

Statistical Errors in Papers in the two Korean Dermatology Journals

Sung Yul Lee, M.D., Kee Soon Nam, M.D., Jong Suk Lee, M.D.

Department of Dermatology, Chunan Soonchunhyang General Hospital, Soonchunhyang University College of Medicine, Chungnam, Korea

Background: Statistical errors have been noted in a large percentage of articles appearing in medical journals. Their incidence in the *Annals of Dermatology* and the *Korean Journal of Dermatology*, however, has not been studied.

Objective: The purpose of this study was to evaluate the accuracy of statistical methods in the *Annals of Dermatology* and the *Korean Journal of Dermatology*

Method: Original papers using the t test in the *Annals of Dermatology*(AD) and in the *Korean Journal of Dermatology*(KJD) from January 1990 to July 1994 were analyzed for correctness of statistical methods.

Results: Of the 376 original papers in the period considered, 78(20.7%) presented t tests and 43(55.1%) of these also contained some errors. Of the 9 papers from the *Annals of Dermatology*, 4(44.4%) included at least one error. Of the 69 papers from the *Korean Journal of Dermatology*, 39(56.5%) included at least one error.

Conclusion: Misuse of statistical methodology may not be uncommon in the two Korean Dermatology Journals, and it would be prudent to give more attention to statistical methodology (*Ann Dermatol* 8:(2)107~109, 1996).

Key Words : Statistical errors, *Annals of Dermatology*, *Korean Journal of Dermatology*

It is now widely accepted that statistical methods have an important role in the interpretation and communication of medical research findings. To serve any useful purpose, however, statistical analysis must be valid, and correctly undertaken. An inappropriate analysis can lead to wrong conclusions and may also lend a false respectability to the conclusions in the eyes of an unwary reader¹. The first important analysis of the quality of statistics to appear in medical journal was published in the *Journal of the American Medical Association*². This article, based on a survey of publications in 10 different journals, suggests that, from a

statistical viewpoint, 41(28%) of 149 analytic studies were acceptable, 7(5%) should have been rejected, and 101(68%) should have been revised before publication. Since this report, a large number of studies³⁻⁷ on the quality of statistics in medical journals have been carried out.

A study has been undertaken to discover the frequency of errors in the two Korean Dermatology Journals. The results of the study are discussed here and it is hoped that the points raised may be of help to other researchers preparing their work for publication.

MATERIALS AND METHODS

The 10 issues of the *Annals of Dermatology* (AD) and the 30 issues of the *Korean Journal of Dermatology*(KJD) from January 1990 to July 1994 were chosen for scrutiny. In total, 46 original papers from the *Annals of Dermatology* and

Received July 6, 1995.

Accepted for publication January 4, 1996.

Reprint request to : Sung Yul Lee, M. D., Department of Dermatology, Chunan Soonchunhyang General Hospital, Soonchunhyang University College of Medicine, Chungnam, Korea

330 original papers from the Korean Journal of Dermatology are included. All papers were reviewed for statistical analysis and then 78 original papers which used the t test (9 from the Annals of Dermatology and 69 from the Korean Journal of Dermatology) were selected. The t test is the most popular statistical test in biomedical research. Therefore we selected papers which used the t test and analyzed these for the correctness of the statistical method used. We defined three statistical error categories and made these as objective as possible. The categories are listed in Table 1. Although design considerations are highly relevant to the value of a study, criticisms of the choice of design were thought to be beyond the scope of this paper and so were not included.

RESULTS

Of the 376 original papers in the period considered, 78(20.7%) presented the t test and 43(55.1%) of these also contained some errors (Table 1). Of the 9 papers from the Annals of Dermatology, 4(44.4%) included at least one error from one or more of the three categories. Of the 69 papers from the Korean Journal of Dermatology, 39(56.5%) included at least one error (Table 1).

DISCUSSION

This survey reveals three abnormal trends. First, the t-test is often wrongly used as a comparative statistic for more than two groups. Where there are more than two groups it is inadvisable to use repeated t-tests between each possible pair, as this increases the likelihood of finding a significant difference by chance. If 5 groups are being compared, there are 10 possible comparisons between

pairs of means, and with 10 groups there are 45. If there are no real differences, the probability of at least one being statistically significant at the 5 percent level can be considerably greater than 0.05. Thus the likelihood of erroneous claims is increased. The appropriate procedures for dealing with this problem are known as the analysis of variance with multiple comparisons⁸. Second, the t-test is used as a comparative statistics in ordinal data. Ordinal data represent a somewhat stronger level of measurement than do nominal. Ordinal data usually exist when a ranking or scaling system has been used to subgroup individuals or responses. Numbers can easily be assigned to each of the categories on an ordinal scale. The actual value of the number is unimportant. Note that the distance between subgroups need not be equal when an ordinal scale or classification system is used. One must recognize that t-tests are intended for use only on data of the interval level. They should not be used when the measurement has reached only ordinal level. Where there are ordinal data it is inadvisable to use the t-test. Ordinal data require a non-parametric test such as the Mann-Whitney U-test⁹. Third, the t-test is often used as a comparative statistics in repeated measurement data. Where there are repeated measurement data it is inadvisable to use the multiple paired t-test. Repeated measurement data require repeated measure analysis of variance (ANOVA)⁹.

