REVIEW

Common Statistical Methods and Reporting of Results in Medical Research

Guoshuang Feng[1,2], Guoyou Qin[3], Tao Zhang[4], Zheng Chen[5] and Yang Zhao[6]

[1]*Big Data Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, 100045, China*

[2]*Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University & Capital Medical University, Beijing, 100083, China*

[3]*Department of Biostatistics, School of Public Health, Fudan University, Shanghai, 200032, China*

[4]*Department of Biostatistics, School of Public Health, Shandong University, Jinan, 250012, China*

[5]*Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, 510515, China*

[6]*Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 211166, China*

Abstract

Statistical analysis is critical in medical research. The objective of this article is to summarize the appropriate use and reporting of commonly used statistical methods in medical research, on the basis of existing statistical guidelines and the authors' experience in reviewing manuscripts, to provide recommendations for statistical applications and reporting.

Keywords: Medical statistics; statistical application; statistical reporting

Introduction

In medical research, statistical analysis is essential, and it involves two aspects: correct application of statistical methods, and correct presentation of statistical results. The former ensures the reliability of results [1–3], and the latter is equally important in the publication of articles. Non-standard results may not clearly express the authors' intentions and may increase the difficulty of future utilization of articles by researchers.

Although many methods for medical statistical analysis exist, clinical studies commonly use comparisons of multiple groups (such as *t* test, analysis of variance (ANOVA) or chi-square test), correlation analysis and regression analysis (such as linear regression or logistic regression) [4–7]. Although these methods are not complicated, they are among the most error-prone in practical applications. Many suggestions or guidelines have been made regarding statistical reports [8–12], primarily in scientific research design and data preprocessing, such as population selection, variable selection, randomization and outliers. In contrast to previous studies, this article describes the correct application and presentation of statistical results for the comparison of multiple groups, correlation analysis, regression analysis and survival analysis, according to

Correspondence: Guoshuang Feng, Ph.D., Associate Professor, Big Data Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, 56 Nanlishi Road, Xicheng District, Beijing 100046, China, E-mail: glxfgsh@163.com

the given study purpose, to provide a reference for clinical researchers.

Descriptive Methods

Descriptive Statistics

The most commonly used descriptive statistics for quantitative data are the mean, standard deviation, median and interquartile range (Q1 for 25th percentile and Q3 for 75th percentile). The statistics of qualitative data primarily comprises the frequency, proportion and rate. For quantitative data, the data distribution must be considered. If the data are normally distributed, reporting as mean (standard deviation, SD) or mean ± SD is recommended. If the data do not follow a normal distribution, reporting as median (Q1–Q3) is recommended, e.g., 135 (128–143).

Reporting Descriptive Statistics

The general tabular form for statistical description is shown in Table 1 [13].

The following aspects must be emphasized for statistical descriptions:

(1) Basic principles of the statistical description table: First, the group factor is usually used as the column head, and the characteristics being compared are listed in the leftmost column of the table (stub column), because many baseline characteristics are usually present. Second, the corresponding units of measurement (such as ng/ml or age) should be listed for the different variables. Providing this information is particularly important for variables with multiple units of measurement. Third, any further explanation, if required, is usually indicated at the bottom of the table. For example, Table 1 may state why the sum of percentages does not equal 100%. Fourth, the number of cases in each group should be listed in the table. If the number of missing cases varies among variables, the number of missing cases should be listed for each variable. Fifth, if two very clear classifications of categorical variables are present, and one of them is of greater interest, only one type of data may be listed. For example, the variable of previous conception in Table 1 is divided into yes and no, and only the frequency with the percentage of the yes category is listed.

Table 1 Characteristics of the Participants at Baseline[†].

