

Problems in reporting of statistics: comparison between journal related to basic science with journal related to clinical practice

Jaykaran, P Yadav, P Bhardwaj, J Goyal

Citation

Jaykaran, P Yadav, P Bhardwaj, J Goyal. *Problems in reporting of statistics: comparison between journal related to basic science with journal related to clinical practice*. The Internet Journal of Epidemiology. 2008 Volume 7 Number 1.

Abstract

Introduction- For any research paper most important part is method section. Results and conclusions derived from research articles are of no validity if methods are not appropriate. Statistics is very important part of method section. It is observed that many articles published in various journals have one or more statistical faults. **Methods-** We appraised research paper published in Indian journal of pharmacology and Indian pediatrics in 2007 and 2008 on the basis of modified checklist. **Results-** We observed that calculation of sample size was not mentioned in any article in Indian journal of pharmacology. It was mentioned in 24% of the article of Indian pediatrics. Confidence interval was mentioned in only one article of Indian journal of pharmacology but it was mentioned in 6 (13%) article of Indian pediatrics. Distribution of data was analyzed in only three percentages of articles in Indian journal of pharmacology and 15% of article in Indian pediatrics. Average sample size for clinical trials was 18.9 in Indian journal of pharmacology whereas in Indian pediatrics it is 66.6. Appropriate statistical tests were used in 91% of the articles in Indian journal of pharmacology whereas 93% in Indian pediatrics. Study protocol was followed in 100% of articles and two tailed tests were used in 100% articles. Parametric tests were used in 89% of articles while nonparametric tests were used in 11% of articles in Indian journal of pharmacology. In Indian pediatrics these values are 84% and 15% respectively. **Conclusion-** We believe that awareness need to be generate regarding the use of confidence interval, distribution of data, adequate sample size and nonparametric statistics in both of the journals.

INTRODUCTION

The factor which decides the worthness of reading a research paper is not the interesting hypothesis, nature and potential of result or the speculations in the discussion but the design of method section. Statistics is very important part of method section. Statistics gives range of powerful tools to help researcher to understand the biological structure within the data.

Article published regarding the assessment / reporting of statistical analysis in various journals revealed that many researchers appears to misunderstand the fundamental concept of statistics.¹ This misunderstanding and misuse of statistics jeopardize the focus of scientific discovery and accumulation of scientific knowledge. Statisticians have documented that statistical errors are common in the scientific literature: roughly 50% of published articles have at least one error.²³

So we decided to appraise the various statistical methods described in journal related to basic science that is "Indian

journal of pharmacology" and journal related to clinical practice "Indian pediatrics" and try to get any message by comparing the statistical methods described in them.

MATERIAL AND METHODS

All the authors independently surveyed the original contributions published in the Indian journal of pharmacology and Indian pediatrics in 2007 and 2008. We noted the statistical techniques in the modified proforma on the basis of checklist (see box).⁴ Original contributions related to pharmacoepidemiology, pharmacoeconomics and cell line studies, secondary studies and case reports were not included. Discrepancies between the findings of all reviewers were resolved by consensus.

Figure 1

Check list for the statistical aspects of a paper	
1. Have the authors set the scene correctly?	
a. Have they determined whether their groups are comparable and, if necessary adjusted for baseline differences?	
b. What sort of data they got and have they used appropriate statistical tests?	
c. If statistical tests in the paper are obscure why have the author chosen to use them?	
d. Have the data been analyzed according to original study protocol?	
2. Paired data, tails and outliers:	
a. Were paired test performed on paired data?	
b. Was a two tailed performed whenever the effect of an intervention could conceivably be a negative one?	
c. Were outliers analyzed with both common sense and appropriate statistical adjustments?	
3. Correlation, regression and causation:	
a. Has correlation been distinguished from regression and has the correlation coefficient (r value) been calculated and interpreted correctly?	
b. Have assumptions been made about the nature and directions of causality?	
4. Probability and confidence	
a. Have P value been calculated and interpreted appropriately?	
b. Have confidence interval calculated and do the authors' conclusions reflect them?	
5. Have authors expressed their results in terms of the likely harm or benefit that an individual patient can expect, such as:	
a. Relative risk reduction	
b. Absolute risk reduction	
c. Number needed to treat	
d. Odds ratio	

RESULTS

Figure 2

Table 1: Reporting of statistical methods in Indian journal of pharmacology.

