International Journal of Pharmacology

ISSN 1811-7775

science alert

ansinet
Asian Network for Scientific Information

International Journal of Pharmacology 6 (4): 354-359, 2010
ISSN 1811-7775
© 2010 Asian Network for Scientific Information

Some Issues Related to the Reporting of Statistics in Clinical Trials Published in Indian Medical Journals: A Survey

Jaykaran, P. Yadav, N. Chavda and N.D. Kantharia
Department of Pharmacology, Govt. Medical College, Surat, 395001, India

Abstract: Standard of reporting of statistics in clinical trials published in Indian medical journals generally low. A growing body of literature points to persistent statistical errors, flaws and deficiencies in clinical trials published in most of the medical journals published in India. In this study, we present a short review of few frequently observed statistical errors in clinical trials published in Indian Medical Journals. Seven potential statistical errors and shortcomings, differentiated for the distinct phases of medical research are presented and discussed. Statisticians should be included in early phase of study design, as mistakes at this point can have major impact, negatively affecting all subsequent stages of medical research. Consideration of issues discussed in this short review, when planning, conducting and preparing medical research manuscripts, should help further enhance statistical quality in clinical trials published in Indian medical journals.

Key words: Sample size, inappropriate statistical tests, intention to treat principle, multiple endpoints, baseline comparisons, confidence interval

INTRODUCTION

Randomized Controlled Trials (RCTs) are considered to be credible evidence of efficacy of a new drug or intervention hence considered as gold standard of evidence based clinical practice. Clinicians make treatment decisions based on RCTs published in peer reviewed journals. It is observed that there are some problems existed with the reporting of statistics in clinical trials and quality of reporting of statistics in clinical trials is less than satisfactory (Hopewell *et al*., 2010; Agha *et al*., 2007; Lai *et al*., 2007; Gagnier *et al*., 2006; Rios *et al*., 2008; Anttila *et al*., 2006; Scales *et al*., 2008; Wang *et al*., 2007a; Mills *et al*., 2005; Pienaar *et al*., 2002; Karan *et al*., 2010). This is a serious issue as inappropriate or wrong use of statistics may lead to false conclusions, fake research results and wastage of resources (Strasak *et al*., 2007).

Problem of inappropriate statistics is addressed in various forums and it is pointed out that inappropriate reporting of statistics in clinical trials not only considered as unethical but also may lead to use of wrong therapeutic decisions (Altman, 1998). Efforts at various levels done by biomedical journal editors to improve the quality of reporting of statistics in the clinical trials like providing various guidelines related to reporting of statistics or including statistician in editorial team etc. (Gore *et al*., 1992; Altman, 1998; Altman *et al*., 1983; Murray, 1991). Despite all these efforts clinical trials with poor reporting of statistics are still being published (Hopewell *et al*., 2010; Agha *et al*., 2007; Lai *et al*., 2007; Gagnier *et al*., 2006; Rios *et al*., 2008; Anttila *et al*., 2006; Scales *et al*., 2008; Wang *et al*., 2007a; Mills *et al*., 2005; Pienaar *et al*., 2002; Karan *et al*., 2010).

In this short study, we are discussing some important statistical errors which we found in reporting of statistics in clinical trials published in Indian medical journals. Our aim for writing this study is not only the exploration of statistical pitfalls but also provide guidance to researchers so that these mistakes can be prevented.

INADEQUATE SAMPLE SIZE

In clinical trials sample should be big enough to have a high chance of detecting, as statistically significant, a worthwhile effect if it exists and thus to be reasonably sure that no benefit exists if it is not found in trial. For sample size calculation in hypothesis testing researcher must know the effect size, standard deviation, significant level and power of study (Altman, 1991). Effect size and standard deviation of new interventions can be calculated by pilot study. Not only the sample size should be calculated but it should also be mentioned that how it was calculated. Very small sample size results in parameter estimates that are unnecessarily imprecise, enhances potential for failed randomization, yield hypothesis testing that is underpowered and biased because assumptions underlying the applied statistical methods could not be examined adequately (Tyson, 2004). In present study

