J Vet Clin 27(2) : 170-174 (2010)

Journal of Veterinary Clinics
www.ksvc.or.kr

Statistical Issues in the Articles Published in the Journal of Veterinary Clinics

Son-Il Pak[1] and Tae-Ho OH*

College of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chunchon, 200-701, Korea
**College of Veterinary Medicine, Kyungpook National University, Daegu, 702-701, Korea*

(Accepted: April 04, 2010)

Abstract : With the ease availability of statistical software and powerful computers the application of statistical methods in domestic veterinary journals is on the increase. In parallel with this benefit, statistical errors are not uncommon even in renowned scientific and medical journals. These errors may lead to misinterpretation of the data, thereby, subjected to faulty conclusions. A systematic review of articles published in 8 issues of the Journal of Veterinary Clinics during 2006-2007 was performed to assess the statistical methodology and reporting. Ninety-four (72.9%) articles of the 129 original articles screened included any inferential statistical analysis in the article, including comparison of 3 or more groups (53 or 56.4%), comparison of independent 2 groups (40 or 42.6%), and paired t-test (9 or 9.6%) in order. Of the 94 articles in which statistical analysis was done 62 (or 66.0%) had at least 1 statistical error. Errors included failure to apply or incorrectly applying independent Student's t-test for paired data or vice versa, inappropriate use of t-test for more than 3 groups and failure in chi-square test to consider continuity-correction for small expected frequencies. The common errors in ANOVA were failure to validate assumption of the test, inappropriate post-hoc multiple-comparison and incorrect assumption of independence of data in repeated measures design. Reporting errors included failure to state statistical methods and failure to state specific test if more than 1 test was done. It is suggested that an editorial effort would be necessary to achieve the improvement of appropriate statistical procedures through the publication of statistical guidelines to author(s).

Key words **:** statistical method, error.

Introduction

Statistical hypothesis testing is primarily to justify or support the conclusions presented by the author(s) and, if possible, to generalize the characteristic of other large population based on findings with a well-designed small sample study. This methodology has widely been used in veterinary medical literatures. With the widespread availability of statistical software and powerful computers, papers with more sophisticated techniques on a variety of experimental designs are increasingly published in domestic peer-reviewed journals. However, misuse and/or abuse of statistical procedures are not uncommon. In an earlier study of 94 articles of volumes from 37 to 39 (No. 1) of Korean Journal of Veterinary Research, 29-66.7% were committed at least one statistical error depending on the statistical procedures, demonstrating that the veterinary literature may be subjected to flaws in statistical methods (11). Reviews on statistical problems have also been reported in many scientific journals (2,5,7,9,10,12).

The aim of the present paper was to critically appraise the articles published in a peer-reviewed journal regarding statistical appropriateness. The issues discussed here are intended to aware veterinary researchers about statistical errors or shortcomings in key aspect of a study including data collection, inferential method applied, data analysis, and interpretation of study results, in the hope that improvement of statistical quality of researches and skills in critical appraisal pf published literatures can be expected.

[1]Corresponding author.
E-mail : paksi@kangwon.ac.kr

Materials and Methods

Selection of publications and review of statistical methods

Articles published in the Journal of Veterinary Clinics (hereafter, JVC) during 2006-2007 were initially screened. Because no statistics or simple descriptive statistics were applied in clinical case reports and short communications the author excluded these papers, thereby, only original articles were included for further review. To evaluate the statistical procedures for common errors the author focused only on inferential statistics for the type of data measured and transparent reporting of the results of statistical analysis. Since no formalized statistical guidelines are provided by JVC, the author developed evaluation criteria based on the current instructions to authors recommended by the Journal of Veterinary Internal Medicine (http://jvim.allentrack.net. accessed June 24, 2008) and accepted statistical principles (1,3,4,6).

Table 1. Use of statistical methods in articles published in the Journal of Veterinary Clinics during 2006-2007

	2006	2007	Total
No. of total articles[1]	58	71	129
No. (%) of articles using inferential statistical methods	41 (70.7)	53 (74.6)	94 (72.9)
Mean inferential statistics/article	1.4	1.3	1.3

[1]Only full articles but clinical case reports were included for further review.

