TMJM THE MINDANAWAN JOURNAL OF MATHEMATICS
OFFICIAL JOURNAL OF THE MINDANAO STATE UNIVERSITY - ILIGAN INSTITUTE OF TECHNOLOGY

Vol. 7 (2025), no. 1, pp. 17–36 ISSN: 2094-7380 (Print) | 2783-0136 (Online)
http://doi.org/10.62071/tmjm.v7i1.721.

Generalized Estimating Equations in longitudinal data analysis in the presence of missing data

Edralyn Marie W. Rufino[1] and Bernadette F. Tubo[2,*]

[1]Philippine Statistics Authority, Municipality of Nabunturan, Davao de Oro, Philippines
e.rufino.psa@gmail.com

[2] Department of Mathematics and Statistics
MSU-Iligan Institute of Technology, 9200 Iligan City, Philippines
bernadette.tubo@g.msuiit.edu.ph

Received: 2nd January 2025 Revised: 8th April 2025

Abstract

Generalized Estimating Equations (GEE) are a statistical approach used to estimate the parameters of Generalized Linear Models (GLMs) in the presence of potential correlations among observations, particularly across different time points. GEE adjusts for within-cluster correlations, enabling more accurate and efficient parameter estimation when fitting regression models. Correctly specifying the correlation structure in a statistical model enhances the efficiency of parameter estimates. However, the challenge of missing data, which is common in many studies, can significantly impact the reliability of inferences drawn from GEE-based models. This paper explores recently developed selection criteria for identifying the underlying correlation structure, focusing on longitudinal studies with varying degrees of missingness ($\Delta m \in 5\%, 10\%, 15\%$). The criteria under investigation include: (a) Rotnitzky and Jewell Criterion (RJ), (b) Gaussian Pseudolikelihood Criterion (GP), (c) Quasi-likelihood under Independence Model Criterion (QIC), (d) Correlation Information Criterion (CIC), (e) Pardo and Alonso Criterion (PAC), and (f) Gaussian Bayesian Information Criterion (GBIC). The study examines performance across varying cluster sizes, highlighting the importance of accounting for different degrees of correlation in both complete and incomplete datasets. Both the simulation and empirical studies, across all scenarios with positive results, show that GBIC demonstrates robust and consistent performance with up to 5% to 10% missing observations.

1 Introduction

Longitudinal data are measurements repeatedly taken from a particular subject for subsequent points which gives a longitudinal aspect to the data [17]. Examples of longitudinal data are as follows: (1) CT scan values acquired for 6 years, which were 1 year apart in time interval [10]; (2) Annual body mass index (BMI) of more than 14000 boys and girls in United States, which were collected from the year 1996 to 1999 [1] and; (3) Levels of mental distress gathered from patients in every year cycle for 5 years [11].

*Corresponding author
2020 Mathematics Subject Classification: 62-08, 62J12 and 62P10
Keywords and Phrases: Selection Criteria, Correlation Structures, Missing Data, Generalized Estimating Equations, Generalized Linear Model
This research was supported by the DOST-SEI ASTHRDP Scholarship Program awarded to the first author during her MS Statistics studies at MSU-IIT.

In longitudinal data, measurements on a particular subject are naturally correlated. It is expected that the measurement of a subject at a particular point of time is affected by its previous measurement and at the same time, it will mostly likely affect its future measurements. Observations in longitudinal data are assumed to be dependent within-subject but independent from between-subject [17]. Thus, an incorporation of within-subject correlation and between-subjects variation into the model fitting is a necessity to improve the efficiency of the estimation and to enhance power [19]. The key to modeling longitudinal data is on how to account for natural correlation of the data and incorporating this in the model [17].

An extension of the GLM is usually applied for longitudinal data is the Generalized Estimating Equations or GEE. The GEE method was developed by Liang and Zeger [8] in order to produce regression estimates when analyzing repeated measures with non-normal response variables. Hence, GEE is a general statistical approach that directly fits a marginal model into a longitudinal analysis and this has been widely applied into clinical trials and biomedical studies [18].

GEE has some defining features. First, if the primary research of interest is population-based, GEE treats the variance-covariance matrix of the response variable as the nuisance of the parameter [18]. In addition, GEE is not very particular in the assumption of the distribution as long as marginal mean and variance are correctly specified as well as the link function. Lastly, GEE specifies the working correlation structure in the model. If the correlation structure of responses was misspecified, the score equation is still consistent but loses its efficiency [18]. Thus, misspecification of the working correlation structure decreases the efficiency of the derived parameter estimates, which may lead to biased estimates of the regression parameters [9]. Its efficiency may be as low as 60% compared with an estimator that correctly specified the working correlation structure [13]. Thus, a correct correlation structure will further enhance the efficiency of the parameter estimates [14].

Several authors have proposed various criteria for accurately specifying the working correlation structure in GEE. Wang [18] reviewed these developed criteria, to name: (1) Rotnitzky and Jewel's *RJ(R)* criterion; (2) Quasi-likelihood under the independence model criterion *QIC(R)* of Pan; (3) Gaussian pseudo-likelihood or the *GP(R)* of Carey and Wang; and (4) Correlation Information Criterion or the *CIC* of Hin and Wang. Aside from the different criteria reviewed by Wang [18], some authors have also proposed their criterion and the following are widely known: (1) Pardo and Alonso's Criterion *PAC(R)* and; (2) Xialou and Zhongyi's Gaussian Pseudolikelihood in the Bayesian Information Criterion *(GBIC(R))*.

On the other hand, missing data is one of the common methodological problems in longitudinal data. The longer the length of the study, the greater the likelihood of observations being dropped. It is hard to avoid or even inevitable in longitudinal studies and might influence the results of statistical inference [4]. Thus, missing data challenges the performance or the reliability of the model used to draw such conclusions in a particular study.

The goal of this study is to further investigate the performance of the different selection criteria in terms of their classification rate to correctly specify the underlying correlation structure when a complete dataset is induced with varied percentages of missing observations ($\Delta m \in \{5\%, 10\%, 15\%\}$). Also, missing data used in this study assumes a missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) mechanisms. Also, varied cluster sizes ($c \in \{3, 5, 9\}$) is explored. The simulated longitudinal dataset is generated using the model by Wang [18] given as

$$log(\mu_{ij}) = \beta_0 + \beta_1 x_{ij}$$

where $\beta_0 = \beta_1 = 0.5$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, c_i$. Moreover, the covariates x_{ij} are independent and identically distributed from a standard uniform distribution $U(0, 1)$. Several features of longitudinal data were considered and was adapted from related literature to integrate in this study , namely: (1) magnitude of correlation: $\alpha \in \{0.3, 0.7\}$ [18]; (2) size of sample: $n \in \{30, 50\}$ [9]; (3) size of cluster: $c \in \{3, 5, 9\}$ [9]; (4) underlying correlation structure: $R_i \in \{\text{IND, EXC, AR(1)}\}$, where IND denotes independence, EXC as exchangeable and AR(1) as first-order autoregressive [8]. Moreover, a correlation coefficient that is close to 0 indicates a weak correlation while a correlation coefficient that is close to 1 indicates a strong correlation. Thus, this study adapted the workings of Wang [18] to explore on the performances of selection criteria with different magnitudes of correlation. Further, due to the complexity of longitudinal data features, this study limited the outcome variable to binary response.

2 Correlation Structures

2.1 Independence (IND)

An independent working correlation structure assumes that measurements are uncorrelated within a subject and the values of its off-diagonal are 0. Thus, $R_i^I(\alpha) = I$ where I is an identity matrix written in matrix notation as

$$R_i^I(\alpha) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

It is often convenient and attractive to utilize the independent correlation due to its simplicity. Its estimates are relatively efficient on cases when the magnitude of correlation is not that large [8]. However, for time-varying covariates, the use of an independent correlation structure yields an inefficient estimator that may be as low as 60% [13].

2.2 Exchangeable (EXC)

An exchangeable working correlation structure assumes the same correlation (α) in any two responses within a subject. One correlation parameter is to be estimated in this structure. Under the assumption of exchangeable correlation structure, an arbitrary number of observations and cluster size is possible [8]. A general form of exchangeable working correlation structure is given as follows:

$$R_i^E(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{pmatrix}$$

2.3 First Order Auto-Regressive(AR(1))

The AR(1) structure allows the representation of arbitrary numbers and spacing of observations. It allows the covariance between measurements within a subject to decay over time [8]. A general form of the AR(1) working correlation structure is given as follows:

$$R_i^A(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{c_i-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{c_i-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{c_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{c_i-1} & \alpha^{c_i-2} & \alpha^{c_i-3} & \dots & 1 \end{pmatrix}$$

3 Methods

To evaluate the performance of various criteria in identifying the true underlying correlation structure, particularly under varying levels of missingness ($\Delta m \in \{5\%, 10\%, 15\%\}$) across the MCAR, MAR, and MNAR mechanisms, we assess the frequency with which each criterion correctly identifies the true correlation structure in 1,000 simulated longitudinal data replications.The wgeesel package in R is utilized to generate longitudinal datasets while allowing the modification of key features, such as sample size, cluster size, magnitude of correlation (α), correlation structure, and response type. This approach ensures a flexible and controlled framework for simulating data and testing the effectiveness of each selection criterion. The R code is available in the GitHub repository for reference.

3.1 Gaussian Pseudolikelihood (GP)

Carey and Wang [2] investigated the choice of working covariance models in the analysis of repeated measures in a GEE approach. Since the covariance model is a function of both chosen correlation structure and unobservable correlation structure, Carey and Wang [2] proposed the Gaussian Pseudolikelihood criterion for working covariance model regarded as a combination of variance function and correlation model to attain efficient regression parameter estimates in GEE. It is based on the asymptotic properties of correlation parameter and regression estimates established by Hall and Severini [5] to the solution of their proposed Extended GEE where the solutions share some features with Multivariate Gaussian likelihood function [2].

