Turkish Journal of Thoracic and Cardiovascular Surgery 2023;31(1):8-10

Assessment of the factors that affect fast-track or early extubation following pediatric cardiac surgery: Logistic regression in clinical studies

Pediatrik kalp cerrahisi sonrası fast-track veya erken ekstübasyonu etkileyen faktörlerin değerlendirilmesi: Klinik çalışmalarda lojistik regresyon

Meral Yay

Department of Statistics, Mimar Sinan Fine Arts University, Faculty of Arts and Sciences, Istanbul, Türkiye

In recent years, logistic regression, which is an extension of linear regression, is often used to investigate the independent effect of a variable on the binomial outcome (dependent or response) in medical research. The fact that it allows the use of both continuous and categorical variables allows logistic regression to be preferred. It also provides the ability to adjust for multiple predictors (independent or explanatory) and this makes logistic regression particularly useful for analysis of observational data, when adjustment is needed to reduce the potential bias resulting from differences in the groups compared.[1] The use of standard linear regression for a two-level outcome in scientific research leaves researchers with very unsatisfactory results. When linear regression is used, the outcome variability is the same for all values of the predictors. However, when the outcome variable has two categories, the assumption of constant variability does not work and, in this case, linear regression is insufficient. The logistic regression developed to fill this gap assumes a relationship between the predictor variable and the logit of the outcome. According to the outcome variable and the structure of the predictor, the researcher should decide which type of regression to use. For instance, when the predictor variable is continuous or categorical, different types of regression are used depending on the nature of the outcome variable. If the outcome variable is continuous, linear regression is used, but it is not easy to encounter a continuous outcome in practice. To illustrate, a researcher who is willing to find out the odds of having cardiovascular disease in a patient with a smoking habit is faced with a binary outcome, such as cardiovascular disease (present/absent). At this time, it is not possible to use linear regression, as linearity, normality and continuity assumptions of linear regression are violated when the outcome variable is not continuous. This need, which is the starting point of logistic regression, is defined in different ways as binary, multinomial or ordinal, depending on the type of outcome variable. Binary logistic regression is used, when the outcome variable has two outcomes, such as (present/absent) or (dead/alive). However, the outcome variable (drug X, drug Y, drug Z) may have more than two categories. There is no order of importance for drugs in themselves and, in this case, multinomial logistic regression is preferred. However, if the outcome variable is ordered such as cancer stages (Stage I-II-III-IV), then ordinal logistic regression should be used. Examples of logistic regression are quite common in the literature.[2,3] As in all models, certain assumptions are made to fit the model to the data. Logistic regression does not assume a linear relationship between the dependent and independent variables, but between the logit of the outcome and the predictor values.[4] The outcome variable must be categorical; the predictor variables need not be interval; nor normally distributed, nor linearly related, nor of equal variance within each group, and finally, the categories (groups) must be mutually

Corresponding author: Meral Yay.
E-mail: meral.yay@msgsu.edu.tr

Doi: 10.5606/tgkdc.dergisi.2023.98550

Received: January 19, 2022
Accepted: January 19, 2022
Published online: January 30, 2023

Cite this article as: Yay M. Assessment of the factors that affect fast-track or early extubation following pediatric cardiac surgery: Logistic regression in clinical studies. Turk Gogus Kalp Dama 2023;31(1):8-10. doi: 10.5606/tgkdc.dergisi.2023.98550.

©2023 All right reserved by the Turkish Society of Cardiovascular Surgery.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (http://creativecommons.org/licenses/by-nc/4.0/).

exclusive and exhaustive. A case can only be in one group and every case must be a member of one of the groups.[5] If predictor variables are normally distributed and outcome has a linear relationship with the variable, the power of the analysis increases. With small sample sizes, the Hosmer-Lemeshow test has low power and is unlikely to detect subtle deviations from the logistic model. The Hosmer-Lemeshow recommend sample sizes greater than 400 and a minimum number of cases per predictor variable is 10.[6,7]

Logistic regression used to estimate the probability (or risk) of a particular outcome, given the value(s) of the independent variable, assumes a linear relationship with the logit of the outcome (natural logarithm of odds). The logistic regression model is initially defined as the probability of a two-level result of interest. In the model, the probability of the event of interest is expressed, as the ratio of the probability of the event to the probability of its not happening. The logistic regression model takes the natural logarithm of the probabilities as a regression function of the estimators. If there is only one predictor in the model, it takes the form *"ln[odds(Y=1)]=β_0+β_1X"*. Herein, *"Y"* is the outcome variable and "Y" takes the value *"1"* when the event occurs and "0" when the event does not occur. *"0"* is also the cut-off term and represents the regression coefficient representing the change in the logarithm of the probabilities of the event with a 1-unit change in *"X"*. It is also possible to extend the model to the form containing more than one predictor, in which case the model becomes *"ln[odds(Y=1)]=β_0 +β_1X_1+β_2X_2+...+β_kX_k"*. This equation is the general form of logistic regression and *"k"* is the number of predictors. The selection of predictors can be only achieved by a careful review of the literature in relation to the outcome to ensure that the full range of potential predictors is included.[8] The main goal of logistic regression is to estimate the *"k+1"* unknown parameters *"β"* in the equation. It is possible to estimate the *"β_s"* coefficients in the model using the maximum likelihood or weighted least square methods. The fact that these coefficients, which show the relationship between *"X_s"* and logit of *"Y"*, take a value greater than *"0"* indicates an increase in logit of *"Y"*; a value less than *"0"* indicates a decrease in the logit of *"Y"* and a value of *"0"* indicates that there is no linear relationship between logit of *"Y"* and *"X_s"*.