Corresponding Author: Dr. Jaykaran, Department of Pharmacology, Govt. Medical College, Surat, 395001, India

done for reporting of statistics in clinical trials published in Indian Medical Journals we found that information related to calculation of sample size was not mentioned in many of clinical trials and median sample size per group was less as compared to clinical trials published in western journals (Hopewell *et al*., 2010; Agha *et al*., 2007; Lai *et al*., 2007; Gagnier *et al*., 2006; Rios *et al*., 2008; Anttila *et al*., 2006; Scales *et al*., 2008; Wang *et al*., 2007b; Mills *et al*., 2005; Pienaar *et al*., 2002; Karan *et al*., 2010; Jaykaran *et al*., 2009). In a study done by us for negative clinical trials published in Indian Medical Journals, on the basis of post hoc power calculation we have found that about half of the clinical trials were underpowered for the 30 and 50% difference between the outcome and exact sample size was not calculated in any of the clinical trial (Jaykaran *et al*., 2010a). Labeling a clinical trial negative on the basis of inappropriate statistics is a matter of concern and unethical (Halpern *et al*., 2002).

Sample size should be calculated before the start of clinical trial at the study design phase. If the investigators are from the non statistical background than help of statistician should be sought. This calculated sample size should be adjusted for the expected loss to follow up data.

INAPPROPRIATE STATISTICAL TESTS

This is considered as major problems in reporting of statistics in clinical trials published in Indian Medical Journals. Wrong/inappropriate statistical tests are used to analyze the data. Wrong statistical tests are used when there is incompatibility of statistical test with type of data examined or when unpaired tests are used for paired data or vice versa or when of parametric methods used at the place of nonparametric methods and inappropriate test used for the hypothesis under investigation (Strasak *et al*., 2007). It is very important for the investigator to understand that each statistical test has some underlying assumptions. These assumptions should be fulfilled before using the statistical test. Violation of assumptions leads to serious threat to validity of results obtained by that particular statistical method. These assumptions should not only be fulfilled but should also be mentioned in the manuscript (Lang, 2004). Even basic statistical tests like t-test and Chi-square tests are misused because of violation of their assumptions (Goodman and Hughes, 1992). In present study, we found that not a single clinical trial mentioned about the fulfillment of assumptions of statistical methods used in analysis of data. Distribution of data was checked in only few of the clinical trials (Karan *et al*., 2010).

Investigator should be very clear about the type of data, distribution of the data, type of various statistical methods and there assumptions. Help of an experienced statistician should be taken. Decision regarding the method of analysis should be taken at the design phase. When using t test or Chi-square test correct version of these tests should be selected as they are available in various versions (Goodman and Hughes, 1992). In the case of Chi-square test if expected count in cell is less than 5 than Yates-continuity-correction or preferably exact test should be used (Goodman and Hughes, 1992).

Complete information regarding the statistical methods should be included in manuscript like whether the test is one tailed or two tailed, paired or unpaired etc. if some unusual or obscured method is used than author should mention the justification of that method and suitable reference should be included. One common problem we observed in statistics section of clinical trials published in Indian Medical Journal is where appropriate statement. In many clinical trials it is mentioned that appropriate statistical tests were used to analyze the data or t-test were used for quantitative data and chi square test used for qualitative data. These kinds of statements should be avoided as they are providing the insufficient information (Goodman and Hughes, 1992).

FAILURE TO REPORT ADJUSTMENT FOR MULTIPLE ENDPOINTS

Multiplicity of inferences can arise in a various ways in clinical trials like multi-sample comparisons, interim and subgroup analysis and particularly, multiple endpoints. This multiplicity is associated with the false positivity, i.e., likelihood of getting the significant result just by chance. Hence, the appropriate threshold to declare a test's p-value significant becomes complex when more than one test is performed as it is difficult to maintain type I error in this condition. If, with one statistical test, the chance of a significant result is 5%, then after 20 tests, it will increase to 40%. These result are, however, not based on a true treatment effect, but, rather, on the play of chance (Neuhauser, 2006).

