CME Article

Effective Medical Writing

Pointers to getting your article published

Ng K H, Peh W C G

Presenting the statistical results

ABSTRACT

Statistical methods are reported in a scientific paper to summarise the data that has been collected for a study and to enable its analysis. These methods should be described with enough detail to allow a knowledgeable reader who has access to the original data to verify the reported results. This article provides basic guidelines to aid authors in reporting the statistical aspects of the results of their studies clearly and accurately.

Keywords: biostatistics, medical writing, scientific paper, statistics

Singapore Med J 2009;50(1):11-14

INTRODUCTION

Statistical methods are increasingly being reported in scientific studies and medical publications. They are used in a scientific paper to summarise the data that has been collected for a study and enable its analysis. These methods should be described with enough detail to allow a knowledgeable reader who has access to the original data to verify the reported results. However, the misuse of techniques for statistical analysis remains prevalent. The widespread availability of easy-to-use statistical software in fact promotes the proliferation of confusing data and results.

The aim of this article is to provide authors with practical information to help them in presenting the statistical results of their research clearly and accurately. Selected recommendations of textbooks and journal articles describing statistical methods in detail are listed at the end of this article. However, the best way to design and carry out appropriate statistical analyses is to consult an experienced biostatistician during the early stages of the study design. This article guides the reader through the various sections of a scientific paper in which statistical methods and results are commonly reported.

REPORTING STATISTICAL METHODS IN THE MATERIALS AND METHODS SECTION

Study design

Describe the main features of the study design and specify the outcome variables. The study design should be described accurately and in detail so that the study can be reproduced, if required. Study/sampling/data flow charts can be used to describe complicated sampling/study designs to facilitate presentation of information clearly and concisely.

Population and sample size

Describe the target population of the study and present an adequate description of the type of sample, how the sample was selected (e.g. random sample, any stratification used), the pool from which the sample was drawn, the inclusion and exclusion criteria, the assignment mechanism to different treatments (randomised or non-randomised) and any blinding techniques used (e.g. single or double blind).

Describe the expected sample size and the outcome variables. State the assumptions made on the distribution of data, the choice of significance levels and power upon which this sample size was based. State the method of sample size calculation clearly and justify the assumptions made.

Data collection

For a survey type of study, describe the method of data collection and provide pertinent details, such as the number of questions, range of questionnaire response scores and the meaning of each score. Give the number of observations. Report losses to observation, such as dropouts from a clinical trial.

Statistical analysis

Provide clear descriptions of the main features of the statistical analysis (e.g. confidence interval, including degree of confidence; hypothesis tests, including null and alternative hypotheses; level of significance; particular

Biomedical Imaging and Interventional Journal, c/o Department of Biomedical Imaging, University of Malaya, Kuala Lumpur 50603, Malaysia

Ng KH, PhD, MIPEM, DABMP
Editor

Singapore Medical Journal, 2 College Road, Singapore 169850

Peh WCG, MD, FRCP, FRCR
Editor

Correspondence to:
Prof Ng Kwan Hoong
Tel: (60) 3 7949 2069
Fax: (60) 3 7949 4603
Email: dwlng@tm.net.my

tests and test statistics). State the statistical techniques that were used. If sophisticated advanced techniques were used, provide suitable references. References should be to standard works (with page numbers provided) when possible, rather than to papers in which the techniques were originally reported.

Describe the procedures that were put in place to handle missing values and data, as well as any outliers (provide the definition used for an outlier). Describe the model assumption tests that were performed (e.g. test of normality or goodness-of-fit tests). Specify the statistical software and the version used in the analysis, and provide in parenthesis the manufacturer of the software, city, and country of origin.

REPORTING STATISTICS IN THE RESULTS SECTION

Present the results of the main analysis carefully and clearly. Explain how the results address the study objectives. Illustrate the main characteristics of the key variables in tables and/or graphs, if appropriate. Use tables to report summary statistics, summaries of results, or the quantities associated with a p-value. Use appropriate measures of central tendency and spread as summary statistics; for instance, report medians for highly skewed data. If using the notation a ± b, clearly state what b is – it could be standard deviation or standard error.