Characteristic	Group 1 (*n* = 1077)		Group 2 (*n* = 1080)	
	No. of patients[*]	Value	No. of patients[*]	Value
Age, yr		28.5 ± 3.0		28.4 ± 3.1
BMI		22.0 ± 3.0	1079	22.2 ± 3.1
Blood pressure, mm Hg				
Systolic		118.6 ± 11.9		118.4 ± 12.4
Diastolic		73.0 ± 8.3		72.7 ± 8.4
Fertility history				
Duration of attempt to conceive, yr		3.4 ± 2.0	1079	3.4 ± 2.1
Previous conception, no.(%)		368(34.2)		399(36.9)
Indications for IVF, no.(%)[#]				
Tubal factor		665(61.7)		660(61.1)
Male factor		277(25.7)		280(25.9)
Combined factors		135(12.5)		140(13.0)
Total testosterone, ng/ml	1038	0.28 ± 0.13	1036	0.28 ± 0.14

[†]Plus–minus values are means ± SD. No significant differences were observed between groups (*P* > 0.05) in any baseline characteristics.
[*]The number of patients included in each analysis is provided if it differs from the total number in the trial group.
[#]The total percentage of classified variables may not be 100%, owing to rounding in the calculation.

(2) How many decimal places should be retained for quantitative data? No clear rules exist regarding this issue. For example, the Statistical Analyses and Methods in the Published Literature (SAMPL) guidelines [14] recommend rounding to a reasonable extent for ease of comprehension and simplicity. The European Association of Science Editors (EASE) guidelines [15] recommend providing numbers with two to three effective digits. Habibzadeh has provided additional suggestions [16]: the precision for reporting of each statistic depends on how that statistic is derived; moreover, the number of decimal places reported for the mean, SD, median and IQR in scientific reports should not exceed that of the precision of the measurement in the raw data. We recommend following this suggestion, such that the number of decimal places depends on the accuracy of the original data. For example, if the measurement precision of a red blood cell count is one digit after the decimal point, and the hemoglobin level is an integer, the following could be reported: "the mean (standard deviation) red blood cell count is 4.7 (0.4) $\times 10^{12}$/L, and the mean (standard deviation) hemoglobin is 136 (12) g/L."

(3) How many decimal places should be reported for percentages? In most cases, percentages can be reported with one decimal place, and two decimals can be used for the main variables of interest. If the number of cases in the denominator is less than 100, the percentage has been recommended to be reported as an integer without retaining the decimal point [17, 18]. For example, if 20 of 80 people were positive, the data can be reported as follows: "20 (25%) of 80 people had positive outcomes." When the denominator is less than 100 cases, the change range is greater than 1% for each increase or decrease in the number of cases in the numerator.

(4) For percentage reporting, first, if the total number of cases is too small (e.g., the denominator is less than 20), some articles have recommended not reporting percentages at all, because they can easily be misleading [19, 20]. For example, if six of ten cases are effective, the conclusion that "60% of cases are effective" is not convincing. Reporting the percentage together with the number of cases or 95% confidence interval (*CI*) is added and conclusions should be drawn carefully. Second, if the reporting rate is the main research focus, reporting the 95% *CI* is recommended, to reflect the precision of the results. If only the rate is reported, the information provided is insufficient. For example, for the same incidence rate of 30%, the 95% *CI* for 30 of 100 cases is 21–39%, and that for 3000 of 10000 cases is 29.1–30.9%. The precision of the two results differs by 10 fold.

(5) Must statistics and *P* values be reported for baseline comparisons? The requirements depend on the study design. For randomized controlled trials, reporting *P* values is not recommended, because such trials are randomly grouped, and randomization ensures that any differences between groups are by definition due to chance. In this case, statistical analysis is unnecessary and illogical [21–24]. For observational studies, however, owing to the lack of randomization, group differences may occur because of the selection of cases or exposures. Therefore, statistical analysis can be performed, and the statistics and *P* values can be reported.

(6) How should the percentage of classified variables be presented? The percentages of categorical variables are usually displayed in two ways (as shown in Table 2), which convey different meanings. In Table 2, when the total amount of the row is 100%, the incidence rate in men and women is emphasized. When the

Table 2 Two Ways to Present Percentages.