Table 2. Number of articles (%) by type of analysis reported in 2006 (n=41) and 2007 (n=53)

Type of analysis	2006	2007	Total
2 group comparison			
Student's t-test (independent)	20 (48.8)	20 (37.7)	40 (42.6)
Student's t-test (paired)	5 (12.2)	4 (7.5)	9 (9.6)
Chi-square (Fisher's)	3 (7.3)	5 (9.4)	8 (8.5)
≥ 3 group comparison			
ANOVA (1 or 2-way, cross-over)	29 (43.9)	24 (45.3)	53 (56.4)
Multiple comparison	9 (22.0)	18 (34.0)	27 (28.7)
Regression (correlation)	2 (4.9)	6 (11.3)	8 (8.5)
Others			
Diagnostic test evaluation	2 (4.9)	1 (1.9)	3 (3.2)
Sample size	0 (0)	1 (1.9)	1 (1.1)

Each article was counted more than once if multiple analyses were used.

Results

Of the 129 original articles screened 94 (or 72.9%) included any inferential statistical analysis in the description of materials and methods or somewhere in the article (Table 1). The proportion increased from 70.7% in 2006 to 74.6% in 2007. The mean number of inferential statistics per article was 1.3, ranging from 1 to 3. Statistical comparison of 3 or more groups (53 or 56.4%) was the single most commonly identified statistical test reported by the author(s), followed by comparison of independent 2 groups (40 or 42.6%) and paired t-test (9 or 9.6%) (Table 2). The quality of the analysis reported was less than optimal. Of the 94 articles in which statistical analysis was done 62 (or 66.0%) had at least 1 statistical error (Table 3). Counting each article once if multiple errors in a paper were identified, the percentage of errors in papers with Student's t-test, chi-square test, ANOVA, and regression were 53.1%, 25.0%, 47.6%, and 50.0%, respectively. Approximately half (41 or 43.6%) of the articles did not specify the analytical software and its version.

In reports with inappropriate statistical tests the most common errors were failing to apply or incorrectly applying unpaired tests for paired data or vice versa (Table 4). Statistical errors in chi-square test were failure to consider continuity-correction for small expected frequencies and inappropriate use of the test for continuous data. Common errors encountered in papers with ANOVA were failure to prove assumption of the test and failure to include or inappropriate post-hoc multiple-comparison. The most frequent correlation error was parametric Pearson correlation coefficient and over-fitting for small sample size. Other reporting errors were failure to state statistical methods used, failure to state the direction of hypothesis testing, wrong names for the statistical test, failure to state specific test if more than 1 test was done and misinterpretation of test results.

Table 3. Percentage of statistical errors identified in 2006 (n=41) and 2007 (n=53)

Type of analysis	2006	2007	Total
Articles with at least 1 error[1]	58.5	71.7	66.0
Student's t-test (independent, paired)	56.0	50.0	53.1
Chi-square (Fisher's)	33.3	20.0	25.0
ANOVA[2]	44.4	50.0	47.6
Regression (correlation)	0.0	50.0	50.0
Not specified analytical package	46.3	41.5	43.6

[1]Each article was only counted once even if multiple errors were identified. Thus, the percentage of articles with at least 1 error may be less than the sum of individual error types.
[2]Errors in multiple-comparison of ANOVA were not considered.

