Avoiding negative reviewer comments: common statistical errors in anesthesia journals

Sangseok Lee

Department of Anesthesiology and Pain Medicine, Sanggye Paik Hospital, Inje University College of Medicine, Seoul, Korea

Manuscripts submitted to journals should be understandable even to those who are not experts in a particular field. Moreover, they should use publicly available materials and the results should be verifiable and reproducible. Readers and reviewers will want to check the strengths and weaknesses of the research study design, and ways to make this determination should be clear through proper analysis methods. Studies should be described in detail so as to help readers understand the results. Statistical analysis is one of the key methods by which to do this. The inappropriate application of statistical methods could be misleading to readers and clinicians. While many researchers describe their general research methods in detail, statistical methods tend to be described briefly, with certain omissions or errors or other incorrect aspects. For instance, researchers should describe whether the median or mean was used, whether parametric or non-parametric tests were used, whether the data meet the normality test, whether confounding factors were corrected, and whether stratification or matching methods were used. Statistical analysis regardless of the program should be reported correctly. The results may be less reliable if the statistical assumptions before applying the statistical method are not met. These common errors in statistical methods originate from the researcher's lack of knowledge of statistics and/or from the lack of any statistical consultation. The aim of this work is to help researchers know what is important statistically and how to present it in papers.

Key Words: Biomedical research, Research, Statistical analysis, Statistical errors, Statistics.

Introduction

Statistics is necessary at all steps in a study so as to obtain scientifically accurate and reliable results. Statistical analysis should not be neglected in clinical studies in that the inappropriate application of statistical methods severely damages research ethics. Improperly designed and calculated studies can represent a waste of time and funding intended for the study. Too small a sample size may not lead to significant results, whereas a too large sample size runs the risk of harming subjects and causing them discomfort [1]. On the other hand, incorrect conclusions may be derived from even a sufficiently well designed study if the results are improperly reported or misinterpreted. This article intends to help researchers to emphasize and appropriately report statistically significant results when planning, conducting, and reporting a study.

Common Mistakes in Statistical Methods

The International Committee of Medical Journal Editors

(ICMJE) established in 1978 on the basis of Vancouver Group prepared the document entitled 'Uniform Requirements for Manuscripts Submitted to Biomedical Journals' (1979) and later revised it a few times. The revision published in 1988 included not only statistical methods and instructions for describing results but also guidelines with respect to the principles of the application of essential statistical methods to which researchers should conform. Most medical and scientific journals in Korea, including the *Korean Journal of Anesthesiology* (KJA), provide independent instructions to authors, referring to the Uniform Requirements. Nevertheless, statistical errors are strikingly common in medical articles; Altman and Bland [2] estimated that more than 50% of medical reports published at that time included statistical errors. On the other hand, an analysis of 164 articles published in British psychiatry journals showed that 40% of articles included statistical errors [3]. Articles published in Korean journals are not very different. Ko et al. [4] analyzed KJA articles from Vol. 1 in 1981 to Vol. 6 in 1990, and reported that statistical errors were included initially in 97% of the articles and in about 67% of the articles published later. Ahn [5] analyzed KJA articles published in five years starting in 1994 and reported that 60% of the articles included various types of errors.

Types of errors vary and occur in all types of statistical analysis; however, certain types of errors are commonly found when analyzed by researchers. Glantz [6] analyzed all of the original articles published in the journal entitled *Circulation* and reported that the most common statistical error is the inappropriate use of a t-test for a multi-group hypothesis test. This result is consistent with another report which found that the most common statistical error occurred with data to which an ANOVA or paired t-test should be applied but which were tested with Student's t-test [7]. According to work by Olsen [8], of the approximately 141 articles published in the journal *Infection and Immunity*, 54% included statistical errors. These were found in the detailed statistical methods applied (20%), in the descriptions of the statistical results (22%), or in both (12%). An assessment of the articles published in the Journal of the Korean Medical Association in the 1980s showed that 97.8% of the analyzed articles included one or more statistical errors, including insufficient descriptions of the statistical power of the test method and confidence interval (91.9%), duplicated testing due to the incorrect statistical method applied (65.2%), insufficient descriptions of the statistical method itself (58.2%), and deductions from unreasonable statistical conclusions (52.2%) [9]. In addition, a parametric test method was often applied to a variable for which normality is doubted, or analytical results were omitted without an appropriate explanation despite being described in the methods section. Most of these statistical errors were commonly observed in articles published in the KJA.