Gender	Total of row is 100%		Total of column is 100%	
	Case	Control	Case	Control
Male	26 (49.1%)	27 (50.9%)	26 (48.2%)	27 (64.3%)
Female	28 (65.1%)	15 (34.9%)	28 (51.8%)	15 (35.7%)

total of the column is 100%, the data indicate the proportions of men and women in the case and control groups.

The general principle for displaying percentages is that the total percentage for each group variable is 100%. As shown in Table 2, if gender is used as a group variable, the total percentage for each row should be 100%. If the outcome (case or control) is the group variable, the total percentage of each column should be 100%.

Methods for Comparison of Groups

The comparison of groups can be used not only for the main research variables but also for the baseline characteristics. In experimental studies, cohort studies, case-control studies and cross-sectional surveys, comparison of groups can be used according to different purposes, and the meaning of the groups in various study types differs [25–27]. In experimental studies, the groups are usually intervention and non-intervention groups; in cohort studies, the groups are usually exposed and non-exposed groups; and in case-control studies, the groups are case and control groups.

Introduction to Methods

A variety of methods can be used for the comparison of groups [28]. Common methods and applications are shown in Table 3.

Presentation of Results

In most cases, because more than one research variable is compared between groups, the variables are listed in the leftmost column of the table, and the group variable is displayed as a column spanner, as shown in Table 4.

Correlation Analysis

Methods for Correlation Analysias

The use of correlation coefficients depends on the data type and data distribution [29–31]. In general, the Pearson correlation coefficient can be used for quantitative data conforming to a normal distribution; Spearman correlation can be used for quantitative data that do not follow a normal distribution. The correlation analysis between two nominal variables can be described by, e.g., the Pearson contingency coefficient or phi coefficient. The correlation between ordinal variables can be described by, e.g., the Kendall correlation coefficient or Spearman correlation coefficient.

Correlation Reporting

In cases with only several variables, reporting the mean, SD, correlation coefficient and 95% *CI* is recommended (Table 5). If many variables are present, the mean and SD may not be listed, but reporting the correlation coefficient and 95% *CI*, instead of the correlation coefficient and *P* value, is recommended. Reporting values to two decimal places is

Table 3 Descriptive Statistics and Methods for Comparing Multiple Groups.

Data	Descriptive statistics	Methods for comparison of groups
Quantitative data		
Normal distribution	Mean (SD)	t test for two groups (t statistic); ANOVA for multiple groups (F statistic)
Non-normal distribution	Median (Q1–Q3)	Wilcoxon rank sum test for two groups (Z statistic); Kruskal-Wallis rank sum test for multiple groups (χ^2 statistic)
Qualitative data		
Binary	Frequency (percentage)	χ^2 test or Fisher's exact test (χ^2 statistic)
Nominal	Frequency (percentage)	χ^2 test or Fisher's exact test (χ^2 statistic)
Ordinal	Frequency (percentage)	Wilcoxon rank sum test for two groups (Z statistic); Kruskal-Wallis rank sum test for multiple groups (χ^2 statistic)

Table 4   Presentation of a Comparison of Two Groups.

	Group 1 (*n* = n1)	Group 2 (*n* = n2)	Statistics	P value
Normally distributed variable	Mean (SD)	Mean (SD)	*F* or *t*	Actual *P* value
Non-normally distributed variable	Median (Q1–Q3)	Median (Q1–Q3)	*Z* or *H*	Actual *P* value
Binary variable	Frequency (rate)	Frequency (rate)	χ^2	Actual *P* value
Polytomous variable			χ^2 or *H*	Actual *P* value
Class 1	Frequency (percentage)	Frequency (percentage)		
Class 2	Frequency (percentage)	Frequency (percentage)		
Class *k*	Frequency (percentage)	Frequency (percentage)		

Table 5   Mean, SD and Correlation Analysis for Four Variables.

