International Journal of Clinical and Experimental Medicine Research, 2025, 9(6), 680-683
https: //www.hillpublisher.com/journals/ijcemr/
ISSN Online: 2575-7970
ISSN Print: 2575-7989

The Application and Challenges of Regression Analysis Methods in Clinical Research

Haicheng Fei, Hongying Jia*

Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250100, Shandong, China.

How to cite this paper: Haicheng Fei, Hongying Jia. (2025) The Application and Challenges of Regression Analysis Methods in Clinical Research. *International Journal of Clinical and Experimental Medicine Research*, 9(6), 680-683.
DOI: 10.26855/ijcemr.2025.11.019

Received: October 30, 2025
Accepted: November 28, 2025
Published: December 25, 2025

***Corresponding author:** Hongying Jia, Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250100, Shandong, China.

Abstract

Regression analysis is a fundamental statistical method widely employed in clinical research to explore the quantitative relationships between exposure variables and various outcomes. By using regression models, researchers can identify key associations between predictors and outcomes, which is essential for making data-driven clinical decisions. The main types of regression models include linear regression, logistic regression, Cox proportional hazards regression, Poisson regression, and negative binomial regression. These models are tailored to different types of dependent variables, such as continuous outcomes, binary outcomes, survival times, and event counts. Each regression model has its own assumptions and limitations, which must be carefully considered to ensure valid results. This study provides an in-depth overview of these regression models, their application characteristics, and common assumptions in clinical research. Additionally, it evaluates their potential contributions to precision medicine and discusses the challenges and solutions in clinical practice, aiming to provide theoretical and application reference for clinical researchers and biostatisticians.

Keywords

Regression analysis; Clinical research; Regression models

1. Introduction

Regression analysis is a type of method in statistics used to explore the relationship between a dependent variable and one or more independent variables, and can also identify risk factors related to prognosis and calculate the risk score of an individual's prognosis [1]. With the accumulation of clinical research, especially a large number of electronic health records and multi-center randomized trial data, the importance of regression models in disease risk prediction, treatment effect evaluation, prognosis analysis, and other aspects is increasing day by day [2-5]. Through regression analysis, researchers can adjust confounding factors, extract true associations between variables, estimate effect indicators, and transform individual risks into decision support, which is conducive to promoting the development of precision medicine [6]. This study introduces multiple regression analysis methods, with a focus on selecting the appropriate regression analysis approach for specific clinical research.

2. The Basic Principles and Methods of Regression Analysis

2.1 Linear Regression Model

Linear regression is used to explore the linear relationship between the dependent variable and the independent variable. It can be divided into simple linear regression and multiple linear regression. In clinical research, linear regression is widely applied in fields such as disease risk prediction, treatment effect evaluation, and health indicator analysis [7]. Chen et al. used multiple linear regression to explore the clinical and psychological factors influencing the

sleep quality of patients with chronic migraine [8]. The advantage of the linear regression model lies in its intuitive parameter interpretation, which can effectively estimate the influence of the independent variable on the dependent variable [9]. However, when using a linear regression model, certain assumptions must be met, including the linear relationship between the dependent variable and the independent variable, the normal distribution of the error term, and the homogeneity of the variance of the residuals etc. If these assumptions do not hold true, it may be necessary to make model adjustments or use other, more complex regression methods.

2.2 Logistic Regression Model

Logistic regression is a regression analysis method often used for binary or multi-classification problems. When the dependent variable is a categorical variable and linear regression is not applicable, Logistic regression is used to model the probability of the dependent variable through the log-odds function. In clinical research, Logistic regression is widely applied in disease prediction, patient risk assessment, and the analysis of treatment effects, etc. [10]. Ma et al. used Logistic regression to analyze the risk of diabetes in middle-aged and elderly adults and found that factors such as smoking, drinking, hypertension, and mental illness were independent predictors of diabetes [11]. One significant advantage of the Logistic regression model is that it can directly output the probability that the dependent variable is of a certain category, which is convenient for clinical decision-making and risk assessment. In addition, Logistic regression is particularly effective for non-normally distributed data and categorical dependent variables [12]. However, the application of Logistic regression also has certain limitations. It is necessary to check whether the model assumption that the relationship between the independent variable and the dependent variable is a linear relationship of logarithmic probability is met. If the sample size is small or the number of a certain category in the sample is too small, the estimation results of Logistic regression may be unstable.

