JSCIRES RESEARCH ARTICLE

Statistics: From arbitrary to accurate reporting is the need of the hour: Results of a preliminary study

S. Kannan*, S. Gowri[1]

Department of Health Sciences, [1]Department of Oral Health, College of Medicine, Nursing and Health Sciences, Fiji National University, Samabula, Fiji

ABSTRACT

Statistics has to be used appropriately in order to draw credible result. Statistical flaws have been detected in many biomedical journals but have not been evaluated for and published in dental journals. Hence, the present study was envisaged. Original Research Articles published in Contemporary Clinical Dentistry from January to December 2012 were downloaded from the journal's website and were assessed in the present study. A structured, validated questionnaire was developed to assess the quality of reporting statistics. Descriptive statistics were used to represent the data. A total of 62 articles were retrieved for the assessment. Only 1 out of 55 articles mentioned a *prior* sample size calculation. One-sixth of randomized clinical studies have mentioned the method used for generating random sequence, and none explained the process of allocation concealment. Only one-tenth of the articles with nonrandomized study design compared baseline characteristics. The majority of the studies (97.3%) employing parametric tests have not mentioned whether the assumptions have been checked. Except for one, no other articles have reported confidence interval at least for the primary outcome. None, except two in our study, albeit having mentioned multiple *P* values, made an attempt to adjust the same using any of the tests. We found poor reporting of statistics and inferred that more attention is needed from both the researcher and journal editor.

Keywords: Dental journals, statistics, quality

INTRODUCTION

Statistics is an essential component in biomedical research. The quality of research work depends both on the study design and application of accurate statistics.[1] The use of statistics in biomedical field has increased significantly in recent times, and more complex and modern statistical tools are also being employed.[2] If the statistics used in a study is inappropriate, the conclusion drawn becomes invalid. Hence, biostatisticians need to be consulted before starting and while the study is being conducted which is hardly given consideration by the researchers. Furthermore, only a few journals have statistical experts on their editorial board.[3] Hence, statistical flaws have been detected in many biomedical journals[4-9] including those published in the field of animal research[10] and ranges 30–90%. Although, studies have been conducted to evaluate the quality of randomized controlled studies in dentistry,[11,12] there was none pertaining to the quality of reporting statistics in particular. Hence, the present study was conducted as an initial attempt to evaluate the same. The results presented here are from an ongoing study where multiple journals are being assessed.

***Address for correspondence:**
E-mail: skannandr@gmail.com

Access this article online

Quick Response Code:

Website: www.jscires.org

DOI: 10.4103/2320-0057.167245

METHODS

Original Research Articles published between January and December 2012 in Contemporary Clinical Dentistry[13] were downloaded from their respective website and were assessed in the present study. Case Reports, Review

Articles, Letter to Editor, and Short Communications were excluded. A structured, validated questionnaire was developed to assess the quality of reporting statistics. The quality was assessed independently by both the authors (KS and GS). If there was any disagreement between the authors, it was resolved by a discussion. The level of agreement achieved was excellent (kappa = 0.90). Descriptive statistics (proportion of articles in each of the items as described in the questionnaire was calculated for each of the journals) was used to represent the data.

RESULTS

Demographics

As of January 30, 2014, four issues and two supplements were available in the journal website. A total of 62 Original Articles (12 – issue 1; 12 – issue 2; 12 – issue 3; 16 – issue 4; 2 – supplement 1; 8 – supplement 2) were downloaded and analyzed.

Types of Statistical Tests Used

Of the 62 articles, 6 (9.7%) did not use any statistics (including descriptive) while 3 (4.9%) have used only descriptive statistics. The details of the type of inferential statistics used are mentioned in Table 1. Of those that had used statistical tests (n = 51), only 18 (35.3%) have mentioned the name of the statistical software used.

Details about the Reporting of Statistics

Only 1 (2%) study has mentioned that *prior* sample size calculation was performed. Many studies have used multiple statistical tests with median (range) number of 6 (1–30) of which only two have used correction tests for the multiplicity of P values. Of the 51 studies, only 2 (3.9%) have reported the number of tails of tests and 16/51 (31.4%) have reported exact P values. Details of other information regarding each items of the questionnaire are represented in Table 2.