Variable	Mean ± SD	AHI	OAI	OAHI	CAI
AHI	4.08 ± 10.02	–			
OAI	3.01 ± 8.39	0.67(0.62–0.71)	–		
OAHI	2.91 ± 8.08	0.86(0.84–0.88)	0.78(0.75–0.81)	–	
CAI	1.23 ± 2.75	0.17(0.09–0.25)	0.13(0.05–0.22)	0.08(−0.01 to 0.16)	–

recommended for the correlation coefficient and its 95% *CI*.

When describing correlation analysis results, directly reporting the correlation coefficient is recommended, without subjectively describing the correlation as high, moderate or low. For example, "the correlation coefficient between OAI and AHI is 0.67 (0.62–0.71)" is recommended rather than "there is a high correlation between OAI and AHI."

Regression Analysis

In medical research, regression analysis is commonly used in three applications [32]: (1) exploring risk factors, (2) correcting confounding factors and (3) establishing predictive models. The commonly used regression analysis methods are linear regression, logistic regression, Poisson regression and Cox regression, which correspond to the dependent variables for continual data, categorical data, count data and survival data, respectively.

Methods and Methodology

Before the application of regression models, the relevant assumptions must be met [33]. A linear regression model must satisfy the LINE assumption, that is, linearity, independence, normality and equal variance; logit regression with ordinal outcome must meet the proportional odds assumption; and Cox regression must meet the proportional hazards (PH) assumption.

A regression model should report different content according to the research purpose [34]. For example, for analyses aimed at correcting confounding factors, the main research factors and confounding factors must be clearly stated. For analyses aimed at exploring risk factors, the method for variable screening (such as the stepwise regression method or optimal subset method) must be explained. For analyses aimed at establishing a predictive model, the indicators used to reflect the goodness of fit of the model must also be explained; these indicators may include R-squared, the Akaike information criterion, the Bayesian information criterion, root mean square error, area under the ROC curve and 95% *CI*, specificity or sensitivity.

Presentation of Results

In linear regression analysis, the parameter estimation and its 95% *CI*, standardized regression coefficient (preferred if the numerical units of the respective variables are substantially different), standard error, *t* value and *P* value must usually be reported. If space is limited, reporting at least the parameter estimation and its 95% *CI*, instead of the parameter estimation and *P* value, is recommended. In logistic

regression, parameter estimation, standard error, the Wald χ^2, P value, and odds ratio (OR) with 95% CI must usually be reported. If space is limited, reporting at least the OR and its 95% CI is recommended. The reporting form for Poisson regression and Cox regression is similar to that for logistic regression, but the OR is substituted by the risk ratio and hazard ratio (HR), respectively.

Table 6 and Table 7 show routine reporting of linear regression and logistic regression results, respectively.

In a regression model, polytomous variables must be noted, such as living status in Table 7. In most cases, polytomous variables should be included in the form of dummy variables with a pre-specified reference category [35], and the comparison results between other categories and the reference category should be reported. As shown in Table 7, compared with that of living alone (reference category), the life satisfaction rating of living with a spouse (excluding other families) is 4.262 higher on average, that of living with family (excluding spouse) is 1.946 higher on average, and that of living with a spouse and family is 4.748 higher on average. As shown in Table 8, compared with that of the < 35 year age group, the OR values for cardiovascular disease in the 35–55 year age group and > 55 year age group are 3.34 and 3.61, respectively.

Table 6   Results of Linear Regression for Life Satisfaction Rating.

	Parameter estimation	95% CI	Standard error	t value	P value
Age, yr	−0.181	−0.243, −0.118	0.032	−5.73	<0.001
Living status					
Living alone	Ref				
Living with a spouse (excluding other family members)	4.262	1.690, 6.834	1.312	3.25	0.001
Living with family (excluding spouse)	1.946	−1.066, 4.958	1.537	1.27	0.206
Living with spouse and family	4.748	2.141, 7.355	1.330	3.57	<0.001
Social family status rating	3.888	3.418, 4.358	0.240	16.21	<0.001

Table 7   Results of Logistic Regression for Cardiovascular Disease.