Though ICH E9 guideline on biostatistics in clinical trials recommends selection of one primary variable, it is observed that investigators typically examine more than one primary endpoint and a single primary endpoint may not be sufficient in many disease areas (Neuhauser, 2006).

In present study, we observed that in many of clinical trials published in Indian Medical Journals there were no division of primary and secondary endpoints which increases the chance of type I error (Karan *et al*., 2010;

Jaykaran *et al*., 2010b). We observed that median number of end points reported in clinical trials published in Indian Medical Journals as well as median number of end points which were used for testing of significance was four (Karan *et al*., 2010). None of the clinical trials mentioned about adjustment of multiple endpoints (Karan *et al*., 2010). This raises serious questions regarding the validity of results of these clinical trials as many of the clinical trials out of these may be false positive.

Investigator should use various methods described to adjust multiple end points and hence type I error (Neuhauser, 2006). Not only ICH E9 guideline but also CONSORT statement demands use of procedure for adjustment of multiple endpoints (Altman *et al*., 2001).

Cleophas and Zwinderman (2006) observed that out of the 16 positive clinical trials published in British Medical Journal (BMJ) in 2004 only 8 remain positive after Bonferroni adjustment. In present study for clinical trials published in Indian Medical Journals we found that out of 49 positive clinical trials, positive significant test after the Bonferroni adjustment for multiple endpoints was seen in only 22 (Jaykaran *et al*., 2010c).

More subgroup also increases the chance of type I error. In order to prevention of type I error subgroups should be stated in advance in the protocol and should be limited in number with the primary goal of evaluating qualitatively the internal consistency (Moreira *et al*., 2001). Wang *et al*. (2007b) also observed that among the original articles published in the New England Journal of Medicine during the period from July 1, 2005, through June 30, 2006, subgroup analysis was reported in 61% of trials. We observed that for the clinical trials published in Indian Medical Journals subgroup analysis is not this much common? Various guidelines for the reporting of subgroup analysis are published, which can be followed for the reporting of subgroup analysis (Wang *et al*., 2007b).

REPORTING THE RESULTS ONLY AS P-VALUES

The p-values describe the risk of making a type I error (that is, the risk of concluding that there is a significant difference between groups when in fact there is no such difference). We observed that in most of the clinical trials only p-value was given to show the difference between two groups. It is reported that p-values are often misinterpreted and even if it is interpreted correctly it has some limitations (Bailar and Mosteller, 1988). Results should be explained as absolute difference between the two groups for the endpoint as well as 95% confidence interval around the difference with or instead of p-value. Confidence interval tells the reader exactly the range of values with which the data are statistically compatible. That is, they define all of the potential results that are supported by the data (Gardner and Altman, 1986). When using statistical estimation for comparison of two groups, confidence intervals should rather be given for the differences between groups, than for each group itself (Strasak *et al*., 2007). Though instruction to author of many Indian Medical Journal mention about the reporting of exact p-value still many clinical trials published in Indian Medical Journals reported arbitrary p-value like $p<0.05$ or $p>0.05$ or p = NS. Exact p-values should be reported rather than arbitrary one.

Writing confidence interval also helps in differentiating between clinical significance and statistical 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. Confidence interval around the absolute difference will give idea about the clinical significance (McGuigan, 1995). In study done by us for 33 negative clinical trials published in Indian medical journals it was observed that confidence interval was not reported in any of the clinical trial (Jaykaran *et al*., 2010a).

UNNECESSARILY REPORTING BASELINE STATISTICAL COMPARISONS IN RANDOMIZED TRIALS

We observed that in many of the clinical trials baseline comparisons between the two groups for various endpoints are reported with p-values for each. In randomized controlled trials recruitments of subjects done with proper randomization techniques hence the difference observed between the groups are considered to be chance finding. So, there is no need for reporting of baseline comparisons. Any difference between the groups should be adjusted by various statistical techniques during the analysis of results but p-value need not be reported. supposing alpha is set at 0.05, of every 100 baseline comparisons in trials, 5 should be statistically significant, just by chance. However, in a study it was found that among 1,076 baseline comparisons in 125 trials, only 2% were significant at the 0.05 level (Altman and Dore, 1990). In this study done for clinical trials published in Indian Medical Journals we observed that baseline comparisons between groups were done in 89% trials. Each time comparison was done by statistical tests and p-value was reported (Karan *et al*., 2010).

