Basics, common errors and essentials of statistical tools and techniques in anesthesiology research

Sukhminder Jit Singh Bajwa

Department of Anaesthesiology and Intensive Care, Gian Sagar Medical College and Hospital, Banur, Patiala, Punjab, India

Abstract

The statistical portion is a vital component of any research study. The research methodology and the application of statistical tools and techniques have evolved over the years and have significantly helped the research activities throughout the globe. The results and inferences are not accurately possible without proper validation with various statistical tools and tests. The evidence-based anesthesia research and practice has to incorporate statistical tools in the methodology right from the planning stage of the study itself. Though the medical fraternity is well acquainted with the significance of statistics in research, there is a lack of in-depth knowledge about the various statistical concepts and principles among majority of the researchers. The clinical impact and consequences can be serious as the incorrect analysis, conclusions, and false results may construct an artificial platform on which future research activities are replicated. The present tutorial is an attempt to make anesthesiologists aware of the various aspects of statistical methods used in evidence-based research and also to highlight the common areas where maximum number of statistical errors are committed so as to adopt better statistical practices.

Key words: Confidence interval, research studies, statistics, study design, type-I error, type-II error

Introduction

Statistical methods have become an inseparable part of the modern anesthesiology research. The evidence-based anesthesia research and practice has to incorporate statistical tools in the methodology right from the planning stage of the study itself. Though the medical fraternity is well acquainted with the significance of statistics in research, there is a lack of in-depth knowledge about the various statistical concepts and principles among majority of the researchers. These partially ignorant and inappropriate practices make the results, and observations suffer from numerous errors and statistical limitations.[1-7] Biostatistics is one such specialty subject which is taught minimally at graduate and postgraduate levels, and majority of the researchers are unaware of its significance either. As such, it becomes extremely difficult for the researchers to choose appropriate statistical method to analyze the research results. The clinical impact and consequences can be serious as the incorrect analysis, conclusions, and false results may form an artificial platform on which future research activities are replicated. As a result, many patients are exposed to a higher risk of drug effects and techniques which were inadequately tested during the original study. Further, there are no comprehensive set of universal guidelines which monitor the application and analysis of statistical methods in various research studies. Evidence of wrong reporting of statistics has been cited numerous times in literature by different medical specialties.[1-7] The scope of the present article covers various aspects of statistical methods used in anesthesiology research as well as an attempt to encompass a descriptive review of the various errors committed at different stages of the study.

Types of Research Study

It is extremely difficult to elaborate on the descriptive methodology of designing a study in one short tutorial. However, the present article aims to discuss in brief the various important aspects of statistical tools and techniques which are helpful for designing the study in a concise and appropriate manner. Research in anesthesiology practice mainly involves randomized clinically controlled trials, cross-sectional studies, case-control studies and rarely longitudinal studies. A general classification of research studies can be stated as [Figure 1]:

Address for correspondence: Dr. Sukhminder Jit Singh Bajwa,
House No-27-A, Ratan Nagar,
Tripuri, Patiala, Punjab, India.
E-mail: sukhminder_bajwa2001@yahoo.com

Research Studies

Observational
(non-experimental)
studies

Intervention
(experimental)
studies

– *Cohort studies*
– *Case-control studies*
– *Cross-sectional surveys*
– *Routine-data-based studies*
individual level
aggregated level

– *Clinical trials*
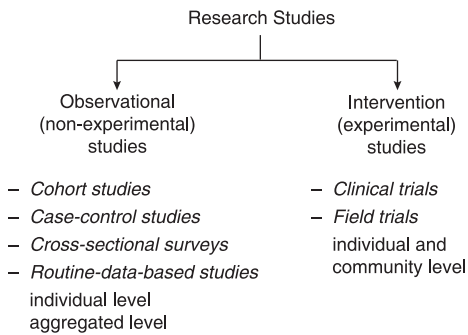– *Field trials*
individual and
community level

Figure 1: A general classification of epidemiological research studies

Cohort (longitudinal) studies

These are observational studies which observe the variables over time and measure the incidence of diseases in a more precise manner. These studies are also called prospective studies as the data is acquired in a prospective manner. For example: These types of study designs are commonly used for intensive care and postoperative patients.