	Parameter estimation	Standard error	Wald χ^2	P value	OR	95% CI
Age, yr						
<35					Ref	
35–55	1.207	0.427	7.990	0.005	3.34	1.42–7.70
>55	1.284	0.423	9.216	0.002	3.61	1.54–8.25
Personal life rating						
>2					Ref	
≤2	1.103	0.104	111.635	<0.001	3.01	2.46–3.69

Table 8   Log-Rank Test for Survival Time.

	Cases	Death (%)	Median survival time (95% CI)	χ^2	P value
Gender				0.468	0.494
Male	106	44(41.5)	3.53(2.03–4.37)		
Female	231	96(41.6)	4.50(3.83–5.00)		
Dose, mg				1.384	0.501
250	124	45(36.3)	4.03(2.83–4.63)		
425–500	198	87(43.9)	4.33(3.53–5.10)		
>500	15	8(53.3)	2.87(1.40–14.10)		

Survival Analysis

Survival analysis is a series of analytic processes [36] including description, comparison of groups and regression analysis.

Methods and Methodology

For the description of survival data, the survival rate and median survival time are usually estimated with the Kaplan-Meier method [37]. For comparison of groups of survival data, the log-rank test and Gehan-Breslow-Wilcoxon test are commonly used. The log-rank test, which tends to perform best toward the right side of the survival curve, is often used when the PH assumption is met [38], whereas the Gehan-Breslow-Wilcoxon test, which tends to perform best on the left side of the survival curve, is the fallback method when the PH assumption fails. For regression analysis of survival data, Cox regression, a semi-parametric method, is widely used, but the assumption of PH must be satisfied [39].

When survival analysis methods are introduced in an article, the following should be noted. (1) The starting time (such as follow-up after surgery) and outcome (such as death) should be clearly defined. (2) Statistical description indicators, usually the median survival time and its 95% *CI*, should be stated. Sometimes the median follow-up time may also be stated. (3) The estimation method for survival rate, such as the Kaplan-Meier method, should be stated. Of note, the Kaplan-Meier method is a method for estimating survival, not a statistical inference method. For example, it can be said that "the survival rate is estimated by Kaplan-Meier method" or "the Kaplan-Meier survival curve is drawn," but it cannot be said that "the Kaplan-Meier method is used to compare the survival curves of the two groups." (4) The method used for statistical inference should be stated. For example, the log-rank test should not be used if a clear intersection is present between survival curves. Cox regression should meet the PH assumption; otherwise, a non-PH model should be considered.

Reporting of Results

Statistical Description

The follow-up profile, such as the number of cases in each group or the number lost to follow up,
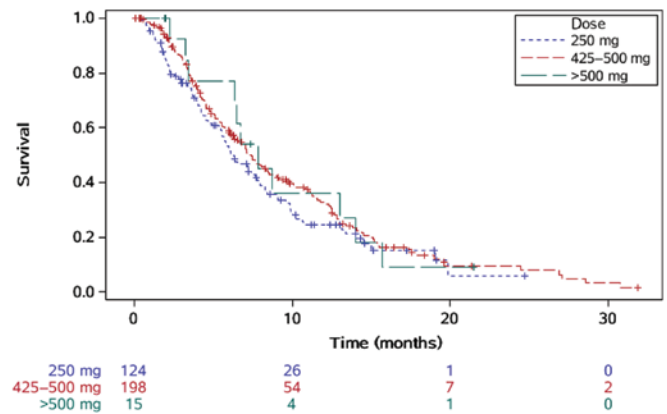

Figure 1 Survival Curves for Three Groups.

should be stated. The median survival time and its 95% *CI* should also be reported; for example, "the median survival times of the three groups were 5.7 (3.7–8.0) months, 7.1 (4.6–7.9) months and 7.9 (2.3–13.0) months." Sometimes, depending the purpose of the study, the survival rate at a fixed time point (95% *CI*) can also be reported. For example, "the 1-year Kaplan Meier survival rates in the treatment group and the control group were estimated to be 0.677 (0.588–0.766) and 0.206 (0.173–0.239), respectively."

