prone to fall in skewed distribution, hence checking of distribution become very important. Variable measured from small samples are also more prone to fall under skewed distribution.^[1] Nonparametric statistical tests should be used for data having skewed distribution. It is also observed that parametric tests were used for data related to rank and scores (ordinal data), which is also wrong, in this type of data nonparametric tests should be used. Another observation is use of unpaired tests for paired data and vice versa. So goal of analysis should be very clear and paired test should be used for paired data and vice versa. Even most simple tests like Chi-square and *t* test are found to be used inappropriately.

In most of the clinical trials published in Indian medical journals multiple end points are measured and hence multiple statistical tests are used to measure the difference between groups. This may lead to increase in false positivity or Type I error. The *P* value is based on principal of probability, if with one statistical test

the chance of having a significant result is 5%, then after 20 statistical tests it may be more than 40%. This is called inflation of type I error. More the number of statistical tests more inflation of Type I error.^[6,11] Inflation of type I error can be prevented dividing the endpoints at the design phase itself into primary and secondary endpoints. Most important endpoint should be considered as primary endpoint and other endpoints should be considered as secondary endpoints. If the primary endpoint cannot be restricted to one then multiple endpoints should be adjusted with the help of various adjustment methods like Bonferroni method, least significant difference test, composite endpoint method etc. In a study done by Tom J *et al.* it was found that out of 16 positive randomized controlled trials published in British Medical Journal in 2004 only 8 clinical trials remain significant after the adjustment of multiple end points.^[12] In a similar study done for clinical trials published in Indian medical journals it was found that about one third of the clinical trials published in four Indian journals were false positive and statistical methods used to adjust this error was not mentioned in even a single trial.^[11]

It is difficult to decide a single best method out of all method for adjustment of multiple endpoints. As each method has its unique advantage and disadvantage. According to International Conference on Harmonization E 9 guideline if in a clinical trial adjustment of multiple endpoints is not reported then the reason for this should be mentioned clearly in the manuscript.^[13] Consolidated Standards of Reporting Trials (CONSORT) statement also favors adjustment of multiple endpoints.^[14]

Post hoc subgroup analysis too can only be justified if decision regarding analysis is taken in design phase of the study itself. Subgroup analysis should not be done just to find significant difference in some specific group. That will lead to “fishing” of significant results and will cause inflation of type I error.^[15]

It is also a common practice during analysis of data continuous data to divide it into ordinal categories. This leads to the reduction of precision of measurement and variability. Authors should mention the reason for this transformation and how boundaries of ordinal categories were decided. Sometime cut off boundaries are chosen in so as to favor some results.^[16]

During analysis of data all subjects who were randomized should be included in the analysis this is called “Intent to treat” (ITT) principle. ITT is done to avoid the effect of crossover or dropouts of study subjects which lead to braking of randomization to the treatment groups in a trial. Aim of ITT is to analyze the

study subjects in the same in group in which they were randomized at the start of study and to prevent bias caused by unequal dropouts from both the groups. Not accounting missing data may lead to bias in the favor of intervention.^[17] Reasons for withdrawal should be documented according to the group in which subjects were randomized. ITT principle is very important for pragmatic clinical trials where subjects are observed in real life situation and results are used to make policy decisions.^[17] Findings related lost to follow up subjects should not be discarded as such and various methods for adjustment of missing data should be used to include these subjects in analysis.^[18] A flow chart should be given with manuscript so that patient randomization, recruitment, lost of follow up, withdrawal, no. of subjects included in final analysis can be seen visually. CONSORT statement instructed all author to publish such flow chart with manuscript.^[14] All statistical methods should be decided in advance in design phase of study and data dredging should be avoided just to get significant *P* values.

MISREPORTING RESULTS OF A STATISTICAL ANALYSIS

Use of all the statistical method in a trial needs to be documented with relevant description to facilitate the readers to validate and recalculate the finding as narrated by the authors. Common statistical methods can be described in brief but some less common or obscure tests should be explained in detail. Reference should be given for less common or obscure statistical tests.^[19] If more than one statistical tests are used it should be clearly mentioned and specified with reference to the which endpoint and data.^[20] Many clinical trials that are being published are found with nonspecific statements for describing statistical tests for example “Chi-square test is used for categorical data or student *t* test is used for quantitative data” or “appropriate statistical tests were used for analysis of data”. Such inappropriate statements create lots of confusion in the mind of readers, particularly those from non-statistical backgrounds. Hence specific information regarding conditions under which a test is used should be mentioned. Authors should also take due precautions in narrating the version of test like paired or unpaired.