Tips for Avoiding Negative Reviewer Comments on Statistics

Consulting about statistics should start with the planning and design of the study

Study planning and design are among the most important steps in research. Errors or mistakes generated at these stages have a significantly negative effect on the validity and reliability of the research results. Therefore, it is critical to reduce statistical errors by positively accepting advice on statistics from the stage of designing a study. A review by the Ministry of Food and Drugs Safety (MFDS) showed that a total of 796 errors were found per 100 clinical trial protocols published in 2012.[1] The most common errors were found in the statistical methods used (35.3%); followed by the research sample size calculation (26.8%), test planning errors (19.4%), and endpoint errors (18.6%). Common problems raised in this review included definitions of the research hypothesis and the primary validity assessment endpoint, the choice of the primary analysis subjects, the basis of calculating the research sample size, and insufficient descriptions of covariate variables. Good studies come from good research designs. Efforts should be made to prevent errors through statistical consulting from the research design stage. The initial study plan should describe the objective, hypothesis, the endpoint measured to test the hypothesis, and the statistical method applied, which should be also included in the final research report or article. A study is a series of procedures to generate, prove, or dispute a hypothesis. A study objective should be specifically described as an answer to a question which is intended to be obtained by conducting the study. To prove a hypothesis, researchers should clearly establish measurement variables such as a primary endpoint and a secondary endpoint and specifically describe the statistical methods used to analyze the variables [10].

What type of study should I conduct?

The first step is to choose the study type which may best support the desired conclusion. Results obtained from an inappropriate study type are less precise in terms of their estimation power. Each study type has its own pros and cons. Randomized controlled clinical trials are the most powerful study type in medical research, but they are associated with high costs and considerable investments in time. The Statistical Round article introduced methods for properly designing and reporting randomized controlled clinical trials [10]. On the other hand, well planned observation studies require less time and cost. Cross-

---

[1] Clinical Statistics Fact Sheet (2012), Ministry of Food and Drug Safety.

sectional studies provide a snapshot of a disease or condition at a specific time point; thus, attention should be paid to the inference of disease progression on the basis of the results obtained from these studies. Questionnaire survey studies, when properly conducted, enable clinicians to understand current manipulations and concepts. In questionnaire surveys and observation studies, the choice of control group and pairing should be described in detail. The direction of a study (retrospective, cross-sectional, prospective) should be clearly noted. When applying results from one study to another general population through extrapolation, a proper choice of research subjects and a high participation rate (response rate) are emphasized. Case-series studies should be used only to highlight the need for future planned studies.

Sample size calculation errors and ethics

Appropriate calculation of the sample size is essential for the cost-effective, effort-effective, and ethical implementation of a study. This increases the number of opportunities to observe the expected effects. Calculation of the sample size is directly related to research ethics. Registration of too many uncalculated participants in a study can expose the subjects to unidentified risks. Too small a sample size is also unethical in the sense that the power of test of the study is decreased, thus limiting the scientific value of the research. Consequently, patients can be harmed by incorrect clinical decision-making based on incorrect study results. Power of test is determined by the sample size, the magnitude of the type I error (α), and the effect size. These values are interrelated. With an increase in the significance level (type I error), that is, with worsening reliability, the power of test level is increased. As the standard deviation increases, the power of test decreases. A smaller difference between two populations decreases the power of test, while a larger sample size increases the power of test. Out of these, the effect size is the most critical factor with regard to the power of test [11]. It is recommended to use a "clinically meaningful empirical effect size" as an effect size. For example, the correlation between two variables, the regression constant in a regression analysis, the difference in mean values, or a risk such as the ratio of heart attack survivals to heart attack casualties can be used as an effect size. If there is no available information about the effect size, an effect size reported from previous observational studies may be used [11]. Establishing a clinically meaningless or preposterous effect size to reduce the sample size is obviously erroneous and unethical.