Reporting the survival curve of the main analysis indicators (Figure 1) is strongly recommended because it can visually indicate the changes in the survival rates in two or more groups. If possible, the number of people at risk at different follow-up times in each group in the survival curve should be reported. At the bottom of Figure 1, the number of risk sets of the three dose groups at 0, 10, 20 and 30 months is shown. Reporting the 95% confidence band is recommended if only one survival curve is shown. Of note, the survival curve corresponds to the confidence band rather than the confidence interval. The confidence interval is the interval for each time point, and the confidence band is the interval of the entire survival function.

Statistical Inference

When comparing survival data between groups, the median survival time should be reported if only one grouping variable is compared. The results of statistical analysis can be stated as text in the results, such as "the median survival times

Table 9 Results of Cox Regression for Survival Time.

	Parameter estimation	Standard error	Wald χ^2	P value	HR	95% CI
Treatment						
Control					Ref	
Treatment	−1.000	0.271	13.637	<0.001	0.368	0.216–0.625
Age, yr						
≤40					Ref	
41–60	−0.355	0.502	0.500	0.479	0.701	0.262–1.875
>60	−0.426	0.502	0.718	0.397	0.653	0.244–1.749

of the treatment group and the control group are 280 (159–352) days and 99 (67–151) days, respectively, and the difference between groups is statistically significant ($\chi^2 = 16.126$, $P < 0.001$)." If multiple grouping variables are present, the results of each variable should be displayed in a table, as shown in Table 8.

If Cox regression is used for multiple analysis, the test output of the PH assumption must first be reported to validate that the model is applicable, followed by reporting of the results of the regression analysis. For Cox regression, parameter estimation, the standard error, Wald χ^2, P value, *HR* and its 95% *CI* must usually be reported. If space is limited, reporting at least the *HR* and its 95% *CI* is recommended. The reported results of Cox regression multiple analysis are given in Table 9.

Summary

We provide a summary of the appropriate application and reporting of commonly used statistical methods, such as comparison of groups, correlation analysis, regression analysis and survival analysis. These recommendations do not include all statistical methods, nor do they establish a comprehensive standard. Instead, they are aimed at providing suggestions for clinical researchers, to avoid statistical application errors in medical articles. No single document can cover all statistical methods. Clinical researchers should consult a statistician with experience if necessary.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. Biochem Med (Zagreb) 2015,25(1):5–11.
2. Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools, and a research agenda. Account Res 2002,9(2):65–74.
3. Nyirongo VB, Mukaka MM, Kalilani-Phiri LV. Statistical pitfalls in medical research. Malawi Med J 2008;20:15–8.
4. Sebastião YV, St Peter SD. An overview of commonly used statistical methods in clinical research. Semin Pediatr Surg 2018,27(6):367–74.
5. Nour-Eldein H. Statistical methods and errors in family medicine articles between 2010 and 2014-Suez Canal University, Egypt: A cross-sectional study. J Fam Med Prim Care 2016;5(1):24–33.
6. Narayanan R, Nugent R, Nugent K. An Investigation of the variety and complexity of statistical methods used in current internal medicine literature. South Med J 2015;108(10):629–34.
7. He J. Randomized controlled trial and statistical analysis methods for clinical medicine: the existing problems. Acad J Second Mil Med Univ 2006;27(7):697–700.
8. Althouse AD, Below JE, Claggett BL, Cox NJ, de Lemos JA, Deo RC, et al. Recommendations for statistical reporting in cardiovascular medicine: a special report from the American Heart Association. Circulation 2021;27;144(4):e70–91.
9. American Psychological Association. Publication manual of the American Psychological Association. 6th ed. Washington, DC; 2010.
10. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. Stat Med 2008;27:6547–69.
11. Assel M, Sjoberg D, Elders A, Wang X, Huo D, Botchway A, et al.

