Original Article

USE OF STATISTICAL METHODS AND COMPLEXITY OF DATA ANALYSIS IN RECENT RESEARCH PUBLICATIONS IN BASIC MEDICAL SCIENCES

Swati Patel[1], V D Naik[2], Prakash Patel[3]

Financial Support: None declared

Conflict of interest: None declared

How to cite this article:
Patel S, Naik VD, Patel P. Use of Statistical Methods and Complexity of Data Analysis in Recent Research Publications in Basic Medical Sciences. Natl J Community Med 2014;5(2):253-256.

Author's Affiliation:
[1]Statistician cum Assistant Professor, Department of community Medicine, SMIMER; [2]Principal, SPB Commerce Collage, Surat; [3] Assistant Professor, Department of Community Medicine, SMIMER, Surat

Correspondence:
Ms. Swati Patel,
Email: swatipatel84@gmail.com

Date of Submission: 01-06-14

Date of Acceptance: 29-06-14

Date of Publication: 30-6-14

ABSTRACT

Objective: This study was carried out to evaluate the use of various statistical methods; statistical software; complexity of statistical analysis; and appropriateness of use of statistical methods in recent research publications of basic medical sciences.

Methods: Original articles published in Journals of Basic medical Sciences, i.e Anatomy, Physiology and Pharmacology subscribed by the central library of our institute (SMIMER, Surat) were taken into consideration were reviewed for statistical applications in the manuscript.

Results: Total 143 original articles were reviewed; out of them 89.51% (95% CI 84.49 - 94.53) had used Statistical techniques. Most frequently used statistical method was ANOVA (42.97%) and it's use is lower in Anatomy compared to other two subjects. One out of seven publications used either t-test (independent and paired) or contingency table (chi-square and fisher exact) or nonparametric test. Confidence intervals were estimated in approximately one fifth of all research papers. Use of statistical software was much more in Physiology and Pharmacology compare to Anatomy. Basic methods of statistics were used 63 times while "modern Analysis" techniques were used 88 times. Only 10% (15 out of 143) research papers had mention the study design; 7% (10 out of 143) research papers had mentioned appropriate formula for calculation of sample size; 1.4% (2 out 143) had defined sampling technique. In 7.03% (9 out of 128) papers statistical measures were presented inappropriately.

Conclusion: Statistics methods are used widely in basic biomedical research. However, conceptual understanding of methodology and biostatistics was lacking.

Keywords: Bio-statistics, basic medical sciences, research publications, statistical test

INTRODUCTION

Pierre-Charles-Alexander Louis, WilliaFarr, and John Snow were the first who introduced and applied the statistical methods in medical research.(1) Over the time there has been an immense increase in the use of statistics in bio medical research. Though the use of simple statistical tests like t –test and chi- square test is very common, use of more advance and complex statistical tests are also on increase, especially after the availability of sophisticated statistical software.

Application of bio-statics give meaning to raw data generated during the research studies. Results of various statistics tests help to draw valuable inference or conclusion from the observations. However, a good understanding of basic statistics is required to draw proper inference from the outcome of statistical tables. During our routine academic exercises we have observed that improper test were applied in many research articles, even interpretation was also unjustifiable. Even after having possibility of application of advance statistical test, some author limited application to common statistical tests only. All these lead to improper inference or limited inference of data, which were collected after lot of efforts and resources.

With this background, we planned to review published bio-medical research literature available in our library.

OBJECTIVE

This study was carried out to evaluate the use of various statistical methods; statistical software; complexity of statistical analysis; and appropriateness of use of

statistical methods in recent research publications of basic medical sciences.

METHODOLOGY

All the indexed journals subscribed by the central library of our institute (SMIMER, Surat) were taken into consideration. However, considering huge number of journals in our library we have limited our study to original research paper of Basic medical Sciences, i.e Anatomy, Physiology and Pharmacology only. All available issues of the selected journal published in the year 2013 were included in the study.

