Letter to the Editor

JOURNAL
OF HEPATOLOGY

Critical analysis of statistical methods in albumin treatment study: Advocating for non-linear approaches in biological data analysis

To the Editor:
Schleicher *et al.* conducted a study to investigate the effect of albumin treatment duration on response rates and outcomes in patients with cirrhosis and acute kidney injury.[1] To identify the variables associated with responses to albumin treatment beyond 48 hours, they employed multivariable logistic regression models that incorporated all variables demonstrating significance in univariable analyses.[1]

While logistic regression remains a powerful and widely adopted statistical tool in medical research, its application in biological data analysis requires careful consideration of inherent limitations. This paper raises critical concerns about potential violations of fundamental assumptions in logistic regression analysis. The method assumes a linear relationship between independent variables and log odds of the outcome, independence of observations, and absence of multicollinearity among predictors – assumptions that are often challenging to meet in biological systems.[2–7] When these assumptions are violated, the model can produce distorted outcomes through overfitting or underfitting, leading to misguided interpretations and conclusions.[2–7] Furthermore, biological data typically exhibits non-linear and non-parametric characteristics that fundamentally conflict with the rigid assumptions of traditional logistic regression, potentially obscuring genuine relationships and complex interactions among variables. In the context of albumin treatment studies, such statistical misalignments could lead to incorrect inferences about treatment efficacy and patient outcomes, particularly concerning for vulnerable populations where accurate analysis is crucial for clinical decision-making.

Additionally, feature importances derived from logistic regression can be biased due to the method-specific nature of model interpretation, especially in the absence of ground truth values. To more accurately assess true associations without these standard benchmarks, three critical components must be considered: data distribution, the statistical relationships between variables, and the validation of statistical significance through *p* values.

Therefore, it is essential for researchers to thoroughly understand the data analysis tools they employ, including the assumptions of parametric testing, independence of observations, and the necessity to check for multicollinearity. To achieve accurate data analysis, researchers should consider utilizing non-linear and non-parametric approaches or implementing multifaceted strategies that cross-validate multiple outcomes. By doing so, they can better capture the complexity of biological data and obtain more reliable and valid results.

In light of these concerns, this paper advocates for the use of random forests for predictive modeling, alongside non-linear, non-parametric statistical methods such as Spearman's rank correlation and Kendall's tau for examining monotonic relationships,[8] complemented by MI (mutual information) analysis for exploring non-monotonic relationships and complex interactions among multiple variables.[9] By employing these approaches, researchers can enhance their understanding of variable interdependencies and improve the robustness of their conclusions in studies related to biological data.

Yoshiyasu Takefuji*
Musashino University, Data Science Department, Tokyo, Japan
*Corresponding author. Address: 3-3-3 Ariake, Koto-Ku, Tokyo
135-8181, Japan.
E-mail address: takefuji@keio.jp

Received 31 March 2025; Accepted 7 April 2025; Available online xxx
https://doi.org/10.1016/j.jhep.2025.04.009

Financial support

The authors did not receive any financial support to produce this manuscript.

Conflict of interest

The author has no conflict of interest.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Yoshiyasu Takefuji completed this research and wrote this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2025.04.009.

References

[1] Schleicher EM, Karbannek H, Weinmann-Menke J, et al. Effect of albumin treatment duration on response rates and outcomes in patients with cirrhosis and acute kidney injury. J Hepatol 2025. https://doi.org/10.1016/j.jhep.2025.03.008.

[2] Dey D, Haque MS, Islam MM, et al. The proper application of logistic regression model in complex survey data: a systematic review. BMC Med Res Methodol 2025;25:15. https://doi.org/10.1186/s12874-024-02454-5.

[3] Pinheiro-Guedes L, Martinho C, O Martins MR. Logistic regression: limitations in the estimation of measures of association with binary health outcomes. Acta Med Port 2024;37(10):697–705. https://doi.org/10.20344/amp.21435.

[4] Wang T, Tang W, Lin Y, et al. Semi-supervised inference for nonparametric logistic regression. Stat Med 2023;42(15):2573–2589. https://doi.org/10.1002/sim.9737.

[5] Rifada M, Chamidah N, Ningrum RA. Estimation of nonparametric ordinal logistic regression model using generalized additive models (GAM) method based on local scoring algorithm. AIP Conf Proc 2022;2668(1):070013. https://doi.org/10.1063/5.0111771.

[6] Work JW, Ferguson JG, Diamond GA. Limitations of a conventional logistic regression model based on left ventricular ejection fraction in predicting coronary events after myocardial infarction. Am J Cardiol 1989;64(12):702–707. https://doi.org/10.1016/0002-9149(89)90751-0.

[7] van Maanen L, Katsimpokis D, van Campen AD. Fast and slow errors: logistic regression to identify patterns in accuracy–response time relationships. Behav Res 2019;51:2378–2389. https://doi.org/10.3758/s13428-018-1110-z.

[8] Okoye K, Hosseini S. Correlation tests in R: pearson cor, kendall's tau, and spearman's rho. In: R programming. Singapore: Springer; 2024. https://doi.org/10.1007/978-981-97-3385-9_12.

[9] Erman B. Mutual information analysis of mutation, nonlinearity, and triple interactions in proteins. Proteins 2023;91(1):121–133. https://doi.org/10.1002/prot.26415.