SPECIAL ARTICLE
J Korean Neurosurg Soc 40 : 304-308, 2006

A Review of the Statistical Analysis used in Clinical Articles Published on Journal of Korean Neurosurgical Society

Weechang Kang, Ph.D.
Department of Applied Statistics, Daejeon University, Daejeon, Korea

Statistical analyses used in clinical articles published on the Journal of Korean Neurosurgical Society were identified and appropriateness of statistical aspects in reporting results was assessed. Forty seven clinical articles were selected in this study, which were published from February, 2005 to February, 2006 on the journal. The frequency of statistical analysis was as follows : descriptive statistics only 24 (51.1%), one type of statistical method 10 (21.3%), two or more methods 13 (27.6%). An assessment of statistical aspects was performed in 24 clinical articles reporting inferential statistics. Ten articles (41.7%) did not adequately describe or reference all statistical methods used. There were six articles (25.0%) not reporting the confidence level used as the critical criteria of the statistical significance. In thirteen articles (54.2%) it seems more appropriate to implement multivariate analyses in addition to univariate analyses. We recommend that the journal readers should concentrate on improving their knowledge of basic statistics and statistical review for manuscripts submitted should be sought from professionals in the fields of biostatistics and epidemiology.

KEY WORDS : Statistical analysis · Clinical articles · Appropriateness.

Introduction

In reporting results of medical studies, statistical analyses such as statistical modeling and significance testing have become essential features to ensure that conclusions are based on evidence rather than opinions. This has resulted in an frequent use of statistical tests, making interpretation of scientific publications more difficult[1,10,18]. As a consequence the lack of statistical support and knowledge has become one of the most critical obstacle to scientific reporting of studies and correct understanding of articles published. Therefore, it has become important for investigators, reviewers and editors to ensure the validity of statistical methods used and the correct interpretation of their statistical results. Also, journal readers should be familiar with the statistical methods used and improve comprehension of clinical findings[13].

The frequency of use of statistical methods in general or specialty-oriented medical journals have previously been surveyed[2,4-8,10-16]. Descriptive statistics, t-tests for continuous variables, Pearson's chi-square test for contingency tables, simple correlation/regression analysis and corresponding nonparametric tests are the most common statistical tests reported on the medical journals. The aim of this study is to identify which statistical analyses are used in clinical articles published on the Journal of Korean Neurosurgical Society (JKNS) and to assess whether statistical aspects are appropriate in the articles.

Materials and Methods

We examined 108 articles published from February, 2005 (Vol. 38, No. 2) to February, 2006 (Vol. 39, No. 2) on the JKNS. Among them, 47 clinical articles were selected in this study, excluding 43 case reports, 2 technical notes and 2 special articles, 2 review articles and 12 laboratory investigations because they rarely used statistical methods and not contained clinical data.

• Received : August 31, 2006 • Accepted : September 6, 2006
• Address for reprints : Weechang Kang, Ph.D., Department of Applied Statistics, Daejeon University, 96-3 Yongun-dong, Dong-gu, Daejeon 300-716, Korea Tel : +82-42-280-2446, Fax : +82-42-280-2444, E-mail : weechang@dju.ac.kr

Each article was reviewed to evaluate statistical aspects in reporting the study and to determine the methods of statistical analyses used. Statistical aspects were assessed according to a check list similar to that used for statistical review of general papers on the British Journal of Medicine(BJM)[9] and are given Fig. 1. As shown in Fig. 1, the checklist differs slightly from that of the BJM. For example, the question 5 in the BJM asking the response rate achieved is removed because of its less relevance to the JKNS, on the other hand, the questions on confidence level (question 6), analysis for patient's characteristics (question 7) and univariate/multivariate analysis for the primary variable (question 8) are added. Statistical methods were categorized based on characteristics of the variable analyzed and are summarized Table 1. Other categorizations for statistical methods have been found in previous investigations[5,10,13]. The cumulative accessibility[10,13], defined as the percentage of articles to which a journal reader would have full access by being familiar with a given statistical methods as well as all simpler methods, was also calculated (see Table 2).