Case-control studies

In case-controlled studies, the disease forms the basis of sampling rather than the exposure status. The population in whom some disease is present constitutes the cases and the other group in which no disease is present forms the control group. These studies are based on the principle that controls represent the population at risk of the disease. In unmatched case-control studies odds ratio is used to analyze the binary outcomes as a measure of association starting with 2×2 tables and progressing to Mantel-Hanzel methods and logistic regression to control the effect of confounding variables. In matched case-control studies, the matching of cases is done with controls having same values for analyzing the confounding variables. McNemar's test is used to measure the paired proportions while multivariate analysis is based on the linear logistic model which is employed for the analysis of case-control studies with pair-wise matching.

Cross-sectional studies

These are descriptive studies which are carried out for a short duration period and measures the clinical features of patients at just one point in time. They differ from case-control studies in that they observe data of the entire population under study whereas in case-control studies only population with specific selected characteristics is studied. As such the nature of these studies makes them eligible not only for measuring the odds ratio, but also to measure the absolute and relative risks from prevalence.

Randomized controlled trials

Randomized controlled trials (RCT's) are considered to be important as they provide the best evidence of the various anesthetic techniques, drugs and interventions. The various elements of RCT's include randomization scheme, allocation concealment, double blinding of the study where both the investigator and the participant are blind to the nature of the study and intention to treat analysis. RCT's are also useful for establishing the adverse effects of the drug. The most significant merit of choosing RCT's include elimination of allocation bias thus achieving a balance among both known and unknown prognostic factors. However, the external validity of RCT's may be limited by certain demographic characteristics such as geographical areas, patient characteristics, study procedures and outcome measures. Furthermore, these trials can prove to be expensive and sometimes take years to get complete in all aspects. In a bid to improve study design of RCT's, Consolidating Standards of Reporting Trials guidelines came into existence in 1996 which have been revised from time to time. In addition, the various other aspects related to RCT's, meta-analyses and diagnostic studies are illustratively summarized in QUOROM and STARD statement.[8-12]

Methodology of Study Design

The main aim in describing the study design in the methods section is to primarily make the reader understand that whatever the information is being conveyed will possibly help in replicating the methodology in their own respective settings. In any research study, it is important to specify in the beginning only about the primary outcomes currently being studied and also to study the secondary outcomes. Therefore, the information pertaining to various aspects should be completely given in the materials and methods section which include but is not limited to:
- Aims and objective of the study.
- Hypothesis to be tested and the null hypothesis should be mentioned.
- The size of the groups and the number of patients selected.
- Process of randomization and concealment of allocation groups.
- Process of blinding.
- Sources and demographic profile of the patients.
- Type of surgery and anesthesia.
- The dose and method of drug administration if a drug is selected.
- Details of the technique (if techniques are compared).
- Inclusion and exclusion criteria.
- Various parameters and techniques to assess the parameters to be observed.
- Different scales with references if used in the study.
- Study design to be specified whether, cross-sectional, prospective and retrospective and so on.
- Sample size estimation methodology.

- All the statistical methods and tests should be described in this section.

Lack of stringent universal statistical guidelines is one of the major contributory factors for the inappropriate use of statistical tools by the medical researchers.[2-6] As a result, the standard of statistical analysis and application has not improved, and errors are committed frequently.[4,5,13,14] During drafting of the study, any error, limitation, shortcoming or some basic flaw can lead to the generation of nonreliable and weak research conclusions. The aims and objectives of the study should be thoroughly evaluated and formulated on the basis of hypothesis being tested and statistical tools to be used.[15]

Types of errors
Type-I error
In statistical terminology, type-I error pertains to a wrong decision when a test rejects a true null hypothesis and is also known as an error of the first kind. It can be compared with the occurrence of false positive in test situations. It is represented with Greek letter α (alpha) and denotes the probability of type-I error.