List of the journal and issues included in the study were as follows:

- The journal of cytology Volume 30 Issue 1 to 4
- Indian Journal of Physiology and Pharmacology Volume 57, Issue 1 to 4
- American Journal of Applied physiology Volume 114
- Indian journal of Pharmacology Volume 45 Issue 1 to 3
- Journal of Pharmacology and Pharmacotherapeutic Volume 4 Issue 4

The review was also limited to statistical part of the manuscript only. Original research articles published in above listed journal were included in the study. As very few statics were used in editorials, letters, case reports, short communications and other type of article, they were excluded.

All papers were manually reviewed by the authors. Categories of statistical methods given by Emerson and Colditz[2] were used to classify statistical methods, with certain modification. Details of different categories are given along with observation tables.

As there were no statistical method like cost benefit analysis, life table, Regression of survival and other survival analysis were used, these categories were not considered for final categorization table.

If more than one statistical method was used in a single paper all were considered separately. If the same statistical method applied more than once in the same paper, it was counted only once.

RESULTS

Total 143 original articles were reviewed; out of them 33 were from the journal of cytology; 42 were from the Indian Journal of Physiology and Pharmacology; 24 were from American Journal of Applied physiology; and 37 were from the Indian journal of Pharmacology; and 7 were from Journal of Pharmacology and Pharmacotherapeutic. Table 1 show the Frequency of Modified Categorization of statistical Analysis in all selected original research papers of basic medical science.

Table:-1 Frequency of Modified Categorization of statistical Analysis

	Anatomy	Physiology	Pharmacology	Total
Total original papers reviewed	33 (3.07)	68 (47.55)	42 (29.37)	143
No statistics or descriptive* statistics only	7(21.2)	8(11.76)	0(0)	15(10.49)
No .papers used statistical technique	26(78.7)	60(88.24)	42(100)	128(89.51)
Checked the normality	1 (3.85)	12(20)	3(7.14)	16(12.5)
One sample t-test	0 (0)	0 (0)	1(2.38)	1(0.78)
Independent t-test:-	2 (7.69)	20 (33.33)	0 (0)	22(17.19)
Paired t-test	0 (0)	12 (20)	8(19.05)	20(15.63)
Contingency table$	12(46.15)	6(10)	4(9.52)	22(17.19)
Correlation (Pearson)	2(7.69)	16(26.67)	2(4.76)	20(15.63)
Correlation (Spearman)	2(7.69)	1(1.67)	1(2.38)	4(3.13)
Nonparametric test@	4 (15.38)	13(21.67)	5 (11.90)	22(17.19)
Contingency table (kappa)	1 (3.85)	0 (0)	0(0)	1(0.78)
Regression (linear, logistic)	3(11.54)	4(1.67)	1(2.38)	8(6.25)
Multiple Regression(linear, logistic)	1(3.85)	2(7.69)	1(2.38)	4(3.13)
ANOVA(one way , two- way, Repeated measure)	2(7.69)	25(41.67)	28(66.67)	55(42.97)
Multiple comparison test(post –hoc)	2(7.69)	22(36.67)	26(61.90)	50(39.06)
ROC	0(0)	1(1.67)	0(0)	1(0.78)
Epidemiological Analysis**	7(26.92)	0(0)	0(0)	7(5.47)
Confidence interval	5(19.23)	12(20)	9(21.43)	2(1.56)
Level of significance	21(80.77)	48(80)	34(80.95)	95(74.22)
Log transformation	0(0)	1(1.67)	0(0)	1(0.78)
Others#	0(0)	5(8.33)	0(0)	5(3.91)
Use of Software##	8(30.77)	34(56.67)	25(59.52)	67(52.34)
One software	7(87.5)	29(85.29)	24(96)	60(46.88)
More than one	1(12.5)	5(14.71)	1(4)	7(5.47)

