Statistical Errors in Medical Journals (A Critical Appraisal)

Asif Hanif,[1] **Tahira Ajmal**[2]

Abstract

Background: Importance of Statistics and its appropriate appliance in every field of life are inevitable. Research conducted in all sciences, particularly in medicine stipulates the usage of statistics to make the results unswerving and authentic. Any research, how so ever expensive and hefty, has no worth and fails to accomplish the tasks for which it is conducted, if statistics involved in it contains error. It has been attested now, that many articles published in medical journals restrain errors in them. Particularly in Pakistan, the use of statistics is not given the suitable worth, resulting many statistical flaws. Thus it has become the need of hour to investigate these errors, highlight and consequently make an effort to eradicate them from medical research. This study has been designed to scrutinize such errors in local medical journals of Pakistan.

Study Design: Cross – sectional study design was used.

Methodology: 80 research articles published in indexed and recognized local journals of Pakistan were reviewed. The selection of those research articles was random and those were easily and freely available online. Case series and case reports were excluded.

Results: Descriptive statistics was only used in 32 (28.75%), inferential statistics along with the descriptive statistics was used in 33 (41.25%) articles. Design of a study was not mentioned in 42 (52.5%) articles, in 74 (92.5%) articles there was no criteria for sample size estimation and power calculation. Wrong statistical analysis was done in 23 (28.75%) articles. Out of 80 articles evaluated, 57 (71.2%) articles used inferential statistics but did not mention whether the assumptions for statistical test used, met or not.

Conclusion: In order to endorse medical research with authentication and reliability, the role of biostatistician needs to be divulged. Thus biostatistician should be taken in editorial boards or articles should be referred by the biostatisticians or statisticians for approval in order to keep the standards of research upgraded.

Keywords: Statistical error, biomedical research and Statistical methods.

Hanif A.[1]

Official Biostatistician Annals of K.E.M.U., Department of Orthopedics Surgery and Traumatology and Students of M. Phil Biostatistics and Epidemiology at King Edward Medical University / Mayo Hospital, Lahore – Pakistan

Ajmal T.[2]

M.Sc. Biostatistics (PU), Department of Cardiology King Edward Medical University / Mayo Hospital, Lahore – Pakistan

Introduction

It has become an established fact today, that role of statistics in scientific research process is inevitable. Appliance of statistics is the most powerful tool to investigate scientific verity and interpret convoluted explanations and justifications in medical research findings.[1-3] As a variety of statistical methods are employed to inspect and interpret undisclosed information in medical sciences, it is important for these statistical methods to be valid and correctly undertaken[4]. A good research focuses on all components of statistical procedures including data collection, selection and suitable implication of statistical tests, correct analysis and reporting.[5-7] According to medical literature, most often researches rely on descriptive statistics for comprehensive presentation and understanding of data. These descriptive statistics include frequency, percentages,

Mean ± S.E / S.D, etc. Moreover, inferential statistics like t-tests, chi-square tests, and regression and correlation analysis are also widely used in many researches.[8-10] Though wide usage of statistics is crucial and should be practiced to achieve valid results. Latest advancements provoke the need of using latest procedures that are compatible with this era of classic techniques.[11] Unfortunately, statistics is not given the worth it deserves that has led to a range of statistical errors and shortcomings in articles published in medical journels.[12-14] Various clinical reviews have been consistently finding many statistical flaws and reported that nearly 50% of medical articles contain statistical errors and most of them are in selection of wrong statistical and / or sampling methods.[15-16] The misuse of statistics has been discussed extensively and this fact has been made clear that inappropriate usage is both unethical and constitutes serious clinical hazards.[17] The topmost reason for such statistical flaws in Pakistan is deficient knowledge among professionals, slipshod practices of statistics and incompetent teaching of statistics in health sector.[18] Hence, the prime requirement is to point out such errors and make efforts to eradicate them as these may give incorrect conclusions and cause wastage of valuable resources. That ultimately may result in unspoken yet detrimental consequences.[19-20]

Objectives

The objective of this study was to assess the quality of statistical procedures used, the appropriateness of study design, sampling criteria /technique, relevant statistical analysis and interpretation.