The Gaussian Pseudolikelihood criterion examines the choice among working covariance models through

$$GP(R) = -0.5 \times \sum_{i=1}^{n} (Y_i - \mu_i)' V_i^{-1} (Y_i - \mu_i) + log(|V_i|)$$

where

$$V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2} \tag{1}$$

and $A_i = Diag\{v(\mu_{i1}), \dots, v(\mu_{ic_i})\}$. GP(R) is computed in each of the various combinations of V_i and $R_i(\alpha)$ and the candidate of combination that maximizes GP(R) will be regarded as the working covariance model [2].

3.2 Gaussian Bayesian Information Criterion (GBIC)

Xiaolu and Zhongyi [20] modified the Gaussian Pseudolikelihood Criterion (GPC) of Carey and Wang [2] by substituting for the parametric likelihood in the Bayesian Information Criterion of Schwarz [16] and named it Gaussian Bayesian Information Criterion (GBIC). The GBIC accounts for the dimension of the model because the hypothesized correlation structures might have different numbers of correlation to be estimated [3].

Let β be a $p \times 1$ vector of covariates and $\alpha = (\alpha_1, \ldots, \alpha_{c_1-1})'$. Then the GBIC is given by

$$GBIC = -2GP + log(n)dim(\theta)$$

where $\theta = (\beta', \alpha')'$ [3].

In contrast to the GP criterion of Carey and Wang [2] where the correlation structure that maximizes GP is chosen, GBIC chooses the correlation structure that yields its value to a minimum. In addition, the traditional Bayesian Information Criterion of Schwarz [16] requires a correct specification of likelihood function while GBIC treats the GP criterion as a working likelihood.

3.3 Pardo and Alonso Criterion (PAC)

Pardo and Alonso [14] proposed a criterion which measures the discrepancy between the covariance matrix estimator and the specified working covariance matrix. It is defined by

$$PAC(R) = \left| \frac{det\left(\frac{1}{n}\sum_{i=1}^{n} \hat{S}_i \hat{S}_i^T\right)}{det\left(\frac{1}{n}\sum_{i=1}^{n} \hat{V}_i\right)} - 1 \right|$$

with $\hat{S}_i$ and $\hat{V}_i$ the estimates of S_i and V_i where $S_i = Y_i - \mu_i$ and variance-covariance matrix V_i defined in equation (1). Accordingly, the smaller the value of the PAC(R), the closer the true correlation structure to the corresponding working correlation structure.

3.4 Quasi-likelihood criterion (QIC)

Pan[13] proposed a model selection criterion based on the extension of Akaike's Information Criterion. AIC is a likelihood-based model anchored on the asymptomatic properties of the maximum likelihood estimator (MLE). However, AIC cannot be directly used in GEE since GEE is non-likelihood based, that is, it does not have an assumption on distribution. Thus, Pan proposed a criterion that is an extension of AIC to GEE and is called quasi-likelihood under the independence model criterion or the QIC and is defined by

$$QIC(R) = -2Q(\hat{\beta}(R); I, B) + 2\ trace(\hat{\Omega}_I \hat{V}_{LZ}) \tag{2}$$

where B denotes the data on hand and the quasi-likelihood under the working independence model is defined by

$$Q(\beta, \phi; I, B) = \sum_{i=1}^{n} \sum_{j=1}^{c_i} Q(\beta, \phi; \{Y_{ij}, X_{ij}\})$$

with the (log) quasi-likelihood function $Q(\mu, \hat{\phi}; y) = \int_y^u \frac{(y-t)}{\hat{\phi} V(t)} dt$. $\hat{V}_{LZ}$ is defined by

$$\hat{V}_{LZ} = \left[\sum_{i=1}^{n} D_i' V_i^{-1} D_i\right]^{-1} \hat{M}_{LZ} \left[\sum_{i=1}^{n} D_i' V_i^{-1} D_i\right]^{-1} \tag{3}$$

with $\hat{M}_{LZ} = \sum_{i=1}^{n} D_i' V_i^{-1} Cov(Y_i) V_i^{-1} D_i$. $\hat{\Omega}_I$ is defined by

$$\hat{\Omega}_I = \sum_{i=1}^{n} D_i' V_i D_i|_{\beta=\hat{\beta}, R_i^I = I}, \tag{4}$$

$D_i = \partial\mu_i/\partial\beta'$ and V_i defined in (1). The derivation of the quasi-likelihood equation is particular to $R_i^I = I$ where I is an identity matrix. However, the derivation of quasi-likelihood function in a more general correlation matrix is not within the scope of their study since there is no guarantee that a corresponding correlation matrix exists. Accordingly, one needs to calculate the QIC of different working correlation structure and choose the correlation structure that yields to smallest QIC value.

3.5 Correlation Information Criterion (CIC)

Hin and Wang [7] proposed a modification of QIC(R) by Pan[13] to improve its performance profile in working correlation structure modeling. Accordingly, the first term of QIC, defined in equation (2) does not contain any information on the correlation structure. On the other hand, the second term does contain information on the correlation structure. Thus, it makes sense to ignore the first term of QIC(R). For these reasons, a modification of QIC(R) was proposed and is called Correlation Information Criterion (CIC). It is defined by

$$CIC(R) = tr(\hat{\Omega}_I \hat{V}_{LZ})$$

where $\hat{\Omega}_I$ and $\hat{V}_{LZ}$ are defined in equations (4) and (3), resepectively.

3.5.1 Rotnitzky and Jewell RJ(R) crtierion

Rotnitzky and Jewell [15] introduced an approach to measure the adequacy of the particular choice of the correlation structure. Let β be a $p \times 1$ vector of regression coefficients and consider the partitioning of β such that $\beta' = (\gamma', \sigma')$ where γ is an $r \times 1$ vector of the first r components of $\beta(1 \leq r \leq p)$. Define $Q = Q_o^{-1}Q_1$ where $Q_o = n^{-1}\Sigma D_i' V_i^{-1} D_i$ and $Q_1 = n^{-1}\Sigma D_i' V_i^{-1} cov(Y_i) V_i^{-1} D_i$, the adequacy of a working correlation structure can be examined through the weights q_j for $j = 1, \ldots, r$. Thus, $q_1 = tr(Q)/r$ and $q_2 = tr(Q^2)/r$, where the values of q_1 and q_2 should be approximately 1. Hin [6] defined the RJ criterion and is given by

$$RJ(R) = \sqrt{(1-q_1)^2 + (1-q_2)^2}.$$

An accurately specified correlation structure should lead the RJ value close to 0 [3].

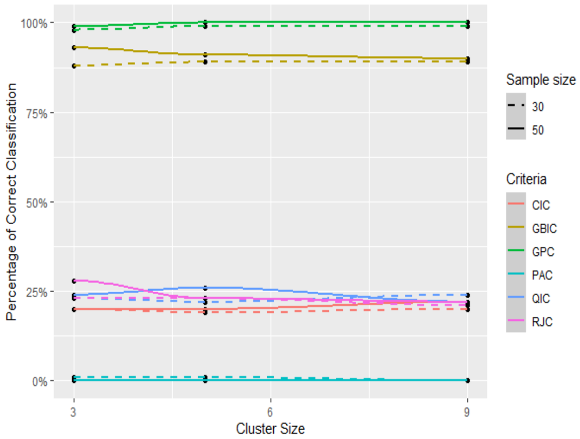
4 Results and Discussion

This section presents a detailed discussion of the simulation results under the MCAR mechanism, considering various correlation structures in the dataset. It is worth noting that similar simulation studies were also conducted for MAR and MNAR scenarios. The results under these mechanisms exhibited patterns similar to those observed under MCAR, with a few notable differences. Only the most significant findings are discussed here, with supporting figures included in the appendices.

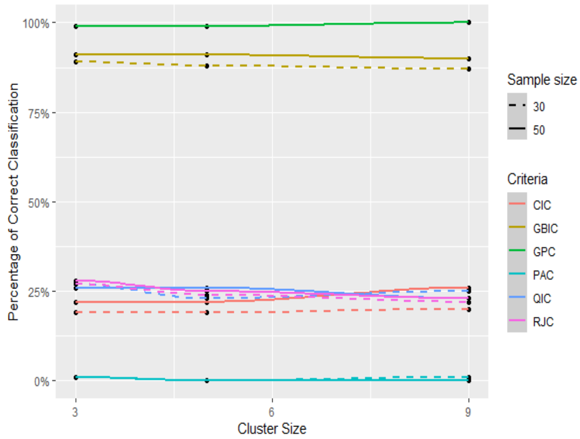
4.1 Simulation Study

4.1.1 With weak correlation $\alpha = 0.3$

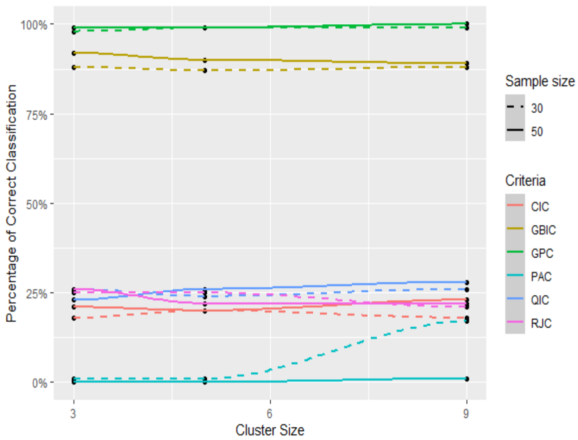
This section discusses the results of the performance of the selection criteria across three defined correlation structures, namely, INC, EXC, and AR(1), in both complete data $\Delta m = 0\%$ and data with missingness generated under a missing completely at random (MCAR) mechanism, $\Delta m \in \{5\%, 10\%, 15\%\}$. The discussion highlights the criteria that demonstrated the highest



