

Reiterating rationale use of statistical tools in research articles

Manandhar N

Naresh Manandhar, Associate Professor; Department of Community Medicine, Kathmandu Medical College, Kathmandu, Nepal.

DOI: <https://doi.org/10.3126/jkmc.v8i3.29712>

Statistics is a branch of science that deals with the collection, organization, presentation, analysis and interpretation of numerical data from the samples to the whole population¹. This requires proper designing of the study, appropriate selection of the study sample and choice of suitable statistical tests. An adequate knowledge of statistics is necessary for proper designing of an epidemiological study or a clinical trial. Improper statistical methods may result in erroneous conclusions which may lead to unethical practice². Statistical methods involved in carrying out a study include planning, designing, collecting data, analyzing, drawing meaningful interpretation and reporting of the research findings. The statistical analysis gives meaning to the meaningless numbers, thereby breathing life into a lifeless data. The results and inferences are precise only if proper statistical tests are used. Over the past decades, a great increase in the use of statistical tools has been documented for a wide range of medical journals. The trend toward the usage of more sophisticated techniques can be observed and favored by the availability of statistical software packages such as MS Excel, Statistical Package for Social Science (SPSS), Statistical Data (STATA), Statistical Analysis System (SAS), EPI Info, Minitab, R, etc. This editorial shall acquaint the reader with a few basic statistical concepts that are utilized while conducting various researches.

First, researchers must be familiar with their study variables. A variable is a characteristic that varies from one individual member of population to another individual². Researchers have to know the scale of each variable to be included in the analysis. Variables can be of nominal /categorical, ordinal, interval and ratio

scale. This needs to be done for both independent and dependent variables. A nominal scale variable is a variable where the categories just have names, such as sex, religion, caste, or whether participants are in the control or experimental group. Ordinal scale data is data that is ordered, like result of exam: first, second, third position etc.; the main point is that the distance between first and second and between second and third are different from each other. When the distance between units is the same, scale will be interval data. Interval scale data have equally spaced units, such as a temperature 30-35 and 40-45. Ratio scale data are similar to interval scale data, except that the data have a zero point in it, like age, time or weight.

Second, the researchers must know whether they are doing descriptive or analytic statistics. Descriptive statistics is a type of statistics that helps researchers and readers to understand the information of data collected through its organization and summarization³. The use of raw data in research article is uncommon and may impair its interpretation and make readers not interested. It is used to describe data using numbers or statistical measures that may best represent all collected data during research. Inferential statistics, on the other hand, is used to draw conclusion and make inference after analyzing data collected in research³. It includes hypothesis testing and estimates to make comparison and prediction and draw conclusion.

Third, it is important to find out whether the numerical data (quantitative variables) follows normal distribution. If numerical data (quantitative variables) follows normal distribution, parametric tests are used to analyze data. Two most basic prerequisites for parametric statistical analysis are: the assumption of normality which specifies that the means of the sample group are normally distributed, and the assumption of equal variance which specifies that the variances of the samples and of their corresponding population are equal. However, if the

Address for correspondence

Dr. Naresh Manandhar
Associate Professor, Department of Community Medicine
Kathmandu Medical College, Kathmandu, Nepal
E-mail: nareshsayami@yahoo.com
ORCID: <https://orcid.org/0000-0002-4113-7076>

distribution of the sample is skewed towards one side or the distribution is unknown due to the small sample size, non-parametric statistical techniques are used.

Failure to acknowledge and address the above-mentioned concepts is common. The following five are common statistical errors and shortcomings that are frequently observed in research papers⁴:

- Use of a wrong or suboptimal statistical test
- Failure to adjust for multiple comparisons

- Failure to indicate confidence intervals for main effect size measures
- Failure to define all statistical tests being used clearly and correctly
- Failure to conduct a priori sample size estimation

With regards to the proper statistical tests, the following table provides a list of tests to be used appropriately according to measurement scale of variable, types of variable, type of distribution and number of sample group.

