ORIGINAL ARTICLE

Assessment of errors and misused statistics in dental research

Jee Soo (Jay) Kim[1,2], Dong-Kie Kim[2] and Suk Jin Hong[3]

[1]*Yonsei University College of Dentistry, Seoul, South Korea;* [2]*Chosun University School of Dentistry, Gwangju, South Korea;* [3]*Chonnam National University School of Dentistry, Gwangju, South Korea.*

Objectives: The goal of the study is to assess the level of misused statistics or statistical errors in dental research, and identify the major source of statistical errors prevalent in dental literature. **Methods:** A total of 418 papers, published between 1995 and 2009, was randomly selected from 10 well established dental journals. Every paper in the sample underwent careful scrutiny for the correct use of statistics. Of these, there were 111 papers for which we were unable to judge whether or not the use of statistics was appropriate, due to insufficient information presented in the paper; leaving 307 papers for this study. A paper with at least one statistical error has been classified as 'Misuse of statistics', and a paper without any statistical errors as 'Acceptable'. Statistical errors also included misinterpretation of statistical analytical results. **Result:** Our investigation showed that 149 were acceptable and 158 contained at least one misuse of statistics or a statistical error. **Conclusion:** This gave the misuse rate of 51.5%, which is slightly lower than that reported by several studies completed for the medical literature.

Key words: *Misused statistics, ordinal data, descriptive statistics, hypothesis test, significance level, P-value*

Advancing technology has enabled us to observe and quantify a variety of phenomena that occur in dental and medical laboratories and clinics allowing the investigators to collect prodigious data. It is difficult for anyone to make good sense of a confusing and chaotic array of raw data by visual inspection alone. The data must be properly processed and analysed in a technically sound and meaningful way to uncover the hidden clues. Methods of statistical analysis are absolutely indispensable and are powerful tools providing the techniques for drawing objective and scientific conclusions for the investigators.

Statistics is an integral part of the dental and medical curricula. An important aspect of students' training is to develop the ability to critically read literature in their specialist areas, and correctly interpret the published research findings so that they can implement the results in their practice for the benefit of the patients[1].

As reported by Dawson-Saunders and Trapp, many published scientific articles have shortcomings in statistical design and analysis[2]. Statisticians have long been aware of widespread misuses and abuses of statistics in scientific literature, Altman being one of the first statisticians to bring the issue to the public[3].

Statistics are required at every phase of dental science research from the beginning with an experimental design, a sampling technique, and collection of mean-

ingful and valid data to the ensuing data analysis and interpretation of the statistical results. The improper use of statistical method and data analysis could lead the investigators to inappropriate conclusions that could be clinically detrimental. In most circumstances the wrong dental or medical treatments are easily detectable and recognised even by the individuals with no background in dentistry or medicine, while statistical mistakes can go unnoticed for a long time. Statistical errors are subtle, technical and difficult to detect while the impact of statistical errors is not easy to measure. Although the damage caused by a statistical mistake can be catastrophic, it may not be easy to pin the responsibility on such an error.

In 1966, Schor and Karten reviewed 295 papers published in 10 medical journals and concluded that only 28% of the papers were statistically acceptable, about 68% were deficient and 5% were 'un-salvageable'[4]. Gore *et al.* reported that of 62 analytical reports published in *The British Medical Journal* 32 (51.6%) included at least one type of statistical error category[5]. Kuo reviewed 178 articles published from January to June, 2000 in the *BMJ*, *JAMA*, *The Lancet* and *The New England Journal of Medicine*. His study revealed that almost 60% of the articles involved a statistical error[6]. McGuigan examined 164 articles published in a psychiatry journal and reported that 'serious' statistical

errors were found in 40% of the articles[7]. Welch and Gabbe's investigation showed that in 19% of 145 articles published in an obstetrics and gynecology journal contained 'serious' statistical errors which could have led to misleading conclusions[8]. Gardner and Bond undertook an exploratory study of statistical assessment of papers submitted to *BMJ*. Of the 45 papers, only five (11%) were considered statistically acceptable at submission, but increased to 38 (84%) after a revision and the subsequent publication[9]. Felson *et al.* evaluated two groups of *Arthritis and Rheumatism* papers; one group of papers published in 1967–1968 and the other published in 1982[10]. The study showed that statistical misuse occurred in 60% of the papers in the first group and 66% in the second group[10].

To our knowledge, no similar studies have been done regarding the prevalence of misused statistics or statistical errors in dental literature. It is important to emphasise that the aim of our study is not to pillory any particular journal paper or its authors. The primary goal of this paper is to assess the level of statistical mistakes that are believed to be widespread in dental literature, and highlight several common mistakes that are easily rectifiable.