Checklist points	Frequency in Indian journal of Pharmacology	Frequency in Indian pediatrics
Sample size calculation	0/58	11/45 (24.4%)
Average sample size (Human studies)	18.9	66.6
Baseline comparison	4/58 (2.3%)	43/45 (95.5%)
Ratio data	50/58 (86%)	9/45 (20%)
Ordinal	4/58 (7%)	2/45 (4%)
Mixed	4/58 (7%)	34/45 (74%)
Endpoints (average)	5	4.7
Appropriate statistical test	53/58 (91%)	42/45 (93%)
Parametric test	52/58 (89%)	38/45 (84%)
Nonparametric test	6/58 (11%)	7/45 (15%)
Normal distribution analysed	2/58 (3%)	7/45 (15%)
Obscure test	0/58 (0%)	0/45 (0%)
Study protocol followed	58/58 (100%)	45/45 (100%)
Two tailed test	58/58 (100%)	45/45 (100%)
Outliers	0/58 (0%)	0/45 (0%)
Correlation and regression	2/58 (1.1%)	9/45 (20%)
r value	2/58 (1.1%)	7/45 (15%)
p value <0.05	56/58 (96%)	45/45 (100%)
p value <0.01	2/58 (4%)	0/45 (0%)
Confidence interval	1/58 (1.7%)	6/45 (13%)
Relative risk reduction	0/58 (0%)	0/45 (0%)
Absolute risk reduction	0/58 (0%)	0/45 (0%)
Number needed to treat	0/58 (0%)	0/45 (0%)
Use of software	17/58 (29.3%)	31/58 (29.3%)

Values in parenthesis are percentages

From the issues of 2007 and 2008, 58 original contributions were collected from Indian journal of pharmacology and 45 from Indian pediatrics. Among the articles from Indian journal of pharmacology 49 (84%) were animal studies and 10 (16%) were clinical (human) studies. From articles of Indian pediatrics 15 were related to clinical trials, 14 were cohort studies, 6 studies were cross sectional survey and 10 were other prospective and retrospective studies. We observed that in Indian journal of pharmacology in none of the articles calculation of sample size was mentioned whereas in Indian pediatrics calculation of sample size was mentioned in more than 24 percent of articles. Average sample size in animal studies was 5.5 per group and in clinical study it was 18.9 per group in Indian journal of pharmacology. In clinical studies of Indian pediatrics average sample size is 66.6.

In Indian journal of pharmacology baseline comparison table between the study group and control group was not mentioned in any of the animal study. 4 (44%) article among the total 9 article of clinical studies were having the baseline comparison table. Among these 4 articles the statistical analysis between the baseline variables done in 2 articles.

None of the article mentioned about the adjustment of baseline variables whereas in Indian pediatrics baseline comparison table between the study group and control group was mentioned in more than 95 percent of articles. Except one, in all of the article adjustment of baseline variables was done.

In Indian journal of pharmacology among the types of data majority of the articles were dealing with ratio data (86%). Ordinal data were seen in 4 (7%) articles. Mixed data for different variable observed in 4 (7%) articles and in Indian pediatric most common (48%) was ratio and nominal data in combination in one article. Nominal data were seen in 11 (24%) articles. Ratio data was seen in 20% of articles.

Average endpoints per study were 5 in the Indian journal of pharmacology whereas it was 4.7 in Indian journal of pediatric. We found that normal distribution of the data was assessed in 3% articles in Indian journal of pharmacology and in Indian pediatrics it was assessed in 15% of articles. Appropriate statistical tests were used in 91% articles and 93% of the articles in Indian journal of pharmacology and Indian pediatric respectively. Parametric statistical tests were used in 89% studies in Indian journal of pharmacology and 84% studies in Indian pediatrics. Nonparametric tests were used in 11% studies of Indian journal of pharmacology and 15% in Indian pediatrics. Though obscure test was not used in any article but none of the article mentioned about the reference of any statistical test. No violation of original protocol noticed in any article.

Paired test use for unpaired data or vice versa was not observed in any study. All articles used the two tailed test. No article mentioned about the outliers.

In Indian journal of pharmacology correlation and regression test was used in only 2 studies and r value was calculated. No assumptions were made regarding the causality. In Indian pediatric correlation and regression test was used in 9 studies and r value was calculated in 7 of them. Assumptions were made regarding the causality in one study.

P value was calculated in almost all articles. It was the only method to see the difference between study group and control group. Most of the article mentioned about the P value as <0.05 . . Exact P value was not calculated in any article in Indian journal of pharmacology but it was mentioned in most of the article published in Indian pediatric. In Indian journal of pharmacology confidence interval was calculated in only one article. In Indian

pediatric it was calculated in 13% of articles.

None of the study measures other outcomes like Relative Risk, Relative risk reduction, Absolute Risk Reduction and Number Needed to Treat.

In Indian journal of pharmacology in 17 articles statistical software was used. Most frequently used software was SPSS (47%) followed by Graph pad (35%). Others were sigma state, instate and SAS. In 31 articles statistical software was used. Most frequently used software was SPSS (77%) followed by stata (13%).

