DOI: 10.2478/bile-2025-0007

Biometrical Letters
Vol. 62 (2025), No. 2, 107 -110

Incompatibility of Model Specifications in GLMs: A Response to the Misclaim of Neuhaus and Jewell (2025)

Changyong Feng[1,2], Hongyue Wang[1], Honghong Liu[1]

[1]Departments of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA, e-mail: changyong feng@urmc.rochester.edu

[2]Department of Anesthesiology and Perioperative Medicine, University of Rochester, Rochester, NY 14642, USA

SUMMARY

In our paper (Feng et al., 2024), we identified a critical flaw in the work of Neuhaus and Jewell (1993). Although Neuhaus and Jewell (2025) argue that we misinterpreted their results, we respectfully disagree. In this paper, we provide a detailed analysis demonstrating that their claim is unfounded.

Key words: Conditional expectation; Regression analysis; Generalized linear model.

1. Introduction

In regression analysis, given a set of covariates, many different model specifications are possible. However, some specifications may fail to satisfy the fundamental properties of conditional expectation; such models are said to be *incompatible*. Conclusions drawn from an incompatible model are invalid and lack meaningful interpretation. In a recent paper (Feng et al., 2024), we examined this issue in the context of generalized linear models (GLMs), using Neuhaus and Jewell (1993) as an illustrative example to show that their model specifications are incompatible. In response, Neuhaus and Jewell (2025) have claimed that we misinterpreted their work. In this paper, we demonstrate that their claim is unfounded.

2. The claim of Neuhaus and Jewell (2025) is invalid

In this section, we rigorously demonstrate why the claim made by Neuhaus and Jewell (2025) is invalid. We follow their notation for clarity.

Let Y be an outcome variable and let X and Z be two covariates. Neuhaus and Jewell (2025) assume the "data arises from a true model with two covariates"

$$g(\mathrm{E}(Y \mid X, Z)) = \mu + \beta X + \gamma Z, \tag{1}$$

Here, g is "an appropriate link function" (Neuhaus and Jewell, 2025). Since we are working in the context of GLMs, we assume without loss of generality that g is nonlinear and strictly increasing. Let g^{-1} denote its inverse. To avoid trivial cases, we assume the distributions of Y, X, and Z are nondegenerate.

Neuhaus and Jewell (2025) further state that "we can focus on the implications of fitting a GLM with a single covariate"

$$g(\mathrm{E}(Y \mid X)) = \mu^* + \beta^* X. \tag{2}$$

Note that the link function g is the same in both models (1) and (2).

The key question is: Given that the true model (the regression of Y on X and Z) is of the form (1), can we assume that the regression of Y on X is of the form (2)? Since Model (2) is a marginal model derived from Model (1), compatibility requires that it be mathematically deducible from (1). If not, the specification of Model (2) is incorrect, and any conclusions drawn from it – as in Neuhaus and Jewell (2025) – are invalid. The same reasoning applies to Neuhaus and Jewell (1993).

Let us examine this derivation. From Model (1) we have

$$\mathrm{E}(Y \mid X, Z) = g^{-1}(\mu + \beta X + \gamma Z). \tag{3}$$

By the properties of conditional expectations (see, e.g., Dudley, 2002), it follows from (3) that

$$\mathrm{E}(Y \mid X) = \mathrm{E}(\mathrm{E}(Y \mid X, Z) \mid X) = \mathrm{E}(g^{-1}(\mu + \beta X + \gamma Z) \mid X),$$

which implies that

$$g(\mathrm{E}(Y \mid X)) = g(\mathrm{E}(g^{-1}(\mu + \beta X + \gamma Z) \mid X)). \tag{4}$$

If Model (2) were correctly specified, combining (2) with (4) would give

$$\mu^* + \beta^* X = g(\mathrm{E}(g^{-1}(\mu + \beta X + \gamma Z) \mid X)). \tag{5}$$

The left-hand side of (5) is a linear function of X. For a nonlinear link function g, it is unlikely that the right-hand side is always linear in X. If we cannot show that the right-hand side of (5) is linear in X, then Model (2) is incompatible with the true Model (1), and any inferences based on (2) will be invalid.