MATERIALS AND METHODS

The total of 418 papers, published between 1995 and 2009, was randomly selected from the following 10 well established journals in dental sciences: *American Journal of Orthodontics and Dentofacial Orthopedics*, *Angle Orthodontists*, *Clinical Oral Implant Research*, *The International Journal of Oral & Maxillofacial Implants*, *The Journal of American Dental Association*, *Journal of Endodontics*, *The Journal of Prosthodontics*, *Journal of Periodontal Research*, *Journal of Periodontology* and *Pediatric Dentistry*.

Case reports and papers void of statistics were excluded from sampling. Every paper in our sample of 418 manuscripts underwent careful scrutiny for the correct use of statistics. Of these there were 111 papers for which we were unable to judge whether or not the use of statistics was appropriate, due to insufficient information presented in the paper. After removing the 111 undetermined papers from further consideration 307 papers were left for evaluation. A paper with at least one statistical error was classified as 'Misuse of statistics', while a paper without any apparent statistical error was termed 'Acceptable'. Statistical errors also included misinterpretation of statistical results. Some papers committed multiple errors. We did not count the number of statistical errors committed in these papers, nor did we tabulate the specific types of statistical mistakes. The primary goal of our study was to simply count the number of papers that contain at least one misuse of statistics mimicking the previous investigators who reported the misuse rates for medical journals, as discussed above. Our secondary goal was to compare the misuse rates between dental and medical journals.

RESULTS

From the sample of 307 papers, our investigation revealed that 149 were acceptable and 158 contained at least one misuse of statistics, yielding a misuse rate of 51.5%. A little more than a half of the papers published in the established dental journals committed at least one misuse of statistics from minor to major. Some papers failed to identify the statistical tests that were performed. We did not consider this as a misuse of statistics. In fact, if the other statistical methods used in the paper were sound and correct, the paper was judged to be acceptable. The extent to which statistical procedures were used by the authors in these papers was broad and diverse, from simple descriptive statistics to sophisticated applications of survival analysis techniques. *Table 1* summarises the misuse rate of statistics reported by various investigators by year and journal examined.

In addition, we counted the number of times certain statistical methods have been used in the 307 papers we reviewed. For this count we did not make an attempt to distinguish between the acceptable paper and those with misused statistics. *Table 2* illustrates the frequency distribution of the statistical methods applied by the authors of these papers in our study sample.

Table 1 The misuse rate of statistics for different journals

Authors	Year	Study journals	Sample size	Misuse rate (%)
Schor & Karten	1966	10 medical journals	295	72
Gore *et al.*	1977	British Medical Journal	62	51.6
Felson *et al.*	1984	Arthritis and Rheumatism (published in 1967–68)	47	60
		Arthritis and Rheumatism (published in 1982)	74	66
Gardner & Bond	1990	British Medical Journal	45	16*
McGuigan	1995	British Journal of Psychiatry	164	40*
Welch & Gabbe	1996	American Journal of Obstetrics and Gynecology	145	19*
Kuo	2002	BMJ, JAMA, Lancet, NEJM	178	60
Kim *et al.*	2010	10 dental journals	307	51.5

*Only the 'serious' errors were considered. The authors did not specify what constitutes the 'serious' error.

Table 2 Statistical methods used in the sampled papers

Statistical methods	No. times used
Descriptive statistics (mean, median, SD, percentile, range and coefficient of variation)	384
Correlation analysis (Pearson correlation, Spearman rank correlation, Kendall's partial rank correlation)	41
t-test (one-sample, two-sample, paired)	129
Confidence interval	20
ANOVA (one-way, two-way, three-way, multiple comparison procedures)	124
Regression analysis (simple linear, multiple, logistic regression)	33
Categorical data analysis (Fisher exact test, rxc contingency table, measures of association; Cochran–Mantel–Haenszel test, McNemar test, Kappa statistic, ICC)	51
Nonparametric statistics (Mann–Whitney *U*-test, Wilcoxon signed rank test, Kruskal–Wallis test, Friedman test)	112
Survival analysis	55
Others (rates and proportions, ANCOVA, longitudinal analysis)	9
Total	958

Major source of misused statistics

We enumerate common misuses of statistical concepts we found from our assessment of the published papers[11]. If these errors were prevented, the misuse rate would be substantially lower.

Selected sample, not random sample

Nearly every paper that presents survey data begins by stating '… based on the responses from a random sample of 186 subjects …'. In survey research it is not easy to obtain a random selection of respondents. The subjects decide whether or not to respond and return the survey. The samples are thus self-selected. There is a difference between a self-selected sample and a random sample. Self-selected samples are often known to be biased. Therefore, survey research requires special attention because of the potential harm that misleading results can cause.

