Errors in use of multivariable regression analysis

Madam,

Hiremath and Kamdod published a retrospective study and applied multivariable (linear and logistic) regression analysis to find the association of change in MAP level, serum creatinine level and survival benefit with various risk factors. I have some remarks regarding the application of multivariable regression methods in his study.

Issues Related to Multivariable Logistic Regression

Thirty-one subjects were included in the study, as per rule of thumb derived from the simulation study for logistic regression at least 10 events per variable (EPV) for the minimum outcome. The author calculated on the basis of 10 subjects per variable, which is not correct for logistic regression. For instance, 11 subjects out of 31 were survived on the end of the study. Thus, minimum outcome event is survival at end of the study. Therefore, survival benefit outcome had only 5.5 EPV as author included two risk factors in the model. In addition to this condition when a risk factor is rare which means positivity of this risk factor is small, even 10 EPV may be inadequate.

Insufficient sample size in regression models yield unstable risk estimates, wider confidence interval (CI), and can reflect inaccurate association as author found in their study for in vitro fertilizations (IVFs) variable higher odds ratio (OR), wider CI and P value closest to 0.05. Furthermore, the author reported log-odds and its 95% CI 2.42 (1.06, 121.13), instead of reporting 95% CI of log-odds, the author reported 95% CI of OR. The correct value of log-odds and its 95% CI was 2.42 (0.06, 4.797). The OR is equal to exp(log-odds). In logistic regression, log-odds and OR has different meaning former give additive effect while later provide the multiplicative effect. The wider CI of OR usually indication of over-fitting of model and results of the model is not trustworthy. The H-L goodness of fit test is also sensitive to small sample size because this is based on the Chi-square test and usually divided into deciles (10 parts). Chi-square test needed more the 5 excepted frequencies in at least 80% of cells. Thus, this test is not reliable in a small size.

Issues Related to Multiple Linear Regression Analysis

For multiple linear regression rules of thumb state that at least 20 subjects per eligible variable were included in the model. Whereas, the author applied multiple linear regression only in seven subjects which are too small to yield correct model results. The terminology in multiple regression is “regression coefficient” not “regression correlation” as reported by the author in results. Author reported seven subjects were included and reported F-statistics as F(2,6) = 6.27 which is also wrong because it should be F (p, n-p-1) where P is number of parameters and n is number of sample size, according to this it should be F(2,4). The basic assumptions for multiple linear regression: Normality of residuals and homogeneity of variance were not reported by the author.

Author reported that negative Pearson correlation between volume of IVF used and change in serum creatinine ($r = -0.816$) but in multiple regression its regression coefficients was positive and too small 0.000 although significant, zero regression coefficient appeared because volume of IVF was measured in milliliter, and the effect of 1 ml volume of IVFs change had very small influence on change in serum creatinine level. In such situations, it would be better to convert the ml into liter to get the regression coefficient into a meaningful and in a presented way.

Furthermore, Pearson correlation needs both variables should be metric scale, the author reported Pearson correlation ($r = 0.042$) between sex and change in MAP, which also seems to be inaccurate. In addition, the ratio between male to female is 30–1 that showed only 1 female in this study.

The multivariable regressions are an important tool in medical literature to find the association between outcome and risk factors, or to predict the outcome from a set of predictors. Reliability of regression model depends on the fulfillment of model associated assumptions. The quality of logistic regression was evaluated on bases of well-established 10 points criteria and found the quality in Indian medical journal is far lagging behind as compared to quality of logistic regression in articles.

E-mail: ak.arunava@gmail.com

Correspondence to:
Dr. Arunava Kali,

1st Floor, 31/33, Jor Bagh, New Delhi, India
Letters to the Editor

1. Hiremath SB, Kamdod MA. Effect of various drugs used in conservative therapy of hepatorenal syndrome: A retrospective drug utilization study. Indian J Pharmacol 2014;46:538-42.

2. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-9.

3. Katz MH. Multivariable analysis: A primer for readers of medical research. Ann Intern Med 2003;138:644-50.

4. Kumar R, Indrayan A, Chhabra P. Reporting quality of multivariable logistic regression in selected Indian medical journals. J Postgrad Med 2012;58:123-6.

5. Kumar R, Chhabra P. Cautions required during planning, analysis and reporting of multivariable logistic regression. Curr Res Pract 2014;4:31-9.

References

Nil.

Conflicts of Interest

There are no conflicts of interest.

Rajeev Kumar

Department of Biostatistics and Medical Informatics,
University College of Medical Sciences, New Delhi, India

Correspondence to:
Dr. Rajeev Kumar,
E-mail: rajeev.kumar.malhotra@gmail.com

Financial Support and Sponsorship

Nil.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

Access this article online

Website: www.ijp-online.com

DOI: 10.4103/0253-7613.165187

Quick Response Code:

Cite this article as: Kumar R. Errors in use of multivariable regression analysis.
Indian J Pharmacol 2015;47:571-2.