When reporting the results of statistical tests, be explicit about any assumption that has been made. Provide complete details of the results as follows: state the test name, followed by a colon, then the test statistic (together with any degree of freedom) and the p-value. E.g. ANOVA: $F_{2,6} = 5.6$, $p = 0.02$ and chi-square test: $\chi^2 = 16.81$, $p = 0.01$.

Report the effects of variables using measures that are clinically relevant. For example, report the effects of age in ten-year increments rather than one-year increments, and the effects of weight in 10 kg rather than 1 kg increments.

Avoid non-technical use of technical terms in statistics, such as "random" (which implies a randomising device), "normal", "correlations" and "significant". All statistical terms, abbreviations and symbols should be defined.

Significant figures

Use an appropriate number of significant figures to report the means and other measured or calculated values. The significant figures should reflect the degree of precision of the original measurement; a measurement or calculation should not have to be qualified with the terms, "about" or "approximate".

Here are some examples:

- The mean age of adults only needs to be reported to one decimal place, not four (e.g. 68.1 years, not 68.1276 years).
- For summary statistics (e.g. mean and standard deviation), report one digit more than was presented in the raw data.
- If age is recorded to the nearest whole year, report the mean age to the nearest tenth of a year (e.g. mean = 54.3 years).
- For percentages, the nearest whole percent (e.g. 25%) is usually adequate, although some journals prefer percentages to the nearest tenth of a percent (e.g. 25.4%).
- For test statistics, such as the chi-square statistic, *t* statistic and F statistic, report to two decimal places for accuracy, e.g. *t* statistic = 2.56.

p-values and confidence intervals

Report exact p-values for all main analyses. Exact p-values, such as $p = 0.65$, are preferable to the term "N.S." or "not significant". When a p-value is stated in a table or graph, it need not be repeated in the text. Report the results of *t*-tests (and other tests) in detail, by providing the confidence interval. A simple statement, such as $p < 0.05$, is insufficient. Refer to Box 1 for an example.

Box 1. Reporting *t*-test results: an example.

The difference between the sample mean systolic blood pressure in diabetic patients and non-diabetic patients was 6.0 mmHg, with a 95% confidence interval from 1.1 to 10.9 mmHg; the *t*-test statistic was 2.4, with 198 degrees of freedom and an associated p-value of 0.02.

Avoid sole reliance on statistical hypothesis testing, such as the use of p-values, which fail to convery important qualitative information.

Statistical and clinical significance

Statistical significance must be distinguished from clinical significance. Statistical significance is a statistical calculation that reflects how far a given association exceeds that which would be expected by chance, while clinical significance refers to the potential for the research findings to make a real and measurable difference to patients or clinical practice.

A non-significant result does not necessarily prove the null hypothesis. A "statistically significant" finding is one in which researchers are 95% confident that an

association exists. However, a statistically significant finding does not necessarily prove a cause-effect association or that the results are clinically significant.

SUMMARY

Statistical methods are reported in a scientific paper to summarise the collection and analysis of data that have been presented in the study. The accuracy of the statistical methods used and the way in which they are presented in the scientific paper, affect the integrity of the study. Therefore, statistical judgment must always be exercised.

It is strongly recommended that all authors consult a biostatistician when designing the study, to ensure that the appropriate statistical methods are chosen. The biostatistician could also be called upon to help review the statistical analysis as well as the presentation and interpretation of the results, particularly for studies that employ more advanced or sophisticated statistical methods.

Box 2. Common errors:

- Making the assumption that if $p < 0.05$, the results are worth publishing as they are statistically significant.
- Using statistical significance to prove that there is clinical significance.
- Making the assumption that a non-significant result proves the null hypothesis.

Box 3. Take home points:

1. Exercise statistical judgment at all times.
2. Seek the advice of a biostatistician before beginning the research, instead of waiting until the results have been obtained.