NOT FOLLOWING INTENTION TO TREAT PRINCIPLE (ITT)

In clinical trials published in Indian Medical Journals ITT principle is usually not followed. Meaning of intention to treat principle is that all patients randomized into clinical trial are to be accounted for the primary analysis and all primary event observed during the follow up period are to be accounted for as well. If either of these aspects is not adhered to, the analysis of results may easily be biased in unpredictable directions and thus the interpretation of the results compromised (May *et al*., 1981). There are two common reasons that are commonly given to exclude randomized patients from analysis, post hoc ineligibility assessment and lack of patient compliance. In some clinical trials it is proved that how not accounting all subjects may lead to serious bias in the results (Anturane Reinfarction Trial Research Group, 1978, 1980).

As perfect compliance is often not attainable and exclusion of noncompliant patients may introduce serious bias, those who design trials must attempt to adjust the design and increase the sample size to compensate for non adherence. The first goal should be to minimize noncompliance by choosing the best tolerated intervention strategy and then to monitor noncompliance during the conduct. An increase in sample size will often be necessary to compensate for the dilution effect of noncompliance. For example, a 10% noncompliance in the intervention arm can require an increase of 23% in the sample size. A 20% intervention noncompliance may require a 56% sample size increase. Whilst such an increase is most costly and time consuming, the analysis will not be biased and thus the conclusions not compromised. Our study for clinical trials published in Indian Medical Journals revealed that Intension to treat principle was mentioned in only 8% of trials (Saurabh *et al*., 2010a).

GIVING SEM INSTEAD OF SD TO DESCRIBE DATA

Use of descriptive statistics is very common in clinical trials published in various Indian Medical Journals. For the ratio and interval data following the normal distribution most common descriptive statistics are Mean and Standard Deviation (SD) and for data not following the normal distribution it is Median and Range. For the ordinal and nominal data the most common descriptive statistics are median and frequency respectively. It is observed in clinical trials published in various medical journals that Mean and Standard Error of Mean (SEM) are used to describe the variability within the sample. Investigators should understand the difference between SEM and SD. The SD represents the variability within the sample. More the SD more will be variability in the sample. So, when authors mention in there article Mean±SD, it means they want to express the mean and variability around the mean in the form of SD. The SEM represents the uncertainty of how sample mean represent the population mean. It shows the probability of sample mean representing the population mean. It does not represent variability within the sample (Lang, 2004). The value of SD is less than SEM so when data is represented in the form of Mean±SEM, because of small value of SEM readers may assume variability within the sample very less which is actually not so. Even to show the probability of sample value to represent population value confidence interval is better than SEM (Gardner and Altman, 1986). So to represent the sample variability use of Mean±SEM is not statistically correct and it should be avoided (Gardner and Altman, 1986; Nagele, 2003).

During the reporting of mean and SD, Instead of writing Mean±SD the better way of representation is Mean(SD) as it will decrease the chance of confusion with confidence interval (Gardner and Altman, 1986).

In present study done for clinical trials published in Indian Medical Journals we observed that in few clinical trials SEM is still reported to show variability within the sample (Saurabh *et al*., 2010b).

CONCLUSIONS

Present aim for this study was to highlight some frequently observed statistical errors in clinical trials published in Indian Medical Journals. Various kinds of statistical errors can be observed in clinical trials published in India but here we report some important and frequently observed errors. Similar kind of errors also observed in clinical trials published in non Indian Medical Journals. These statistical errors can easily be handled by careful planning at the design phase of clinical trials. Statistician should be included at this stage for the better planning. Now a day various kinds of statistical software are available which helps investigator to perform the analysis of there own data but insufficient knowledge of statistical concepts may lead to the problem in interpretation. So, role of statistician is important.