A total of 78 papers used the t-test and 43(55.1%) of these also contained some errors. Of the 9 papers from the Annals of Dermatology, 4(44.4%) included at least one error from one or more of the three categories. Of the 69 papers from the Korean Journal of Dermatology, 39(56.5%) included at least one error. Perhaps because

Table 1. Numbers of papers with errors, according to error categories, out of 78 presenting the t-test

Error categories	Number		
	AD ^a	KJD ^b	Total
More than two groups	3	18	21
Ordinal data	0	7	7
Repeated measurement data	1	14	15
Total	4	39	43

a: Annals of Dermatology(Seoul)

b: Korean Journal of Dermatology

it is such a well-known method, some researchers are careless in its application and do not check that the assumptions underlying the test are even approximately satisfied¹. Parametric tests, such as the t-test and the analysis of variance, are based upon assumptions that the data from both groups has been drawn from normally distributed populations and that these populations have the same variance⁹. 78 original papers were performed with the t-test but all 77 papers except for one paper with a large sample size neglected these two important assumptions.

One assumption of the t-test is that the distribution of the population is normal. The test is still satisfactory('robust') for moderate departures from normality, but in small samples from skewed distributions it is preferable to transform the data or to use a non-parametric test⁹. For example, ordinal data require a non-parametric test such as the Mann-Whitney U-test. Interval data from a skewed distribution can sometimes be made more nearly normal by applying a transformation(e.g. log, cube, square...) to the values¹.

The researcher may expect in advance that the variances of the two populations will differ or this may be discovered after the data have been collected. In either case the Student' t-test should not be used without question as it assumes equal variances. For the unpaired t-test the variances of observations in the two populations should be equal, although some difference may be acceptable if the samples are large and of approximately the same size. The alternatives are to use a different form of t-test with the Satterthwaite or Welch approximation to the degrees of freedom, or to use a non-parametric test¹.

This study has shown that papers recently published in the journal contain a considerable number of statistical errors. The proportion of papers found to have committed some error(55.1% of 78) is similar to that found in a survey of the British Medical Journal³(53 percent of 62). How can these errors be prevented? Remedies suggested have included more consultation with medical statisticians and elementary instruction in statistical methods during clinical training³. These two actions will force medical investigators to learn enough elementary statistics to design their exper-

iments and analyze their data correctly and to recognize cases that require help from a professional statistician. A further possibility is of more extensive statistical refereeing of papers submitted to the journal¹. This would at least prevent the worst errors from appearing in print and need not unduly delay publication. Consulting statisticians may be helpful, but, most of the errors are misapplications of elementary techniques that the investigators themselves should learn to use correctly. Most professional(and even amateur) statisticians are not especially interested in grinding out garden-variety statistics for other people. More importantly, with rare exceptions, no one can do a better job of analyzing the data from a clinical protocol or experimental procedure than the investigator.

REFERENCES

1. White SJ: Statistical errors in papers in the British Journal of Psychiatry. *Brit. J. Psychiat* 1979; 135; 336-342
2. Schor S, Karten I: Statistical evaluation of medical journal manuscripts. *JAMA* 1966, 195; 145-150
3. Gore SM, Jones IG, Rytter EC: Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J* 1977, 1; 85-87
4. Emerson JD, Colditz GA: Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med* 1983, 309; 709-713
5. Gardner MJ, Altman DG, Jones DR et al: Is the statistical assessment of papers submitted to the British Medical Journal effective? *Br Med J* 1983, 286; 1485-1488
6. Pocock SJ, Hughes MD, Lee RJ: Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med* 1987, 317; 426-432
7. Murray GD: The task of a statistical referee. *Br J Surg* 1988, 75; 664-667
8. Glantz SA: Biostatistics: How to detect, correct and prevent errors in the medical literature. *Circulation* 1980, 61; 1-7
9. Elenbaas RM, Elenbaas JK, Cuddy PG: Evaluating the medical literature part 2: Statistical analysis. *Ann Emerg Med* 1983, 12; 610-620