USE OF INAPPROPRIATE STATISTICS WHILE PRESENTATION

Good presentation is important part in enhancing the quality of research.^[21] Appropriate methods of central tendency or dispersion should be used in presenting the findings. Ratio and interval data which

follows the normal distribution should be described as Mean (SD). Ratio and interval data not following the normal distribution should be described as Median or Range, as range is affected much by outliers so median should be considered as better method. Ordinal data should be described as median and nominal data as frequencies and percentage. As mentioned earlier in clinical trials usually the endpoints are biological variables and biological variables usually follows non normal or skewed distribution so median should be more frequently found in published literature. It is observed not only in western journals but also in Indian journals that “Mean±SEM” is used to describe the ratio and interval data which follows the normal distribution.^[22,23] Standard error of mean (SEM) is not a descriptive statistics, SEM indicates probability of falling of population mean around the range of sample mean. It does not show variability within the sample. Variability within the sample is shown by standard deviation (SD). Value of SEM is always less than SD so if group is described as Mean±SEM it can be falsely concluded that variability within the sample is less. So SD should be used in the place of SEM. Even for prediction for population mean better method is confidence interval not SEM. Instead of writing “Mean±SD” better method of presentation is Mean (SD) as with this presentation confusion with confidence interval can be avoided.^[24]

Confidence interval should be given for the main outcome or important outcome measured in the study. This confidence interval should be the confidence interval of difference between two means not of individual group means. 95% Confidence interval shows the probability of population mean value around a range in sample mean with 95% probability. Probability in terms of *P* values has several shortcomings and can not be relied completely for decision making. *P* value may be significant but the difference between two groups for endpoints may not be as large to have some clinical significance. Small differences between large groups can be statistically significant but clinically meaningless and large differences between small groups can be clinically important but not statistically significant. So confidence interval should be mentioned with the *P* value or instead of *P* value alone.^[25] In many of the clinical trials published in Indian medical journals confidence interval is not narrated.^[2,4,8]

Authors should also try to avoid using terms like “*P* values as “<0.05” or “=NS”, but exact *P* values should be mentioned. As per review of “Instruction to authors” section of most of the Indian medical journals, authors are instructed to write exact *P* value but still finding exact *P* value is not common in clinical trials. Equally important is to remember that the reporting

should not be done with unnecessary precision for both test statistic and *P* value.^[26]

MISINTERPRETING STATISTICAL ANALYSIS RESULTS

Interpretation of the data obtained is very important for arriving to a valid conclusion. If there is no significant difference between two study groups it cannot be concluded that there is “No effect” or “No difference”. Non significant results may be because of any of the two reasons: Either there is no actual effect of intervention or the study is underpowered to show that any worthwhile effect exists. One of the important reasons for less power of study is inadequate sample size. So during interpretation of results of clinical trial, sample size should also be considered. This issue becomes more important in the case of negative clinical trials as the non-significance may be a result of less power. So sample size should be calculated in design phase and in the case of negative clinical trials post hoc power calculation should be done.^[8] If multiple tests are used in the trials sufficient information regarding type I error should be mentioned in the manuscript. If some confounding factors are present than whether they are adjusted or not, should also be mentioned in the manuscript.

Below mentioned are few tips that are to be considered while reading, analyzing, interpreting or planning a Clinical trial:

RECOMMENDATIONS

- Endpoints in clinical trials should be divided into primary and secondary endpoints
- Sample size should be calculated before starting of clinical trial and all components of sample size calculation should be reported in manuscript
- Before applying any statistical test all the assumptions for that statistical test should be fulfilled and information regarding the same should be mentioned in the manuscript
- Baseline demographic data of study groups should be compared to see if any heterogeneity is there but statistical tests should not be used. Any heterogeneity should be corrected by adjusting the results through multivariate analysis
- Appropriate statistical tests should be selected on the basis of aim of the study, type of data and distribution of data
- Multiple endpoints should be adjusted by methods like Bonferroni, composite endpoint method etc.
- Post hoc subgroup analysis should be avoided
- Missing data analysis should be done. In pragmatic

clinical trials “Intent to treat” principle should be utilized

- CONSORT statement should be followed during reporting of clinical trials
- Reference should be given for the less common or obscure statistical tests
- Statistical tests should be mentioned in enough detail so they can be replicated by reader
- “Mean (SD)” should be used in the place of “Mean±SEM”
- Confidence interval should be given for difference between the means for primary or important endpoints and should be given with or instead of *P* value
- Exact *P* values should be mentioned in the manuscript.