Medical students should be encouraged to learn statistical concepts at undergraduate as well as postgraduate level. Incorporation of short course of research methodology and biostatistics at the postgraduate level may help student to understand and write the research properly.

Editors of medical journals should broaden the editorial team to include few statisticians. Each article should also be send for statistical review before sending it to peer review. Peer reviewers should be trained for the statistical concepts. Various efforts are done to enhance the quality of statistical reporting in medical journals and improvement is observed over time. Similar or more aggressive efforts should be done for the Indian medical journals.

REFERENCES

Agha, R., D. Cooper and G. Muir, 2007. The reporting quality of randomized controlled trials in surgery: A systematic review. Int. J. Surg., 5: 413-422.

Altman, D.G., S.M. Gore, M.J. Gardner and S.J. Pocock, 1983. Statistical guidelines for contributors to medical journals. BMJ., 286: 1489-1493.

Altman, D.G. and C.J. Dore, 1990. Randomisation and baseline comparisons in clinical trials. Lancet, 335: 149-153.

Altman, D.G., 1991. Practical Statistics for Medical Research. Chapman and Hill, London.

Altman, D.G., 1998. Statistical reviewing for medical journals. Stat. Med., 17: 2661-2674.

Altman, D.G., K.F. Schulz and D. Moher, 2001. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Ann. Int. Med., 134: 663-694.

Anttila, H., A. Malmivaara, R. Kunz, I. Utti-Ramo and M. Makela, 2006. Quality of reporting of randomized, controlled trials in cerebral palsy. Pediatrics, 117: 2222-2230.

Anturane Reinfarction Trial Research Group, 1978. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction: The anturane reinfarction trial. New Engl. J. Med., 298: 289-295.

Anturane Reinfarction Trial Research Group, 1980. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. New Engl. J. Med., 302: 250-256.

Bailar, J.C. and F. Mosteller, 1988. Guidelines for statistical reporting in articles for medical journals. Ann. Intern. Med., 108: 266-273.

Cleophas, T.J. and A.H. Zwinderman, 2006. Clinical trials are often false positive: A review of simple methods to control this problem. Curr. Clin. Pharmacol., 1: 1-4.

Gagnier, J.J., J. DeMelo, H. Boon and P. Rochon, 2006. Bombardier C. Quality of reporting of randomized controlled trials of herbal medicine interventions. Am. J. Med., 119: 800-811.

Gardner, M.J. and D. Altman, 1986. Confidence intervals rather than P values: Estimation rather than hypothesis testing. BMJ., 292: 746-750.

Goodman, N.W. and A.O. Hughes, 1992. Statistical awareness of research workers in british anaesthesia. Br. J. Anaesth., 68: 321-324.

Gore, S.M., G. Jones, S.G. Thompson, 1992. The Lancet's statistical review process: Areas for improvement by authors. Lancet, 340: 100-102.

Halpern, S.D., J.H. Karlawish and J.A. Berlin, 2002. The continuing unethical conduct of underpowered clinical trials. JAMA., 288: 358-362.

Hopewell, S., S. Dutton, L.M. Yu, A.W. Chan and D.G. Altman, 2010. The quality of reports of randomized trials in 2000 and 2006: Comparative study of articles indexed in PubMed. BMJ., 340: c723-c723.

Jaykaran, P. Yadav, P. Bhardwaj and J. Goyal, 2009. Problems in reporting of statistics: comparison between journal related to basic science with journal related to clinical practice. Internet J. Epidemiol., Vol. 7(1). 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

Jaykaran, M. Saurabh, P. Yadav, N. Chavda and P. Mody, 2010a. False positive clinical trials published in four Indian medical journals: A survey. J. Pharm. Res., 3: 822-823.

Jaykaran, M.K. Saurabh, S. Gohiya, V. Gohiya, G. Sharma and N. Chavda, 2010b. Reporting of power, sample size and confidence interval in negative clinical trials published in four Indian medical journals. J. Pharmacy Res., 3: 298-300.

