GEE Report

Kevin Josey

September 9, 2018

Introduction

The recent reviews provided by Statistical Methods in Medical Research referred to literature that use a GEE approach to approximate power in study designs with clustered data. There also appears to be literature that makes use of these approaches in settings with anticipated missing data. We contend that the approximations are highly biased in studies with relatively small sample sizes (N $<$ 100). Further compounding the issue is the presumption of missing data which operationally decreases the sample size even further.

To motivate our method given these less than appealing alternatives, we conducted a small simulation study to demonstrate that fitting GEE models to correlated data with small sample sizes produce inflated Type I error rates proportionate to 1) the sample size and 2) the amount of missing data.

Simulation Design

The simulation scenarios follow a subset of conditions presented in Scenario 4 and Scenario 5 in the manuscript. For both scenarios, the mean difference scale is set to 0 ($\delta_\beta = 0$) so that the power approximations are equivalent to approximating the Type I error. The error variance scale is set to 2 ($\delta_\sigma = 2$) in Scenario 4 and 1 ($\delta_\sigma = 1$) in Scenario 5. For Scenario 4, the planned sample sizes include $N \in \{40, 80, 120, 160, 320, 400\}$. The planned sample sizes for Scenario 5 are $N \in \{20, 40, 80, 160, 200\}$. Note that Scenario 4 contains four treatment groups and 3 repeated measures and Scenario 5 contains 2 treatment groups and 5 repeated measures. In addition the Type I error rate (α-level) for Scenario 4 is 0.01 and 0.05 for Scenario 5.

The missing data patterns follow a conditional linear process with either $\boldsymbol{\pi} = \begin{bmatrix}\pi_1 & \pi_2 & \dots & \pi_2\end{bmatrix}$ or $\boldsymbol{\pi} = \begin{bmatrix}\pi_1 & \dots & \pi_1 & \pi_2\end{bmatrix}$ and $\pi_1, \pi_2 \in \{0.8, 0.9, 1\}$. Thus there a total of six missing data scenarios for either Scenario 4 or Scenario 5. The missing data indicators follow a compound symmetric covariance pattern with a correlation of $\gamma = 0.25$.

Simulation Method

For each replication, we attempted to fit a GEE model, with $\boldsymbol{y}^*$ as the response vector. Models were fit in SAS 9.4 using the PROC GENMOD procedure. The modeling approach used an unstructured working correlation and the REPEATED statement and the KR2 (Kenward Roger) option for the denominator degrees of freedom.

For each model that converged, we calculated the Wald statistic and computed a p-value according to the chi square distribution with either 12 degrees of freedom (Scenario 4) or 10 degrees of freedom (Scenario 5). Empirical power was computed as the number of p-values less than or equal to α, divided by the number of replicates for which the model converged. The simulation study used $10,000$ replications so that the error in the estimation of the empirical power would occur in the second decimal place.

Simulation Results

The results of the simulations for Scenario 4 and Scenario 5 appear in Table 1 and Table 2, respectively. Graphs showing the relationship between the planned sample size, the amount of missing data, and the Type I error rate are shown in Figure 1 and Figure 2.

