

Editorial

REPORTING STATISTICAL INFORMATION IN A MANUSCRIPT PREPARED FOR PUBLICATION IN A MEDICAL JOURNAL

Manuscripts submitted to IJPP invariably have major deficits in the statistical information, especially when the papers are written by junior scientists or young researchers. Statistical analysis of data for understanding and highlighting the importance of research work is different from reporting statistical data in a manuscript submitted for publication in a medical journal. A good research paper should not only have proper statistical analysis of data, but also clear presentation of statistical data (1). Statistical information provided in a paper should be clear enough for the reviewers to understand the significance of the work from the data presented. Sometimes, papers are rejected due to paucity in its precision in statistical presentation of data. Also, data presented should be clear enough for any general medical person to understand what the data signify. Therefore, statistical information in a medical journal should be amply clear and informative with less use of uncommon terminologies. It should be remembered that many of our readers are not statisticians or experts in statistics.

Statistical Description of Data

Statistical description in a paper should include the following (2):

1. Sample size
2. Type of study
3. Drop-outs
4. The form of data depiction as mean \pm SD/SEM
5. Confidence interval
6. Level of significance (P values)
7. Correlation analysis
8. Method of analysis of data
9. Software used for analysis
10. References for uncommon methods used

Details of each of the above components are described in different sections below.

Places of Statistical Description in a Manuscript

Statistical description in a manuscript usually appears in four sections :

1. Abstract
2. Materials and Methods
3. Results
4. Discussion

Statistical mention in 'Abstract' of the manuscript

In abstract, statistical information is not described, but briefly mentioned. Sample size should be noted in the abstract. If it is a structured abstract, sample size should be mentioned in the 'Methods' component of the abstract. Values of important data (mean \pm SD / SEM) and level of

significance (P values) for major findings/ comparisons/ correlations should be highlighted in the 'Results' component of the abstract. Statistical method used for analysis should not be mentioned in abstract.

Example of statistical mention in 'Abstract'

Abstract : Chronic hypertension has been found to be associated with altered pulmonary function tests (PFTs). However, alteration in PFTs and the gender difference in PFTs in pre-hypertensive subjects have not yet been studied. Therefore, the present study was undertaken to assess the difference in change in pulmonary functions in male and female pre-hypertensives compared to their age and gender matched controls. Prehypertensive (n=30) and normotensive (n=34) subjects were recruited from Medicine OPD of JIPMER based on the blood pressure values. PFT parameters were recorded in them with the help of a computerized spirometer and comparison was made between prehypertensive and normotensive subjects and between male and female prehypertensive subjects. There was no significant difference in diastolic BP of male (85.4 \pm 4.2 mmHg) and female (84.6 \pm 3.7 mmHg) prehypertensive subjects. In female prehypertensive subjects, there was a significant decrease in FVC (P<0.001), FEV1 (P<0.001), PEF (P<0.05), FEF₂₅₋₇₅ (P<0.05) and MVV (P<0.01) compared to male prehypertensive subjects. Decrease in FVC was significantly correlated (r="0.48, p=0.02) with level of diastolic BP in female subject, but not in males. We conclude that female prehypertensive subjects are more prone to respiratory dysfunction than male prehypertensive subjects.

Statistical description in 'Materials and Methods' of manuscript

In 'materials and methods', statistical description of data is depicted at two places: first at the beginning of the 'methods', where the type of study and sample size should be mentioned, and second, usually at the end of 'methods', as separate sub-heading as 'statistical analysis of data'.

Total sample size of the study and sample size in each group should be mentioned (2). It is nice to provide how the sample size was calculated. Drop-outs are usual in human studies. Hence, the number of drop-outs should be clearly mentioned. It will be nice to mention in what stage of the study drop-out had occurred. Data of drop-outs (depending on the time of drop-out) should be excluded from the study and should be mentioned in the 'methods'. Accordingly, the final sample size (after excluding the drop-outs), should be mentioned. Randomization (of subjects/of intervention) in the group, if any should be mentioned. Brief description should be given about the type of randomization. The word 'random' should not be used loosely.

In description of 'Statistical analysis of data' that usually appears at the end of 'methods', as a separate sub-section, following aspects should be addressed:

1. Expression of data :

It should be clearly mentioned how the data are presented in the manuscript such as Mean/Mode/Median/SD/SE. In human studies (as they have larger sample size), data is usually presented as mean \pm SD, and in animal studies (as they have smaller sample size), data is usually presented as mean \pm SE (3).

2. Statistical tests used :

The statistical test used should be described, such as Student's t test (paired/unpaired), ANOVA (one-way, two-way, repeated measures) etc. It is nice to mention in detail about which parameter was assessed by which test. Also, name of the post-test used, if any such as Tukey-Kramer post-hoc, should be mentioned.

3. Correlation/regression analysis :

If any important data was significantly correlated with the alteration in a physiological variable, the method of correlation/regression analysis used such as linear (Pearson) correlation, Spearman nonparametric correlation, linear regression, logistic

regression should be used. It should be clearly mentioned about the data/parameter correlated with variable. For, example, correlation of Lp(a) with diastolic blood pressure was performed by Pearson correlation analysis. In a case-control study, associations are commonly estimated using Odd's ratio, as such studies are typically done when the study outcome is uncommon in the population in which the cases and control arose; and Odd's ratio approximate the risk ratio. The 95% confidence interval should be mentioned for risk ratio (2).