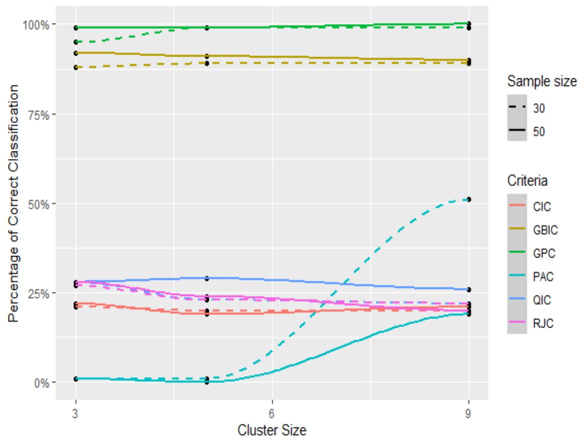
(a) Complete Data



(b) 5% MCAR



(c) 10% MCAR



(d) 15% MCAR

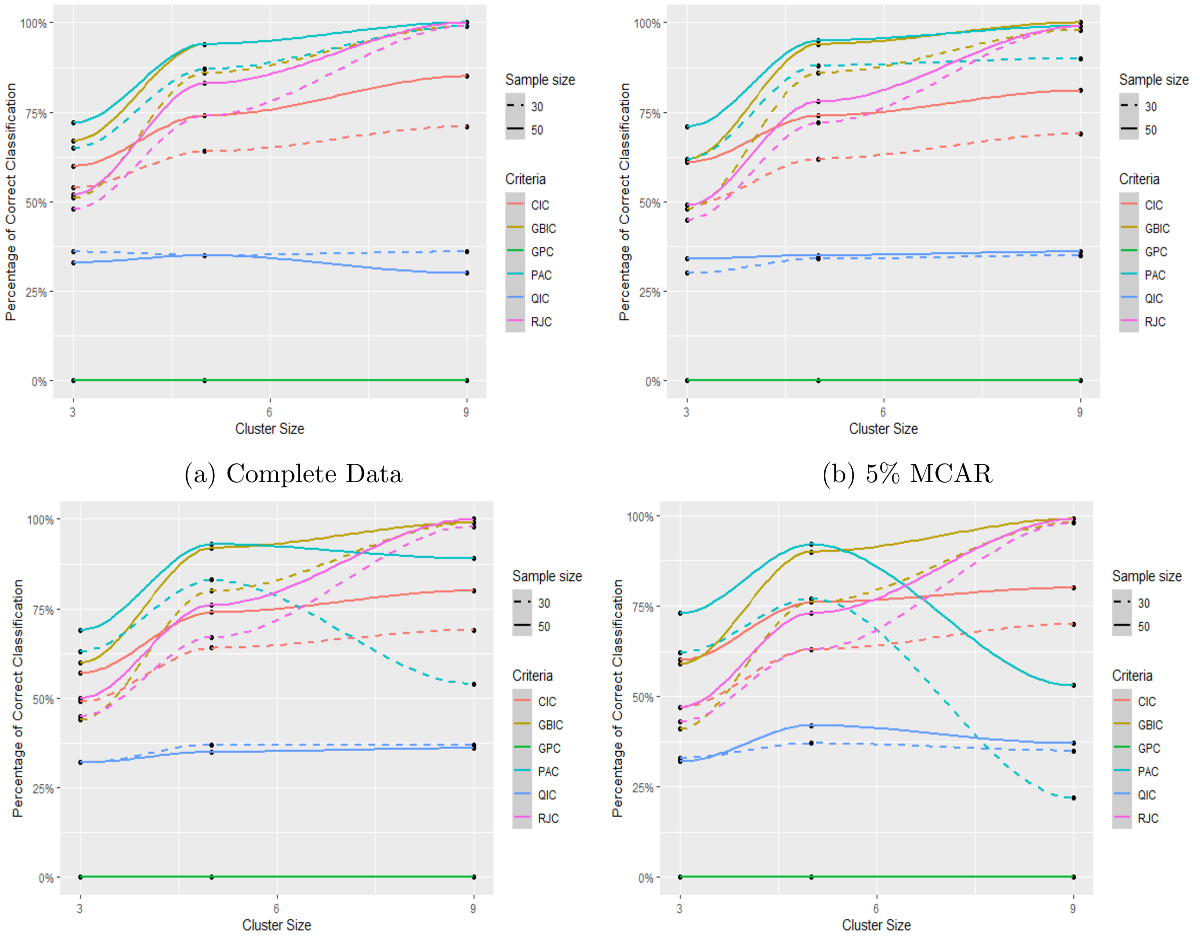
Figure 1: Performance of Selection Criteria to Classify $R_i = IND$ with $\alpha = 0.3$

percentage of correctly classified underlying correlation structures, emphasizing those that are most effective in identifying the true correlation structure, even in the presence of missing data.

Figure 1 illustrates the performances of the different selection criteria to capture an IND underlying correlation structure with a weak correlation, $\alpha = 0.3$. Regardless of cluster size and missingness structure (0%, 5%, 10% and 15%), the GP criterion performed best among other selection criteria and is followed by GBIC, for sample sizes $n \in \{30, 50\}$. Both GP and GBIC displayed similar performances regardless of the increasing percentage of missingness compared to their performances with complete data.

Figure 2 illustrates the performance of various selection criteria in capturing exchangeable correlation (EXC) with a correlation value of $\alpha = 0.3$. For a sample size of $n = 30$ in the complete dataset scenario, the PAC criterion demonstrated the best performance in identifying the underlying correlation structure for cluster sizes of $c = 3$ and $c = 5$ (see Figure 2(a)). CIC followed closely for $c = 3$. When the cluster size increased to $c = 5$, the performance of GBIC improved significantly, with its accuracy in correctly classifying EXC nearly matching that of PAC. This indicates that, while GBIC shows weaker performance for smaller cluster sizes, it becomes more effective for moderate cluster sizes. Additionally, for $n = 30$ and $c = 9$, the performances of PAC, GBIC, and RJC showed substantial improvement, with their accuracy in correctly classifying EXC approaching nearly 100%.

http://doi.org/10.62071/tmjm.v7i1.721



(a) Complete Data

(b) 5% MCAR

(c) 10% MCAR

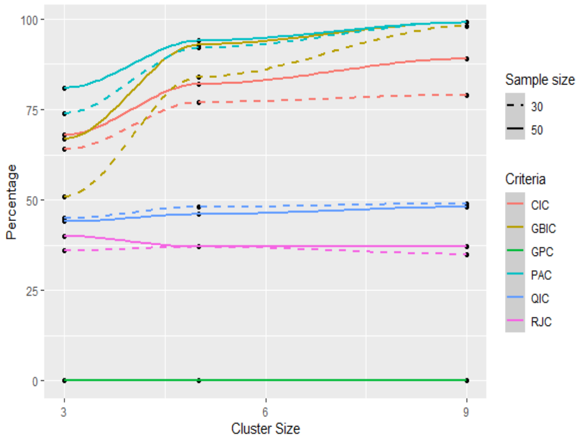
(d) 15% MCAR

Figure 2: Performance of Seelection Criteria to Classify $R_i = EXC$ with $\alpha = 0.3$

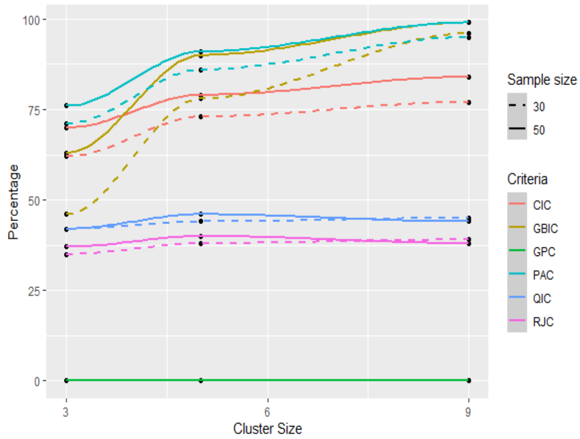
For a sample size of $n = 50$ and cluster sizes of $c = 3$ and $c = 5$, PAC outperformed the other selection criteria, with GBIC following closely. Furthermore, for a cluster size of $c = 9$, the performance of RJC was comparable to that of GBIC and PAC, achieving nearly 100% accuracy in correctly classifying the underlying correlation structure.

In the presence of missing observations, for sample sizes of $n = 30$ and $n = 50$ with a small cluster size of $c = 3$, PAC exhibited minimal sensitivity to the potential effects of missingness on its performance. At a cluster size of $c = 5$, both PAC and GBIC maintained strong performance compared to other criteria, although their accuracy was somewhat affected by 10% and 15% MCAR observations. For a larger cluster size of $c = 9$, GBIC and RJC displayed consistent and robust performance, regardless of the percentage of missing data. This indicates that GBIC and RJC are more effective for larger cluster sizes and are robust to increasing levels of missingness. While PAC is often considered one of the most reliable criteria for complete data, its performance becomes unstable as the percentage of missingness increases, particularly in larger cluster sizes.