The sample size should be calculated using the primary endpoint. When there are multiple endpoints, the type I error due to multiple tests should be calibrated to estimate the sample size, as the number of hypotheses to be proved may increase, eventually equaling the number of primary endpoints. Bonferroni correction and Šidák correction are generally performed. If a study has a secondary endpoint which is important, the sample size should be sufficiently large for the analysis of the variables. In this case, the sample size may be ideally calculated for the individual endpoints which are considered important. Multiple comparison which is not appropriately corrected has been reported as one of the most commonly discovered statistical errors (multiple comparison error) [6]. Below is one example of a multiple comparison error. Assume that an experiment is planned to compare the effect of two drugs regulating blood sugar levels. The researcher measured the blood sugar level in three patient groups: a control group, the drug A group, and the drug B group. In such a case, three t-tests are usually performed to compare the control group with the drug A group, the control group with the drug B group, and the drug A group with the drug B group. However, special attention is necessary with this type of analysis. This example is a typical case of false positive generation by multiple comparison error. Because attempts were made to test three hypotheses at a significance level of 5% through one experiment, the significance level needs to be corrected. However, given that no correction was performed in this case, the applied significance level was 15% (5% × 3). If Bonferroni correction is applied, a significance level of 5% / 3 = 1.7% should be used for each test. Adjusting the power of test level after performing an experiment is very difficult. Therefore, it is important to consult with a statistical expert to examine if the study plan has sufficient power of test before starting to collect data. Another problem is that researchers tend to use a smaller sample size than planned for reasons which are to be used in upcoming article publications and/or conference presentations.

Errors associated with study methods and the applied analytical method

A study article should be elaborately described to secure verification and reproducibility so that the readers and reviewers may easily understand it. In particular, when a complicated statistical method, i.e., not a common method (e.g., a t-test), is applied, an additional explanation should be given as well as references, providing information about the application of the statistical method. Readers may easily understand if a brief explanation is provided about why a specific statistical method, not a general one, has been applied.

It is important to apply an analytical method appropriate for the type of data constituting the measured variables. Generally, there are three types of research data: discrete, ordinal, and continuous. Discrete type data represent the quality or group, not the quantity, and are also classified as nominal data or qualitative data. This type of data represent, for example, the gender (male and female), the anesthetic method (general, re-

gional, and local anesthesia) or the location (Seoul, Busan, and/ or Daejeon), and are often used for grouping. The second data type is ordinal, including data representing the order and/or rank. Examples include test scores expressed in ranks (first rank, second rank, and third rank), height ranks, and weight ranks. Raw data as well can represent ranks; for example, grades (e.g., A, B, and C) are ordinal-type data. An error may be made by treating ordinal-type data as continuous-type data, or becoming confused with the figures representing the orders. The figures representing ranks may not undergo arithmetic manipulation. Finally, there are continuous-type data having a quantitative meaning. Test scores, weights, and heights are included in this type of data. Continuous-type data are the data type most suitable for statistical analysis, but continuous-type data are often dichotomized (i.e., divided into two or more separate domains) to simplify the analysis in some studies. For example, in a study related to obesity, the weights of patients are measured, but they may not be used as continuous-type data, instead being divided into two groups entitled 'normal weight' and 'overweight.' Such a conversion of continuous-type data into dichotomized data may enable a comparison of two groups with simple statistics such as a t-test instead of a complex regression analysis. However, the problem in such a case is that the measurement precision of the original data is decreased, as is the variability of the data, resulting in a reduction of the information included in the data and the power of test in the study. Moreover, most researchers do not apply common boundaries or cut points when dividing data. Therefore, to dichotomize continuous-type data, a researcher should explain why the data need to be dichotomized despite the sacrifice of data precision, as well as how the cut points were established.

It is ironical that one of the causes of errors made by researchers originates from statistical software programs, which typically help with statistical analyses. Errors from statistical software are often made when researchers use the software without consulting with a statistics expert or without obtaining sufficient statistical knowledge. Some researchers use convenient methods to analyze data and calculate P values without sufficiently considering the data characteristics or statistical assumptions. Once a significant P value is secured, researchers believe that their results are valid. However, it is important to bear in mind that statistical software programs always give a P value regardless of the sample size, data type and scale, or statistical methods used. Various analytical methods applied to statistics are based on the fundamental statistical assumptions. If an analysis is performed without satisfying the fundamental assumptions, incorrect conclusions may be made on the basis of erroneous analytical results.