Type-II error
A type-II error pertains to a wrong decision when a test fails to reject a false null hypothesis and is also known as an error of the second kind. A type-II error may be compared with the so-called false negative in other test situations. It is represented with Greek letter β (beta), and it forms the complement of the power of the test.

The goal of the test is to determine if the null hypothesis can be rejected. A statistical test can either reject (prove false) or fail to reject (fail to prove false) a null hypothesis, but never prove it true (i.e., failing to reject the null hypothesis does not prove it true).

Sample size estimation
Sample size estimation is very crucial to determine the significance and impact of the outcome. Conventionally chosen alpha and beta errors are arbitrary and have come to be used by tradition rather than any scientific validity. Furthermore, these statistical tools and techniques should be chosen individually for each research question. However, going by tradition, a small sample size may not be able to detect the true difference in the study which can be termed as false -ve or type-II or β-error. The maximum amount of false -ve results should be 20% for any sample population studied and keeping the test of significance (i.e., $P > 0.05$). This computation gives the power of the study which can be simply expressed as $1 - \beta$ or 80% or more for detection of true differences in the variables studied. Though a large sample size may be appropriate to diminish the type-II error, it increases the cost of the entire project and also delays the completion of the research activities in a stipulated time period. In addition, large sample size may not adhere to the estimated costs of the project and can result in undue delay in the completion of the research study. Choice of particular statistical test is governed by few important factors such as comparison of mean or percentages, the number of study groups, type of data, paired or unpaired data and the distribution of data.[16-19]

Comparison of characteristics and parameters
The blinding of the research activity ensures nonbiased results and observations.[20] The process of randomization and sampling should be elaborated in the material and methods section so as to eliminate any bias during data collection which is an essential part of the research methodology.[1,21] While selecting the groups, comparability factors that are specified in the inclusion criteria should be chosen strictly so as to minimize the differences and errors in results obtained.[21,22] These differences in results can be further minimized by application of multivariate analysis during computation of the results.[23] The errors in statistical tests are easily remedied, if the raw data is available, but it requires a re-analysis. The comparison of demographic and other attributes in the study and control group may show insignificant differences but for validating the comparison, calculating the statistical power of the study can help in achieving the accurate results in a small study group.[24] It is, therefore, essential that during the study designing, the sample size calculation, participants withdrawing from the study, clear description of the null hypothesis, description of the randomization process, methods of blinding, appropriate selection of study and control group and appropriate selection of statistical tests for comparing the baseline characteristics are to be formulated in clear and elaborative manner.

Application of statistical tests
This is another potential area where maximum number of errors are encountered during validation of the observations during research. The type of the statistical test applied for a particular data should be clearly mentioned.[13,25] Any vague statement regarding the application of various statistical tests such as "wherever applicable" or "where appropriate" should always be avoided.[25]

Ignorance about the correct application of even simple tests such as Chi-square and *t*-test leads to widespread misuse of these tests.[13,21,26] The small numerical values may yield incorrect results on application of Chi-square test. The comparison of multiple groups mandates application of analysis of variance. Variations in the study group or presence of confounding factors should be rectified at the earliest by multivariate techniques.[27]

Common errors encountered during statistical application include but are not limited to:

- Choosing wrong test for a particular data.
- Choosing a wrong test for the proposed hypothesis.
- Falsely elevated type-I error during *post-hoc* significance analysis.
- Inappropriate use of Chi-square test when numerical value (NV) in a cell is <5.
- Failure to apply Yates' continuity correction to the Chi-square test especially when the number analyzed is small.
- Unevenly matched group size for Student's *t*-test.
- Application of unpaired *t*-test for paired data.

In fact, one of the major problems regarding Student's *t*-test is the extreme imbalance rather than minor imbalance. In a simple randomization method, which can be termed as best scheme if the size of the sample is large, the chances are few that one can get equal numbers in all the groups. Block randomization is recommended in studies of small sample sizes so as to ensure allocation of equal numbers. However, adoption of this method does not guarantee that equal numbers will be followed up to complete data collection.