Guidelines for reporting of statistics for clinical research in urology. BJU Int 2019;123:401–10.

12. Vickers AJ, Assel MJ, Sjoberg DD, Qin R, Zhao Z, Koyama T, et al. Guidelines for reporting of figures and tables for clinical research in urology. Urology 2020;142: 1–13.

13. Shi Y, Sun Y, Hao C, Zhang H, Wei D, Zhang Y, et al. Transfer of Fresh versus Frozen Embryos in Ovulatory Women. N Engl J Med 2018;378(2):126–36.

14. Lang TA, Altman DG. Basic Statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines". Nurs Stud 2015;52:5–9.

15. European Association of Science Editors. EASE guidelines for authors and translators of scientific articles to be published in English. 2018. Available from: https://www.ease. org.uk/wp-content/uploads/2018/11/ doi.10.20316.ESE_.2018.44.e1.pdf.

16. Habibzadeh F, Habibzadeh P. How much precision in reporting statistics is enough? Croat Med J 2015;56:490–2.

17. Lang TA, Secic M. How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers. 2nd ed. Philadelphia: American College of Physicians; 2006.

18. Habibzadeh F. Common statistical mistakes in manuscripts submitted to biomedical journals. Eur Sci Edit 2013;39:92–4.

19. Priebe HJ. The results. In: Hall GM. editor. How to write a paper. 3rd ed. London: BMJ Publishing Group; 2003. pp. 22–35.

20. Habibzadeh F. Statistical data editing in scientific articles. J Korean Med Sci 2017;32:1072–6.

21. Knol MJ, Groenwold RHH, Grobbee DE. P-values in baseline tables of randomised controlled trials are inappropriate but still common in high impact journals. Eur J Cardiovasc Prev Rehabil 2011,19(2):231–2.

22. Keriazes GA. Misuse of the p Value for baseline characteristics. Pharmacotherapy 2012;32(9):e172–3.

23. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche P, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28–55.

24. Wadgave U, Khairnar MR, Wadgave Y. Statistical issues in randomized controlled trials: an editorial. Electron Physician 2018;10(10):7293–8.

25. Cataldo R, Arancibia M, Stojanova J, Papuzinski C. General concepts in biostatistics and clinical epidemiology: observational studies with cross-sectional and ecological designs. Medwave 2019;19(8):e7698.

26. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J 2003;20(1):54–60.

27. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg 2010;126(6):2234–42.

28. Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. Application of student's t-test, analysis of variance, and covariance. Ann Card Anaesth 2019;22(4):407–11.

29. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018;126(5):1763–8.

30. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol Methods 2012;17(3):399–417.

31. Zou KH, Tuncali K, Silverman SG. Correlation and simple linear regression. Radiology 2003;227(3):617–22.

32. Bender R. Introduction to the use of regression models in epidemiology. Methods Mol Biol 2009;471:179–95.

33. Ali P, Younas A. Understanding and interpreting regression analysis. Evid Based Nurs 2021;24(4):116–8.

34. Singh SK, Kaplan B, Kim SJ. Multivariable regression models in clinical transplant research: principles and pitfalls. Transplantation 2015;99(12):2451–7.

35. Cleophas TJ, Atiqi R, Zwinderman AH. Handling categories properly: a novel objective of clinical research. Am J Ther 2012;19(4):287–93.

36. Benítez-Parejo N, Rodríguez del Águila MM, Pérez-Vicente S. Survival analysis and Cox regression. Allergol Immunopathol (Madr) 2011;39(6):362–73.

37. Barakat A, Mittal A, Ricketts D, Rogers BA. Understanding survival analysis: actuarial life tables and the Kaplan-Meier plot. Br J Hosp Med (Lond) 2019;80(11):642–6.

38. Koletsi D, Pandis N. Survival analysis, part 2: Kaplan-Meier method and the log-rank test. Am J Orthod Dentofacial Orthop 2017;152(4):569–71.

39. George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol 2014;21(4):686–94.