Discussion

This review demonstrated that statistical procedures were often applied incorrectly. The overall frequency (66%) of statistical errors in JVC article was higher than that observed in other studies; 54% in infection and immunity literature (10); 31.7% in obstetrics and gynecology literature (14); 38% in Nature articles (5); 25% in articles in British Medical Journal (5). In contrast, 71% in urology literature (13); 71% in Australian veterinary literature (9) are documented. However, the error rate in this study would be much underestimates in that the ability to assess the appropriateness of usage of statistics was severely limited in many articles due mainly to incomplete or incorrect descriptions of statistics or misinterpretation or unjustifiable interpretation of results within the article. This was not considered an error. In addition, the author focused on statistical methodology such as failure to account for basic assumption, sample size and data type for each procedure, not for the overall quality of the published articles. Further, inappropriate use of descriptive statistics and failure to post-hoc multiple-comparison in ANOVA were not included to calculate overall error rate. Accordingly, the proportion of errors would be much higher than the figure presented if taken all these limitations into account.

Repeated measures ANOVA would be appropriate in 17 articles, but the majority of papers had not specified the

Table 4. Statistical errors related to data analysis and documentation of methods applied in the Journal of Veterinary Clinics during 2006-2007

Category	Description
Errors in Student's t-test	Unpaired tests for paired data or vice versa
	Parametric test for small sample size
	Inappropriate use for more than 3 groups
	Failure to prove assumption of t-test
	Failure to examine normality of the data
Errors in chi-square test	No continuity-correction for small expected frequency
	Inappropriate use of the test for continuous outcomes
	Failure to use techniques to adjust for confounders
	No explicit explanation of the null hypothesis
	Failure to include trend test for ordinal variables
Errors in ANOVA	Failure to prove assumption of ANOVA
	Not consider dependence of data in repeated ANOVA
	Failure to include post-hoc multiple-comparison
	Inappropriate method of multiple-comparison
	Inappropriate use of test for ordinal data
	Inappropriate use of procedures for unequal sample size
	Too many comparison groups
	Parametric test for small sample size
Errors in regression	Parametric correlation for small sample size
	Incompatibility of test with type of data measured
	Failure to consider partial correlation coefficient
Others	Failure to state statistical methods used
	Failure to state the direction of hypothesis testing
	Use of wrong names for statistical test
	Failure to state specific test if more than 1 test was done
	Misinterpretation of test results
	Incomplete description of experimental design
	Not consider sample size or power for non-significance
	Conclusion without conducting a statistical test
	Not specified analytical package

detailed procedures for analysis. The major benefit of taking multiple samples from the same experimental unit is that variability among units can be reduced when each unit is compared to itself. In this case each measurement no longer independent, and therefore, statistical approaches on this type of data would be those applied to paired samples. Another important error is inappropriate use of t-test for more than 3 groups. This approach raises concerns over an increased likelihood of type I error (significance level, α), defined as the probability of wrongly rejecting the true null hypothesis (11,13). Assuming significance level of 0.05 per test and 3 tests are performed. The probability of at least 1 type I error could be 14.3%, almost 3 times higher than predefined error rate. This indicates that statistical significance can be obtained more frequently when in fact there is no difference. This error should be avoided whenever multiple-comparison is under consideration.

Nonparametric techniques were used in 7 (7.4%) articles, most commonly Mann-Whitney U test and the Kruskal-Wallis test. The Mann-Whitney U test and Wilcoxon rank sum test are identical but in a paper these terms were used as different statistical procedures. The central tendency and location of the data analyzed by nonparametric test should be reported as median or range not mean and standard deviation. One of the most common analytical errors is inappropriate use of parametric test for skewed data. Since many variables such as some hematological variables, somatic cell counts and titers in medical research are not normally distributed, these variables should be assessed on normality prior to applying parametric test (10,11). Based on the results whether the data shows normal distribution parametric on the transformed data or nonparametric test should be used.

Presentation of statistical procedures varied depending on the author(s); statistical procedures were not explained in the

methods section of the articles or no information at all despite the presence of p value in the text; procedures are explained only as footnotes to tables or figures. Few studies have provided information on the choice of parametric versus nonparametric test and on the distribution of normality of the data. Finally, the analytical software with version needs to be stated because this gives some information on the methods applied (5). Although evaluation of experimental design was not considered in this study, it was found that randomized block design instead of simple ANOVA would be appropriate in many instances because this design can include source of variation called a blocking factor, whose variation can be separated from the error variation to give more precise group comparisons.