Reporting of data

Normally, numerical data are expressed as mean ± standard deviation (SD). While ordinal data are preferably summed up as median and interquartile range at a minimum. Percentages are used to express the nominal data and are a part of inferential tests which gives the value of P after the test is applied. $P < 0.05$ is considered significant while $P > 0.05$ as nonsignificant. However, it is important to calculate and display the 95% confidence intervals around any estimated spot percentages. It is highly recommended that exact observed values be reported rather than mentioning $P <$ or > 0.05 or P as $<$ or > 0.0001. The reporting data should be precise with regards to various qualitative tests whether it may be the proportion, the correlation coefficient or mean value. Reporting of $P > 0.05$ as nonsignificant may also obscure the results and as such it is not recommended. Percentages should also be reported up to one-decimal point only. For a small sample size, the reporting up to even one decimal point is not needed. However, one can express the values of t, χ^2 and r to two decimal places.

Parametric and nonparametric tests

The assumptions which are formulated at the beginning of the study provide a base on which analysis is pertaining to the distribution of variables can be performed. Data can be either normally distributed, or it can have variable distribution for which either some transformation before the analysis is required. The data analysis should be preceded by a detailed and thorough description of the variables measured during the

study. Specific variables which are important for the study should preferably be described in detail in order to validate the statistical analysis and hypothesis being tested.

The illustration with various methods such as tables, graphs, figures, scatter diagrams, pie chart and histograms is of immense significance. The underlying assumptions should indicate whether the data collected has a normal distribution, or the distribution is highly skewed. If the data is asymmetrical or highly skewed in distribution, the application of nonparametric tests such as Mann-Whitney U-test is mandatory. Highly skewed observations are difficult to analyze statistically and needs mathematical transformation so as to precisely analyze the observed parameters with available statistical tests.

During statistical analysis, significant differences among various study groups (more than 2 groups) mandate application of ANOVA and *post-hoc* significance testing for multiple comparison. The nonparametric tests are to be applied on ordinal and nominal data and include but are not limited to Mann-Whitney U-tests, Wilcoxon, Kruskal-Wallis and Freidman test. They can also be applied to numerical data as well if the distribution of the observed values is not normal.

The statistical significance of the difference of means in study groups should be measured by *t*-test or ANOVA. The association between variables can be measured statistically by using Chi-square or Fisher's exact test. Risks and outcome association can be studied using odds ratio, risk ratio and number needed to treat analysis. For measuring the correlation between variables, correlation analysis can be performed with Spearman correlation, Karl Pearson and correlation coefficient. Log-rank test, Kaplan-Meir's curve, Mantel-Hanzel test and cox proportional hazard can be used to observe the difference in the occurrence of an event over a period, time, term. The "*t*-test" should not be used for data that is not normally distributed. It is invariably observed that for comparing the mean between two groups, multiple *t*-tests are used. When the value of minimal data is <5 in a cell, Fisher's exact test should be used instead of Chi-square test. Spearman rank tests should be used for data that is not normally distributed (non-Gaussian distribution) instead of using Karl Pearson correlation.

Confidence intervals

In studies where mean differences are estimated with an emphasis on relative risk, it will be prudent to measure confidence intervals in these studies. The measurement of confidence intervals is hugely linked to results of the hypothesis being measured in that study. The basic underlying fact is that measurement of confidence interval allows near accurate measurement of the observations being studied. The more is the width of the confidence interval; more are the chances of

the information being inappropriate. This should be calculated with great care as there are chances of over-interpretation of the results and observations especially in studies of small sample size. During reporting of confidence interval, more stress should be exercised with regards to reporting of differences between the groups under study.

Paired and unpaired data

A common error is made during the computation of paired and unpaired data. It is necessary for the measurements of two different groups that unpaired observations should be distinguished — for example, patients receiving alternative therapeutic regimens — from that of paired observations, when the comparison is done between two measurements made on the same individuals at different time intervals. For unpaired data, two sample t-test, Mann-Whitney U-test and Chi-square test are useful whereas for paired data the common paired t-test, Wilcoxon test and McNemar's test are used.