Selected reading list

Books

1. Knapp RG, Miller MC. Clinical Epidemiology and Biostatistics. National Medical Series. Baltimore: Williams and Wilkins, 1992.
2. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. Biostatistics in Clinical Medicine. 3rd ed. New York: McGraw-Hill, 1994.
3. Altman DG. Practical Statistics for Medical Research. New York: Chapman & Hall/CRC, 1999.
4. Mould RF. Introductory Medical Statistics, 3rd ed. Boca Raton: Taylor and Francis, 1998.
5. Campbell MJ, Machin D. Medical Statistics. A Commonsense Approach. 3rd ed. Chichester: John Wiley & Sons, 1999.
6. Armitage P, Berry G. Statistical Methods in Medical Research. 4th ed. Oxford: Blackwell Scientific Publications, 2002.
7. Dawson-Saunders B, Trapp RG. Basic and Clinical Biostatistics. 4th ed. Norwalk: Appleton and Lange, 2004.

Journal papers

1. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. Br Med J (Clin Res Ed) 1983; 286:1489-93.
2. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. Ann Intern Med 1988; 108:266-73.
3. Murray GD. Statistical guidelines for The British Journal of Surgery. Br J Surg 1991; 78:782-4.
4. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. J Bone Joint Surg Am 2004; 86:607-20.
5. Chan YH. Randomised Controlled Trials. Series of two articles on Basic Statistics for Doctors. Singapore Med J 2003; 44: 60-3, 172-4.
6. Chan YH. Biostatistics. Series of 16 articles on Basic Statistics for Doctors. Singapore Med J 2003; 44:280-5, 391-6, 498-503, 614-9, 2004; 45:55-61, 149-53, 249-56, 354-9, 456-61, 558-66, 2005; 46:54-62, 153-60, 259-69, 377-86, 514-8, 675-80.

SINGAPORE MEDICAL COUNCIL CATEGORY 3B CME PROGRAMME
Multiple Choice Questions (Code SMJ 200901A)

	True	False
Question 1. Statistical methods are described in scientific papers to:		
(a) Summarise the collection and analysis of data that have been presented in the study.	☐	☐
(b) Enable a knowledgeable reader who has access to the original data to verify the reported results.	☐	☐
(c) Report how the data has been collected and interpreted.	☐	☐
(d) Enable a biostatistician to check on the validity of the study.	☐	☐
Question 2. Significant figures:		
(a) Should reflect the degree of precision of the original measurement.	☐	☐
(b) Should be at least three decimal places.	☐	☐
(c) For mean age, should be to one decimal place.	☐	☐
(d) For percentage, should be to at least two decimal places.	☐	☐
Question 3. In reporting statistical results:		
(a) p-value alone is sufficient.	☐	☐
(b) If $p < 0.05$, then the results are worth publishing.	☐	☐
(c) Confidence interval should also be provided with the p-value.	☐	☐
(d) Exact p-values are recommended.	☐	☐
Question 4. Adequate description of the sample should include:		
(a) How the sample was selected.	☐	☐
(b) The pool from which the sample was drawn.	☐	☐
(c) Details for each case in the sample.	☐	☐
(d) Inclusion and exclusion criteria used.	☐	☐
Question 5. The following statements are true:		
(a) A non-significant result proves the null hypothesis.	☐	☐
(b) A statistically significant finding does not necessary prove that there is clinical significance.	☐	☐
(c) In reporting *t*-test results, it is recommended to report the confident interval as well.	☐	☐
(d) The advice of a biostatistician should only be sought after the results are obtained.	☐	☐

Doctor's particulars:

Name in full: _____

MCR number: _____ Specialty: _____

Email address: _____

SUBMISSION INSTRUCTIONS:
(1) Log on at the SMJ website: http://www.sma.org.sg/cme/smj and select the appropriate set of questions. (2) Select your answers and provide your name, email address and MCR number. Click on "Submit answers" to submit.

RESULTS:
(1) Answers will be published in the SMJ March 2009 issue. (2) The MCR numbers of successful candidates will be posted online at www.sma.org.sg/cme/smj by 15 March 2009. (3) All online submissions will receive an automatic email acknowledgment. (4) Passing mark is 60%. No mark will be deducted for incorrect answers. (5) The SMJ editorial office will submit the list of successful candidates to the Singapore Medical Council.

Deadline for submission: (January 2009 SMJ 3B CME programme): 12 noon, 8 March 2009.