



Review Article

## Statistical Aspect in Medical and Paramedical Research Articles

Balkishan Sharma

Associate Professor (Biostatistics), Department of Community Medicine, Sri Aurobindo Medical College & P. G. Institute, Indore (M. P.), India

Received: 15/07/2014

Revised: 02/08/2014

Accepted: 24/09/2014

### ABSTRACT

**Background:** The purpose of this study was to introduce the appropriate application of statistical concept and methodology in medical and paramedical science articles. Author represented statistical methods that have been utilized in the literature.

**Materials and Methods:** The concepts of descriptive, inferential, basic and advance statistics, p-values, clinical and statistical significance are discussed with examples to project their application to the interpretation of various medical and paramedical researches. In addition, descriptions of Student's t test, Chi-square test, Fisher's exact test, Logistic regression and Cox regression are presented. These techniques are described with adequate detail to allow a reader who has access to the original data to verify the reported results.

**Results:** Statistical literacy required for scientific evolution of the research. Statistical methods are increasingly a necessary and an inseparable part of medical and paramedical research. Due to wide application of statistical tools in medical and paramedical science, statistics may be rightly called as biostatistics which represents a critical methodological skill for clinicians. Bio-statistical training has enabled academic investigators to contribute significantly to the body of statistical research literature. Researchers may have a potential to use significantly the descriptive statistics after collateral reading of this article while inferential statistics to draw valid inferences. Student's t-test, Pearson's Chi-square/Fisher's Exact test will be comprehended to read and interpret by an investigator.

**Conclusion:** This article offers an understanding of application of statistical tools and basic guidelines in reporting the statistical aspects of the results of their research studies accurately and scientifically.

**Key Words:** Biostatistics, Statistical significance, P-Value, Health Statistics, Statistical methods, parametric test, nonparametric test

### INTRODUCTION

Biostatisticians and epidemiologists play a vital role when more complex statistical methods are used in the analysis.

[1] The increasing sophistication of statistical methods in journals, the reader must continue to expand his statistical

understanding. [2] An understanding of statistical concepts and methods is essential for the clinician who wishes to interpret the results of clinical studies. [3] Published original medical research articles now nearly always include statistics, and the complexity of the statistical methods has increased over

time. <sup>[4-5]</sup> Appropriate utilization of biostatistical methods is becoming increasingly important in biomedical research. More research may be required to develop or refine an assessment instrument more appropriate for statistical methods in evidence based medicine. <sup>[6]</sup> Various journals have dedicated statistical committee that scrutinizes the methods used in analyzing data. In the last decade, several papers addressing statistical analysis approaches and study design issues in different clinical fields have been published underpinning the importance of robustness in methodology. <sup>[7-11]</sup>

Statistical methods allow studying diseases, patients, and epidemiological events and a deeper understanding of statistics is required to avoid colossal mistakes of misleading. <sup>[12]</sup> Despite its importance, a recent study showed that medical residents lack the knowledge to understand the most common statistics found in clinical journals. <sup>[13]</sup> This suggests that the level of bio-statistical education in India adopted for graduate and under graduate medical and paramedical may not be enough to adequately comprehend the broad range of statistics.

Used statistical tests should be described and if it is a confirmatory analysis or if it is an exploratory test that needs to be strengthened with cross validation. Reporting results, any exclusion during analysis should be expressed. Data analysis should be reported with appropriate confidence intervals or P value. <sup>[14-17]</sup> It is strongly recommended that all authors consult a biostatistician when designing the study, to ensure that the appropriate statistical methods are chosen. <sup>[18]</sup>

Various studies showed a significant use of an appropriate statistical method has been associated with more chance of publication in prestigious journals. The author hypothesized that help from

biostatistician will be more fruitful for those who are willing to conduct a research especially in the field of medicine. Therefore, Medical and Para-medicals have been taught methods for reasoning under uncertainty and the use of sensitivity, specificity and other diagnostic statistics. Overall, they needed to understand the intricacies of the statistical methods. In this article, the author aimed to identify the significance of statistical concepts and methods in medical and paramedical research articles.

## **MATERIALS AND METHODS**

This section covers statistical methods. The methodological frame work consists of four factors: (1) Study design; (2) population and sample size; (3) data collection and (4) classification of statistical methods. Health research is essential in developing evidence-based interventions that will make a difference in mitigating health problems, promoting health and ultimately improving the quality of life. Data collected and compiled from experimental work, records and surveys should be accurate and complete. They must be checked for accuracy and adequacy before processing further. So far they lie in masses, are scattered in the records and in other words they are mixed and unsorted.