Mean and standard deviation of ordinal data

Virtually every investigator assigns the values 1, 2, 3, 4, and 5 to the five pain categories which are known as a 5-point Likert scale: 1 = No pain, 2 = Mild pain, 3 = Slight pain, 4 = Severe pain and 5 = Extremely severe pain. With few exceptions, the investigators report the mean and SD of the ordinal data, calculated from the assigned values. The numbers 1, 2, 3, 4, and 5 are completely arbitrary and are nothing more than convenient labels for the purpose of data analysis. The numbers have absolutely no sensible quantitative meaning. We may have five different social security numbers, instead of 1, 2, 3, 4, and 5, and assign the smallest social security number to 'no pain' and the largest social security number to 'extremely severe pain'. Because the categories are not quantifiable, algebraic operations make no sense for nominal and ordinal data. Thus, the average and the standard deviation based on the arbitrary numeric assignments have no meaning at all, and should not be calculated.

Pearson correlation coefficient for ordinal data

To evaluate a statistical relationship between two variables or two questions in a survey questionnaire, investigators often mistakenly utilise the Pearson correlation coefficient when the data are ordinal. Suppose patients were asked to respond to the following two survey items:

Q1. I have anxiety associated with a dental treatment: Strongly disagree, Disagree, Neutral, Agree and Strongly agree.

Q2. How much pain did the treatment cause you?: No pain, Mild pain, Slight pain, Severe pain, Extremely severe pain.

We may assign −10.4, −7.8, 16, 28.5 and 41, instead of 1, 2, 3, 4 and 5 to the five pain categories. Different labelling systems give rise to totally different values of the Pearson correlation coefficient for the identical data. This statistical absurdity occurs because Pearson correlation is calculated from the arbitrary label values that have absolutely no quantitative meaning. Pearson correlation coefficient is a measure of the degree of linear association between two continuous variables. The Pearson correlation coefficient for ordinal data should be avoided. In this case the Spearman rank correlation is an appropriate statistical measure to use.

The one-sample t-test, two-sample t-test, and paired t-test

One of the most common misuses of statistical methods in the dental research papers was the application of the *t*-tests and paired *t*-test when the samples are from non-normal populations. For a valid application of these test methods the measurements or outcome responses must be normally distributed. Numerous papers reported the statistical results from the *t*-tests to compare two groups or pairs of responses within the same subjects when the data are ordinal.

Chi-squared contingency table

Chi-squared contingency table is a widely used method to study statistical relationships between two variables. Using the two survey questions above, with five choices for a response, the related contingency table has five

rows and five columns yielding 25 cells in the table, and thus, is referred to as a 5×5 contingency table. The name chi-squared in a chi-squared contingency table comes from the test statistic for an analysis having an approximate chi-squared probability distribution. To ensure a reasonably good approximation two technical requirements must be satisfied:

- No more than 20% of the cells have an expected cell frequency <5.
- No cells have an expected cell frequency <1.

To satisfy these requirements it is necessary to have a sufficiently large sample size, i.e., a large number of respondents. We have come across a number of papers which violated the above conditions.

Simple linear regression analysis

The simple linear regression model is an area of inferential statistics which explores the nature of a relationship between two quantitative variables so that one variable (the outcome variable) can be predicted from the other (the explanatory variable or independent variable). As discussed above, the numbers assigned to the pain categories have no quantitative meaning. There were papers that have erroneously used the ordinal or even nominal measurement scales as explanatory variables. Often ignored is an underlying assumption that the dependent variable (outcome variable) is normally distributed with the same variance for each value of the explanatory variable. A typical example is a regression model in which the VAS (visual analogue scale) scores are being predicted from the 5-point Likert scale. The investigators mistakenly assumed that the pain categories are quantified by the values 1, 2, 3, 4 and 5. Such regression models make little sense and could be very misleading.

One-way ANOVA

One-way analysis of variance (ANOVA) is an extremely useful technique and a frequently applied statistical procedure to compare three or more treatment groups. This is one of the most egregiously misused statistical methods not only in dental but also in other scientific research. Suppose three treatments are to be compared with respect to their effectiveness. The following three conditions must be satisfied for the proper use of one-way ANOVA:

- The treatments are normally distributed
- Their variances are equal
- The error terms are independent.

These conditions clearly imply that one-way ANOVA is an inappropriate statistical procedure to apply to ordinal data. The most prevalent misuse of one-way ANOVA occurred when there was a large discrepancy among the variances.

DISCUSSION

Our review of the papers for this study revealed that countless authors stated 'There is no statistically significant difference between the control and experimental groups ($P > 0.05$).'

There are no misused statistics in this statement. However, it must be stressed that '$P > 0.05$' implies that the P-value of the test is anywhere between 0.05 and 1.0, a huge range of values. It is advisable that a specific P-value be provided, say $P = 0.6572$. It is critically important to state the level of significance for the test as this would determine whether or not to reject or accept the null hypothesis at the given level. We found many papers with confusing statements such as: 'The degree of apical leakage from the teeth prepared by laser was not significantly less than that from control teeth ($P > 0.01$),' with no indication of the level of significance. If the significance level were specified at $P = 0.05$ and the P-value were 0.04, the above conclusion would be wrong.