Statistical Expression of Results

The results of various statistical tests should be described in a descriptive manner in the results section. Whatever data is presented, it should clearly convey the various statistical measures of central tendency and dispersion. If the distribution of data is skewed, usage of median and quartile range will be more appropriate along with mentioning of measures of variability.[15,26] All the symbols and abbreviations related to statistics should be explained during the first appearance in the text.[28,29] The description of confidence interval is very important for all the primary and main results.[1,22] The application of confidence interval largely overcomes many weaknesses in the study with the measurement of the difference between groups.

Reporting data with precision

The reporting and expression of numbers, especially numerical values should be reported after rounding off the digits to improve the expression.[30] Quite often an error is encountered when continuous data is expressed in ordinal category thereby compromising with precision of data presentation.[31] Paired data represents data from the same patient, and it has a tendency to get hidden when group mean values are reported for various patients. Descriptive statistics for continuous variables is commonly presented as mean and SD in a normal or Gaussian distribution of value [Figure 2]. SD can be expressed as plus or minus 1, 2 and 3 for 68%, 95% and 99% of the normal distribution while for nonnormal distribution, median and interquartile range are better options.[32] Limitation of mean and SD include inappropriate estimation during small sample study when the biological data is not normally distributed.[33] Standard error of the mean is a measure of precision as
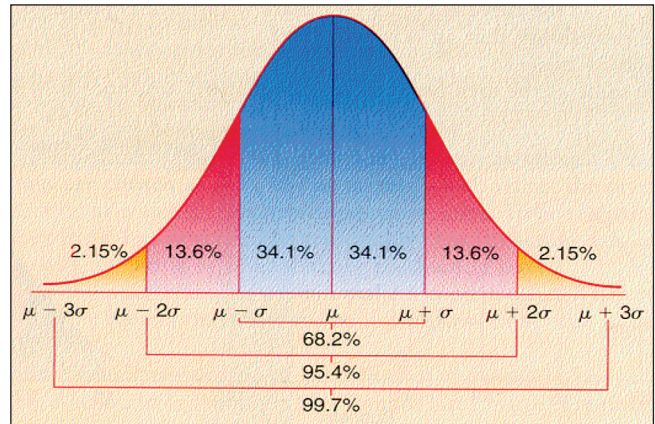

Figure 2: Showing a bell-shaped curve of normal (Gaussian) distribution of the data

compared to SD and will always have a smaller value than SD. Hence, as to make results look more precise.

The complex statistical tests require an explanation and should be stated with appropriate reference. The more important variables require a detailed description as the main outcome of the study is dependent on them and can be expressed by bar charts, graphs, scatter plots or histograms. Reporting of proportions, such as American Society of Anesthesiologists grading, is an appropriate method of reporting qualitative data.

Correlation analysis

Correlation coefficient is one of the most measured statistical entities. The correlation matrix is a useful adjunct to express the correlation between the variables when the number is significant. Rank correlation should be used instead of Pearson product moment correlation for the data which is variable in distribution, for the variables that are constrained to be above or below certain values and when the relation between the variables is not linear. Regression analysis should not be confused with correlation, and both cannot accompany each other. The most difficult situation to compute is the presence of more than one outcome variables which may require the adoption of multivariate techniques for statistical analyses. These multivariate techniques are very difficult to teach to the readers of the anesthesiology articles.

Presenting standard deviation, SE and the numerical value

While presenting the results, it should be ensured that mean values should be quoted with some measures of variability or precision.[34] The variability can be expressed with SD while precise measurement can be expressed with a standard error of the mean. Instead of using symbols (±) SD and SE should simple be expressed as (SD/SE NV). Similarly, confidence interval should be given as NV to NV rather by using symbols such as ±. The denominator should be clearly expressed whenever percentages are used.