2.3 Cox Proportional Hazards Regression Model

Cox Proportional Hazards Regression is a widely used statistical method in survival analysis, especially suitable for studying the relationship between time-event data (such as patient survival period, event occurrence time, etc.) and a set of independent variables. The Cox regression model does not require a specific distribution of dependent variables. It measures the impact on the risk of event occurrence by estimating the Hazard Ratio (HR) of each independent variable. In clinical research, Cox regression is widely used to evaluate clinical factors, treatment effects, and other related variables that affect patients' survival period. In a chronic kidney disease cohort study, the Cox proportional hazards model was used to assess the time-event relationship between 24-hour urinary calcium excretion and the risk of cardiovascular disease (CVD) and all-cause mortality. The model adjusted for multiple confounding factors and revealed the association between different urinary calcium levels and the risk of endpoint events [13]. The advantage of the Cox regression model lies in its flexibility, which allows for the handling of deleted data and does not require specific assumptions about the baseline risk function. Cox regression can handle the time-dependent and right-censoring issues of survival data and is widely applied in fields such as cardiology and chronic disease management [14]. However, Cox regression also has certain limitations. The most important assumption is the proportional risk assumption, that is, the risk ratio between different groups should remain constant. If the proportional hazards assumption does not hold, other methods need to be adopted, such as the Accelerated Failure Time model (AFT) or the Cox regression model with time-dependent covariates.

2.4 Poisson Regression Model

The Poisson regression model is a generalized linear regression model used to analyze counting outcome variables, which is suitable for establishing the relationship between independent variables and outcomes when the dependent variable is the number of events (such as the number of hospitalizations, emergency department visits, and outpatient visits). Poisson regression assumes that the dependent variable follows a Poisson distribution and relates the independent variable to the expected count of the outcome through a logarithmic link function. It is often used for risk estimation and prediction of count data in clinical studies, especially when the outcome variable is not a continuous value but a non-negative integer. In clinical research practice, Poisson regression is widely used to analyze count outcomes such as days of stay, number of hospitalizations, and health care utilization. Weaver et al. used Poisson regression to analyze the hospitalization days of dialysis patients and explored the influence of different treatment methods on hospitalization rates. At the same time, they compared the applicability of the Poisson model with other counting models and revealed the actual role of the Poisson model in the analysis of counting hospital indicators [15]. The advantage of the Poisson regression model is that it can effectively model the event-count outcome, and the

interpretation parameter is the Incidence Rate Ratio (IRR), which is convenient to describe the influence of independent variables on the event rate [16]. However, the model needs to satisfy the assumptions of conditional variance equal to the mean and independence between events. If there is excessive dispersion or a large number of zero values in the data, more complex methods such as negative binomial regression or zero-inflation models need to be considered to improve the fitting effect.

2.5 Negative Binomial Regression Model

Negative Binomial Regression is a generalized linear model used for analyzing counting outcome data and is an extended form of the traditional Poisson regression. Compared with Poisson regression, which assumes that the variance of the dependent variable is equal to the mean, when the count data shows excessive dispersion, that is, when the variance is significantly greater than the mean, Poisson models often produce biased estimation results. In this case, negative binomial regression introduces an additional dispersion parameter, enabling the model to better fit the distribution characteristics of the data, thereby enhancing the stability and accuracy of the estimation. Its model coefficients are usually linearly associated with covariates through logarithmic link functions and can be interpreted as IRR after exponential transformation. In clinical research, negative binomial regression models are widely used to analyze count-based outcome variables such as the number of hospitalizations, the frequency of cardiovascular events, and the utilization rate of medical resources [17]. In a study from the RESPOND multicenter cohort, researchers used multivariate negative binomial regression to evaluate the association between integrase strand-transfer inhibitors (INSTIs) exposure and the incidence of CVD events, and adjusted traditional cardiovascular risk factors in the model, showing that drug exposure at certain time periods was significantly associated with the incidence of cardiovascular events [18]. This study demonstrates the effectiveness and practical value of negative binomial regression in handling event count data and estimating IRR. The advantage of the negative binomial regression model lies in its flexibility and robustness, and it is especially suitable for overly discrete cases of counting data. It can provide more accurate coefficient estimates and standard errors, especially in cases where the event frequency is too high or the frequency distribution is uneven. However, although negative binomial regression has strong applicability, model diagnosis is still necessary when using this model. Especially when choosing covariates, too many or too few independent variables may lead to overfitting or underfitting of the model. In addition, researchers should also examine the degree of excessive dispersion of the data and ensure that the model Settings conform to the characteristics of the data.