Project the main features of the research design accurately in detail (experimental/observational /others) and specify the outcome variables (independent/dependent). The design of the research should be described in such a way that the study can be reproduced, if required. Illustrate the target population of the study and how the sample was selected (randomly/stratified/clustery/any other), the inclusion and exclusion criteria, the assignment mechanism to different treatments and any blinding techniques (single/double/triple) used. Describe the

expected sample size and its calculation clearly with the outcome variables. Justify the choice of level of significance and power upon which the sample size was based. Illustrate the method of data collection and provide pertinent details. Provide the actual number of observations and report losses to observation like dropouts from a clinical trial.

In medical and paramedical field of research, statistical techniques are provided parametric and non-parametric methods which further classified as either basic or advanced. The details of basic and advanced statistical methods presented in result section.

## RESULT

This section covers analysis of data. A set of mathematical procedures for describing, synthesizing, analyzing, and interpreting quantitative data is termed as statistics and the selection of an appropriate statistical technique is determined by the research design, hypothesis, and the data collected. Research on used statistical techniques in medical and paramedical research identified with a wider range of statistics. A research design and choosing of appropriate statistical techniques are the two vital factors with varying levels of complexity. Choosing the appropriate study design with relevant statistical analysis technique is largely dependent on the complexity of the aim and objectives of the proposed research.

Data analysis converts data into information which becomes knowledge and explores the relationship between variables. Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. According to Hamoo and Resnik (2003) various analytic procedures “provide a way of drawing inductive inferences from data

and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data”.

[19]  
Understanding of the data analysis procedures will enable researchers to appreciate the meaning of the scientific method which includes testing of hypotheses and statistical significance in relation to research questions. There are a number of issues that researchers should be cognizant of with respect to data analysis. Some of the key considerations in analysis and selection of the right test of significance are as follows:

- ✓ Having the necessary skills to analyze.
- ✓ Distinguishing data types.
- ✓ Distinguishing different types of statistical tests.
- ✓ Identify the selection of a right test.
- ✓ Determining statistical significance.
- ✓ Distinguishing between Parametric and Non-Parametric test with their applying criteria.
- ✓ Distinguishing between Correlation and Regression.
- ✓ Drawing unbiased inference.
- ✓ Inappropriate subgroup analysis.
- ✓ Lack of clearly defined and objective outcome measurements.
- ✓ Reliability and validity and extent of analysis.

The two common type of analysis are exploratory (descriptive) and confirmatory (inferential). Exploratory data analysis involves examination of the data errors and describing data using summary statistics and graphical techniques while confirmatory statistics have used to verify the significance of hypothesis. The common exploratory statistics and valuable precautions while presenting research contents are described next.

The mean is most commonly usable measure of central tendency in medical and

paramedical research while median is less common. The standard deviation is the most suitable choice among all measures of variation as it indicates the extent to which the individual values fall away from the average or the central value accurately. Correlation quantifies the strength of the linear relationship between a pair of variables but when aim is forecasting use regression.

Here are some suggestions to present statistics accurately, for example:

- (i) The mean age of medicos only needs to be presented to second place of decimal, not four (e.g. 28.75 years, not 28.7524 years).
- (ii) Consider summary statistics up to second place of decimal with

depiction of spread of (mean±standard deviation) parametric findings.

- (iii) The nearest whole percent (e.g. 25%) is usually sufficient to present prevalence/ percentage but some journals prefer percentages to the nearest tenth of a percent (e.g. 25.4%).
- (iv) Show probability value to two places of decimal for applied test of significance such as  $\chi^2$ , t or F test, for accuracy, e.g. t-test =2.56.
- (v) Present p-values for all applied test with 95% confidence intervals of mean.
- (vi) Do not duplicate illustrations.

**Table 1: Common statistical tests**

Outcome Variable	Type of observations		Assumptions
	Independent	Correlated	
Continuous (e.g. Hemoglobin, age, height, pain score)	t-test ANOVA Linear correlation Linear regression	Paired t-test Repeated-measures ANOVA Mixed models/GEE modeling	Outcome is normally distributed especially for small samples. Outcome and predictor have a linear relationship.
Binary or categorical (e.g. hypertension yes/no)	Chi-square test Relative risks Logistic regression	McNemar's test Conditional logistic regression GEE modeling	Chi-square test assumes sufficient numbers in each cell ( $\geq 5$ )
Time-to-event (e.g. time-to-death, time-to-fracture)	Kaplan-Meier statistics Cox regression	n/a	Cox regression assumes proportional hazards between groups

[SD-Standard deviation; ANOVA-Analysis of Variance; GEE-Generalized Estimating Equations; LR-Linear Regression]