Jaykaran, S. Gohiya,V. Gohiya, G. Sharma, M.K. Saurabh and P. Yadav, 2010c. Reporting of the methodological quality and ethical aspects in clinical trials published in Indian Journals: A survey. J. Pharm. Res., 3: 307-309.

Karan, J., N.D. Kantharia, P. Yadav and P. Bhardwaj, 2010. Reporting statistics in clinical trials published in Indian Journals: A survey. Pak. J. Med. Sci., 26: 212-216.

Lai, T.Y., V.W. Wong, R.F. Lam, A.C. Cheng, D.S. Lam and G.M. Leung, 2007. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. Ophthalmic Epidemiol., 14: 390-398.

Lang, T., 2004. Twenty statistical errors even you can find in biomedical research articles. Croat. Med. J., 45: 361-370.

May, G.S., D.L. DeMets and L.M. Friedman, 1981. The randomized clinical trial: Bias in analysis. Circulation, 64: 669-673.

McGuigan, S.M., 1995. The use of statistics in the British Journal of Psychiatry. Br. J. Psychol., 167: 683-688.

Mills, E.J., P. Wu, J. Gagnier and P.J. Devereaux, 2005. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. Contemp. Clin. Trials, 26: 480-487.

Moreira, E.D., Z. Stein and E. Susser, 2001. Reporting on methods of subgroup analysis in clinical trials: A survey of four scientific journals. Brazilian J. Med. Biol. Res., 34: 1441-1446.

Murray, G.D., 1991. Statistical guidelines for the British Journal of Surgery. Br. J. Surg., 78: 782-784.

Nagele, P., 2003. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. Br. J. Anaesth., 90: 514-516.

Neuhauser, M., 2006. How to deal with multiple endpoints in clinical trials. Fundamental Clin. Pharmacol., 20: 515-523.

Pienaar, E.D., J. Volmink, M. Zwarenstein and G.H. Swingler, 2002. Randomised trials in the South African Medical Journal, 1948-1997. South Afr. Med. J., 92: 901-903.

Rios, L.P., A. Odueyungbo, M.O. Moitri, M.O. Rahman and L. Thabane, 2008. Quality of reporting of randomized controlled trials in general endocrinology literature. J. Clin. Endocrinol. Metab., 93: 3810-3816.

Saurabh, M.K. Jaykaran, N. Chavda, P. Yadav, N.D. Kantharia and L. Kumar, 2010a. Misuse of standard error of mean (SEM) when reporting variability of a sample: A critical appraisal of four Indian journals. J. Pharm. Res., 3: 277-279.

Saurabh, M.K., Jaykaran, P. Yadav, N. Chavda, R.V. Lokesh, S. Deodhare and M. Chaudhari, 2010b. Reporting of missing data in clinical trials published in four Indian journals: A survey. J. Pharm. Res., 3: 277-279.

Scales, Jr. C.D., R.D. Norris, G.M. Preminger, J. Vieweg, B.L. Peterson and P. Dahm, 2008. Evaluating the evidence: Statistical methods in randomized controlled trials in the urological literature. J. Urol., 180: 1463-1467.

Strasak, A.M., Q. Zaman, P.P. Karl, G. Göbel and H. Ulmer 2007. Statistical errors in medical research-a review of common pitfalls. Swiss MED Wkly., 137: 44-49.

Tyson, H., 2004. Ten categories of statistical errors: A guide for research in endocrinology and metabolism. Am. Physio. Endocrinol. Metab., 286: 495-501.

Wang, G., B. Mao, Z.Y. Xiong, T. Fan and X.D. Chen *et al.*, 2007a. The quality of reporting of randomized controlled trials of traditional Chinese medicine: A survey of 13 randomly selected journals from mainland China. Clin. Ther., 29: 1456-1467.

Wang, R., S.W. Lagakos, J.H. Ware, D.J. Hunter and J.M. Drazen, 2007b. Statistics in medicine-reporting of subgroup analyses in clinical trials. N. Engl. J. Med., 21: 2189-2194.