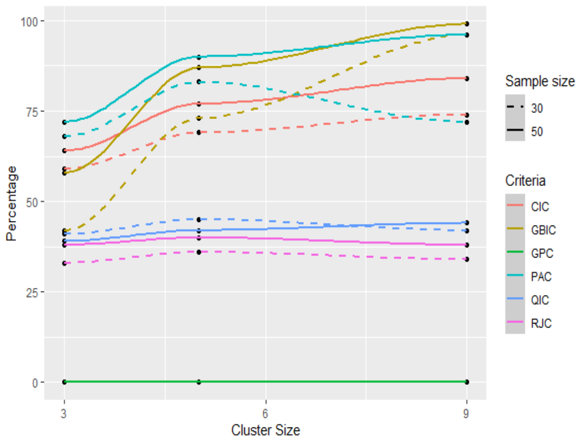
Figure 3 illustrates the performance of various selection criteria for an AR(1) underlying correlation structure with a weak correlation magnitude ($\alpha = 0.3$). Specifically, in the case of complete data (see Figure 3(a)), for both sample sizes ($n = 30$ and $n = 50$) with a small cluster size ($c = 3$), PAC outperformed all other criteria, followed by CIC. For larger cluster sizes



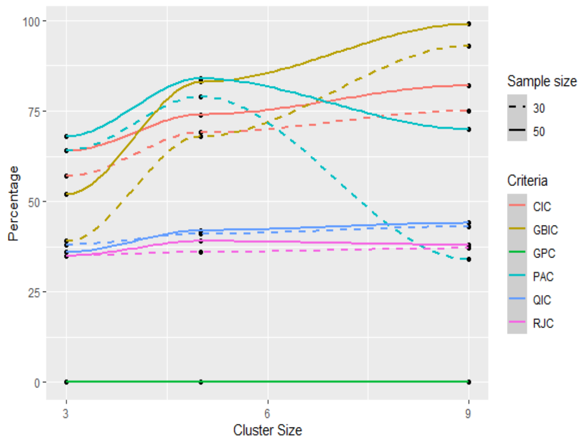
(a) Complete Data



(b) 5% MCAR



(c) 10% MCAR



(d) 15% MCAR

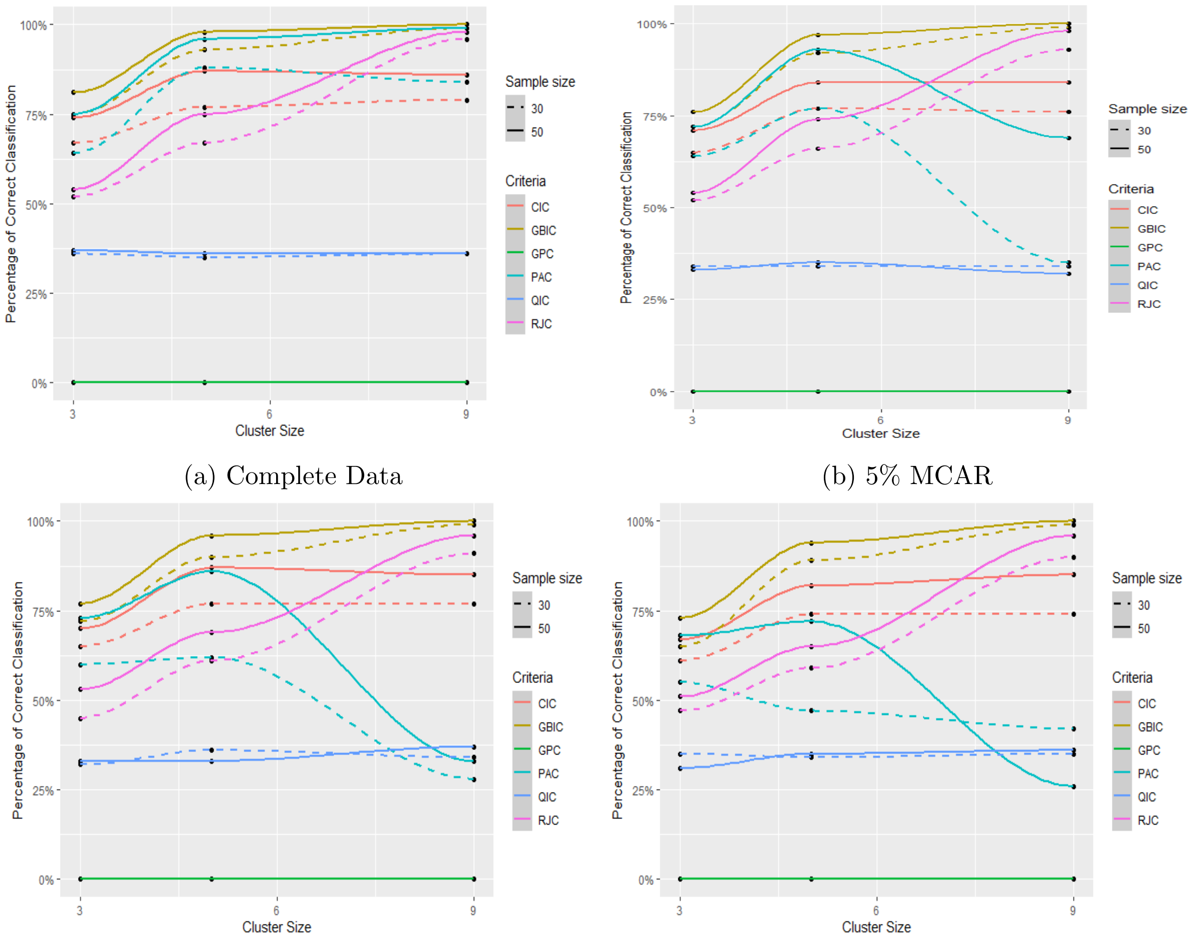
Figure 3: Performance of Selection Criteria to Classify $R_i = AR(1)$ with $\alpha = 0.3$

($c = 5$ and $c = 9$), PAC remained the best-performing criterion, with GBIC ranking second. Notably, GBIC exhibited weak performance for a small cluster size ($c = 3$) with $n = 30$, but its performance improved significantly for a moderate cluster size ($c = 5$). While CIC performed well for a cluster size of $c = 3$, it was outperformed by other criteria for larger cluster sizes ($c = 5$ and $c = 9$).

However, in the presence of missing observations, the performance of the selection criteria for a cluster size of $c = 3$ across both sample sizes was largely consistent with their performance on complete data, with only minimal declines due to missingness. For a cluster size of $c = 5$, both GBIC and PAC showed a gradual decline in performance as the percentage of missing data increased, becoming most noticeable with 15% missing observations. In contrast, for a cluster size of $c = 9$, GBIC demonstrated excellent performance, particularly for $n = 50$, regardless of the percentage of missing data. This suggests that GBIC is more robust for larger sample sizes ($n = 50$) and larger cluster sizes ($c = 9$), even with increasing levels of missingness. However, PAC exhibited unstable performance for $c = 9$ as the percentage of missingness grew. CIC, on the other hand, remained minimally affected by missing data, making it a more reliable criterion than PAC when 15% of observations are missing.

Similarly, for a sample size of $n = 50$, GBIC demonstrated the best performance for a cluster size of $c = 3$, followed by PAC and CIC. For $c = 5$, GBIC maintained its top performance,

http://doi.org/10.62071/tmjm.v7i1.721



(a) Complete Data

(b) 5% MCAR

(c) 10% MCAR

(d) 15% MCAR

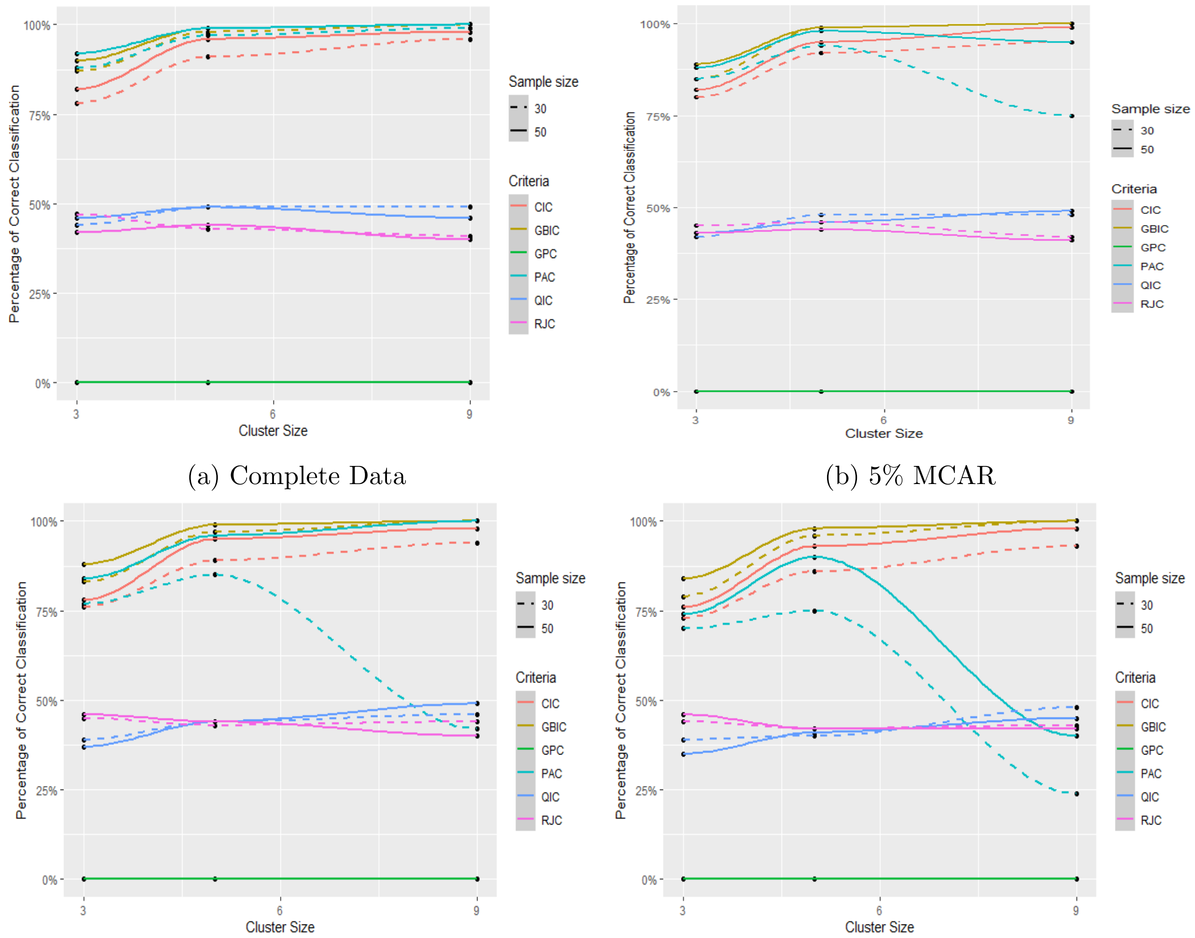
Figure 4: Performance of Selection Criteria to Classify $R_i = EXC$ with $\alpha = 0.7$

followed by PAC. At $c = 9$, GBIC remained the most effective criterion, followed by PAC, with RJC ranking third. This consistent performance shows that GBIC as the most reliable criterion across varying sample sizes and cluster sizes for a strong correlation magnitude.