One common error is that a nonparametric method is not applied in cases where the data are severely skewed, not follow-ing a normal distribution. When analyzing continuous-type data, a normality test should be performed with the analyzed data, and the method and result of the test should be described. The t-test, which is generally used with continuous-type data, is a parametric method which can be applied only when normality, equal variance, and independence are tested and satisfied. If these statistical assumptions are satisfied, the author may state the following:

"*The data were approximately and normally distributed and thus did not violate the assumptions of the t-test.*"

If data which do not satisfy these assumptions were analyzed by a nonparametric method, the author may state the following:

"*The number of subjects was small and normality was found to be contrary to the normal distribution test results. A nonparametric test was performed using the Wilcoxon rank-sum test.*"

For appropriate understanding and application, parametric tests and nonparametric tests were discussed in two articles of the KJA Statistical Round in detail [12,13]. Another common error with t-tests is that Student's t-test, rather than a paired t-test, is performed to analyzed paired independent samples.

In an analysis of the results with categorical-type data, Fisher's exact test or asymptotic methods with appropriate adjustments should be used if the event is rare and the sample size is small. A standard chi-squared test and a difference-in-proportions test may be performed, provided that the number of samples and the number of events are sufficiently large. Data for which both rows and columns are dichotomous, an extreme type of discontinuous data, follow a complex distribution consisting of a product of two conditional probabilities (a binomial distribution), which approximately follows a chi-square distribution if the number is sufficiently large. Because such data are basically discontinuous, continuity correction is necessary in the approximation into a continuous chi-square distribution. Although controversial among statisticians, an approach with a direct probability calculation such Fisher's exact test is more feasible if the results from Pearson's chi-square test and Yate's correction differ. In addition, if the expected frequency of at least one of the four cells is less than 5, Fisher's exact test should be used.

A correlation analysis is a method of analyzing the linear relationship between two variables. The calculated correlation coefficient represents the measure of the degree of linearity between two variables. If the correlation between two variables is more curved rather than linear, the correlation coefficient may be very small. In contrast, when some observation data are positioned very differently from the rest, the correlation coefficient may be great. Neither of these cases represents a proper analysis. Hence, it is necessary visually to examine the data distribution using a scatter plot before performing a correlation analysis. A correlation coefficient merely represents the degree of correlation between two variables; it does not explain a causal

relationship. 'Correlation' does not necessarily mean that the two variables are in a cause-and-effect relationship; rather, it is simply one of the conditions of a cause-and-effect relationship. Nevertheless, researchers often make the "post hoc, ergo propter hoc" mistake, in which a temporal relationship between two independent variables is considered as a causal relationship, leading to the erroneous conclusion of "B occurred after A; therefore, B occurred due to A." For example, a researcher observed yearly Coke sales trend as well as yearly drowning casualties. A strikingly high correlation was found in the correlation analysis of the two variables. Can the researcher make the conclusion that "the number of drowning casualties is increased because of Coke"? Before believing a research result, researchers should initially check if the result is in accordance with common sense. In this example, the real cause of the increase in the two variables is not between the two variables but is a third cause, which is the summer season. Generally, when there is correlation between A and B, a few more interpretations are possible, besides the third cause mentioned above. For example, "B may be the cause of A" (reverse) or "A is the cause of B and B is also the case of A at the same time" (interactive), and "They occurred at the same time coincidentally without any causal relationship." A correlation itself is not an implication of a causal relationship but is simply one of the necessary conditions of a causal relationship. A correlation analysis is better used as a method of producing a hypothesis rather than testing one, and should be accepted as a proposal of a follow-up study to identify a causal relationship. An additional test should be performed to identify a causal relationship between variables through a well-planned experiment, which is a randomized controlled trial. Equivalence of experimental groups is employed to prove the existence of a causal relationship statistically. Samples are randomly taken from two or more groups and then allocated to a study group and a placebo or control group, making the two groups as homogenous as possible. If the effect by the treatment is greater than the effect by the placebo treatment (greater than a predetermined effect size), it may be concluded that the treatment has a causal effect.