**Table 2: Significant figures with decision**

Null Hypothesis	Probability value (p-value)	
	Lies between/or	Decision
Accepted	p>0.05	Insignificant/Not Significant
Rejected	0.05<p<0.10	Suggestively/Poorly Significant
Rejected	0.02<p<0.05	Moderately Significant
Rejected	0.01<p<0.001	Highly/Strongly Significant

Table 1 showed common statistical tests. Statistical methods used as either basic or advanced. Methods classified as basic included Student t-test, Chi-Square and Fishers Exact test, Mann-Whitney, Kruskal-Wallis, Wilcoxon, simple one-way ANOVA and correlation statistics. Use normal distribution for large sample but

significance of difference between means of small sample can be observed by student's t-test (Parametric test) and if the observations are correlated than use paired t-test otherwise use independent sample test or unpaired t-test. When to compare more than two groups use analysis of variance (ANOVA) or F-test. Chi-Square and Fishers

Exact test (Non-parametric test) is used to observe the significance of an association between two attributes/parameters/variables if samples are drawn randomly and size is sufficiently large but Mann-Whitney, Kruskal-Wallis and Wilcoxon test may be taken into account if the sample size is small and assumption of normality is violated.

Modeling approaches such as Logistic Regression, Conditional Logistic Regression, Cox regression, time-varying Cox-regression and Cox regression with frailty and epidemiologic statistics were classified as advanced. Logistic regression is useful to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. [20] Cox DR (1979) observed that Cox regression analysis is a survival analysis technique in time-to-event data that incorporates follow-up time and fixed covariates. [21] The method assumes risk of an event is homogeneous. Extensions of the Cox regression exist which include time dependent Cox regression and Cox regression analysis with frailty described by Lin DY (1994). [22] Time dependent Cox regression analysis accounts for the inherent correlation that may exist when covariates change over time.

Table 2 showed the criteria either for accepting or rejecting the assumption. Scientifically, risk for rejecting the null hypothesis is fixed at 5% level of significance but in medical research when samples are very rare like cardiovascular event in younger patient, the criteria may be lowered.

It is worth advisable that exact p-values for all main analyses should be reported instead of a simple statement ( $p < 0.05$ ) which is insufficient. For example, the exact probability-values such as  $p = 0.85$  are preferable to the term insignificant or not significant. Present the results of t-tests/other tests in detail by providing the confidence interval such as “the difference

in mean level of Hs-CRP between healthy controls and coronary heart disease (CHD) patients was 0.77 mg/L with a 95% confidence interval from 0.2 mg/L to 3.2 mg/L. The probability value for t-test was 3.2 with 198 degrees of freedom and reported  $p = 0.001$ ”.

Lastly, it was noticed that the clinicians making a common attitude that if  $p < 0.05$ , the results are worth publishing as they are statistically significant. Therefore, distinguishing between clinical and statistical significance is necessary. A “statistically significant” result is one in which investigators are 95% confident that a relation/correlation/association exists. However, clinical significance may be supported with statistical significance to show the chance and the degree of bonding for cause-effect relationship.

## DISCUSSION

This article offers an in-depth understanding with use of an appropriate statistical method to approve the significance of scientific evolution of research in medical and paramedical science. However, the decisive instrumental (i.e., applied) knowledge is expressed together with some statistical degree of confidence. The authors, [1] Becker et al. (1995) appealed for statistical excellence actually applies to the researcher, to the clinician journal reader, and to the editor. Significantly applied statistical tools to a research study may be providing an optimal chance of publication. The concepts of descriptive statistics and inferential statistics are discussed with examples to illustrate their application to the interpretation of clinical trials observed by [2] Pace NL (1986).

A statistical calculation evidenced how far a given correlation/relation/association exceeds that would be expected by chance termed as statistical significance

and statistically significant result does not necessarily prove a cause-effect relationship until and unless the results are clinically significant. The use of statistical studies allows to evaluate the causality and to draw a conclusion about the effect of an independent variable on the dependent variables examined. <sup>[4]</sup> Reed III JF (2003) highlighted that physician needed educational efforts should focus on appropriate study design and analysis while <sup>[5]</sup> Hellems MA (2007) reported that there was no validated knowledge and skills assessment for graduate level biostatistics for academic medical researchers.