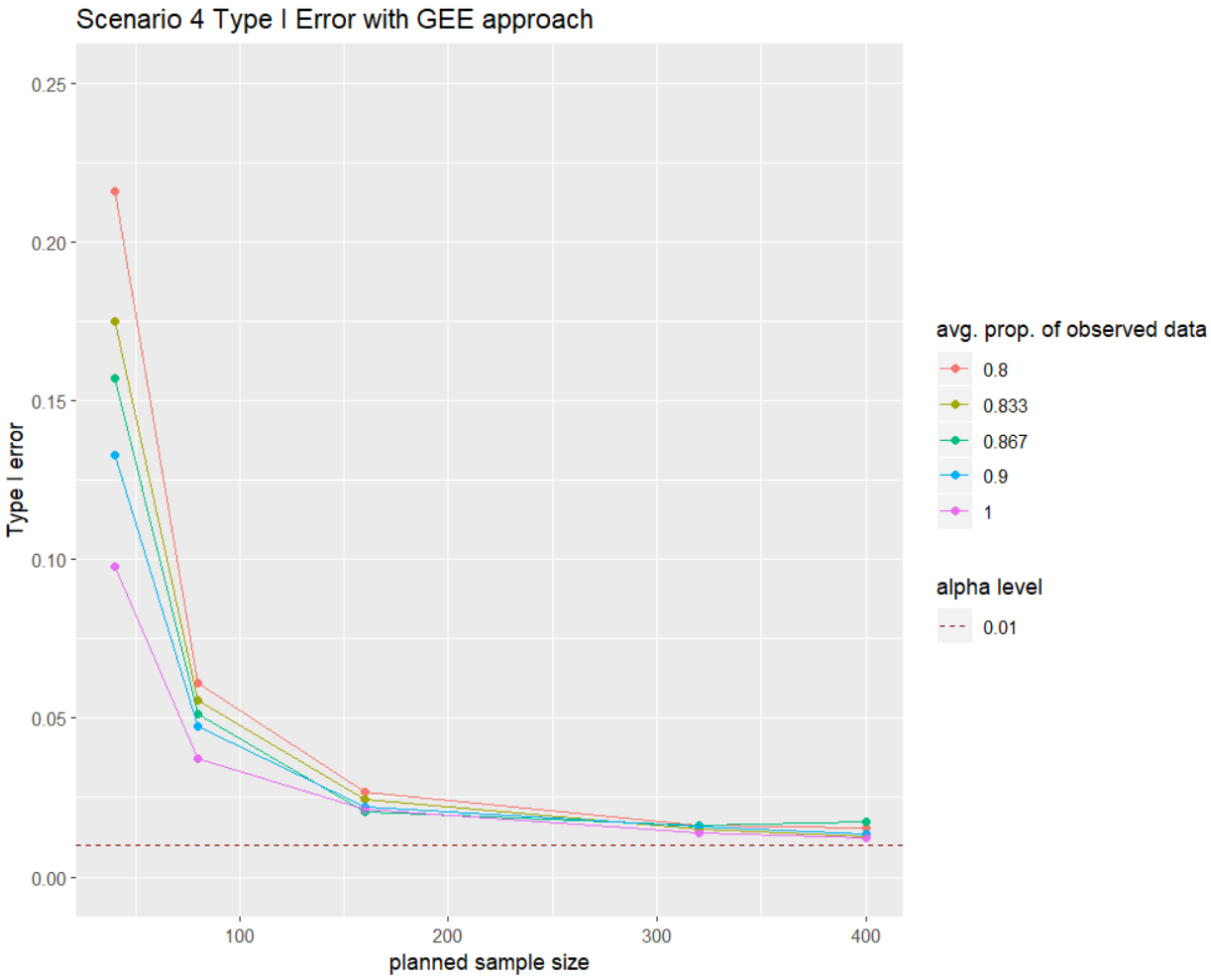
Missing Model	δ_σ	δ_β	N	$\mathcal{E}(N_m)$	Type I Error
CLP	2	0	40	32	0.216
CLP	2	0	40	33.333	0.175
CLP	2	0	40	34.667	0.157
CLP	2	0	40	36	0.133
CLP	2	0	40	40	0.098
CLP	2	0	80	64	0.061
CLP	2	0	80	66.667	0.056
CLP	2	0	80	69.333	0.052
CLP	2	0	80	72	0.048
CLP	2	0	80	80	0.037
CLP	2	0	160	128	0.027
CLP	2	0	160	133.333	0.024
CLP	2	0	160	138.667	0.021
CLP	2	0	160	144	0.022
CLP	2	0	160	160	0.021
CLP	2	0	320	256	0.016
CLP	2	0	320	266.667	0.015
CLP	2	0	320	277.333	0.016
CLP	2	0	320	288	0.016
CLP	2	0	320	320	0.014
CLP	2	0	400	320	0.015
CLP	2	0	400	333.333	0.013
CLP	2	0	400	346.667	0.017
CLP	2	0	400	360	0.014
CLM	2	0	400	400	0.013