Methodology

We reviewed 80 research articles published in indexed and recognized local journals of Pakistan. The selection of those research articles was random and those were easily and freely available online. We searched these articles by using common search Google and Bing by typing the key words, medical research, use of biostatistics in medicine, and health improvement in Pakistan and statistical errors in medical journals of Pakistan. Case series and case reports were excluded. The objective assessment was made by the investigators their self. The criteria of assessment the statistical errors was adopted from Curran – Everet D., et al (2004), Strasak, A., et al (2007) and Lang, T., et al (2003).[6,15,24]

Results

In this manuscript we assessed 80 online research articles and investigated them according to the defined criteria. Among, 80 articles, in twenty one (26.25%) articles, the statistical methods used were not reported. descriptive statistics was only used in 32 (28.75%), inferential statistics along with the descriptive statistics was used in 33 (41.25%) articles. There was no article with inferential statistics only. In 33 articles which were loaded with inferential statistics, t-test was used in 11 (13.75%) articles, contingency tables with chi-square and fisher test were used in 24 (30%) articles, analysis of variance tests were seen in 6 (7.5%) articles and there were 3 (3.75%) articles in which non-parametric tests were applied. Correlation coefficient was seen in 6 (7.5%) articles, and 8 (10%) articles were based on regression analysis, in which basic, simple linear regression was in only one (1.25%) article and 7 (8.75%) articles were having logistic type regression. There were only 8 (10%) articles in this research, with the application of epidemiologic methods. Survival analysis was applied in not a single article while in only 12 (15%) articles confidence interval was given. Repeated measurement analysis of variance was seen in 3 (3.75%) articles and complexity of statistical analysis (multiple statistical analyses, more than one analysis) was seen in 11 (13.75%) articles. Statistical errors, flaws and deficiencies related to the design of a study and statistical analysis were assessed as mentioned in the methodology. Design of a study was not mentioned in 42 (52.5%) articles, in 74 (92.5%) articles there was no criteria for sample size estimation and power calculation. Only six studies mentioned the appropriate formula for sample size calculation. Sampling selection criteria was not given in 60 (75%) articles. Wrong statistical analysis was done in 23 (28.75%) articles, in which 16 (20%) were inappropriate due to data examined, 3 (3.75%) were due to sample selection, 1 (1.25%) was not compatible with study design, and 2 (2.5%) were due to inappropriate parametric test.

There were only 39 (48.75%)articles, in which data collection technique was defined, in 17 (21.25%) articles statistical techniques was given but not used, and in 13 (16.25%) articles the cut point (for the significance) of *p-value* was not given, and these articles were not compared with any specific *p-values.* Wrong name of statistical test was found in 10 (12.5%) articles, in 56 (70%) articles no statistical package was mentioned and overall inappropriate interpretation

Table 1: Types and frequencies of statistical methods.

Types and Frequencies of Statistical Methods	f (%) n = 80
No statistical methods	21 (26.25%)
Descriptive statistics only	32 (28.75%)
Inferential methods with descriptive	33 (41.25%)
Inferential only	0 (0%)
Contingency table analysis	24 (30%)
t-tests	11 (13.75%)
Basic Chi-square, Fisher's Test	24 (30%)
Non-Parametric tests	3 (3.75%)
Analysis of Variance	6 (7.5%)
Basic One way ANOVA	6 (7.5%)
Advanced	0 (0%)
Correlation coefficient	6 (7.5%)
Regression	8 (10%)
Basic (simple linear regression)	1 (1.25%)
Advanced (logistic)	7 (8.75%)
Epidemiological Methods	8 (10%)
Survival Analysis	0 (0%)
Reputed Measurement ANOVA	3 (3.75%)
Confidence interval	12 (15%)
Complexity of Statistical analysis (Multiple test)	11 (13.75%)
No. of different inferential methods	30 (37.5%)
Only 1 method	19 (23.75%)
2 0r 3 methods	3 (3.75%)
4 or 5 methods	6 (7.5%)
More than 5	2 (2.5%)

were made in 11 (13.75%) articles. According to other errors, in 43 (53.7%) articles only, the variables were defined in measurable terms, 43 (53.7%) articles gave the reference for transforming (or categorizing the data). Additionally, it was seen that in 13 (16.25%) articles mean and S.D were used for non-continuous data and the use of S.E was seen only in 4 articles. Errors in interpreting the p-values were present in 26 (32.5%) articles, only p-value (without statistical test value) was given in 38 (47.5%) articles. Out of 80 articles evaluated, 57 (71.2%) articles used inferential statistics but did not mention whether the assumptions for statistical test used, met or not.