3. Conclusion

Regression analysis methods have significant application value in clinical research, especially in areas such as disease risk prediction, treatment effect evaluation, and health indicator analysis [19]. By using different types of regression models, researchers can effectively reveal the relationship between independent variables and dependent variables, and provide quantitative risk assessment and treatment guidance for precision medicine.

However, each regression model has certain assumptions in practical applications and faces challenges in data preprocessing, model diagnosis, and result interpretation. For instance, linear regression assumes a linear relationship between the dependent variable and the independent variable, while Cox regression requires verifying the proportional hazards assumption. Poisson regression requires that the variance of the data be consistent with the mean, while negative binomial regression can address the issue of excessive discreteness. Facing these challenges, researchers should adjust the model based on the actual situation and select appropriate methods for hypothesis testing and model optimization.

In the future, with the development of big data and artificial intelligence technologies, regression analysis methods will become more precise and diverse. Combining the advantages of machine learning and traditional regression models can provide more powerful analytical tools for clinical research. Further research should focus on how to handle high-dimensional data, how to optimize the hypothesis testing and result interpretation of regression models, and how to combine regression analysis with precision medicine strategies to promote the progress of personalized treatment and disease prevention.

References

[1] Zapf A, Wiessner C, König IR. Regression analyses and their particularities in observational studies—part 32 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2024;121(4):128-134.

[2] Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with

electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198-208.

[3] Matheny ME, Ricket I, Goodrich CA, et al. Development of electronic health record-based prediction models for 30-day readmission risk among patients hospitalized for acute myocardial infarction. JAMA Netw Open. 2021;4(1):e2035782.

[4] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-1931.

[5] Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-73.

[6] Gao Y, Xiang L, Yi H, Song J, Sun D, Xu B, et al. Confounder adjustment in observational studies investigating multiple risk factors: a methodological study. BMC Med. 2025;23(1):132.

[7] Schober P, Vetter TR. Linear regression in medical research. Anesth Analg. 2021;132(1):108-109.

[8] Chen YM, Wang JH, Liang CS, Lin YK, Yang FC. Clinical and psychological predictors of sleep quality in chronic migraine: a preliminary retrospective analysis study. BMC Neurol. 2025;25(1):156.

[9] Roustaei N. Application and interpretation of linear-regression analysis. Med Hypothesis Discov Innov Ophthalmol. 2024;13(3):151-159.

[10] Garwe T, Choi J. An introduction to clinical prediction models using logistic regression in acute care surgery research: methodologic considerations and common pitfalls. J Trauma Acute Care Surg. 2025;98(5):699-703.

[11] Ma S, Sun Q, Xu Y, Tu Q, Xu S. Identification of risk factors for diabetes in Chinese middle-aged and elderly adults. Sci Rep. 2025;15(1):12550.

[12] Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. Int J Epidemiol. 2014;43(3):962-970.

[13] Yuan QJ, Xie YY, Wang JW, et al. [Study on the correlation between urinary calcium levels and severity and prognosis of chronic kidney disease]. Zhonghua Liu Xing Bing Xue Za Zhi. 2025;46(2):264-272.

[14] Deo SV, Deo V, Sundaram V. Survival analysis-part 2: Cox proportional hazards model. Indian J Thorac Cardiovasc Surg. 2021;37(2):229-233.

[15] Weaver CG, Ravani P, Oliver MJ, Austin PC, Quinn RR. Analyzing hospitalization data: potential limitations of Poisson regression. Nephrol Dial Transplant. 2015;30(8):1244-1249.

[16] Hutchinson MK, Holtman MC. Analysis of count data using poisson regression. Res Nurs Health. 2005;28(5):408-418.

[17] Teo YH, Tam WW, Koo CY, et al. Sleep apnea and recurrent heart failure hospitalizations after coronary artery bypass grafting. J Clin Sleep Med. 2021;17(12):2399-2407.

[18] Neesgaard B, Greenberg L, Miró JM, et al. Associations between integrase strand-transfer inhibitors and cardiovascular disease in people living with HIV: a multicentre prospective study from the RESPOND cohort consortium. Lancet HIV. 2022;9(7):e474-e485.

[19] Zakhary BL, Choi J. Regression analysis in clinical research. J Trauma Acute Care Surg. 2025;99(1):29-35.