4.1.2 With strong correlation $\alpha = 0.7$

Figure 4 illustrates the performance of selection criteria for an EXC underlying correlation structure with a strong correlation magnitude ($\alpha = 0.7$). As shown in Figure 4(a), for a sample size of $n = 30$ and a cluster size of $c = 3$, GBIC outperformed all other criteria, followed by CIC for complete dataset situation. For a cluster size of $c = 5$, GBIC continued to show the highest rate of correctly classified EXC, followed by PAC. For a larger cluster size of $c = 9$, GBIC again achieved the best performance, with RJC ranking second.

In the presence of missing observations, GBIC, RJC, and CIC showed minimal sensitivity to the effects of missingness. Furthermore, these criteria performed better with a strong magnitude of correlation compared to their performance with a weak magnitude of correlation. This suggests that these criteria are more effective at capturing an exchangeable correlation structure when the dependency between observations within clusters is strong, and they remain robust to the influence of missing data. In contrast, PAC exhibited unstable performance as the percentage of missingness increased.



(a) Complete Data (b) 5% MCAR

(c) 10% MCAR (d) 15% MCAR

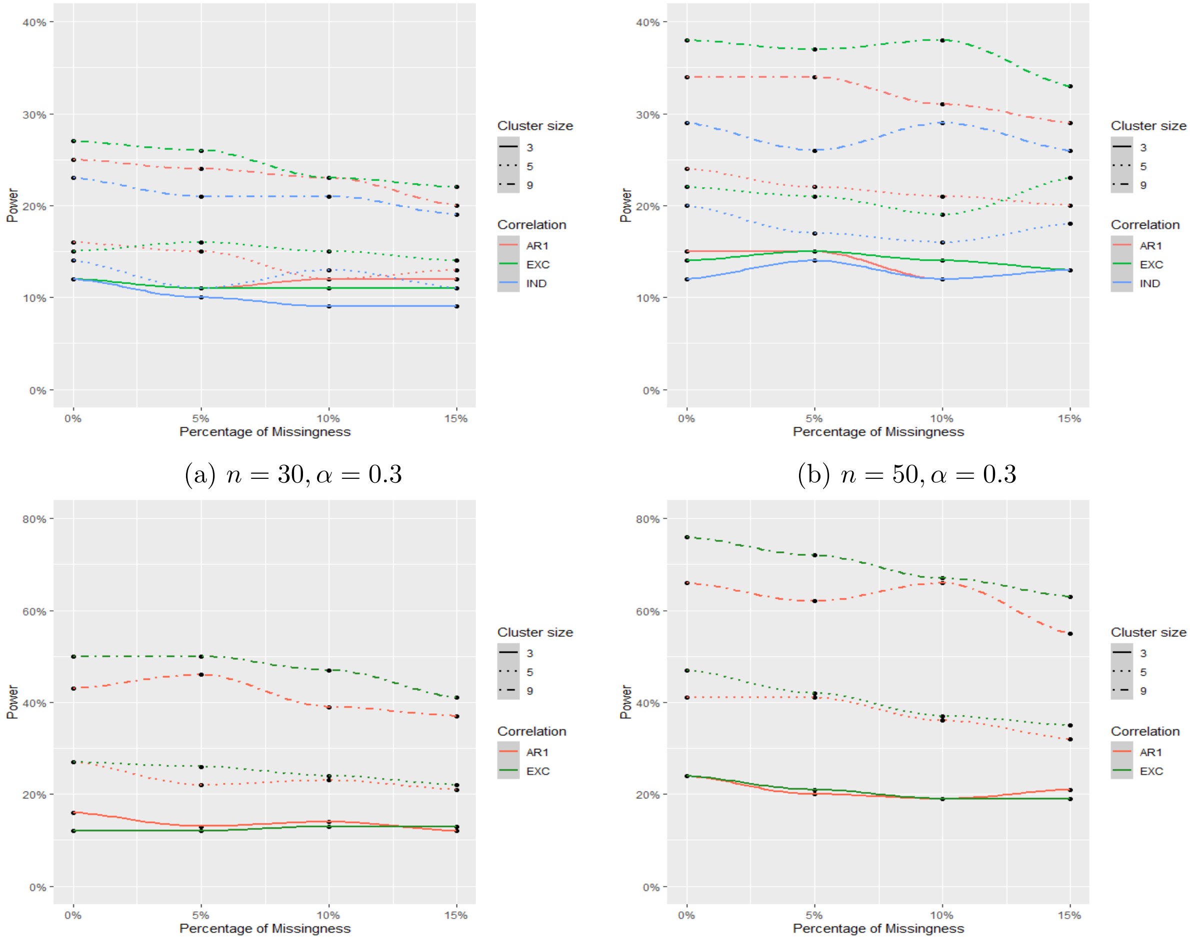
Figure 5: Performance of Selection Criteria to Classify $R_i = AR(1)$ with $\alpha = 0.7$

Figure 5 illustrates the performance of selection criteria for an AR(1) underlying correlation structure with a strong magnitude of correlation ($\alpha = 0.7$). As shown in Figure 5(a), for a sample size of $n = 30$, both PAC and GBIC performed best across all cluster sizes, with minimal differences in their performance. CIC also performed well but was slightly outperformed by PAC and GBIC. The same trend was observed for a sample size of $n = 50$, with PAC and GBIC continuing to perform better. Furthermore, their performance improved as the cluster size increased, demonstrating a more pronounced improvement in a larger sample size.

In the presence of missing observations, the performance of the criteria exhibited similar trends to those observed with complete data, but with some decline due to missingness. For a cluster size of $c = 3$, regardless of the sample size, all criteria—PAC, GBIC, and CIC—showed a gradual decrease in performance as the percentage of missingness increased. The decline in PAC's performance was most noticeable as the missingness percentage rose from 5% to 15%, with this trend becoming more pronounced for a cluster size of $c = 9$ across both sample sizes. In contrast, GBIC and CIC demonstrated minimal sensitivity to the increasing percentage of missing data, making them reliable criteria for capturing an AR(1) correlation structure with a strong magnitude of correlation, even when the missingness percentage increased.

Figure 6 illustrates the impact of missing observations on the power of GEE for parameter β_1. Overall, a decreasing trend is observed: as the percentage of missingness increased from 5%

http://doi.org/10.62071/tmjm.v7i1.721



(a) $n = 30, \alpha = 0.3$ (b) $n = 50, \alpha = 0.3$

(c) $n = 30, \alpha = 0.7$ (d) $n = 50, \alpha = 0.7$

Figure 6: Power of GEE with MCAR Observations

to 15%, the power of GEE correspondingly decreased. This decline in power results from the reduction in available observations, leading to fewer data points and, consequently, a decreased power of GEE.

Additionally, it is clear that regardless of the correlation magnitude, a sample size of $n = 50$ with a cluster size of $c = 9$ achieved the highest power compared to $n = 30$, even with increased missingness. Furthermore, a stronger correlation magnitude resulted in a higher power in the presence of missing data compared to a weaker correlation.

Thus, it is preferable to have a larger sample size and more repeated measurements from subjects, with minimal or no missing observations, to reduce the likelihood of committing a Type II error when drawing inferences from samples using GEE methodology.

A similar simulation procedure was conducted in this study to evaluate the performance of selection criteria under MNAR and MAR mechanisms. The results indicated that, in the presence of MNAR observations, the performance of the selection criteria was comparable to that observed under the MCAR mechanism. However, a steeper decline in power was noted as the percentage of missingness increased to 15%.

For selection criteria performance under MAR observations, the results showed that complete case deletion compromises the robustness of the criteria, including power and its estimates. Therefore, complete deletion of missing observations is not recommended.

4.2 Empirical Study

The dataset used in this study was Shoulder Pain after Laparoscopic Cholecystectomy which can be retrieved in the *glmtoolbox* package of R Software. Measurements were taken from patients who had undergone laparoscopic cholecystectomy and received either abdominal suction or without abdominal suction that removes residual gas which causes shoulder pain. Patients were asked to rate the level of their shoulder pain after the laparoscopic surgery every morning and afternoon for three consecutive days. 41 patients were repeatedly measured six times which makes up a total of 246 observations.

Furthermore, the dataset consists of 7 variables: (1) ID, the identifier of the respondents; (2) Age, a numerical variable; (3) Treatment, a categorical variable where a value of "A" represents abdominal suction and a value of "P" represents placebo; (4) The time when the patient was asked to rate his or her shoulder pain, it consists of integers 1 to 6 where a value of 1 indicates that the measurement was taken in the morning of the first day, 2 where the measurement was taken in the afternoon of the first day and so on; (5) Gender which is a categorical variable where M represents male while F represents Female; (6) Pain, an ordinal variable, pertains to the level of pain rated by the patients from 1 to 5. with 1 indicating a low level of pain and 5 indicating a high level of pain and; (7) Pain2, a categorical variable and outcome variable, consists of values 0 and 1. A value of 1 indicates that the patient has either felt a level 1 or 2 kind of pain, as described by the pain variable, while a value of 0 indicates a pain level of 3, 4, or 5.