Regression analysis is an analytical method which is used to derive a mathematical relationship which expresses a correlation between an independent variable and a dependent variable. Regression analysis may explain a correlation between two variables and make a statistical prediction through an established model. While correlation analysis refers to the identification of a correlation between two variables, regression analysis serves to calculate the contributions of the correlations of multiple independent variables with a single dependent variable (multiple regression analysis). One error commonly found in medical research papers is that regression analysis is used without clearly showing the necessary statistical assumptions. The simple linear regression analysis, a typical form regression analysis, requires of

a model the basic assumptions that a dependent variable and an independent variables should have a linear relationship, and that mutually independent error terms should have a mean value of 0 (zero) and should be in equal in terms of variance and be normally distributed. In addition, the absence of multicollinearity among variables should be assumed. The linear relationship between two variables may be visually determined through a scatter plot. Violation of the basic assumptions of a linear regression equation may be determined on the basis of a residual scatter plot.

Satisfaction of statistical assumptions is a prerequisite of a statistical analysis. Data analysis without satisfying these assumptions can raise questions about the reliability of the results and severely damage the reproducibility of the research. Repeated-measures analysis of variance, which is often used in articles submitted to the KJA, requires various statistical assumptions to be satisfied before the analysis, as in the regression analysis mentioned above, but most of the articles omit an explanation of the necessary assumptions, instead simply providing only the analytical results [14,15]. Finally, detailed information about the computer software programs used for the statistical analysis should be provided in the article. If the raw data are provided, it will help readers or reviewers who want to reproduce the results. It should be noted that the same statistical model may give different results depending on the statistical software program used.

Presentations & Interpretations of Results

As mentioned above, a research article should include a detailed description of applied statistical methods. Access to raw data enables readers and peer reviewers to test the results contained in the article. Many scientists report that reproducing experiments is the most important part of scientific advancement. This type of reproduction allows for the filtering of false positives. According to Pitkin [16], a review of the validity of the descriptions of statistical methods in abstracts published in six prominent medical journals, including the *British Medical Journal*, *The Journal of the American Medical Association*, and the *New England Journal of Medicine* showed that the statistical results described in the abstracts were different from or were not mentioned in the main text in 18 to 68% of the reviewed articles. Because most readers judge the results and values of studies through abstracts before reading the full-text version, this review result may not be regarded as a mere mistake. It is herein emphasized that correctly describing the results is as important as appropriately performing the statistical analysis. When two or more analytical methods are applied, detailed descriptions should be provided about the data set applied to each of the analytical methods. It is not enough simply to say "*where appropriate.*"

In the description of the results, the standard deviation or standard error of mean is used along with the mean in order to explain the data distribution pattern. However, the standard deviation and or standard error of the mean are often confused with each other and are interchangeably used. Moreover, some articles do not mention which is which. Standard deviation is used to explain the characteristics of samples, which are the center of a normal distribution and a varied distribution, whereas the standard error of the mean represents the estimate (mean) and the precision of the estimate with respect to the population. The standard error of mean is decreased as the sample size increases. Some researchers obtain significant results by increasing the sample size and thus decreasing the standard error of the mean, which is unethical. In addition, because the standard error of the mean is usually smaller than the standard deviation, some researchers intentionally present only the standard error of mean of the data. The previous KJA Statistical Round also discussed the differences between the standard deviation and the standard error of the mean as well as proper interpretations of both [17].