Working with expert in statistics requires some knowledge of appropriate methods, sufficient common language for collaboration, and the ability to interpret statistical results. Recently, <sup>[11]</sup> Romano R, Gambale E (2013) identified that an observational study evaluates the inferences about the effect of a treatment on patients, when the option of giving a patient to a treated group versus a control group is not an option for the investigator. The accuracy of the statistical methods used and the way in which they are presented in the scientific paper, affect the integrity of the study reported by <sup>[18]</sup> Ng KH Peh WCG (2009). Therefore, statistical judgment must always be exercised. The general assumption was noticed among investigators that a non-significant result necessarily prove the null hypothesis which is not true. <sup>[20]</sup> Sharma BK and Jain RK (2013) revealed with the fact that clinicians must have to develop an ability to understand the intricacies of the statistical methods for their alignment with core competencies necessary for successful use of statistics.

Descriptive statistics has been used to summarize the features and the characteristics of the data while inferential statistics has been used to draw inferences about the population. Currently, in an article

<sup>[23]</sup> Otwombe KN et al. (2014) identified that descriptive statistics and survival analysis techniques remain the most common methods of analysis in publications on predictors of all-cause mortality in HIV-infected cohorts while on Canadian community health survey based study the authors <sup>[24]</sup> Yergens DW et al. (2014) showed that descriptive and regression statistical methods increased at a faster rate than elementary statistics, and the distance between elementary statistics and the rest grew over time. On the other hand, <sup>[25]</sup> Scotch M et al. (2010) determined that as core competencies and credentialing in biomedical informatics are developed, scientists in this field should have, as a minimum, proficiency in descriptive and elementary statistics.

An understanding of statistical concepts and methods is essential for the clinician who wishes to interpret the results of clinical studies agreed by <sup>[26]</sup> Lewis RJ and Bessen HA (1991) while increasing attention has been given to teaching and using statistics in medical education across the continuum of lifelong learning observed by <sup>[27]</sup> Rao G and Kanter SL (2010). <sup>[28]</sup> Gou L et al. (2005) reported that it is the key of choosing rational statistical methods to distinguish the type of design and variable. The scientific quality of psychiatry research and submitted reports could be greatly improved if researchers became sensitive to, or sought consultation on frequently encountered methodological and analytic issues determined by <sup>[29]</sup> Harris AH et al. (2009). Therefore, bio-statistical competencies for medical and paramedical researchers need to be enhanced.

An integrated approach by undertaking the statistical concept and technique in-depth would ensure the scientific dissemination of empirical findings to patients, health-care providers and regulatory agencies, and a more

effective protection of public safety and promotion of medical and paramedical research. Lastly, author do hope that this article will provide clinicians with hands-on experience to promote the use of statistical thinking and techniques to apply them to make educated decisions whenever there is variation in data and contribute to maintain the standard of statistical work reported in medical and paramedical publications.

## CONCLUSION

This article dedicated to an understanding of application of statistical tools and basic guidelines in reporting the statistical aspects of the results of their research studies accurately and scientifically.

The main objective of an application of proper statistical methods in any scientific research is to provide a factual account of the empirical findings. Provided guidelines in the article may contribute to an improvement in the standard of statistical work reported in medical and paramedical articles. Author do hope that this article will enable clinicians to enhance the statistical skills and experience essential to design medical and paramedical studies, carry out worth statistical analyzes and disseminate the findings in a clear and objective manner.

### ***Implication of the study:***

This article provide guidelines to statistical aspect and methods in medical and paramedical research that motivates for more training in the use of advanced and basic statistical methods. Collateral reading of the article will be fruitful for researchers to improve the standard of carry out statistical work includes in medical and paramedical articles and therefore awareness among professionals regarding the improvement of statistical components in research may be created.

## REFERENCES

1. [Becker PJ, Viljoen E, Wolmarans L, IJsselmuiden CB: An assessment of the statistical procedures used in original papers published in the SAMJ during 1992. S Afr Med J. 1995; 85(9):881-884.
2. Pace NL: Learning statistical methods. Int J Clin Monit Comput. 1986;3(1):3-10.
3. Lewis RJ, Bessen HA: Statistical concepts and methods for the reader of clinical studies in emergency medicine..J Emerg Med. 1991;9(4):221-232.
4. Reed III JF, Salen P, Bagher P: Methodological and statistical techniques: what do residents really need to know about statistics? J Med Syst. 2003;27(3):233-238.
5. Hellems MA, Gurka MJ, Hayden GF. Statistical literacy for readers of Pediatrics: a moving target. Pediatrics. 2007;119(6):1083-1088.
6. Enders F: Evaluating Mastery of Biostatistics for Medical Researchers: Need for a new assessment tool. Clin Transl Sci. 2011;4(6):448-454.
7. Wu S, Jin Z, Wei X, Gao Q, Lu J, Ma X, Wu C, He Q, Wu M, Wang R, Xu J, He J: Misuse of statistical methods in 10 leading Chinese medical journals in 1998 and 2008. ScientificWorldJournal 2011:2106-2114.
8. Al-Benna S, Al-Ajam Y, Way B, Steintraesser L: Descriptive and inferential statistical methods used in burns research. Burns. 2010;36:343-346.
9. Harris AHS, Reeder R, Hyun JK: Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: What editors and reviewers