An underlying AR(1) correlation structure is assumed in generating the empirical results, based on the data exploration conducted. The result on the correlation structure identification of selection criteria are shown in Table 1.

Table 1 shows that CIC, QIC, GP, PAC, and GBIC gave a coherent result with respect to their specified correlation structure regardless of the degree of missingness while RJ criterion has a discrepancy on its specified correlation structure when the percentage of missingness is at 15%.

Furthermore, Table 2 shows that the specified correlation structure of the selection criteria differs as the percentage of missingness increases. Only the GBIC consistently identified the true underlying correlation structure, even when up to 5% of the observations were missing.

Finally, for the MAR mechanism, a significant difference was observed between the parameter estimates from complete and incomplete data. As a result, the generation of empirical data results was not conducted, since biased estimates and unstable performance of the selection criteria would likely be contributed by the given data.

Table 1: Classification of Correlation Structure with MCAR Observations

Selection Criteria	**Correlation Classification**	**5% Missing**	**10% Missing**	**15% Missing**
CIC	IND	IND	IND	IND
QIC	IND	IND	IND	IND
GP	IND	IND	IND	IND
PAC	AR(1)	AR(1)	AR(1)	AR(1)
GBIC	AR(1)	AR(1)	AR(1)	AR(1)
RJ	AR(1)	AR(1)	AR(1)	EXC

Table 2: Classification of Correlation Structure with MNAR Observations

Selection Criteria	**Correlation Classification**	**5% Missing**	**10% Missing**	**15% Missing**
CIC	IND	IND	AR(1)	AR(1)
QIC	IND	IND	AR(1)	AR(1)
GP	IND	IND	EXC	EXC
PAC	AR(1)	AR(1)/EXC	IND	IND
GBIC	AR(1)	AR(1)	IND	IND
RJ	AR(1)	IND	AR(1)	AR(1)

5 Concluding Remarks

Based on the study, this paper concludes that in the complete data scenario and regardless of the degree of correlation, the performance of selection criteria varies across different correlation structures: (1) For the independent (IND) correlation structure, GP and GBIC performed consistently well across all cluster sizes. (2) Under the exchangeable (EXC) correlation structure, PAC demonstrated superior performance regardless of cluster size. Additionally, CIC performed well for small cluster sizes ($c = 3$), while GBIC excelled for larger cluster sizes ($c = 6$ and $c = 9$). (3) For the AR(1) correlation structure, PAC again showed superior performance across all cluster sizes. Similar to the EXC structure, CIC performed well at $c = 3$, while GBIC performed best at $c = 6$ and $c = 9$.

It is noteworthy that missing observations significantly impact the performance of selection criteria, with GBIC demonstrating notable robustness against the effects of missing data. PAC, while exhibiting strong classification performance for small cluster sizes ($c = 3$), becomes increasingly unstable as cluster sizes grow and as the percentage of missingness rises. Conversely, GBIC shows weaker performance at $c = 3$ but improves substantially for larger cluster sizes ($c = 5$ and $c = 9$). CIC remains a reliable criterion for small cluster sizes, particularly at $c = 3$, though GBIC and PAC outperform it at larger cluster sizes ($c = 5$ and $c = 9$). Additionally, the RJ criterion aligns closely with exchangeable structures and achieves its best performance when the cluster size is $c = 9$.

As anticipated with any statistical methodology, GEE demonstrated a gradual decline in statistical power as the percentage of missingness increased. To address this issue and prevent a significant power reduction, it is recommended that the allowable percentage of missing observations be limited to no more than 10%.

Additionally, further research is encouraged to explore the data imputation methods in longitudinal analysis. Such studies could evaluate the performance of selection criteria in handling datasets with imputed missing observations, providing insights into their robustness and accuracy under various imputation strategies. Moreover, further exploration on the performances of selection criteria in a non-binary response is highly encouraged. Such studies are valuable in providing insights on the flexibility of the selection criteria to handle other types of responses.

Acknowledgements

The authors of this paper would like to acknowledge the Department of Science and Technology-Science Education Institute for the MS Statistics scholarship grant through the Accelerated Science and Technology Human Resource Development Program.

References

[1] C. Berkey, et.al., *Longitudinal Study of Skipping Breakfast and Weight Change in Adolescents*, International Journal of Obesity **27** (2003), no. 10, 1258–1266, doi:10.1038/sj.ijo.0802402.

[2] V. Carey and Y. Wang, *Working Covariance Model Selection for Generalized Estimating Equations*, Statistics in Medicine **30** (2011), no. 26, 3117–3124, doi:10.1002/sim.4300.

[3] L. Fu, et.al, *Working correlation structure selection in generalized estimating equations*, Computational Statistics **33** (2018), 983–996, doi:10.1007/s00180-018-0800-4.

[4] K. Gustavson, et.al., *Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study*, BMC Public Health **12** (2012), no. 918, 1–11, doi:10.1186/1471-2458-12-918.

[5] D. Hall and T. Severini, *Extended generalized estimating equations for clustered data*, Journal of the American Statistical Association **93** (1998) no. 444, 1365–375, doi:10.2307/2670052.

[6] L. Hin, et.al., *Criteria for working–correlation–structure selection in GEE: Assessment via simulation*, The American Statistician **61** (2007) no. 4, 360–364, doi:10.1198/000313007X245122.

[7] L. Hin and Y. Wang, *Working-correlation-structure identification in generalized estimating equations*, Statistics in Medicine **28** (2009), no. 4, 642–658, doi:10.1002/sim.3489.

[8] K. Liang and S. Zeger, *Longitudinal data analysis using generalized linear models*, Biometrika **73** (1986), no 1, 13–22, doi:10.1093/biomet/73.1.13.

[9] K. Lu and D. Mehrotra, *Specification of covariance structure in longitudinal data analysis for randomized clinical trials*, Statistics in Medicine **29** (2010), no. (4), 474–488, doi:10.1002/sim.3820.

[10] M. Mott, et.al., *Progression of early structural lung disease in young children with cystic fibrosis assessed using ct*, Thorax **67** (2012), no. 6, 509–516, doi:10.1136/thoraxjnl-2011-200912.

[11] P. Pahwa and C. Karunanayake, *Modeling of longitudinal polytomous outcome from complex survey data-application to investigate an association between mental distress and non-malignant respiratory diseases*, BMC Medical Research Methodology **9** (2009) no. 84, doi:10.1186/1471-2288-9-84.

[12] Paik, *Repeated measurement analysis for nonnormal data in small samples*, Communications in Statistics-Simulation and Computation **17** (1988),no. 4, 1155–1171, doi:10.1080/03610918808812718.

[13] W. Pan, *Akaike's information criterion in generalized estimating equations*, Biometrics **57** (2001), no. 1,120–125, doi:10.1111/j.0006-341x.2001.00120.x.

[14] M. Pardo and R. Alonso, *Working correlation structure selection in gee analysis*, Statistical Papers **60** (2019), 1447–1467, doi:10.1007/s00362-017-0881-0.

[15] A. Rotnitzky and N. Jewell, *Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data*, Biometrika **77** (1990), no. 3, 485–497, doi:10.2307/2336986.

[16] G. Schwarz, *Estimating the dimension of a model*, The Annals of Statistics **6** (1978), no. 2,461–464, doi:10.1214/aos/1176344136.

[17] C. Sotto, *Longitudinal data analysis and missing data[reading materials]*, University of the Philippines Cebu, 2021, fb.watch/n6HTp9oNwS/.

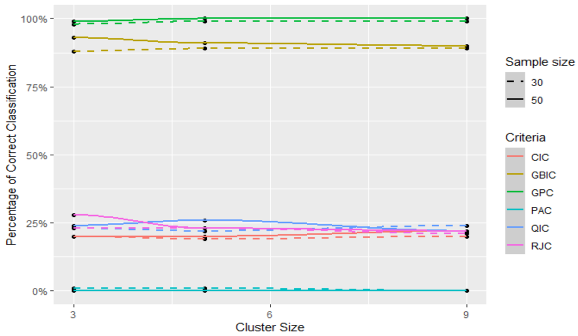
[18] M. Wang, *Generalized estimating equations in longitudinal data analysis: a review and recent developments*, Advances in Statistics **2014** (2014), no. 1, 1–11, doi:10.1155/2014/303728.

[19] M. Wang, et.al., *Covariance estimators for generalized estimating equations (GEE) in longitudinal analysis with small samples*, Statistics in Medicine **35** (2016), no. 10, 1706–1721, doi:10.1002/sim.6817.

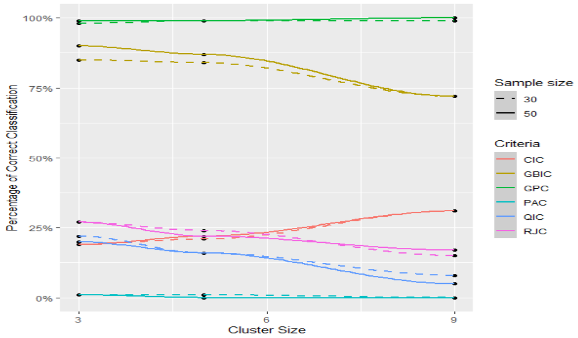
[20] Z. Xiaolu and Z. Zhongyi, *Comparison of criteria to select working correlation matrix in generalized estimating equations*, Chinese Journal of Applied Probability and Statistics **29** (2013), no. 5, 515–530.