Discussion Section

However, it is extremely difficult for a researcher or academician to go through the entire book of statistics for his or her dedicated research. Therefore, the present tutorial is an attempt to outline the designs, errors and application of various statistical tests in a brief manner during the research activity. Though numerous software are available on the internet for analyzing the results, but still majority of researchers are unaware or ignorant to the fact that which set of statistical tests are appropriate for the type of data collected.[35,36] One possible alternative to this problem is the employment of a biostatistician for every journal and editor should ensure that highest quality of statistical reporting is carried out. These steps and checking can be done during the peer review stage of the article where one can add statistical review stage. This allows the biostatistician to have a deeper look at the various mathematical observations.[37] The ideal situation is to involve the bio-statistician at the planning stage of the study itself.

Clinical significance verses statistical significance

One of the most common errors made by the researchers during the study is that when clinical significance is presumed to be synonymous with statistical significance. As such, a significant observation should not be considered as a real effect. Similarly, on obtaining nonsignificant results it cannot be presumed that there are actually no effects. In all the studies, it is implied that $P <$ or > 0.05 is considered a significant or nonsignificant entity. But in reality a $P = 0.04\text{-}0.049$ or $0.051\text{-}0.06$ almost leads to similar inferences from the results although with minor variations rather than drastically different values. These scientific limitations mandate that during reporting of clinical observations exact values of P should be reported rather than reporting P as $<$ or > 0.05. In such scenarios, especially when the population being studied, or the sample size is small, usage of confidence interval is considered essential as it can predict the degree of uncertainty related to results. The usefulness of confidence interval is highly significant when used in conjunction with nonsignificant results.

Secondary results

On numerous occasions, besides getting the results of hypothesis being tested, one can get subsidiary or secondary results which were not anticipated at the time of hypothesis formulation especially when the number of hypothesis are being tested. However, these results should not be given much importance for the ongoing study rather weightage should be given to the primary results only. However, these secondary findings can be helpful in formulating hypothesis for future research work. A statistically significant association between the variables being studied does not convey necessarily about the relationship between two variables. Compared to RCT's, it is difficult to establish a causal relationship in observational studies which can be done only on the basis of nonstatistical grounds.

Even regression analysis has few shortcomings especially when regression equation is used in individual cases so as to predict the numerical level of one variable over the other. The solution to this limitation is to calculate the prediction interval for estimated value of one variable corresponding to a specific value of the other variable.

At the end of the study or anywhere along the text wherever appropriate, limitations and weaknesses of the study should always be addressed. In general, the limitations can vary from the source and type of subjects, research deigns and methodology, impact on the observations, implementation pattern of the study design to a better solution of the present limitations.

Conclusion Section

The conclusion of the study and the inferences derived depends largely upon the use of appropriate and powerful statistical test. One major error which commonly come across at this stage is the lack of reporting exact conclusion if the statistical test applied turn out to be insignificant.[13,15,22] The role of type-II error is significant in this scenario when insignificant results are obtained in a study population of small sample size.[38,39]

A thorough knowledge of these statistical tools and tests can really go a long way in improving the research design thereby producing concrete and evidence-based interpretations. However, the acquisition of these skills and knowledge is an uphill task, but efforts to acquire optimal knowledge about these tools are the first step in the right direction for all the academicians and researchers in modern day research activities. Bio-statistician can play a vital role in educating the editors, reviewers and authors. It will be immensely rewarding for the patients and the mankind if all the teachers and researchers get themselves updated about these statistical aspects through various seminars and workshops on a regular basis.