- want authors to know. *J Psychiatr Res.* 2009;43:1231-1234.
10. Strasak AM, Zaman Q, Marinell G, Pfeiffer PK, Ulmer H: The use of statistics in medical research: A comparison of the New England Journal of Medicine and Nature Medicine. *Am Stat.* 2007;61:47-55.
  11. Taback N, Krzyzanowska MK: A survey of abstracts of high-impact clinical journals indicated most statistical methods presented are summary statistics. *J Clin Epid.* 2008; 61:277-281.
  12. Romano R, Gambale E: Statistics and Medicine: the Indispensable Know-How of the Researcher. *Transl Med UniSa.* 2013;5:28-31.
  13. Windish DM, Huot SJ, Green ML: Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* 2007;298:1010-1022.
  14. Fukuda H, Ohashi Y: A Guideline for Reporting Results of Statistical Analysis in Japanese Journal of Clinical Oncology. *Jpn J Clin Oncol.* 1997;27(3):121-127.
  15. Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical Guidelines for Contributors to Medical Journals. *BMJ.* 1983;286:1489-1493.
  16. Daniel WW: Biostatistics: a Foundation for Analysis in the Health Sciences. John Wiley & Sons Inc; 1995.
  17. Bailar JC, III, Mosteller F: Guidelines for Statistical Reporting in Articles for Medical Journals. *Ann Intern Med.* 1988;108:266-273.
  18. Ng KH, Peh WCG: Presenting the statistical results. *Singapore Med J.* 2009; 50(1):11-14
  19. Shamoo, AE, Resnik, BR: Responsible Conduct of Research. 2003;Oxford University Press.
  20. Sharma BK and Jain RK: Right choice of a method for determination of cut-off values: A statistical tool for a diagnostic test. *Asian J Med Sci.* 2014;5(3):21-30.
  21. Cox DR: Regression models and life-tables. *Journal of the Royal Statistical Society Series B* 1972;34:187-220.
  22. Lin DY: Cox regression analysis of multivariate failure time data: the marginal approach. *Stat Med.* 1994;13:2233-2247.
  23. Otworld KN, Petzold M, Martinson N, Chirwa T: A review of the study designs and statistical methods used in the determination of predictors of all-cause mortality in HIV-infected cohorts: 2002-2011. *PLoS One* 2014;9(2):e87356.
  24. Yergens DW, Dutton DJ, Patten SB: An overview of the statistical methods reported by studies using the Canadian community health survey. *BMC Med Res Methodol.* 2014;14:15
  25. Scotch M, Duggal M, Brandt C, Lin,Z, Shiffman R :Use of statistical analysis in the biomedical informatics literature. *J Am Med Inform Assoc.* 2010;17(1):3-5.
  26. Lewis RJ, Bessen HA: Statistical concepts and methods for the reader of clinical studies in emergency medicine. *J Emerg Med.* 1991; 9(4):221-232.
  27. Rao G, Kanter SL: Physician numeracy as the basis for an evidence-based medicine curriculum. *Acad Med.* 2010;85:1794-1799
  28. Gou L, Yu J, Yang Z: Appropriate application of statistical analysis method in medical science and technology articles. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi.* 2005;19(11):925-927.



29. Harris AH, Reeder R, Hyun JK:  
Common statistical and research  
design problems in manuscripts  
submitted to high-impact psychiatry

journals: what editors and reviewers  
want authors to know. J Psychiatr  
Res. 2009;43(15):1231-1234.

How to cite this article: Sharma B. Statistical aspect in medical and paramedical research articles. Int J Health Sci Res. 2014;4(10):268-276.

\*\*\*\*\*

**International Journal of Health Sciences & Research (IJHSR)**

**Publish your work in this journal**

The International Journal of Health Sciences & Research is a multidisciplinary indexed open access double-blind peer-reviewed international journal that publishes original research articles from all areas of health sciences and allied branches. This monthly journal is characterised by rapid publication of reviews, original research and case reports across all the fields of health sciences. The details of journal are available on its official website ([www.ijhsr.org](http://www.ijhsr.org)).

Submit your manuscript by email: [editor.ijhsr@gmail.com](mailto:editor.ijhsr@gmail.com) OR [editor.ijhsr@yahoo.com](mailto:editor.ijhsr@yahoo.com)