Assessment of a Modified Sandwich Estimator for Generalized Estimating Equations with Application to Opioid Poisoning in MIMIC-IV ICU Patients

Paul B Rogers, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, FDA
Julie A Stoner*, Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center

Introduction

Longitudinal regression models for correlated binary outcomes are frequently fit using generalized estimating equations (GEE). The Liang and Zeger sandwich estimator is often used in GEE to produce unbiased standard error estimation for regression coefficients in large sample settings, even when the covariance structure is misspecified. The sandwich estimator performs optimally in balanced designs when the number of participants is large with few repeated measurements. The sandwich estimator's asymptotic properties do not hold in small sample and rare-event settings. Under these conditions, the sandwich estimator underestimates the variances and is biased downwards. The goal of this research is to construct a hybrid sandwich estimator able to correctly model the variances in rare outcome and finite sample situations. Only a handful of statisticians have attempted improving the performance of the sandwich estimator under these conditions. Here, the performance of a modified sandwich estimator is compared to the traditional Liang-Zeger estimator and alternative forms proposed by authors Morel, Pan, and Mancl-DeRouen. Each estimator's performance was assessed with 95% coverage probabilities for the regression coefficients using simulated data under various combinations of sample sizes and outcome prevalence values with independence and autoregressive correlation structures.

Materials and Methods

We created a hybrid sandwich estimator and compared it to others by Pan, Mancl-DeRouen, Morel, and the traditional Liang-Zeger. The comparisons were done via simulations with real-world EHR SpO2 measurements from the MIMIC-IV ICU dataset after patient repositioning.

Literature Cited

Rogers P, Stoner J. Modification of the Sandwich Estimator in Generalized Estimating Equations with Correlated Binary Outcomes in Rare Event and Small Sample Settings. Am J Appl Math Stat. 2015;3(6):243-251. doi:10.12691/ajams-3-6-5

Pan W. On the Robust Variance Estimator in Generalised Estimating Equations. Biometrika, 2001;88(3), 901–906. http://www.jstor.org/stable/2673458

Results and Discussion

We demonstrated in simulations with sample sizes of 100 subjects and an autoregressive covariance structure with higher correlation settings (0.10 and 0.15) that all sandwich estimators produced coverage probabilities that fell below 95%. This was not observed in our earlier simulations with low correlation values. As the sample sizes dropped under these same correlation conditions, the Liang-Zeger continued to perform abysmally while the Rogers-Stoner and Pan estimators adjusted. As the sample sizes decreased under a 0.10 correlation with 10% and 5% outcome prevalences, the coverage probabilities of the Liang-Zeger continued to deteriorate, while the Rogers-Stoner and Pan estimators recovered, almost achieving 95% coverage probabilities at 40 subjects and lower.

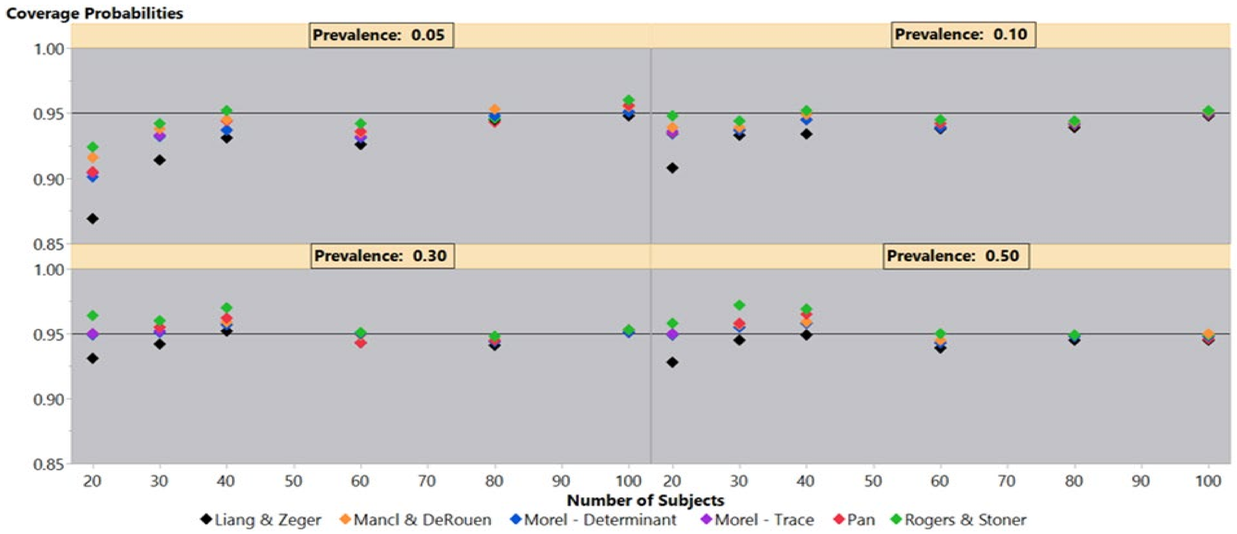

Figure 1. Coverage probabilities when estimating the regression coefficient β_1 under a simulated independence covariance structure for 5% through 50% prevalence values.

Table 1. Estimated regression coefficients, odds ratios (OR), standard errors, 95% confidence intervals (CI) from analysis of repositioning events in recorded ICU SpO_2 measurements within the first 24 hours for 20 patients.

Estimated Regression Coefficients Under an Autoregressive Correlation Structure						
Method	$\hat{\beta}_1$	$\widehat{SE(\hat{\beta}_1)}$	OR	95% CI for OR	Z-Score	p-value
Liang-Zeger	-0.2712	0.0748	0.7624	0.6583, 0.8828	-3.6238	0.0002
Mancl-DeRouen	-0.2712	0.1413	0.7624	0.5779, 1.0057	-1.9195	0.0549
Morel – (Trace)	-0.2712	0.1047	0.7624	0.6209, 0.9361	-2.5900	0.0095
Morel-(Determinant)	-0.2712	0.0998	0.7624	0.6268, 0.9272	-2.7156	0.0066
Pan	-0.2712	0.2183	0.7624	0.4969, 1.1696	-1.2423	0.2141
Rogers-Stoner	-0.2712	0.2283	0.7624	0.4872, 1.1928	-1.1877	0.2349

Conclusion

In our previously published research on simulations involving only low values of correlation, we concluded that, the choice of estimators matters, and alternatives to the Liang-Zeger estimator should be considered. This research further assessed the qualities of our hybrid sandwich estimator, reaffirming its superior performance over that of the Liang-Zeger estimator in simulations involving finite samples and low outcome prevalence. The Liang-Zeger sandwich estimator's performance suffered as the sample sizes dropped below 60 subjects under correlation settings as low as .01, when outcome prevalence values were less than 30%. This drop-off in performance was exacerbated further by lower outcome prevalence values, smaller sample sizes, and higher correlation settings. The real-world ICU practice of patient repositioning to improve opioid-induced respiratory depression showed the danger in hypothesis testing for low-prevalence outcomes with small sample sizes in GEE models with binary outcomes. Our approach provides a method for modeling rare events in finite samples on the effects of medications, drugs, and poisons.

Additional Information

Please see https://physionet.org/for more information. Contact Paul.Rogers@fda.hhs.gov for further questions or comments.

Acknowledgements

*Julie Ann Stoner (1972-2020) Dedicated to the memory of Julie Stoner, a committed public health warrior. Deceased 18 June 2020.

Division of Bioinformatics and Biostatistics, NCTR

Dr. Dong Wang
Dr. Weida Tong