Most research journals, including the KJA, use P < 0.05 (or P < 0.001) to indicate the significance of the results. Results that are not significant have been presented with the description P > 0.05. However, such a description does not allow further interpretation. Specific P values should be provided such that readers can judge on the basis of the individual critical values or cut-off values. However, given that it is difficult intuitively to understand results only with P values, using a confidence interval (Equations 1 and 2) is recommended to provide more information, as follows:

95% confidence interval of population mean:

$$\overline{X} \pm 1.96 \times \frac{s}{\sqrt{n}} \qquad (Eq.\ 1)$$

95% confidence interval of population proportion:

$$p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}} \qquad (Eq.\ 2)$$

The confidence interval is the sum of an estimate and the uncertainty accompanied by the estimate, representing the uncertainty of the research conclusion. The confidence interval represents the range of values in which unknown parameter values of the population derived from the sample statistical quantities may be included. While the P value is difficult to interpret and clearly conveyed, the confidence interval may complement such shortcomings. When the entire confidence interval includes the clinically significant range, the treatment performed in the study may be concluded to have been clinically effective. When the entire confidence interval is out of the clinically significant range, the treatment is concluded to have been clinically ineffective. In addition, when some part of the confidence interval is out of the clinically significant range, a clinical conclusion should be withheld considering that the sample size may not be sufficiently large.

The significance level itself does not represent the probability that the study hypothesis is true. In addition, a P value of less than 0.05 does not indicate that a conclusion is incorrect at a probability of 5%. A P value is not a measure of effect size. A similar P value does not mean a similar effect size. Many researchers have long misinterpreted the P value. To correct these year-long customs in academic areas, the American Statistical Association eventually published a statement on significance levels, in 2016 (Table 1) [18]. The reader can refer to the previous KJA Statistical Round to find information about the proper understanding and utilization of significance levels [19].

A rejection region, which is a region in which a null hypothesis is rejected, is determined as the range of the significance level value. A two-tailed test or one-tailed test can be performed depending on the location. Except in the case where a one-tailed test is required because the alternative hypothesis indicates a direction of difference (small or large) (e.g., a non-inferiority test), all significance levels should be obtained by a two-tailed test. The P value should be described to three decimal places (and not as "P < 0.05"). If the P value is less than 0.001, it should be described as "P < 0.001." Scientifically significant figures should be used to describe the results. A calculated or estimated value may not have a significant figure at a decimal point lower than that of the original measurement. Some articles list unnecessarily accurate figures to interfere understanding [20].

A randomized clinical trial should be reported according to

Table 1. Six Principles for Using P-values

1. P values can indicate how incompatible the data are with a specified statistical model.
2. P values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a P value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A P value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a P value does not provide a good measure of evidence regarding a model or hypothesis.

Adopted from the American Statistical Association (ASA) statement on P values.

guidelines such as CONSORT, which includes a flow diagram and a checklist and which clearly states the types of information that should be included in an article for reproduction of its experiment. Details of the guideline can be found on the CONSORT website [21]. A number of scientific journals have already required conforming to the CONSORT reporting guidelines before authors submit a manuscript. Editors of the journals check whether the relevant information is included in the submitted manuscripts. Missing data will allow readers to doubt the quality of the study and the accuracy of the presented results. Readers are also requested to remember the guidelines when reading scientific articles.

Conclusions and Recommendations

Statistics is an essential methodology for medical research and is the basic language by which medical knowledge is acquired. However, a number of medical research articles are published which nonetheless contain statistical errors (Table 2). Most clinicians have a very low level of statistical knowledge, mainly because the medical school curriculum and the internship and residency for medical specialists do not provide a systematic statistical education. Editorial boards have identified these problems and emphasized the need for peer reviews on the statistics included in submitted manuscripts. The KJA is undertaking pioneering work by employing statistics-specialized editors and regularly publishing statistical educational materials (the Statistical Round) for the education of the society members, journal readers, and authors. Moreover, the KJA is currently preparing new instructions for authors, for which the specific instructions about statistical methods and description are elaborated and intensified. Providing specific and clear author instructions may improve the quality of the articles and reduce the effort and time required of the reviewers. The publication of error-free, high-quality, and well-written articles will help the KJA to develop into one of the world's prominent journals. We herein thoroughly discussed the statistical errors that are often found in articles submitted to the KJA. The objective of the present article is neither to criticize the quality of the research articles nor to blame the contributors for their unethical and insincere aspects. We hope that the present article will make a small contribution to the production of statistically better works by reviewing frequently found errors and mistakes.

