LETTER TO THE EDITOR

Methodological Limitations in Breast Cancer Survivorship Research: Beyond Linear Regression

To the Editor:

Pritzl et al[1] conducted a comprehensive investigation into the trajectories of global mental health (GMH) and global physical health (GPH) among breast cancer survivors over a 4-year period. To evaluate changes in PROMIS-10 scores throughout this longitudinal study, the researchers employed linear regression modeling techniques. Specifically, the team constructed linear regression models that assessed the association between PROMIS-10 scores and time since survey completion, while carefully accounting for individual variability by incorporating subject ID as a blocking term in their statistical analysis.

The article raises significant theoretical and empirical concerns regarding the use of linear regression due to the linear parametric nature of linear regression, leading to potentially erroneous interpretations and conclusions against the nonlinear nonparametric nature of biologic data analysis.[2–5] Linear regression assumes a straight-line relationship between variables (linearity assumption), which is often inappropriate for biologic processes that frequently follow complex, nonlinear patterns with plateaus, thresholds, or exponential changes. Additionally, parametric assumptions in these models—including normally distributed residuals, homoscedasticity (constant variance), and independence of observations—are rarely satisfied in biologic systems where variability often increases with the mean, distributions are frequently skewed, and measurements over time are inherently correlated.

Pritzl et al[1] must recognize that violating assumptions of data analysis tools such as linear regression potentially distorts the outcomes and skews conclusions. When linear methods are applied to nonlinear data and parametric models are applied to nonparametric data, the outcomes including important metric scores such as *P* values are severely skewed, with type I and type II errors becoming uncontrollable. This is particularly problematic in longitudinal health research where recovery patterns, treatment responses, and disease progression typically follow nonlinear trajectories. Numerous peer-reviewed articles have documented these critical issues on assumption violations, emphasizing the need for more appropriate nonlinear and nonparametric approaches to accurately model biologic reality.[2–5]

Supervised machine learning models validate prediction accuracy against ground truth targets, but feature importances lack similar validation benchmarks. High prediction accuracy doesn't guarantee reliable feature importance rankings, because these represent distinct evaluation dimensions. Linear regression struggles with complex, nonmonotonic variable interactions that characterize biologic systems. This paper advocates for nonlinear nonparametric methods such as effective transfer entropy (ETE), which can effectively capture multidimensional, nonmonotonic patterns in data where traditional linear approaches fail to reveal true underlying relationships despite potentially strong predictive performance.

Yoshiyasu Takefuji, PhD[a]

[a]*Faculty of Data Science, Musashino University, Tokyo, Japan.*
Email: takefuji@keio.jp

J Natl Compr Canc Netw 2025;23(8):e257060
doi:10.6004/jnccn.2025.7060

References

1. Pritzl SL, Vierkant RA, Ruddy KJ, et al. Trajectory of global mental health and global physical health in breast cancer survivorship. J Natl Compr Canc Netw 2025;23:141–146.
2. Zuur AF, Ieno EN, Walker NJ, et al. Limitations of linear regression applied on ecological data. In: Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health. Springer; 2009:11–33.
3. Anandhi P, Nathiya E. Application of linear regression with their advantages, disadvantages, assumption and limitations. Int J Stat Appl Math 2023;8:133–137.
4. Bansal S, Singh G. Multiple linear regression based analysis of weather data: assumptions and limitations. In: Shaw RN, Paprzycki M, Ghosh A, eds. Advanced Communication and Intelligent Systems. Springer; 2023:221–238.
5. Jarantow SW, Pisors ED, Chiu ML. Introduction to the use of linear and nonlinear regression analysis in quantitative biological assays. Curr Protoc 2023;3:e801.

AUTHORS' REPLY

To the Letter to the Editor by Takefuji: Methodological Limitations in Breast Cancer Survivorship Research: Beyond Linear Regression

We appreciate the opportunity to respond to Dr. Takefuji's letter. He raises "theoretical and empirical concerns" regarding our use of parametric linear regression models to compare within-participant changes in self-reported mental and physical health scores over time among breast cancer survivors. To recap our approach for the reader, we first employed linear regression models to compare global within-participant changes in health scores collected annually via survey over a 4-year period. For these models, our *y* variable was the continuously-distributed physical health score and our *x* variables were the years of survey. Participants could contribute up to 5 surveys: one at baseline and one for each of years 1, 2, 3, and 4 post–breast cancer diagnosis. Our experimental