A model is created which includes all the predictor variables useful in logistic regression to accurately predict the outcome category for individual cases using the best model. It calculates the probability of success over the probability of failure and gives an odds ratio (OR) for the analysis results. While interpreting logistic regression, OR is also included in the interpretation to evaluate not only the regression coefficients, but also the relationships between outcome and predictors. When a predictor variable *"X_i; i=1,2,...,k"* increases by one unit *"X_i+1"*, which is referred to as the independent variable, the probabilities of the dependent variable, called the outcome variable, increase by the factor *"exp(β_i)"*, while all other factors remain constant. This is called the OR and ranges from zero to positive infinity. It indicates the relative amount by which the odds of the outcome variable increase (OR>1) or decrease (OR<1) when the value of the corresponding predictor variable increases by one (1) unit. The OR is a measurement that finds wide use particularly in epidemiology. For instance, when *"Y"* indicates liver cancer status (have=1; have not=0), *"X"* indicates alcohol use status (using=1; not using=0); an OR calculated as "OR=3" means that the probability of liver cancer among alcohol users is three times as likely among non-alcoholics in any given population.

The validity of the model obtained with the help of selected variables should be checked by evaluating the goodness-of-fit of the logistic regression model. First, the overall evaluation of the model is made by examining the relationship between all predictor variables in the model and the outcome variable. For this purpose, evaluation is made by comparing the compatibility of two models with and without predictor variables. If the predictor variable model shows an improvement over the predictor variable model, it is stated that it fits the data better. For this purpose, the *"H_0: $\beta_1=\beta_2=...=\beta_k=0$"* hypothesis is tested by using "likelihood ratio test" for the general fit of the model. If the "p" value for the general model fit statistic is lower than the significance level of the test ("p<0.05"), the *"H_0"* hypothesis is rejected and at least one of the predictor variables is speculated to contribute to the prediction of outcome. The Hosmer-Lemeshow test, another indicator of good overall model fit, is used to examine whether the observed proportions of events are similar to the predicted probabilities of occurrence in subgroups of the model population. The Hosmer-Lemeshow test statistic asymptotically follows a *"χ^2"* distribution and small values (with large p-value closer to 1) indicate a good fit to the data, thus, good overall model fit. On the other hand, large

values (with $p<0.05$) indicate a poor fit to the data. If the general model obtained in logistic regression works well, the second step is how important each of the predictor variables is. The "Wald test" of the logistic regression coefficients for each predictor is performed and their contribution to the model is evaluated. For each model coefficient, the null hypothesis that the coefficient is zero is tested against the alternative that the coefficient is not zero using a "Wald test". The Wald statistic is asymptotically distributed as a "χ^2" distribution and each Wald statistic is compared with a "χ^2" critical value with degrees of freedom "1". If the H_0 hypothesis is rejected, it is stated that the contribution of the coefficient to the model is statistically significant. The classification table, which should be prepared in the third step, is a method to evaluate the predictive accuracy of the logistic regression model.[9] In the 2×2 classification table, the observed values for the outcome and the predicted values are cross-classified. The classification table, which should be prepared in the third step, is a method to evaluate the predictive accuracy of the logistic regression model. In the generated classification table, the number of observations in cells a, b, c and d, respectively, is expected to be high in cells a and d, and very few in cells b and c. High sensitivity and specificity calculated with the help of medical diagnostic test indicate better fit to the model. While specifying a single cut-off point using the 2×2 classification table, by extending this table, all possible cut-off points in the range 0-1 can be observed. Plotting the pairs of sensitivity and one minus specificity on a scatter plot provides a Receiver Operating Characteristic (ROC) curve. The area under this curve (AUC) provides an overall measure of fit of the model.[10] The larger the AUC, the greater the predictability, which is an indicator of the success of the model.

In this study, factors affecting fast-attack and early extubation were evaluated with the help of multivariate logistic regression analysis. While outcome was variable (fast-track or early extubation) in the established multivariate model, age, body weight, presence of genetic syndrome, operational risk category and procedure time variables were found to be statistically significant as predictor variables. As a result of the analysis, fast-attack and early extubation can be successfully applied with low reintubation rates in selected cases with congenital heart surgery. Age over 30 days, less complex procedures (RACSH-1<4), absence of genetic syndrome, shorter duration of cardiopulmonary bypass and lower vasoactive-inotropic score (<8) may predict rapid follow-up and early extubation.

REFERENCES

1. Kirkwood BR, Sterne JAC. Essential Medical Statistics. Oxford: Blackwell Science Ltd.; 2003.
2. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons, Inc.; 2000.
3. LaValley MP. Logistic regression. Circulation 20086;117:2395-9. doi: 10.1161/CIRCULATIONAHA.106.682658.
4. Hosmer DW, Lemeshow S, Sturdivant RX. The multiple logistic regression model. Applied Logistic Regression 1989;1:25-37.
5. Boateng EY, Abaye DA. A Review of the logistic regression model with emphasis on medical research. JDAIP 2019;7:190-207. doi: 10.4236/jdaip.2019.74012.
6. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley & Sons, Inc.; 2013.
7. Hosmer DW, Jovanovic B, Lemeshow S. Best Subsets Logistic Regression. Biometrics 1989;45:1265-70. doi: 10.2307/2531779
8. Reed P, Wu Y. Logistic regression for risk factor modelling in stuttering research. J Fluency Disord 2013;38:88-101. doi: 10.1016/j.jfludis.2012.09.003.
9. Peng CYJ, So TSH. Logistic regression analysis and reporting: A primer, understanding statistics. 2002;1:31-70. doi: 10.1207/S15328031US0101_04.
10. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care 2004;8:508-12. doi: 10.1186/cc3000.