Table 2. Summary of Common Errors in Statistics

\<Statistical errors in the design of a study\>
Error in choosing the research type that can best prove the conclusion
Unclear descriptions of study objectives, hypotheses, and variables measured to test the hypotheses
Absence of hypotheses description
Error in sample size calculation
Absence of a description of the effect size
Inaccurate description or missing description of a randomized trial
Insufficient description of a blind study
Missing information about the homogeneity between compared groups with respect to basic characteristics
\<Statistical errors in data analysis\>
Application of analytical methods which are inappropriate for the type of data
Unnecessary dichotomization of continuous-type data
Error in the application of parametric/non-parametric test methods
Basic statistical assumptions unchecked
Generation of type I error: multiple comparison error, with corrections not implemented
Exact test or continuity correction not implemented with categorical data having a small sample size
Misinterpretation of correlation as a causal relationship
\<Errors in the documentation, presentation and interpretation of statistical analysis\>
Absence of a detailed description of each statistical method applied to each data set
Omission of two-tailed/one-tailed test information
Reason for applying an unusual statistical method and a detailed explanation of the method not given
Incorrect names of statistical methods
Confusing the standard deviation with the standard error of mean or not mentioning which is which
Providing results with only the significance level without mentioning the confidence interval
Significance level presented as 'P = NS' or 'P < 0.05'
Misinterpretation of 'insignificance' as 'ineffective' or 'no difference'
Not considering the possibility of type II errors when reporting insignificant results
Making conclusions not derived from the results
Not reporting missing data
Nonconformity to the CONSORT reporting requirements

References

1. Altman DG. Statistics and ethics in medical research. Misuse of statistics is unethical. Br Med J 1980; 281: 1182-4.
2. Altman DG, Bland JM. Improving doctors' understanding of statistics. J R Stat Soc Ser A 1991; 154: 223-67.
3. McGuigan SM. The use of statistics in the British Journal of Psychiatry. Br J Psychiatry 1995; 167: 683-8.
4. Ko H, Kwak IY, Kim KW, Ham BM, Choe IH. Statistical methods in the articles of the Journal of the Korean Society of Anesthesiologists from 1981 to 1990. Korean J Anesthesiol 1993; 26: 22-7.
5. Ahn W. Statistical methods in the articles in the Korean Journal of Anesthesiology Published from 1994 to 1998. Korean J Anesthesiol 2000; 39: 706-11.
6. Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation 1980; 61: 1-7.
7. Wang Q, Zhang B. Research design and statistical methods in Chinese medical journals. JAMA 1998; 280: 283-5.
8. Olsen CH. Review of the use of statistics in infection and immunity. Infect Immun 2003; 71: 6689-92.
9. Lee HK, Ahn YO. An assessment of methodological and statistical validity of medical articles published in Korea, from 1980, to 1989. Korean J Med Educ 1991; 3: 52-69.
10. Lee S, Kang H. Statistical and methodological considerations for reporting RCTs in medical literature. Korean J Anesthesiol 2015; 68: 106-15.
11. Nahm FS. Understanding effect sizes. Hanyang Med Rev 2015; 35: 40-3.
12. Kim TK. T test as a parametric statistic. Korean J Anesthesiol 2015; 68: 540-6.
13. Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. Korean J Anesthesiol 2016; 69: 8-14.
14. Park SI, Lee DK, In J. Statistical review of 95 studies employing repeated-measures analysis of variance published in the Korean Journal of Anesthesiology. Korean J Anesthesiol 2016; 69: 97-9.
15. Lee Y. What repeated measures analysis of variances really tells us. Korean J Anesthesiol 2015; 68: 340-5.
16. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. JAMA 1999; 281: 1110-1.
17. Lee DK, In J, Lee S. Standard deviation and standard error of the mean. Korean J Anesthesiol 2015; 68: 220-3.
18. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat 2016 [Epub ahead of print]. Available from http://dx.doi.org/10.1080/00031305.2016.1154108
19. Park S. Significant results: statistical or clinical? Korean J Anesthesiol 2016; 69: 121-5.
20. Lang T. Twenty statistical errors even you can find in biomedical research articles. Croat Med J 2004; 45: 361-70.
21. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001; 357: 1191-4.