Study Design Features

1. Was there the objective of study sufficiently described?
   Yes Unclear No
2. Was an appropriate study design used to achieve the objective?
   Yes Unclear No
3. Was there a satisfactory statement given of source of subjects?
   Yes Unclear No
4. Was there a power based assessment of adequacy of sample size?
   Yes Unclear No

Statistical Analysis and Presentation

5. Was there a statement adequately describing or referencing all statistical procedures used?
   Yes No
6. Was the confidence level used given?
   Yes No
7. Were statistical analyses for the patient's characteristics appropriately considered?
   Yes Unclear No
8. Were univariate or multivariate analyses for the primary variable appropriately considered?
   Yes No
9. Were the statistical analyses used appropriate?
   Yes Unclear No
10. Was the presentation of statistical material (i.e. tables or figures including descriptive and inferential statistics) satisfactory?
    Yes No
11. Were confidence intervals or p-values given for the main results?
    Yes No
12. Was the conclusion drawn from the statistical analysis justified?
    Yes Unclear No

Fig. 1. Check list for assessment of statistical aspects of clinical articles for JKNS.

Statistical analysis

Statistical analysis was conducted with SAS for windows, Release 8.02 and p-value less than 0.05 was considered to indicate a statistical significance. Frequency table was reported to summarize the spectrum of use of statistical methods and the response of the check list of statistical aspects. Independent samples t-test was implemented to compare the sample size between articles using descriptive statistics only and those using inferential statistics.

Table 1. Classification of statistical procedures relevant to the JKNS

Category	Statistical methods
None	No statistical procedures employed
Descriptive statistics only	Such statistics as mean, standard deviation, median, inter-quartile range, range, frequency, percentage, ratios, histogram, scatter diagram, frequency polygon, graph of the means
Comparing Means Basic methods	One-sample t-test, Paired t-test, Wilcoxon signed-rank test, Independent samples t-test, Mann-Whitney test, One-way ANOVA, Kruskal-Wallis test, Multiple comparison
Comparing Proportions Basic methods	Pearson's chi-square tests, Fisher's exact test McNemar test Mantel-Haenszel method
Advanced methods Correlation analysis Basic methods	Pearson's correlation coefficient, Spearman rank correlation, Kendall's tau
Association analysis Advanced methods	Relative risk, Odds ratio, Linear-by-linear association
Regression analysis Basic methods	Simple linear regression, Multiple linear regression
Advanced methods	Simple logistic regression, Multiple logistic regression
Survival analysis Basic methods	Kaplan-Meier estimator, Logrank test, Cox regression model

Table 2. Statistical content by category and accessibility

Category	No (%) of articles containing methods (n=47)	Cumulative accessibility by methods (%)
Descriptive statistics only	24 (51.1)	24 (51.1)
Basic statistical methods		
Independent samples t-test	13 (27.7)	28 (59.6)
Paired t-test	1 (2.1)	29 (61.7)
Mann-Whitney test	3 (6.4)	31 (66.0)
Kruskal-Wallis test	1 (2.1)	32 (68.1)
Pearson's chi-square tests	9 (19.1)	38 (80.9)
Fisher's exact test	4 (8.5)	40 (85.1)
Pearson's correlation coefficient	1 (2.1)	41 (87.2)
Simple linear regression	2 (4.2)	43 (91.5)
Advanced statistical methods		
Mantel-Haenszel method	2 (4.2)	44 (93.6)
Linear-by-linear association	1 (2.1)	45 (95.7)
Multiple logistic regression	1 (2.1)	46 (97.9)
Kaplan-Meier estimator	1 (2.1)	46 (97.9)
Cox regression model	1 (2.1)	47 (100.0)

Results

Frequency and spectrum of use of statistical methods

Forty seven clinical articles were included in the study. The frequency of statistical analysis was as follows : descriptive statistics only 24 (51.1%), one type of statistical method 10 (21.3%), two or more methods 13 (27.6%). Inferential statistics such as t-tests, ANOVA, Pearson's chi-square tests, correlation/regression analyses and corresponding nonparametric tests were found in 48.9% of the articles. The first two frequent statistical procedures reported were independent samples t-test used in 27.7% and Pearson's chi-square tests performed in 19.1% of the articles. Advanced statistics including Mantel-Haenszel method, linear-by-linear association, multiple logistic regression, Kaplan-Meier estimator and Cox regression models accounted for 8.5% (Table 2). In Table 2 the value of the cumulative accessibility column indicates the percentage of articles which an individual reader could understand given their knowledge of statistics. For example, a reader with knowledge of descriptive statistics and independent samples t-test only would have informed access to 59.6% of clinical articles on the JKNS.