* i.e. included mean, SD, median, IQR, percentage, range; $ Chi-square, fisher-exact test, Mc nemar test; @Mann Whitney, Wilcoxen rank sum test, kruskal walish; **Odds Ratio, Relative risk, Sensitivity Analysis; # Generalised linear model, mathematica models ect; ##use different version of SPSS,STATA, Graph pad ,Sigmastat , Sigmaplot , Statistica ,Matlab, Cruncher Statistical System 2007 and Power Analysis and Sample Size 2008 Statistical Software (Utah, USA),.SAS, statmate freeware; Figure in parenthesis indicate percentage

Table 2: List of errors in application of statistical analysis

	Anatomy (n=26)	Physiology (n=60)	Pharmacology (n=42)	Total (n=128)
Didn't mention the applied test statistics #	3(11.54)	2(3.33)	1(2.38)	6(4.69)
Confusion regarding test statistics $	1(3.85)	3(5)	1(2.38)	5(3.9)
Inappropriate measure of parameter for presentation†	-	9(15)	-	9(7.03)
Inappropriate presentation of data ‡	-	2(3.33)	-	2(1.56)
Incorrect name of test statistics (i.e nonparametric independent t-test)¥	-	1(1.67)	-	1(0.78)
Didn't mention the name of software by which analyzed the data. (n*)	18(54.54)	26(38.3)	17(40.48)	61(44.65)
Didn`t mention the version of software*	5(15.15)	10(14.70)	24(57.14)	39(30.47)

n *= total reviewed original articles

#= The data were analyzed but didn`t mention the statistical method by which method they calculated the p-value for significant or non significant result, it should be mentioned.

$ = In part of methodology author confused about define the specific statistical method which should be applied i.e independent t- test or Mann Whitney , chi-square or Fisher exact, independent t-test or ANOVA.

† = SE(Standard error) calculated instead of SD(standard deviation)

‡ = i.e for ordinal data calculated mean ± SD.

¥ = Author gave the name of the applied statistical test like non parametric independent t-test.

Table:-3 Complexity of statistical data analysis

	Anatomy (n=26)	**Physiology (n=60)**	**Pharmacology (n=42)**	**Total (n=128)@**
Use of Basic statistics#	14(28.85)	36(60)	13(30.95)	63(49.22)
Use of Modern Statistics*	13(50)	40(66.66)	35(83.33)	88(68.75)
Both(basic and Advance)	7(26.92)	20(33.33)	6(14.28)	33(25.78)
Used Only one statistical method	6(23.07)	16(26.66)	9(21.43)	31(6.24)
Two statistical method	6(23.07)	24(40)	28(66.67)	57(44.53)
Three statistical method	3(11.54)	14(23.33)	2(4.76)	19(14.84)

#t-test (one sample, independent t-test, paired t-test), contingency table (chi-square, fisher), correlation (Pearson, spearman), Nonparametric test

*contingency table (kappa), regression (linear, logistic), multiple regression, ANOVA, multiple comparison test (post hoc), Epidemiological analysis, Confidence interval

@Sum of percentage is more than 100 as single publication was counted for more than once if more than one type of statistical test were used.

Out of 143, 89.51% (95% CI 84.49 - 94.53) original articles used Statistical techniques, and among them subject wise use of statistics were 78.79% (95% CI 64.84 - 92.74) for Anatomy, 88.24% (95% CI 80.58 – 95.89%) for Physiology and 100% for Pharmacology.

Most frequently used statistical method was ANOVA (including one way, two way and repeated measure ANOVA) which were used in 42.97%(95% CI 34.39 – 51.54) of research papers. It was also found that 90.91% (95% CI 83.31- 95.51) author has used Multiple comparison test (Post hoc) after getting significance result from ANOVA. Use ANOVA test was least in Anatomy compared to other two subjects. In one publication (0.78%) it was found that authors used multiple comparison test before checking significance among mean of independent groups (i.e before used the ANOVA) which is statistically incorrect way of application of statistical test.