Table 1: Appropriate parameters by its measurement scale^{5,6}

Characteristics	Group	Scale	Distribution	Statistical Tests/ Parameters
Description	one sample	R/I	ND	Mean/ Standard Deviation
			NND	Median/ Inter quartile range
		O	-	
		N	-	Proportion
Comparison of one sample group to hypothetical value	One	R/I	ND	One sample t test
			NND	Wilcoxon test
		O	-	
		N	-	Chi square/ Binomial test
Comparison of two sample groups	Unpaired	R/I	ND	Unpaired t test
			NND	Man Whitney test
		O	-	
		N	-	Fisher exact / Chi square for large sample
	Paired	R/I	ND	Pair t test
			NND	Wilcoxon test
		O	-	
		N	-	Mc Neymar test
Comparison of three or more groups	Unmatched	R/I	ND	One way ANOVA
			NND	Kruskal Wallis test
		O	-	
		N	-	Chi square test
	Matched	R/I	ND	Repeated Measure / ANOVA
			NND	Friedman test
		O	-	
		N	-	Cochrane Q
Measure of association between	Two variables	R/I	ND	Pearson's Correlation
			NND	Spearman's rank Correlation
		O	-	
		N	-	Contingency Coefficients
Prediction	From another measure variable	R/I	ND	Simple linear Regression
			NND	Non parametric Regression
		O	-	
	From Several measure variables	N	-	Simple Logistic Regression
		R/I	-	Multiple linear Regression
		N	-	Multiple Logistic Regression

R/I: Ratio and Interval, O: Ordinal, N: Nominal, ND: Normal Distribution, NND: Non-Normal Distribution

The quality of statistical reporting and data presentation in research papers is generally poor. Researchers must understand that the accurate communication of scientific findings depends on transparent reporting of methods and results. Specifically, information on data variability and results of statistical analysis are required to make accurate inferences.

Various reporting guidelines have been developed, endorsed and mandated by various journals to improve the quality of research reporting. Furthermore, journals have published editorial advice to advocate better reporting standards. Nevertheless, it is arguable whether reporting standards have improved substantially⁷. For example, Drummond and Vowler recommended to:

(1) report variability of continuous outcomes using standard deviations instead of standard errors of the mean,

(2) report exact p-values for primary analyses and post-hoc tests, and
(3) plot raw data used to calculate variability.

These recommendations were made so authors would implement them in future research reports. However, it is not known whether reporting practices in these journals have improved since the publication of this editorial advice⁷.

To conclude, selection of appropriate statistical tools is very important in research. It is important that the researchers know the basic concept of the statistical methods used to conduct research that yield valid and reliable results. There are various statistical tools that can be used in different situations. Each test makes certain assumptions about the data. These assumptions should be taken in consideration when deciding the most appropriate statistical test.

REFERENCES

1. Mahajan BK, Methods in Biostatistics, 7th edition, New Delhi, India, Jaypee brothers medical publishers (P) Ltd, 2010
2. Zulfiqar Ali Z, Bhaskar SB: Basic statistical tools in research and data analysis Indian J Anaesth. 2016 Oct; 60(10): 790.
3. Rodrigues CF, de Lima FTC, Barbosa FT. Importance of using basic statistics adequately in clinical research. Rev Bras Anesthesiol, 2017 Apr; 67(6): 619-25. [PubMed]
4. Alexander MS, Qamruz Z, Gerhard M, Karl PP, Hanno U. The Use of Statistics in Medical Research: A Comparison of The New England Journal of Medicine and Nature Medicine. The American Statistician. 2007 Feb;61(155) [DOI]
5. Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. Ann Card Anaesth. 2019;22:297-301. [DOI]
6. Jaykaran: How to select appropriate statistical test? Technical Note. Journal of Pharmaceutical Negative Results. 2010 Oct; 1(2):61-3. [DOI]
7. Diong J, Butler AA, Gandevia SC, Héroux ME. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice, PLoS One. 2018; 13(8): e0202121. [DOI]