DISCUSSION

In this study we found that sample size calculation is ignored in most of the articles from both of the journals. If sample size is less, then there are more chances of occurring of the type II error and wider range of confidence interval.⁵ less sample size responsible for the less power of study. Roberts and colleagues (2002) did a metaanalysis of 44 animal experiments on fluid resuscitation and found that none of them had sufficient power to reliably detect a halving of death rate.⁶

In clinical trials sample should be big enough to have a high chance of detecting, as statistically significant, a worthwhile effect if it exists, and thus to be reasonably sure that no benefit exists if it is not found in trial.⁷ for sample size calculation in hypothesis testing researcher must know the effect size, standard deviation, significant level and power of study.⁷ effect size and standard deviation of new agents can be calculated by pilot study. Not only the sample size should be calculated but it should be mentioned that how it was calculated. For example-

“A sample of 30 patient per group was required to detect a difference in IOP of 4 mmHg (and an SD of 4mmHg) with 90% power and an α error of 0.05”.⁸

A recent review finds that the most common potential cause of statistical error is small sample size (< 10). Very small sample size 1. Result in parameter estimates that are unnecessarily imprecise. 2. It enhances potential for failed randomization 3. Yield hypothesis testing that is underpowered 4. Yield hypothesis testing that is biased because assumptions underlying the applied statistical methods could not be examined adequately.⁹ Our study shows that average sample size is very less (for animal studies- 5.5) and for human – in Indian journal of pharmacology- 18.9) and so chances of errors are more. In

Indian pediatrics sample size is more (66.6) as compared to human studies of Indian journal of pharmacology.

In animal studies usually baseline characteristics of study group and control group are not mentioned. But baseline characteristics should be given in tabular form and should be matched. Unequal distributions of variables like male female ratio, weight, age, strain etc may affect the results. If randomization is perfect, than generally comparable groups are formed. But sometimes factors which determine outcome might be unequally distributed in study and control group. This may threaten the validity of statistical analysis. Appropriate statistical tests should be used to see any significant difference between the two groups. If different is there than it should be adjusted by statistical methods.¹⁰

It was observed that data were checked for normal distribution in only 3% of studies in Indian journal of pharmacology and 15 % in Indian pediatrics . This raise concern as knowing the type of distribution is one of the prerequisite for the selection of the statistical test. Biological variables are notorious to fall in distribution free (nonparametric) statistics. Small sample size enhances this problem. Majority of researcher don't want to test there data for normal distribution as they might struck in non-parametric tests and getting significant result through nonparametric test is difficult as compared to parametric tests.⁴

The most common problems associated with the appropriateness of statistical tests are¹¹:-

1. Using parametric tests when the data are not normally distributed (skewed). In particular, when comparing two groups, Student's paired t test is often used when the Wilcoxon rank-sum test (or another nonparametric test) is more appropriate.
2. Using tests for independent samples on paired samples, which require tests for paired data. Again, Student's unpaired t test is often used when a paired t test is required.

It is encouraging that in none of the article paired test were used on unpaired data and vice versa. But in few studies parametric test were used at the place of nonparametric. Mainly these data were dealing with rank and scores. The incidence of inappropriate tests are less as compared to other studies published in this area.²³¹² Reference for statistical test was not given in any article. Reference should be given from the standard work not from the previous article published

related to similar study.¹³ Nonparametric tests were used in 11% of the studies in Indian journal of pharmacology and 15% in Indian pediatrics. Similar studies observed more extensive use.¹⁴

Other encouraging thing observed in statistics of both of the journal is that none of the study violated the original protocol. There was no premature termination of study. There was no retrospective subgroup analysis. Retrospective subgroup analysis many times leads to false conclusion.⁴

Two tailed test were used in all the articles. That is also very encouraging. One tailed test justifiable in two circumstances 1. A difference between two mean in one of two directions is known to be impossible or 2. A difference in one direction is of no interest whatsoever the circumstance. But these instances are rare. So use of two tailed test is a right thing.

It is seen that multiple hypothesis testing was done in most of the articles. Separate test was used for each variable. Average end points in article published in 2007 and 2008 are 5 in Indian journal of pharmacology and 4.7 in Indian pediatrics. When separate tests were used for each variable than each test leads to some type I error. This type I error get compounded with every new variable and that may seriously threat the validity of using statistical test. There are some powerful statistical methods which can take care of this problem.¹⁶ No article mentioned about the outliers. Unexpected results may reflect idiosyncrasies in the subject, error in measurement, error in interpretation or error in calculation. Only first of these should be included in analysis. Outliers can be corrected statistically.⁷

Correlation and regression in not utilized much in these journal. Only two studies are deals with r value in Indian journal of pharmacology and seven studies in Indian pediatrics. But the r value (Pearson's coefficient) is most overused statistical instrument in books. A r value is not valid unless it follows these criterias- 1. Data should be normally distributed 2. Two variables should be strictly independent 3. Only single pair of measurement should be made on each subject 4. Every r value should be accompanied by P value or confidence interval.⁴ assumptions related to causality should not be made on the basis of r value. Some criteria should be fulfilled to established causality.¹⁶

We observed that P value was the sole criteria for assessment of outcome. Confidence interval was calculated

in only few articles. We believe that awareness regarding the use of confidence interval needs to be generated in authors.