Table 1: Types of statistical tests employed *n* (%)

Name of the inferential test	*n* (%)*
Student's *t*-test	
Unpaired	9 (11.4)
Paired	10 (12.7)
ANOVA	
One-way	14 (17.7)
Repeated measures	1 (1.3)
Correlation	
Pearson	6 (7.6)
Spearman	2 (2.6)
Regression	
Simple	4 (5.2)
Multiple	3 (3.9)
Logistic	2 (2.6)
Wilcoxon signed rank sum test	3 (3.9)
Mann-Whitney U-test	4 (5.2)
Krukal-Wallis H-test	4 (5.2)
Chi-square test	12 (15.2)
Fisher's exact probability test	3 (3.9)
Intra-class correlation co-efficient test	1 (1.3)
OR	1 (1.3)

*The total shall be more than 51 as more than one statistical test was used in some studies. OR=Odds ratio

Table 2: Details of the responses collected for the questionnaire

Items	Number/total number of applicable studies (%)
Mention of sample size calculation	1/55 (1.8)
Mention of method of randomization and allocation concealment in case of randomized studies	1/6 (17)
Comparability of baseline characteristics in case of nonrandomized studies	1/10 (10)
Use of SE to describe the variability of the data set rather than SD	1/44 (2.3)
Statement that the assumptions of normality have been checked in case of use of a parametric test	1/37 (2.7)
Correction of inflation of type I error in case of multiplicity of the statistical tests	2/51 (3.9)
Use of two group test for three or more than three groups	4/24 (16.7)
Use of three group test for two groups	1/20 (5)
Yate's correction in case of Chi-square with size less than 100	1/11 (9)
Use of Chi-square when the expected frequency of a cell is less than 5	7/11 (63.6)
Correlation tests used for testing agreement	3/8 (37.5)
Mention of confidence interval at least for the primary outcome variable	1/51 (2)

SD=Standard deviation, SE=Standard error

DISCUSSION

The present study is the first of this kind to the best of our knowledge which assesses the quality of reporting statistics in a dental journal. We found poor reporting corroborating many other studies in the biomedical literature.[4-9]

Only 1 out of 55 articles has mentioned that *a priori* sample size calculation has been performed. Sample size has to be appropriate and essentially should be calculated before the start of the study so that the result of the study is credible. Inflation of Type I and II errors may be seen in studies with small sample size.[14] While studies have reported that

around 50% of the articles have mentioned sample size calculation,[15] the reporting rate is very low in the present study. Furthermore, the discrepancy between calculated and recruited sample size was observed to be 15–30% in a recent study in the specialty of orthodontics and in other multi-specialty dental journals.[16,17] It is mandatory to calculate sample size in any research and many software and online tools are available for the same.[18]

In the present study, only one-sixth of randomized clinical studies have mentioned the method used for generating random sequence and none about the process of allocation concealment. Randomization allocates study participants based on chance and not by choice, thus avoiding selection bias.[19] Studies with such design are considered to be the gold standard in evidence-based medicine.[20] Different types of randomization such as simple, stratified, block can be used depending on the needs.[21] Although random sequence generation is proper, unless the allocation of the participants has been concealed, bias can creep into leading to the unreliability of the study results.[22] Hence, studies employing randomization design are expected to report the methods used for generating such sequence as well as allocation concealment. Although selection bias may be difficult to control in nonrandomized clinical studies, comparable baseline characteristics/outcome affecting the study results may reduce the chances of selection bias.[23] However, we found that only one-tenth of articles with nonrandomized study design did so.

We found that a majority of studies employing parametric tests have not mentioned whether the assumptions for performing such tests have been checked with the collected data in the study. Parametric tests should be applied only when their assumptions (normal distribution and homogeneity of the variance) are satisfied.[24] Studies have shown that the use of parametric statistical tests for a nonparametric data inflates Type II error leading to false-negative conclusions.[25] Similarly, we also found that majority of the studies have used Chi-square tests even when the expected frequency of a cell is less than 5, again violating the assumptions of the test. Additionally, tests assessing correlation were used for testing agreements between different diagnostic methods in three studies, which is again misleading.[26,27] Using statistical tests without ensuring that the assumptions have been met or violating the same will result in inaccurate results.[28]

Except for one, no other article in the present study has reported confidence interval at least for the primary outcome. Almost all studies have reported *P* values conveying whether the results of the study have emerged by chance.[29] The confidence interval is an estimate of the precision of the study and gives a range in which the true population value lies.[30] Thus, it provides more information that is clinically important than *P* value and many researchers advocate the report of confidence intervals.[31] Also, when the number of statistical comparison increases in a study, the Type I error occurs. Hence, several methods for correcting such inflation have been recommended.[32] None of the articles, except two, in our study albeit having mentioned multiple *P* values, did attempt in adjusting the same using any of these tests. Hence, there are many chances that the conclusions that have been drawn by these studies may not hold true.

Many studies[5-9] including the present one have been found to have many lacunae in reporting statistics in biomedical literature. Studies have shown that only around 40% of the dental journals are noncompliant with regard to instructions regarding the data management/statistical analysis.[33] Authors have suggested that journals have statistical guidelines and our previous study detected that only 1 out of the 10 journals had.[3] Additionally, studies do suggest that a statistical expert shall be included in the editorial committee of a journal for an effective review of the submitted manuscripts.[34,35]

CONCLUSION

The present study is limited in having assessed only one journal in the specialty, but interestingly we found out various statistical flaws in the published articles in the same. This emphasizes the requirement of more attention from both the researchers and journal editor related to the aspects of analyzing and reporting statistics in a scientific method instead arbitrarily.

