Education

Brief guideline on use of statistical analysis and reporting data in articles submitted for consideration for publication in Heart Vessels and Transplantation

Samat Kadyrov
Vienna, Austria

These are general guidelines for using statistics in medical papers in accurate and precise way (1-3). Hereafter you can find the brief excerpt of main rules for reporting statistical data (1-5).

The main rule is clearly defining the research question/hypothesis and design of the study, which should guide you in selection of appropriate tests.

General guiding principles:

1. There is an excellent statement by the International Committee of Medical Journal Editors for reporting statistical analyses:

"Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size. References for the design of the study and statistical methods should be to standard works when possible (with pages stated). Define statistical terms, abbreviations, and most symbols. Specify the computer software used" (4).

2. "Provide enough detail that the results can be incorporated into other analyses. In general, this principle requires reporting the descriptive statistics from which other statistics are derived, such as the numerators and denominators of percentages, especially in risk, odds, and hazards ratios. Likewise, P values are not sufficient for re-analysis. Needed instead are descriptive statistics for the variable being compared, including sample size of the groups involved, the estimate (or "effect size") associated with the P value, and a measure of precision for the estimate, usually a 95% confidence interval" (1).

Principles for reporting statistical methods (1-3)

Preliminary analyses: Check for normality of distribution of your data and identify any statistical procedures used to modify raw data before analysis. Examples include mathematically transforming continuous measurements to make distributions closer to the normal distribution, creating ratios or other derived variables, and collapsing continuous data into categorical data or combining categories.

Primary analyses (1-3)

• Describe the purpose of the analysis.

• Identify the variables used in the analysis and summarize each with descriptive statistics.

• When possible, identify the smallest difference considered to be clinically important or calculate using data from previous studies (5).

• Describe the main methods used for analyzing the primary objectives of the study in the statistical analysis section. Describe how the data are presented – continuous variables, categorical/ discrete variables.

• Make clear which method was used for each analysis, rather than just listing all the statistical methods used. For example: Unpaired or t-test for independent samples was used to compare continuous variable between groups with and without coronary heart disease.

• Verify that that data conformed to assumptions of the test used to analyze them. For example: 1) skewed data were analyzed with nonparametric tests, 2) comparison of continuous data within one group e.g. before and after treatment were analyzed with paired tests, or in case of several measurements over time – analysis of variances for repeated measurements was used, and 3) the underlying linear relationship analyzed with linear regression models.

• Indicate whether and how any adjustments were made for multiple comparisons (performing multiple hypothesis tests on the same data).

• If relevant, report how any outlying data were treated in the analysis.

• Specify whether tests were one- or two-tailed and justify the use of one-tailed tests.

• Report the alpha level (e.g., 0.05) that defines statistical significance.

Principles for Reporting Statistical Results

Reporting numbers and descriptive statistics (1-3)
•Report numbers—especially measurements—with precision. For simplicity, round as much as is reasonable. For example, mean age can often be rounded to the nearest year without compromising either the clinical or the statistical analysis. If the smallest meaningful difference on a scale is 5 points, scores can be reported as whole numbers; decimals are not necessary.
•Report total sample and group sizes for each analysis.

Reporting risk, rates, and ratios (1-3)
•Report the type of rate (incidence rates; survival rates), ratio (odds ratios; hazards ratios), or risk (absolute risks; relative risk differences).
•Identify the quantities represented in the numerator and denominator (e.g., the number of men with coronary heart disease divided by the number of men capable of having coronary heart disease).
•Report the time over with each rate applies.
•Report unit of population (that is, the unit multiplier: e.g., x 100; x 10,000) associated with the rate.
•Report a confidence interval for estimated risks, rates, and ratios.

Reporting association analyses (1-3)
•Describe the association of interest.
•Describe the variables used and summarize each with descriptive statistics.
•Report the test of association used.
•Indicate whether the test was one- or two-tailed.
•Justify the use of one-tailed tests.
•For tests of association (e.g., a Chi-square test), report the P value of the test (because association is defined as a statistically significant result).

•Report numerators and denominators for all percentages.
•Summarize data that are approximately normally distributed with means and standard deviations (SD). Use the form: mean (SD), not mean ± SD.
•Summarize data that are not normally distributed with medians and interpercentile ranges, ranges, or both. Report the upper and lower boundaries of interpercentile ranges and the minimum and maximum values of ranges.
•Display the data in tables or figures (2, 3).

•Report the value of the coefficient and a confidence interval for measures of association.

Reporting correlation analyses (1-3)
•Describe the purpose of the analysis.
•Summarize each variable with the appropriate descriptive statistics.
•Report the correlation coefficient used in the analysis (e.g., Pearson, Spearman).
•Confirm that the assumptions of the analysis were met.
•Report the alpha level (e.g., 0.05) that indicates whether the correlation coefficient is statistically significant.
•Report the value of the correlation coefficient. To understand strength of correlation the following might guide: value 0 means no relationship, value 1 means linear relationship and value close -1 means negative linear relationship, any value between 0.5 and 1 – moderate relation. There might be nonlinear relationship, which can be seen from scatterplot – see below.
•Report the (95%) confidence interval for the correlation coefficient, whether or not it is statistically significant.
•Consider reporting the results as a scatter plot. The sample size, correlation coefficient (with its confidence interval), and P value can be included in the data field.

Reporting analyses of variance (ANOVA) or of covariance (ANCOVA), or Kruskal-Wallis test for nonparametric continuous variables (1-3)

These analyses usually are used for comparison of more than 3 groups, or multiple repeated measurements within 1 or 2 groups.

•Describe the purpose of the analysis in methods sections, describe the statistical test and type of post-hoc or post-test where available.

•Identify the variables used in the analysis and summarize each with descriptive statistics.

•Confirm that the assumptions of the analysis were met. For example, indicate whether an analysis of residuals confirmed the assumptions of linearity.

•If relevant, report how any outlying data and any missing data were treated in the analysis.

•Specify whether the explanatory variables were tested for interaction, and if so how these interactions were treated.

•In a table, report the P value for each explanatory variable, the test statistics and, where applicable, the degrees of freedom for the analysis. Report in a table the post-hoc or post-test significance if comparisons were performed between 2 groups within multiple groups

•Provide an assessment of the goodness-of-fit of the model to the data, such as R^2.

•Specify whether and how the model was validated. It is recommended to collaborate with departments of biomedical statistics, you need to explain precisely clearly your hypothesis/ research question you seeking answers for and design of the study – in a simple way explain what you want to do. Consult biomedical statistician for advanced statistical analyses. According to the journal`s rules you have to acknowledge names of contributors - here specialists who helped with statistical analyses.

We recommend using SAMPL guidelines for reporting statistical analysis provided on website of the journal and further references listed below.

References
1.Lang T A, Altman DG. Basic statistical reporting for articles published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. Int J Nurs Stud 2015; 52: 5–9.
2.Schriger DL, Arora S, Altman,DG. The Content of medical journal instructions for authors. Ann Emerg Med 2006; 48: 743–9.e4.
3.Lang TA. How to Write, Publish, & Present in the Health Sciences: A Guide for Clinicians & Laboratory Researchers. American College of Physicians 2010: 389.
4. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated December 2016. Available at: URL: www.icmje.org.   Accessed on 9.01.2017
5. Whitley E, Ball J. Statistics review 4: Sample size calculations. Critic Care 2002; 6: 334-41.