The International Committee of Medical Journal Editors issued these guidelines for reporting statistics¹³

“Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid sole reliance on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information. . . . Give numbers of observations. . . . References for study design and statistical methods should be to standard works (with pages stated) when possible rather than to papers where designs or methods were originally reported. Specify any general-use computer programs used”

This guideline clearly mentions the importance of confidence interval and weakness of P value. P value is arbitrary cut off point. It shows the statistical significance not clinical significance as it does not reflect the size of effect. In spite of using “ $P < 0.05$ ” authors should mention exact P value. Comparison between two P values can be done easily by this. Gordon Guyatt et al concluded that “why use a single cut off point (for statistical significant) when choice of such a point is arbitrary? Why make the question of whether a treatment is effective a dichotomy (a yes – no decision) when it would be more appropriate to view it as a continuum?”¹⁰

Confidence interval gives us range of the value around the effect size of sample where population mean actually may lie. Much information may be gained by careful observation of confidence interval. In positive studies lower boundary of confidence interval shows the adequacy of sample size and definitiveness of study. Same things can be seen by observing the upper boundary of negative studies.⁵ confidence interval should be calculated for about any statistical test as t value, r value, relative risk reduction, odds ratio, number needed to treat etc.

Calculation of parameters like relative risk, relative risk reduction, absolute risk reduction, odds ratio and number needed to treat are useful tools to get more information. Whenever possible these parameters should be calculated.⁴

We believe that this guideline given in instructions to authors regarding statistics reporting need to be more elaborate as it doesn't talk about confidence interval, distribution of data etc. for authors of poor statistics background this information is not sufficient.

We observed that as compared to previous studies in other journals the statistics of Indian journal of pharmacology and Indian pediatrics is better.^{12,14} Editors should generate more awareness regarding confidence interval, distribution of data, nonparametric statistics and calculation of sample size.

References

1. Curran-Everett D, Taylor S, and Kafadar K. Fundamental concepts in statistics: elucidation and illustration. *J Appl Physiol* 1998 ; 85: 775–786.
2. Altman DG. Statistics in medical journals: some recent trends. *Stat Med* 2000;19: 3275–3289,
3. Altman DG and Bland JM. Improving doctors' understanding of statistics. *J R Stat Soc Ser* 1991;154: 223–267.
4. Trisha Greenhalgh. How to read a paper? The basics of evidence based medicine. 3rd ed. London. BMJ Publishing house; 1997.
5. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians. 2. Interpreting study results: confidence intervals. *Can Med Assoc J* 1995: 152: 169-73.
6. Roberts I, Kwan I, Evans P, Haig S. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. *BMJ*. 2002 ;324:474-6.
7. Altman D. Practical statistics for medical research. London: Chapman and Hall, 1991. 456.
8. Sudheera KS, Bhardwaj N, Yaddanapudi S. Effect of intravenous metoclopramide on intraocular pressure: A prospective, randomized, double- blind, placebo- controlled study. *J Postgrad Med* 2008; 54: 195-8.
9. Tyson H. Ten categories of statistical errors: a guide for research in endocrinology and metabolism. *Am Physio Endocrinol Metab* 2004; 286: 495-501.
10. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians. 1. Hypothesis testing. *Can Med Assoc J* 1995: 152: 169-73.
11. Tom lang. Twenty Statistical Errors Even You Can Find in Biomedical Research Articles. *Croat Med J* 2004; 45:361-370.
12. Mantha S. Statistical reporting of study results: is there a scope for improvement. *Journal of Anaesthesiology Clinical Pharmacology*. 1994 Apr; 10(2): 81-4.
13. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med* 1997;126: 36–47,.
14. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Eng J Med* 1987; 317: 426-32.
15. Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics* 1987; 43: 487-98.
16. Bradford Hill, A. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300.

Author Information

Jaykaran, MD (Pharmacology)

Tutor, Department of Pharmacology, Govt. Medical College

Preeti Yadav, MD (Pharmacology)

Associate professor, Department of Pharmacology, Govt. Medical College

Pankaj Bhardwaj

Assistant professor, Department of community medicine, Era medical college

J.P. Goyal

Assistant professor, Department of pediatrics, Govt. medical college