REFERENCES

1. Jaykaran, Yadav P, Bhardwaj P, Goyal J. Problems in reporting of statistics: Comparison between journal related to basic science with journal related to clinical practice. *Internet J Epidemiol* 2009;7. Available from: (http://www.ispub.com/journal/the_internet_journal_of_epidemiology/volume_7_number_1_24/article/problems-in-reporting-of-statistics-comparison-between-journal-related-to-basic-science-with-journal-related-to-clinical-practice.html, [Last accessed on 2010 Oct 7]).
2. Karan J, Goyal JP, Bhardwaj P, Yadav P. Statistical reporting in Indian Pediatrics. *Indian Pediatr* 2009;46:811-2.
3. Jaykaran, Yadav P, Chavda N, Kantharia ND. Some issues related to the reporting of statistics in clinical trials published in Indian medical journals: A survey. *Int J Pharmacol* 2010;6:354-9.
4. Karan J, Kantharia ND, Yadav P, Bhardwaj P. Reporting statistics in clinical trials published in Indian journals: A survey. *Pak J Med Sci* 2010;26:212-6.
5. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials – A survey of three medical journals. *N Engl J Med* 1987;317:426-32.
6. Jaykaran, Saurabh M, Yadav P, Chavda N, Mody P. False positive clinical trials published in four Indian medical journals: A survey. *J Pharm Res* 2010;3:822-3.
7. Olsen CH. Review of the use of statistics in Infection and Immunity. *Infect Immun* 2003;71:6689-92.
8. Jaykaran, Saurabh M, Gohiya S, Gohiya V, Sharma G, Chavda N. Reporting of Power, Sample Size and Confidence interval in negative clinical trials published in four Indian medical journals. *J Pharm Res* 2010;3:298-300.
9. White SJ. Statistical errors in papers in the British Journal of Psychiatry. *Br J Psychiatry* 1979;135:336-42.
10. Tom L. Twenty statistical error even YOU can find in biomedical research articles. *Croat Med J* 2004;45:361-70.
11. Dar R, Serlin R, Omer H. Misuse of Statistical Tests in three Decades of Psychotherapy Research. *J Consult Clin Psychol* 1994;62:75-82.
12. Ton J, Aeilko H. Clinical trials are often false positive: A review of simple methods to control this problem. *Curr Clin Pharmacol* 2006;1:1-4.
13. ICH E9 Expert Working Group. ICH harmonised tripartite guideline: Statistical principles for clinical trials. *Stat Med* 1999;18:1905-42.
14. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
15. Moreira ED, Stein Z, Susser E. Reporting on methods of subgroup analysis in clinical trials: A survey of four scientific journals. *Brazilian J Med Biol Res* 2001;34:1441-6.
16. Lang T, Secic M. How to report statistics in medicine: Annotated guidelines for authors, editors, and reviewers. Philadelphia (PA): American College of Physicians; 1997.
17. Saurabh M, Jaykaran, Yadav P, Chavda N, Lokesh RV, Deodhare S, et al. Reporting of missing data in clinical trials published in four Indian journals- a survey. *J Pharm Res* 2010;3:277-9.
18. Mu MC, Albert PS, Wu BU. Adjusting for drop-out in clinical trials with repeated measures: Design and analysis issues. *Stat Med* 2001;20:93-108.
19. Altman DG. Statistics in medical journals: Some recent trends. *Stat Med* 2000;19:3275-89.
20. Welch II GE, Gabbe SG. Review of statistics usage in the American Journal of Obstetrics and Gynecology. *Am J Obstet Gynecol* 1996;175:1138-41.
21. Evans M. Presentation of manuscripts for publication in the British Journal of Surgery. *Br J Surg* 1989;76:1311-4.
22. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br J Anaesth* 2001;90:514-6.
23. Saurabh M, Jaykaran, Chavda N, Yadav P, Kantharia ND, Lokesh RV. Misuse of standard error of mean (SEM) when reporting variability of a sample: A critical appraisal of four Indian journals. *J Pharm Res* 2010;3:799-800.
24. Jaykaran. “Mean±SEM” or “Mean (SD)”?. *Indian J Pharmacol* 2010;42:329.
25. Gardner MJ, Altman D. Confidence intervals rather than *P* values: Estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.
26. Tom L. Twenty statistical error even you can find in biomedical research articles. *Croat Med J* 2004;45:361-70.

How to cite this article: Charan J, Saxena D. Suggested Statistical Reporting Guidelines for Clinical Trials Data. *Indian J Psychol Med* 2012;34:25-9.

Source of Support: Nil, **Conflict of Interest:** None.