



Linear and logistic regression models: when to use and how to interpret them?

Horacio Matias Castro^{1,2} , Juliana Carvalho Ferreira^{1,3} 

PRACTICAL SCENARIO

A secondary analysis⁽¹⁾ of a study designated “Integrating Palliative and Critical Care,” a cluster randomized trial, was conducted to explore differences in receipt of elements of palliative care among patients who died in the ICU with interstitial lung disease (ILD) or COPD in comparison with those who died of cancer. The authors used two methods of multiple regression analysis: linear regression to estimate the impact of COPD and ILD, in comparison with that of cancer, on the length of ICU stay, and logistic regression to evaluate the effects of COPD and ILD on the presence or absence of elements of palliative care. All regression models were adjusted for confounders (age, sex, minority status, education level, among others) of the association between the patient diagnosis and palliative care outcomes.

INTRODUCTION

Linear and logistic regressions are widely used statistical methods to assess the association between variables in medical research. These methods estimate if there is an association between the independent variable (also called predictor, exposure, or risk factor) and the dependent variable (outcome).⁽²⁾

The association between two variables is evaluated with simple regression analysis. However, in many clinical scenarios, more than one independent variable may be associated with the outcome, and there may be the need to control for confounder variables. When more than two independent variables are associated with the outcome, multiple regression analysis is used. Multiple regression analysis evaluates the independent effect of each variable on the outcome, adjusting for the effect of the other variables included in the same regression model.

WHEN TO USE LINEAR OR LOGISTIC REGRESSION?

The determinant of the type of regression analysis to be used is the nature of the outcome variable. Linear regression is used for continuous outcome variables (e.g., days of hospitalization or FEV1), and logistic regression is used for categorical outcome variables, such as death. Independent variables can be continuous, categorical, or a mix of both.

In our example, the authors wanted to know if there was a relationship between cancer, COPD, and ILD (baseline disease; the independent variables) with two

different outcomes. One outcome was continuous (length of ICU stay) and the other one was categorical (presence or absence of elements of palliative care). Therefore, two models were built: a linear model to examine the association between baseline disease (chronic pulmonary disease or cancer) and length of ICU stay, and a logistic regression analysis to examine the association between the baseline disease and being in receipt of elements of palliative care.

HOW TO INTERPRET RESULTS OF REGRESSION ANALYSIS?

Regression models are performed within statistical packages, and the output results include several parameters, which can be complex to interpret. Clinicians who are learning the basics of regression models should focus on the key parameters presented in Chart 1.

In our example, the baseline disease—COPD, ILD, or cancer (the reference category)—is the independent variable, and length of ICU stay and receipt of palliative care elements are the outcomes of interest. In addition, the regression models also included other independent variables considered as potential confounders, such as age, sex, and minority status. In the linear regression model, the length of ICU stay for patients with ILD was longer than for those with cancer ($\beta = 2.75$; 95% CI, 0.52-4.98; $p = 0.016$), which means that, on average, having ILD increased the length of ICU stay in 2.75 days when compared with the length of ICU stay among cancer patients. In the logistic regression model, the authors found that patients with ILD, when compared with cancer patients, were less likely to have any documentation of their pain assessment in the last 24 h of life (OR = 0.43; 95% CI, 0.19-0.97; $p = 0.042$), which means that having ILD decreased the odds of documentation of pain assessment by more than half.

KEY POINTS

- Linear and logistic regressions are important statistical methods for testing relationships between variables and quantifying the direction and strength of the association.
- Linear regression is used with continuous outcomes, and logistic regression is used with categorical outcomes.
- These procedures require expertise in regression model building and typically require the assistance of a biostatistician.

1. Methods in Epidemiologic, Clinical, and Operations Research-MECOR-program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.

2. Pulmonary Medicine Department, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

3. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.

Chart 1. Most important parameters in regression analyses and their interpretations.

Parameter	Linear regression	Logistic regression
Direction and strength of the association between the independent variable and the dependent variable (outcome)	Beta coefficient: Describes the (expected) average change in the outcome variable for each one-unit change in the independent variable for continuous variables, or the average change in the outcome variable for one category of the independent variable compared with a reference category for categorical variables	OR: The OR for a continuous independent variable is interpreted as the change in the odds of the outcome occurring for every one-unit increase in the independent variable The OR for categorical independent variables is interpreted as the increase or decrease in odds between two categories (e.g., men vs women) OR = 1: no association; OR > 1: positive association or risk factor; and OR < 1: negative association or protective factor
Example (for a continuous independent variable)	The expected increase in FEV ₁ for each centimeter increase in height	The expected increase in the odds of death for each increase of one year of age among patients with sepsis
Example (for a categorical independent variable)	The expected increase in FEV ₁ for men compared with women with the same height and age	The expected increase in the odds of death for men compared with women among COVID-19 patients
Precision of the estimate	The 95% CI of the beta coefficient	The 95%CI of the OR
Statistical significance	The p value (significant when < 0.05)	The p value (significant when < 0.05)

REFERENCES

1. Brown CE, Engelberg RA, Nielsen EL, Curtis JR. Palliative Care for Patients Dying in the Intensive Care Unit with Chronic Lung Disease Compared with Metastatic Cancer. *Ann Am Thorac Soc.* 2016;13(5):684-689. <https://doi.org/10.1513/AnnalsATS.201510-667OC>
2. Bzovsky S, Phillips MR, Guymer RH, Wykoff CC, Thabane L, Bhandari M, et al. The clinician's guide to interpreting a regression analysis. *Eye (Lond).* 2022;36(9):1715-1717. <https://doi.org/10.1038/s41433-022-01949-z>