Assessment of statistical aspects in the reporting of clinical studies

An assessment of statistical aspects was performed in 24 clinical articles including 23 articles used one or more statistical methods and one article which reported only descriptive statistics but mentioned p-value in the Results Section without the corresponding statistical method. The first four questions of the check list (Fig. 1) asking study design features were excluded in this study because it might be more desirable for a biostatistician having application to neurosurgery to evaluate them. The results of the assessment were as follows (Table 3).

Table 3. Assessment of statistical aspects of clinical articles on JKNS

Question	Item of answer No (%) of articles (n=24)		
5. Was there a statement adequately describing or referencing all statistical procedures used?	Yes 14 (58.3)		No 10 (41.7)
6. Was the confidence level used given?	Yes 18 (75.0)		No 6 (25.0)
7. Were statistical analyses for the patient' s characteristics appropriately considered?	Yes 14 (58.3)	Unclear 3 (12.5)	No 7 (29.2)
8. Were univariate or multivariate analyses for the primary variable appropriately considered?	Yes 11 (45.8)		No 13 (54.2)
9. Were the statistical analyses used appropriate?	Yes 4 (16.7)	Unclear 4 (16.7)	No 16 (66.6)
10. Was the presentation of statistical material (i.e. tables or figures including descriptive and inferential statistics) satisfactory?	Yes 2 (8.3)		No 22 (91.7)
11. Were confidence intervals or p-values given for the main results?	Yes 21 (87.5)		No 3 (12.5)
12. Was the conclusion drawn from the statistical analysis justified?	Yes 16 (66.7)	Unclear 3 (12.5)	No 5 (20.8)

Statement of all statistical procedures used

Ten articles (41.7%) did not adequately describe or reference all statistical methods used. Mistakes frequently found were as follows : absence of the Statistical analysis Section describing statistical methods used and the reason of using them, lack of description for some of statistical methods and a statistical package implemented, too sort statement for the methods used, no reporting the results of statistical analyses mentioned in the article.

Statement of confidence level used

There were six articles (25.0%) not reporting the confidence level used as the critical criteria of the statistical significance. The remainder 18 (75.0%) used the confidence level of 0.05.

Statistical analyses for the patient's characteristics

Seven articles (29.2%) did not adequately perform statistical analyses to evaluate the comparability of patient's baseline characteristics, such as age, sex, hospital, the times of year, etc., between the interesting groups.

Appropriateness of statistical analyses used

In thirteen articles (54.2%) it seems more appropriate to implement multivariate analyses in addition to univariate analyses because these articles reported two or more statistically significant variables expected to be relevant to the primary endpoint from the univariate analyses performed. Statistical methods were not suitable in sixteen articles (66.6%). Frequent misapplication were as follows : use of independent samples t-test instead of Mann-Whitney test for a ordinal categorical variable in data set with small sample size, reporting of Pearson's correlation instead of Spearman rank correlation in its analysis for ordinal variables with small sample, regarding data that the majority of the patients have one observation but some patients have two or more observation of the endpoint as independent data.