To our knowledge, this study is the first of its kind to assess the level of statistical misuses and abuses in dental literature. The misuse rate of 51.5% for dental literature, based on the 307 papers published over the period 1995–2009 in 10 dental journals, is comparable to the study performed by Gore, *et al.* for medical literature[5]. Compared to some studies done for medical journals (*Table 1*), it appears that the misuse rate in dentistry is lower than that in medicine. We deliberately have not made any attempts to perform a significance test between the dental and medical misuse rates, as the criteria for misuse may differ from study to study. For example, we did not consider the papers with no indication of the statistical test method performed as 'misuse of statistics', but other authors for the medical research may have.

The misuse of statistics in the data analysis may lead to erroneous conclusions and make the research findings difficult to replicate. It should be noted that in certain circumstances both the correct and incorrect statistical tests lead the investigators to the same conclusion. But of course, the P-values will be different. Though the conclusions may be the same, the improper use of statistics in research endeavour will not only negatively impact the value of the findings, but greatly damage the credibility of the investigators involved and the most seriously, it will tarnish the reputation of the entire dental community and dental science. Unfortunately, the consequences of research conclusions based on misused statistics are slow to show their effects.

Why the rampant misuse of statistics?

We may ask why so many investigators in dental research misuse statistical concepts. How do we explain this prevalent misuse of statistics?

• No licence is required to practice statistics.

To practice dentistry, medicine, nursing, physical therapy, law, accounting, plumbing, etc., one must earn a licence to practice his/her chosen profession. However, statistics is not a licensed profession. There is no system in place to validate the competence of an individual who performs the statistical analysis. Anyone can claim to be a statistician.

• Detection of statistical errors.

Unlike dentistry and medicine, statistical mistakes can go unnoticed for a long time, and the impact of statistical mistakes is not easy to quantify. The errors in statistics are more subtle and technical, and therefore, they can easily escape reviewers' attention.

• Availability of statistical software packages.

In the last few decades we have seen amazing development and advances made in statistical software packages. Upon entering data into a worksheet, a few clicks can instantly produce impressive looking outputs, such as ANOVA tables, regression equations, graphs, P-values, etc. As convenient as a software package is, it can be a double-edged sword. Statistical packages do not tell us whether or not one-way ANOVA procedure is appropriate to apply to ordinal data. They are programmed to faithfully produce an output, regardless. The statistically less sophisticated users of a software package tend to have blind trust and complete confidence in the output the package provides.

CONCLUSION

We took a random sample of 307 articles published in 10 dental journals during 1995–2009. These papers were reviewed and examined for the proper use of statistical methods. The result of our study showed that a 51.5% of the dental journal papers contained at least one statistical error.

Acknowledgements

REFERENCES

1. Kim JS, Dailey RJ. *Biostatistics for Oral Healthcare*. Ames, IA: Wiley-Blackwell Publishing; 2007.
2. Dawson-Saunders B, Trapp RO. *Basic and Clinical Biostatistics*. 2nd ed. Norwalk, CT: Appleton & Lange; 1994.
3. Altman DG. Statistics and ethics in medical research. *Br Med J* 1980 281: 1182–1184.
4. Schor S, Marten I. Statistical evaluation of medical journal manuscripts. *J Am Med Assoc* 1966 195: 1123–1128.
5. Gore S, Jones I, Rytter E. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *BMJ* 1977 278: 85–87.
6. Kuo Y. Extrapolation of correlation between 2 variables in 4 general medical journals. *J Am Med Assoc* 2002 287: 2815–2817.
7. McGuigan S. The use of statistics in the *British Journal of Psychiatry*. *Br J Psychiatry* 1995 167: 683–688.
8. Welch G, Gabbe S. Review of statistics usage in the *American Journal of Obstetrics and Gynecology*. *Am J Obstet Gynecol* 1996 175: 1138–1141.
9. Gardner M, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *J Am Med Assoc* 1990 263: 1355–1357.
10. Feldon DT, Cupples LA, Meenan RF. Misuse of statistical methods in *Arthritis and Rheumatism*. 1982 vs. 1967–1968. *Arthritis Rheum* 1984 27: 1018–1022.
11. Kim JS. Misused statistics in dental research and challenges of teaching statistics. *Loma Linda University Dentistry Alumni Magazine* 2010 21: 11–16.

Correspondence to:
Dr Suk Jin Hong,
Professor of Preventive Dentistry,
Chonnam National University School of Dentistry,
Gwangju, South Korea.
Email: sjhong@chonnam.ac.kr