Considering the complexity of statistical analyses, it was observed that basic methods of statistics were used 63 times while "modern Analysis" techniques were used 88 times. In 33 (25.78%) articles used the both methods (i.e Basic and Advanced).

Table -2 summarized the appropriateness of statistical analysis used in selected research papers. Only 10%

Independent t- test and Paired t- test were not used in any publication of Pharmacology and Anatomy respectively. Approximately out of one of seven publications used either t-test (independent and paired) or contingency table (chi-square and fisher exact) or nonparametric test. The pearson's correlation was used more frequently than Sperman`s correlation and the difference was 12.5%.

One sample t-test, advance contingency table (kappa), log transformation after checking the normality and ROC were used only once. Confidence intervals were estimated in approximately one fifth of all research papers. Use of statistical software was much more in Physiology and Pharmacology compare to Anatomy.

(15 out of 143) research papers had mention the study design; 7% (10 out of 143) research papers had mentioned appropriate formula for calculation of sample size; 1.4% (2 out 143) had defined sampling technique. Out of total 128, 6 papers (4.67%) didn't mention the statistical method used to calculate p-value and in

3.90% (5 out of 128) papers author seems confused in mentioning statistical tests in methodology.

In 7.03%(9 out of 128) papers statistical measures were presented inappropriately; in 0.7% (1 out of 143) papers statistical method was mentioned incorrectly; in 1.40% (2 out of 128) papers data were presented inappropriately. Out of total 143, around one third research paper didn't mention the name of software.

DISCUSSION

This study shows 89.51% original articles used statistical techniques to analyze data; remaining 10.48% articles were without any statistical analysis. This shows high use of statics in basic biomedical publication. However, statics is two sided sword, if used appropriately it can lead to good inference from data but if it is used inappropriately it can lead to false results and research end up with wastage of resources.

This study shows that many study published in basic medical science research journals (included Anatomy, Physiology and Pharmacology) without following General Principles for Reporting Statistical Methods [3].

We observed that many papers included inferential statistics. However, authors were heavily relied upon the application of well-established statistical methods only and it seems that they are avoiding using new statistical methods which are many a time more suitable than using routine statistical methods. By avoiding use of modern techniques, we may miss many possible important inferences from the same data.

However, in the long run, with the ongoing development of new statistical methods and more user friendly statistical software packages, basic medical research is likely to progress in statistical science and will discover new directions in terms of data analysis. [5]

This survey found common errors in reporting inference of statistical test, especially in complex of statistical analysis. This highlights the importance of biostatistician in reporting statistical results. These results reinforce the fact that bio-Statistician must be included in any research at all steps.

It is also important to note that a very few articles had mention study design, sample size and sampling techniques properly.

This study has some limitations like only basic medical science original research articles were considered and article published only in 2013 were analyzed. Another limitation is that only journals subscribed by our library were included.

CONCLUSION

Based on this study we conclude that statistics methods are used widely in basic biomedical research. However, conceptual understanding of methodology (study design, sample size justification and use of sampling techniques) and bio-statistics (parametric test, non–parametric test, linear regression, presentation of data and use of statistical software) was lacking. There are lots of scope of improvement in use and reporting of bio-statistics in medical research.

REFERENCES

1. Bland M. (1987) An Introduction to Medical Statistics Oxford University Press, Oxford. (Reprinted 1988, 1989, 1990, 1991, 1992, 1994, English Language Book Service edition for developing countries 1989.)
2. Altman DG: Statistical reviewing for medical journals. Statist Med 1998;17:2661-2674.
3. Lang TA, Altman DG. Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines" in: Smart P, Maisonneuve H, Polderman A (eds). Science Editors' Handbook, European Association of Science Editors, 2013
4. Altman DG, Goodman S: Transfer of technology fromstatistical journals to the biomedical literature: past rends and future predictions. J AmMed Assoc 1994; 272:129-132.
5. Qamruz Zaman et al. Statistical methods and complexity of data analysis in recent surgical research Elixir Hum. Physio. 35 (2011) 2961-2963.