Appropriateness of the presentation of statistical materials

Inappropriateness of the presentation of statistical materials was founded on twenty two articles (91.7%). Statistical materials were only reported as a form of description on the body

Table 4. Comparison of sample size between articles using descriptive statistics only and those using inferential statistics

Group	n	Minimum	Median	Maximum	Mean	SD*	p-value†
Descriptive statistics only	23	6	16	144	26.3	28.9	0.01
Inferential statistics	24‡	15	50.5	789	135.7	194.425	

* standard deviation, † independent samples t-test., ‡ 23 articles with inferential statistics plus one article which reported only descriptive statistics but mentioned p-value in the Results Section without the corresponding statistical method

of the article without tables or figures summarizing them at most articles. Some descriptive statistics were not clearly interpreted because of lack of statement of their meaning. For example, in the statement of "The mean age of the patients was 33±10 years…", the value 10 is not clear whether the standard deviation of age or the standard error of the mean age. Descriptive statistics summarizing the distribution of data conventionally consist of two or more statistics representing the location and dispersion of data. In general mean value representing the location is followed by standard deviation representing the dispersion[3] and the median value followed by range or interquartile range. However, in some articles the distribution of data was summarized with the mean value and the range only. In those cases it could be more appropriate to additionally describe the median and the standard deviation.

Statistical problems in articles reporting descriptive statistics only

Tables or figures summarizing statistical results were not given in most articles with descriptive statistics only. There were some articles asserting the conclusion not justified from a statistical analysis. The sample sizes of the articles were significantly smaller than those of the articles with inferential statistics. However, the sample size was not too small to apply statistical method to justify the conclusion in many cases, that is, 50% of the articles have a larger sample size than 16 (Table 4).

Discussion

In this study, the frequency of use of statistical methods in clinical articles on recent two volumes of the JKNS has been identified and statistical aspects in these articles have been assessed. A reader familiar with basic statistical methods (independent samples t-test, paired t-test, Mann-Whitney test, Kruskal-Wallis test, Pearson's chi-square tests, Fisher's exact test, Pearson's correlation coefficient, simple linear regression) could successfully understand about 92% of the articles reviewed. This result suggest that the JKNS's readers need to make an effort to improve their knowledge of basic statistics in order to have a complete understanding of the majority of clinical articles published on the JKNS.

The study of the frequency of use of statistical methods did not completely guarantee that there were not "errors" in appropriateness of statistical techniques used[11,13]. As shown in the result of the assessment of statistical aspects, there were some statistical errors or mistakes. Determining "appropriateness" of statistical aspects within a manuscript is a complex process, involving authors, reviewers and editorial staff[17,20,21]. This result based on only one biostatistician's perspective may be restrictive in the assessment of the appropriateness, although the assessment was performed with a standardized check list and concentrated on statistical analyses and presentation rather than study design features. The inappropriateness of the presentation of statistical materials found in most articles was not due to invalid methods of statistical analyses but poor method of summarization of statistical results as mentioned in the Results Section. Also, the inappropriateness was often resulted from reiteration of errors induced from other statistical aspects such as no description of confidence intervals or p-value.

Statistical tests are used to determine statistical significance in the differences of the primary endpoint demonstrated between groups. However, statistical significance dose not necessarily imply a clinical significance[21]. Statistical significance is just a necessary condition for the clinical significance. It is important to describe a confidence interval of the magnitude of the differences in addition to p-values, since indications for subsequent clinical practice rather depend on the magnitude (clinical significance) than p-value (statistical significance)[18].

The quality of a study cannot be examined by the number of statistical tests used nor their degree of complexity. Frequent use of statistical tests in an article may lead to a loss of credibility of them since, inevitably, the risk of reporting some false positive finding increases. Also, the more complexity of statistical methods increase, the more difficulty of interpretation of their results increase, which may mislead investigators into a wrong conclusion. For an enhancement of the quality of a study, the clearly defining of study objective(s) before the start of an investigation, the selection of the suitable study design to meet these and the use of prudential methods to avoid potential biases toward distorting their results are as important as the choice of a valid statistical test. In even descriptive study, careful attention to study design is needed to ensure the validity of the observations[10,18,19].

Authors should report all statistical aspects considered in the study as clear as possible. They should describe the main study objective(s), the study design used to achieve its objective, and the source and rationale for the data set chosen in an easily comprehensible format. The issue of sample size such as mathematical justification (e.g. power calculations) should be well documented in Methods Section, but many studies are still too small to detect anything other than large, and often unre-

alistic, effects[10,17].

