

Limitations of logistic regression in analyzing complex ambulatory blood pressure data: a call for non-parametric approaches

Yoshiyasu Takefuji  *

Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

Online publish-ahead-of-print 28 July 2025

This commentary refers to ‘Ambulatory blood pressure monitoring, European guideline targets, and cardiovascular outcomes: an individual patient data meta-analysis’, by D.-Y. Zhang et al., <https://doi.org/10.1093/eurheartj/ehaf220> and the discussion piece ‘Clarifying the use of multinomial logistic regression in analysis: a methodological response’, by D.-Y. Zhang et al., <https://doi.org/10.1093/eurheartj/ehaf543>.

Zhang et al.¹ on behalf of the International Database of Ambulatory Blood Pressure (ABP) in Relation to Cardiovascular Outcomes Investigators, performed an individual patient data meta-analysis to assess the relationship between office blood pressure (OBP) and ABP, both measured at baseline.¹ Through multinomial logistic regression, they analyzed the probability of individuals belonging to each of four descending quartiles of the percentage of time that blood pressure remained within the target range, as established by the 2024 European Society of Cardiology guidelines. Their analysis encompassed a wide range of systolic and diastolic OBP values in both untreated and treated study participants.

However, Zhang et al.¹ do not appear to address the critical importance of verifying the underlying assumptions of their data analysis methods—a key step to ensure valid results. As outlined in the Edinburgh Medical School’s data analysis textbook and supported by numerous peer-reviewed articles,^{2–9} violating fundamental assumptions of logistic regression—such as linearity of the logit for continuous predictors, independence of observations, lack of multicollinearity, and appropriate distribution of residuals—can render statistical results, including *P*-values and confidence intervals, unreliable and potentially misleading against non-linear non-parametric biological data analysis. Because logistic regression is a parametric technique, it is particularly vulnerable to these violations; failure to assess and address these assumptions may compromise both the validity and interpretability of their study’s findings.

Given the inherently non-linear and non-parametric nature of much biological data, reliance on logistic regression in this context raises significant methodological and empirical concerns. Logistic regression assumes linearity of the logit for continuous variables, absence of multicollinearity, independence of observations, and sufficient sample size for each category. However, biological phenomena often involve

complex interactions in which multiple variables influence each other simultaneously and in context-dependent ways that cannot be captured by simple additive or multiplicative models. Additionally, many biological relationships are non-monotonic, meaning the associations between variables do not consistently increase or decrease but may change direction or magnitude at different levels—a limitation that logistic regression cannot adequately address due to its assumption of monotonicity in predictor effects. Furthermore, logistic regression does not capture the directionality of information flow between variables, which can be critical when studying causality or influence in physiological systems.

When these assumptions are violated, as frequently happens with complex, heterogeneous biomedical data, conclusions drawn from logistic regression may be biased, incomplete, or misleading. To ensure rigorous inference and valid conclusions, it is therefore essential for researchers to assess, report, and—if necessary—adjust their modelling approach accordingly. Alternatives such as non-linear, non-parametric, or information-theoretic methods—like Effective Transfer Entropy¹⁰—offer superior handling of complex variable interactions, non-monotonic patterns, and the explicit quantification of directional information transfer, making them especially suitable for the nuanced analysis required in contemporary biomedical research.

Declarations

Disclosure of Interest

Nothing to declare.

References

1. Zhang DY, An DW, Yu YL, Melgarejo JD, Boggia J, Martens DS, et al. Ambulatory blood pressure monitoring, European guideline targets, and cardiovascular outcomes: an individual patient data meta-analysis. *Eur Heart J* 2025;**46**:2974–87. <https://doi.org/10.1093/eurheartj/ehaf220>
2. Dey D, Haque MS, Islam MM, Aishi UI, Shammy SS, Mayen MSA, 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**:697–705. <https://doi.org/10.20344/amp.21435>

* Corresponding author. Tel: +81-3-6865-7681, Email: takefuji@keio.jp

© The Author(s) 2025. Published by Oxford University Press on behalf of the European Society of Cardiology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

4. Wang T, Tang W, Lin Y, Su W. Semi-supervised inference for nonparametric logistic regression. *Stat Med* 2023;**42**:2573–89. <https://doi.org/10.1002/sim.9737>
5. Akturk B, Beyaztas U, Shang HL, Mandal A. Robust functional logistic regression. *Adv Data Anal Classif* 2025;**19**:121–45. <https://doi.org/10.1007/s11634-023-00577-z>
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**:702–7. [https://doi.org/10.1016/0002-9149\(89\)90751-0](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**: 2378–89. <https://doi.org/10.3758/s13428-018-1110-z>
8. Osborne J. A practical guide to testing assumptions and cleaning data for logistic regression. In: *A Practical Guide to Testing Assumptions and Cleaning Data for Logistic Regression*. SAGE Publications, Ltd, 2015, 84–130.
9. Zulfadhli M, Budiantara IN, Ratnasari V. Nonparametric regression estimator of multi-variable Fourier series for categorical data. *MethodsX* 2024;**13**:102983. <https://doi.org/10.1016/j.mex.2024.102983>
10. Ekhlasi A, Nasrabadi AM, Mohammadi M. Improving transfer entropy and partial transfer entropy for relative detection of effective connectivity strength between time series. *Commun Nonlinear Sci Numer Simul* 2023;**126**:107449. <https://doi.org/10.1016/j.cnsns.2023.107449>