Discussion

This study was aimed to point out most common and treacherous statistical errors present in articles published in local medical journals. In many large – scale studies conducted previously, it has been demonstrated that many studies constitute incompatible and inappropriate statistical procedures that may lead to deceivable and false conclusions.[21] Nearly 50% of the articles use statistical methods with selection or application faults. Though many guidelines and suggestions have been put forth by reviewers to enhance the quality of these procedures yet the widespread and common statistical mistakes have made it difficult for peer reviewers to control the situation.[22] In order to exterminate such errors from scientific research, it is preferable to have review by professional statisticians.[23] A total of 171 articles were reviewed in 2007 by Hellems, M.A.[24] He reported that in 171 articles only 1 article used no statistical procedure. The frequency of articles that owned only descriptive statistics decreased from 23% to 10% from 1982 to 2005. In our study, in 28.75% articles only descriptive statistics was used. This frequency is more close to the percentage that was observed in 1982. With advent of more comprehensive techniques, usage of inferential statistics has become imperative. But unfortunately, this is not well understood yet in our country. As recently shown, there were only 18% articles in 2005 used only descriptive statistics or no statistics. Inferential statistics like t-test and chi-square or correlation coefficient increased to 65% in 1982. Moreover, he explained the mean number of inferential statistics increased from 2.5% in 1982 to 3.9 in 2005.[23] In our study, t-test was used in 11 (13.75%) articles, contingency tables with chi-square and fisher test were used in 24 (30%) articles, analysis of variance tests were seen in 6 (7.5%) articles Correlation coefficient was seen in 6 (7.5%) articles, and 8 (10%) articles were based on regression analysis. This indicates that trend of inferential statistics is not as much prevalent as it should be, because of its power of drawing inferences in more ample manner. Other reasons for this increment are due to the development of new study designs, need of rapid investigations and statistical procedures.[1,3,23] Furthermore errors in randomization, improper sample

Table 2: Statistical Errors, flaws and deficiencies related to the design of a study and statistical analysis.

Categories	f (%) n = 80
Design of study not given	42 (52.5%)
No Sample size calculation/ power calculation (overall)	74 (92.5%)
Sampling Selection criteria no given	60 (75%)
Use of wrong statistical test/ analysis	23 (28.75%)
In compatibility of statistical test with type of data examined	16 (20%)
In compatibility of statistical test with sample selection	3 (3.75%)
In compatibility of statistical test with study designed	1 (1.25%)
Inappropriate use of parametric test	1 (1.25%)
Data analysis technique defined	39 (48.75%)
Statistical technique defined but not used	17 (21.25%)
Statistical technique used but not defined	8 (10%)
p-value cut point defined	13 (16.25%)
Wrong name of statistical test	10 (12.5%)
No statistical package defined with version	56 (70%)
Over all inappropriate interpretation	11 (13.75%)

Table 3: Other errors.

Other Errors	f (%) n = 80
Defining Each Variable in Measurable Terms	43 (53.7%)
Providing the Level of Measurement of Each Variable	23 (28.7%)
Use reference for categorizing the data	43 (53.7%)
Using the mean and standard deviation for non-continuous data	13 (16.25%)
Using the Standard Error of the Mean (SEM) as a descriptive statistic	3 (10%)
Errors in interpreting probability (p) values, not clear	26 (32.5%)
Reporting only *p* values for results	38 (47.5%)
Not Confirming That the Assumptions of Statistical Tests Were Met	57 (71.2%)

size, reporting flaws and other methodological errors have also been reported in some studies.[24] In this particular study, in 21.25% articles statistical techniques was given but not used, wrong statistical analysis was done in 28.75% articles, only six studies mentioned the appropriate formula for sample size calculation and wrong name of statistical test was given in 12.5% articles. On top of that, in 70% articles no statistical package was mentioned and overall inappropriate interpretation were made in 13.75% articles. This shows the extent of slapdash reporting and analysis procedures in Pakistan. With the passage of time and con-

tinuously increasing advancements in improving statistical technique and enhancing their reliability blend with the appropriate usage to result the accurate findings. For this purpose, Human research committees should step on and enforce that an investigator defines and authenticates an appropriate strategy for data analysis involving some professionals before approving their protocols or proposals. In addition to that some principles should be defined and their implication should be enforced to follow the prearranged guidelines.[25]