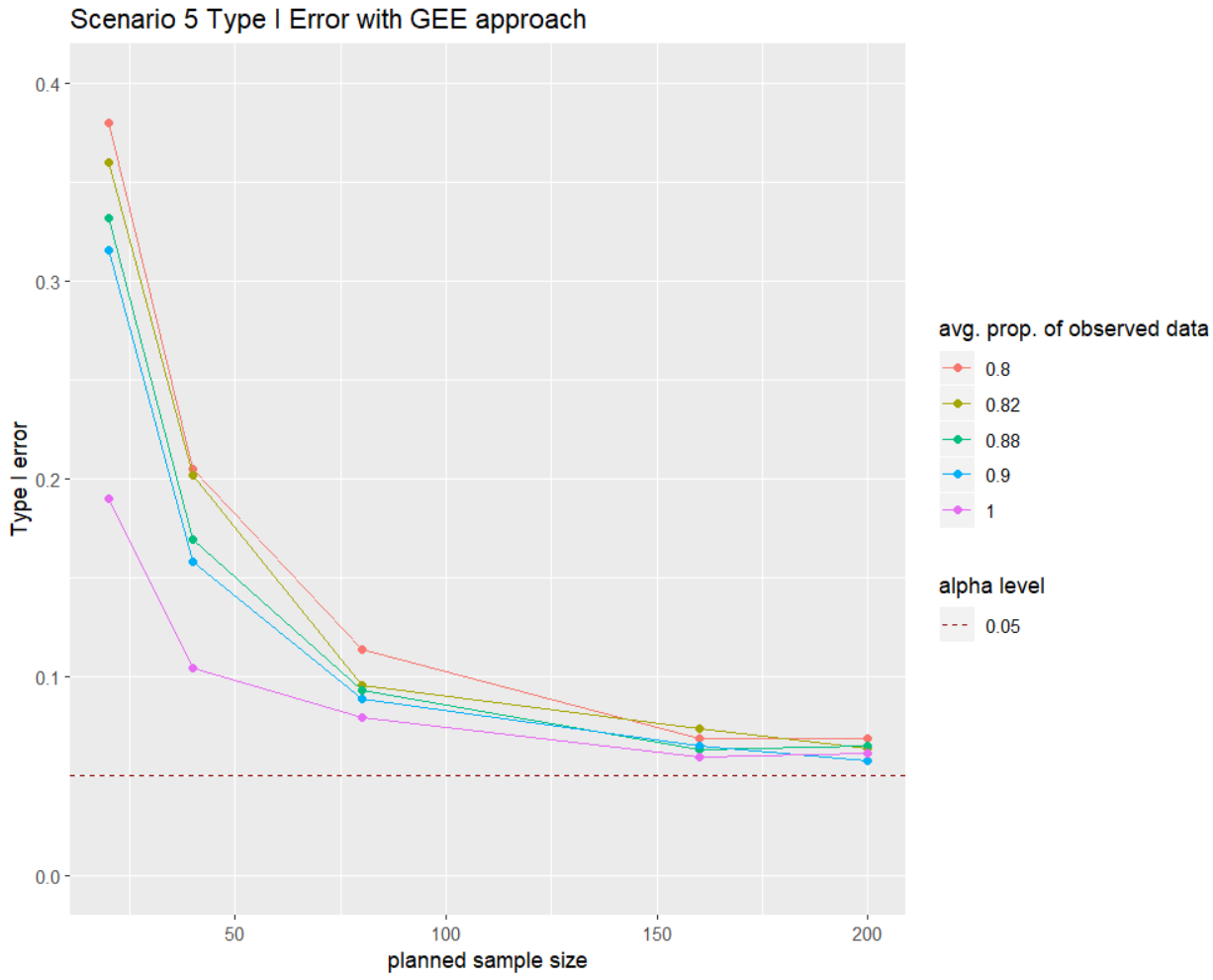
Table 1: Scenario 4 Type I error inflation

Notice that as the planned sample size increases, the Type I error stabilizes though never achieves the exact rate. Further as the rate of missing data decreases so does the Type I error inflation. The issue abounds, however when N_c, i.e the expected amount of obseverved data at each of the repeated measurements, is less than 100. These are preciously the study designs that we worry are not adequately powered. Hopefully, this small presentation is suitable to convince readers that a GEE approach given small planned sample sizes or given an experiment that incurs several missing measurements is not appropriate.

Missing Model	δ_σ	δ_β	N	$\mathcal{E}(N_m)$	Type I Error
CLP	1	0	20	16	0.380
CLP	1	0	20	16.4	0.360
CLP	1	0	20	17.6	0.332
CLP	1	0	20	18	0.316
CLP	1	0	20	20	0.190
CLP	1	0	40	32	0.205
CLP	1	0	40	32.8	0.202
CLP	1	0	40	35.2	0.170
CLP	1	0	40	36	0.158
CLP	1	0	40	40	0.105
CLP	1	0	80	64	0.114
CLP	1	0	80	65.6	0.096
CLP	1	0	80	70.4	0.093
CLP	1	0	80	72	0.089
CLP	1	0	80	80	0.080
CLP	1	0	160	128	0.069
CLP	1	0	160	131.2	0.074
CLP	1	0	160	140.8	0.063
CLP	1	0	160	144	0.065
CLP	1	0	160	160	0.060
CLP	1	0	200	160	0.069
CLP	1	0	200	164	0.064
CLP	1	0	200	176	0.065
CLP	1	0	200	180	0.058
CLP	1	0	200	200	0.062

Table 2: Scenario 5 Type I error inflation


Scenario 4 Type I Error with GEE approach
0.25
0.20
0.15
0.10
0.05
0.00
Type I error
100
200
300
400
planned sample size
avg. prop. of observed data
0.8
0.833
0.867
0.9
1
alpha level
0.01


Scenario 5 Type I Error with GEE approach
Type I error
0.4
0.3
0.2
0.1
0.0
50
100
150
200
planned sample size
avg. prop. of observed data
0.8
0.82
0.88
0.9
1
alpha level
0.05