6 Appendix

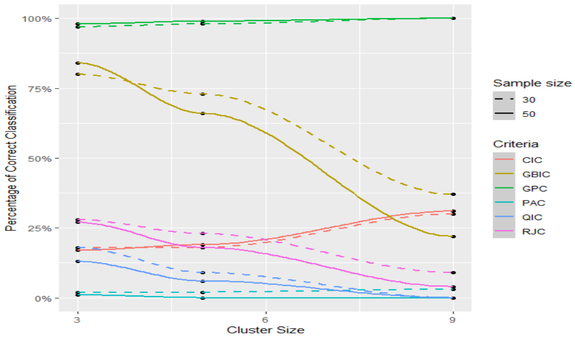
The following figures show the results of the simulation studies conducted for the MAR mechanism, considering various correlation structures and degrees of missingness in the dataset. A similar pattern is observed in the MNAR scenario, and as a result, it is not included further. Additionally, the power graph of GEE with MNAR observations is presented.



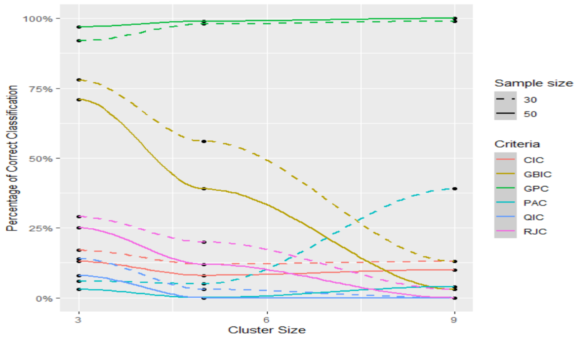
(a) Complete Data



(b) with 5% MAR observations

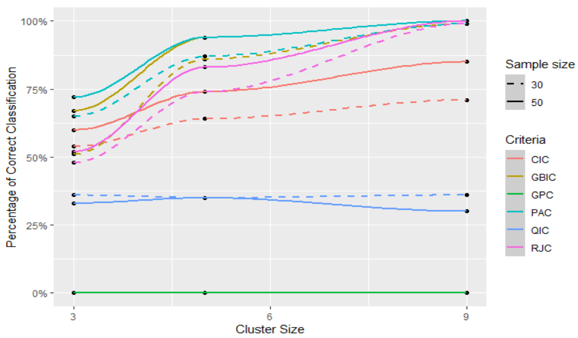


(c) with 10% MAR observations

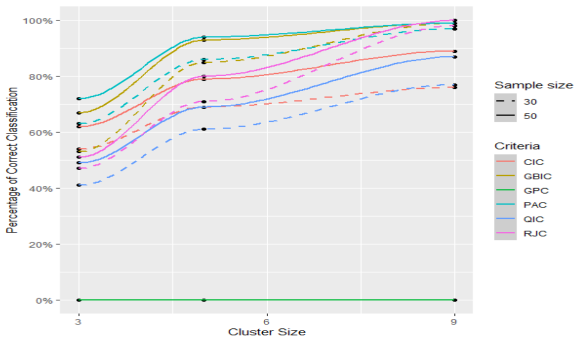


(d) with 15% MAR observations

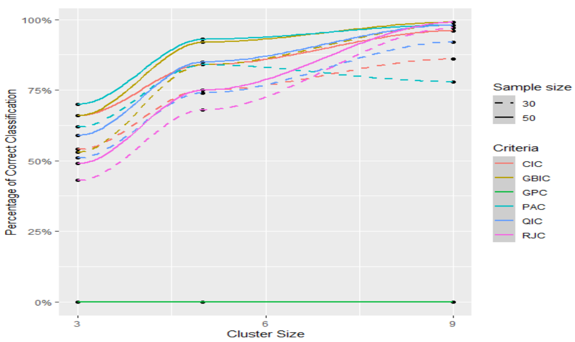
Figure 7: Performance of Selection Criteria to Classify $R_i = IND$ with $\alpha = 0.3$



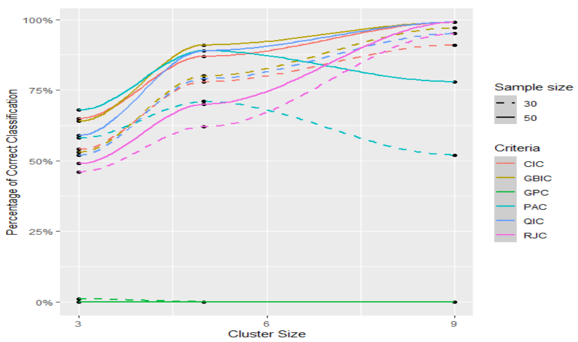
(a) Complete Data



(b) with 5% MAR observations



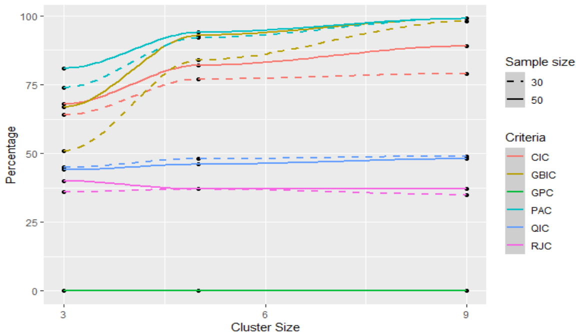
(c) with 10% MAR observations



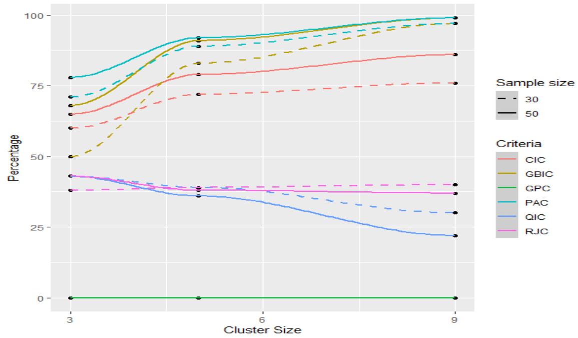
(d) with 15% MAR observations

Figure 8: Performance of Selection Criteria to Classify $R_i = EXC$ with $\alpha = 0.3$

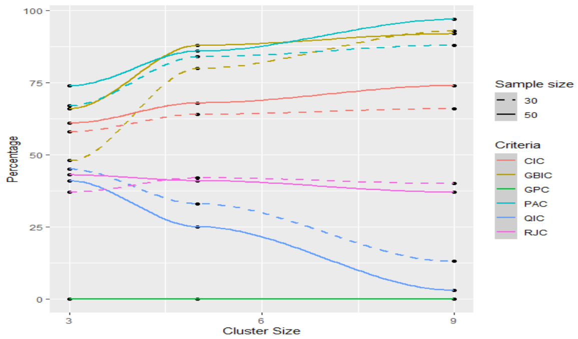
http://doi.org/10.62071/tmjm.v7i1.721



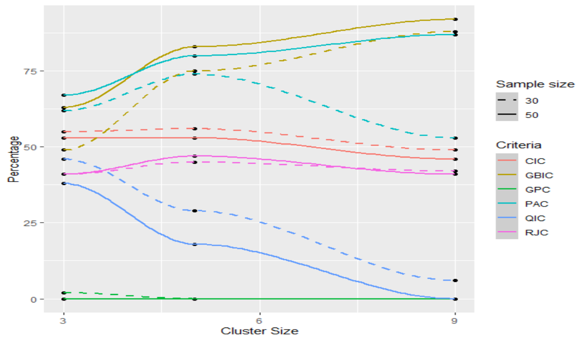
(a) Complete Data



(b) with 5% MAR observations

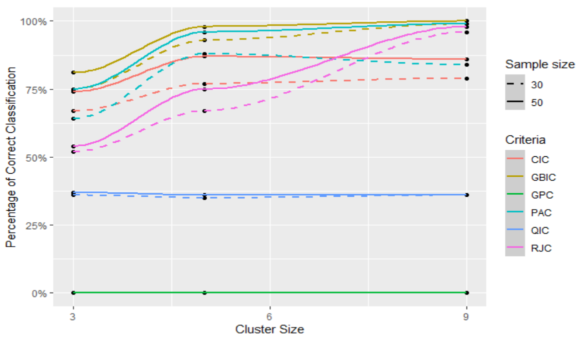


(c) with 10% MAR observations

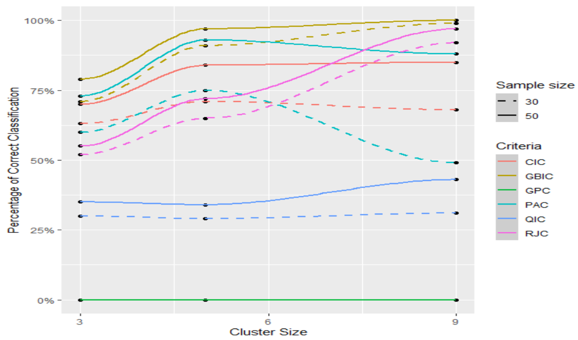


(d) with 15% MAR observations

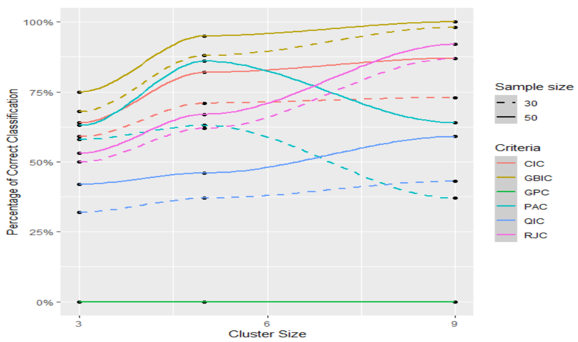
Figure 9: Performance of Selection Criteria to Classify $R_i = AR(1)$ with $\alpha = 0.3$