Conclusion

In order to endorse medical research with authentication and reliability, the role of biostatistician needs to be divulged. Thus biostatistician should be taken in editorial boards or articles should be refereed by the biostatisticians or statisticians for approval in order to keep the standards of research upgraded. It will help the readers to trust the presented research and believe on the inferences made. Also, it will ensure proper utilization of resources and further anticipations could be made in the field of medicine.

References

1. Altman DG. Statistics in medical journals. Statistics in Medicine 1982; 1: 59-71.
2. Altman DG. Statistics in medical journals: Developments in the 1980s. Statistics in Medicine 1991; 10: 1897-913.
3. Altman DG. Statistics in medical journals: Some recent trends. Statistics in Medicine 2000; 19: 3275-89.
4. White SJ. Statistical errors in papers in the British Journal of Psychiatry. The British Journal of Psychiatry 1979; 135: 336-42.
5. Evans M. Presentation of manuscripts for publication in the British Journal of Surgery. British Journal of Surgery 1989; 76: 1311-4.
6. Lang T, Secic M. How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers. Philadelphia (PA): American College of Physicians; 1997.
7. DerSimonian R, Charette LJ, McPeek B, et al. Reporting on methods in clinical trails. N Engl J Med. 1982; 306: 1332-7.
8. Reed III JF, Salen P, Bagher P. Methodological and statistical techniques: what do residents really need to know about statistics? Journal of Medical Systems 2003: 27: 233-8.
9. Hand D, Sham P. Improving the Quality of Statistics in Psychiatric Research. British Journal of Psychiatry 1995; 167: 689-91.
10. Dar R, Serlin R, Omer H. Misuse of Statistical Tests in three Decades of Psychotherapy Research. J Consult Clin Psychol. 1994; 62: 75-82.
11. Huang W, LaBerge JM, Lu Y, et al. Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods. Journal of Vascular and Interventional Radiology 2002; 13: 247-55.
12. Cooper RJ, Schriger DL, Close RJH. Graphical literacy: the quality of graphs in a large – circulation journal. Annals of Emerg Med. 2002; 40: 317-22.
13. García – Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Method. 2004; 4: 13-7.
14. Porter AM. Misuse of correlation and regression in three medical journals. J Roy Soc Med. 1999; 92: 123-8.
15. Curran – Everett D, Benos J. Guidelines for reporting statistics in journals published by the American physiological association. American Journal of Physiology and cell physiology 2004; 287: 243-5.
16. Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation 1980; 61: 1-7.
17. Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools and a research agenda. Account Res. 2002; 9: 65-74.
18. Hanif A, Butt A, Awais SM. Importance and understanding of Bio-statistics among post graduate students at King Edward Medical University Lahore, Pakistan. ANNALS of KEMU 2009; 15 (3): 107-10.
19. Murray GD. Statistical guidelines for the British Journal of Surgery. British Journal of Surgery 1991; 78: 782–4.
20. Altman DG, Schulz KF, Moher D, et al. The CONSORT statement: revised recommendations for improving the quality of parallel – group randomized trials. Annals of Internal Medicine. 2001; 134: 657-62.
21. Reznick R, Dawson-Saunders E, Folse R. A rationale for the teaching of statistics to surgical residents. Surgery 1987; 101: 611-7.
22. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: areas for improvement by authors. Lancet 1992; 340: 100-2.
23. Altman DG, Goodman SN, Schroter S. How Statistical Expertise is used in Medical Research. JAMA 2002; 287: 2817-20.
24. Strasak AM, Zaman Q, Pfeiffer KP, et al. Statistical errors in medical research – a review of common pitfalls. Swiss medicine weekly 2007; 137: 44-9.
25. Altman DG. The scandal of poor medical research. British Medical Journal 1994; 308: 283-4.