Gastroenterology 2026;■:1–2

Evaluating Linear Parametric Poisson Regression vs Nonparametric Unsupervised Learning in Ulcerative Colitis Data

Dear Editors:

Li Wai Suen et al[1] conducted the PREDICT-UC study to investigate whether early infliximab levels and clearance could predict outcomes following infliximab rescue therapy in patients with acute severe ulcerative colitis. Risk ratios (RR) and confidence intervals (CI) were estimated using modified Poisson regression with cluster-robust standard errors. Multivariable analysis identified independent predictors of day 7 nonresponse, including greater early infliximab clearance (RR, 3.96 per L/d; 95% CI, 1.83–8.58; $P = .0005$), higher Lichtiger score at infliximab administration (RR, 1.08 per point; 95% CI, 1.03–1.14; $P = .0038$), and baseline thiopurine use, which reduced nonresponse risk (RR, 0.54; 95% CI, 0.35–0.82; $P = .0043$).

Researchers must critically understand the limitations of their analytical tools, including Poisson regression, which is a linear parametric method operating within generalized linear models. It imposes strict assumptions: linearity between predictors and the log-transformed outcome, observation independence, Poisson-distributed outcomes, and mean variance equality. These assumptions are rarely fully satisfied in real-world data. When linear methods are applied to data violating underlying assumptions—such as overdispersion, zero inflation, or non-Poisson distributed counts—resulting estimates, including feature importance, RRs, and P values, may be severely distorted, leading to erroneous conclusions. To ensure a safer and more robust analysis, researchers should adopt a conservative stance by assuming data are nonlinear and nonparametric unless sufficient theoretical and empirical justification confirms otherwise. This precautionary approach prioritizes nonparametric and nonlinear methods by default, reducing assumption violation risks and enhancing the credibility and reproducibility of findings.

For validating true associations, researchers must examine two mandatory criteria: consistency and dose–response relationships.[2–6] Consistency requires that the same associations be reproducibly identified across different analytical conditions, whereas dose–response relationships necessitate that changes in predictor magnitude correspond systematically with outcome changes, together providing evidence that an observed association reflects a genuine underlying relationship rather than a statistical artifact.

To operationalize these criteria, this paper implements a novel leave-top-1-out approach to test consistency and dose–response relationships by comparing 2 ordered sets of feature importance rankings across iterative analytical conditions with a publicly available dataset.[7] Cross-validation assesses each algorithm's predictive quality, while comparing feature importance rankings evaluates consistency and dose–response relationships between iterations. This technique systematically evaluates feature importance stability by identifying top-ranked features from a complete dataset, removing the highest-ranked feature to create a reduced dataset, and comparing ranking changes between iterations. Stable and consistent rankings across iterations provide evidence of true associations, whereas substantial reorganization suggests unreliable findings driven by methodological sensitivity. This approach addresses a critical gap in traditional feature selection methods, which typically fail to verify consistency and dose–response properties under varying conditions and neglect predictive quality assessment through cross-validation.

Supplementary Table 1 reveals a striking contrast in feature ranking stability between supervised and unsupervised approaches. Supervised models such as random forest (RF) and XGBoost, despite achieving perfect cross-validation accuracy (CV6 = 1.00), demonstrate notable instability in feature ranking orders, with RF prioritizing Surgery_Cat and Surgical_Complications, whereas XGBoost shifts toward Age_dx, Age_cat5, and Age_cat10, reflecting inconsistent and divergent feature importance hierarchies.

This instability is further amplified in Poisson regression, which not only suffers from poor minority class detection, F1(1) = 0, but also produces an entirely different set of top-ranked features, including Ts_progression and Paris_E3Lumped, appearing in neither RF nor XGBoost rankings. Similarly, although feature agglomeration and highly variable gene selection yield identical feature rankings to each other, their top-ranked features—dominated by Date_dx, Age_dx, and Yr_dx—diverge substantially from supervised model findings, underscoring broader inconsistency across the methodological landscape.

In sharp contrast, the unsupervised methods feature agglomeration and highly variable gene selection exhibit perfect internal feature ranking consistency, demonstrating that removing label dependency stabilizes feature selection patterns. Most notably, the Spearman-based approach achieves the unique distinction of combining perfect predictive accuracy, CV6 = 1.00, F1(0) = 1.00, F1(1) = 1.00, with perfect feature ranking stability, consistently identifying clinically meaningful features such as Surgical_Complications, Biol_Type, and Biologics across both top 6 and top 5 ranking thresholds—a coherence unmatched by any other method. Collectively, these results suggest that supervised models trade feature ranking stability for predictive power, whereas Spearman uniquely delivers both, positioning it as the most reliable and interpretable method for this clinical prediction task.

Owing to the inherent complexity of accurately identifying true associations in observational data, no single analytical method is sufficient, and multifaceted approaches encompassing multiple complementary algorithms are, therefore, necessary to strengthen the reliability and validity of findings. For purposes of reproducibility and transparency, the Python code cv.py used in this study is publicly available at GitHub.[8]

YOSHIYASU TAKEFUJI
SciencePark Corporation
Zama-shi, Kanagawa, Japan

Supplementary Material

To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2026.04.011.

References

1. Li Wai Suen CFD, et al. Gastroenterology 2026;170:118–131.
2. Ioannidis JP. Epidemiology 2008;19:640–648.
3. Prasad V, Jena AB. JAMA 2013;309:241–242.
4. Ma Y, et al. BMC Public Health 2025;26:297.
5. Roberts MR, et al. J Invest Dermatol 2019;139:502–511.e1.
6. Lai Q, et al. Nat Commun 2025;16:8744.
7. Ihekweazu F. Mendeley Data Published April 26, 2021. https://doi.org/10.17632/fcwgv4kc5w.1.
8. GitHub. cv.py. https://github.com/y-takefuji/gastroenterology.

Conflicts of interest
The author discloses no conflicts.

https://doi.org/10.1053/j.gastro.2026.04.011