Acknowledgment

References

1. Altman DG. Statistics in medical journals: Some recent trends. Stat Med 2000;19:3275-89.
2. 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;317:426-32.
3. McKinney WP, Young MJ, Hartz A, Lee MB. The inexact use of Fisher's Exact Test in six major medical journals. JAMA 1989;261:3430-3.
4. García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 2004;4:13.
5. Cooper RJ, Schriger DL, Close RJ. Graphical literacy: The quality of graphs in a large-circulation journal. Ann Emerg Med 2002;40:317-22.
6. Porter AM. Misuse of correlation and regression in three medical journals. J R Soc Med 1999;92:123-8.
7. Gardenier JS, Resnik DB. The misuse of statistics: Concepts, tools, and a research agenda. Account Res 2002;9:65-74.
8. 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.
9. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896-900.
10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Ann Intern Med 2003;138:W1-12.
11. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The CONSORT statement: Revised recommendations for improving the quality of parallel-group randomized trials. Ann Intern Med 2001;134:657-62.
12. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-12.
13. Olsen CH. Review of the use of statistics in infection and immunity. Infect Immun 2003;71:6689-92.
14. Marshall SW. Testing with confidence: The use (and misuse) of confidence intervals in biomedical research. J Sci Med Sport 2004;7:135-7.
15. Klijnsma MP, Cameron ML, Burns TP, McGuigan SM. Out-patient alcohol detoxification — Outcome after 2 months. Alcohol Alcohol 1995;30:669-73.
16. Gogtay NJ. Principles of sample size calculation. Indian J Ophthalmol 2010;58:517-8.
17. Julious SA. Sample sizes for clinical trials with normal data. Stat Med 2004;23:1921-86.
18. Devane D, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. J Adv Nurs 2004;47:297-302.
19. Karlsson J, Engebretsen L, Dainty K, ISAKOS Scientific Committee. Considerations on sample size and power calculations in randomized clinical trials. Arthroscopy 2003;19:997-9.
20. Ogundipe LO, Boardman AP, Masterson A. Randomisation in clinical trials. Br J Psychiatry 1999;175:581-4.
21. MacArthur RD, Jackson GG. An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. J Infect Dis 1984;149:349-54.
22. McCance I. Assessment of statistical procedures used in papers in the Australian Veterinary Journal. Aust Vet J 1995;72:322-8.
23. Krzanowski WJ. Recent trends and developments in computational multivariate analysis. Stat Comput 1997;7:87-99.
24. Dar R, Serlin RC, Omer H. Misuse of statistical test in three decades of psychotherapy research. J Consult Clin Psychol 1994;62:75-82.
25. Welch GE 2nd, Gabbe SG. Review of statistics usage in the American Journal of Obstetrics and Gynecology. Am J Obstet Gynecol 1996;175:1138-41.
26. Goodman NW, Hughes AO. Statistical awareness of research workers in British anaesthesia. Br J Anaesth 1992;68:321-4.
27. Moreira ED Jr, Stein Z, Susser E. Reporting on methods of subgroup analysis in clinical trials: A survey of four scientific journals. Braz J Med Biol Res 2001;34:1441-6.
28. Andersen B, Forrest M. Misuse of statistics. If neither SD nor SE — what then? Nord Med 1987;102:141-2.
29. Hoffmann O. Application of statistics and frequency of statistical errors in articles in acta neurochirurgica. Acta Neurochir (Wien) 1984;71:307-15.
30. Ehrenberg AS. The problem of numeracy. Am Stat 1981;286:67-71.
31. Lang T, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. Philadelphia (PA): American College of Physicians; 1997.
32. Murray GD. The task of a statistical referee. Br J Surg 1988;75:664-7.
33. Feinstein AR. X and iprP: An improved summary for scientific communication. J Chronic Dis 1987;40:283-8.
34. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. Br J Anaesth 2003;90:514-6.
35. Laopaiboon M, Lumbiganon P, Walter SD. Doctors' statistical literacy: A survey at Srinagarind Hospital, Khon Kaen University. J Med Assoc Thai 1997;80:130-7.
36. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. JAMA 2002;287:2817-20.
37. Gardner MJ, Altman DG, Jones DR, Machin D. Is the statistical assessment of papers submitted to the "British Medical Journal" effective? Br Med J (Clin Res Ed) 1983;286:1485-8.
38. Kuzon WM Jr, Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. Ann Plast Surg 1996;37:265-72.
39. Kanter MH, Taylor JR. Accuracy of statistical methods in TRANSFUSION: A review of articles from July/August 1992 through June 1993. Transfusion 1994;34:697-701.