We respectfully request that Drs. Neuhaus and Jewell provide a proof that the right-hand side of (5) is always linear in X. Their claim hinges on the validity of this derivation. We attempted to establish this result but were unsuccessful–even in the special case where g is the identity function.

Theorem 1. *In* general, *Models (2) and (1) are incompatible, even in the simplest case where g is the identity function.*

Proof. If g is the identity function, then Models (1) and (2) become

$$\mathrm{E}(Y \mid X, Z) = \mu + \beta X + \gamma Z, \tag{6}$$

and

$$\mathrm{E}(Y \mid X) = \mu^* + \beta^* X, \tag{7}$$

respectively. From (6) we obtain

$$\mathrm{E}(Y \mid X) = \mathrm{E}(\mathrm{E}(Y \mid X, Z) \mid X) = \mu + \beta X + \gamma \mathrm{E}(Z \mid X).$$

Comparing this with (7) gives

$$\gamma \mathrm{E}(Z \mid X) = (\mu^* - \mu) + (\beta^* - \beta) X,$$

or equivalently,

$$\mathrm{E}(Z \mid X) = ((\mu^* - \mu) + (\beta^* - \beta) X)/\gamma.$$

This implies that $E(Z \mid X)$ must be a linear function of X. However, unless Z is conditionally linear in X, this is not guaranteed. Therefore, Models (6) and (7) are generally incompatible. □

This derivation shows that Model (2) is, in general, incompatible with the true Model (1).

3. Conclusion

Based on our rigorous mathematical derivations, we have demonstrated that the assumption made by Neuhaus and Jewell (2025) concerning the regression of Y on X in Model (2) cannot be derived from their true Model (1), which specifies the regression of Y on (X, Z). This inconsistency reveals a fundamental flaw in their model specification, rendering it mathematically invalid. We therefore respectfully request that Drs. Neuhaus and Jewell provide a detailed, step-by-step derivation showing how their formula (2) logically follows from formula (1) in Neuhaus and Jewell (2025). The conclusions drawn in both Neuhaus and Jewell (2025) and Neuhaus and Jewell (1993) rest entirely on the validity of this derivation. Without such a proof, we cannot accept their characterization of our work. A correction of the relevant results in Neuhaus and Jewell (2025) and Neuhaus and Jewell (1993) may also be warranted.

REFERENCES

Dudley, R. M. (2002): Real Analysis and Probability. New York: Cambridge University Press.

Feng, C., Wang, H., and Liu, H. (2024): Incompatibility of model specifications in generalized linear models. Biometrical Letters 61(1): 33–50.

Neuhaus, J. and Jewell, N. (1993): A geometric approach to assess bias due to omitted covariates in generalized linear models. Biometrika 80(4): 807–815.

Neuhaus, J. and Jewell, N. (2025): There is no Flaw in Neuhaus and Jewell (1993). Biometrical Letters 62(1): 101-105.

Editorial note:

As part of the discussion on the article by Feng, Wang and Liu (2024), the editors received comments from both the authors of the earlier paper (Neuhaus and Jewell, 1993) and the authors of the new paper. Both present different interpretations of the statistical methods used in generalised linear models (GLMs).

The editors have decided to publish both comments to permit open and transparent scientific debate. The dispute concerns the details of the assumptions of the statistical models, which are of both theoretical and practical importance in data analysis. The presentation of the two positions allows readers to judge for themselves the weight of the arguments, and contributes to the depth of the discussion in the field.

At the same time, the editors remind readers that differences in the interpretation of statistical methods are a natural part of scientific development, and should be presented in a spirit of mutual respect and informed criticism.