The statistical methods used should be stated together with its purpose. If a technique is novel or unfamiliar, a description of an outline of the method should be given together with a suitable reference. An appropriate reporting standard might be that the statistical analyses used be described in sufficient detail that a reader could reproduce the calculations if the data were available. If statistical software packages are used, specific packages and procedures should be identified[9,13]. It might be desirable that the statement related to statistical analyses should be separately specified in the Statistical analysis Section.

Conclusion

We suggest that consumers of the JKNS should make an effort to improve their knowledge of basic statistics and technique of description of statistical results. We hope that the checklist in Fig. 1 be used to review for statistical aspects of a manuscript submitted to the JKNS. It might be desirable for a biostatistician or epidemiologist to document the checklist for the manuscript.

• Acknowledgement
We would like to thank Mr. Dae Hyung Lee for his assistance in the preparation of this manuscript.

References

1. Altman DG : Statistics in medical journals : developments in the 1980s. **Stat Med 10** : 1897-1913, 1991
2. Avram MJ, Shanks CA, Dykes MHM, Ronia AK, Stiers WM : Statistical methods in anesthesia articles : an evaluation of two American journals during two six-month periods. **Anesth Analg 64** : 607-611, 1985
3. Bartko JJ : Rationale for reporting standard deviations rather than standard errors of the mean (editorial) : **Am J Psychiatry 142** : 1060, 1985
4. Colditz GA, Emerson JD : The statistical content of published medical research : some implication for biomedical education. **Med Educ 19** : 248-255, 1985
5. Elster AD : Use of statistical analysis in the AJR and Radiology : frequency methods and subspecialty differences. **AJR AM J Roentgenol 163** : 711-715, 1994
6. Emerson JD, Colditz GA : Use of statistical analysis in the New England Journal of Medicine. **N Engl J Med 309** : 709-714, 1983
7. Feinstein AR : Clinical biostatistics : XXV : A survey of the statistical procedures in general medical journals. **Clin Pharmacol Ther 15** : 97-107, 1974
8. Fromm BS, Snyder VL : Research design and statistical procedures used in the Journal of Family Practice. **J Fam Pract 23** : 564-566, 1986
9. Gardner MJ, Machin D, Campbell MJ : Use of check lists in assessing the statistical content of medical studies. **Br Med J 292** : 810-812, 1986
10. Goldin J, Zhu W, Sayre JW : A review of the statistical analysis used in papers published in Clinical Radiology and British Journal of Radiology. **Clin Radiol 51** : 47-50, 1996
11. Gore SM, Jones IG, Bytter EC : Misuse of statistical methods : critical assessment of articles in BJM from January to March 1976. **Br Med J 1** : 85-87, 1977
12. Greenland S : Quantitative methods in the review of epidemiologic literature. **Epidemiol Rev 9** : 1-30, 1987
13. Hokanson JA, Bryant SG, Gardner R Jr, Luttman DJ, Guernsey BG, Bienkowski AC : Spectrum and frequency of use of statistical techniques in psychiatric journals. **AM J Psychiatry 143** : 1118-1125, 1986
14. Hokanson JA, Ladoulis CT, Quinn FB Jr, Bienkowski AC : Statistical techniques reported in pathology journals during 1983-1985 : implications for pathology educators : **Arch Pathol Lab Med 111** : 202-207, 1987
15. MacArthur RD, Jackson GG : An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. **J Infect Dis 149** : 349-354, 1984
16. Menegazzi JJ, Yealy DM, Harris JS : Methods of data analysis in the emergency medicine literature. **Am J Emerg Med 9** : 225-227, 1991
17. O'Fallon JR, Dubey SD, Salsburg DS : Should there be statistical guidelines for medical research papers? **Biometrics 34** : 687-695, 1978
18. Pocock SJ, Hughes MD, Lee RJ : Statistical problems in the reporting of clinical trials : a survey of three medical journals. **N Engl J Med 317** : 426-432, 1987
19. Rothman KJ : Significance testing. **Ann Intern Med 105** : 445-447, 1987
20. Vaisrub N : Manuscript review from a statistician's perspective. **JAMA 253** : 3145-3147, 1985
21. Weiss GB, Bunce H : Are we ready for statistical guidelines for medical research papers? (letter). **Biometrics 35** : 911, 1979