Despite increasingly use of inferential statistics in the majority of papers reviewed incorrect or inappropriate use of them is also common, indicating that statistical tests should be reviewed more closely prior to editorial decision for publication. While errors stated above do not necessarily lead to faulty conclusions, it may threaten the overall validity of studies (8). Several guidelines on the use and reporting of statistics in many peer-reviewed journals are available (1,3,4,6). The authors believe that it is high time statistical guidelines for JVC publication enacted for potential author(s) to improve quality of papers, and educational efforts should be necessary as well.

Acknowledgement

This study was supported by the Institute of Veterinary Science, Kangwon National University.

References

1. Altman DG. Statistical reviewing for medical journals. Stat Med 1998; 17: 2661-2674.
2. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. BMJ 1983; 286: 1489-1493.
3. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. Ann Intern Med 1988; 108: 266-273.
4. Davidoff F, Godlee F, Hoey J, Glass R, Overbeke J, Utiger R, Nicholls MG, Horton R, Nylenna M, Hojgaard L, Kotzin S. Uniform requirements for manuscripts submitted to biomedical journals. J Am Osteopath Assoc 2003; 103: 137-149.
5. García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 2004; 4: 13.
6. Gardner M, Machin D, Campbell M. Use of check lists in assessing the statistical content of medical studies. BMJ 1986; 292: 810-812.
7. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the British Medical Journal. JAMA 1990; 263: 1355-1357.
8. Greenhalgh T. How to read a paper. Statistics for the non-statistician. I: Different types of data need different statistical tests. BMJ 1997; 315: 364-366.
9. McCance I. Assessment of statistical procedures used in papers in the Australian Veterinary Journal. Aust Vet J 1995; 72: 322-328.
10. Olsen CH. Review of the use of statistics in Infection and Immunity. Infect Immun 2003; 71: 6689-6692.
11. Pak SI. An assessment of statistical errors in articles in the Korean Journal of Veterinary Research. Korean J Vet Res 1999; 39: 1187-1196.
12. Porter AM. Misuse of correlation and regression in three medical journals. J Roy Soc Med 1999; 92: 123-128.
13. Scales CD Jr, Norris RD, Peterson BL, Preminger GM, Dahm P. Clinical research and statistical methods in the urology literature. J Urol 2005; 174: 1374-1379.
14. Welch GE, Gabbe SG. Review of statistics usage in the American Journal of Obstetrics and Gynecology. Am J Obstet Gynecol 1996; 175: 1138-1141.

한국임상수의학회지에 발표된 논문의 통계분석 검토

박선일[1] · 오태호*

강원대학교 수의과대학 및 동물의학종합연구소, *경북대학교 수의과대학

요 약 : 본 연구는 2006-2007년 한국임상수의학회지에 발표된 논문을 대상으로 자료 분석과 보고방법의 오류를 중심으로 검토하였다. 총 129편 중 94편이 적어도 한가지 이상의 통계분석을 수행하였으며, 분석기법으로는 세 집단 이상 비교 (53편, 56.4%), 두 독립표본 검정 (40편, 42.6%), 짝지은 표본 검정 (9편, 9.6%) 순으로 나타났다. 94편 중 62편 (66%)의 논문에서 적어도 한가지 이상의 통계적 오류가 발견되었다. 주요 오류로는 짝지은 표본에 대한 독립표본 검정, 세 집단 이상에 대한 t 검정의 반복, 카이제곱 검정에서 연속성 보정 무시, 분산분석에서 정규성 검토와 다중비교 방법 선택의 오류, 반복측정 자료에 대한 의존성 가정 무시, 통계분석 방법에 대한 부적절한 설명, 적용한 분석기법에 대한 구체적인 설명 부재 등으로 나타났다. 이러한 문제점을 개선하기 위해서는 학회차원에서 통계처리와 기술방법에 대한 가이드라인을 시급히 마련할 필요가 있을 것으로 사료된다.

주요어 : 통계분석, 오류.