Although the study is limited in having assessed only one journal in the specialty, to conclude, we found poor reporting of statistics and more attention is needed from both the researcher and journal editor.

REFERENCES

1. Yates F, Healy MJ. How should we reform the teaching of statistics. J R Stat Soc 1964;127:199-210.
2. Jin Z, Yu D, Zhang L, Meng H, Lu J, Gao Q, *et al.* A retrospective survey of research design and statistical analyses in selected Chinese medical journals in 1998 and 2008. PLoS One 2010;5:e10822.

3. Kannan S, Deshpande SP, Gogtay NJ, Thatte UM. Policy of reviewing statistics in Indian medical and surgical journals. J Pharmacol Pharmacother 2013;4:139-40.
4. Altman DG. Statistics in medical journals: Developments in the 1980s. Stat Med 1991;10:1897-913.
5. Yim KH, Nahm FS, Han KA, Park SY. Analysis of statistical methods and errors in the articles published in the Korean journal of pain. Korean J Pain 2010;23:35-41.
6. Kotur PF. Statistics in biomedical journals. Indian J Anaesth 2006;50:166-8.
7. Jaykaran, Yadav P. Quality of reporting statistics in two Indian pharmacology journals. J Pharmacol Pharmacother 2011;2:85-9.
8. Simundic AM, Nikolac N. Statistical errors in manuscripts submitted to Biochemia Medica journal. Biochem Med 2009;19:294-300.
9. Milewska AJ, Miewski R, Gorska U, Jankowska D. Statistical Methods in Polish Medical Publications. Studies in logic, grammar, and rhetoric. 2010;21:81-90.
10. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, *et al.* Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One 2009;4:e7824.
11. Sjögren P, Halling A. Quality of reporting randomised clinical trials in dental and medical research. Br Dent J 2002;192:100-3.
12. Cioffi I, Farella M. Quality of randomised controlled trials in dentistry. Int Dent J 2011;61:37-42.
13. Contemporary Clinical Dentistry. Available from: http://www.contempclindent.org/backissues.asp. [Last accessed on 2014 Jan 31].
14. Sakpal TV. Sample size estimation in clinical trial. Perspect Clin Res 2010;1:67-9.
15. Shahravan A, Haghdoost AA, Rad M, Hashemipoor M, Sharifi M. Sample size calculation of clinical trials published in two leading endodontic journals. Iran Endod J 2014;9:56-60.
16. Koletsi D, Fleming PS, Seehra J, Bagos PG, Pandis N. Are sample sizes clear and justified in RCTs published in dental journals? PLoS One 2014;9:e85949.
17. Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: Are numbers justified? Eur J Orthod 2014;36:67-73.
18. McCrum-Gardner E. Sample size and power calculations made simple. Int J Ther Rehabil 2010;17:10-4.
19. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: Chance, not choice. Lancet 2002;359:515-9.
20. Zhao W. Selection bias, allocation concealment and randomization design in clinical trials. Contemp Clin Trials 2013;36:263-5.
21. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. J Hum Reprod Sci 2011;4:8-11.
22. Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: Importance of allocation concealment and masking. Fam Med 2007;39:132-7.
23. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al.* Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii-x, 1-173.
24. Chan YH. Biostatistics 102: Quantitative data – Parametric and non-parametric tests. Singapore Med J 2003;44:391-6.
25. Qualls M, Pallin DJ, Schuur JD. Parametric versus nonparametric statistical tests: The length of stay example. Acad Emerg Med 2010;17:1113-21.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
27. Zaki R, Bulgiba A, Nordin N, Azina Ismail N. A systematic review of statistical methods used to test for reliability of medical instruments measuring continuous variables. Iran J Basic Med Sci 2013;16:803-7.
28. Lang T. Twenty statistical errors even you can find in biomedical research articles. Croat Med J 2004;45:361-70.
29. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or P value?: Part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009;106:335-9.
30. Gupta SK. The relevance of confidence interval and P value in inferential statistics. Indian J Pharmacol 2012;44:143-4.
31. Akobeng AK. Confidence intervals and P values in clinical decision making. Acta Paediatr 2008;97:1004-7.
32. Pocock SJ. Clinical trials with multiple outcomes: A statistical perspective on their design, analysis, and interpretation. Control Clin Trials 1997;18:530-45.
33. Mathur VP, Dhillon JK, Kalra G, Sharma A, Mathur R. Survey of instructions to authors in Indian and British Dental Journals with respect to ethical guidelines. J Indian Soc Pedod Prev Dent 2013;31:107-12.
34. Petrovecki M. The role of statistical reviewer in biomedical field journal. Biochem Med 2009;19:223-30.
35. Young J. Statistical errors in medical research – A chronic disease? Swiss Med Wkly 2007;137:41-3.

How to cite this article: Kannan S, Gowri S. Statistics: From arbitrary to accurate reporting is the need of the hour: Results of a preliminary study. J Sci Res 2015;4:115-8.

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