4. *Level of significance :*

The level of P value, below which the statistical significance of the data was determined, should be clearly mentioned.

5. *Software used :*

Name of the software used for data analysis with version (such as SPSS version 13, Graph-Pad version 5), year and company name should be mentioned.

6. Give references for methods which are not commonly used.

Example of poor/inadequate description of statistical data

Statistical analysis

The results expressed are mean \pm SD. The data were analyzed by Student's *t* test. The P values <0.05 was considered significant. The significance of correlation of Lp(a) with diastolic BP was established by using appropriate analytical methods.

Example of better description of statistical data

Statistical analysis

The results are presnted as mean \pm SD. Lp(a) was categorized into tertiles based on the cut-off points of the entire distribution. Proportion and means for baseline cardiac risk factors were calculated. The significance of any difference was tested with the Kruskal Wallis test, and using analysis of variance (two-way ANOVA). All P values were two-tailed. Differences were considered statistically significant for P values <0.05 . A multivariate logistic regression analysis was performed for establishing

correlation (SPSS version 13, SPSS Inc. Chicago, IL, USA). The results of the multivariate analysis were expressed as Odd's ratio with 95% confidence intervals. The relative Odd's ratios were calculated by cross tabulation using the χ^2 test. Bivariate correlation and linear regression analyses were used to assess the correlation between Lp(a) and other physiological variables.

Statistical description in 'Results' of manuscript

In 'results' section, the level of significance for various parameters between the groups should be clearly depicted, both in the text part and in the table/bar-diagrams/graphs. In the text/description of the results, the level of significance for each parameter should be mention very categorically. The level of P values such as $P<0.05$, $P<0.01$ etc. may be mentioned with the bracket. Absence of significance for important parameters, if any should also be mentioned. Even values of important parameters with their mean \pm SD/SE should be mentioned in the text. For details, the corresponding table or figures within bracket may be referred.

A nicely prepared table or a figure not only explains the level of statistical significance for various parameters between the groups, but also the important concepts/applications of the work. A researcher must learn the art and science of preparing tables and graph. A table/graph should be independent and complete in its expression. The author must know that the reviewer should be able to understand the result and application of the work from the table/figures without referring the text of the manuscript. Therefore, tables/graphs should be self explaining with proper title, the way data are presented (mean \pm SD/SE), sample size in each group, the methods used for analysis of data and post-hoc, the level of significance of the entire group for a particular data, the level of significance between the groups for the data and P values for each asterix indicated in the footnote of the table for various groups.

Example of good presentation of data in a table:

TABLE 1: Age, body weight, BMI and basal cardiovascular parameters (in supine resting condition) of the subjects of control group (normotensives) and study groups (pre-hypertensives and hypertensives).

<i>Parameters hypertensives values</i>	<i>Normotensives Hypertensives</i>	<i>P r e - P</i>
<i>(n=36)</i>	<i>(n=35)</i>	<i>(n=32)</i>
Age (years) 32.1 \pm 5.1	34.3 \pm 4.3 33.4 \pm 4.5	0.780

BW (Kg)	55.4±7.6		SBP (mmHg)	110.5±10.8	
61.4±6.5*	68.8±7.1**,#	0.032	121.5±12.5*	134.8±10.2**,##	0.002
BMI (Kg/m ²)	22.5±2.4		DBP (mmHg)	70.5±6.5	
27.5±3.6*	31.4±4.1**,#	0.020	82.5±8.2*	106.6±7.8***,##	0.001
BHR (mmHg)	65.5±8.6				
67.4±7.6	71.4±8.2 #	0.048			

Data presented are mean±SD. Statistical analysis of data was done by one-way ANOVA and post-hoc by Tuke-Kramer test. The * depicts comparison with normotensives and # depicts comparison with pre-hypertensive subjects. * P < 0.05; ** P < 0.01; *** P<0.001; # P < 0.05; ## P < 0.01. BW: body weight; BMI: body mass index; BHR: basal heart rate; SBP: systolic blood pressure; DBP: diastolic blood pressure.

Statistical mention in 'Discussion' of manuscript

In 'discussion', detail description of data is not needed, as the data have already been described in 'results'. However, for important observations, the level of significance between the groups may be mentioned with P values in bracket, while discussing the importance or implication of the data obtained or comparing the same in reference to other works in the field. However, unnecessary repetition of data in 'discussion' that has already been described in 'results', should be avoided.

CONCLUSION

The primary objectives of reporting statistical information in a medical journal article are to clearly highlight the significance or importance of the work, so that the paper is accepted (by the reviewer) for publication, and once published, to make

the reader easily understand the validity of the work. As most of readers have only basic knowledge in statistics, authors should make a simpler presentation of statistical data in a paper.

3. Kelsey JL, Whittermore AS, Evans AS, Thompson WD. In: *Methods in observational epidemiology*. New York, Oxford Univ Press, 1996: 136.

REFERENCES

1. Pal GK. Effective writing of research paper. *Biomedicine* 2007; 27: 1-3.
2. Peter C. Reporting statistical information in medical journal article. *Arch Ped Adplesc Med* 2003; 157: 321-324.

G. K. PAL

Editor (Physiology), IJPP,
Professor of Physiology,
JIPMER, Pondicherry – 605 006
 E-mail: physiologyeditorijpp@rediffmail.com
gopalpravati@sify.com