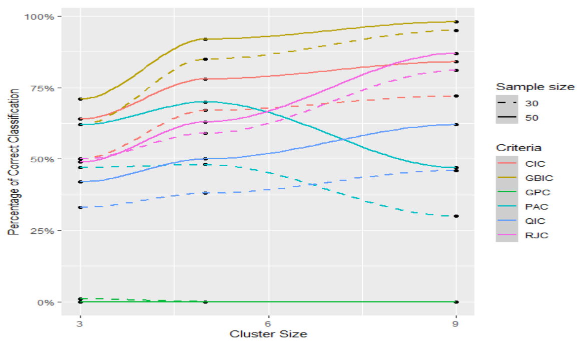
(a) Complete Data



(b) with 5% MAR observations

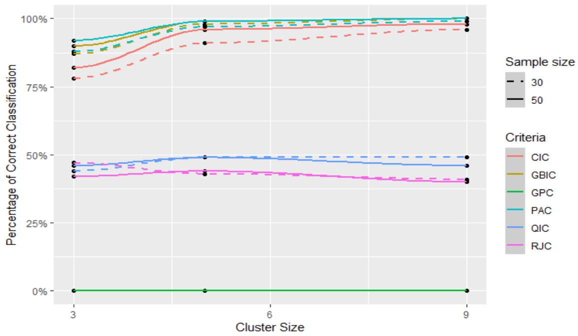


(c) with 10% MAR observations

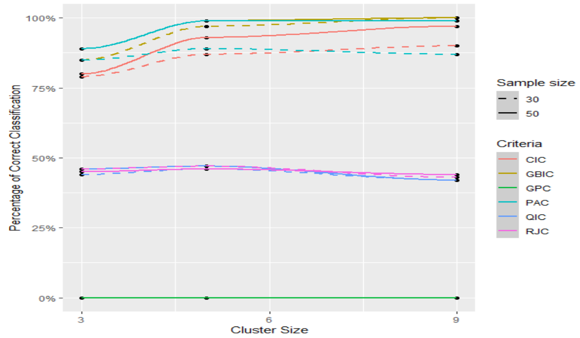


(d) with 15% MAR observations

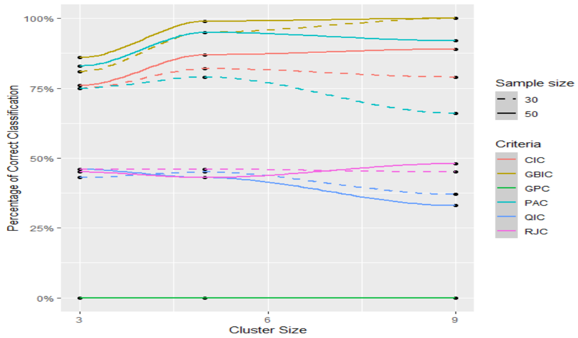
Figure 10: Performance of Selection Criteria to Classify $R_i = EXC$ with $\alpha = 0.7$



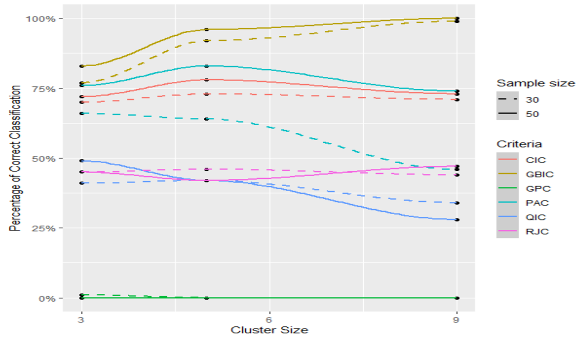
(a) Complete Data



(b) with 5% MAR observations

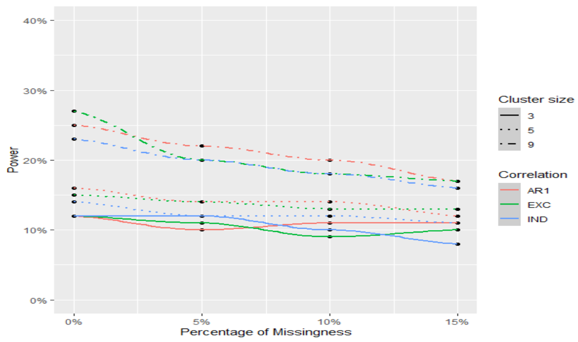


(c) with 10% MAR observations

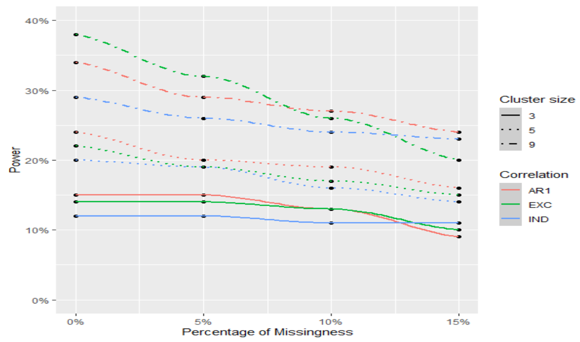


(d) with 15% MAR observations

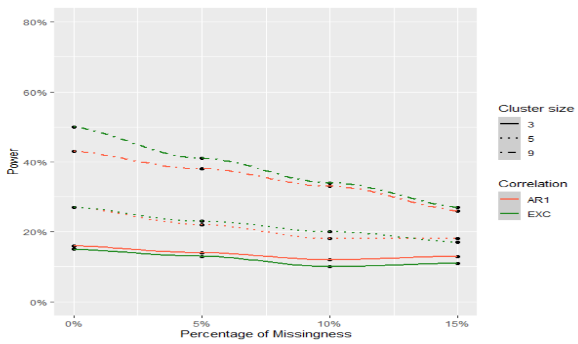
Figure 11: Performance of Selection Criteria to Classify $R_i = AR(1)$ with $\alpha = 0.7$



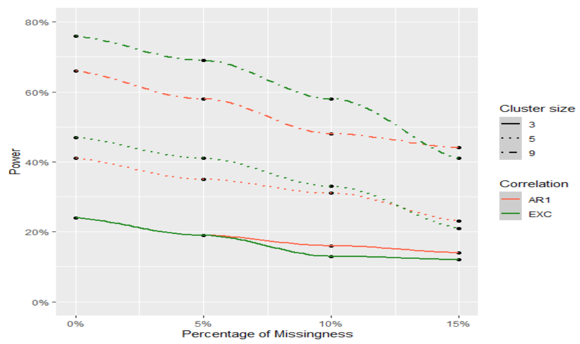
(a) $n = 30, \alpha = 0.3$



(b) $n = 50, \alpha = 0.3$

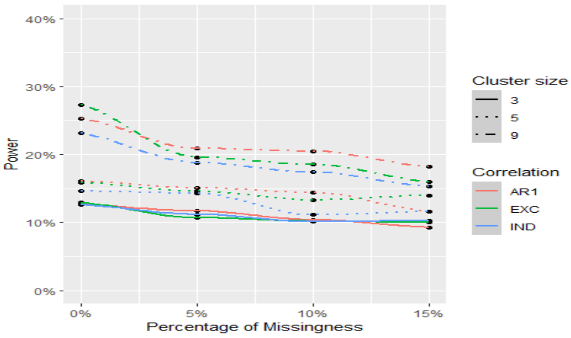


(c) $n = 30, \alpha = 0.7$

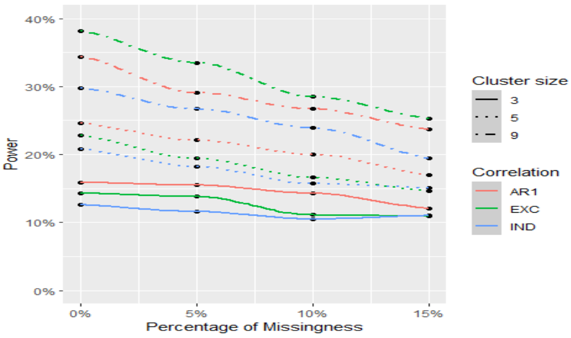


(d) $n = 50, \alpha = 0.7$

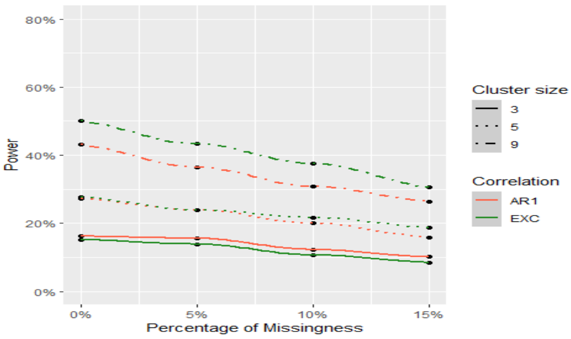
Figure 12: Power of GEE with MAR observations



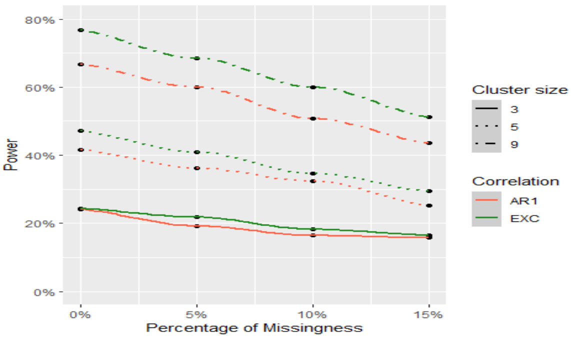
(a) $n = 30, \alpha = 0.3$



(b) $n = 50, \alpha = 0.3$



(c) $n = 30, \alpha = 0.7$



(d) $n = 50, \alpha = 0.7$

Figure 13: Power of GEE with MNAR observations